Application of Multivariate Volatility Models in Estimating Optimal Dynamic Hedge Ratios for Crack Spreads with Volatility Spillovers in Energy Markets

JANE AKINYI ADUDA

Reg No. I80/95667/2014

December 2017

Doctoral thesis submitted for the degree of
Doctor of Philosophy in Financial Mathematics
in the School of Mathematics,
University of Nairobi.

SCHOOL OF MATHEMATICS

UNIVERSITY OF NAIROBI

Declaration

This thesis is my own work which has not been submitted in any other institutions.

Signed: ..... Dated: .....

Jane Akinyi Aduda

Supervisor's statements

We declare that this thesis has been submitted with our approval as supervisors.

Signed: ..... Dated: .....

Prof. Patrick Weke
School of Mathematics,
University of Nairobi.

Signed: ..... Dated: .....

Dr. J.I. Mwaniki
School of Mathematics,
University of Nairobi.

Signed: ..... Dated: .....

Dr. Philip Ngare
School of Mathematics,
University of Nairobi.

Acknowledgement

Foremost, I would like to thank God for his infinite mercy, without which, I would not amount to much.

My most sincere gratitude goes to my supervisors Prof. Patrick Weke, Dr. Joseph Mwaniki and Dr. Philip Ngare for their continuous support of my Ph.D study and research, for their patience, motivation, enthusiasm, immense knowledge, their encouragement, insightful comments, and hard questions. Their guidance helped me during my research and when writing this thesis. I could not have imagined having better mentors.

I am grateful for the funding sources that allowed me to pursue my studies. My appreciation goes to Jomo Kenyatta University of Agriculture and technology (JKUAT) for providing research grants. My sincere thanks also goes to Lappeenranta University of Technology, Finland, who, through the North-South -South Mobility programme, awarded me a research exchange opportunity that exposed me to diverse and exciting ideas.

I would also like to appreciate in a special way my colleagues in JKUAT, and especially in the department of Statistics and Actuarial Sciences, and my friends George, Oscar, victor and Edward for their undying support and encouragement.

Last but not the least, I would like to thank my parents Michael Otieno Aduda and Mary Anyango Otieno for supporting me spiritually throughout my life. Many thanks also go to my two wonderful children Fred Michael and Mikaela for their perseverance when I had to be away, and all my siblings who always stand in the gap for me. The support of family provided the much needed peace of mind when I had to study and work on my research.

ABSTRACT

The effectiveness of hedging is examined using crack spreads on crude oil and distillate fuels traded at the New York Mercantile Exchange (NYMEX). Using both static and dynamic hedge ratios, and allowing for volatility transmission and cross market linkages, it is found that dynamic cross-market hedges are more effective. We use multivariate volatility models to account for spillover between the markets. We consider the statistical properties of energy futures and spot prices and investigate the trends that underlie the price dynamics in order to gain further insights into possible nuances of price discovery and energy market dynamics. The family of autoregressive moving average (ARMA)-generalised autoregressive conditional heteroscedastic (GARCH) models are explored. The trends depict time varying variability and persistence of oil price shocks. The return series conform to a constant mean model with GARCH variance. We consider the joint behaviour of all the series using co-integration to analyse the long-term equilibrium relationship between the commodities. Analysis using the Engle-Granger (E-G) and Johansen co-integration test reveal co-integrating relations and a vector error correction model (VECM) is used to capture the short term dynamics of energy prices. Finally, multivariate GARCH (MGARCH) models are used to capture the source and magnitude of volatility spill overs. BEKK, constant conditional correlation (CCC) and dynamic conditional correlation (DCC)-GARCH models are used to capture the volatility transmission between the series and estimate the parameters using maximum likelihood estimation (MLE). The hedge ratios are estimated under a mean variance and utility maximization framework.

TABLE OF CONTENTS

ABBREVIATIONS

ACF autocorrelation function.

ADF augmented Dickey-Fuller.

AIC Akaike's Information Criterion.

AR autoregressive.

ARCH autoregressive conditional heteroscedastic.

ARIMA autoregressive integrated moving average.

ARMA autoregressive moving average.

BIC Bayesian Information Criterion.

CCC constant conditional correlation.

CDF cumulative distribution function.

CNLRM classical normal linear regression model.

CVAR co-integrated vector autoregressive.

D-W Durbin-Watson.

DCC dynamic conditional correlation.

DF Dickey-Fuller.

DGP data generating process.

E-G Engle-Granger.

ECM error correction model.

EIA Energy Information Administration.

FDCC Flexible Dynamic Conditional Correlation.

GARCH generalised autoregressive conditional heteroscedastic.

GSV generalized semi-variance.

i.i.d independent and identically distributed.

IGARCH integrated GARCH.

JB Jarque-Bera.

KS Kolmogorov-Smirnov.

LB Ljung-Box.

LM Lagrange multiplier.

LS least squares.

MA moving average.

MAE mean absolute error.

MARMA multivariate autoregressive moving average.

MGARCH multivariate GARCH.

ML maximum likelihood.

MLE maximum likelihood estimation.

MSE mean square error.

NYMEX New York Mercantile Exchange.

OK WTI Oklahoma West Texas Intermediate.

OLS ordinary least squares.

OPEC Organization of Petroleum Exporting Countries.

PACF partial autocorrelation function.

PD positive definite.

QML quasi-maximum likelihood.

RBOB Reformulated Blendstock for Oxygenate Blending.

RMSE root mean square error.

RMSFE root mean squared forecasting errors.

SSE sum of squared errors.

SST total sum of squares.

SUR seemingly unrelated regression.

VAR vector autoregressive.

VARMA vector autoregressive moving average.

VEC vector error correction.

VECM vector error correction model.

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS

[1] **Jane Aduda**, Patrick Weke, Philip Ngare and Joseph Mwaniki. 2016 "Financial time series modelling of trends and patterns in the energy markets", *Journal of Mathematical Finance*, 6(2).

[2] **Jane Aduda**, Patrick Weke and Philip Ngare. "A co-integration analysis of the interdependencies between between crude oil and distillate fuel prices ". `-Accepted for publication by Journal of Mathematical Finance`.

[3] **Jane Aduda**, Philip Ngare and Patrick Weke. "Multivariate GARCH models for estimating dynamic hedge ratios that account for volatility spillover effects in energy markets". `-under review`.

1. INTRODUCTION

Energy use is behind virtually everything we interact with daily. It dictates the activities in our lives now more than ever. The global energy market has rapidly expanded and become increasingly interdependent. In developed countries, there is heavy reliance on energy and other related products for sustainable economic growth and development as is evidenced by the increase in consumption of energy products. Similarly, for developing countries, there is heavy reliance on energy resources to steer economic growth and development.

Initially, energy was viewed just as an enabler, a utility with very limited consumer interest, but now, it has become key in the achievement of a sustainable future and affects all, from household production to professional consumption (Dorsman et al., 2013; Ramos and Veiga, 2014). Globally, the energy trading marketplace has also developed along with the growing demand for energy that is worldwide. New energy trading products are being introduced in exchanges around the world at a record pace. Most of these new products aid energy companies in offsetting various risks.

Energy firms are increasing their trading and risk management activities to ensure enhanced and uninterrupted supplies of energy products to their consumers and also protect themselves against economic shocks and political influence. Energy industry activities entail four interrelated functions, which constitute the energy cycle around which the entire industry revolves. These include exploration and production, transportation and storage, refining and processing and finally distribution and sales. These functions entail complex risks which inform mitigation strategies. The first two steps are generally referred to as upstream processes, while the last two are downstream (GARP, 2009).

Energy markets globally have undergone rapid deregulation, implying, more competition and increased price volatility which exposes participants to potentially much greater risks. Before, regulators, who were mainly governments, set prices. Prices were generally stable, though consumers paid highly for inefficiencies such as complex cross-subsidies from surpluses areas to shortage or inefficient technology areas (Clewlow and Strickland, 2000). Deregulation gave rise to a free market implying more competitive prices revealing increased energy price volatility across all commodities.

The impact of deregulation is felt by both consumers and producers. It has increased awareness for the need of risk management to control exposure to energy price fluctuations (Dooley, 1998; Winston, 1993). Risk in energy markets could be due to unexpected jumps in demand and supply, a reduction in production capacity, resource reserve policy, Organization of Petroleum Exporting Countries (OPEC) share capacity and policy, regional and global economic policies, environmental regulations and stringent product specifications among others (Chang et al., 2011), and other geopolitical risks. Other investors have also been drawn in as they seek new markets to operate. This combination of producers and consumers, coupled with the sheer size of the market contributes to the unparalleled growth experienced in energy derivatives, which is considered by most market participants as a new phenomenon.

Crude oil is a fossil fuel composed of hydrocarbons formed from animals and plants that eons ago. It is found in underground pools or reservoirs, in tiny spaces within sedimentary rocks,in liquid form , and close to the surface in tar (or oil) sands. Petroleum products are made from crude oil and other hydrocarbons contained in natural gas. After mining, it is sent to a refinery and separated into usable petroleum products such as gasoline and other distillates such as heating oil, diesel fuel, jet fuel, waxes, asphalt, lubricating oils and petrochemical feed-stocks.

The oil refining process cracks crude oil into its constituent products. The refiner's profits are directly linked to the "crack spread" (the difference between the revenue from mainly gasoline and distillate fuels and cost of crude oil). The term "crack" refers to the process a refiner uses to make money by breaking up the long hydrocarbon chains that make up crude oil into petroleum products which are made of shorter chains. As a result, refiners are simultaneously exposed to both sides of the market, and hence, exposed to greater risks compared to oil producers or fuel consumers. They are caught between two markets: the raw materials they must purchase and the finished products they sell. Figure 1.1 displays the products of the refining process (EIA, 2015).

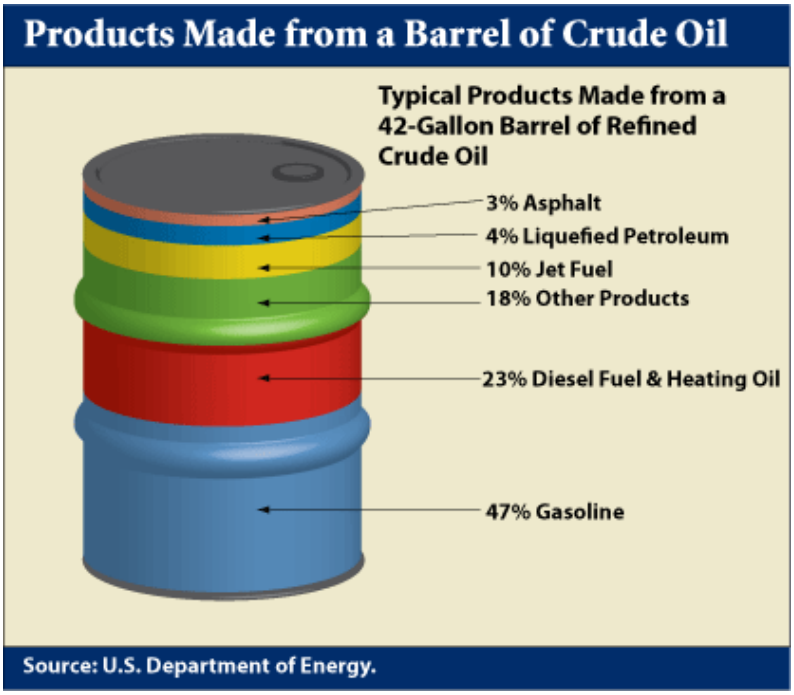

Figure 1.1: Products of the refining process

Gasoline is used mainly as an engine fuel for vehicles, with most vehicles operating on the least expensive grade, regular gasoline. The distillate fuels include diesel, used in diesel engines found in most trucks, trains, buses, boats, and farm and construction vehicles and heating oil used for heating homes.

This section gives a brief history of energy markets and their evolution; it discusses the risks faced in these markets and how they can be managed. It introduces futures and

crack spread trading; and gives a brief insight into hedging against energy market risks using crack spreads. The problem statement, objectives and significance of the study are also presented in this section.

1.1 Futures and Forwards

Over the last 30 years, pricing, expansion of global markets and significant changes in demand and supply, have affected most commodity markets and especially the energy markets. Changing economic patterns, war, international politics and other structural changes within the energy sector have introduced considerable uncertainty with respect to the future direction of market conditions (Dorsman et al., 2013). This has caused an upsurge in market volatility, hence, creating a need for an effective hedge for the risk of adverse exposure to price vulnerabilities. The main risk management tools available to participants in energy markets are derivative contracts listed in exchanges worldwide.

A futures contract is a commitment to trade a specified quantity and quality of a particular asset at a specified price to be delivered on a specified future date. Essentially, it allows the investor to “lock in" a price now so as to potentially benefit in future if prices fluctuates. The price of the asset is agreed upon at the time the commitment is made. The long party, who is the buyer, agrees to receive the underlying commodity, while the short party, who is the seller, agrees to deliver the underlying commodity. Futures are standardized, in terms of dates, quality and amounts traded, and can be re-traded during their life time on a futures exchange. Forward contracts, which are direct agreements between two parties made over the counter, are also agreements to transact on fixed terms at a future date.

Although forwards and futures all involve an agreement to trade at a certain future date for a certain price, they have significant differences. Foremost, futures are standardized meaning they are exchange traded, whereas forwards trade between individual institutions (Over-the-counter). Secondly, cash flows of the two contracts occur at different times. Forwards are settled once at maturity whereas futures are marked to market daily with

cash-flows exchanging between the long and short positions to reflect the daily price variations. However, if future interest rates are known with certainty then both forwards and futures can be treated as similar for purposes of pricing.

1.2 Hedging using futures

A financial position set to offset a significant loss (or gain) that could be incurred in a portfolio is called a hedge. Hedging is one of the most important and practical applications of futures contracts. In the event of any adverse movements in the market, hedging is a simple work around to protect a trader's trading positions from making a loss. Futures can also be used to make profit. All the terms of the contract are standardized except for the price, which is discovered through the bids (demand) and offers (supply). Ultimately, all contracts are settled either by delivery of the actual physical commodity or through liquidation by an offsetting transaction (a sale after an initial purchase or a purchase after an initial sale).

The principle of hedging in energies is based on the co-evolution of spot and futures market prices. This evolution is not necessarily identical, but usually close enough that minimising risk of a loss in the spot market by taking an opposite position in the futures market is possible. This way, losses in one market are offset by gains in the other by the hedger. In this manner, the hedger is able to establish a price level for a spot transaction that may not actually take place for several months.

A perfect hedge involves a strategy that eliminates the risks associated with a future market commitment completely. It is established by matching the holding period to the maturity date of the futures contract, and the physical characteristics of the commodity to be hedged must exactly match the underlying commodity in the futures contract, otherwise, a perfect hedge is not possible. If there exists a mismatch, risk could still be reduced but not entirely eliminated. With futures, a trader uses short and long hedges to

replace price risk with basis risk.

1.2.1 Spread trading

Spread trading involves the simultaneous purchase and sale of different financial instruments with the objective of profiting from the movement of the spread between the prices of the instruments, and not from the movement of the absolute values of the prices of the instruments (Kanamura et al., 2011). In energy markets, apart from temporal spread traders who use calendar spreads, and traders who try to hedge transportation risk exposure using locational spreads, most spread traders use at least two different physical commodities. These markets use spreads to quantify the production cost of refined products after cracking the crude oil. Crack and spark spreads are the most commonly used in the markets. Crack spreads are also known as paper refineries while spark spreads are also known as paper plants.

1.2.2 Crack spreads

The oil refining process that cracks crude oil into its constituent products. Processes in these markets are independently subject to various levels of supply, demand, transportation costs among other factors, exposing refiners to greater risks as compared to oil producers and fuel consumers, especially when crude prices increase with those of refined products remaining static or even dropping. In order to manage risks, a refiner straddles the crude oil it buys and the petroleum product it sells.

Crack spreads involve a position of purchase and sale of futures contracts on crude oil and the petroleum products simultaneously. This allows the refiner to lock in any price differentials between their input and output prices. A crack spread seller (refiner) buys crude oil contracts and sell gasoline and heating oil contracts. The refiner has to observe both sides of the market on more than one commodity namely the crude on the costs

side and the refined products on the revenue side. In crack spread hedging, the profit maximizing portfolio for the refiner is sought in terms of taking various positions (long and short) in the market.

1.3 Statement of the Problem

Over the past three decades, the oil market has grown to become the biggest commodity market in the world. It evolved from a market engaging primarily in physical activities into sophisticated financial markets with trading horizons extending over 10 years forward (Chang et al., 2010; Fleming and Ostdiek, 1999). In the process, a wide range of participants including hedge funds, investment banks, insurance companies, asset managers for mutual funds, pension funds, physical oil traders among others have been involved. In line with this expansion, oil price volatility and risk mitigation needs of industry participants triggered the development of derivative contracts (futures, forwards, swaps and options) which now dominate market activities.

Oil is physically traded in two levels; first, as feedstock for the refinery , and, secondly, as a finished product. Even though crude oil and refined petroleum products have different characteristics, they are inextricably linked by the economics and technology of the refining process. Despite the fact that product prices may, and do, fluctuate, they must be related to crude oil prices because refineries can not continue to operate on negative margins while competition on the other hand will also drive down high margins.

Using historical data from energy markets, this study looks at strategies used to estimate optimal dynamic hedge ratios that minimize the effects of exposure to adverse price changes of crude oil and the refined petroleum products, using crack spread differentials while taking into account volatility spill over effects and time varying volatility as the key determinants of the dynamics in the oil industry.

1.4 Objectives of the Study

In many markets, energy markets included, the concept of perfectly replicating contracts by continuously trading the underlying asset is unrealistic. Some commodities like spot electricity cannot be easily stored, and therefore continuously adjusting trading positions may not be possible and hence not optimal. Most energy derivatives actually depend on the futures rather than spot prices, and futures contracts can be used to replicate trading positions that minimize exposures the price variation vulnerabilities. This study applies multivariate volatility models to estimate optimal dynamic hedge ratios for crack spreads with volatility spillovers in energy markets, specifically using crude, gasoline and distillate fuel prices.

1.4.1 Specific objectives

(i) Analysing the trends and patterns in energy markets and investigating the stylised features of the financial time series.

(ii) Investigate co-integration and multivariate heteroscedasticity in energy markets so as to capture volatility spill overs or transmissions in the markets.

(iii) Incorporating volatility spill over effects in the futures and spot markets market using MGARCH models.

(iv) Estimating static and dynamic hedge ratios that can account for spillover effects in energies and compare their hedge effectiveness.

1.5 Significance of study

Oil production and consumption is one of the greatest drivers of most economies the world over. As a transportation fuel, oil has virtually no substitute and there doesn't seem to be any immediate prospects for a substitute at current prices. The need for a reliable source of energy is key for any economy as is evidenced by the continuous explorations for oil prospects by most countries. This seems to be the only path for industrialization. Energy

Risk management is therefore an emerging area and strategies to hedge these risks would also be worth considering. Due to the heavy reliance of all economies around the world, on oil products as the main sources of energy, ability to carefully harness this energy, and handle the risks that come with this by turning them into opportunities will steer growth in any economy.

This study is also intended to establish the better hedging strategy for international trade affected by oil price fluctuations, demand and supply, and transportation among other factors. For an investor wishing to hedge their market risk exposure, crack spread trading in oil markets, with optimal dynamic hedge ratios would be an interesting alternative. Precise recognition of the trends and patterns in energy markets would also provide the much needed information in understanding market dynamics and the data generating process (DGP).

1.6 structure of the document

The rest of the document is structured as follows; In chapter 2, there is a review of previous literature, chapter 3 looks a univariate analysis of energy prices, chapter 4 discusses co-integration in energy models and uses MGARCH models to describe the dynamics of energy prices. In chapter 5, crack spread trading is discussed and hedge ratios are estimated based on the models developed in chapter 4. Finally, in chapter 6, we give Conclusions and make recommendations.

2. LITERATURE REVIEW

For any economy, investment in energy and management of risks in the energy markets is of paramount importance. Salman and Atya (2014) in their paper on the role of financial development and energy consumption on economic growth, state the importance of the existing causal relationship between energy consumption, financial development and economic growth, to policy makers, as it directly affects production, economic growth and development. According to Dorsman et al. (2013), oil shocks of the 1970's demonstrated the vulnerability of the world's economy to interruptions of energy supply, the steady increase in global energy demand, price volatilities, the recent energy price increases, political instability experienced in energy producing regions and the threats of terrorism related strikes against energy infrastructure have significantly led to a growing concern over security in the energy sector.

Moshiri and Foroutan (2006) observed that oil prices are complex and unpredictable and cannot therefore be subjected to traditional linear models since large errors would be expected if a linear model were to be applied to a non-linear DGP. They applied ARMA and GARCH models to describe the trends of, forecast crude oil prices. ARMA models provide insight into many areas of time series forecasting. An ARMA model entails order identification, parameter estimation and forecasting (Box and Jenkins, 1976). In ARMA modeling, the time series must be linear and stationary (Chan, 2011), though in real life, time series data are by nature non-linear and non-stationary. If the present can be modelled plausibly in terms of only the past values of the independent inputs, then, forecasting will be possible (Shumway and Stoffer, 2010). They are most widely used for the prediction of second-order stationary processes (Box and Jenkins, 1976; Chan, 2011; Montgomery et al., 2011; Taylor, 2007; Wei and Wei, 1990).

Engle (1982) introduced ARCH model, and later Bollerslev (1986) introduced the GARCH which was an important generalization of the ARCH model. This collection of GARCH volatility models have come up as a very important tool kit for empirical asset pricing and financial risk management. After the work of Engle (1982) and Bollerslev (1986), volumes of econometric literature have developed on the estimation and forecasting of volatility. Some literature in empirical finance such as Bollerslev (1986), Campbell and Hentschel (1992), French et al. (1987), Glosten et al. (1993), Pagan (1990), and Schwert (1989) apply GARCH models.

This family of models have been extended in various directions in order to increase the flexibility of the original ARCH model that was introduced by Engle (1982) and later generalization to GARCH by Bollerslev (1986). Herwartz and Reimers (2002) modelled and analysed daily log returns of exchange rates. To account for volatility clustering, they fit a GARCH(1,1)-model with leptokurtic innovations and found that this model accurately described the empirical distribution of foreign exchange returns. Similarly persuaded by the past success of GARCH models in fitting log returns and by the failure of deterministic volatility models in fitting derivative prices, researchers have broadened the GARCH models into the area of derivatives pricing.

With regard to the application of GARCH in derivatives pricing, Duan (1995) developed a risk-neutral model within the GARCH framework. He used the non-linear GARCH in mean (NGARCH-M) that was developed by Engle and Ng (1993). He characterized the transition between the real world and the risk-neutral probability distributions if the asset price dynamics of the underlying followed a GARCH process, and hence he established the foundation for option valuation under GARCH. An important feature all GARCH models is that they are non-Markovian in nature and hence can be used to explain some of the systematic biases associated with existing pricing models such as the Black-Sholes model. In essence, asset prices depends on the information set generated by the past and

present prices of the underlying stock, and its past, present and one-step-ahead values of the conditional variances.

Duan and Pliska (2004) built a co-integration option valuation model. Their approach was based on a discrete-time model where the assets were co-integrated process with MGARCH volatility. They showed that the co-integrating variable enters the pricing model only when volatility is dynamic or time-varying. They performed a numerical study for European-style spread options, compared the option prices obtained with their model to those obtained with a constant volatility model and with those obtained using an MGARCH model without co-integration. Despite GARCH models being convenient parametric models that can capture positive excess kurtosis and volatility clustering, Baillie and Bollerslev (1990) and Bollerslev (1987) among others note that GARCH models conditioned on the normal distribution of errors generally fail to capture the positive excess kurtosis which usually characterizes asset returns. Benavides (2004) applied the GARCH, the BEKK model (named after Baba, Engle, Kraft and Kroner, Engle and Kroner (1995)), an option implied and a composite forecast model to investigate the accuracy of volatility forecasts for corn and wheat futures and found that the option implied model was superior to the historical models and the composite forecast model was the most accurate.

Time varying hedging strategies constructed using MGARCH models have, to some certain extent, had empirical success with respect to risk reduction. These time varying hedge ratios are usually estimated using CCC (GARCH) model of Bollerslev (1990) and DCC (GARCH) model of Engle and Sheppard (2001). These models have the flexibility of univariate GARCH models and the parsimony of parametric models for correlations. They are non linear but can simply be estimated using methods based on the likelihood function. (Alizadeh et al., 2004; Cecchetti et al., 1988; Chang et al., 2011; Chen et al., 2003; De Jong et al., 1997; Haigh and Holt, 2002) among others discuss the concept of optimal hedging, optimal hedge ratios and hedging efficiency in futures markets.

Haigh and Holt (2002) allowed for time-varying volatility spillovers when finding the optimal hedge ratio and noted that from a risk management perspective, there had been no attempt to combine the estimation of dynamic hedge ratios while accommodating dynamic co-variability between related energies. They therefore built upon previous work to examine, the effectiveness of using crude oil, heating oil and unleaded gasoline futures contracts in helping reduce price uncertainty for energy-traders, for the first time. In particular, to directly consider time-varying volatility spillovers between related markets they employed an MGARCH model that also allowed for the direct incorporation of the time to maturity effect often found in futures markets. They note that considerable improvements over all other hedging procedures may be obtained if informational linkages among energy markets are directly accounted for in a time varying manner using an MGARCH model that directly incorporates maturity effect.

Alizadeh et al. (2004) also compared the effectiveness of constant versus time-varying hedge ratios and noted that differences in hedging effectiveness across regional markets are attributed to varying regional supply and demand factors in each market. Byun and Min (2011) investigated the disparity between the data generating and risk neutral conditional volatilities. They allowed the risk-neutral one-day ahead conditional volatility to be different from physical one-day-ahead conditional volatility so as to accurately describe the risk neutral dynamics of asset returns and volatility implied by a cross section of prices. Their results showed that using day-ahead conditional volatilities estimated under the real world measure did not indicate any advantage of the discrete GARCH option pricing model over continuous-time stochastic volatility models unless risk neutral day-ahead volatilities is adjusted for non-normality and risk premium.

Carmona and Durrleman (2003) surveyed the theoretical and computational problems associated with pricing and hedging financial instruments for which closed form formulae cannot be derived such as spread options. They also looked at their role as speculation devices and risk management tools. In considering risks associated with below-target

returns, Fishburn (1977) considers a mean-risk dominance model where a threshold is measured by a probability weighted function of deviations below a specified target return and later De Jong et al. (1997) uses this to study out of sample hedging effectiveness for currency futures.

3. TRENDS AND PATTERNS IN ENERGY MARKETS

Most financial data occur sequentially through time creating a time series. A time series is a chronological or a time-oriented sequence of observations on a variable of interest. When building dynamic econometric models, the first step is to carry out a detailed analysis of the attributes of the variables involved so that when modelling the DGP of a system of potentially related variables, properties of the individual series are considered (Lütkepohl and Krätzig, 2004). Econometricians are faced with the task of developing reasonably simple models that can predict future values, interpret and test hypothesis of economic series (Enders, 2008). There are two methods in time series analysis, namely: frequency-domain which entails the analysis of mathematical functions or signals with respect to frequency, and time-domain which entails the analysis of mathematical functions, physical signals or time series of economic or environmental data, with respect to time. This research is mainly concerned with time-domain.

Here an observed price series, $p_1, p_2, \ldots, p_T$, is regarded as a particular realisation of a stochastic process $\{p_t\}$. A realization of stochastic process is just a sample path of T real numbers which could for example be a stock price process; and if history took a different course, we would have observed a totally different sample path. A stochastic process can be described as a model which describes the probability structure of a sequence of random data, the simplest being specified as $p_t \overset{i.i.d.}{\sim} D$ for some distribution D.

The important concepts that underlie standard time series analysis include stationarity, ergodicity, autocorrelation, white noise and innovation processes. Time series analysis is grounded on a central family of models known as ARMA models (Francq and Zakoian, 2011). In addition to these, for the analysis of financial time series, the concept of vola-

tility, which is pivotal in finance has to be incorporated. Along side volatility, the main stylized facts which include unpredictability of returns, volatility clustering and hence predictability of squared returns, leptokurticity of the marginal distributions, asymmetries, etc. concerning financial time series, are also considered (Francq and Zakoian, 2011; Tsay, 2010; Zivot and Wang, 2007).

This chapter is structured into three main sections. Section 3.1 introduces time series concepts and important models together with extensions to model financial time series data. Section 3.2 discusses financial time series concepts and models and in section 3.3, an exploratory analysis of the US energy markets is carried out based on the models discussed in section 3.1.

3.1 Time Series Concepts and Models

Figure 3.1 is a time series plot of the daily prices of Cushing OK WTI crude oil futures for contract 1 for the period 03/01/2006 to 22/05/1015. In this section, we discuss the basic

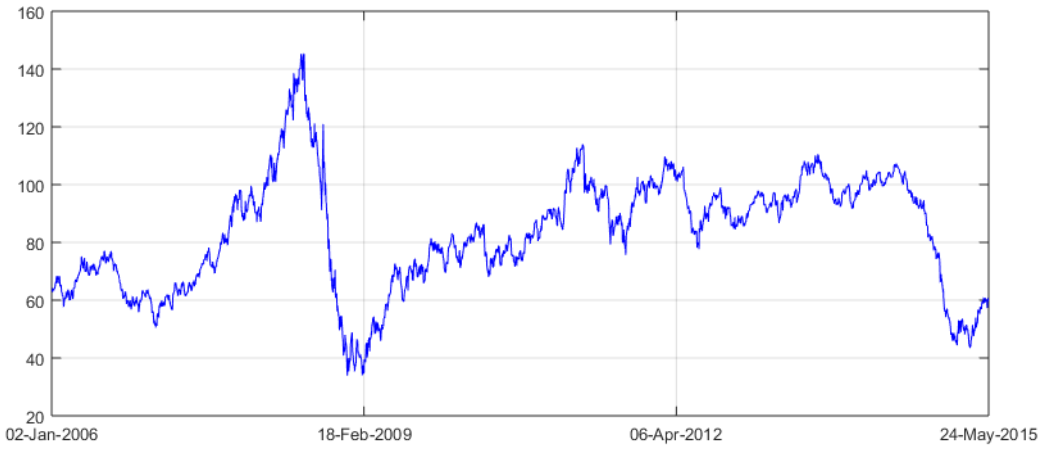

Figure 3.1: Cushing OK WTI crude oil futures contract 1 daily prices

concepts of time series analysis which are the building blocks for complex or advanced time series models. A time series is a stochastic process which can be defined as follows:

Definition 3.1.1 (Time series model)**.** *For some observed data* $\{p_t\}$ *at time* t*, a time series model is a specification of the joint distributions (or possibly only the means and*

covariances) of a sequence of random variables $\{P_t\}$ of which $\{p_t\}$ is postulated to be a realization at time t (Brockwell, 2002).

3.1.1 Stationarity

One of the most basic requirements for any statistical data analysis is the existence of some statistical properties of the data which remain stable over time. This is important for parameter estimation, and for characterization of the dependence between observations over time. Stationarity of a time series relates to its statistical properties in time. In a more strict sense, a stationary time series exhibits similar "statistical behaviour" in time, often characterized by a constant probability distribution in time (Montgomery et al., 2011).

There are two forms of Stationarity, namely, strict and weak or covariance stationarity.

Definition 3.1.2 (Strict Stationarity)**.** *A stochastic process $\{p_t\}$ is strictly stationary if it's properties are not affected by a change in the time origin. The joint probability distribution associated with t observations $p_1, p_2, \ldots, p_t$, at times $1, 2, \ldots, t$ and the joint distribution associated with t observations $p_{k+1}, p_{k+2}, \ldots, p_{k+t}$, at times $k{+}1, k{+}2, \ldots, k{+}t$ for $k \in \mathbb{Z}$ are the same.*

The joint distribution of any such set of observations remains unchanged even if all the times of the observations are shifted either forward or backwards by any integer amount k. The only factor which affects the relationship between the two sets of observations is the distance between them. For a strictly stationary process, since the distribution function is the same for all t, the mean function $\mu_t = \mu$, a constant provided $E[p_t] < \infty$, the variance function $\sigma_t^2 = \sigma^2$ for all t provided $E[p_t^2] < \infty$, the covariance function $\gamma(i,j) = \gamma(i+k, j+k)$ and the correlation function $\rho(i,j) = \rho(i+k, j+k)$. Also if we let $i = t-k$ and $j = t$, we will have $\gamma(i,j) = \gamma(t-k,t) = \gamma(t,t+k) = \gamma(k)$ and $\rho(i,j) = \rho(t-k,t) = \rho(t,t+k) = \rho(k)$ since $\gamma(k)$ and $\rho(k)$ are even functions and hence symmetric (Wei and Wei, 1990). The first two moments are also finite and the covariance and correlation between p_t and p_{t+k} depend only on the distance between them or the

time difference k.

Definition 3.1.3 (Covariance Stationarity)**.** *A stochastic process* $\{p_t\}$ *is covariance stationary if the unconditional mean and unconditional variance are finite and do not change over time. This means*

$$E[p_t] = \mu \quad \forall\, t, V[p_t] = \sigma^2 < \infty \quad \forall\, t \quad E[(p_t - \mu)(p_{t-k} - \mu)] = \gamma_k \quad \forall\, t \text{ and any } k. \tag{3.1}$$

From (3.1), We notice that a strictly stationary process which has the first two moments being finite is also covariance stationary. The white noise process is the simplest example covariance stationarity. This process facilitates the construction of more complex stationary processes making it particularly important. Covariances are particularly important as input to, for example, a portfolio analysis, though, in order to understand the relationship between variables, examining their sample correlations is more informative.

Definition 3.1.4 (White noise)**.** *The process* $\{a_t\}$ *is called weak white noise if, for some positive constant* σ^2*; (i)* $E[a_t] = 0 \quad \forall\, t$*, (ii)* $V[a_t] = \sigma^2 \quad \forall\, t$ *and (iii)* $Cov(a_t, a_{t+k}) = 0 \quad \forall\, t$ *and for* $k \neq 0$.

It is sometimes necessary, particularly for financial time series, to replace hypothesis (iii) by the stronger hypothesis that $\{a_t\}$ and $\{a_{t+k}\}$ are independent and identically distributed (i.i.d). In this case the process $\{a_t\}$ is said to be strong white noise. For strong white noise, $\{a_t\}$ cannot be predicted either linearly or non-linearly and it reflects the true innovation in the series. For a general white noise process, the innovation term may not be predictable linearly, yet it can be probably predictable non-linearly using ARCH or GARCH models. Figure 3.2 shows a simulated white noise process of 5,000 values with $\sigma^2 = 4$

In analysis of time series data, the initial step is to make the non-stationary data stationary by removing the trend or any other inherent pattern, since, a strong trend will usually obscure the behaviour of the stationary process and also most statistical and econometric models only apply to stationary time series. This is achieved either by de-

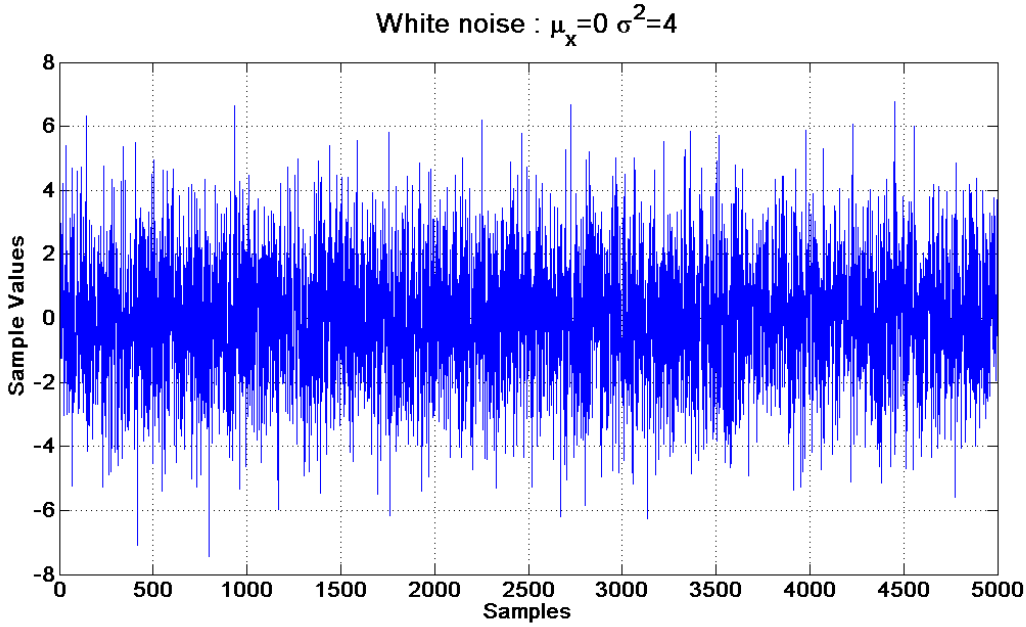

Figure 3.2: White noise process

trending or differencing the data. The simplest form of stationarity to work around is trend stationarity, where, the process is stationary around a trend. This type of model is written as

$$p_t = \mu_t + a_t \tag{3.2}$$

where p_t represents the observations through time, μ_t denotes the trend, and a_t denotes the innovations which form the stationary process. The trend is removed by estimating the trend component $\widehat{\mu_t}$ using an appropriate method like say, OLS, and then working with the residuals $\widehat{a_t}$. This is known as de-trending, given as

$$\widehat{a_t} = p_t - \widehat{\mu_t} \tag{3.3}$$

With respect to differencing, the first difference, denoted by $\nabla p_t = p_t - p_{t-1}$ is used to eliminate the linear trend while a second difference eliminates a quadratic trend, and so on. To define higher order differences, we can use the back-shift operator, given by $Bp_t = p_{t-1}$, and its powers can be extended so that $B^2 p_t = B(Bp_t) = Bp_{t-1} = p_{t-2}$ and so on, so that $B^k p_t = p_{t-k}$. With this, we can rewrite the first difference equation as

$\nabla p_t = (1 - B)\, p_t$ and differences of order d can be written as $\nabla^d = (1 - B)^d$. The second difference for example becomes $\nabla^2 p_t = (1 - B)^2\, p_t = (1 - 2B + B^2)\, p_t = p_t - 2p_{t-1} + p_{t-2}$.

Instead of modelling a fixed trend as in (3.2), we can use the random walk with drift to model the trend as a stochastic process.

$$\mu_t = \delta + \mu_{t-1} + w_t \tag{3.4}$$

so that if the appropriate model is (3.2), then differencing p_t, yields a stationary process; that is,

$$\begin{aligned} p_t - p_{t-1} &= (\mu_t - a_t) - (\mu_{t-1} - a_{t-1}) = \delta + \mu_{t-1} + w_t + a_t - \delta - \mu_{t-2} - w_{t-1} - a_{t-1} \\ &= \mu_{t-1} - \mu_{t-2} + w_t - w_{t-1} + a_t - a_{t-1} \\ \text{but } \mu_{t-1} &= \delta + \mu_{t-2} + w_{t-1},\ \text{so } \mu_{t-1} - \mu_{t-2} = \delta + w_{t-1} \end{aligned} \tag{3.5}$$

and $p_t - p_{t-1} = \delta + w_{t-1} + w_t - w_{t-1} + a_t - a_{t-1} = \delta + w_t + a_t - a_{t-1}$ which is stationary.

One advantage of differencing over de-trending is that when differencing, there is no need to estimate any parameters, though, it does not estimate the stationary process $\{p_t\}$. When analysing time series, we must check for the presence of unit roots. The order of integration of each variable in the model is tested to check for stationarity and establish the number of times the variable needs to be differenced to yield stationarity.

If to achieve stationarity, a series, p_t must be differenced d times, then this series is said to be integrated of order d, written as $p_t \sim I(d)$ so that if $p_t \sim I(d)$, then, p_t contains d unit roots, and $\nabla^d p_t \sim I(0)$. While an $I(1)$ series contains one unit root, an $I(0)$ series is stationary. Most financial time series contain one unit root. They are $I(1)$ processes, although some are stationary. From literature however, consumer prices have been argued to have two unit roots. There are several ways of testing for unit roots, but, the most popular is the Dickey-Fuller (DF) test, because of it's simplicity and general nature. It tests the null hypothesis that the series contains a unit root, against the alternative of

stationarity.

3.1.2 Ergodicity

In the analysis of time series , the fundamental challenge is that the realization of a process is observed only once out of the many that could have arisen from the stochastic process had history taken a different course. Egordicity describes a random process where the time average of one sequence of events is the same as the average of the whole ensemble.

Definition 3.1.5 (Ergodicity)**.** *If a time series* $\{p_t\}$ *is ergodic and it's* r^{th} *moment* μ_r *is finite, then, the ensemble mean can be determined by the time average based on a single realization given the following limit for the ensemble average.* $T^{-1}\sum_{t=1}^{T} p_t^r \xrightarrow{p} \mu_r$ *where* T *is the time interval of 1 realization and p represents the convergence in probability.*

Ergodicity has to do with asymptotic independence. If a stochastic process, which is ergodic, is sampled at two points far apart in time, the samples will be independent. Using a single realisation to infer unknown parameters of a joint probability distribution only works if the process is ergodic, meaning, sample moments for finite stretches of the realisation approach their population counterparts as time tends to infinity (Mills and Markellos, 2008). The ergodicity theorem states that averages will converge to their expectation, provided the expectation exists. For an ergodic process, there is no need observing separate independent replications of the entire process when estimating its mean or other moments. One sufficiently long sample path enables us to estimate the underlying moments (Chan, 2011).

3.1.3 Autocorrelation and partial autocorrelation

For stationary time series, the joint probability distribution of any two observations, p_t, and p_{t+k} is similar for any two time periods t and $t+k$ that are separated by the same time difference or lag length k (Montgomery et al., 2011). The auto-covariance of a time series measures the dependence between observations. For a stationary process $\{p_t\}$, the auto-covariance function at lag k is $\gamma(k) = \mathrm{Cov}\,(p_{t+k}, p_t) = E\left[(p_{t+k} - \mu)(p_t - \mu)\right]$ for $k \neq 0$. When there is interest in the linear dependence between $\{p_{t+k}\}$ and its past values $\{p_t\}$,

the correlation concept is generalized to autocorrelation (Tsay, 2010). The autocorrelation coefficient for a time series measures of the auto-covariance restricted to lie between -1 and 1. For a process $\{p_t\}$, the ACF at lag k is denoted by

$$\rho(k) = \text{Cor}\,(p_{t+k}, p_t) = \frac{E\left[(p_{t+k} - \mu)(p_t - \mu)\right]}{E\left[(p_t - \mu)^2\right]} = \frac{\gamma(k)}{\gamma(0)} \text{ for } k \neq 0. \tag{3.6}$$

For a stationary process, the variance $\sigma^2 = \gamma(0)$ is the same at time $t + k$ as at time t, implying that $\rho(0) = 1$. Classical time series analysis is focused on the second-order structure of the processes. Gaussian stationary processes are completely characterized by their mean and auto-covariance function. For non-Gaussian processes, the mean and auto-covariance give a first idea of the temporal dependence structure. In practice, these moments are unknown and are estimated from a time series (Francq and Zakoian, 2011).

For a series with stationary DGP, the sample autocorrelations usually die out fast as the lag length k increases. The correlation between $\{p_{t+k}\}$ and $\{p_t\}$ after their mutual linear dependency on the intervening variables $\{p_{t+k-1}, p_{t+k-2}, \ldots, p_{t+1}\}$ has been removed, gives the partial autocorrelation which provides important information on the properties of the DGP of a given series (Wei and Wei, 1990). Partial autocorrelation between $\{p_{t+k}\}$ and $\{p_t\}$, is the autocorrelation conditional on the in–between values of the time series. Formally,

$$\text{Cor}\ \ (p_t, p_{t+k} | p_{t+1}, p_{t+2}, \ldots, p_{t+k-1}) \tag{3.7}$$

represents the partial autocorrelation in the time series analysis. The partial correlation between $\{p_t\}$ and $\{p_{t+k}\}$ will equal the ordinary correlation between $(p_t - \hat{p}_t)$ and $(p_{t+k} - \hat{p}_{t+k})$. If $\varrho(k)$ denotes the partial autocorrelation between $\{p_t\}$ and $\{p_{t+k}\}$, then, the PACF is given by

$$\varrho(k) = \frac{\text{Cov}\left[(p_t - \hat{p}_t), (p_{t+k} - \hat{p}_{t+k})\right]}{\sqrt{\text{Var}\,(p_t - \hat{p}_t)}\sqrt{\text{Var}\,(p_{t+k} - \hat{p}_{t+k})}} \tag{3.8}$$

where $\hat{p}_t$ and $\hat{p}_{t+k}$ are the best linear estimates of p_t and p_{t+k} respectively and

$$\begin{aligned} \hat{p}_t &= \beta_1 p_{t+1} + \beta_2 p_{t+2} + \ldots + \beta_{k-1} p_{t+k-1} \text{ and} \\ \hat{p}_{t+k} &= \alpha_1 p_{t+k-1} + \alpha_2 p_{t+k-2} + \ldots + \alpha_{k-1} p_{t+1} \end{aligned} \tag{3.9}$$

and β_i and α_i for $(1 \leq i \leq k-1)$ are the mean square linear regression coefficients obtained by minimizing

$$\begin{aligned} E\left(p_t - \hat{p}_t\right)^2 &= E\left(p_t - \beta_1 p_{t+1} - \beta_2 p_{t+2} - \ldots - \beta_{k-1} p_{t+k-1}\right)^2 \text{ and} \\ E\left(p_{t+k} - \hat{p}_{t+k}\right)^2 &= E\left(\alpha_1 p_{t+k-1} - \alpha_2 p_{t+k-2} - \ldots - \alpha_{k-1} p_{t+1}\right)^2 \end{aligned} \tag{3.10}$$

3.1.4 Autoregressive moving average (ARMA) Models

ARMA models are a very useful class of models used to describe time series dynamics. An ARMA model allows the dependent variable to depend on the past or lagged values of the independent variables and possibly its own past or lagged values. Forecasting is possible if the present can be modelled in terms of only the past or lagged values of the independent variable (Shumway and Stoffer, 2010). In order to predict second–order stationary processes, ARMA models are the most widely used (Box and Jenkins, 1976; Chan, 2011; Montgomery et al., 2011; Taylor, 2007; Wei and Wei, 1990). These ARMA models can be further decomposed into the autoregressive (AR) and the moving average (MA) processes.

3.1.4.1 Autoregressive (AR) Models

AR models stem from the idea that for a time series $\{p_t\}$, the current values can be described by p past values, $p_{t-1}, p_{t-2}, \ldots, p_{t-p}$, where p represents the number of steps back into the past needed to predict the current value. Current values are therefore functions of the past p values of the time series $\{p_t\}$. AR models closely resemble the traditional regression models with the predictor variables replaced by the past (lagged) values of the time series, a property that makes them so appealing. An AR model of order

p, abbreviated $AR(p)$, for a stationary series $\{p_t\}$ takes the form

$$\begin{aligned} p_t &= \phi_1 p_{t-1} + \ldots + \phi_p p_{t-p} + a_t, \quad a_t \sim \text{WN}(0, \sigma^2) \\ &= \sum_{i=1}^{p} \phi_i p_{t-i} + a_t, \end{aligned} \tag{3.11}$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are constants with $\phi_p \neq 0$. The mean μ of $\{p_t\}$ in (3.11) is zero. If $\mu \neq 0$, then replace $\{p_t\}$ by $\{p_t - \mu\}$ in (3.11) giving

$$\begin{aligned} p_t - \mu = \phi_1 (p_{t-1} - \mu) + \phi_2 (p_{t-2} - \mu) + \ldots + \phi_p (p_{t-p} - \mu) + a_t \text{ so that} \\ p_t = \phi_0 + \phi_1 p_{t-1} + \phi_2 p_{t-2} + \ldots + \phi_p p_{t-p} + a_t, \end{aligned} \tag{3.12}$$

where $\phi_0 = \mu (1 - \phi_1 - \ldots - \phi_p)$, so that $\mu = \dfrac{\phi_0}{1 - \phi_1 - \ldots - \phi_p} = \dfrac{\phi_0}{1 - \sum_{i=1}^{p} \phi_i}$. The general $AR(p)$ process can also be written using the back-shift operator, as

$$\left(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p\right) p_t = a_t, \tag{3.13}$$

or more succinctly, as $\phi(B)\, p_t = a_t$, where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ is the AR operator. For the polynomial $\phi(B) = 0$, if the solution lies outside the unit circle, then the $AR(p)$ process will be stationary. However, if a variable contains a unit root, then it is non-stationary. Some of the properties of a general $AR(p)$ model like the one in (3.11) include:

(i) The mean, $E[p_t] = \dfrac{\phi_0}{1 - \sum_{i=1}^{p} \phi_i}$, hence the mean is only finite if $\sum_{i=1}^{p} \phi_i < 1$.

(ii) The variance, $V[p_t] = \dfrac{\sigma^2}{1 - \sum_{i=1}^{p} \phi_i \rho(i)}$ where $\rho(i)$ is the i^{th} autocorrelation. The variance is only finite if $\sum_{i=1}^{p} \phi_i < 1$.

(iii) The auto-covariance, $E[(p_t - E[p_t])\,(p_{t-k} - E[p_{t-k}])] \neq 0$ for any k (generally, though certain parametrizations could produce 0 auto-covariances)

Following the definitions of ACF and $PACF$ in section 3.1.3, the ACF of an $AR(p)$ process can be a combination of exponential decay and diminishing sinusoid expressions

based on the roots of associated polynomial, and, the $PACF$ between p_t and p_{t-k} for an $AR(p)$ model for $k > p$ should equal zero.

3.1.4.2 Moving average (MA) Models

Let $\{a_t\}$ be a sequence of i.i.d random variables with zero mean and variance σ^2. If we require $\{a_t\}$ only to be uncorrelated, and not necessarily independent, then $\{a_t\}$ is just a white noise sequence denoted by $a_t \sim \text{WN}(0, \sigma^2)$. This means this system is random with no systematic structures. Forming a weighted average of $\{a_t\}$ gives an $MA(q)$ time series model represented as

$$\begin{aligned} p_t &= a_t + \psi_1 a_{t-1} + \ldots + \psi_q a_{t-q}, \quad a_t \sim \text{WN}(0, \sigma^2) \\ &= a_t + \sum_{i=1}^{q} \psi_i a_{t-i} \end{aligned} \tag{3.14}$$

This model has q lags and $\psi_1, \psi_2, \ldots, \psi_q$ are parameters with $\psi_q \neq 0$. The $MA(q)$ process may also take the the form $p_t = \psi(B)a_t$ where, $\psi(B)$, the MA operator, given by $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots + \psi_q B^q$ defines a linear combination of values in the shift operator $B^k a_t = a_{t-k}$. In contrast with the AR process, the MA process is stationary for any values of the parameters $\psi_1, \psi_2, \ldots, \psi_q$.

If an MA model is algebraically equivalent to a converging infinite order AR model, then it is said to be invertible. Converging here implies the AR coefficients diminish to 0 as we move back in time. To ensure that this happens, $|\psi| < 1$. The condition for invertibility of a MA process is the counterpart to the condition of stationarity of an AR. An $MA(q)$ model is invertible if the roots of the polynomial $\psi(B) = 0$ all lie outside the unit circle. Some of the properties of a general $MA(q)$ model include,

(i) The mean, $E[p_t] = 0$.

(ii) The variance, $V[p_t] = \left(1 + \sum_{i=1}^{q} \psi_i^2\right) \sigma^2$.

(iii) The auto-covariance, $E[(p_t - E[p_t])\,(p_{t-k} - E[p_{t-k}])] = \sigma^2 \sum_{i=0}^{q-k} \psi_i \psi_{i+k}$ for $k \leq q$ and 0 otherwise.

ACF and $PACF$ have very distinct and indicative properties for AR and MA processes respectively, and are thus necessary for identification of the model. The $PACF$ of an $MA(q)$ process is a combination of exponential decay and diminishing sinusoid expressions based on the roots of associated polynomial, and, the ACF between a_t and a_{t-k} for an $MA(q)$ model for $k > q$ should be equal to zero.

Although both AR and MA processes have their own merits, in order to capture the complexity of the structure oa a time series, a relatively long AR and MA model is required. This compromises parsimony and to correct this, the AR and MA parts are combined form an ARMA process (Chan, 2011).

Definition 3.1.6 (ARMA process)**.** *A time series* $\{p_t\}$ *is* $ARMA(p, q)$ *if it is stationary and*

$$\begin{aligned} p_t &= \phi_1 p_{t-1} + \ldots + \phi_p p_{t-p} + a_t + \psi_1 a_{t-1} + \ldots + \psi_q a_{t-q}, \\ &= \sum_{i=1}^{p} \phi_i p_{t-i} + \sum_{j=1}^{q} \psi_j a_{t-j} + a_t \end{aligned} \tag{3.15}$$

with $\phi_p \neq 0$, $\psi_q \neq 0$ *and* $\sigma_a^2 > 0$*, and, where* a_t *is a white noise process with the additional property that* $E_{t-1}[a_t] = 0$ *and* p *and* q *are the AR and the MA orders, respectively.*

The $ARMA(p, q)$ model can be expressed as $\phi(B)p_t = \psi(B)a_t$ using the back-shift operator. The stationarity of an $ARMA(p, q)$ model is related to the AR component in the model and if the absolute values of all the roots of the polynomial $\phi(B) = 0$ are less than one, then, the $ARMA(p, q)$ process is stationary. Similarly, the invertibility of an $ARMA(p, q)$ process is related to the MA component and if the absolute values of all the roots of the polynomial $\psi(B) = 0$ are less than one, then, the $ARMA(p, q)$ process is invertible.

The efficiency of ARMA models lie in their parsimonious representation, and their properties can be characterized using their ACF and PACF just like in the case of AR and MA models. Similar to the stationarity and invertibility conditions, the ACF and $PACF$ of

$ARMA$ processes are driven by the AR and MA components respectively. The ACF and $PACF$ of an $ARMA(p,q)$ both exhibit exponential decay and/or diminishing sinusoid patterns, which makes model identification of the $ARMA(p,q)$ model comparably more difficult. Table 3.1 summarises the theoretical behaviours of the ACF and $PACF$ for stationary time series models.

Table 3.1: Nature of the ACF and the $PACF$ for stationary time series models

Model	ACF	$PACF$
$AR(p)$	Exponential decay and/or diminishing sinusoid	Cuts off after lag p
$MA(p)$	Cuts off after lag q	Exponential decay and/or diminishing sinusoid
$ARMA(p,q)$	Exponential decay and/or diminishing sinusoid	Exponential decay and/or diminishing sinusoid

A homogeneous non-stationary process $\{p_t\}$ whose first differences $a_t = p_t - p_{t-1} = (1-B)\,p_t$ or higher-order differences $a_t = (1-B)^d\,p_t$ produce a stationary time series is an ARIMA process of orders $p, d,$ and q. The $ARIMA(p,d,q)$ model can therefore be written as:

$$\phi(B)(1-B)^d p_t = \psi(B)a_t \tag{3.16}$$

After differencing to obtain the stationary time series $a_t = (1-B)^d\,p_t$, methods provided previously can be employed to obtain the full model. Usually, for most cases, $(d=1)$ and sometimes $(d=2)$ suffices to achieve stationarity. However, occasionally, other transformations such as taking logarithms of the original series can be used to achieve stationarity. After model identification, methods such as least squares, MLE and method of moments can be employed for the estimation of parameters for the provisionally identified model. After fitting a provisional model to the data, its adequacy is examined through residual analysis and potential improvements proposed.

3.2 Financial Time Series Concepts and Models

So far, in the previous section, we have discussed general time series concepts and models. In this section, we discuss the basic concepts of financial time series analysis which is an extension of the time series models, to cater for peculiar properties inherent in financial data.

In most financial studies, we consider asset returns, rather than asset prices, because, foremost, for an investor, the return represents a complete and scale-free summary of the investment opportunity and secondly, return series have desirable statistical properties (Tsay, 2010). Rather than analyse price series $\{p_t\}$ which often displays unit-root behaviour, they are transformed to log–return series and analysed. For a price series $\{p_t\}$, the continuously compounded return, or simply, the log–return r_t is represented as

$$r_t = \log p_t - \log p_{t-1} = \ln\left(\frac{p_t}{p_{t-1}}\right) = \ln\left(1 + \frac{p_t - p_{t-1}}{p_{t-1}}\right) = \ln(1 + R_t), \tag{3.17}$$

where R_t is the simple or relative return. Log–returns approximately equal relative price changes and conform to the stationarity hypothesis, at least in the short run. Log–returns rather that relative returns are considered for analysis because of their additivity property, not shared by relative returns.

3.2.1 Stylised facts in finance

Modelling financial time series is complex mainly as a result of the existence of stylised facts which are challenging to artificially reproduce using stochastic models. These stylised facts were first discussed by Mandelbrot (1963) and have since then been extensively researched on (Campbell et al., 1997; Cont, 2001; Danielsson, 2011; Francq and Zakoian, 2011; Pfaff, 2012; Tsay, 2010). They can be observed more or less clearly depending on the nature of the time series and its frequency. These observed properties have important implications for assessing how appropriate a chosen model is. Figure 3.3 shows the return

series plot for the daily prices of Cushing OK WTI crude oil futures for contract 1 for the period 03/01/2006 to 22/05/1015.

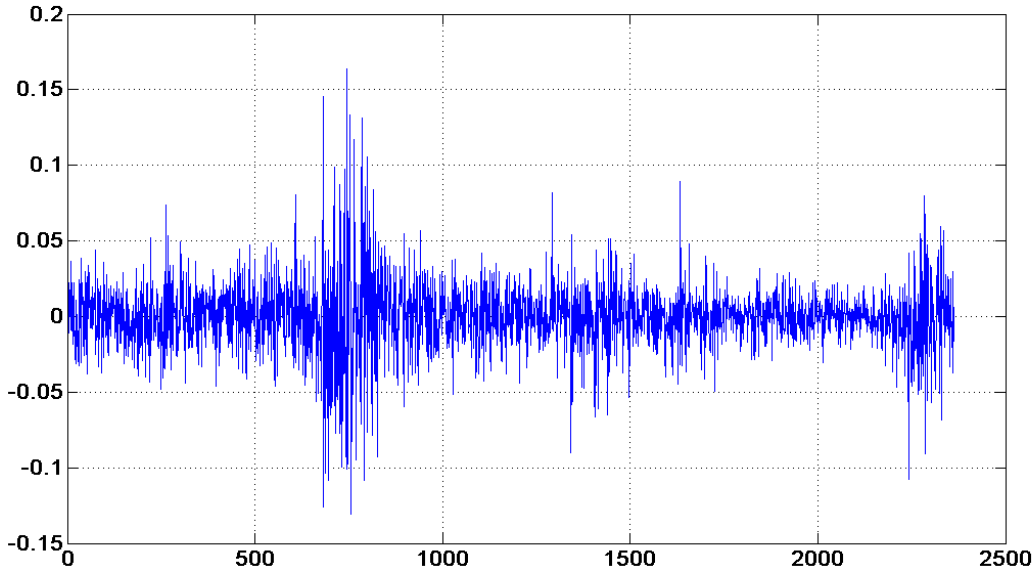

Figure 3.3: Cushing OK WTI crude oil futures contract 1 daily return series

From figure 3.3, we see that sample paths of prices generally look similar to random walks as were depicted in figure 3.1, with oscillations varying in magnitude but being almost averagely constant over long sub-periods. For observed financial market data, the following stylised facts can be stated.

3.2.1.1 Volatility clusters

The observation that the magnitudes of financial return volatilities tend to cluster together, so that turbulent (high-volatility) periods are followed by quiet (low-volatility) periods relates to volatility clustering. Extreme returns are observed closely in time as shown in figure 3.3. Changes between periods of low, medium and high turbulence occurring randomly and can not be modelled using any systematic pattern. This implies that shocks due to volatility presently will influence the expectations of volatility in several future periods (Engle and Patton, 2001). Volatility clustering is further evidenced in figure 3.4 which depicts the sample ACF of the Cushing OK WTI crude oil futures contract 1

daily return and the squared return series. The ACF of the squared returns show conside-

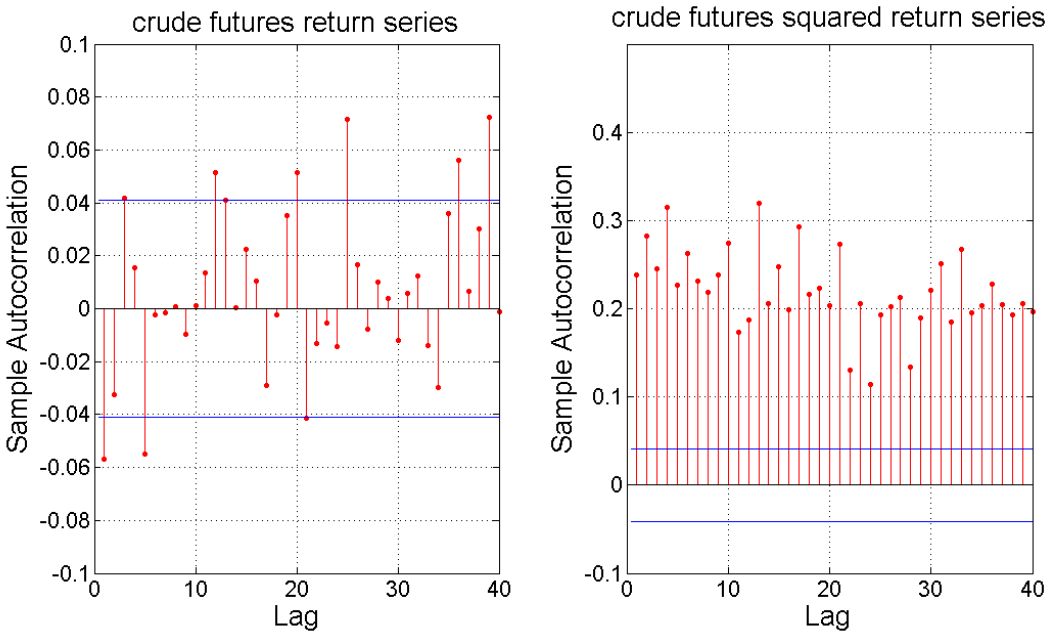

Figure 3.4: Sample ACF for Cushing OK WTI crude oil futures contract 1 daily return and squared return series

rable positive autocorrelation which is attributed to volatility clustering. This causes the similarities of values of the squared returns between some days say t to $t-k$, at least for small values of k. GARCH and stochastic volatility models are stochastic processes with a non-constant conditional variance, a property necessary when modelling volatility clustering. This makes GARCH and stochastic volatility models appropriate for modelling volatility clustering.

3.2.1.2 Fat tails (leptokurticity)

The empirical distribution of daily returns does not resemble a Gaussian distribution. The distribution of financial market returns is leptokurtic. The densities have heavy tails, meaning that extreme values occur more often than is implied in the Gaussian distribution. They are also sharply peaked at zero. Heavy tails, point to the fact that financial returns occasionally have very large negative or positive returns, which are very unlikely to be observed, if returns were to be normally distributed. The excess kurtosis for the returns series is greater than 0. When the time interval over which returns are computed increases, excess kurtosis vanishes and the empirical distributions tend towards

a Gaussian distribution.

3.2.1.3 Non-linear dependence

The autocorrelation coefficient of linearly dependent return series describes how the series move together. Non-linear dependence of return series addresses how they relate with each other, and usually, the absolute or squared returns are highly autocorrelated if there exists serial dependence.

3.2.1.4 Leverage effect

Leverage effect, which involves an asymmetry of the impact of past positive and negative values on the current volatility, was first noted by Black (1976). Negative returns (corresponding to price decreases) tend to increase volatility by a larger amount as compared to positive returns (price increases) of a similar magnitude.

3.2.1.5 Seasonality

Calendar effects manifest in financial returns data. The effects of the day of the week and the closeness of holidays among other seasonal phenomenon, may significantly affect returns. After a period of market inactivity, volatility tends to rise, reflecting the news or information that cumulated during the break. However, it can be seen that this boom is much less than if the information had been accrued at a steady pace. Seasonal effects can also be observed in intra-day data.

3.2.2 Conditional Heteroscedastic Models

The effects of heteroscedasticity observed in a time series process typically in form of heavy tails, volatility clusters, and the leverage effect cannot be accounted for by the mean equation, hence the need for conditional heteroscedastic models. Consider a log-return series r_t with mean given by $\mu_t = E\left[r_t|\mathcal{F}_{t-1}\right]$ and variance given by $\sigma_t^2 = \mathrm{Var}\left[r_t|\mathcal{F}_{t-1}\right] = E\left[\left(r_t - \mu_t\right)^2|\mathcal{F}_{t-1}\right]$, where $\mathcal{F}_{t-1}$ consists of the historical information (all linear functions

of past returns). The dependence, if seen is weak (weakly stationary). So we can have

$$\begin{aligned} r_t &= \mu_t + a_t \\ \mu_t &= \sum_{i=1}^{p} \alpha_i Y_{t-i} - \sum_{i=1}^{q} \beta_i a_{t-i} \text{ and} \\ Y_t &= r_t - \alpha_0 - \sum_{i=1}^{k} \theta_i x_{it}, \end{aligned} \tag{3.18}$$

where a_t is the shock at time t also known as the innovation process, x_{it} are explanatory variables and Y_t represents the adjusted series of returns after the effect of explanatory variables have been corrected for. Considering (3.18), a_{t-1} can be regarded as new information obtained at time t regarding the time series .

Notice that

$$\begin{aligned} a_t &= r_t - \mu_t, \\ E\left[a_t\right] &= 0, \text{ and} \\ \text{Var}\left[a_t\right] &= E\left[a_t^2\right] - 0 = E\left[a_t^2\right]. \end{aligned} \tag{3.19}$$

Conditional heteroscedastic models are concerned with how the conditional standard deviation of an asset return σ_t^2, which measures the volatility of an asset, evolves. σ_t^2 can not be directly observed, but possesses some attributes seen commonly in asset returns. These include Volatility clustering, continuous evolution of volatility over time, (jumps are not common), stationarity (volatility does not infinitely diverge) and leverage effect (volatility reacts differently to big price changes, compared to small price changes.)

Volatility models can generally be classified into two, namely:

1. The type that formulates the conditional variability directly as a function of observable phenomenon. GARCH models fall in this category.

2. The types that do not formulate volatility models as functions purely of observables, rather, they use stochastic equations to describe σ_t^2. These are the latent or

stochastic volatility models.

In this work, we will concentrate mainly on the first category. Conditional heteroscedastic models, which model volatility, play an important role in today's financial risk management. Continuous time stochastic volatility models, which can be a bit difficult to implement (observations are discrete), are effective for option pricing.

Understanding changes in σ_t^2 is of importance in financial markets since agents and market players require higher compensation for holding riskier assets, and this compensation takes the form of expected returns. Further, a time series with variance changing over time definitely has implications on the validity and efficiency of statistical inference about the parameters. It is partly for these reasons that time series models with heteroskedastic errors came into being.

3.2.2.1 Generalised autoregressive conditional heteroscedastic (GARCH) Models

Energy price returns distribution tend to be leptokurtic (fat tailed) and crude oil prices also exhibits volatility clustering. These two behaviours are consistent with the process where volatility is both stochastic and autoregressive. Models that attempt to incorporate and explain these behaviours are the GARCH models. The GARCH model framework, which was developed by Bollerslev (1986), explains variance using two distributed lags, the first one on past squared residuals so as to capture high frequency effects, and the second one on lagged values of the variance itself, so as to capture longer term influences. This enables the model th capture volatility clustering as well as the positive excess kurtosis that is prevalent in the unconditional distribution of returns.

A GARCH model is a parametric error distribution model, commonly used in the analysis of financial time series. It captures volatility clustering and also accommodates positive excess kurtosis (i.e. thick tails). The GARCH model expresses the conditional variability at time t as a linear combination of it's past values and of the squared variance values in a manner similar to an $ARMA$ model for conditional mean. The simplest $GARCH$

model is the $ARCH(1)$ process.

ARCH Models

Engle (1982) proposed a stationary non-linear model for the return series r_t which he termed the ARCH model. This model provides the first systematic framework used in modelling volatility. Formally, the ARCH model is simply a regression model that models the conditional volatility as the response variable and the past lags of squared innovations are modelled as covariates. For an ARCH model basically, (i) innovations or shocks are dependent but serially uncorrelated, and (ii) this dependence of the shocks can be described simply using a quadratic function of its lagged values. The conditional errors are normally distributed and the conditional variance is a linear function of past squared errors. This causes extreme values to be followed by other extreme values, albeit of unpredictable signs.

Let a_t be the innovation or shock of an asset return at time t be, so that $a_t = r_t - \mu_t$, μ_t is the expected return and $r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$ is the conditional return. a_t can also represent the unexpected return. In its simplest form, the $ARCH(p)$ model represents an innovation series a_t as

$$a_t = \sigma_t \varepsilon_t \text{ where } \varepsilon_t \sim N(0,1), \text{ and } \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2. \tag{3.20}$$

$\alpha_0 \ldots, \alpha_p$ are constants, volatility depends on historical news hence the conditional variance is evolves according to the previous values a_{t-i}. The effect of a shock i periods ago, for $(i \leq p)$ on the volatility today is governed by α_i and as expected, $\alpha_i < \alpha_j$ for $i > j$ since the effect of historical information on current volatility diminishes with time. Any information more than p periods old does not affect current volatility

For an $ARCH(1)$ where: $a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2$, and- $\alpha_0 > 0, \alpha_1 \geq 0$

1. $E[a_t]$, which is the unconditional mean of a_t is 0, since $E[a_t] = E[E(a_t|\mathcal{F}_{t-1})] = E[\sigma_t E(\varepsilon_t)] = 0$, where $\mathcal{F}_{t-1}$ represents the information set at time $t-1$.

2. $\text{Var}(a_t) = \dfrac{\alpha_0}{1-\alpha_1}$, which gives the unconditional variance of a_t, is obtained as below.

$$\text{Var}(a_t) = E(a_t^2) = E\left[E\left(a_t^2|\mathcal{F}_{t-1}\right)\right] = E\left[\alpha_0 + \alpha_1(a_{t-1}^2)\right] = \alpha_0 + \alpha_1 E(a_{t-1}^2)$$

Due to stationarity of a_t, $E(a_t) = 0$, $\text{Var}(a_t) = \text{Var}(a_{t-1}) = E\left[a_{t-1}^2\right]$

$$\begin{aligned} \therefore \text{Var}(a_t) &= \alpha_0 + \alpha_1 \text{Var}(a_t) \\ \text{Var}(a_t) - \alpha_1 \text{Var}(a_t) &= \alpha_0 \\ \text{Var}(a_t)\{1-\alpha_1\} &= \alpha_0 \\ \text{Var}(a_t) &= \frac{\alpha_0}{1-\alpha_1} \end{aligned}$$

and since it must be positive, $0 \leq \alpha_1 < 1$

3. Higher moments of a_t must also exist to study skewness and kurtosis.

The evolution through time of the conditional variance of a_t is governed by the previous values of a_t like is the case in an $AR(p)$ model as shown by the identity below. If the sigma field generated by historical information until time $t-1$ is denoted by $\mathcal{F}_{t-1} = (a_{t-1}, a_{t-2}, ...)$, then,

$$E(a_t^2|\mathcal{F}_{t-1}) = E(\sigma_t^2 \varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2 E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2.$$

Conditions need to be imposed on the coefficients so as to obtain a properly defined process for (3.20). To ensure that $\sigma_t^2 \geq 0$ and a_t is well defined, one sufficient condition is

$$\alpha_i \geq 0 \text{ for } i = 1, \ldots, p \text{ and } \alpha_1 + \cdots + \alpha_p < 1.$$

To test for conditional heteroscedasticity or ARCH effects, the squared series a_t^2 is used. To this end, two tests can be used,

(i) The Ljung-Box (LB) test (Ljung and Box, 1978), which tests the null hypothesis that all the autocorrelations of the first p lags of the series $\{a_t^2\}$ are zero. The lag p

LB Q-test statistic is given as:

$$Q(p) = n(n+2)\sum_{i=1}^{p}\frac{\hat{\rho_i^2}}{n-i}, \tag{3.21}$$

where ρ_i represents the estimated autocorrelation of the series at the i^{th} lag, and p denotes the number of lags being tested. The LB statistic$Q(p)$ will follow an asymptotic χ^2 distribution with p degrees of freedom if the data are white noise then. A statistically significant value of $Q(p)$ provides evidence that the conditional volatility varies with time. The LB test rejects the null hypothesis at an α level of significance if $Q(p) > \chi^2_{1-\alpha,p}$

(ii) The Lagrange multiplier (LM) test (Engle, 1982), which is similar to applying the usual F statistic that tests the null hypothesis that $\alpha_i = 0$ for $i = 1, \ldots, p$ in a linear regression

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \ldots + \alpha_p a_{t-p}^2 + e_t \text{ for } t = p+1, \ldots, T$$

where T is the sample size, p is a pre-specified positive integer and e_t is the error term. The F statistic is computed as

$$F = \frac{(SSR_0 - SSR_1)/p}{SSR_1/(T-2p-1)} \text{ where, } SSR_0 = \sum_{t=p+1}^{T}\left(a_t^2 - \bar{a}\right) \text{ and, } \bar{a} = \frac{1}{T}\sum_{t=1}^{T} a_t^2$$

and F is asymptotically distributed as χ^2 with p degrees of freedom.

If significant ARCH effects are evident in the time series, the order of ARCH can be determined using the $PACF$ of a_t^2, since, a_t^2 is an unbiased estimator of σ_t^2, and, a_t^2 is expected to be linearly related to $a_{t-1}^2, \ldots, a_{t-p}^2$, in a manner similar to an $AR(p)$ process.

GARCH Models

Despite their simplicity, in order to adequately represent the asset return volatility dynamics, ARCH models often require several parameters. Bollerslev (1986) sought an

alternative to this by generalising the $ARCH(p)$ model to the $GARCH(p,q)$ such that:-

$$a_t = \sigma_t \varepsilon_t \text{ where } \varepsilon_t \sim N(0,1) \textbf{ and } \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2. \tag{3.22}$$

where $\alpha_0, \ldots, \alpha_p$ and $\beta_1, \ldots, \beta_q$ are constant parameters. The model also requires that conditions be imposed on $\alpha's$ and $\beta's$ for (3.22) to be well defined. These conditions are that $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ for the unconditional variance to exist, and $\left(\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j\right) < 1$, for the unconditional variance to be finite. For $i > p$ and $j > q$, it should be noted that $\alpha_i = 0$ and $\beta_j = 0$ respectively.

The GARCH model is just an ARCH model of infinite order, and the effects of shocks on the present day volatility decline over time following a geometric fashion. σ_t^2, which is the conditional variance of p_t, given the available news up to time $t-1$ has an autoregressive structure and is correlated positively to it's own recent past and the recent past values of the squared innovations a^2. This implies that the conditional variance or volatility is "persistent", and that, small (large) values of a_t^2 are likely to be followed by small (large) values. It can be shown that $E(a_t^2) = \dfrac{\alpha_0}{1 - \left(\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j\right)}$ provided the denominator remains positive. The standard GARCH model can also be more conveniently represented using the back-shift operator as shown in (3.23)

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha(B)\, a_t^2 + \beta(B)\, \sigma_t^2 \text{ where} \\ \alpha(B) &= \sum_{i=1}^{q} \alpha_i B^i \text{ and } \beta(B) = \sum_{j=1}^{p} \beta_j B^j \text{ and} \end{aligned} \tag{3.23}$$

$\alpha(B)$ and $\beta(B)$ are polynomials of degrees q and p. In standard form, it is assumed that the GARCH model driven by normally distributed innovations. Empirical studies have shown the success of the family of GARCH models. From the survey of Bollerslev et al. (1992), it is noted that in most cases, of these models, the $GARCH(1,1)$ is most preferred since adopting lower orders for the lags p and q seems sufficient in modelling variance dynamics.

In the basic ARCH model shown in (3.20) as well as the GARCH model in (3.22), we see that only the squared residuals a_{t-j}^2 enter the equation. This implies that directions of the shocks or residuals do not affect the conditional volatility, only the magnitude does. The standard GARCH assumes that both good and bad news have equal effects on the volatility. In reality however, good news (positive shocks) tends to have a smaller impact on volatility as compared bad news (negative shocks). Black (1976) attributed this to the fact that negative news tends to push down stock prices, hence increasing the leverage (debt-equity ratio) which results in an increase in price volatility.

After the introduction of ARCH and GARCH models, researchers have suggested a number of modifications and extensions. One of the main reasons for considering GARCH model extensions is to achieve greater parsimony. The leverage effect also known as the asymmetric news impact, is for example not captured by the standard GARCH model. For example if a_t is a $GARCH(1,1)$ then $\sigma_t^2 = \alpha_0 + \alpha a_{t-1}^2 + \beta\sigma_{t-1}^2$. Suppose $\alpha_1 + \beta_1 = 1$, then it means there exists a unit root in the AR polynomial of the GARCH model representation, and hence the underlying process a_t is no longer stationary and we have an integrated GARCH (IGARCH)(1,1), which can be given by,

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t \text{ where } \varepsilon_t \sim N(0,1) \text{ and} \\ \sigma_t^2 &= \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1-\beta_1)\, a_{t-1}^2. \end{aligned} \tag{3.24}$$

For the $IGARCH(1,1)$ model, volatility is persistent since it is not covariance stationary. This means the impact of past squared shocks is prolonged. Although the $IGARCH(1,1)$ model strongly resemblance a random walk model, extreme caution is necessary for such an analogy since the autocorrelations still decay exponentially unlike the autocorrelations of a random walk model which are approximately equal to one. Empirically, the volatility reacts asymmetrically to the direction (sign) of the shocks, and as a result, several extensions of the GARCH model aim at accommodating this asymmetry in the variance response. Due to these considerations, many GARCH model generalizations that can cap-

ture some of these phenomena have been proposed. These include EGARCH, FIGARCH, TGARCH, GJR-GARCH, and NGARCH models, among others. The Gaussian assumption of ε_t facilitates the estimation process, but it is also not crucial. It can be relaxed to accommodate heavy-tailed distributions, such as a t-distribution.

3.2.2.2 Combining ARMA and GARCH Models

If some first order autocorrelation exists in the standardized residuals, it implies that returns are correlated and as such, they do not follow a GARCH process. This problem can be remedied by considering a combination of an ARMA model to describe the conditional mean and a GARCH model to describe the conditional variance. This combined model is referred to as the ARMA-GARCH model. The application of ARMA-GARCH models in time series analysis, is a common approach in that considers volatility clustering, heteroscedasticity and autocorrelation. An $ARMA(P,Q) - GARCH(p,q)$ model for a return series r_t, is specified as follows:

$$\begin{aligned} r_t &= \sum_{i=1}^{P} \phi_i r_{t-i} + \sum_{i=1}^{Q} \psi_i a_{t-i} + a_t \text{ and} \\ a_t &= \sigma_t \varepsilon_t \text{ where } \varepsilon_t \sim N(0,1) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \end{aligned} \tag{3.25}$$

The possible autocorrelations of the return values are captured by the first summation term in the first level of (3.25) whereas heteroscedasticity and volatility clustering are captured in the third level of this equation. The GARCH component of this model plays an important role in estimating and forecasting volatility (Beck et al., 2013).

3.2.3 Methods for evaluating model sufficiency

The appropriateness of a chosen model is measured by the sufficiency of the model. The methods of evaluating model sufficiency can give critical guidance to the choice of appropriate models. These methods are mainly categorized into two,

1. The first category include methods that evaluate the goodness of fit for the fitted models. They include:

   (a) Root mean square error (RMSE), which is shown in (3.26). It is a function of the sum of squared errors (SSE), the number of observations n, and the number of independent variables $k \leq p+1$ where p represents the number of independent variables so that k includes the intercept.

   $$RMSE = \sqrt{\frac{SSE}{n-k}} \tag{3.26}$$

   It is calculated for all possible subset models, and, the best linear model is the model with the smallest RMSE.

   (b) Adjusted R^2 which is shown in (3.27). It reflects the proportion of variability in a data set accounted for by the corresponding statistical model. It is a function of the SSE, the total sum of squares (SST), the number of observations n, and the number of independent variables $k \leq p+1$. It's value ranges between 0 to 1 and it measures how well a model can predict the future outcome. It is a modification of the R^2 and it considers the number of explanatory variables in a fitted model.

   $$adjR^2 = 1 - \frac{(n-i)SSE}{(n-k)SST} \tag{3.27}$$

   It is calculated for all possible subset models, and, the best linear model is the model with the largest adjusted R^2. It is only if the new variable improves the model more than would be expected by chance that the adjusted R^2 increases. However, when a predictor improves the model by less than expected by chance, It decreases . However, for models with different numbers of independent variables, it cannot provide a very meaningful comparison (Liu et al., 2011).

   (c) F-test which is calculated as shown in (3.28). It is a function of the SSE, the mean square error (MSE) and the SST. It is a ratio of the explained variability

(as reflected by R^2) and the unexplained variability (as reflected by 1-R^2), each divided by the corresponding degrees of freedom. The F statistic

$$F = \frac{(SST - SSE)/k}{MSE} = \frac{R^2}{(1 - R^2)/\left[n - (k + 1)\right]} \tag{3.28}$$

where $MSE = \dfrac{SSE}{(n-1)}$. The larger the F statistic, the more useful the model.

The calculated value is compared with the critical value and we reject H_0 if $F_{critical} > F_{calculated}$. If we reject the null hypothesis, then, at least one of the predictors is linearly associated to the response. The larger the F statistic, the more useful the model.

(d) Mallows' C_p which calculated as shown in (3.29). It is a function of the full model pure error estimate σ^2, the SSE, the number of independent variables $k \leq p + 1$ and the number of observations n.

$$C_p = \frac{SSE}{\sigma^2} + 2k - n \tag{3.29}$$

Mallows' C_p is calculated for all possible subset models and the best linear model is the model with the smallest C_p.

(e) AIC and Schwarz's Bayesian Information Criterion (BIC) or (SC, or SBC): Information criteria are measures of goodness of fit or the difference between the "true" underlying model and the fit model. The AIC is is calculated as shown in (3.30) and is a function of the number of observations n, the SSE and the number of independent variables $k \leq p + 1$.

$$AIC = n. \ln\left(\frac{SSE}{n}\right) + 2k. \tag{3.30}$$

BIC is calculated as shown in (3.31), and, it is a function of the full model pure error estimate σ^2, the SSE, the number of independent variables $k \leq p+1$ and

the number of observations n.

$$BIC = n.\ln\left(\frac{SSE}{n}\right) + \frac{2(k+2)n\sigma^2}{SSE} - \frac{2n^2\sigma^4}{SSE^2}, \tag{3.31}$$

and SBC is calculated using (3.32)

$$BIC = n.\ln\left(\frac{SSE}{n}\right) + k\ln.n. \tag{3.32}$$

The "best" model is the one with the smallest AIC, BIC or SBC is since it minimizes the difference between the "true" model and the fit model.

2. The methods for the diagnostic checking of fitted models, and they include the LB statistic.

3.3 Exploratory Analysis of Energy Markets

Time series analysis is concerned with model identification, parameter estimation, diagnostic checking of model adequacy, and forecasting. In this section, we discuss the exploratory analysis of energy market data. We specifically explore various price series for both spot and futures markets from the United States of America's energy markets. The series being discussed include the daily futures and spot prices for crude oil, gasoline and distillate fuels for the period running from 2^{nd} of January 2006 to 22^{nd} of May 2015. Exploratory data analysis and model estimation and fitting are carried out using MATLAB. We analyse the historical performance of the futures and spot prices of the refiners raw material crude oil, and the refinery products gasoline and heating oil, which are traded on the NYMEX and investigate the characteristics and evolution dynamics of the energy market.

The data considered are the official daily closing prices as at 2:30 p.m. American time, for a specific delivery month for each product as listed in the NYMEX. The data was obtained from the U.S. Energy Information Administration (EIA) website http://www.eia.gov/

petroleum/data.cfm, accessed on June 25, 2015. The EIA is the principal agency of the U.S. Federal Statistical System responsible for collecting, analysing, and disseminating energy information to promote sound policy-making, efficient markets, and public understanding of energy and its interaction with the economy and the environment (EIA, 2015).

Figure 3.5 shows various time series plots for daily spot and futures contract 1 prices for Cushing OK WTI, Reformulated Blendstock for Oxygenate Blending (RBOB) gasoline and No. 1 heating oil as well as the spot price and futures price 3:2:1 crack spreads that will be considered throughout this research. From figure 3.5, we see that generally

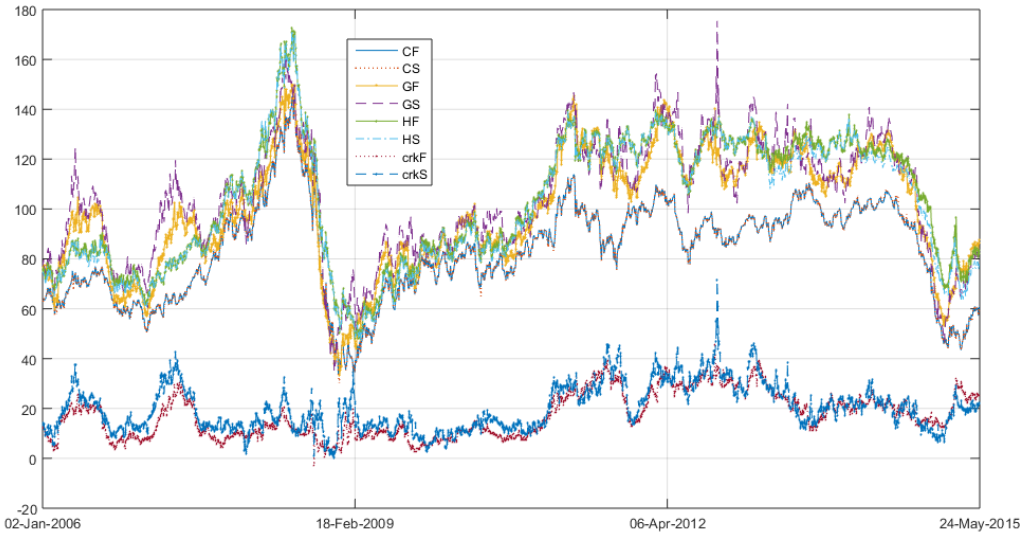

Figure 3.5: Time series plots for spot, futures prices and crack spreads in energy markets. CF=crude futures, CS=crude spot, GF=gasoline futures, GS=gasoline spot, HF=heating futures, HS=heating spot, crkF= crack spread on futures and crkS= crack spread on spot.

the series are heteroscedastic, and, they co-evolve with some spikes and jumps at various points in time. The series seem to be co-integrated and this means the volatilities from one commodity could have a spill over effect on the others. Table 3.2 shows a strong positive correlation between the spot and futures price series of all the products under consideration.

We also see that the data is definitely not stationary and hence in order to be able to

Table 3.2: Correlation analysis of energy spot and futures prices

	CF	CS	GF	GS	HF	HS
CF	1.0000	0.9997	0.9132	0.8908	0.9427	0.9417
CS	0.9997	1.0000	0.9129	0.8903	0.9421	0.9408
GF	0.9132	0.9129	1.0000	0.9662	0.9381	0.9353
GS	0.8908	0.8903	0.9662	1.0000	0.9054	0.9078
HF	0.9427	0.9421	0.9381	0.9054	1.0000	0.9970
HS	0.9417	0.9408	0.9353	0.9078	0.9970	1.0000

apply any time series models, we need to try and make it stationary. Below is a discussion of the model building process.

3.3.1 Model Building

The analysis carried out in ARIMA modelling is performed in three stages, namely, identification, estimation and diagnostic checking and forecasting.

3.3.1.1 Model Identification

In the model identification stage, a candidate ARIMA model that fits the data is identified. This begins by performance of stationarity tests to determine if differencing or de-trending is necessary. If necessary, the data is differenced, then autocorrelations, partial autocorrelations, inverse autocorrelations, and cross correlations are computed. From this kind of analysis, one or several ARIMA models that could be suggested for fitting.

3.3.1.2 Estimation of the Model and Diagnostic Checking

At this point, we fit each candidate models that has been identified and estimate it's parameters. Once the models for the conditional mean and variance have been identified from the differenced series, we employ the MLE method to fit parameters for the specified model of the differenced series. Given a sample time series consisting of n i.i.d observations$\{p_1, p_2, \ldots, p_n\}$ which follow a distribution $f(p)$ with unknown parameters θ, the joint density function is $f(p_1, p_2, \ldots, p_n|\theta) = f(p_1|\theta) \times \cdots \times f(p_n|\theta)$. By considering the observed values $p_1, p_2, \ldots, p_n$ to be fixed 'parameters' of this function and θ as the function's variable, the likelihood function can be given as $L(\theta|p_1, p_2, \ldots, p_n) =$

$f(p_1, p_2, \ldots, p_n|\theta) = \prod_{i=1}^{n} f(p_i|\theta)$. In practice, the logarithm of the likelihood function $\ln L(\theta|p_1, p_2, \ldots, p_n) = \sum_{i=1}^{n} \ln f(p_i|\theta)$ is more convenient to work with.

The MLE concept interprets the density as a function of the set of parameters, conditional on an outcome set (Wurtz et al., 2009). This method infers the process innovations by inverse filtering, so that it transforms the the observed innovations a_t into a white noise process ε_t. the log-likelihood unction is then used to infer corresponding conditional variances σ_t^2 from the inferred innovations ε_t using recursive substitution into conditional variance equations which are dependent on the model. The inferred innovations and conditional variances are finally used to evaluate the appropriate log-likelihood objective function to be optimized. By minimizing the 'negative' log-likelihood function, the parameters θ that best fit the model are obtained . Assuming the observations $\{p_1, p_2, \ldots, p_n\}$ follow a normal distribution with unknown parameters $\theta = \{\mu, \sigma^2\}$, and using $a_t = \varepsilon_t \sigma_t$, the log likelihood function is given by

$$\begin{aligned} \ln L(\theta) &= \ln \prod_{n}^{i=1} \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{a_t^2}{2\sigma_t^2}} \\ &= -\frac{1}{2} \sum_{i=1}^{n} \left[\ln(2\pi) + \ln(\sigma^2) + \varepsilon_t^2 \right] \end{aligned} \tag{3.33}$$

Then, diagnostic statistics which help to judge the model adequacy are produced. Significance tests on parameter estimates indicate which parameters in the model are unnecessary. Goodness-of-fit statistics are used to compare the fit model to other possible models. Tests on the residuals show whether there could be any additional information in the residual series, that might be useful for more complex analysis. If a particular model is found to be problematic by the diagnostic tests, another model is tried, and the whole estimation and diagnostic checking process is repeated.

3.3.1.3 Forecasting

The appropriate model fitted is then used to predict future values of the fitted time series model and to produce confidence intervals for these predictions.

3.3.2 Modelling Energy markets using ARIMA models

In this section, we model the U.S. energy data using ARIMA models. This is useful in establishing the conditional mean equation specification, which can be given by, $E_{t-1}[y_t]$, the expectation conditioned on available information at time $t-1$. Typically it takes the form

$$E_{t-1}[y_t] = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \psi_j a_{t-j}, \tag{3.34}$$

where ϕ_0 is a constant and all the other variables are as explained for (3.15).
Depending on the type of data and frequency of the data set used, the conditional mean can be typically specified as zero, a constant or even a low order ARMA process in the case where there is autocorrelation due to say, the market micro-structure or any other effects not related to trading. In cases of unusual or extreme happenings in the market during the period being considered, dummy variables that are associated with these happenings are included in the specification of the conditional mean equation in order to take care of the effects of these happenings. In this case, the conditional mean specification would typically take the form

$$E_{t-1}[y_t] = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \psi_j a_{t-j} + \sum_{k=0}^{K} \beta_k' X_{t-k}, \tag{3.35}$$

where X_t is an $m \times 1$ vector of the exogenous explanatory variables.

Table 3.3 gives a summary of the descriptive statistics along with the Jarque-Bera (JB) test for normality for the data under consideration. The null hypothesis for the JB test for normality is that the data are i.i.d as normal, implying the skewness and excess kurtosis are both zero; against the alternative that the data do not conform to the normal distribution. The JB statistic is asymptotically distributed as χ^2 with 2 degrees of freedom under the

null hypothesis. (3.36) shows the representation of the JB test statistic.

$$JB = \frac{n}{6}\left(\widehat{S}^2 - \frac{(\widehat{K}-3)^2}{4}\right), \tag{3.36}$$

where $\widehat{S}$ denotes the sample skewness given by $S = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{n\left(\hat{\sigma}^2\right)^{3/2}}]$ and $\widehat{K}$ denotes the sample kurtosis given by $K = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^4}{n\left(\hat{\sigma}^2\right)^2}$.

From the JB statistics summarised in Table 3.3 against the JB critical value of 5.9681, the null hypothesis of normality is rejected for all the differenced series under consideration. Part of the non-normality is caused by the jumps and clusters that can be witnessed in Figures 3.6, 3.7 and 3.8. From the measure of skewness, we can see, all the differenced series are skewed to the left, and from the measure of kurtosis, these differenced series are all leptokurtic, and these are some of the stylised facts of financial time series data. The data therefore doesn't conform to the normal distribution.

Table 3.3: Descriptive statistics for energy spot and futures prices

	ΔCF	ΔCS	ΔGF	ΔGS	ΔHF	ΔHS
Mean	-0.0014	-0.0018	0.0048	0.0025	0.0028	0.0010
Median	0.0350	0.0400	0.0420	0.0420	0	0
Mode	0.1400	-0.2500	-0.7140	0.3360	0	0
Maximum	16.3700	18.5600	13.0200	21.3360	12.3060	11.5080
Minimum	-14.3100	-14.7600	-17.7240	-21.7560	-17.3040	-10.6680
Skewness	-0.0508	0.0415	-0.4559	-0.1728	-0.4672	-0.0898
Kurtosis	10.2522	12.0458	8.0076	8.3115	8.6993	5.8373
SE	0.0007	0.0008	0.0009	0.0012	0.0008	0.0008
JB	5177.1	8053.7	2549.7	2788.3	3282.7	795.4796

Δ denotes differenced series.

Figures 3.6, 3.7 and 3.8 show time series plots of the differenced series for the energy prices. The modified series look somewhat mean stationary, though there is evidence that they are not variance stationary. Despite the fact that the series look mean stationary, this has to be confirmed by testing for stationarity in the means of these series. Volatility

clustering can also be observed from the series.

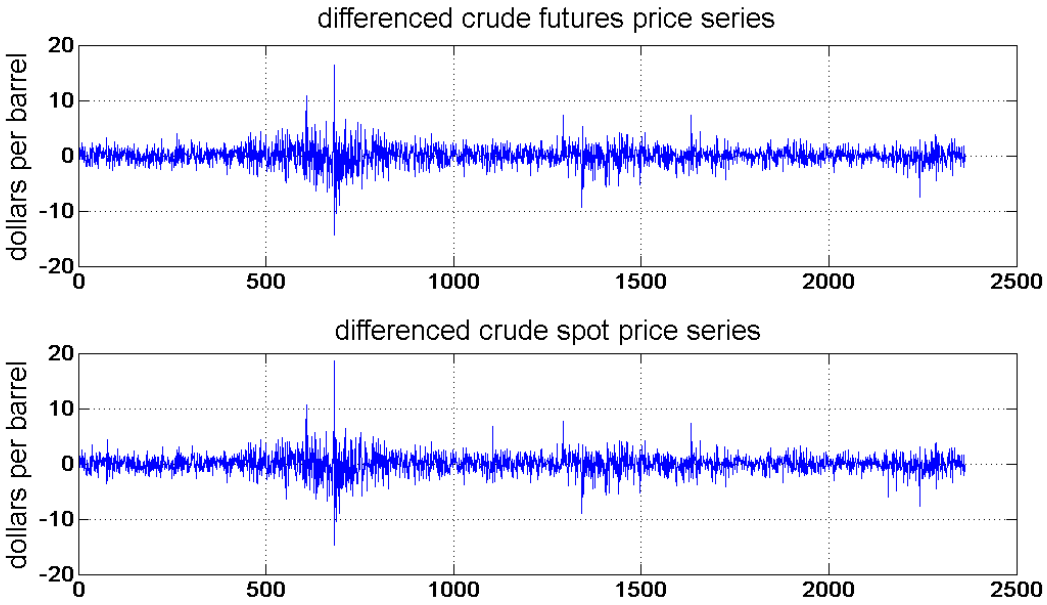

Figure 3.6: Differenced series for crude futures and spot prices

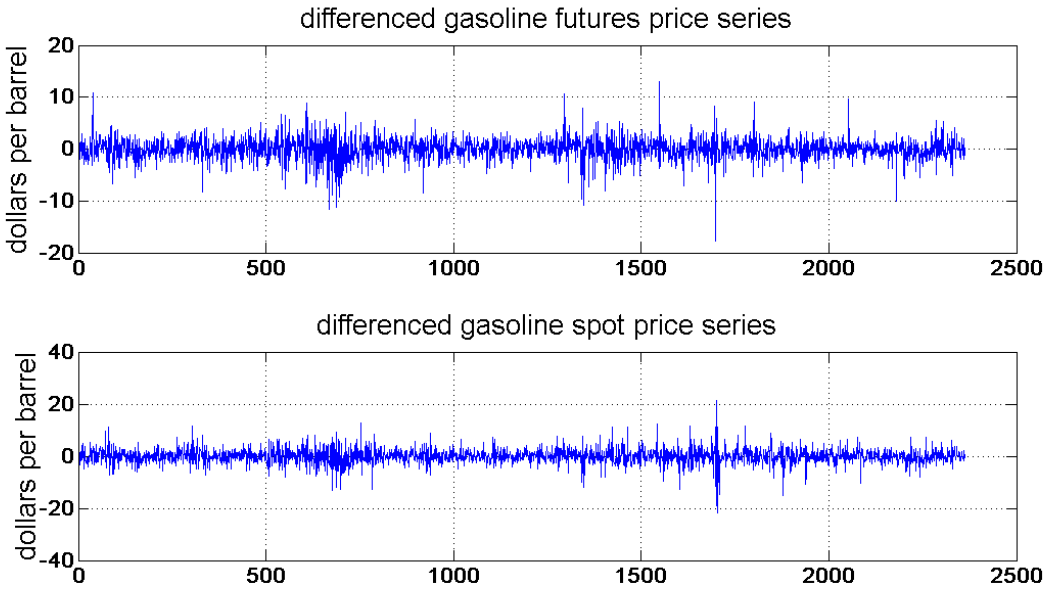

Figure 3.7: Differenced series for gasoline futures and spot prices

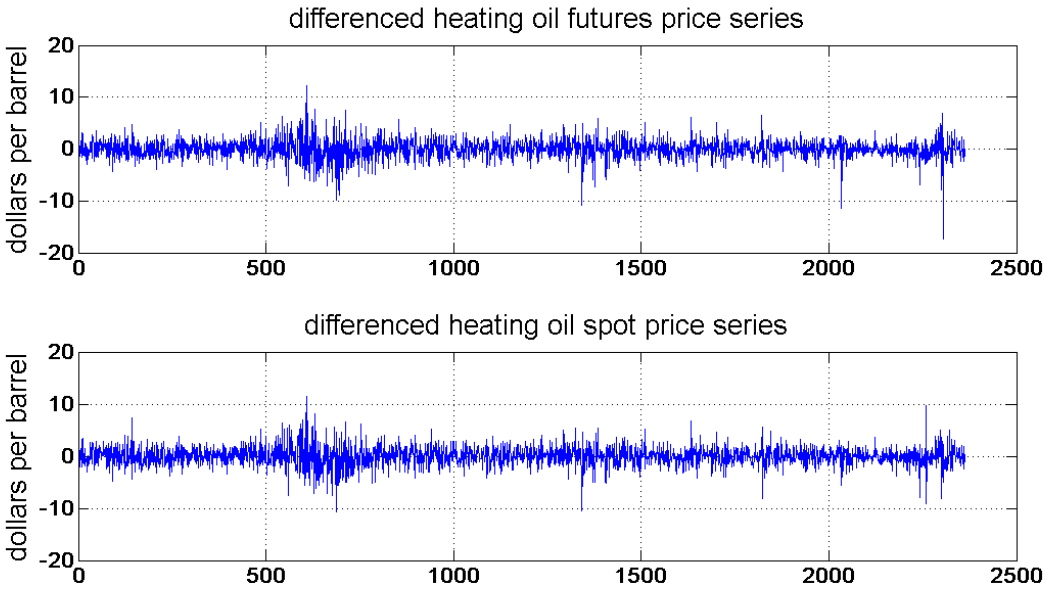

Figure 3.8: Differenced series for heating oil futures and spot prices

3.3.2.1 Test for Stationarity

After differencing the data, the new differenced series have to be tested for stationarity, and in this case, the DF test is used. Dickey and Fuller (1979) developed a method to test if the variable follows a random walk, or equivalently, the null hypothesis that a variable contains a unit root. For a given series $p_t = \phi p_{t-1} + a_t$ for $t = 1, 2, \ldots$, $p_0 = 0$, ϕ is a real number, and $a_t \sim N(0, \sigma^2)$, the following can be observed:

(i) As $t \to \infty$, the series $\{p_t\}$ will converge to a stationary time series if $|\phi| < 1$.

(ii) As $t \to \infty$, the series $\{p_t\}$ will not converge to a stationary time series if $|\phi| = 1$. $\{p_t\}$ will be a homogeneous non-stationary time series and it's variance will be $t\sigma^2$. A time series with $\phi = 1$ is a random walk.

(iii) As $t \to \infty$, the series $\{p_t\}$ will not converge to a stationary time series if $|\phi| > 1$. $\{p_t\}$ will be explosive non-stationary, and the variability of the time series will grow exponentially as time t increases.

The DF test states the hypothesis that

$$\begin{aligned} H_0 &: |\phi| = 1 \Rightarrow p_t \sim I(1) \\ H_a &: |\phi| < 1 \Rightarrow p_t \sim I(0). \end{aligned} \tag{3.37}$$

If the DF test equation exhibits serial correlation, i.e., if the true model is not an $AR(1)$) process, then, an $AR(p)$ process is used in order to get rid of this serial correlation, and hence the augmented Dickey-Fuller (ADF) test is applied and the hypotheses remain the same.

Applying OLS method and finding the estimator for ϕ, the test statistic is given by

$$t_{\phi=1} = \frac{\hat{\phi} - 1}{s.e\left(\hat{\phi}\right)}. \tag{3.38}$$

The test is a one-sided left tailed test. Table 3.4 gives the summary report of the DF test on the six series under consideration. From Table 3.4, we can see that all the differenced

Table 3.4: Results from the Dickey-fuller test on the six series

Series	$p-$ value	Test statistic	Critical value	Decision
ΔCF	0.001	-51.2762	-1.9416	Reject H_0
ΔCS	0.001	-50.9918	-1.9416	Reject H_0
ΔGF	0.001	-49.5626	-1.9416	Reject H_0
ΔGS	0.001	-50.9792	-1.9416	Reject H_0
ΔHF	0.001	-50.0207	-1.9416	Reject H_0
ΔHS	0.001	-50.6481	-1.9416	Reject H_0

series are stationary and hence can be subjected to further time series analysis.

3.3.2.2 Model Identification and Fitting

Once the stationary form of the series has been established, the next step is establish the form of the $ARMA(p, q)$ model for that stationary series. This is done by visually inspecting the ACF and $PACF$ plots of the stationary series. Through this inspection, the numbers of AR and/or MA terms that are necessary for correcting any autocorrelation

that remains after differencing can tentatively be identified. An ACF plot displays the coefficients of correlation between a time series and it's own lags, while a PACF plot is a plot of the partial correlation coefficients between a time series series and it's own lags itself. The partial autocorrelation at lag k is the autocorrelation between $\{p_t\}$ and $\{p_{t-1}\}$ not accounted for by the previous lags 1 up to $k-1$. The effects of any correlation due to the shorter lags are removed.

Figures 3.9 and 3.10, show the ACF and the PACF plots of the crude futures and spot price series. From these plots, it can be seen that it is not very easy to tell the autoregressive or moving average orders of the series. Exponential decay is not clearly visible in either plots and we can see significant lags up to 39 of the 40 plotted. Figures 3.11 and 3.12 show the ACF and PACF plots for the differenced gasoline futures and spot price data. We still see that by merely inspecting the correlogram, we may not be adequately able to get the model that describes this data accurately. There seems to be some decay but then there are surprises of significance at larger lags after some kind of exponential decay. Figures 3.13 and 3.14 are the ACF and PACF plots for the heating oil price series for both the futures and the spot market. They also behave in much the same way as the previous series.

For all the differenced series, when the LB test is carried out, the null hypothesis that all the autocorrelations are zero, is rejected for the crude and gasoline series, and not rejected for the heating oil series as per the results presented in Table 3.5. This implies that a conditional mean model is necessary for the crude and gasoline series, but for the heating oil series, a conditional mean model is not necessary. We therefore start exploring the best conditional mean model for the crude and gasoline series. Since from the ACF and PACF plots, it is not possible to distinctly classify the models for all the series in question, we resort to fitting various models and picking the most appropriate model based on the value of the AIC.

Various ARMA models are fitted for $p = 40$ and $q = 40$ lags, using the Matlab command

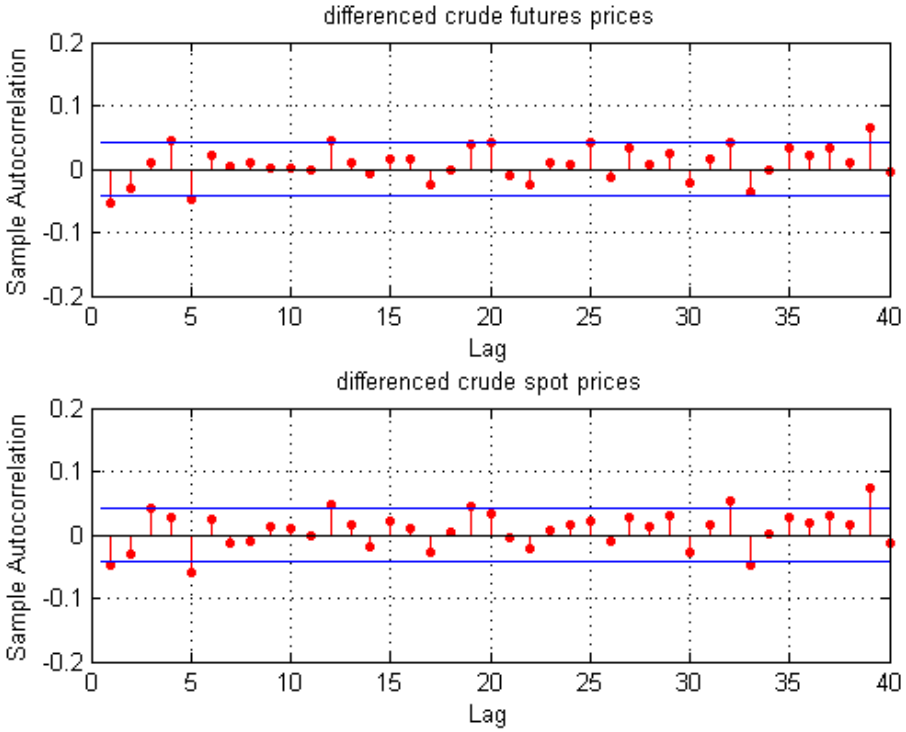

Figure 3.9: ACF plots for the differenced crude futures and spot price series

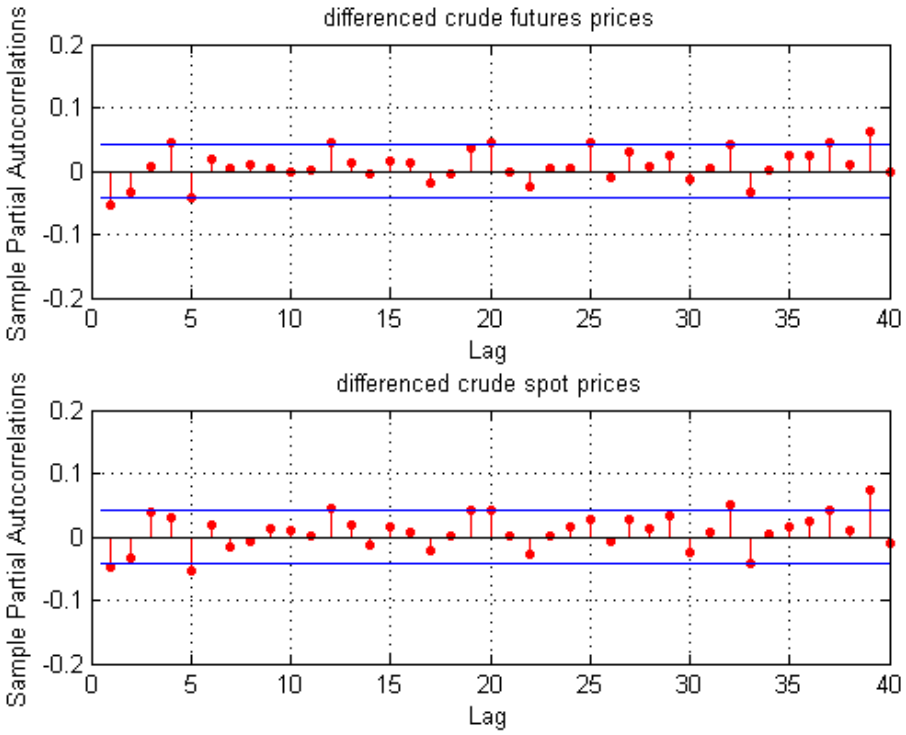

Figure 3.10: PACF plots for the differenced crude futures and spot price series

armaxfilter found in Kevin Sheppard's GARCH Toolbox for Matlab accessed from his website (https://www.kevinsheppard.com/MFE_Toolbox) in May 12, 2015. The AICs

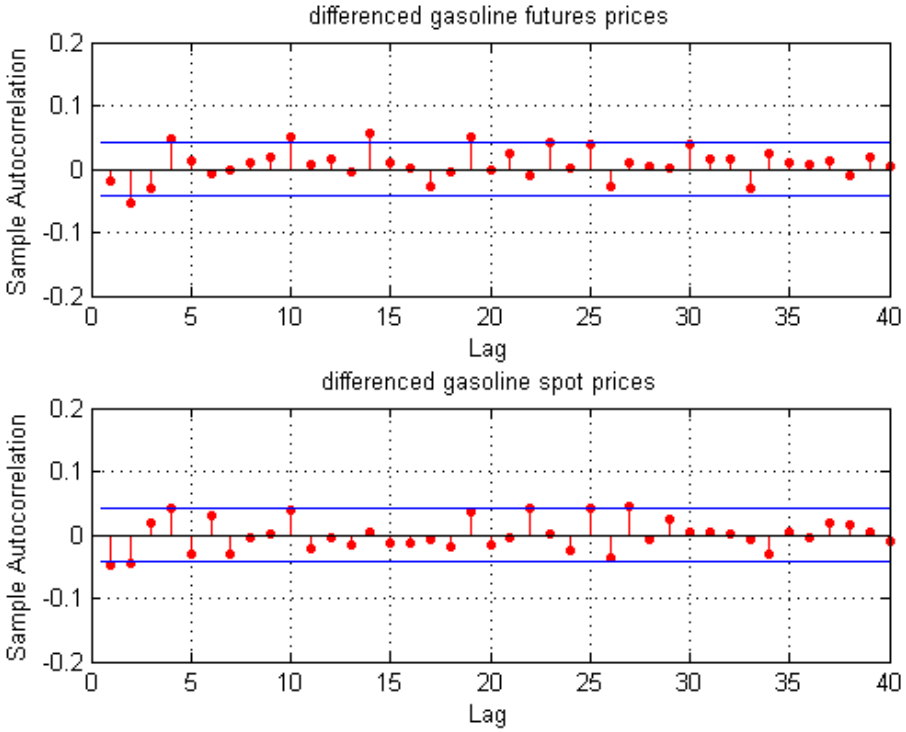

Figure 3.11: ACF plots for the differenced gasoline futures and spot price series

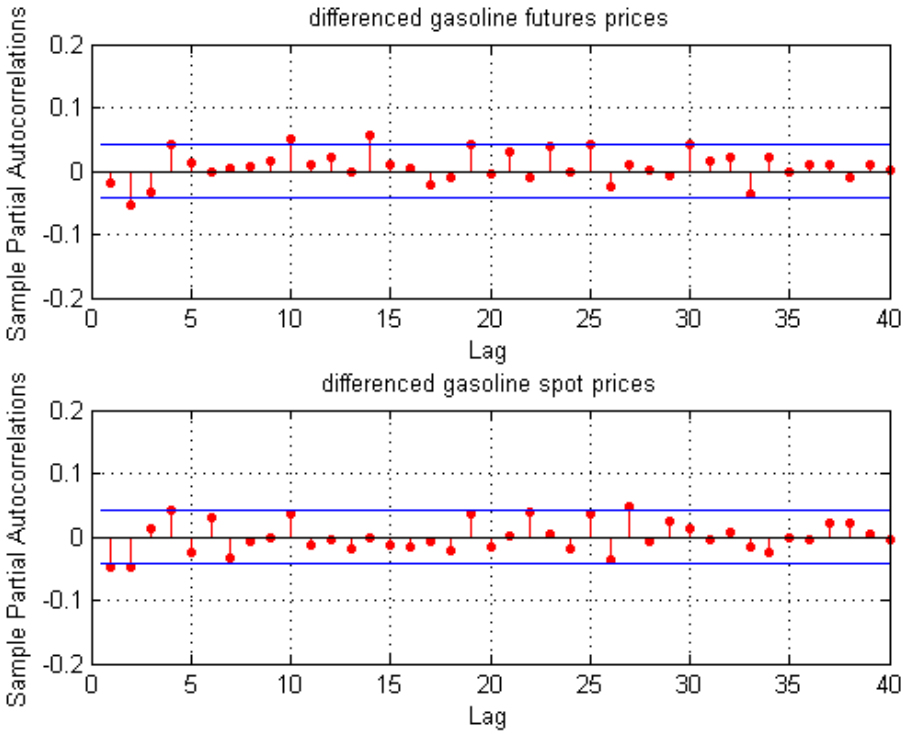

Figure 3.12: PACF plots for the differenced gasoline futures and spot price series

for the models are compared and the best model is selected as that with the smallest AIC. The Matlab code used for model selection can be found in Appendix A

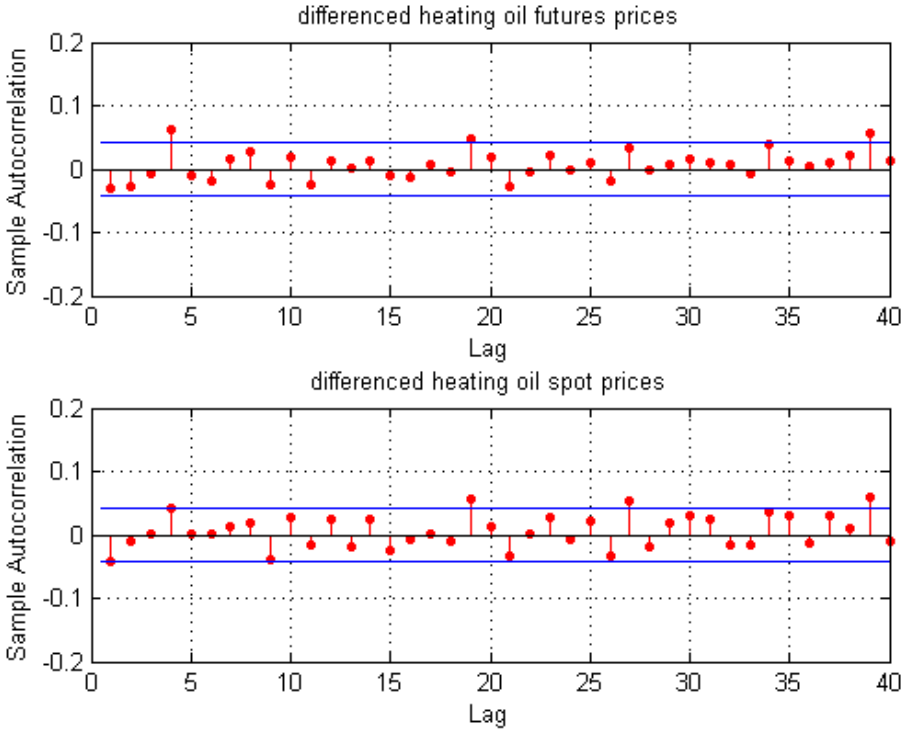

Figure 3.13: ACF plots for the differenced heating oil futures and spot price series

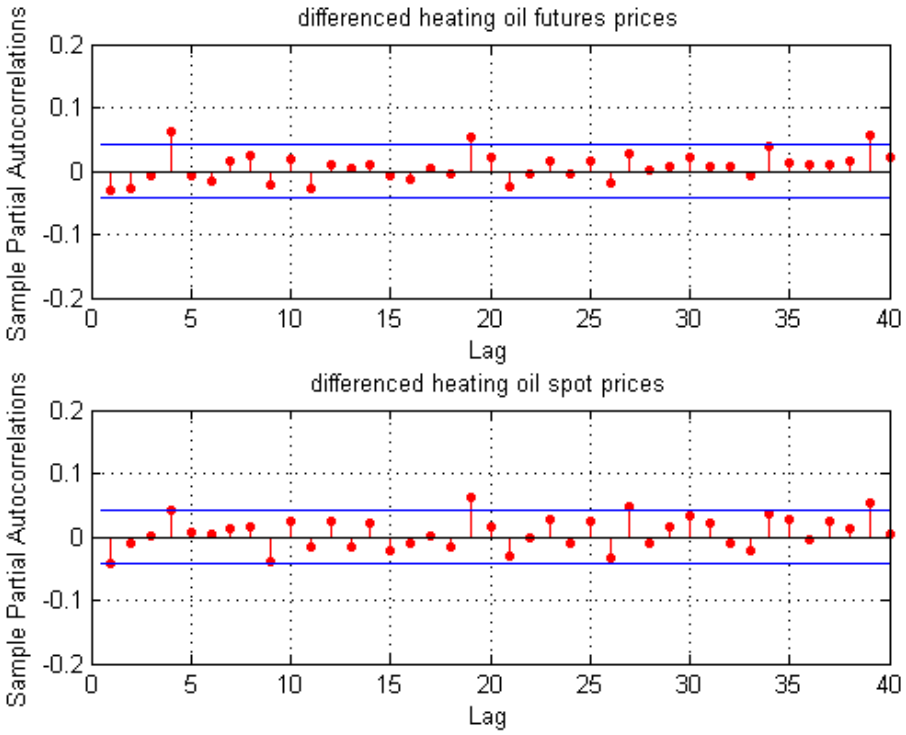

Figure 3.14: PACF plots for the differenced heating oil futures and spot price series

For the differenced series of the crude futures prices ΔCF, an excerpt of the various AICs

Table 3.5: Results from the Ljung-Box test on the six series

Series	$p-$ value	Test statistic	Critical value	Decision
ΔCF	0.0132	36.5559	31.4104	Reject H_0
ΔCS	0.0035	41.1845	31.4104	Reject H_0
ΔGF	0.0065	39.0942	31.4104	Reject H_0
ΔGS	0.0307	33.3769	31.4104	Reject H_0
ΔHF	0.0939	28.6954	31.4104	Do not reject H_0
ΔHS	0.0760	29.6954	31.4104	Do not reject H_0

for the fitted models are shown in Table B.1, and plotted on the graph in Figure 3.15. From these AIC values, and as can be seen from Table B.1 and the lowest point of figure 3.15, the best model is $ARMA(6, 11)$.

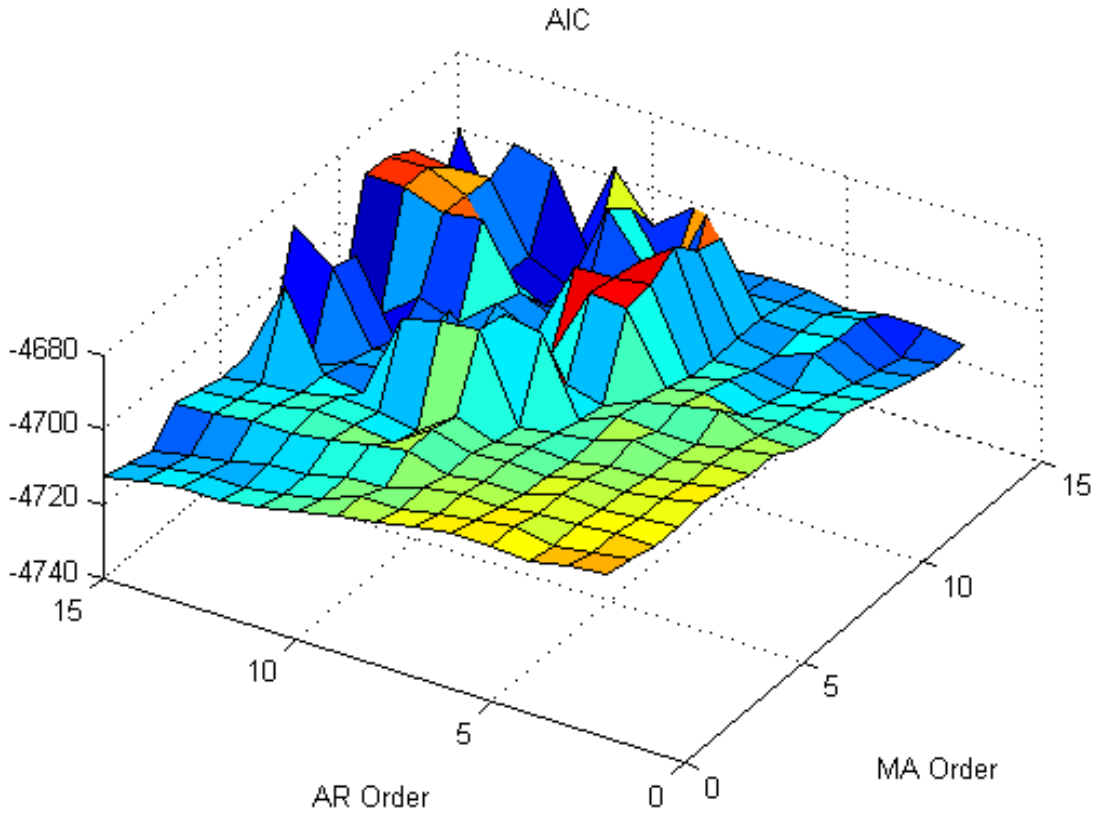

Figure 3.15: Plot for the AICs for the various ARMA models fitted for the differenced crude futures price data

From these fits, the candidate model for ΔCF is the $ARMA(6, 11)$ with an AICs of -4684.1. This model has the parameters estimated via the maximum likelihood using the Matlab Command *arima*, summarized in Table 3.6.

Table 3.6: Results from the $ARMA(6, 11)$ Model fitted to the differenced crude futures price series:

Parameter	Estimate	Standard Error	$t-$ Statistic
Constant	-0.0003	0.0055	-0.0494
AR(1)	0.1064	0.2303	0.4622
AR(2)	0.5423	0.2234	2.4273
AR(3)	0.0277	0.0757	0.366
AR(4)	-0.7031	0.0729	-9.652
AR(5)	0.0806	0.2265	0.3558
AR(6)	0.8426	0.2163	3.896
MA(1)	-0.1625	0.2304	-0.7054
MA(2)	-0.5671	0.2374	-2.3886
MA(3)	0.0127	0.0852	0.1488
MA(4)	0.7652	0.0712	10.7484
MA(5)	-0.1797	0.2365	-0.7597
MA(6)	-0.8514	0.2529	-3.3666
MA(7)	0.0874	0.0380	2.2975
MA(8)	0.0383	0.0189	2.0253
MA(9)	-0.0402	0.0185	-2.1675
MA(10)	-2.5142×10^{-5}	0.0257	-0.001
MA(11)	0.0432	0.0242	1.7835
Variance	3.069	0.0503	61.0522

From Table 3.6, we can use the values of the $t-$statistic to check the significance of the parameters. The greater the absolute value of t, the greater the evidence against the null hypothesis that there is no significance difference between the parameter estimate and 0. The closer to 0 the $t-$statistic is, the more likely the fact that there isn't a significant difference between the parameter and 0. We can use the general rule of thumb that we reject the null hypothesis whenever $|t| \geq 3$

For time series modelling, the assumption of Gaussian innovation distribution is commonly made. After model fitting, the inferred residuals are standardized and normality checks are conducted. If this assumption of normality holds, then the residuals will appear to be approximately normally distributed. Figure 3.16 shows two plots. The first one is a plot of the residual series from which we can still see some significant clustering which suggests non-normality. From this plot, we see that the variance is not constant, there is some indication that there is an excess of large residuals. The second plot of

figure 3.16 is a histogram of the residuals from which it is not very clear to see evidence of non-normality. Figure 3.17 shows a quantile-quantile plot and a box plot of the residuals and from these, we can also deduce non-normality of the residuals largely because of the extreme values.

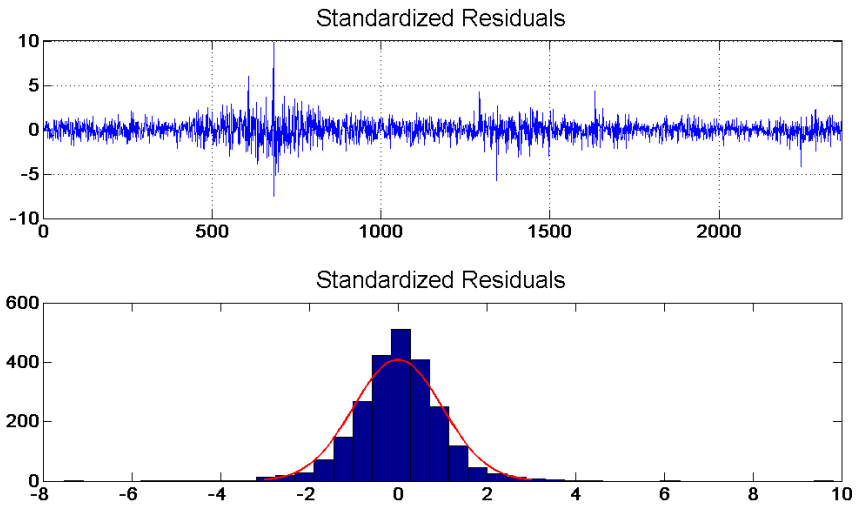

Figure 3.16: Normality plots for residual checks

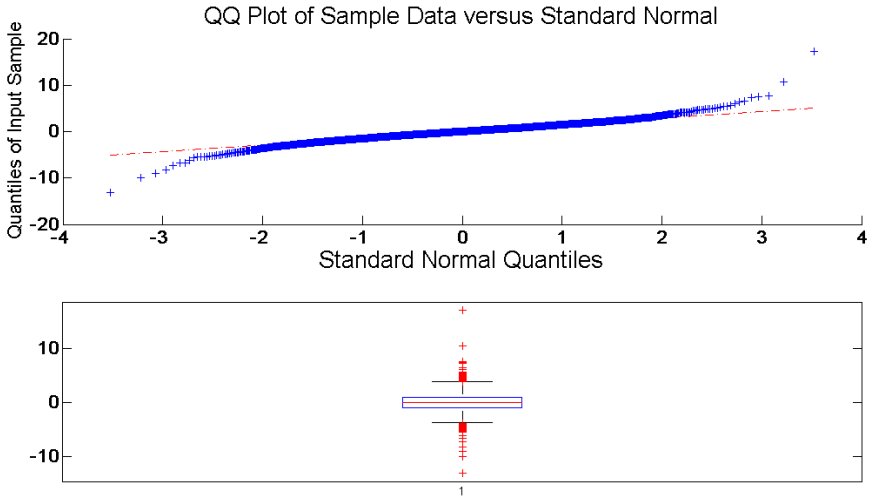

Figure 3.17: QQ-plot and Box plot for residuals from the ARMA(6,11) model

A visual inspection may therefore not be sufficient, so we apply the Kolmogorov-Smirnov (KS) test. The KS-test is a non-parametric test used to determine whether or not two data sets differ significantly. It tests the equivalence of probability distributions which are

continuous and one-dimensional, and which can be used to compare a sample with some reference probability distribution. Given two samples, $X_1, ..., X_m$ and $Y_1, ..., Y_n$, where $F_1(x)$ denotes the cumulative distribution function (CDF) for the distribution of the first population and $F_2(y)$, the CDF for the second. If $\widehat{F_1}$ denotes the CDF for the first sample, and $\widehat{F_2}$, the CDF for the second sample, then, the KS statistic, is found using the formula $KS = \max_w |\widehat{F_1}(w) - \widehat{F_2}(w)|$. The statistic is calculated by finding the maximum absolute value of the differences between the two sample CDFs. It tests the null hypothesis that both data sets come from the same distribution. Formally, the hypotheses can be stated as

$$H_0 : \text{The distributions being compared are the same}$$

$$H_1 : \text{The distributions being compared are not the same}$$

In this case, the reference distribution is the normal probability distribution. This test relies on the fact that the value of the sample CDF is asymptotically normally distributed. The KS statistic results obtained are as shown below.

KS statistic	critical value	p-value
0.0959	0.0279	2.3662×10^{-19}

From these results, the null hypothesis that the residuals are normally distributed is rejected. In this case the $ARMA(6, 11)$ that was fitted does not obey the assumption of Gaussian innovation distribution. The innovation process is also assumed to be uncorrelated. After fitting a model, the residuals are inferred and checked for any unexplained autocorrelation. The sample ACF and $PACF$ of the standardised residuals are plotted as shown in figure 3.18, and in this case, we see that the residuals appear uncorrelated.

In addition to visually inspecting the plotted ACFs, the LB Q-Statistic can be used to quantify the autocorrelation. It tests the residuals for high order serial correlation. The test can be conducted at many values of lags k. To test a residual series, there are $k-p-q$ degrees of freedom, where p and q are the number of AR and MA coefficients in the fitted

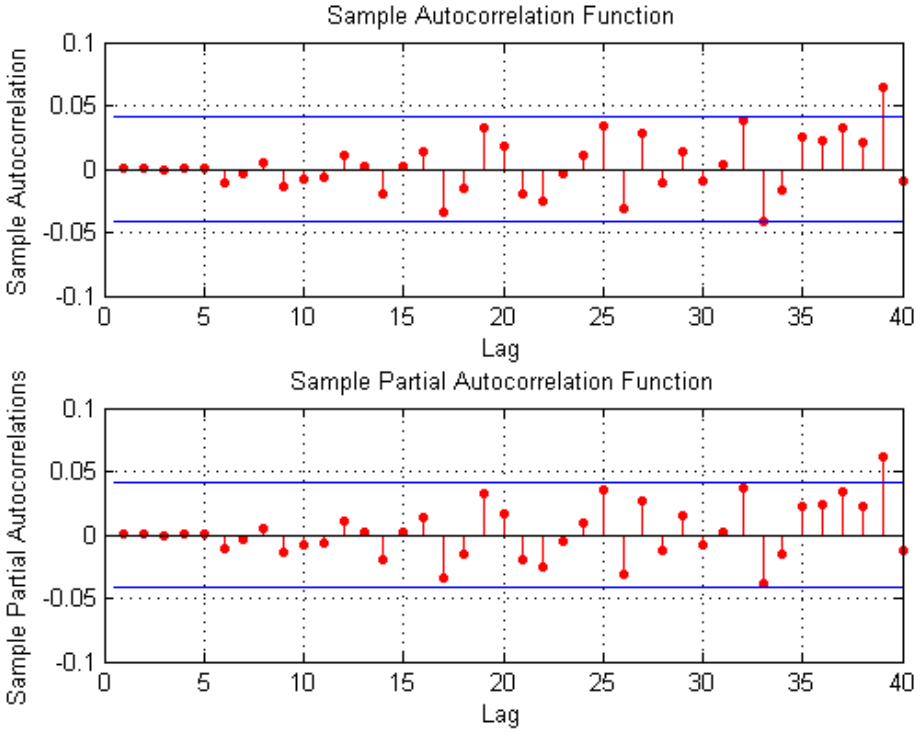

Figure 3.18: ACF and PACF for residuals from the ARMA(6,11) model

model, respectively. For the LB-test on the residual series, the results obtained are as shown below.

$LB-Q$ statistic	critical value	p-value
9.3767	31.4104	0.9781

From this test, we do not have sufficient evidence to reject the null hypothesis that the residuals are not autocorrelated. Based on these results, we test for ARCH effects in the squared residuals and obtain the following results.

$LB-Q$ statistic	critical value	p-value
862.3104	31.4104	< 0.001

The results indicate that there are significant ARCH effects in the residuals of the returns. Figure 3.19 also shows evidence of serial autocorrelation (conditional heteroscedasticity) on the squared residuals. The PACF plot shows autoregression of order one so ARCH(1) should be able to capture the ARCH effects in the residuals.

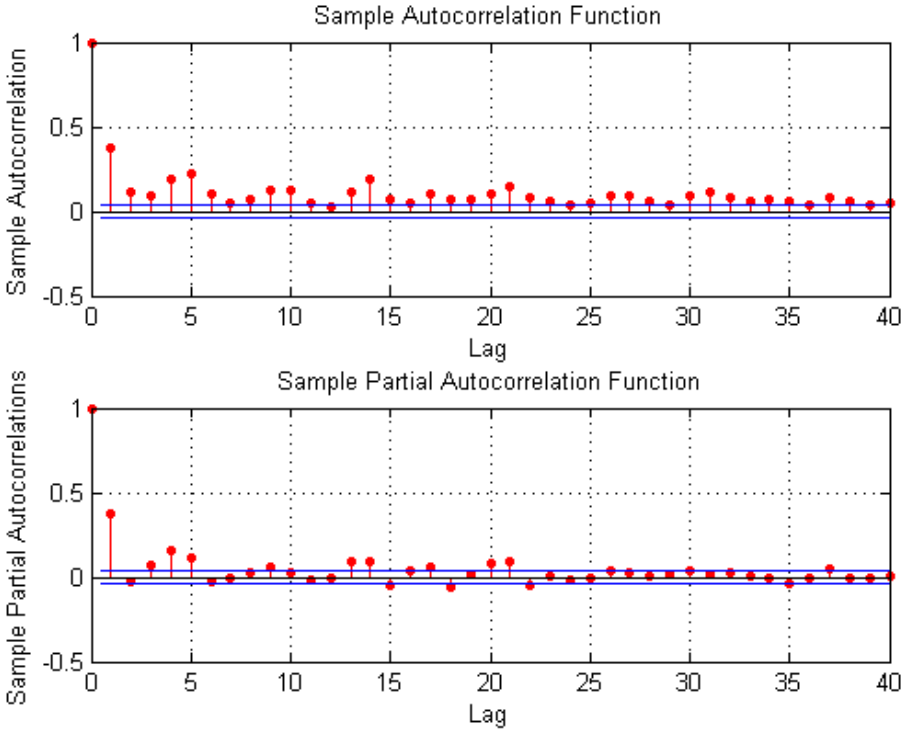

Figure 3.19: ACF and PACF for squared residuals from the ARMA(6,11) model

So we now specify the conditional mean and variance model and fit a combined ARMA-GARCH model so that we can incorporate the ARCH effects present in the variance. We fitted an ARMA(6,11)-GARCH(1,1) to capture these effects and the fit had an AIC of 8810.9. A search for the best combined model revealed that an $ARMA(0,0) - GARCH(1,1)$ model gives the best fit with an AIC 0f -4.4037. From this analysis, we see that a more appropriate model to use would be one with a constant conditional mean and a conditional variance. The most appropriate model for this would be a GARCH model.

For the crude spot market, a similar fit is done and from these AICs, it can be seen from Table C.1, and the lowest point of Figure C.1, the best model is $ARMA(7,11)$. This model has the parameters summarized in Table C.2. A similar analysis was also done on the differenced gasoline futures and gasoline spot prices and they behaved in much the same way as the crude futures data. Some of the results are presented in Figures D.1 and D.2 and Table D.1. The series under consideration here have very high order

AR and MA terms and so since modelling just the evolution of the variance gives a more parsimonious model, in the next section, all the differenced series are now only considered under GARCH in order to capture the heteroscedasticity that is evidently present.

3.3.3 Modelling Energy Markets using using GARCH Models

In the previous section we used ARMA models for the differenced series under the assumption that the variance does not evolve over time. The sample ACF and PACF plots for all the series show virtually no significant autocorrelation. The LB test null hypothesis that all autocorrelations are zero is not rejected for all the series. This implies that a conditional mean model is not necessary for this particular returns series. The analysis in the previous section has shown that the differenced series exhibit the "stylized facts" such as volatility clustering and a non-normal empirical distribution. Non-normality and volatility clustering in financial time series is typically envisaged in high frequency data such as weekly, daily or intra-day data. The persistence of the conditional volatility tends to increase with the sampling frequency.

Volatility clustering in financial time series manifests itself as autocorrelation in squared and absolute returns, or in the residuals of the estimated conditional mean equation. Examining the ACF and PACF plots for the squared differenced series shown in figures 3.20, 3.22, 3.24, 3.21, 3.23 and 3.25 reveals serial correlation of the squared differenced series. The significance of these autocorrelations at various lags was tested using the LB test and the LM test at lags 1,5 and 10, and the results are summarized in Table 3.7

As is evident from Table 3.7, there is enough evidence to reject the null hypothesis of no ARCH effects as per the LB test since the test indicates that there are significant ARCH effects in the all squared series. However, based on Engle's ARCH test, which is the LM test, we do not have sufficient evidence to reject the hypothesis of no ARCH effects for the gasoline futures and the heating oil futures series. For all the other series we reject that hypothesis of no ARCH effects.

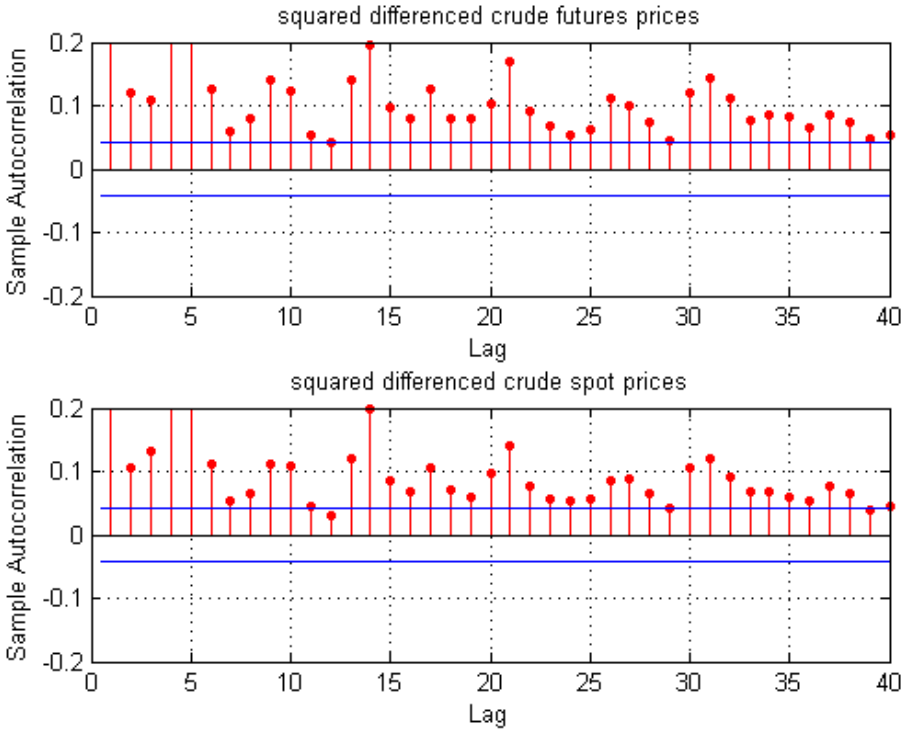

Figure 3.20: ACF plots for the squared differenced crude futures and spot price series

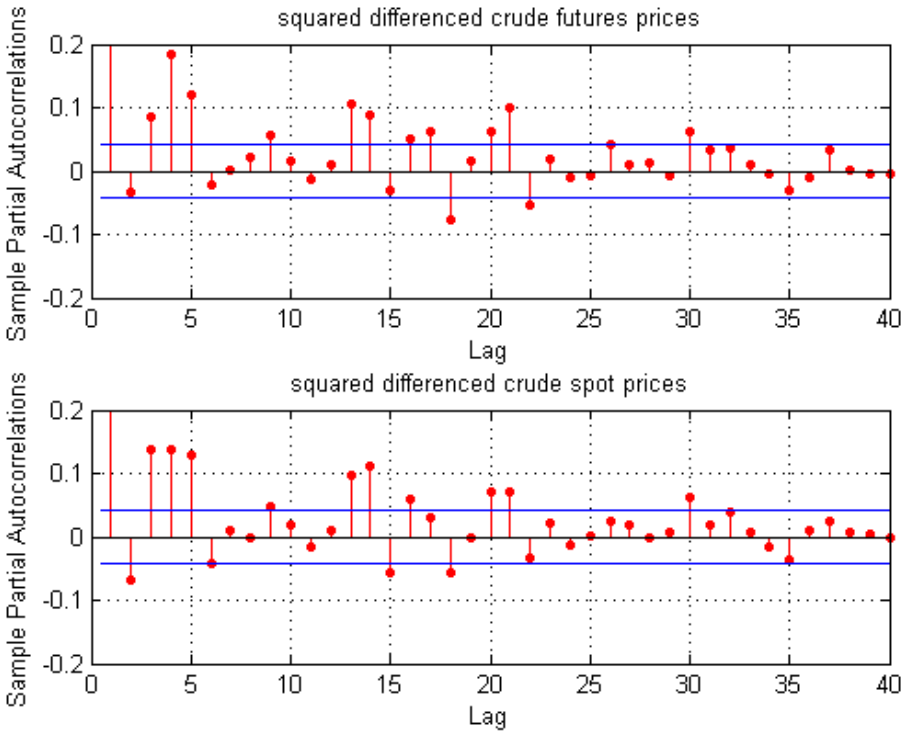

Figure 3.21: PACF plots for the squared differenced crude futures and spot price series

Serial correlation in squared returns, or conditional heteroscedasticity can be modelled

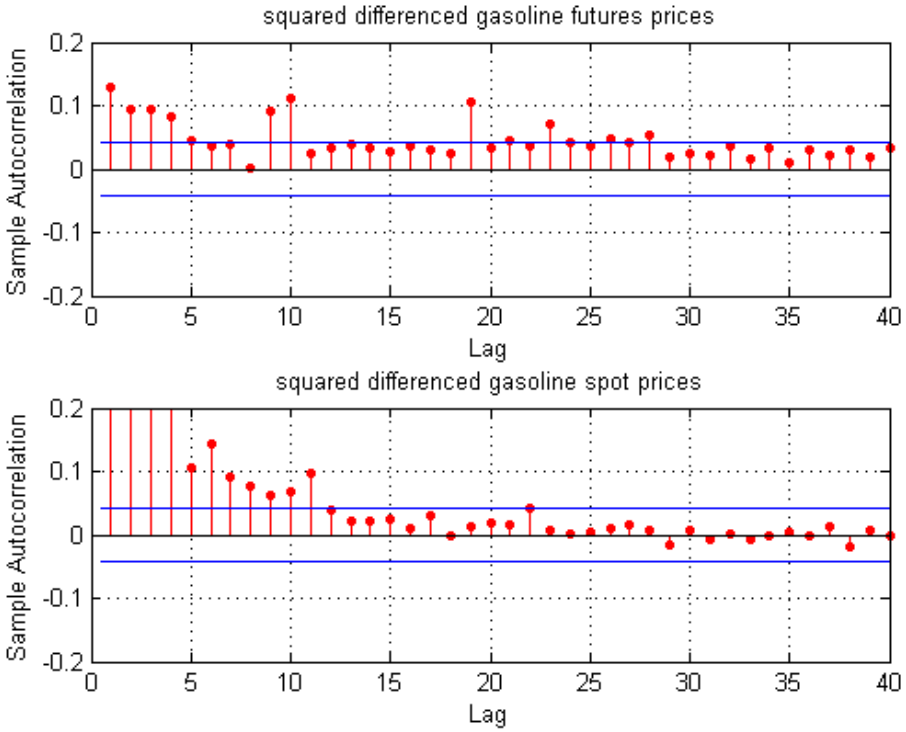

Figure 3.22: ACF plots for the squared differenced gasoline futures and spot price series

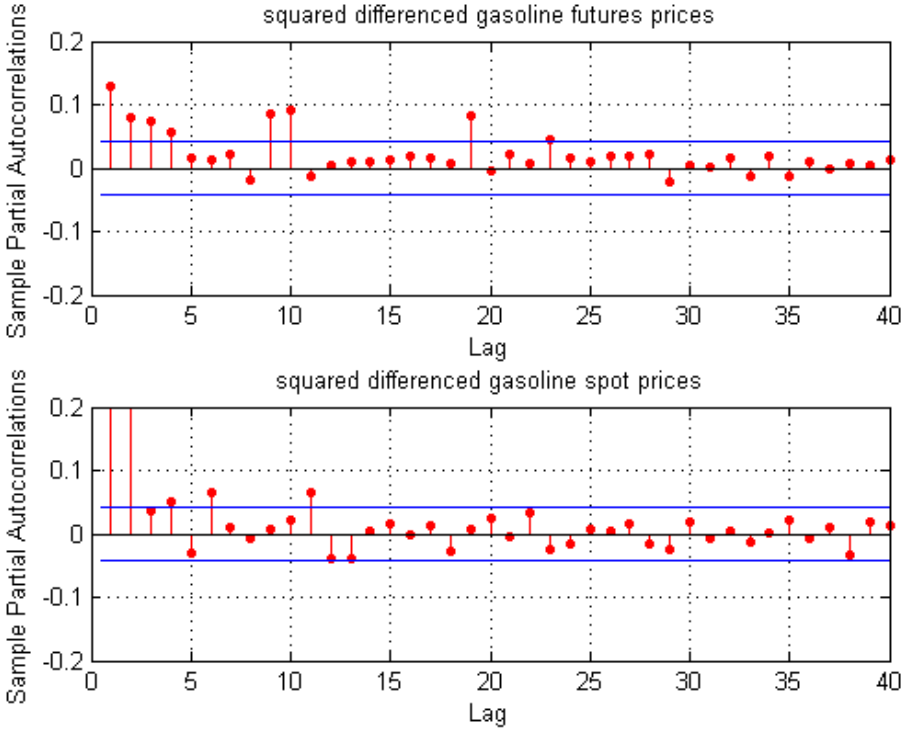

Figure 3.23: PACF plots for the squared differenced gasoline futures and spot price series

using GARCH models. GARCH models allow for the volatility to evolve with time. A GARCH model can be expressed as an ARMA model of squared residuals and hence many

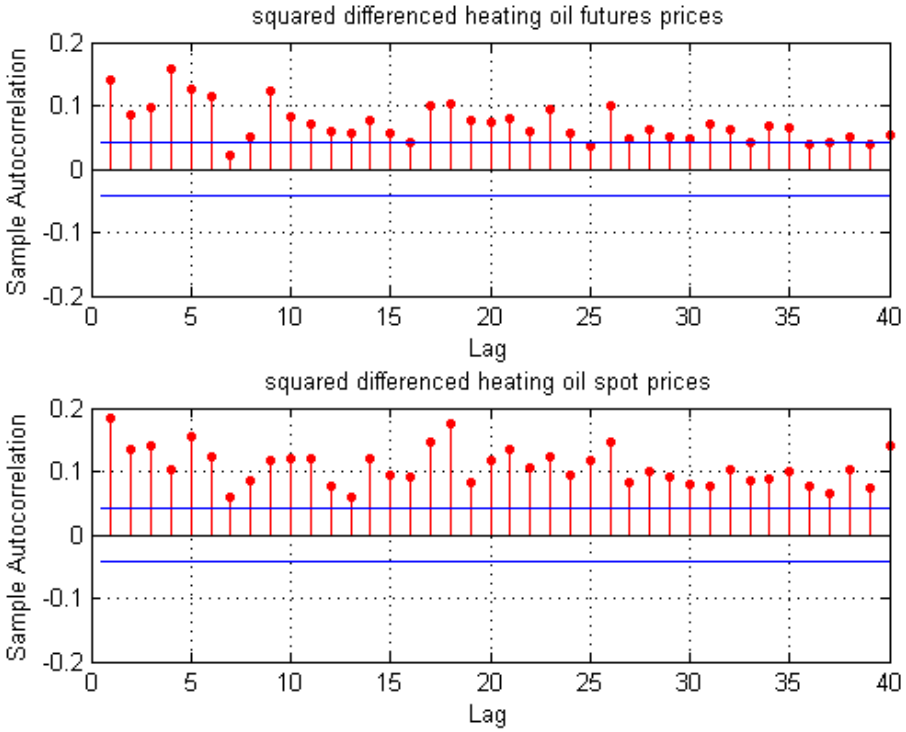

Figure 3.24: ACF plots for the squared differenced heating oil futures and spot price series

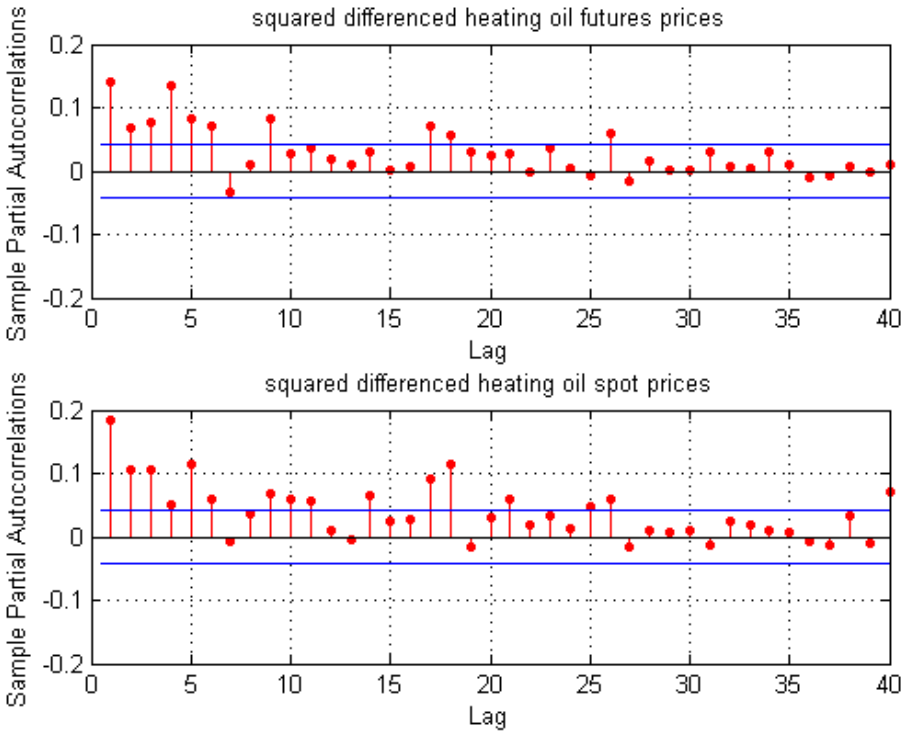

Figure 3.25: PACF plots for the squared differenced heating oil futures and spot price series

of it's properties follow easily from those of the corresponding ARMA process. Bollerslev (1986) showed that if the fourth order moment of a GARCH (1,1) exists, the kurtosis

Table 3.7: Results from the Ljung-Box test and LM test for ARCH effects on the squared differenced series

	$Q(p)$			LM -test		
Lag	1	5	10	1	5	10
$(\Delta CF)^2$	348.7860	677.9798	821.8668	431.6908	564.0678	608.2411
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
$(\Delta CS)^2$	383.8102	696.8113	799.9966	287.6458	346.9196	358.5530
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
$(\Delta GF)^2$	39.1419	101.4364	157.9903	3.6144	4.3235	5.3290
	(< 0.001)	(< 0.001)	(< 0.001)	(0.0573)	(0.5038)	(0.8681)
$(\Delta GS)^2$	317.2978	858.4161	962.2979	311.7861	792.5090	866.7921
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
$(\Delta HF)^2$	45.9233	180.8631	271.3969	1.3445	5.3405	5.7203
	(< 0.001)	(< 0.001)	(< 0.001)	(0.2463)	(0.3758)	(0.8382)
$(\Delta HS)^2$	79.1121	251.7390	379.0275	23.2807	38.8980	46.1772
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)

p-values are indicated in the parenthesis.

implied by a GARCH(1,1) process with normal errors is greater than that of the normal distribution which is 3. For this reason, a GARCH model with normal errors can replicate some of the fat tailed behaviour observed in financial time series, though, most often, a GARCH model with a non-normal error distribution is required in order to fully capture the observed fat tail behaviour in financial time series. Just like exogenous variables can be included in the conditional mean equation as shown in (3.35), exogenous explanatory variables can at times also be incorporated in the conditional variance equation so that (3.22) is modified to take the form

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t \text{ where } \varepsilon_t \sim N(0,1) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + \sum_{k=0}^{K} \delta_k' Z_{t-k}, \end{aligned} \tag{3.39}$$

where δ is an $m \times 1$ vector of positive coefficients and Z_t is an $m \times 1$ vector of variables. Some of these exogenous variables include the trading volume, implied volatility, realized volatility, after hours realized volatility, overnight returns and macroeconomic news announcements.

3.3.3.1 Model Identification and Fitting

In most practical problems, the establishment of the ARCH order p or the GARCH order q can usually pose a significant challenge. Various criteria for model selection such as AIC and BIC can be employed since GARCH models resemble and can be viewed ARMA models for a series of squared residuals. AIC and BIC typically select $ARCH(p)$ models with relatively large values of p when considering high frequency data, whereas for $GARCH(p,q)$ models, the models with $p, q \leq 2$ are typically selected. For most practical applications, the $GARCH(1,1)$ model is usually the most parsimonious.

In this section, we fit several $GARCH(p,q)$ model to the differenced crude futures series for $p = 20, q = 20$. The parameters α_0, α_i and β_j of the $GARCH(p,q)$ model and the in-sample estimates of the volatility are estimated via maximum likelihood. This being high frequency data, the model with the smallest AIC was $GARCH(19,16)$ with an AIC of 8776.4 as shown in Tables E.1, E.2, E.3 and figure 3.26 both showing the AICs of all the models fitted. The best model according to these fits have the parameters summarized in Table 3.8. The Matlab code used to fit these models is in found in Appendix F.

Table 3.8: Results from the $GARCH(19,16)$ Model fitted to the differenced crude futures price series:

Parameter	Value	Standard Error	$t-$ Statistic
Constant	0.1753	0.1030	1.7014
GARCH(16)	0.0898	0.1692	0.5304
GARCH(19)	0.3478	0.1995	1.7432
ARCH(1)	0.1212	0.0221	5.4752
ARCH(2)	0.0710	0.053	1.3404
ARCH(5)	0.0381	0.0523	0.7287
ARCH(6)	0.0374	0.0461	0.8107
ARCH(8)	0.0257	0.033	0.78
ARCH(9)	0.0069	0.0369	0.188
ARCH(10)	0.0498	0.037	1.346
ARCH(13)	0.0217	0.0371	0.5845
ARCH(14)	0.036	0.0376	0.9554
ARCH(15)	0.0125	0.0358	0.3482
ARCH(16)	0.0825	0.0361	2.2876

From these results model it can be seen that most of the in-between lags are not significant, though the important lags at 19 and 16 respectively, are significant. As per the AIC values, most these models do not seem to be significantly different and so the lags can be reduced to gain parsimony. Table 3.9 summarizes the estimated parameters for the GARCH(1,1) model that was fitted to this data.

Table 3.9: Results from the $GARCH(1,1)$ Model fitted to the differenced crude futures price series:

Parameter	Value	Standard Error	$t-$ Statistic
Constant	0.0191	0.0067	2.8406
GARCH(1)	0.9464	0.0074	127.718
ARCH(1)	0.0475	0.0061	7.7615

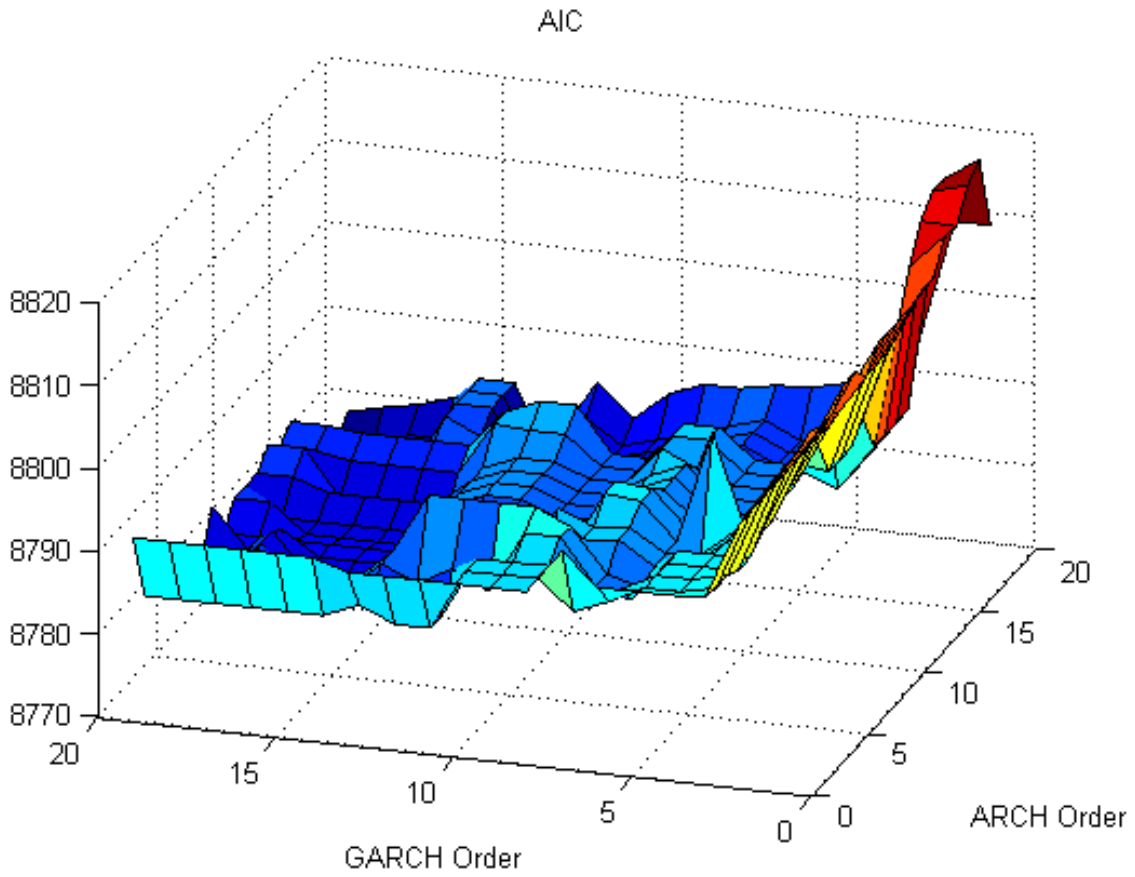

Figure 3.26: A plot for the AICs for various GARCH models fitted

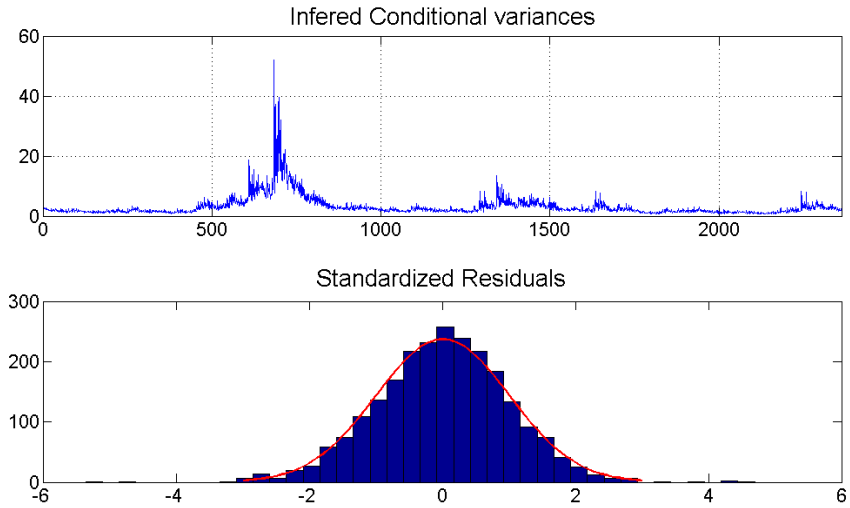

Figure 3.27: GARCH$(19, 16)$ Conditional variances plot and standardized residuals histogram

3.3.3.2 Evaluation of the Estimated GARCH models

After a GARCH model has been fit to the data, it's adequacy is evaluated both graphically and using statistical diagnostic tests. If the model has been correctly specified, then the estimated standardized residuals given by $\hat{\varepsilon}_t/\hat{\sigma}_t$ should typically behave similar to the residuals from classical regression. They must now not display conditional heteroscedasticity, serial correlation, or any other type of non-linear dependence. The distribution of these standardized residuals should also match the specified error distribution used in the estimation. An analysis on the residuals of this model show that this model gives a fairly good fit. Figure 3.27 shows a plot of the inferred conditional variances and a histogram of the standardized residuals. From the histogram, these residuals appear to be normally distributed. Figure 3.28 shows the quantile quantile plot and the box plot. For the quantile quantile plot, most of these values lie along the 45^0 line, depicting a normal distribution. The boxplot also conforms to a normal distribution with a few outliers.

From these two plots we see that although there are a few residuals larger than expected for a Gaussian distribution, the normality assumption is not unreasonable. Generally, An examination of the plots of the standardised residuals after fitting the GARCH(1,1) model for the return series indicates that the residuals are, on the overall, stable with

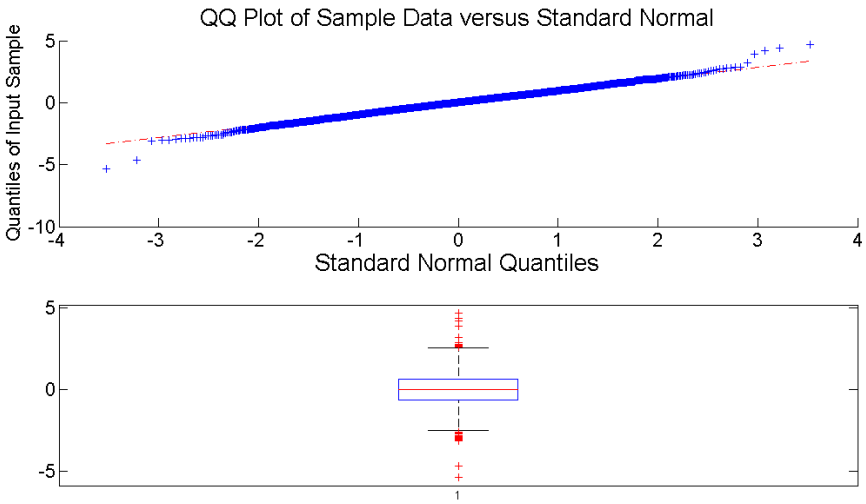

Figure 3.28: $GARCH(19, 16)$ Standardized residuals QQ-plot and boxplot

some clustering as can be seen from the plots in Appendix G. From figure 3.29 of the ACF and the PACF of the standardized residuals, we see that the standardised residuals display no autocorrelation. A residual analysis of the $GARCH(1, 1)$ model gives the exact same result and therefore it makes more sense to just fit a $GARCH(1, 1)$ for this data. The other series also behave in much the same way as the crude futures. Fitting a $GARCH(1, 1)$ model on all the six series gives the results in Table 3.10.

From the measure of persistence, we see the highest persistence of 0.9938 in the ΔCF series, although, all other return series are quite persistent. This indicates that the volatility processes return to their means after some time.

Table 3.10: Results from the $GARCH(1, 1)$ Model fitted to all the differenced price series:

Parameter	ΔCF	ΔCS	ΔGF	ΔGS	ΔHF	ΔHS
α_0	0.0191	0.0265	0.3354	0.5388	0.0743	0.0428
α_1	0.9464	0.9388	0.8236	0.8318	0.915	0.9429
β_1	0.0475	0.058	0.1072	0.1036	0.0642	0.0451
$(\alpha_1 + \beta_1)$	0.9938	0.9916	0.9308	0.9354	0.9792	0.9878

$(\alpha_1 + \beta_1)$ measures the persistence

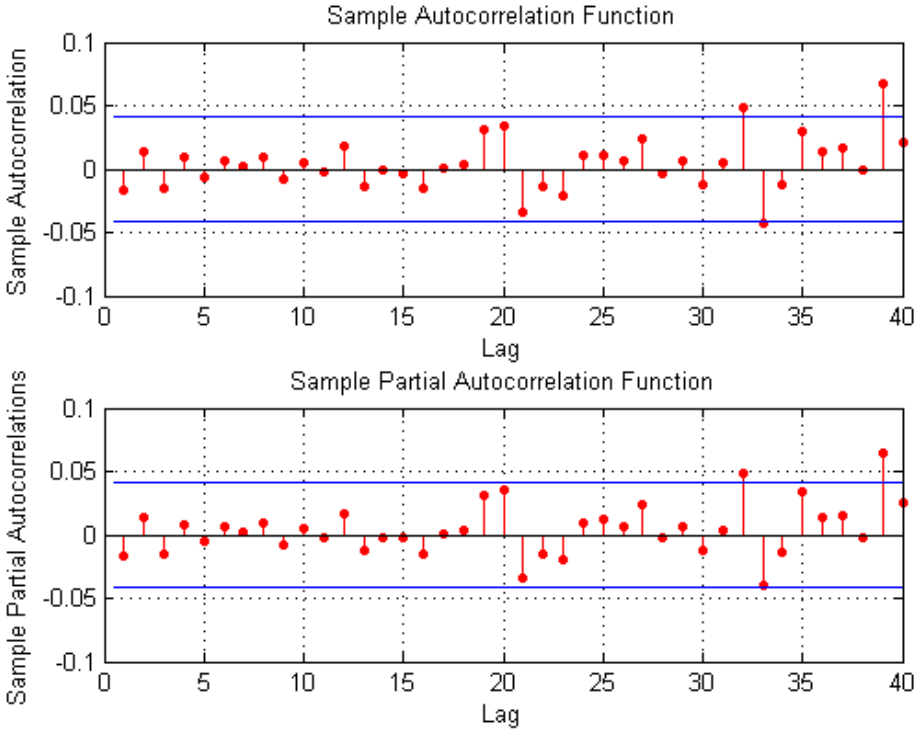

Figure 3.29: $GARCH(19, 16)$ ACF and PACF of standardized residuals

3.3.3.3 *Forecasting volatility with* $GARCH(1, 1)$

Model selection depends on how well the model fits the data, as well as on the objective of the analysis carried out. For purposes of forecasting for example, the best model for in-sample fitting may not necessarily provide better or more accurate predictions (Tsay, 2010). For this reason, the use of the performance of the out-of-sample forecasts to assist in the selection of an adequate statistical model is more common. In the out-of-sample forecasts, the data that is used in model fitting are not the same as those used in forecasting evaluation.

In finance, the sample standard deviation, of returns σ, is used as a simple h time periods ahead forecast of the volatility of returns, over the future period $[t, t + h]$. Suppose this model is estimated using daily stock returns as shown in (3.17), then, the k-periods historical variance will be calculated as

$$\widehat{\sigma}_t^2(k) = \frac{1}{k-1} \sum_{i=0}^{k-1} (r_{t-i} - \bar{r}) \,, \tag{3.40}$$

with the sample mean return $\bar{r} = \frac{1}{k}\sum_{i=0}^{k-1} r_{t-i}t$, representing the estimate of the mean return μ. The average return does not consider the price fluctuations, or the number of prices within the period. Mostly, the mean is usually set to zero in order to achieve a better forecast. Multiplying the variance by the number of trading days in a year, N, and taking the square root results in the annualised volatility

$$\widehat{\sigma} = \sqrt{N.\widehat{\sigma^2}}. \tag{3.41}$$

In most cases, the number of trading days in a year is taken to be $N = 250$. The value $\widehat{\sigma}$ is the best estimator for the volatility from the available price data and the volatility of any period of length, k, can be estimated from it. The k-day volatility forecast can also be found by using the GARCH model. Under the condition that returns are uncorrelated across days, the k-day variance as of $t-1$ is given by

$$E\left[r^2_{t,t+k}|\mathcal{F}_{t-1}\right] = E\left[r^2_t|\mathcal{F}_{t-1}\right] + E\left[r^2_{t+1}|\mathcal{F}_{t-1}\right] + \ldots + E\left[r^2_T|\mathcal{F}_{t-1}\right] \tag{3.42}$$

The predicted volatility over k future periods from $t+1$ to $t+k$, denoted by σ_{Ft}, is the mean of the expected volatility on each and every day from t to $t+k$ i.e.

$$\sigma_{Ft} = \sqrt{\frac{1}{k}\sum_{i=1}^{k} E\left(\sigma^2_{t+k}|\mathcal{F}_t\right)} \tag{3.43}$$

where

$$E\left(\sigma^2_{t+k}|\mathcal{F}_t\right) = \alpha_0 \frac{1-(\alpha_1+\beta_1)^k}{1-(\alpha_1+\beta_1)} + (\alpha_1+\beta_1)^k \sigma^2_t. \tag{3.44}$$

Notice that for $(\alpha_1+\beta_1) < 1$,

$$\lim_{k\to 0} E\left(\sigma^2_{t+k}|\mathcal{F}_t\right) = \frac{\alpha_0}{1-\alpha_1-\beta_1}, \tag{3.45}$$

which is the unconditional variance. The In-Sample and Out-of-Sample forecasting ability of the various volatility models can be measured by the mean absolute error (MAE) and

the root mean squared forecasting errors (RMSFE). These two statistical loss functions are symmetric and have very simple mathematical forms making them more popular for the evaluation the predicting power of a chosen model. The RMSFE has a weakness of assigning more weight to larger forecasting errors. This weakness can be managed using the MAE which contrary to this, assigns equal weights to both scenarios of under and over predictions of volatility.

For comparison of the in-Sample forecasting performance of the various models, the complete data sets for each of the return series will be used. For comparison of out-of-sample forecasting performance of various volatility models, the sample data will be split into two portions. The first part, which is used in training the model and estimating the parameters of the chosen model, which in this case is the $GARCH(1,1)$, contains data from at least two thirds of the entire sample. The remaining one third of the sample is then used to test the forecasting ability of the volatility models. Rolling forecasts are used such that the parameters of the $GARCH(1,1)$ model are estimated each time with a rolling constant sample size. This implies that for a forecasting horizon of k, at each forecast date we add k new data points and subtract k of the old data points. At each forecast date, the realised volatility is then calculated using the expression

$$\sigma_{R,t} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} r_{t+i}^2}, \tag{3.46}$$

so that, for a given model, if the forecast volatility is given by σ_{Ft}, then the RMSFE is given by

$$RMSFE = \sqrt{\frac{1}{n}\sum_{t\in s}\left(\sigma_{Ft} - \sigma_{R,t}\right)^2}, \tag{3.47}$$

and the MAE is given by

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |\sigma_{Ft} - \sigma_{R,t}|\,, \tag{3.48}$$

where n and s denote the number of forecasts and the set of times at which ex ante forecasts are produced respectively in the above expression. Forecasting using ARCH models is obtained recursively (Jondeau et al., 2006). If we left t be the date at which we start the forecasting, then, the $1-$ step ahead forecasts for σ^2_{t+1} is

$$\begin{aligned}\sigma^2_t(1) &= \hat{\alpha_0} + \hat{\alpha_1} a^2_{t+1-1} + \hat{\alpha_2} a^2_{t+1-2} + \ldots + \hat{\alpha_p} a^2_{t+1-p} \\ &= \hat{\alpha_0} + \hat{\alpha_1} a^2_t + \hat{\alpha_2} a^2_{t-1} + \ldots + \hat{\alpha_p} a^2_{t+1-p} \\ &= \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i \hat{a}^2_{t+1-i},\end{aligned} \tag{3.49}$$

where a_t is the estimated residual. For a $2-$ step ahead forecast, we need to forecast a^2_{t+1}, and this is given by $\sigma^2_t(1)$

$$\sigma^2_t(2) = \hat{\alpha}_0 + \hat{\alpha}_1 \sigma^2_t(1) + \hat{\alpha}_2 \hat{a}^2_t + \ldots + \hat{\alpha}_p \hat{a}^2_{t+2-p}. \tag{3.50}$$

The $k-$ step ahead forecast for $\sigma^2_{t+k}$is

$$\begin{aligned}\sigma^2_t(k) &= \hat{\alpha}_0 + \hat{\alpha}_1 \sigma^2_t(k-1) + \ldots + \hat{\alpha}_p \hat{\sigma}^2_t(k-p) \\ &= \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i \sigma^2_t(k-i),\end{aligned} \tag{3.51}$$

where, $\sigma^2_t(k-i) = \hat{a}^2_{t+k-i}$ for $k - i \leq 0$.

Forecasts with GARCH are done in a similar way. The $1-$ step ahead forecast for σ^2_{t+1} is

$$\sigma^2_t(1) = \hat{\alpha}_0 + \sum_{i=1}^{q} \hat{\alpha}_i \hat{a}^2_t + \sum_{i=1}^{p} \hat{\beta}_i \hat{\sigma}^2_t, \tag{3.52}$$

and, since $a_t^2 = \sigma_t^2 \varepsilon_t^2$, the $GARCH(1,1)$ model can be expressed as

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ &= \alpha_0 + (\alpha_1 + \beta_1)\, \sigma_{t-1}^2 + \alpha_1 \sigma_{t-1}^2 \left(a_{t-1}^2 - 1\right),\end{aligned}$$

so that at time $t+2$, we have

$$\sigma_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\, \sigma_{t+1}^2 + \alpha_1 \sigma_{t+1}^2 \left(a_{t+1}^2 - 1\right)$$

with $E\left[\left(a_{t+1}^2 - 1\right)|\mathcal{F}_t\right] = 0$. This means, the $2-$ step ahead forecast for σ_{t+2}^2 is

$$\sigma_t^2(2) = \hat{\alpha}_0 + \left(\hat{\alpha}_1 + \hat{\beta}_1\right) \sigma_t^2(1) \tag{3.53}$$

and the $k-$step ahead forecast for σ_{t+k}^2 is

$$\sigma_t^2(k) = \hat{\alpha}_0 + \left(\hat{\alpha}_1 + \hat{\beta}_1\right) \sigma_t^2(k-1) \text{ for } k > 1. \tag{3.54}$$

An alternative way of writing (3.54) is as shown in (3.55)

$$\sigma_t^2(k) = \hat{\sigma}^2 + \left(\hat{\alpha}_1 + \hat{\beta}_1\right)^{k-1} \left(\sigma_t^2 - \hat{\sigma}^2\right) \text{ for } k > 1. \tag{3.55}$$

where

$$\hat{\sigma}^2 = \frac{\hat{\alpha_0}}{1 - \hat{\alpha_1} - \hat{\beta}_1}.$$

From (3.55), we can see that $\sigma_t^2 \to \hat{\sigma}^2$ as $k \to \infty$

The forecasting ability of the model can only be determined by considering how well a model performs on data not used in estimating the model. It is common practice partition the data into two sets, use the larger portion for estimating the model and the smaller one for testing the model. The test data can be used to measure model accuracy on new data. The size of the test data set should typically be about 20% of the total sample, although this depends on the sample size and the forecast horizon. In an ideal sense, the size of the test sample should be at least as large as the maximum forecast horizon

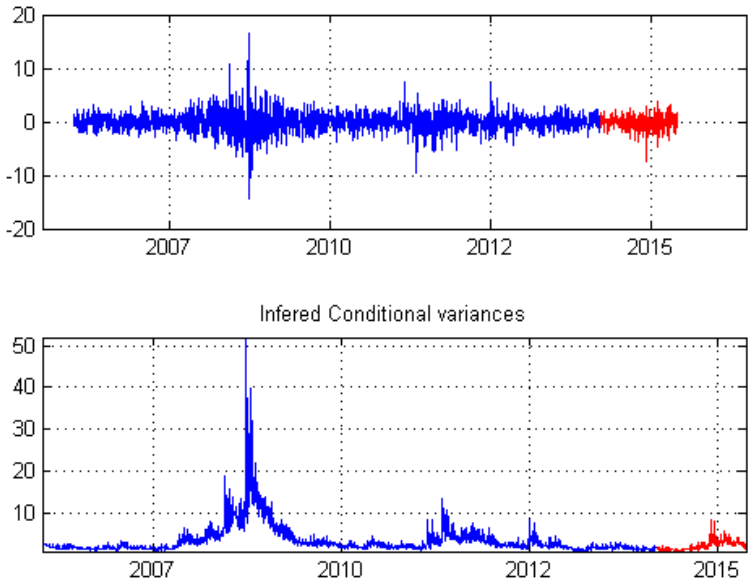

Figure 3.30: Crude futures return series and inferred conditional variances showing the two sub-samples

required (Hyndman and Koehler, 2006).

For this research, the total length of the data under consideration is 2363 data points. this data set was divided into two sets, the first sub-sample, used for model estimation contained 2063 points and the second sub sample, used for forecasting, contained the remaining 300 points. The top graph in figure 3.30 shows how the two sub-samples were produced from the entire data set for the crude oil futures prices. the blue part represents the first sub-sample and the red, the second sub-sample. The second graph shows the inferred conditional variances, inferred from fitting a $GARCH(19, 16)$, divided into the same two sub-samples.

In figure 3.31, we have a graph similar to figure 3.30, except now the red potion on each graph represents the forecast values, both for the returns, and the inferred conditional variances. The red portions of the graph on figure 3.31 are zoomed in on figure 3.32 actually shows our model did some good forecasts.

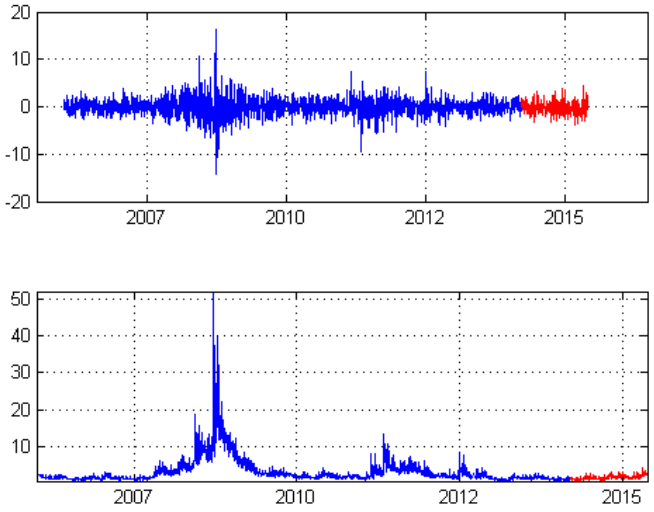

Figure 3.31: Crude futures return series and inferred conditional variances showing the estimation sub-sample and the foretasted values

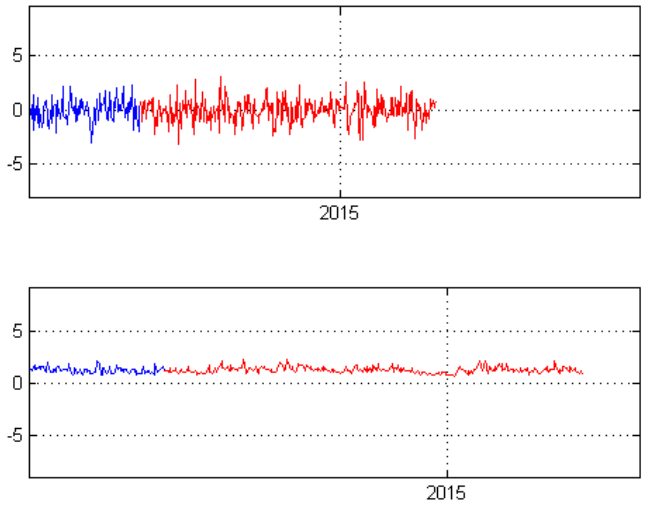

Figure 3.32: Crude futures return series and inferred conditional variances showing the estimation sub-sample and the foretasted values

In order to regain a forecast of the price series, the differencing on the foretasted return series obtained from the foretasted inferred conditional variances is undone. However because the inferred variances are conditioned on a standard normal error, there are several possibilities and, having many different forecasts and averaging would give a stable forecast for the crude futures prices, which can then be used to predict future crude oil futures prices. Figure 3.33 shows on of the possibilities of the foretasted prices, and figure

3.34 just zooms in on figure 3.33 to show the dynamics therein. The RMSFE for this model is 0.2072, an indication that the forecasts are generally good. The MATLAB code used to generate these forecast values is shown in appendix H

Figure 3.33: Crude futures price series showing foretasted prices

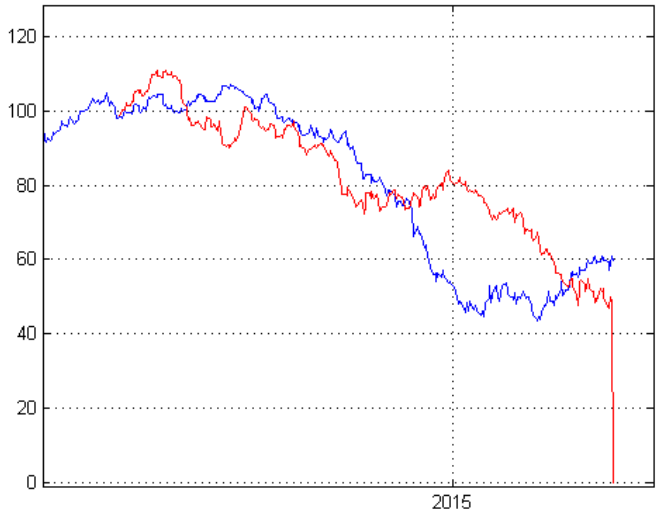

Figure 3.34: Crude futures price series showing foretasted prices

4. MULTIVARIATE ANALYSIS AND CO-INTEGRATION IN ENERGY MARKETS

It is a simple fact is that most realistic applications in empirical finance are actually multivariate, involving more than two assets. Of essential importance in finance is the covariation and co-evolution between the prices of financial assets. The risks faced by a long-run and a short-run investor can usually be quite different since as the investment horizon increases, market fluctuations become the main sources of variability in returns. The ability to understand and model the behaviour of asset or stock returns, especially over long horizons, critically depends on the understanding and modelling of the market dynamics. In order to investigate the common trends seen in multivariate time series, the co-integration approach has emerged as a very powerful technique which provides a sound methodology useful for modelling both short-run and long-run dynamics in a system (Alexander et al., 2002).

Figure 4.1 shows a time series plot of the six series under consideration and figure 4.2 shows a plot of crack spread margins calculated from the ratios discussed and represented in (5.1). Co-evolution of the six series is evident from visual inspection of figure 4.1. This study uses co-integration to analyse the long-run equilibrium relationships between the prices of crude oil and the distillate fuels. If they are co-integrated and have co-integrating residuals which are stationary, then the various spreads are somewhat stationary as can be seen in figure 4.1. If on the other hand they are not co-integrated, then, the spreads can deviate without bounds and using crack spreads for risk management would be questionable (Girma, Paulson, et al., 1999). Co-integration refers to co-evolution in prices, and not the co-evolution in returns.

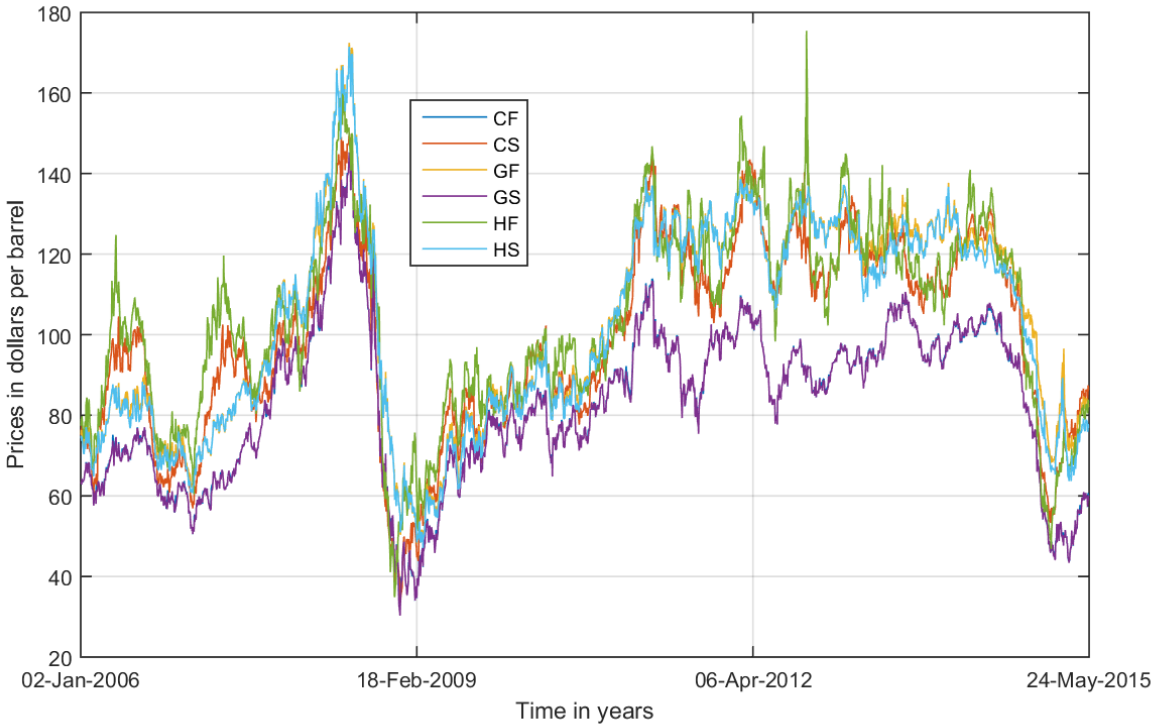

Figure 4.1: Daily futures and spot prices for crude oil and distillate fuels: CF=crude futures, CS=crude spot, GF=gasoline futures, GS=gasoline spot, HF=heating futures, HS=heating spot

4.1 Multivariate Analysis of Energy Markets

The positive correlation of price variations or volatility clustering, as seen on speculative markets is what motivated the introduction by Engle (1982), of the ARCH process, which were later generalized by Bollerslev (1986) to GARCH. The univariate nature of these models implies their capabilities are limited, and as such, they fail to capture any further information embedded in multivariate data that have could have temporal and cross-sectional dependencies in empirical stock price variation and the contemporaneous cross correlation implied by economic theory, such as a set of asset prices, exchange and interest rates, stock market indices among other macroeconomic variables.

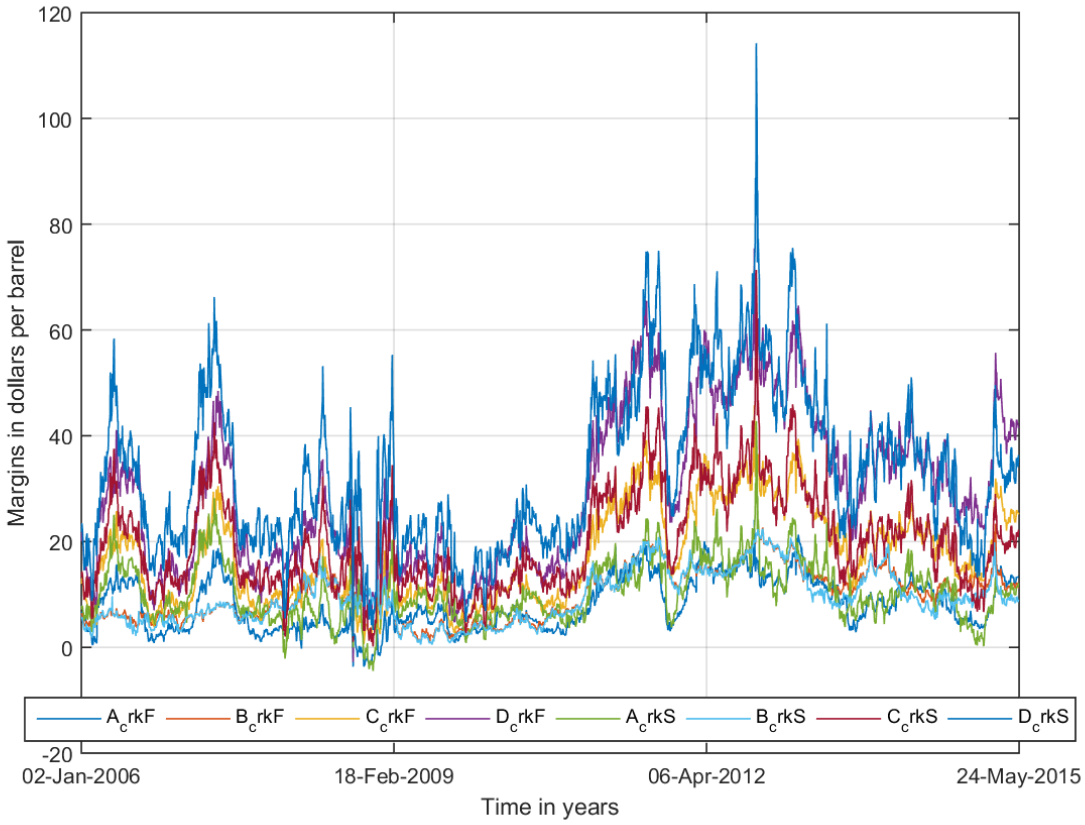

Figure 4.2: Daily crack spread margins from futures and spot prices for crude oil and distillate fuels: AcrkF=1:1:0 futures crack, BcrkF=1:0:1 futures crack, CcrkF=3:2:1 futures crack, DcrkF=5:3:3 futures crack, AcrkS=1:1:0 spot crack, BcrkS=1:0:1 spots crack, CcrkS=3:2:1 spot crack, DcrkS=5:3:3 spot crack

Often we are not interested merely in the behaviour of a single random variable but rather in the joint behaviour of several random variables, for example, several stock returns and a market index. Multivariate distributions are used to describe such joint behaviour since multivariate time series analysis takes into consideration, multiple time series simultaneously. Prices of stocks in the same market usually exhibit similar patterns of behaviour in response to new information that is importance to the whole market. From this perspective, when modelling volatilities that vary with time, a multivariate model seems to provide the natural framework that can be used to take account of cross sectional information. In the evolution of financial markets, asset prices tend to move simultaneously. Over time, asset returns also tend to move simultaneously and so do their respective volatilities.

When studying multivariate processes, a structure is required for depicting the properties

of the individual series as well as the conceivable cross-relationships among these series since they might be interrelated both contemporaneously and across time lags (Reinsel, 1993). Haigh and Holt (2002), employ an MGARCH model that allows for direct incorporation of the time to maturity effect in accounting for the time-varying volatility spill overs between related markets when considering the possibility of simultaneously hedging crude oil, unleaded gasoline and heating oil price risk in a dynamic setting. Their results show that cross market linkages between crude oil, unleaded gasoline and heating oil markets are great importance.

This work uses a multivariate model for time varying volatilities that takes cross sectional information into account. In this section, we discuss some of the basic concepts and models which are deemed crucial in understanding multivariate time series analysis. The main building blocks for the models used in this research are discussed in section 4.2.

4.2 Basic Concepts and Models for Multivariate Time Series

In most practical cases, time series are best considered as components of some vector-valued (multivariate) time series $\{\boldsymbol{p}_t\}$, which has not only serial dependence within each component of the series $\{p_{ti}\}$ but, also has interdependence between different components $\{p_{ti}\}$ and $\{p_{tj}\}$, for $i \neq j$ (Chan, 2011; Tsay, 2013). Most of the theories that apply in univariate time series naturally extend to multivariate time series except for the two major problems of dimensionality and identifiability.

Considering a $k-$dimensional time series $\{\boldsymbol{p}_t = (p_{t1} \ldots, p_{tk})'\}$, the mean vector is defined by

$$\boldsymbol{\mu} = E\left[\boldsymbol{p}_t\right] = \left(E(p_{t1}), \ldots, E(p_{tk})\right)' \tag{4.1}$$

and the covariance matrix

$$\begin{aligned}\boldsymbol{\Gamma}(t+h,t) &= Cov\left(p_{t+h}, p_t\right) \\ &= \begin{pmatrix} \gamma_{11}(t+h,t) & \dots & \gamma_{1k}(t+h,t) \\ \vdots & \ddots & \vdots \\ \gamma_{k1}(t+h,t) & \dots & \gamma_{kk}(t+h,t) \end{pmatrix},\end{aligned} \tag{4.2}$$

where $\gamma_{ij}(t+h,t) = \text{Cov}\left(p_{t+h,i}, p_{t,j}\right)$.

Using matrix notation, we have

$$\boldsymbol{\Gamma}(t+h,t) = E\left(\boldsymbol{p}_{t+h} - \boldsymbol{\mu}_{t+h}\right)\left(\boldsymbol{p}_t - \boldsymbol{\mu}_t\right)'. \tag{4.3}$$

The diagonal elements of the matrix in (4.3) are the auto-covariance functions of the univariate series $\{p_{ti}\}$, while the off-diagonal elements are the covariances between $p_{t+h,i}$ and $p_{t,j}$, $i \neq j$. Note also that $\gamma_{ij}(h) = \gamma_{ji}(-h)$ and the autocorrelation matrix is then defined as

$$\boldsymbol{R}(h) = \begin{pmatrix} \rho_{11}(h) & \dots & \rho_{1k}(h) \\ \vdots & \ddots & \vdots \\ \rho_{k1}(h) & \dots & \rho_{kk}(h) \end{pmatrix}, \tag{4.4}$$

where $\rho_{ij}(h) = \gamma_{ij}(h)\left(\gamma_{ii}(0)\gamma_{jj}(0)\right)^{-1/2}$

4.2.1 Stationarity

A $k-$dimensional time series $\{\boldsymbol{p}_t\}$ is considered to be weekly stationary if the moments $\boldsymbol{\mu}_t$ and $\boldsymbol{\Gamma}(t+h,t)$ are time invariant, meaning, they are both independent of t, in which case

$$E\left[\boldsymbol{p}_t\right] = \boldsymbol{\mu}, \text{ a } k\text{-dimensional constant vector, where } \boldsymbol{\mu} = \left(\mu_1, \mu_2, \dots, \mu_k\right)'$$

and

$$Cov\left(\boldsymbol{p}_t\right) \equiv \Gamma(0) = E\left[\left(\boldsymbol{p}_t - \boldsymbol{\mu}\right)\left(\boldsymbol{p}_t - \boldsymbol{\mu}\right)'\right] = \boldsymbol{\Sigma}_p, \text{ a } k \times k \text{ positive definite matrix.}$$

The probability distribution of observations from a stationary vector process is unchanging with respect to the changes in time.

A $k-$dimensional time series $\{\boldsymbol{p}_t\}$ is said to be strictly stationary if the joint distribution associated with the m collection $(p_{t_1}, \ldots, p_{t_m})$ at times $1, 2, \ldots, t$ is the same as that associated with m observations $(p_{t_{1+j}}, \ldots, p_{t_{m+j}})'$ at times $j+1, j+2, \ldots, j+m$ where m, j and $(t1, \ldots, tm)$ are arbitrary positive integers.

4.2.2 Linearity

In a strict sense, true multivariate time series are non-linear, yet linear models can regularly give precise approximations for making inference.

A $k-$dimensional time series $\{\boldsymbol{p}_t\}$ is linear if

$$\boldsymbol{p_t} = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\psi_i a_{t-i}} \tag{4.5}$$

where $\boldsymbol{\mu}$ is a $k-$dimensional constant vector, $\boldsymbol{\psi_0 - I_k}$ is the $k \times k$ identity matrix, $\boldsymbol{\psi_i}, i > 0$ are $k \times k$ constant matrices and $\boldsymbol{a}_t$ is a sequence of i.i.d random vectors with mean $=0$ and a positive definite covariance matrix $\boldsymbol{\Sigma}_a$.

Definition 4.2.1. $\{\boldsymbol{a_t}\} \sim i.i.d(0, \boldsymbol{\Sigma})$ *if* $\{\boldsymbol{a_t}\}$ *are i.i.d with mean* $\boldsymbol{\mu} = 0$ *and covariance matrix* $\boldsymbol{\Sigma_a}$.

Definition 4.2.2. $\{\boldsymbol{a_t}\} \sim \boldsymbol{WN}(0, \boldsymbol{\Sigma})$ *if and only if* $\{\boldsymbol{a_t}\}$ *is stationary with mean* $\boldsymbol{\mu} = 0$

and

$$\Gamma(h) = \begin{cases} \Sigma & \textit{for } h = 0 \\ 0 & \textit{elsewhere} \end{cases}$$

Alternatively, linearity could be expressed by requiring $\boldsymbol{\psi}_0$ to be a lower triangular matrix with it's diagonal elements being 1, and $\boldsymbol{\Sigma}_a$, being a diagonal matrix. This can be achieved by utilizing the Cholesky decomposition of $\boldsymbol{\Sigma}_a$.

Particularly, the covariance matrix $\boldsymbol{\Sigma}_a$ can be decomposed as

$$\boldsymbol{\Sigma_a = LGL'} \tag{4.6}$$

where L is a $k \times k$ lower triangular matrix with it's diagonal elements being 1 and G is a diagonal matrix.

If we let $\boldsymbol{b_t = L^{-1}a_t}$ so that $\boldsymbol{a_t = Lb_t}$, then

$$\begin{aligned} \text{Cov}(\boldsymbol{b_t}) = \text{Cov}\left(\boldsymbol{L^{-1}a_t}\right) &= \boldsymbol{L^{-1}\Sigma_a}\left(\boldsymbol{L^{-1}}\right)' \\ &= \boldsymbol{L^{-1}}\left(\boldsymbol{LGL'}\right)\left(\boldsymbol{L'}\right)^{\boldsymbol{-1}} \\ &= \boldsymbol{G} \end{aligned}$$

Using the sequence $\{\boldsymbol{b_t}\}$ (4.5) can then be expressed as

$$\begin{aligned} \boldsymbol{p}_t &= \boldsymbol{\mu} + \sum_{i=0}^{\infty} \left(\boldsymbol{\psi_i L}\right) \boldsymbol{b}_{t-i} \\ &= \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\psi_i^* b_{t-i}} \end{aligned} \tag{4.7}$$

where $\boldsymbol{\psi_0^* = L}$, a lower triangular matrix, $\boldsymbol{\psi_i^* = \psi_i L}$ for $i > 0$ and the covariance matrix for $\boldsymbol{b}_t$ ia a diagonal matrix.

For the linear series represented in (4.5) to be stationary, the coefficient matrices must

fulfil the condition

$$\sum_{i=1}^{\infty} \|\boldsymbol{\psi_i}\| < \infty \tag{4.8}$$

where $\|\boldsymbol{\psi}\|$ denotes the norm of a matrix $\boldsymbol{\psi}$. This implies that $\|\boldsymbol{\psi_i}\| \to 0$ as $i \to \infty$. This means that $\boldsymbol{\psi_i} \to 0$ as $i \to \infty$. Further more, we must have

$$E\left(\boldsymbol{p_t}\right) = \boldsymbol{\mu} \text{ and } \operatorname{Cov}\left(\boldsymbol{p_t}\right) = \sum_{i=0}^{\infty} \boldsymbol{\psi_i \Sigma_a \psi_i'}$$

For a stationary multivariate linear time series $\{\boldsymbol{p_t}\}$ represented as in (4.5), for the case of lag $h = 0$, we have

$$\begin{aligned}
\boldsymbol{\Gamma(h)} &= E\left[\left(\boldsymbol{p_t} - \boldsymbol{\mu}\right)\left(\boldsymbol{p_{t-h}} - \boldsymbol{\mu}\right)'\right] \\
&= E\left[\left(\boldsymbol{a_t} + \boldsymbol{\psi_1 a_{t-1}} + \ldots\right)\left(\boldsymbol{a_{t-h}} + \boldsymbol{\psi_1 a_{t-h-1}} + \ldots\right)'\right] \\
&= E\left[\left(\boldsymbol{a_t} + \boldsymbol{\psi_1 a_{t-1}} + \ldots\right)\left(\boldsymbol{a'_{t-h}} + \boldsymbol{a'_{t-h-1}\psi'_1} + \ldots\right)\right] \\
&= \sum_{i=h}^{\infty} \boldsymbol{\psi_i \Sigma_a \psi'_{i-h}}
\end{aligned} \tag{4.9}$$

4.2.3 Covariance and Correlation

For a stationary process $\{\boldsymbol{p_t}\}$, the covariance between $p_{t+h,i}$ and $p_{t,j}$ must depend only on the lag h, not on time, for $i, j = 1, \ldots, t$ and $h = 0, \pm 1, \pm 2, \ldots$. If we let

$$\gamma_{ij}(h) = \operatorname{Cov}\left(p_{t+h,i}, p_{t,j}\right) = E\left[\left(p_{t+h,i} - \mu_i\right)\left(p_{t,j} - \mu_j\right)\right]$$

and denote a $t \times t$ matrix of cross-covariances at lag h by

$$\Gamma(h) = E\left[\left(p_t - \mu\right)\left(p_{t+h} - \mu\right)'\right] = \begin{pmatrix} \gamma_{11}(h) & \gamma_{12}(h) & \ldots & \gamma_{1k}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) & \ldots & \gamma_{2k}(h) \\ \vdots & \ldots & \ddots & \vdots \\ \gamma_{k1}(h) & \gamma_{k2}(h) & \ldots & \gamma_{kk}(h) \end{pmatrix} \tag{4.10}$$

for $h = 0, \pm 1, \pm 2, \ldots$. similarly, the cross-correlation matrix at lag h is denoted by

$$\rho(h) = V^{-1/2}\Gamma(h)V^{-1/2} = \begin{pmatrix} \rho_{11}(h) & \rho_{12}(h) & \ldots & \rho_{1k}(h) \\ \rho_{21}(h) & \rho_{22}(h) & \ldots & \rho_{2k}(h) \\ \vdots & \ldots & \ddots & \vdots \\ \rho_{k1}(h) & \rho_{k2}(h) & \ldots & \rho_{kk}(h) \end{pmatrix} \tag{4.11}$$

for $h = 0, \pm 1, \pm 2, \ldots$, where $V^{-1/2} = Diag\left\{\gamma_{11}(0)^{-1/2}, \ldots, \gamma_{tt}(0)^{-1/2}\right\}$, since $\rho_{ij}(h) = \text{corr}\,(p_{ti}, p_{t+h,j}) = \dfrac{\gamma_{ij}(h)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{-1/2}}$ with $\gamma_{ii} = \text{Var}(p_{ti})$.

For $i = j$, $\rho_{ii}(h) = \rho_{ii}(-h)$ represents the autocorrelation function of the ith series $p_{t,i}$, and, for $i \neq j$, $\rho_{ij}(h) = \rho_{ji}(-h)$ represents the cross-correlation function between the series $p_{t,i}$ and $p_{t,j}$. Note that for $h = 0$, we have the covariance matrix $\Gamma(0)$ of $\boldsymbol{p_t}$ since $\boldsymbol{\Sigma_p = \Gamma(0)}$.

For a positive lag h, $\gamma_{ij}(h)$ can be viewed as a measure of the linear dependence of the ith component $p_{t,i}$ on the hth lagged value of the jth component $p_{t,j}$.

Also note that $\Gamma(h)' = \Gamma(-h)$ and $\rho(h)' = \rho(-h)$, since $\gamma_{ij}(h) = \gamma_{ij}(-h)$ as shown below.

$$\begin{aligned} \boldsymbol{\Gamma}(h) &= E\left[(\boldsymbol{p_t} - \boldsymbol{\mu})\,(\boldsymbol{p_{t-h}} - \boldsymbol{\mu})'\right] \\ &= E\left[(\boldsymbol{p_{t+h}} - \boldsymbol{\mu})\,(\boldsymbol{p_t} - \boldsymbol{\mu})'\right] \text{ by stationarity} \\ &= \left\{E\left[(\boldsymbol{p_t} - \boldsymbol{\mu})\,(\boldsymbol{p_{t+h}} - \boldsymbol{\mu})'\right]\right\}' \text{ since } \boldsymbol{A = (A')'} \\ &= \left\{E\left[(\boldsymbol{p_t} - \boldsymbol{\mu})\left(\boldsymbol{p_{t-(-h)}} - \boldsymbol{\mu}\right)'\right]\right\}' \\ &= \{\boldsymbol{\Gamma}(-h)\}' \text{ by definition} \\ &= \boldsymbol{\Gamma}(-h)' \end{aligned}$$

Therefore, unlike the instance of univariate stationary time series for which the auto-covariances of lag h and lag $-h$ are similar, one must transpose a positive-lag cross-

covariance matrix in order to obtain the negative-lag cross-covariance matrix. The cross-correlation matrices $\rho(h)$ and the cross-covariance matrices $\Gamma(h)$ are positive definite since $\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{b}_i'\Gamma(i-j)\boldsymbol{b}_j \geq 0$ for all positive integers n and all $k-$dimensional vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$, which follows since $\text{Var}\left(\sum_{i=1}^{n} \boldsymbol{b}_i'\boldsymbol{p}_{t-i}\right) \geq 0$

Note that the last equality holds because $\boldsymbol{a_t}$ has no serial covariances and $\boldsymbol{\psi_0 = I_k}$. So for a stationary series $\{\boldsymbol{p_t}\}$, the lag h cross-correlation matrix $\boldsymbol{\rho(h)}$ is defined as

$$\begin{aligned} \boldsymbol{\rho(h)} &= \boldsymbol{D^{-1}\Gamma(h)D^{-1}} \\ &= \boldsymbol{\rho_{ij}(h)} \end{aligned} \tag{4.12}$$

where $\boldsymbol{D} = \text{diag}\{\sigma_1, \ldots, \sigma_k\}$ is a diagonal matrix of the standard deviation of the components of $\{\boldsymbol{p_t}\}$. Specifically, $\sigma_i^2 = \text{Var}(p_{it}) = \gamma_{ii}(0)$, that is the i,ith element of $\boldsymbol{\Gamma(0)}$.

Note that $\boldsymbol{\rho(0)}$ is symmetric, it's diagonal elements are 1 and the off-diagonal elements are the instantaneous cross-variations between the components of $\{\boldsymbol{p_t}\}$.

Not however that for $h > 0$, $\boldsymbol{\rho(h)}$ is not symmetric generally because $\boldsymbol{\rho_{ij}(h)}$ is the correlation coefficient between p_{ti} and $p_{t-h,j}$ whereas $\boldsymbol{\rho_{ji}(h)}$ is the correlation coefficient between p_{tj} and $p_{t-h,i}$, but using the properties of $\Gamma(h)$, we have $\rho(h) = \rho'(-h)$.

Note that for a $k-$dimensional time series $\{\boldsymbol{p_t}\}$ each matrix $\boldsymbol{\rho(h)}$ is a $k \times k$ matrix. Where k is large, it is difficult to simultaneously decipher $\rho(h)$ for several values of h. In order to summarize this information, one can consider k^2 plots of the elements of $\rho(h)$, for $h = 0, \ldots, m$, where m is a specified positive integer. For example for each $\{i,j\}$th position, we plot $\rho_{ij}(h)$ versus h. these plots show the linear dynamic dependence of p_{ti} on $p_{t-h,j}$ for $h = 0, \ldots, m$. The k^2 plots are referred to as cross-correlation plots of $\boldsymbol{p_t}$.

4.2.4 Invertibility

A process is viewed to be invertible if it can be represented as

$$p_t = \sum_{i=1}^{\infty} \psi_i p_{t-i} + a_t. \tag{4.13}$$

In many time series applications such as forecasting, it is prudent to express the time series $\{\boldsymbol{p_t}\}$ as a function of it's lagged values $\{\boldsymbol{b_{t-i}}\}$ for $i > 0$ and some noise $\{\boldsymbol{a_t}\}$. For an invertible series, $\boldsymbol{\psi_i} \to 0$ as $i \to \infty$. A vector autoregressive (VAR) series of order 1 is an example of an invertible time series.

In data analysis, we study the sample cross-correlation matrix $\widehat{\boldsymbol{\rho}}(h)$ in order to observe the linear dynamic dependence in the data.

Given a sample series $\{\boldsymbol{p_t}\}$ for $t = 1, \ldots, T$, the sample mean vector is given by

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{p}} = \frac{1}{T} \sum_{t=1}^{T} p_t \tag{4.14}$$

and the sample covariance matrix is

$$\widehat{\boldsymbol{\Gamma}}(\mathbf{0}) = \frac{1}{T-1} \sum_{t=1}^{T} (\boldsymbol{p_t} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{p}}) (\boldsymbol{p_t} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{p}})' \tag{4.15}$$

The lag h sample cross-covariance matrix is defined as

$$\widehat{\boldsymbol{\Gamma}}(\boldsymbol{h}) = \frac{1}{T-1} \sum_{t=h+1}^{T} (\boldsymbol{p_t} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{p}}) (\boldsymbol{p_t} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{p}})' \tag{4.16}$$

and the lag h sample cross-correlation matrix is given by

$$\widehat{\boldsymbol{\rho}}(\boldsymbol{h}) = \hat{\boldsymbol{D}}^{-\mathbf{1}} \widehat{\boldsymbol{\Gamma}}(\boldsymbol{h}) \hat{\boldsymbol{D}}^{-\mathbf{1}} \tag{4.17}$$

where $\hat{\boldsymbol{D}} = \text{diag}\left\{\hat{\gamma}_{11}^{1/2}(0), \ldots, \hat{\gamma}_{kk}^{1/2}(0)\right\}$ in which $\hat{\gamma}_{ii}(0)$ is the i, ith element of $\widehat{\boldsymbol{\Gamma}}(\mathbf{0})$.

4.3 Co-integration in Energy Markets

Time series that contain some trend have a great potential of creating major problems in empirical econometrics as a result of spurious regressions. A solution to this can be by successively differencing the series until stationarity is achieved and then applying regression analysis to the stationary series. This solution may however not be an ideal one since in addition to differencing the error processes in the regression, it no longer provides a unique long-run solution. For two non-stationary variables, the error can be represented by a combination of two cumulated error processes. These cumulated error processes form a stochastic trend which normally can be expected to combine to produce a new non-stationary process. In the special circumstance that the two variables, say x_t and y_t are truly related, it would be expected that they move together so that their two stochastic trends would be smiler to each other, and in the event that they are combined, there should be a possibility of finding a combination of them which would eliminates non-stationarity. In such a case, we say the two variables are co-integrated.

Co-integration was developed as a means of modelling dynamic co-dependencies in multivariate time series (Chan, 2011). Time series are referred to as co-integrated if they evolve together, while at the same time staying close together, even if as individual systems, they can drift about. If spreads exhibit mean-reversion, and asset prices are entwined by a common stochastic trend in the long-run, then, the price series are said to be co-integrated. This can be interpreted to mean that there exists feedbacks that mutually align the variables (Rachev et al., 2007). Co-integrated processes are characterized by short-run dynamics and long-run equilibria.

Spot and futures prices are entwined since at some maturity date in the future, they converge,and because of this, their tracking error or *basis* must exhibit mean-reversion. For

tow related assets, if the prices are co-integrated then their spreads will usually be stationary. Co-integration simply extends correlation analysis so as to incorporate an initial stage during which the price data are analysed and a second stage which entails dynamic analysis of correlations so that information concerning any lead-lag behaviour between returns is provided (Alexander, 2008). Co-integration theory is useful in estimating and testing long-term equilibrium relationships among non-stationary asset prices and allows for meaningful statistical inference (Girma, Paulson, et al., 1999). Using co-integration, macroeconomic models containing non-stationary stochastic variables can be constructed so that the results are both statistically sound and economically meaningful.

4.4 Co-integration and common trends

Normally, a linear combination of integrated variables will also be integrated of the same order as the individual variables. If two time series x_t and y_t are both integrated of order d ($I(d)$), then generally, any linear combinations of the two series will also be $I(d)$; that is, the residuals obtained from regressing y_t on x_t will be $I(d)$ (Harris and Sollis, 2003). The exception from this rule is called co-integration, where a linear combination of the integrated variables results in a lower order of integration. Co-integration is concerned with the common behaviour of the multivariate time series where each component of a multivariate time series may be non-stationary, but certain linear combinations of these component series are indeed stationary (Chan, 2011).

The main objective of co-integration analysis is to identify any common stochastic (non-stationary) trends and patterns present in price data and use these in the dynamic analysis of correlations in returns. Co-integration involves a two step process. It starts by establishing any long-run equilibrium relationships between prices and then estimates the dynamic correlation models of the returns, through, the error correction model (ECM) since if any two variables are $I(1)$, and are co-integrated, they can be modelled as having been generated by an ECM. An ECM corrects the deviations from the long-run equili-

brium (Alexander et al., 2002).

In standard time series analysis, price data is differenced prior to any analysis, and this eliminates any long-term trends present in the data *a-priori*, and hence, any information that can hitherto be conveyed by there trends. Co-integration analysis on the other hand is applied on raw price data and returns as well. two or more series are co-integrated if each component series is integrated and some linear combination of these series is stationary (Engle and Granger, 1987). For price series to be co-integrated, each must

(i) exhibit random walk behaviour at the price level,

(ii) be integrated of the same order

(iii) have co-integrating residuals which are stationary.

A series that is integrated is one that is stationary in the mean at order d, i.e. $I(d)$. This series contains d unit roots. In the multivariate case, these series may each be integrated processes of some order. A given set of $I(1)$ series are co-integrated if there exists a linear combination of these series that is stationary. Examples of stationary spreads include crack spreads and calendar spreads, which arise from the difference between two futures prices on the same underlying but with different maturities, and are examples. If two series $x_t, y_t \sim I(d)$ then, generally, any linear combination of these two series will also be $I(d)$. These two series $\{x_t\}$ and $\{y_t\}$ will be co-integrated if $x_t, y_t \sim I(1)$ but there exists an α such that $u_t = x_t - \alpha y_t \sim I(0)$. In a more general sense, $\{x_t\}$ and $\{y_t\}$ are considered to be co-integrated of order $CI(d, p)$ if $x_t, y_t \sim I(d)$; and there exists an α such that $u_t = x_t - \alpha y_t \sim I(d - p)$.

Two sequences of random variables $\{x_t\}$ and $\{y_t\}$ are considered to be co-integrated if

(i) they are non-stationary in levels, $\{x_t\}, \{y_t\} \sim I(1)$;

(ii) they are stationary in first differences; and

(iii) there exists a linear combination of the levels, so that there is an α so that the error term $u_t = x_t - \alpha y_t \sim I(0)$, where α is referred to as the co-integrating constant. u_t is also known as the disequilibrium error.

If there are $n > 2$ variables in the model, then there can be more than one co-integration vector. It is possible to have up to $n-1$ linearly independent co-integration vectors. Co-integration is a two step process which begins with the establishment of any long-term relationships between the series involved, and then estimating the time varying correlation model of returns. Note however that, the co-integrating vector is not identified unless we impose some arbitrary normalization.

A $k-$dimensional vector $\{\mathbf{X}_t\}$ is said to be $I(1)$ if $\Delta\mathbf{X}_t - (1-B)\mathbf{X}_t = \mathbf{X}_t - \mathbf{X}_{t-1}$ is $I(0)$ where $B\mathbf{X}_t = \mathbf{X}_{t-1}$.

Let

$$\begin{aligned} \mathbf{X}_{t1} &= \sum_{i=1}^{t} Z_{i1} + Z_{t2} \\ \mathbf{X}_{t2} &= \frac{1}{2}\sum_{i=1}^{t} Z_{i1} + Z_{t3} \\ \mathbf{X}_{t3} &= Z_{t3} \end{aligned} \tag{4.18}$$

where $\boldsymbol{(Z_{t1}, Z_{t1}, Z_{t3}) \sim N(0, I_3)}$, and $\mathbf{I}_3$ is an identity matrix of order 3. $\mathbf{X}_t = (X_{t1}, X_{t2}, X_{t3})'$ is non-stationary since the first two components $\mathbf{X}_{t1}$ and $\mathbf{X}_{t2}$ contain the common trend term $\sum_{i=1}^{t} Z_{i1}$. It is this accumulated sum of Z_{i1} that gives rise to the non-stationarity. On the other hand

$$\begin{aligned} \Delta\mathbf{X}_t &= (1-B)\mathbf{X}_t = \mathbf{X}_t - \mathbf{X}_{t-1} \\ &= \left(Z_{t1} + Z_{t2} - Z_{t-1}, 2, \frac{1}{2}Z_{t1} + Z_{t3} - Z_{t-1}, 3, Z_{t3} - Z_{t-1}, 3\right)' \end{aligned}$$

is stationary since it is easy to verify that $\Delta \mathbf{X}_t \sim I(0), \mathbf{X}_t \sim I(1)$. In addition,

$$\begin{aligned} X_{t1} - 2X_{t2} &= \sum_{i=1}^{t} Z_{i1} + Z_{t2} - \sum_{i=1}^{t} Z_{i1} - 2Z_{t3} \\ &= Z_{t2} - 2Z_{t3} \end{aligned}$$

which is stationary since it is expressed as a finite sum of two uncorrelated white noise sequences. In this case, $\mathbf{X}_t$ is non-stationary, but $\alpha' \mathbf{X}_t$ where $\alpha = (1, -2, 0)'$ is stationary, implying that $\mathbf{X}_t$ is co-integrated with co-integrating vector $\alpha = (1, -2, 0)'$. In this case, stationarity is attained either by differencing or by taking linear combinations.

Definition 4.4.1 (Co-integration)**.** *If $\{\boldsymbol{X}_t\}$ is non-stationary and there exists $\alpha \neq 0$ such that $\alpha' \boldsymbol{X}_t$ is stationary by a suitable choice of $\alpha' \boldsymbol{X}_0$, then $\boldsymbol{X}_t$ is **co-integrated** and α is the **co-integrating vector**. The number of linearly independent co-integrating vectors is the **co-integrating rank** and the space spanned by the co-integrating vectors is the **co-integrating space** (Chan, 2011).*

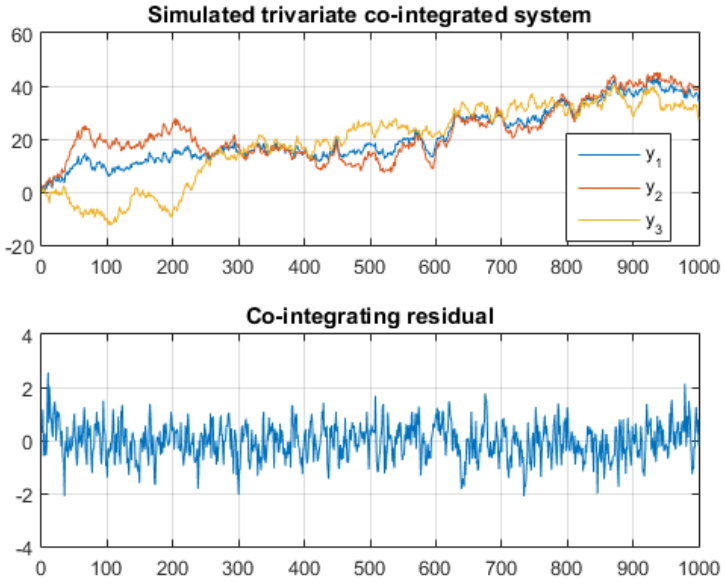

Figure 4.3: Simulated trivariate co-integrated system with one co-integrating vector $\beta = (1, -0.65, -0.35)'$ and two stochastic trends

Figure 4.3 shows simulated data for a trivariate co-integrated system $\mathbf{Y}_t = (y_{1t}, y_{2t}, y_{3t})'$ with one co-integrating vector $\beta = (1, -0.65, -0.35)'$ and two stochastic trends. (4.19)

shows the long-run equilibrium while the next two equations specify the common stochastic trends.

$$y_{1t} = \beta_2 y_{2t} + \beta_3 y_{3t} + z_t \text{ where } z_t \sim I(0) \tag{4.19}$$

$$y_{2t} = y_{2t-1} + u_t \text{ where } u_t \sim I(0) \tag{4.20}$$

$$y_{3t} = y_{3t-1} + e_t \text{ where } e_t \sim I(0) \tag{4.21}$$

Co-integrated prices of energy products have a common stochastic trend since they co-evolve in the long run even though they may drift apart in the short run, the spread or any other linear combination of the two series exhibits mean reversion. Each specific stationary linear combination of the price series acts as a bonding agent for the system so that the more co-integrating vectors in the system, the greater the long rum relationship and association between the price series.

4.5 Error correction model (ECM)

The time paths followed by co-integrating variables are influenced by the extent of any deviations from the long-run equilibria. In any case, for a system to revert back to it's long-run equilibrium, it's movements for at least some of the variables involved must be responsive to the sheer size of the disequilibrium. These movements are captured using an ECM. When two random walk (non-stationary and unit root) $I(1)$ variables are co-integrated, an ECM can be formulated to study their short-run dynamics. In an ECM, the short-run variation of the variables in the system are influenced by the drifts from equilibrium (Enders, 2008).

An ECM is a dynamic model in which the movements of one variable during any period has some relationships with the departures from the long-term equilibrium manifested by the previous period. In such a system, the changes in one variable depend on the deviations from some equilibrium relation (Lütkepohl, 2007).

Consider a simple long-term equilibrium model

$$y_t = \beta_0 + \beta_1 x_t. \tag{4.22}$$

A simple non-constant model of short-run adjustment (with $p = q = 1$) is given by

$$y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \alpha_1 y_{t-1} + z_t \tag{4.23}$$

where $z_t \sim N(0, \sigma^2)$, γ_0 represents the short-term reaction of y_t to changes in x_t and not the long-term effect that would occur if the model had been in equilibrium. This implies that in the long run, the elasticity between $\mathbf{Y}$ and $\mathbf{X}$ is $\beta_1 = \gamma_0 + \frac{\gamma_1}{1-\alpha}$, assuming $\alpha < 1$. However, with this form of the dynamic model, there are several potential problems including the likely high level of correlation between current and lagged values of a variable, which will result in problems such as multicollinearity (Harris and Sollis, 2003), non-standard distributed parameter estimates and spurious correlation. The solution to these problems is in the estimation of first differences of (4.23) obtaining

$$\Delta y_t = \alpha_0 + \gamma \Delta x_{t-1} + \gamma_1 \Delta x_{t-1} + \alpha_1 \Delta y_{t-1} + z_t \tag{4.24}$$

This, however, has the disadvantage of introducing problems of loss of information concerning the long-run equilibrium. A more suitable approach is to adopt the ECM, which is set up as follows:

Subtracting y_t from both sides of the short-run model in (4.23) yields

$$\Delta y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} - (1 - \alpha_1) y_{t-1} + z_t. \tag{4.25}$$

Further subtracting $\gamma_0 x_{t-1}$ from both sides of the resulting equation gives

$$\Delta y_t - \gamma_0 x_{t-1} = \alpha_0 + \gamma_0 x_t - \gamma_0 x_{t-1} + \gamma_1 x_{t-1} - (1 - \alpha_1) y_{t-1} + z_t. \tag{4.26}$$

Rearranging this equation gives

$$\Delta y_t = \alpha_0 + \gamma_0 \Delta x_t + (\gamma_0 + \gamma_1) x_{t-1} - (1 - \alpha_1) y_{t-1} + z_t, \tag{4.27}$$

and re-parametrization then gives

$$\begin{aligned} \Delta y_t &= \gamma_0 x_t - (1 - \alpha_1) \left[y_{t-1} - \frac{\alpha_0}{1 - \alpha_1} - \frac{\gamma_0 + \gamma_1}{1 - \alpha_1} x_{t-1} \right] + z_t \\ &= \gamma_0 x_t - (1 - \alpha_1) \left[y_{t-1} - \beta_0 - \beta_1 x_{t-1} \right] + z_t \end{aligned} \tag{4.28}$$

if we take $\beta_0 = \dfrac{\alpha_0}{1 - \alpha_1}$ and $\beta_1 = \dfrac{\gamma_0 + \gamma_1}{1 - \alpha_1}$. This is an ECM. This model incorporates both short-run and long-run effects so that if the equilibrium holds, then the quantity $[y_{t-1} - \beta_0 - \beta_1 x_{t-1}] = 0$. If this quantity is non-zero, it indicates there is a period of disequilibrium. This term $z_{t-1} = [y_{t-1} - \beta_0 - \beta_1 x_{t-1}] = 0$ is the error correction term which measures the distance that the system has drifted away from equilibrium. This introduction of the equilibrium error from the previous period into this representation as an explanatory variable allows movement to a new equilibrium. The term $(1 - \alpha_1)$ measures the speed of adjustment back to the long-term equilibrium depicted by (4.22). The coefficient of z_{t-1} is required to be negative if the system is to converge to equilibrium. If the values of $(1 - \alpha_1) \to 1$, it indicates that economic agents eliminate large percentages of disequilibrium in each period, if $(1 - \alpha_1) \to 0$, it indicates that adjustment is slow and values close to 2, give an indication that the economic equilibrium has overshot. Positive values would indicate that the system has diverged from the expected long-run path of equilibrium. z_t here is the disequilibrium error or the co-integrating residual. The expected value of z_t defines the long term equilibrium relationship between the variables x_t and y_t and as the observed value of z_t fluctuates around its expected value, periods of disequilibrium occur.

All terms contained in the ECM are stationary, and so standard regression techniques are appropriate, assuming co-integration and that the estimates of β_0 and β_1 have been found (Harris and Sollis, 2003). According to Engle and Granger (1987), if $y_t, x_t \sim CI(1, 1)$,

then an ECM must exist, and conversely, the ECM generates a co-integrated series. In a general sense, if we increase the number of lags p and/or q, the ECM can be reformulated as

$$A(L)\Delta y_t = B(L)\Delta x_t - (1-\pi)\left[y_{t-p} - \beta_0 - \beta_1 x_{t-p}\right] + z_t \tag{4.29}$$

where $A(L)$ is the polynomial lag operator $1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p$, $B(L)$ is the polynomial lag operator $\gamma_0 + \gamma_1 L + \gamma_2 L^2 + \ldots + \gamma_p L^p$ and $\pi = (\alpha_1 + \alpha_2 + \ldots + \alpha_p)$.

Generally, the ECM is important for the following reasons

1. It's convenience lies in it's ability to measure the correction from disequilibrium of the previous periods.

2. With co-integration, ECM models are constructed in terms of first differences, and this ordinarily eliminates trend patterns from the variables involved, hence resolving the problem of spurious regressions.

The ECM can also be specified in multivariate form. In order to do this, we begin with a VAR representation of the time series. VARs are some of the most commonly used multivariate time series models. They can be estimated using least squares (LS), maximum likelihood (ML) or Bayesian methods. VAR models are mainly used for forecasting and structural analysis.

The linear model represented in (4.5) qualifies to be called a MA representation of time series. MAs are mainly useful in forecasting for example when computing the covariance of the forecast error, they come in handy. On the other hand, (4.13) is an AR representation of a time series model. It helps us understand how p_t depends on p_{t-1}.

If a time series has both the characteristics of stationarity and invertibility, then the AR and MA representations will be equivalent and one can easily be obtained from the other. However, neither AR nor MA representations are particularly useful for estimation

if they have too many coefficient matrices. To ease this problem, it can be postulated that coefficient matrices depend only on a finite number of parameters. This consideration leads to the vector autoregressive moving average (VARMA) or multivariate autoregressive moving average (MARMA). A general VARMA(pq) is written as

$$\boldsymbol{p_t} = \boldsymbol{\phi_0} + \sum_{i=1}^{p} \boldsymbol{\phi_i} \boldsymbol{p_{t-i}} + \sum_{j=1}^{q} \theta_j \boldsymbol{a_{t-i}} + \boldsymbol{a_t} \tag{4.30}$$

with p and $q \geq 0$, $\boldsymbol{\phi_0} = k-$ dimensional constant vector, $\boldsymbol{\phi_i}$ and $\boldsymbol{\theta_j} = k \times k$ constant matrices and $\{\boldsymbol{a_t}\} \sim$ i.i.d $(0, \boldsymbol{\Sigma_a})$. Using the back shift operator, we can write $\boldsymbol{\phi(B)p_t} = \boldsymbol{\phi_0} + \boldsymbol{\theta(B)a_t}$ where $\phi(B) = I_k - \phi_1(B) - \ldots - \phi_p B^p$ and $\theta(B) = I_k - \theta_1(B) - \ldots - \theta_q B^q$ are matrix polynomials in B. If we consider a $k-$dimensional process, which is integrated of order d so that $\{\boldsymbol{p_t} \sim I(d)\}$, A general VAR(p) model is written as

$$\boldsymbol{p_t} = \boldsymbol{\phi_0} + \sum_{i=1}^{p} \boldsymbol{\phi_i} \boldsymbol{p_{t-i}} + \boldsymbol{a}_t \tag{4.31}$$

with $p \geq 0$, $\boldsymbol{\phi_0} = k-$ dimensional constant vector, $\boldsymbol{\phi_i} = k \times k$ constant matrix and $\{\boldsymbol{a_t}\} \sim$ i.i.d $(0, \boldsymbol{\Sigma_a})$. Using the back shift operator, we can write $\boldsymbol{\phi(B)p_t} = \boldsymbol{\phi_0} + \boldsymbol{\theta(B)a_t}$ where $\phi(B) = I_k - \phi_1(B) - \ldots - \phi_p B^p$ is the matrix polynomial in B. In a VAR model, each variable is expressed by it's own lagged values and the lagged values of all the other variables in the system. If the variables are co-integrated vector autoregressive (CVAR), we also include the co-integrating vectors that pull the entire system towards a long run equilibrium. When the variables in a VAR model are co-integrated, a VECM model is used.

In general, the variables in a k-dimensional process $\{\boldsymbol{p}_t\}$ are said to be co-integrated of order (d, b), written as, $\boldsymbol{p}_t \sim CI(d, b)$, if all the components of $\{\boldsymbol{p}_t\}$ are $I(d)$ and there exists a linear combination $\boldsymbol{a}_t := \boldsymbol{\beta'} \boldsymbol{p}_t$ with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta k)' \neq 0$ such that $\boldsymbol{a}_t \sim I(d-b)$ (Lütkepohl, 2007). Co-integrated variables are, for the most part unstable, in their levels, although their “spreads" (generalized by the co-integrating relation) exhibit mean-reverting tendencies that make the variables move around common stochastic trends.

Co-integration is also distinguished from the short-run synchronies of positive covariance, which only measures the inclination to co-evolve at each point in time. Am improvement of the VAR model to incorporate co-integrated variables adjusts the short-run dynamics of the system with the long-term inclinations.

If we consider the general VAR(p) model

$$\boldsymbol{p_t} = \sum_{i=1}^{p} \boldsymbol{\phi_i p_{t-i}} + \boldsymbol{a}_t,$$

then, by differencing we have

$$\begin{aligned}
\Delta \boldsymbol{p_t} = \boldsymbol{p}_t - \boldsymbol{p}_{t-1} =& (\phi_1 - 1)p_{t-1} + \phi_2 p_{t-2} + \ldots + \phi_p p_{t-p} + a_t \\
=& -(1 - \boldsymbol{\phi}_1)\,\boldsymbol{p}_{t-1} + \boldsymbol{\phi_2 p_{t-2}} + \ldots + \boldsymbol{\phi_p p_{t-p}} + a_t \\
=& -(1 - \boldsymbol{\phi}_1 - \boldsymbol{\phi}_2 - \boldsymbol{\phi}_3 - \ldots - \boldsymbol{\phi}_{p-1} - \boldsymbol{\phi}_p)\,\boldsymbol{p}_{t-1} \\
& -(\boldsymbol{\phi}_2 + \boldsymbol{\phi}_3 + \ldots + \boldsymbol{\phi}_{p-1} + \boldsymbol{\phi}_p)\,\boldsymbol{p}_{t-1} \\
& -(-\boldsymbol{\phi}_2 - \boldsymbol{\phi}_3 - \ldots - \boldsymbol{\phi}_{p-1} - \boldsymbol{\phi}_p)\,\boldsymbol{p}_{t-2} \\
& -(\boldsymbol{\phi}_3 + \ldots + \boldsymbol{\phi}_{p-1} + \boldsymbol{\phi}_p)\,\boldsymbol{p}_{t-2} \\
& \vdots \\
& -(-\boldsymbol{\phi}_{p-1} - \boldsymbol{\phi}_p)\,\boldsymbol{p}_{t-(p-1)} \\
& -(+\boldsymbol{\phi}_p)\,\boldsymbol{p}_{t-(p-1)} \\
& -(-\boldsymbol{\phi}_p)\,\boldsymbol{p}_{t-p} + \boldsymbol{a}_t \\
\Delta \boldsymbol{p_t} =& \boldsymbol{\Pi p}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{p}_{t-i} + \boldsymbol{a}_t, \text{ where} \\
\boldsymbol{\Pi} =& (1 - \boldsymbol{\phi}_1 - \boldsymbol{\phi}_2 - \boldsymbol{\phi}_3 - \ldots - \boldsymbol{\phi}_{p-1} - \boldsymbol{\phi}_p) \text{ and} \\
\boldsymbol{\Gamma}_i =& -(\boldsymbol{\phi}_i + \ldots + \boldsymbol{\phi}_p)
\end{aligned}$$

If the rank r of the matrix $\boldsymbol{\Pi}$ is zero, there is no co-integration, no stable long-run relationship between variables and VECM is not possible, only VAR in first differences,

so that the equation

$$\Delta \boldsymbol{p_t} = \boldsymbol{\Pi} \boldsymbol{p}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{p}_{t-i} + \boldsymbol{a}_t \tag{4.32}$$

reduces to a VAR model for the differences

$$\Delta \boldsymbol{p_t} = \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{p}_{t-i} + \boldsymbol{a}_t. \tag{4.33}$$

One way of testing for this is to test if the rank of $\boldsymbol{\Pi}$, $r = 0$. If $r = 0$, the the variables are not co-integrated.

If the matrix is full rank i.e. $r = k$, then all the variables in $\boldsymbol{p_t} \sim I(0)$, and there is need to estimate the model as VECM since VAR on untransformed data is okay.

In fact the rank r denotes the number of co-integrating vectors. If $0 < r < k$, there are r co-integrating vectors describing the long-run relationships between variables. in this case, a VECM is okay, and the $k \times k$ matrix $\boldsymbol{\Pi}$ can be written as

$$\boldsymbol{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta'}, \tag{4.34}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $k \times r$ matrices, though this representation is of course not unique. For the sake of interpretation, it is usually convenient to normalize or identify the co-integrating vectors by selecting a specific coordinate system in which to express the variables. An arbitrary normalization, suggested by Johansen (1995), is to solve for the triangular representation of the co-integrated system. The elements of $\boldsymbol{\beta' p}_{t-1}$ may be interpreted as equilibrium relations and the elements of $\boldsymbol{\alpha}$, as adjustment coefficients which multiply the co-integrating relationship $\boldsymbol{\beta' p}_{t-1}$ to help counterbalance the deviations from the equilibrium. α can also be considered a loading matrix since it determines into which equation the co-integrating vectors enter and with what magnitudes. Normally, these coefficients would be expected to be negative.

Suppose $\boldsymbol{p}_t$ is $I(1)$ with $E[\boldsymbol{p}_t] = 0$, then (4.34) represents a VECM and the rank r of the matrix $\boldsymbol{\Pi}$ determines the number of co-integrating relationships. If $\boldsymbol{\Pi}$ is full rank, then the system $\boldsymbol{p}_t$ is stationary in levels. $\boldsymbol{\Pi}\boldsymbol{p}_{t-1}$ is the only potentially non-stationary term in the VECM. But since all the terms are stationary, the elements of $\boldsymbol{\Pi}\boldsymbol{p}_{t-1}$ must be stationary linear combinations of the non-stationary elements of $\boldsymbol{p}_{t-1}$. A VAR representation for the levels may well be consistent with co-integrating relationships. A VECM model with a reduced-rank error-correction coefficient is often called a co-integrated VAR model.

The ECM also fulfils the assumptions of the classical normal linear regression model (CNLRM) which include a linear regression model, normally distributed residuals, no serial autocorrelation of residuals, and no multicollinearity. As such, diagnostic tests must be carried out to ensure these assumptions are not violated. Among the tests employed are JB (Jarque and Bera, 1987) to determine the normality of the ECM, LM (Engle, 1984) test for ARCH effects and LB (Ljung and Box, 1978) for serial autocorrelation or cross correlation in the residuals.

4.5.1 Estimation and testing for co-integration

Testing for co-integration is necessary in checking if the models being built are empirically meaningful and if there is no evidence of cointegration, it is necessary to continue working with time series data in differences. Two or more time series which are integrated of order one are considered co-integrated in the event that it is conceivable to make a meaningful linear regression of one process over the other(s). Generally, it is impossible to make any meaningful linear regression of one integrated process over another. However, if the two processes are co-integrated, regression will be possible. During analysis, sometimes, completely unrelated time series may appear to have a relationship if tests are carried out using conventional testing procedures. Significant relationships may sometimes be obtained from variables that are in fact unrelated when non-stationary time series are used in a regression model. This phenomenon is known as spurious regression (Granger

and Newbold, 1974).

If two random walk series $Y_t = Y_{t-1} + u_t$ and $X_t = X_{t-1} + \varepsilon_t$, where each of the errors are white noise, are used in a regression , for instance, in $Y_t = \alpha + \beta X_t + z_t$, where z_t is a white noise error, spurious regression can be obtained. We could get a highly significant β estimate and a high value for the R^2, and we could be tempted to conclude that X has a significant impact on Y, when in fact there should be none. This regression could well be meaningless, unless the series are co-integrated. It is therefore important to discriminate between spurious regressions between series that are unrelated and genuine relationships which occur when time series are co-integrated. According to Granger and Newbold (1974), an $R^2 > d$ where $d = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$ is the Durbin-Watson (D-W) (Durbin and Watson, 1950, 1951) statistic could be a good sign that there is spurious regression. The D-W statistic tests the null hypothesis that residuals from a regression analysis are not autocorrelated against the alternative that the residuals follow an AR(1) process. In value, the D-W statistic ranges from 0 to 4. A value close to 0 indicates positive autocorrelation, a value close ti 2 indicates no autocorrelation and a value close to 4 implies negative autocorrelation in the sample. Small values of d indicate successive error terms are, on average, close in value to one another, or positively correlated.

The co-integration property allows regression of one integrated series over other integrated series (Fabozzi et al., 2006). Testing for co-integration implies testing the theory that a two or more non stationary series are related in the long run. Because of testing this long run relationship, the length of the data under consideration is very crucial and using a sufficiently long period of data would be important, in order to detect the common long term trends among the series under consideration (Alexander, 2008). The two most common methods used for testing co-integration among series are Engle and Granger (1987), which is based on an OLS regression and Johansen (1988) also found in (Johansen, 1991; Johansen and Juselius, 1990) which is based on eigenvalue analysis.

4.5.1.1 Engle-Granger Test for cointegration

The residual based E-G test (Engle and Granger, 1987) follows a two step procedure. Having verified that the series are $I(1)$, the first step involves fitting a static OLS regression model in order to capture any potential long-run relationship between the series, and then carrying out a stationarity test on the residuals of this OLS regression. The second step describes the dynamic adjustment of the series towards an equilibrium like the one described in (4.35). Consider two series x_t and y_t which are both $I(1)$, with long-run equilibrium relationships that can be captured by the regression models

$$\begin{aligned} x_t &= \alpha_0 + \alpha_1 y_t + u_t \text{ and} \\ y_t &= \beta_0 + \beta_1 x_t + v_t \end{aligned} \tag{4.35}$$

where $\alpha_0, \alpha_1, \beta_0$ and β_1 are co-integrating parameters and u_t and v_t are OLS residuals which capture any divergences between the series from the assumed equilibrium long-run relationship.

Since the E-G method produces only one co-integrating vector, it involves pairwise comparison of two co-integrating regressions and this makes it sensitive to the choice of the dependent variable. Testing for cointegration using the E-G representation test is essentially equivalent to testing for unit roots in the estimated residual series $\hat{u}_t$ and $\hat{v}_t$ for the regressions represented in (4.35) using the DF test. The first difference of the residuals is regressed on the lagged levels of the residuals without a deterministic trend as shown in (4.36) below. If a deterministic trend were to be included, it would mean that the errors would grow steadily over time, and this would violate the very idea of cointegration.

$$\begin{aligned} \Delta\hat{u}_t &= \rho_1 \hat{u}_{t-1} + \varepsilon_t \text{ and} \\ \Delta\hat{v}_t &= \rho_2 \hat{v}_{t-1} + \epsilon_t \end{aligned} \tag{4.36}$$

The E-G test requires the error terms to be serially uncorrelated and this problem can be addressed using the ADF test which accommodates more lags of the first difference of

the residuals to eliminate the serial correlation and can be set up as shown in (4.37)

$$\begin{aligned} \Delta\hat{u}_t &= \rho_1\hat{u}_{t-1} + \sum_{i=1}^{p}\gamma_i\Delta\hat{u}_{t-i} + \varepsilon_t \text{ and} \\ \Delta\hat{v}_t &= \rho_2\hat{v}_{t-1} + \sum_{i=1}^{p}\zeta_i\Delta\hat{v}_{t-i} + \epsilon_t \end{aligned} \tag{4.37}$$

It should however be noted that the distribution for the residual based ADF-test is not the same as the usual DF-distributions and will depend on the number of estimated parameters in the static regression above since additional variables in the static regression will shift the DF-distributions to the left. For the ADF test, it is therefore of extreme importance to find the correct augmentation lag length of $\Delta\hat{u}_{t-i}$ and $\Delta\hat{v}_{t-i}$. A rule of thumb is to count from lag $i = 1$ and include all significant lags plus one additional one (Sjö, 2008). If the hypothesis of the existence of a unit root is rejected then the conclusion is that x_t and y_t are co-integrated. Absence of a co-integrating relationship means the residuals have the same stochastic trends as the dependent variable. Asymptotically, this test is independent of the variable occurring on the left hand side of the co-integrating regression. The co-integrating vector is said to be normalised around the variable chosen to be on the left hand side of the co-integrating regression.

The error correction term, which represents the deviations from the equilibrium in (4.35) may be defined as

$$\begin{aligned} \hat{u}_t &= x_t - \alpha_0 - \alpha_1 y_t \text{ and} \\ \hat{v}_t &= y_t - \beta_0 - \beta_1 x_t, \end{aligned} \tag{4.38}$$

considering the estimated cointegration parameters. Under cointegration, $\hat{u}_t$ and $\hat{v}_t$ are stationary processes, and since their estimates will converge very quickly to their true values, $\hat{u}_{t-1}$ and $\hat{v}_{t-1}$ can be included as fixed regressors in a time varying model. Given $\hat{u}_{t-1}$ and $\hat{v}_{t-1}$, the second step in the E-G procedure therefore is to estimate the error-

correction model as,

$$\begin{aligned}\Delta x_t &= \delta + \lambda_1 \Delta x_{t-1} + \kappa_0 \Delta y_t + \kappa_1 \Delta y_{t-1} + \rho_1 \hat{u}_{t-1} + \varepsilon_t \text{ and} \\ \Delta y_t &= \gamma + \lambda_2 \Delta y_{t-1} + \omega_0 \Delta x_t + \omega_1 \Delta x_{t-1} + \rho_2 \hat{v}_{t-1} + \epsilon_t.\end{aligned} \tag{4.39}$$

All the terms in the error correction models are stationary and therefore, the standard procedures for inference will apply to all the parameters, and as such, $t-$ratios will be $N(0,1)$ asymptotically.

Note that the constant term can be left out of equations (4.39) above in order to improve the efficiency of the estimate. This test has the advantage that it is easy ti perform and it is intuitive. The major drawback of the E-G method is that variables are treated as if they are asymmetric in the sense that one is specified as the dependent variable and the other as the independent variable. However, this problem is addressed by the Johansen (1988) approach.

4.5.1.2 Johansen's test for co-integration

The Johansen (1988) test provides a method to test for co-integration in multivariate time series, where there is a possibility of more than one co-integrating vector being present since the variables in the model might form several equilibrium relationships. The procedure builds co-integrated variables directly on MLE instead of just relying on the OLS estimates. After it has been established that the variables are all $I(1)$, then, Johansen's procedure can be applied to determine the presence of a stable long-term relationship between variables and establish how many co-integrating vectors present. Generally, there can be a possible $k-1$ vectors, where k represents the actual number of variables that the model includes.

The test can be perceived as a multivariate generalization of the ADF test. It is concerned with testing hypotheses about the rank r of the co-integration space. It generally examines linear combinations of variables for unit roots, while following the same princi-

ples as the E-G approach with respect to the order of integration. Based on the results of this test, the long-run coefficients $\boldsymbol{\beta}$ and the speed adjustment coefficients $\boldsymbol{\alpha}$ can be determined, and the resultant ECM produced.

This test is founded on the examination of the long-run coefficient matrix $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, described in (4.34), so that testing for co-integration between variables is achieved by examining the rank of $\mathbf{\Pi}$ through the eigenvalues. The test procedure produces two statistics useful in determining the number of co-integrating vectors. The first Johansen tests is a likelihood ratio test that is based on the maximum eigenvalue of the stochastic matrix and the second test is based on the trace on the stochastic matrix. The two Johansen test are therefore known as the maximum eigenvalue test, and the trace test. Before the estimation of the parameters of a VECM, a choice must be made on the number of lags to be included in the underlying VAR model, specification of the trend, and the number of co integrating relations.

The general steps for Johansen's methodology are detailed below:

1. Test the order of integration of all variables.

2. Chose the lag length that would be appropriate for the model. The is commonly done by specifying and estimating a VAR(p) model for $\boldsymbol{Y}_t$ using a large number of lags, then successively reducing the lags by re-estimating the model for one lag less each time until lag zero is reached. Having done the diagnostic checks attributed to autocorrelation, possible ARCH effects, heteroscedasticity and normality of the residuals, the model with the smallest AIC is then selected as the one with the optimal lag length. Since omitted variables automatically become part of the error term, omitting variables that might only affect the short-run behaviour of the model affects the value of the lag length.

3. Choose an appropriate model with regard the deterministic components in the multivariate system. In doing this, check whether is is necessary to introduce an intercept

and/or trend in either the short-run or the long-run model, or both.

4. Conduct the likelihood ratio tests for the rank of the matrix $\mathbf{\Pi}$ so as to determine the number of co-integrating vectors.

5. Where necessary, impose some normalizing and identification restrictions on the co-integrating vectors.

6. From the normalized co-integrating vectors, get an estimate of the resultant co-integrated VECM using ML.

For both test statistics, the initial Johansen test is a test of the null hypothesis that there is no co-integration against the alternative that there is co-integration.

The tests differ in terms of their alternative hypotheses. For the trace test, the null hypothesis states that the number of co-integrating vectors is less than or equal to the rank r against an alternative that the co-integrating vectors present are more than r. Restricting the number of co-integrating vectors to be r or less implies that where k is the maximum number of possible co-integrating vectors, the remaining $k - r$ eigenvalues are equal to zero. Johansen (1988) derives the distribution of the trace statistic

$$LR_{trace} = -T \sum_{i=r+1}^{k} \ln\left(1 - \hat{\lambda}_i\right) \tag{4.40}$$

where T represents the number of observations and $\hat{\lambda}_i$ represents the $k - r$ estimated eigenvalues. For any given values of r, a large value of the trace statistic is evidence against the null hypothesis of r or fewer co-integrating relations in the VECM.
Note: The test is not based on the trace of $\mathbf{\Pi}$

For the maximum eigenvalue test, the null hypothesis states that the number of co-integrating vectors is r, against an alternative of $r + 1$. The Johansen ML procedure requires that the trace and maximum eigenvalue statistics are first calculated, then compared with appropriate critical values. However, the likelihood ratio statistic does not

have the usual asymptotic χ^2 distribution, instead it follows a multivariate equivalent of the DF unit root distribution which depends on the dimension $n-r$ and the specification of the deterministic terms. Critical values for this distribution for $n-r=1,\ldots,10$ are tabulated in (Osterwald-Lenum, 1992). The maximum eigenvalue statistic is given by

$$LR_{max} = -T\ln\left(1-\hat{\lambda}_{r+1}\right) \tag{4.41}$$

where T represents the number of observations and $\hat{\lambda}_r$ is the i^{th} largest canonical correlation.

Johansen's ML approach presents a number of merits over the two stage E-G approach to cointegration. Since this technique is based on VAR, there is little concern as to whether the explanatory variables are either exogenous or endogenous. Using this method, restrictions can be applied to the co-integrating vectors, something that is not possible with the E-G approach. Johansen's ML test can also be applied in testing *Granger-Causality*, where the lags in the ECM can be jointly tested for significance, hence determining any short-run causality from the explanatory or independent variables to the dependent variable.

However there is a particular issue related with this approach since there will be possibly more than one co-integrating vector. This implies that there are possibly more than one set of long-run coefficients and resultant error correction models. To identify the most appropriate model, either a choice is made of the model with the most appropriate signs, or certain specific over-identifying restrictions are employed, although both approaches can be problematic.

4.5.2 Co-integration analysis of energy markets

Co-integration theory can be applied to study market integration. From the analysis in section 3.3, it was established that all these series are non-stationary and integrated of order one. The two stage E-G method is used to test the existence of any co-integrating

relationship between these six series. In this test, we take the crude futures CF series as the dependent variable and obtain the results summarised in table 4.1.

Table 4.1: Results obtained from the E-G test

Parameter	Value	$t-$statistic	$p-$value	SE
α_0	0.511765	10.69652	0.0000	0.047844
α_1	0.986809	625.7656	0.0000	0.001577
α_2	-0.002571	-1.251454	0.2109	0.002055
α_3	0.001709	1.004433	0.3153	0.001701
α_4	-0.027940	-5.154723	0.0000	0.005420
α_5	0.035578	6.805134	0.0000	0.005228

Though the $R^2 = 0.9994$, the D-W statistic is 0.7545 signalling a spurious regression. The null hypothesis of no co-integration is rejected with a $p-$value < 0.001 implying that indeed these series are co-integrated with the single co-integrating vector given by

$$\boldsymbol{\beta}' = [1, 0.986809, -0.02571, 0.001709, -0.027940, 0.035578] \tag{4.42}$$

and the long-run relationship given by

$$\begin{aligned} CF = {} & 0.5118 + 0.986809CS - 0.02571GF + 0.001709GS - 0.027940HF \\ & + 0.035578HS \end{aligned} \tag{4.43}$$

The co-integrating residuals which give the co-integrating relation are given by the representation

$$\begin{aligned} \hat{a}_t = {} & CF - 0.5118 - 0.986809CS + 0.02571GF - 0.001709GS + 0.027940HF \\ & - 0.035578HS, \end{aligned} \tag{4.44}$$

and the plot of the co-integrating relations (the error correction term), which depicts stationarity, is shown in figure 4.4. Testing for stationarity using the ADF test gives a $p-$value < 0.001 implying that we should reject the null hypothesis of existence of unit roots. However, as has been discussed previously, one of the main drawbacks of the E-G technique of testing for co-integration is that it only identifies a single co-integrating relation, from amongst what could be many such relations. When we consider more than

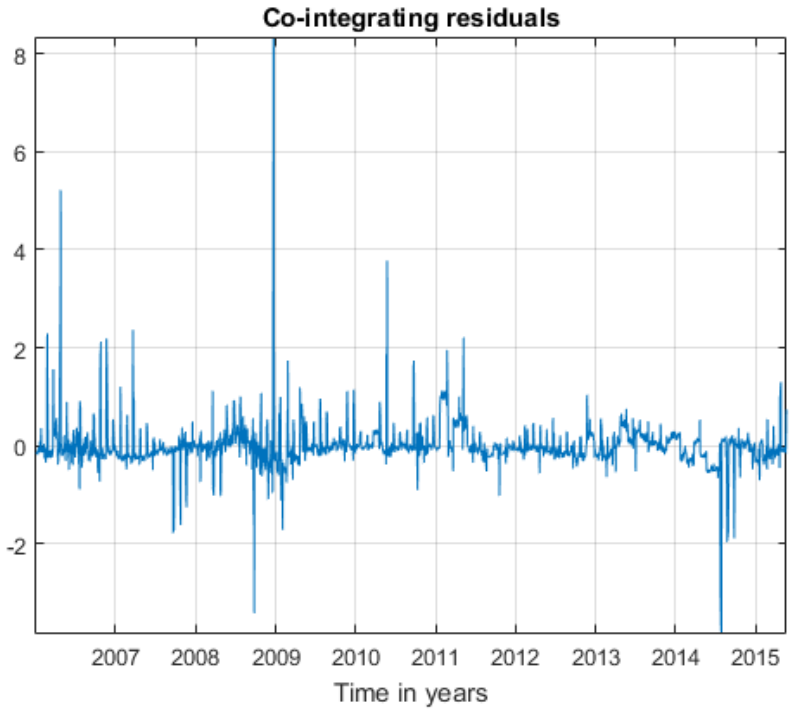

Figure 4.4: Co-integrating residuals for the co-integrating relationship in the six series taking crude futures prices as the dependent variable.

two series, we will have a different co-integrating relationship for every dependent variable specified as shown in figure 4.5 where each of the six price series under consideration is taken to be the dependent variable. In all the six relationships, the null hypothesis of no co-integration is rejected with $p-$values < 0.001 in all the six instances indicating six co-integrating relations.

For the Johansen co-integration trace test, we examine whether the rank of the matrix $\mathbf{\Pi} = r$. As such, testing begins sequentially for $r = 1, 2, \ldots, k$ and the first instance of non-rejection of the null hypothesis is taken as an estimate for r. In our case this happens for $r = 4$ at 4 lags with a $p-$value of 0.0746 as shown in table 4.2. Lag 4 is the smallest lag for which we obtain a rank of 4, making this the optimal lag length. The trace test indicates 4 co-integrating equations at the $\alpha = 0.05$ level and * denotes rejection of the null hypothesis at the 0.05 level

The co-integrating relations are as shown in figure 4.6 and the resultant parameters from

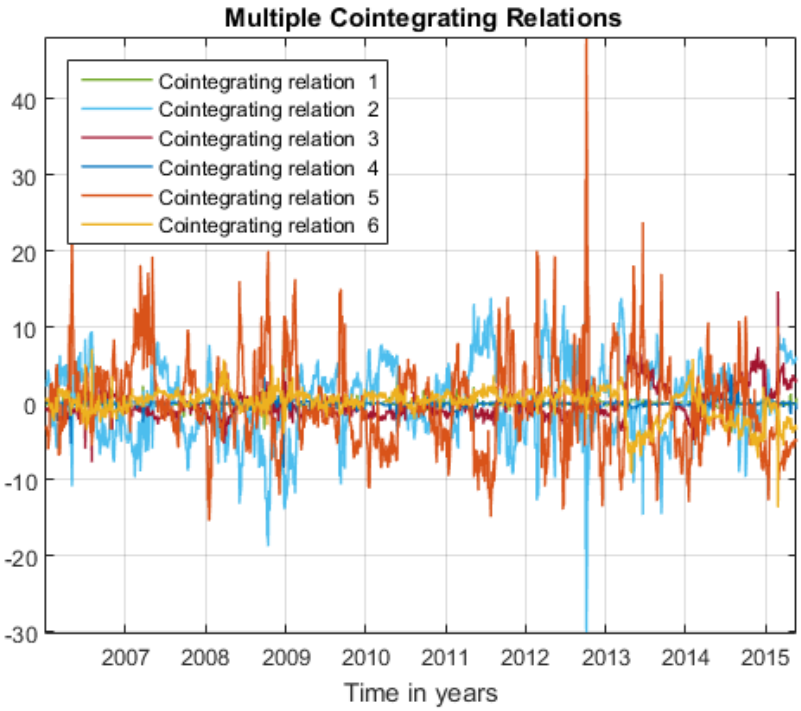

Figure 4.5: Six co-integrating relations obtained when all six series are considered as dependent variables

Table 4.2: Results obtained from Johansen co-integration test

rank	trace statistic	critical value	$p-$value	eigenvalue
0*	444.1103	95.7541	0.0010	0.1150
1*	155.9466	69.8187	0.0010	0.0376
2*	65.5717	47.8564	0.0010	0.0132
3*	34.3072	29.7976	0.0143	0.0084
4	14.3214	15.4948	0.0746	0.0044
5*	3.9785	3.8415	0.0461	0.0017

MLE are the co-integrating speed parameters

$$\boldsymbol{\alpha} = \begin{Bmatrix} 0.0991 & 0.0008 & 0.1401 & 0.0100 \\ 0.1803 & -0.0336 & 0.1195 & 0.1037 \\ 0.0847 & 0.0396 & 0.1639 & -0.0170 \\ 0.2289 & 0.0022 & 0.1364 & 0.0050 \\ 0.1508 & 0.4087 & 0.0775 & 0.0913 \\ 0.1001 & 0.0305 & 0.1070 & -0.0324 \end{Bmatrix}$$

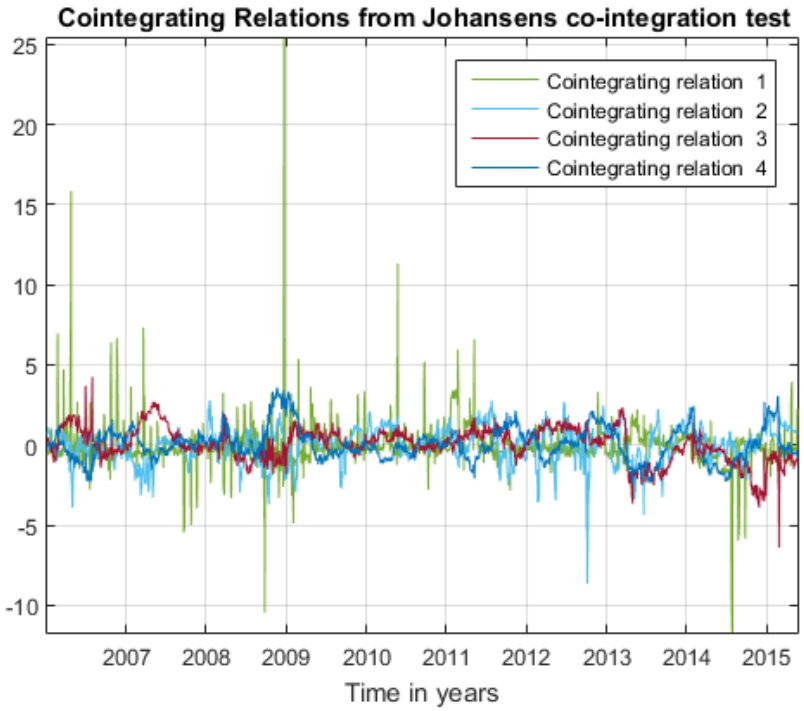

Figure 4.6: Four co-integrating relations obtained using the Johansen's trace test

and the vector of co-integrating relations

$$\boldsymbol{\beta} = \begin{Bmatrix} 3.0080 & 0.1463 & -0.1334 & -0.1482 \\ -0.0017 & 0.1818 & 0.0643 & -0.0817 \\ 0.0870 & -0.1898 & -0.4645 & -0.1300 \\ -2.9847 & -0.1302 & 0.0974 & 0.0821 \\ 0.0066 & -0.1813 & 0.0136 & -0.0022 \\ -0.1008 & 0.1706 & 0.4150 & 0.2435 \end{Bmatrix}$$

so that for most cases, the 6×6 matrix $\mathbf{\Pi}$ represented in (4.34) would be

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' = \begin{Bmatrix} 0.2782 & 0.0082 & -0.0579 & -0.2816 & 0.0024 & 0.0507 \\ 0.5062 & -0.0072 & -0.0469 & -0.5137 & 0.0087 & 0.0509 \\ 0.2412 & 0.0190 & -0.0741 & -0.2434 & -0.0043 & 0.0621 \\ 0.6699 & 0.0084 & -0.0445 & -0.6698 & 0.0029 & 0.0351 \\ 0.4897 & 0.0715 & -0.1123 & -0.4884 & -0.0722 & 0.1089 \\ 0.2962 & 0.0149 & -0.0426 & -0.2951 & -0.0033 & 0.0317 \end{Bmatrix}. \quad (4.45)$$

However, in this analysis, our full VECM is given as

$$\Delta \boldsymbol{p}_t = \boldsymbol{\Gamma}_0 + \sum_{i=1}^{4} \boldsymbol{\Gamma}_i \Delta \boldsymbol{p}_{t-1} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}' \boldsymbol{p}_{t-1} + \delta\right) + \boldsymbol{a}_t \tag{4.46}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are given as above, $\boldsymbol{p}_t' = [CF_t, CS_t, GF_t, GS_t, HF_t, HS_t]$,
$\boldsymbol{\delta}' = [-1.4215, 1.6461, 0.9019, 2.5075]$ are intercepts in the co-integrating relations, $\boldsymbol{\Gamma}_0' = [-0.0045, 0.0017, 0.0021, -0.0009, -0.0002, 0.0021]$, and $\boldsymbol{a}_t$ are the shocks or innovations associated with each of the relations,

$$\boldsymbol{\Gamma}_1 = \begin{Bmatrix} -0.3080 & 0.0253 & -0.1005 & 0.2390 & -0.0317 & 0.1120 \\ -0.5324 & 0.0092 & -0.1000 & 0.4972 & -0.0102 & 0.0732 \\ -0.2669 & 0.0049 & -0.3733 & 0.2746 & -0.0364 & 0.3705 \\ -0.3995 & 0.0306 & -0.1024 & 0.3167 & -0.0291 & 0.1200 \\ -0.3504 & 0.0815 & 0.1126 & 0.3070 & -0.0287 & -0.1938 \\ -0.2601 & 0.0460 & 0.1269 & 0.2649 & -0.0490 & -0.1698 \end{Bmatrix}$$

$$\boldsymbol{\Gamma}_2 = \begin{Bmatrix} -0.1898 & 0.0159 & -0.1731 & 0.1933 & 0.0085 & 0.1007 \\ -0.4415 & -0.0219 & -0.2134 & 0.4629 & 0.0007 & 0.1372 \\ -0.1942 & 0.0021 & -0.3479 & 0.2948 & 0.0129 & 0.2319 \\ -0.2013 & 0.0159 & -0.1646 & 0.1999 & 0.0075 & 0.0996 \\ -0.1517 & 0.0158 & -0.1349 & 0.2206 & -0.0216 & 0.0236 \\ -0.1870 & 0.0092 & -0.1063 & 0.2752 & 0.0168 & -0.0145 \end{Bmatrix}$$

$$\boldsymbol{\Gamma}_3 = \begin{Bmatrix} -0.2912 & -0.0254 & -0.0248 & 0.3225 & 0.0006 & 0.0213 \\ -0.4268 & -0.0866 & -0.0535 & 0.5186 & 0.0095 & 0.0366 \\ -0.2768 & -0.0503 & -0.2320 & 0.3808 & 0.0160 & 0.1827 \\ -0.1805 & -0.0330 & -0.0340 & 0.2683 & -0.0029 & 0.0077 \\ -0.2061 & -0.1020 & -0.0417 & 0.3759 & 0.0643 & -0.0527 \\ -0.3183 & -0.0349 & -0.0491 & 0.4017 & 0.0128 & 0.0058 \end{Bmatrix}$$

and

$$\mathbf{\Gamma}_4 = \begin{Bmatrix} -0.0854 & 0.0188 & -0.0183 & 0.1668 & -0.0166 & -0.0293 \\ -0.0418 & -0.0058 & -0.0020 & 0.1972 & -0.0123 & -0.0532 \\ -0.2077 & -0.0020 & -0.0483 & 0.2911 & -0.0197 & 0.0685 \\ -0.0743 & 0.0190 & -0.0155 & 0.1458 & -0.0163 & -0.0289 \\ 0.0470 & -0.0729 & 0.0273 & 0.1029 & 0.0394 & -0.0047 \\ -0.0756 & 0.0060 & 0.0202 & 0.1828 & -0.0084 & -0.0508 \end{Bmatrix}.$$

$\boldsymbol{\alpha}\left(\boldsymbol{\beta}'\Delta\boldsymbol{p}_{t-1} + \delta\right)$ represents the error correction term.

After fitting this model, an analysis of the residuals show that they are not white noise. Figure 4.7 shows plots for the residuals from the VECM. They do not seem to be multivariate normal, and this is confirmed by the results shown on table 4.3 which reports the multivariate extensions of the JB normality test on the residuals. This is a test of the null hypothesis that the skewness of the underlying distribution is zero, and the kurtosis is three. The measures of these two moments allow one to test hypotheses that are in harmony with the conjecture of multivariate normality.

This test is applicable since individual variables from a set of variables that are jointly multinomial distributed are also normally distributed, although if a number of variables are normally distributed individually, they are not necessarily also multivariate normal (Von Eye and Bogat, 2004). The test statistic is given by (4.47) below where $\hat{S}$ and $\hat{K}$ represent the sample skewness and sample kurtosis respectively.

$$JB = n\left\{\frac{\hat{S}^2}{6} + \frac{\left(\hat{K}-3\right)^2}{24}\right\} \tag{4.47}$$

The null hypothesis for this test is that the residuals are multivariate normal and this is rejected in this instance as can be seen from the $p-$values reported in table 4.3.

A visual inspection of figure 4.7 suggests the presence of volatility clustering. Further investigation of the residual series reveals the presence of ARCH effects as shown in table

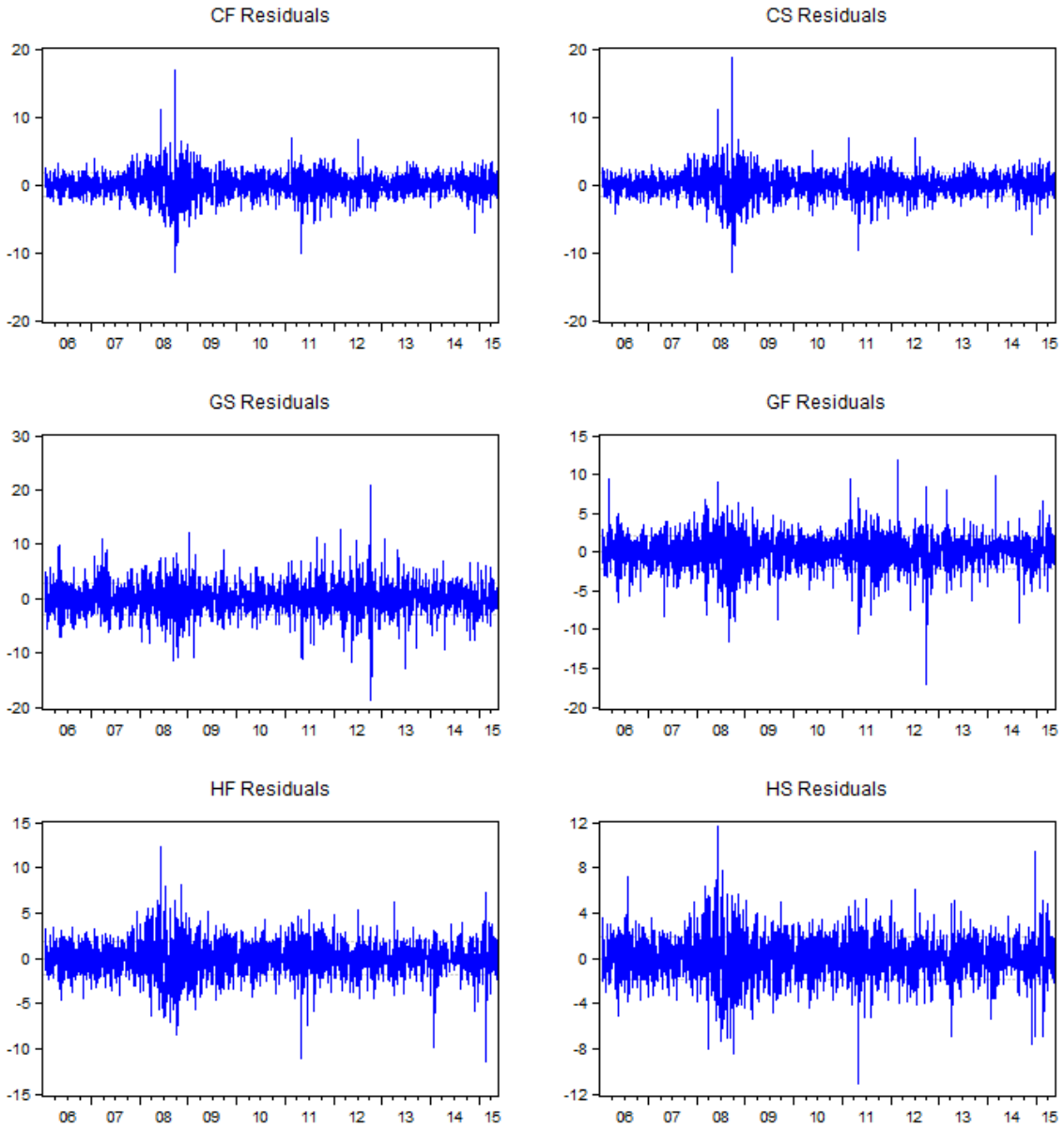

Figure 4.7: Residuals from the VECM

4.4. The null hypothesis of no ARCH effects is therefore rejected and a conclusion that the noise is therefore MGARCH is made.

4.5.3 Granger-Causality

Even when variables are not co-integrated in the long-run, they could still be related in the short-run. One time series can be said to have a causal influence on another if the incorporation of knowledge from the former improves the prediction of the later time series. Granger (1969) proposed a time-series data based approach that can be used to

Table 4.3: Multivariate normality test on residual series

Series	Skewness	χ^2 statistic	df	$p-$ value
CF	0.056937	1.274020	1	0.2590
CS	-4.240484	7066.810	1	0.0000
GF	-0.589093	136.3830	1	0.0000
GS	0.592101	137.7792	1	0.0000
HF	-1.138971	509.8211	1	0.0000
HS	-0.125048	6.145373	1	0.0132
Joint		7858.213	6	0.0000

Series	Kurtosis	χ^2 statistic	df	$p-$ value
CF	10.21375	5112.752	1	0.0000
CS	101.6174	955519.8	1	0.0000
GF	17.20323	19820.15	1	0.0000
GS	15.34322	14968.89	1	0.0000
HF	24.22178	44248.25	1	0.0000
HS	31.65732	80687.03	1	0.0000
Joint		1120357	6	0.0000

Series	JB statistic	df	$p-$ value
CF	5114.026	2	0.0000
CS	962586.6	2	0.0000
GF	19956.53	2	0.0000
GS	15106.67	2	0.0000
HF	44758.07	2	0.0000
HS	80693.18	2	0.0000
Joint	1128215	12	0.0000

determine causality. Specifically, two AR models are fitted to the first time series, in the first instance, with and in the second instance, without including the second time series, and any improvement on the forecasts is measured by the ratio of the variance of the error terms. A ratio bigger than one signifies an improvement, and hence a causal relationship. In the worst case scenario, the ratio is 1, implying causal independence from the second time series to the first. Particularly, if the variance of the AR forecasting error of the first time series is reduced by including past measurements from the second time series, then the second series has a causal influence on the first one. A variable x is said to Granger-causes y if the variable y is better predicted using the the past values of both x and y than it can be predicted, using its past values alone.

Definition 4.5.1 (Granger-Causality)**.** *Given an information set* Ω*, taking the form*

Table 4.4: Test for ARCH effects on residual series

Joint test:

χ^2 statistic	df	$p-$ value
6278.398	1176	0.0000

Individual components:

Dependent	R^2	$F_{(56,2301)}$	$p-$ value	$\chi^2(56)$	$p-$ value
res1*res1	0.233479	12.51559	0.0000	550.5426	0.0000
res2*res2	0.211971	11.05254	0.0000	499.8271	0.0000
res3*res3	0.085742	3.853501	0.0000	202.1805	0.0000
res4*res4	0.260093	14.44374	0.0000	613.2987	0.0000
res5*res5	0.142756	6.842568	0.0000	336.6190	0.0000
res6*res6	0.100405	4.586018	0.0000	236.7544	0.0000
res2*res1	0.222538	11.76124	0.0000	524.7442	0.0000
res3*res1	0.145531	6.998214	0.0000	343.1617	0.0000
res3*res2	0.146352	7.044489	0.0000	345.0987	0.0000
res4*res1	0.134115	6.364236	0.0000	316.2435	0.0000
res4*res2	0.131406	6.216239	0.0000	309.8558	0.0000
res4*res3	0.097981	4.463300	0.0000	231.0398	0.0000
res5*res1	0.127599	6.009785	0.0000	300.8780	0.0000
res5*res2	0.126302	5.939869	0.0000	297.8197	0.0000
res5*res3	0.133311	6.320207	0.0000	314.3473	0.0000
res5*res4	0.119110	5.555916	0.0000	280.8617	0.0000
res6*res1	0.100783	4.605250	0.0000	237.6472	0.0000
res6*res2	0.100042	4.567614	0.0000	235.8994	0.0000
res6*res3	0.101842	4.659118	0.0000	240.1439	0.0000
res6*res4	0.107252	4.936366	0.0000	252.9014	0.0000
res6*res5	0.100744	4.603270	0.0000	237.5553	0.0000

$x_t, \ldots, x_{t-j}; y_t, \ldots, y_{t-i}$ for i and $j = 1, 2, 3, \ldots, n$, we say that x_t Granger-causes y_t w.r.t Ω if the variance of the optimal linear predictor of y_{t+h}, based on Ω, has a smaller variance compared to the optimal linear predictor of y_{t+h} based only on the lagged values of y_t for any h. Thus x Granger causes y if and only if $\sigma_1^2(y_t : y_t - j, x_{t-i}) < \sigma_2^2(y_t : y_{t-j})$, where σ^2 represents the variance of the forecast errors (Foresti, 2006).

Granger causality tests aid in understanding short-term interdependencies among variables. Granger-causality tests can be applied in three different kinds of situations. In the first case where a simple Granger-causality test is considered, it is applied on two variables and their lags. In the second case where we consider more than two variables on the sup-

position that more than one variable can influence the results of another, a multivariate Granger-causality test is applied. Finally, if the multivariate model is extended so that the simultaneity of all included variables is considered, then Granger-causality is tested in a VAR framework. The standard F-test, which seeks to establish whether changing one variable can result in changes in another variable, is the basis of the Granger-causality test. The test statistic is formulated as

$$F = \frac{RSSE - USSE}{USSE}\left(\frac{T-k}{q}\right) \tag{4.48}$$

where $RSSE$ denotes the sum of squared errors from the restricted model, $USSE$ denotes the sum of squared errors the unrestricted model, T represents the sample size, k represents the number of lags and q represents the number of restrictions put in place. Under the null hypothesis, test statistic is distributed as χ^2 with k degrees of freedom. Considering a simple VAR model,

$$\boldsymbol{y_t} = \boldsymbol{\phi_0} + \sum_{i=1}^{p} \boldsymbol{\phi_i} \boldsymbol{y_{t-i}} + \sum_{i=1}^{q} \boldsymbol{\theta}_j x_{t-j} + \boldsymbol{a}_t \tag{4.49}$$

If all the values of θ are significantly different from zero, then X Granger-causes Y. After obtaining the VAR, restrictions are imposed and the hypotheses are such that for the null hypothesis, X does not Granger-cause Y or $H_0 : \theta_1 = \theta_2 = \ldots = 0$ against the alternative that X Granger-causes Y or H_1 : at least one $\theta \neq 0$. Based on the OLS coefficients estimated from the AR equations in a simple Granger-causality test, the results can unidirectional causality, bidirectional causality and independence between x and y. For any set of co-integrated time-series, there must be Granger-causality between them but, the converse is however not true.

For this study, the results for Granger-causality in levels are reported in table 4.5 and results for causality in differences are recorded in table 4.6. From table 4.5, we see bidirectional causality between almost all the series except for gasoline futures which doesn't Granger cause gasoline spot though gasoline spot granger causes gasoline futures.

This result is confirmed by the joint test also. This actually shows the presence of spillover effects across the series. From table 4.6, we see, returns from gasoline futures do not granger cause gasoline spot returns and we also not that gasoline spot returns only Granger causes heating oil futures and maybe heating oil spot returns.

Table 4.5: Results obtained from Granger-causality (F) test in levels

	CF	CS	GF	GS	HF	HS	Joint
CF	2356	254	239	195	1044	951	1819.2
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
CS	321	382	293	159	833	732	1564
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
GF	40	46	33369	6	54	40	146.6
	(< 0.001)	(< 0.001)	(< 0.001)	(**0.2132**)	(< 0.001)	(< 0.001)	(< 0.001)
GS	25	19	49	6077	36	34	103
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
HF	61	125	110	109	13655	148	458.9
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
HS	102	122	78	104	48	14178	455.3
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)

p-values are indicated in the parenthesis.

Table 4.6: Results obtained from Granger-causality (F) test in differences

	ΔCF	ΔCS	ΔGF	ΔGS	ΔHF	ΔHS	Joint
ΔCF	51.5493	81.5114	158.8971	231.7504	871.1525	598.9422	1463.8
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
ΔCS	37.6260	28.3827	210.8375	176.4722	857.9510	579.0799	1212.4
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
ΔGF	23.9313	32.1628	22.5417	3.8876	51.2044	30.3673	175
	(< 0.001)	(< 0.001)	(< 0.001)	(**0.4214**)	(< 0.001)	(< 0.001)	(< 0.001)
ΔGS	6.3281	2.8095	13.2817	6.7035	21.7478	14.5890	72.2
	(**0.176**)	(**0.5902**)	(**0.1**)	(**0.1524**)	(< 0.001)	(**0.0056**)	(< 0.001)
ΔHF	61.3676	192.8939	45.3097	122.6379	173.5127	112.1761	498.4
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
ΔHS	34.5446	90.3696	61.5954	114.1517	30.1818	27.3170	451.6
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)

p-values are indicated in the parenthesis.

4.6 Multivariate ARCH and GARCH Models

MGARCH models are multivariate parametric error distribution models which resemble univariate GARCH models in principle, except that for the multivariate case, the condi-

tional mean is allowed to follow a VAR structure and the conditional covariance matrix of the dependent variables are allowed the to follow a flexible dynamic structure. Financial market volatilities move simultaneously over time across both assets and markets. Viewing these commonalities through a multivariate modelling framework improves efficiency and creates opportunities for better decision making in areas like asset pricing and portfolio selection (Bauwens et al., 2012, 2006).

Let r_t denote a column vector of k asset returns, μ_t denote a column vector of the conditional means of r_t and $\Sigma_t = \sigma_{t,ij}$ be the conditional variance-covariance matrix of r_t. The elements of μ_t and Σ_t must be measurable with respect to the sigma-field $\mathcal{F}_{t-1}$ generated by the past information (here the r_t's) and possibly by any other variables available up to time $t-1$. In other words the sigma field generated by r_{t-j} for $j \geq 1$.
An MGARCH model for r_t can then be defined by

$$r_t = \mu_t + a_t \text{and } a_t = \Sigma_t^{1/2} \varepsilon_t \tag{4.50}$$

where $\Sigma_t^{1/2}$ which is the square root, or Cholesky decomposition, of the covariance matrix, is a $k \times k$ positive definite (PD) matrix such that $\Sigma_t = \Sigma_t^{1/2} \left[\Sigma_t^{1/2}\right]'$ and ε_t is an unobservable $k \times 1$ random vector whose moments are as shown below.

$$E\left[\varepsilon_t\right] = 0 \text{ and } \text{Var}(\varepsilon_t) = E\left[\varepsilon_t \varepsilon_t'\right] = I_k \text{ where } I_k \text{ is an identity matrix of order } k$$

This means ε_t is an unobservable random vector which is an i.i.d process whose mean equal to zero and variance-covariance matrix is an identity matrix.
Note that $E_{t-1}(a_t) = 0$, so that $\Sigma_t = \text{Var}\left(r_t | \mathcal{F}_{t-1}\right) = \text{Var}_{t-1}(r_t)$ so that $\text{Var}_{t-1}(a_t) = \Sigma_t$.

The conditional variance matrix of r_t can be calculated as

$$\begin{aligned} \text{Var}\left(r_t|\mathcal{F}_{t-1}\right) &= \text{Var}_{t-1}(r_t) \\ &= \text{Var}_{t-1}(a_t) \\ &= \Sigma_t^{1/2}\text{Var}_{t-1}\left(\varepsilon_t\right)\left(\Sigma_t^{1/2}\right)' \\ &= \Sigma_t. \end{aligned} \tag{4.51}$$

Notice also that

$$a_t|\mathcal{F}_{t-1} \sim N\left(0, \Sigma_t\right) \tag{4.52}$$

The conditional covariance matrix Σ_t has many possible specifications, but the parameters increase rapidly with the increase in the dimension of a_t, creating a difficulty during model estimation. When constructing MGARCH model, it is important to ensure parsimony is attained while still maintaining flexibility. In addition to this, we must ensure that the conditional covariance matrix Σ_t is PD.

The first form of specification involves the direct Generalization of univariate GARCH models as done by Bollerslev (1986). These include the full VEC and BEKK models (Engle and Kroner, 1995), flexible MGARCH, Factor models, Cholesky & Full factor GARCH models and Riskmetrics. Linear combinations of univariate GARCH models can also be considered. These include latent factor models and generalized orthogonal models. Finally, non-linear combinations of univariate GARCH can also be specified. They include CCC and DCC models, general dynamic covariance models and Copula based GARCH models.

4.6.1 Vech Representation

Let $vech(.)$ denote the vector half operator which converts the unique upper triangular elements of a $k \times k$ matrix into a $\left[\frac{1}{2}k(k+1)\right] \times 1$ vector. Since the conditional covariance

matrix Σ_t is symmetric, $vech\left(\Sigma_t\right)$ contains all the unique elements in Σ_t. For instance

$$vech\left(\begin{bmatrix}\sigma_{11,t} & \sigma_{12,t}\\ \sigma_{21,t} & \sigma_{22,t}\end{bmatrix}\right) = \begin{bmatrix}\sigma_{11,t}\\ \sigma_{12,t}\\ \sigma_{22,t}.\end{bmatrix}$$

The direct generalization of a univariate $GARCH(p,q)$ model gives

$$\begin{aligned} vech(\Sigma_t) &= W + \sum_{i=1}^{p} A_i^* vech(\varepsilon_{t-i}, \varepsilon_{t-i}') + \sum_{j=1}^{q} B_j^* vech(\Sigma_{t-j}) \\ &= W + A^*(L) vech(\varepsilon_t, \varepsilon_t') + B^*(L) vech(\Sigma_t), \end{aligned} \tag{4.53}$$

where, L is the lag operator, W is a $\left[\frac{1}{2}k(k+1)\right] \times 1$ vector, A_i^* and B_j^* are $\left[\frac{1}{2}k(k+1)\right] \times \left[\frac{1}{2}k(k+1)\right]$ matrices. This gives the *vech* representation (Engle and Kroner, 1995). Each element in Σ_{t-1} may affect each element in Σ_t, meaning the number of parameters will be $\left[\frac{1}{2}k(k+1) + (p+q)\left[\frac{1}{2}k(k+1)\right]^2\right]$.

Since for MGARCH models there are $k(k+1)/2$ variances and covariances for a $k-$dimensional process, they usually, suffer from the curse of dimensionality since, as estimating this many free parameters is obviously infeasible, both in terms of data availability and in numerical terms. Moreover, these parameters, need to be restricted for them to yield forecasts of the covariance matrix that are eventually positive semi-definite. This restriction is very difficult to check, let alone impose during parameter estimation. However, most MGARCH models reduce this dimensionality in some way.

4.6.2 Diagonal vech Representation

Under this representation, a restriction is imposed on the model such that every element in the covariance matrix only depends on it's past values, and $\varepsilon_{i,t}, \varepsilon_{j,t}$. A_i^* and B_j^* matrices

are all taken to be diagonal, for instance for $k = 2$ and $p = q = 1$, we have

$$\begin{bmatrix} \sigma_{11,t} \\ \sigma_{12,t} \\ \sigma_{22,t} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} + \begin{bmatrix} a_{11}^* & 0 & 0 \\ 0 & a_{12}^* & 0 \\ 0 & 0 & a_{33}^* \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}, \varepsilon_{2,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{11}^* & 0 & 0 \\ 0 & b_{12}^* & 0 \\ 0 & 0 & b_{33}^* \end{bmatrix} \begin{bmatrix} \sigma_{11,t-1} \\ \sigma_{12,t-1} \\ \sigma_{22,t-1} \end{bmatrix}$$

This trims down the number of parameters to be estimated to $\left[\frac{1}{2}k(k+1)\right](1+p+q)$ but it does not allow for causality in variance, co-persistence in variance and asymmetries.

4.6.3 BEKK Representation

This parametrization was proposed by Engle and Kroner, 1995, and is named after Baba, Engle, Kraft and Kroner (BEKK). Consider the model

$$\Sigma_t = CC' + \sum_{n=1}^{N}\sum_{i=1}^{p} A_{i,N}\varepsilon_{t-i}, \varepsilon_{t-i}' A_{i,N}' + \sum_{n=1}^{N}\sum_{j=1}^{q} B_{j,N}\Sigma_{t-j}B_{j,N}', \tag{4.54}$$

where C, $A_{i,N}$ and $B_{j,N}$ are $k \times k$ coefficient matrices. The intercept matrix is decomposed into CC' where C is the upper triangular matrix. This ensures that the conditional covariance matrix is PD, and this is one of the advantages of this model, though the number of parameters still remains quite high. For $N = 1$, the model in (4.54) is represented as

$$\Sigma_t = CC' + A\varepsilon_{t-1}, \varepsilon_{t-1}' A' + B\Sigma_{t-1}B', \tag{4.55}$$

and in this case,the number of parameters becomes $(p+q)k^2 + \frac{1}{2}k(k+1)$. A diagonal BEKK reduced the parameters to be estimated further, but this is still infeasible for large k. Even though there are a number of MGARCH models, the BEKK is more widely used.

4.6.4 Conditional Correlation Representation

Conditional correlation models are based on the concept of modelling correlations and conditional variances as opposed to modelling the conditional covariance matrix Σ_t. They were proposed by Bollerslev (1990), and they parametrize the dynamic conditional covariances as being proportional to the product of the corresponding conditional standard

deviations. With this assumption the computational burden of estimation in MGARCH models is greatly simplified, and the positive definiteness of Σ_t is assured. The conditional covariance matrix is decomposed into a correlation matrix and conditional standard deviation matrices as shown below.

$$\Sigma_t = D_t R_t D_t \tag{4.56}$$

where $D_t = diag\left(\sigma_{1,t}, \sigma_{2,t}, \ldots, \sigma_{k,t}\right)$ is a diagonal matrix with $\sigma_{i,t}$ as the i^{th} diagonal element. D_t is the conditional standard deviation and $R_t = \rho_{ij,t} = \dfrac{\sigma_{i,j,t}}{\sqrt{\sigma^2_{i,t}\sigma^2_{j,t}}}$ is the dynamic conditional correlation matrix of order k. $\sigma^2_{i,t}$ and $\sigma^2_{j,t}$ follow univariate GARCH processes. Notice that $\rho_{ii,t} = 1 \quad \forall i$ and $\forall t$. Σ_t is PD if $\sigma^2_{i,t}$ is positive $\forall i$ and R_t is PD. In these models therefore, the specification of Σ_t falls into two independent components. The first component is on the model of choice for the conditional variance matrix. This is obtained from quasi-maximum likelihood (QML) estimation of conditional variances as in a $GARCH(1,1)$ framework. The second component is on the model choice for the conditional correlation matrix. Conditional correlation models are either constant, CCC or dynamic, DCC.

For the CCC model, the conditional correlation is assumed to be constant, i.e. $R_t = R$, so that (4.56) can be given as

$$\Sigma_t = D_t R D_t \tag{4.57}$$

For $k = 2$ for instance, (4.57) can be represented as

$$\Sigma_t = \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{22,t} \end{bmatrix} = \begin{bmatrix} \sigma_{1,t} & 0 \\ 0 & \sigma_{2,t} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1,t} & 0 \\ 0 & \sigma_{2,t} \end{bmatrix} = \begin{bmatrix} \sigma_{11,t} & \rho_{12}\sigma_{12,t} \\ \rho_{21}\sigma_{21,t} & \sigma_{22,t} \end{bmatrix}$$

The correlation matrix, $R = \rho_{ij}$ is PD with $\rho_{ii} = 1 \quad \forall i$. The off diagonal elements of Σ_t, given by $[\Sigma_t]_{ij} = \rho_{ij}\sigma_{ij,t}$ for $i \neq j$ are the conditional covariances. The assumption of constant correlations over time is unrealistic, and this process excludes any volatility

transmission or spill over across markets since the conditional correlation is constant. The innovation process $\{a_{i,t}\}$ is modelled as a univariate GARCH so that the conditional variances can be expressed in form of a vector as

$$\sigma_t = c + \sum_{i=1}^{p} A_i a_{t-i}^2 + \sum_{j=i}^{q} B_j \sigma_{t-j} \tag{4.58}$$

where, c is a $k \times 1$ vector, A_i and B_j are diagonal $k \times k$ matrices and $a_{t-i}^2 = a_{t-i} \odot a_{t-i}$ is the element wise product given that $\odot$ is the Hadamard product operator. since R is PD, Σ_t is bound to be PD when he elements of c, A_i, and B_j are positive.

For the DCC, the conditional correlation matrix R_t is a function of t. This class of models was introduced by Engle and Sheppard (2001) and here, the covariance matrix is represented by the equation $\Sigma_t = D_t R_t D_t$ as in (4.56). For this case however, D_t is a $k \times k$ diagonal matrix of conditional standard deviations of the process $\{a_t\}$ at time t, R_t is a $k \times k$ conditional correlation matrix of $\{a_t\}$ at time t, while $\Sigma_t^{1/2}$ is any $k \times k$ matrix at time t such that Σ_t is the conditional covariance matrix of $\{a_t\}$. It may be obtained by Cholesky decomposition of Σ_t. With R_t varying, Σ_t is PD only if R_t is PD at each point in time and the conditional standard deviations $\sigma_{i,t}$ are well defined. This implies that during estimation, the correlation matrix has to be inverted for each time t and hence the advantage of numerical simplicity during estimation is lost. With this model however, volatility transmission or spill over across markets can be incorporated.

4.6.5 MGARCH analysis of energy markets

Analysis begins by checking the energy price returns for the presence of heteroscedasticity. If the process $\{a_t\}$ does not exhibit conditional heteroscedasticity, then it's conditional covariance matrix Σ_t will be constant, implying that Σ_t and hence a_t does not depend on a_{t-1}. This can be done using the LB statistic, which for this case, is be given by

$$Q_N^*(m) = k^2 \sum_{i=1}^{m} \frac{1}{k-i} b_i' \left(\hat{\rho}_0^{-1} \otimes \hat{\rho}_0^{-1} \right) b_i \tag{4.59}$$

where k is the sample size, N denotes the dimension of a_t, $b_i = vec(\hat{\rho}_i')$ and $\hat{\rho}_j$ is the sample cross correlation matrix of a_t at lag$-j$. The test hypothesizes that

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_m = 0 \text{ against } H_a : \rho_i \neq 0.$$

Under this null hypothesis, $Q_N^*(m)$ is asymptotically distributed as $\chi^2_{N^2m}$.

For our case the test statistic = 91252.83 and the $p-$ value < 0.00. This provides evidence that heteroscedasticity is present. Fitting a bivariate full BEKK on the crude futures and crude spot and gasoline futures and gasoline spot returns gives the results shown in Tables 4.7 and 4.8.

Table 4.7: Parameter estimates for BEKK(11) for crude futures and spot.

Coefficient(s):	Estimate	Std. Error	$t-$value	$Pr(>\|t\|)$
μ_{CF}	-1.84×10^{-4}	3.64×10^{-4}	-0.51	0.612
μ_{CS}	-7.41×10^{-5}	3.64×10^{-4}	-0.20	0.839
$A0_{11}$	4.72×10^{-3}	5.11×10^{-4}	9.23	< 0.001
$A0_{21}$	4.53×10^{-3}	5.08×10^{-4}	8.91	< 0.001
$A0_{22}$	1.61×10^{-3}	6.23×10^{-5}	25.8	< 0.001
A_{11}	0.797	0.12	6.48	< 0.001
A_{21}	-0.412	0.12	-3.45	< 0.001
A_{12}	-0.451	0.12	-3.78	< 0.001
A_{22}	0.746	0.12	6.39	< 0.001
B_{11}	0.693	5.75×10^{-2}	12.1	< 0.001
B_{21}	0.130	5.68×10^{-2}	2.3	0.022
B_{12}	0.237	5.67×10^{-2}	4.18	< 0.001
B_{22}	0.797	5.58×10^{-2}	14.3	< 0.001

However because of the problem that the BEKK model has in terms of dimensionality, a model for the entire six series under consideration doesn't converge and as such lacks parsimony. Since in this work, we want to consider a dynamic hedge, we fit CCC and DCC models and they give the following results.

For the CCC, the correlation matrix of the returns is as shown in table 4.9
This is kept constant through out. Table 4.10 shows the constants, ARCH and GARCH parameters after fitting theCCC-GARCH to the six return series of the energies. For this

Table 4.8: Parameter estimates for BEKK(11) for gasoline futures and spot.

Coefficient(s):	Estimate	Std. Error	$t-$value	$Pr(>\|t\|)$
μ_{GF}	-3.02×10^{-4}	4×10^{-4}	-0.75	0.45
μ_{GS}	-3.06×10^{-4}	4.86×10^{-4}	-0.63	0.53
$A0_{11}$	4.83×10^{-3}	2.12×10^{-4}	22.73	< 0.001
$A0_{21}$	3.8×10^{-3}	8.27×10^{-4}	4.593	< 0.001
$A0_{22}$	5.67×10^{-3}	5.95×10^{-4}	9.53	< 0.001
A_{11}	0.299	2.67×10^{-2}	11.23	< 0.001
A_{21}	-0.109	4.72×10^{-2}	-2.31	0.02
A_{22}	2.202×10^{-2}	2.120×10^{-2}	1.04	0.3
A_{22}	0.385	3.36×10^{-2}	11.5	< 0.001
B_{11}	0.946	1.46×10^{-2}	64.7	< 0.001
B_{21}	6.33×10^{-2}	2.30×10^{-2}	2.74	0.006
B_{12}	-2.35×10^{-2}	1.07×10^{-2}	-2.2	0.03
B_{22}	0.882	2.02×10^{-2}	43.7	< 0.001

Table 4.9: Correlation matrix for the energy returns

	ΔCF	ΔCS	ΔGF	ΔGS	ΔHF	ΔHS
ΔCF	1	0.94	0.72	0.53	0.77	0.73
ΔCS	0.94	1	0.72	0.54	0.76	0.72
ΔGF	0.72	0.72	1	0.65	0.75	0.72
ΔGS	0.53	0.54	0.65	1	0.57	0.61
ΔHF	0.77	0.76	0.75	0.57	1	0.89
ΔHS	0.73	0.72	0.72	0.61	0.89	1

model, a total of 18 parameters are estimated using MLE. Figure 4.8 shows a plot of the estimated conditional variances for the series and Figure 4.9 is a plot of the standardised residuals after the fit.

Table 4.10: CCC-GARCH parameter estimates.

Panel A: Constants						
Parameter	a_1	a_2	a_3	a_4	a_5	a_6
Estimate	1.24×10^{-6}	2.17×10^{-6}	1.78×10^{-5}	3.08×10^{-5}	2.47×10^{-6}	1.46×10^{-6}
Std. error	4.55×10^{-7}	5.67×10^{-3}	4.74×10^{-3}	1.35×10^{-3}	6.67×10^{-4}	2.66×10^{-3}
Panel B: ARCH parameters						
Parameter	A_{11}	A_{22}	A_{33}	A_{44}	A_{55}	A_{66}
Estimate	6.34×10^{-2}	6.89×10^{-2}	9.57×10^{-2}	8.88×10^{-2}	5.67×10^{-2}	4.81×10^{-2}
Std. error	2.34×10^{-3}	5.32×10^{-7}	2.35×10^{-3}	7.37×10^{-3}	7.24×10^{-3}	8.51×10^{-3}
Panel C: GARCH parameters						
Parameter	B_{11}	B_{22}	B_{33}	B_{44}	B_{55}	B_{66}
Estimate	0.93	0.92	0.87	0.87	0.94	0.95
Std. error	9.36×10^{-3}	5.86×10^{-3}	4.70×10^{-4}	1.90×10^{-3}	0.003	0.003

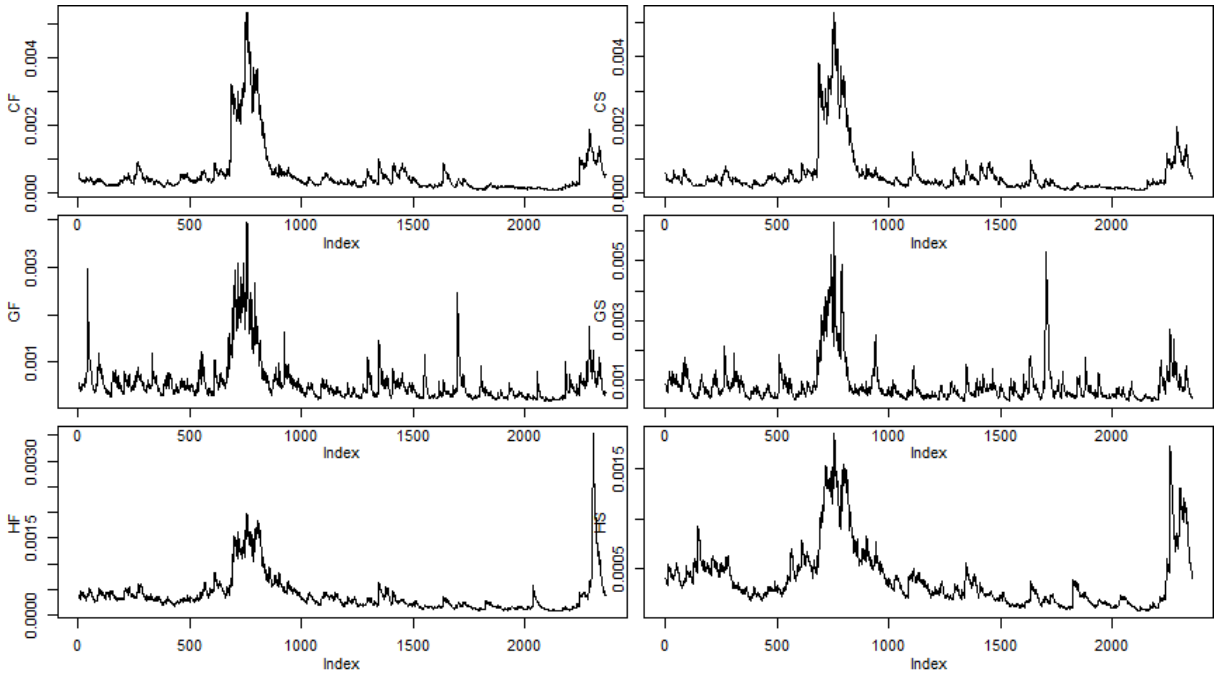

Figure 4.8: Estimated conditional variances

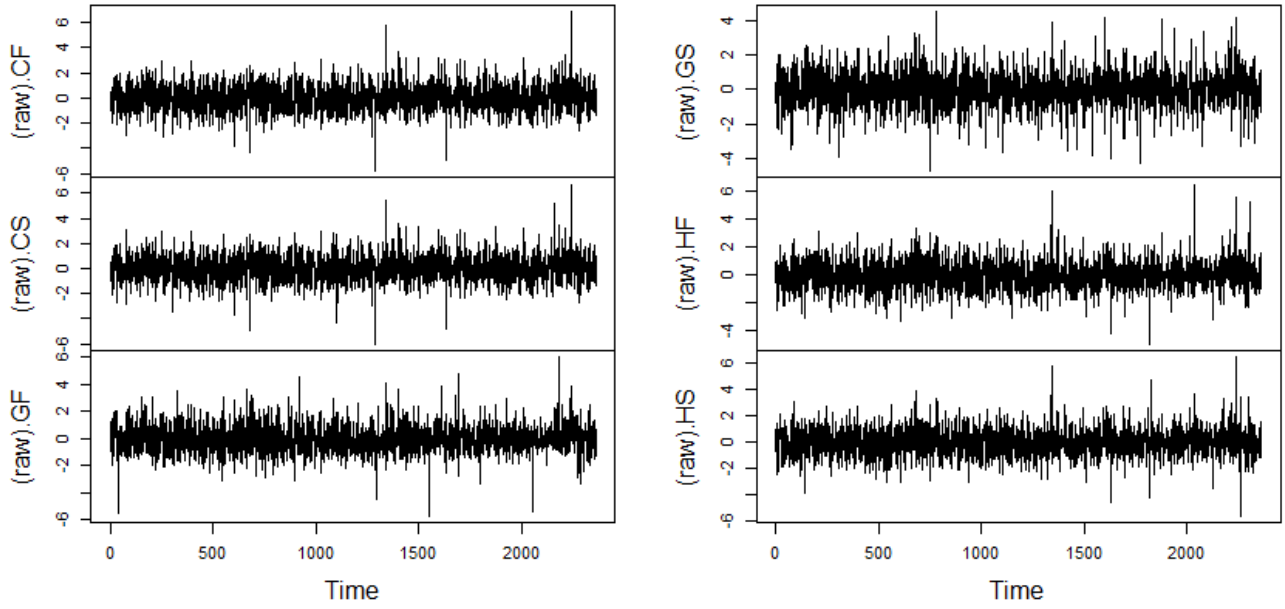

Figure 4.9: Standardized residuals

In order to take care of interdependencies in the series and any volatility spill overs, we also fit the DCC-GARCH and obtain the parameters displayed in Table 4.11. A total of 26 parameters are estimated by MLE, 4 corresponding to each return series and 2 corresponding to the DCC. Focus should be on the joint significance of the ARCH (A_{ii})

and GARCH (B_{ii}) for each of the series, and the joint significance of the dynamic ARCH (dcc_{a1}) and dynamic GARCH (dcc_{b1}) parameters. If A_{ii} and B_{ii} are jointly insignificant, a constant conditional variance rather than GARCH(1,1) should be considered. The dcc_{a1} and dcc_{b1} represent the DCC parameters for the joint models. They provide information as to whether fitting a DCC for a system of series makes sense. If dcc_{a1} and dcc_{b1} are jointly insignificant, we may be better off using a CCC model rather than the DCC(1,1). Since they are significant, the DCC model is appropriate for this data. Figure 4.10 shows plots of the estimated dynamic conditional variances and the estimated dynamic conditional correlations are represented in Figure 4.11 while the estimated conditional covariances are in Figure 4.12.

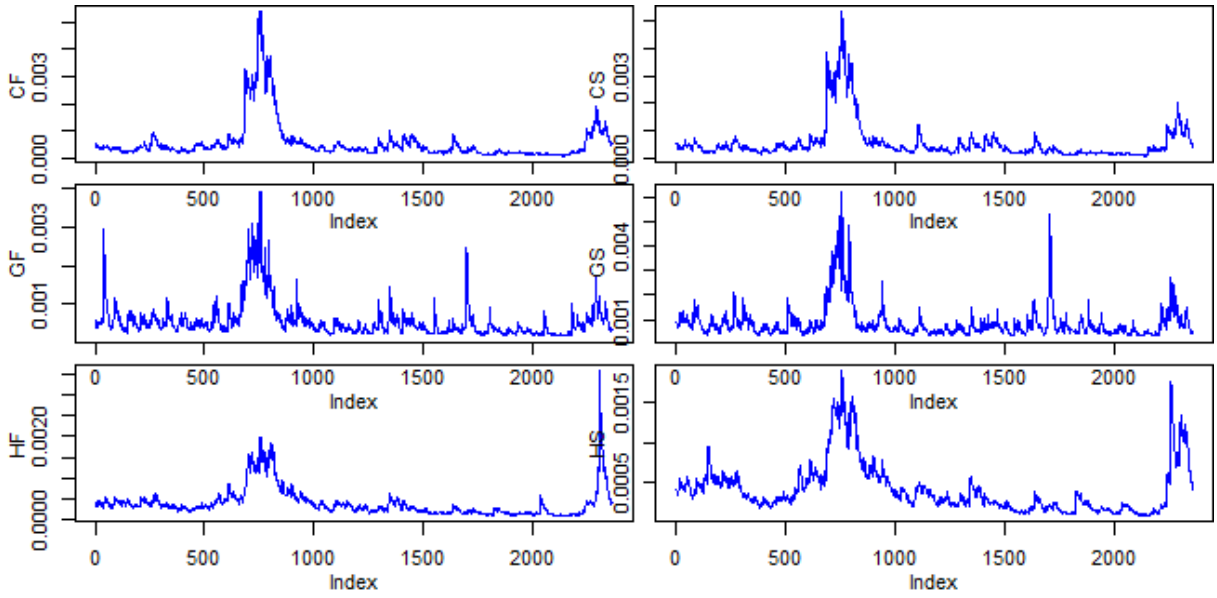

Figure 4.10: Estimated dynamic conditional variances for the six series

Table 4.11: DCC-GARCH parameter estimates.

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
μ_1	-0.000317	0.000384	-0.82547	0.409107
a_1	0.000003	0.000007	0.42512	0.670749
A_{11}	0.063728	0.045161	1.41114	0.158202
B_{11}	0.931956	0.047999	19.41625	0.000000
μ_2	-0.000424	0.000373	-1.13609	0.255918
a_2	0.000004	0.000004	1.01727	0.309027
A_{22}	0.069460	0.022602	3.07322	0.002118
B_{22}	0.923814	0.024906	37.09147	0.000000
μ_3	-0.000205	0.000424	-0.48506	0.627635
a_3	0.000017	0.000009	2.00104	0.045388
A_{33}	0.096072	0.026142	3.67495	0.000238
B_{33}	0.874159	0.017736	49.28633	0.000000
μ_4	-0.000439	0.000532	-0.82442	0.409703
a_4	0.000030	0.000010	2.99635	0.002732
A_{44}	0.089269	0.016844	5.29981	0.000000
B_{44}	0.873938	0.025470	34.31274	0.000000
μ_5	-0.000062	0.000350	-0.17616	0.860171
a_5	0.000003	0.000015	0.20020	0.841325
A_{55}	0.056911	0.094688	0.60103	0.547819
B_{55}	0.936265	0.106665	8.77763	0.000000
μ_6	-0.000261	0.000432	-0.60390	0.545912
a_6	0.000002	0.000017	0.11682	0.906999
A_{66}	0.048404	0.106032	0.45650	0.648027
B_{66}	0.947380	0.114022	8.30872	0.000000
dcc_{a1}	0.053812	0.007595	7.08494	0.000000
dcc_{b1}	0.910125	0.025082	36.28634	0.000000

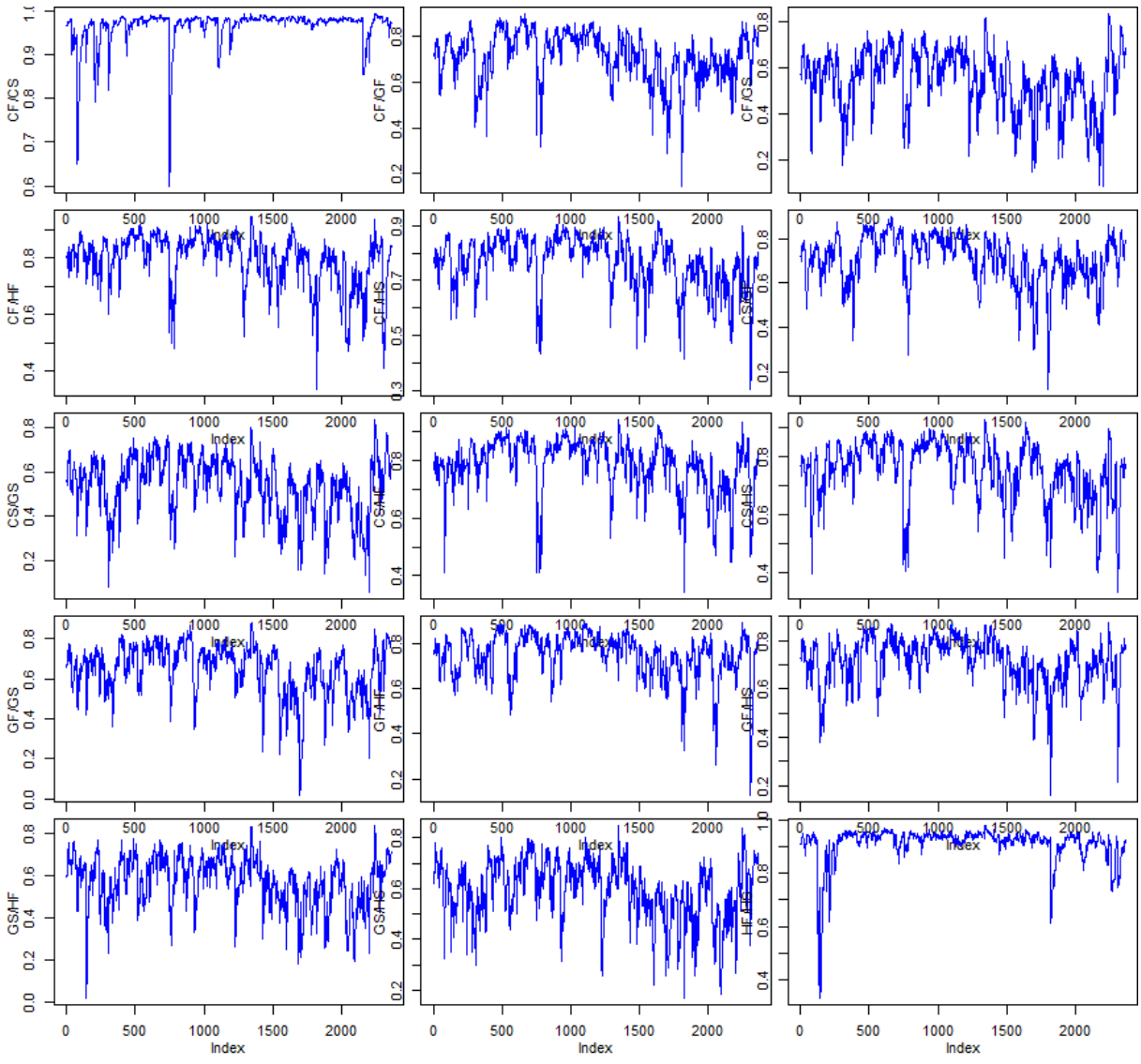

Figure 4.11: Estimated dynamic conditional correlations for the six series

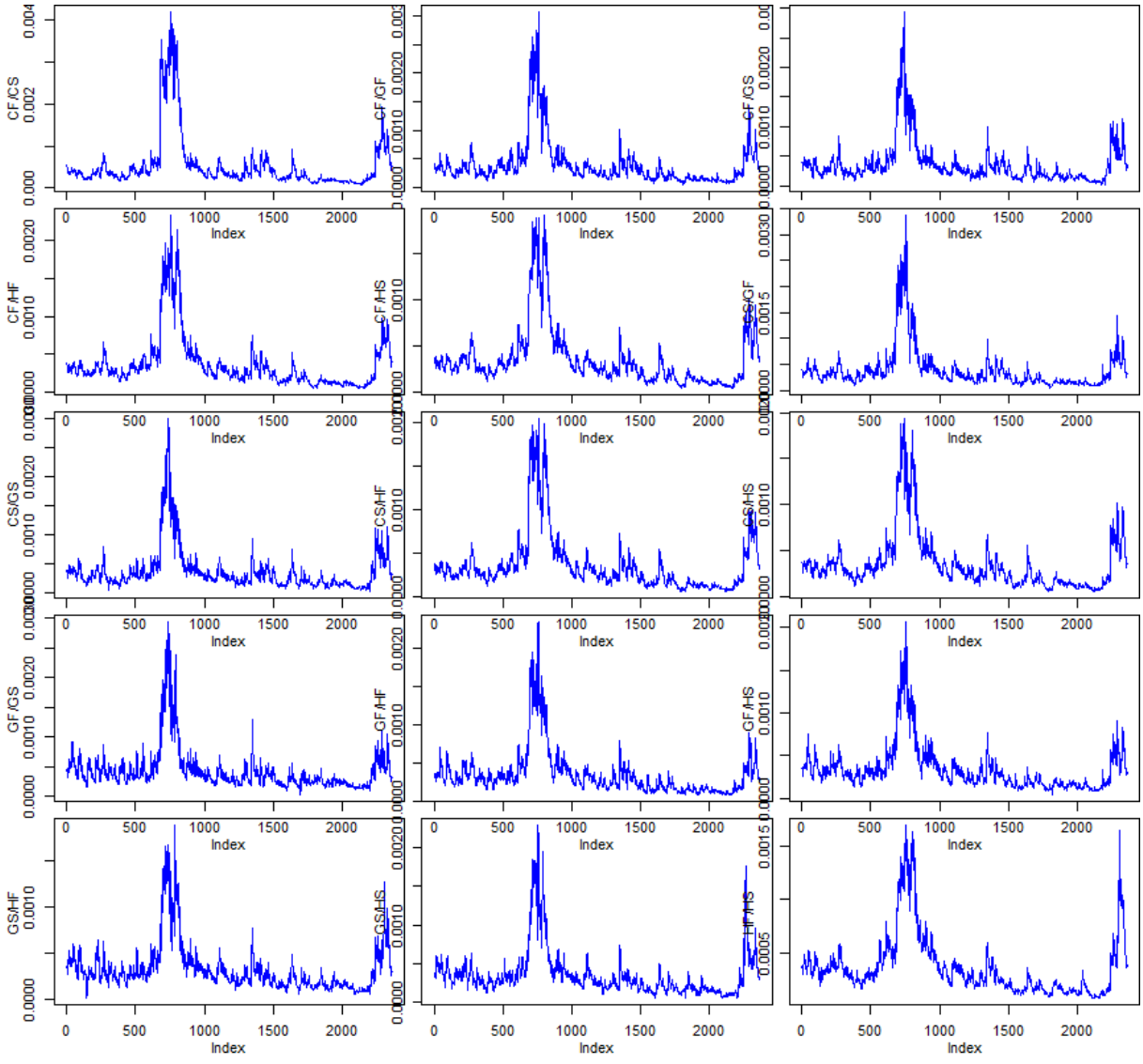

Figure 4.12: Estimated dynamic conditional covariances for the six series

5. CRACK SPREAD TRADING AND DYNAMIC HEDGING IN ENERGY MARKETS

Energy markets the world over are experiencing more competition and increased price volatility as a result of globalization coupled with the expanding regional integration through modern financial systems and international trade. This has increased the exposure of market players to potentially greater risks. As a result, the need for appropriate risk management models and techniques has emerged. The use of Crack spread hedging by refiners is one such initiative. In order to be able to build these models and employ these techniques, a clear understanding of the market dynamics through data is needed. This can be achieved by modelling and understanding the DGP, which is characterized by a non-stationary process and follows a stochastic trend.

Refiners, just as oil producers and consumers of refined products, also have the ability to hedge their exposure to price volatility. In fact, it could be argued that more than producers and consumers, refiners are faced with an even greater need to hedge since their profit margins are based on the prices of several commodities, namely the prices of their raw input (crude oil) and their refined outputs (gasoline, heating oil, fuel oil, gasoil, diesel fuel, jet fuel, etc). A refinery therefore straddles between the raw materials it buys and the finished products it sells and crack spreads represent the refiners theoretical margin. A refiner (crack spread seller) goes “short the crack" by buying crude oil futures and selling gasoline and heating oil futures.

According to Hunjra et al. (2011), volatility is the time variation in market prices due to various market forces. It does play a pivotal role in pricing portfolio selection, hedging,

and risk management. The dynamism of price change variability for various commodities has been accorded considerable attention in the field of finance (Haigh and Holt, 2002). Stock market and commodity returns exhibit clustered and asymmetric time varying volatilities, as discussed by Hunjra et al. (2011). Commodity markets are for the most part highly volatile, and this volatility itself fluctuates over time. Pindyck (2004) has showed that changes in volatility affect the prices of futures and spot contracts, as well as inventories.

The crack spread, also known as a paper refinery, is a good approximation of the refinery margins. It consists of a position that entails the simultaneous purchase and sale of futures contracts on crude oil and petroleum products. This allows the refiner to secure any price differentials between refinery input and output products. The refiner can use the crack spread to “lock in" the gross profit margin. The primary use of Crack spread contracts are speculation and hedging. The refiner who is a crack spread seller would buy crude oil contracts and sell gasoline or any other distillate fuel contracts in certain ratios. A major determinant of these ratios is how much crude oil gets processed into different refined petroleum products, since each type of crude oil yields different products with different values, and in different proportions due to their composition. These ratios therefore vary by region due to the product combination. For a brief discussion on the common crack spread ratios, see the technical reports by Cross et al. (2013) and Dash and Skyba (2011).

The naming of a crack spread is based on the number of futures contracts on crude oil, gasoline and heating oil, held in that order. The most common ratios are the 1:1:0 also known as the gasoline crack spread), 1:0:1 also called the heating oil crack spread, 2:1:1, 3:2:1, 5:3:2 and 6:3:2:1 where residual fuels are also considered (Cross et al., 2013; Dash and Skyba, 2011; Girma, Paulson, et al., 1999). The most common ratio in the US is the 3:2:1, whereas, in Europe including the Atlantic Basin covering Eastern Canadian refineries, a 6:3:2:1 ratio is the most common.

In order to calculate the theoretical refining margin, the combined value of the refined output products is first calculated and then compared with the price of the crude. In the US, crude oil prices are quoted in dollars per barrel and prices for the distillate products are quoted in cents per gallon, so for purposes of comparison, the product prices are converted to dollars per barrel by multiplying them by 42 (there are 42 gallons in a barrel). Some of the crack spread ratios are calculated as follows:

$$\begin{aligned} 1:1:0 \text{ crack spread } &= G_t - C_t \\ 1:0:1 \text{ crack spread } &= H_t - C_t \\ 3:2:1 \text{ crack spread } &= \frac{2}{3}G_t + \frac{1}{3}H_t - C_t \\ 5:3:2 \text{ crack spread } &= \frac{3}{5}G_t + \frac{2}{5}H_t - C_t \end{aligned} \tag{5.1}$$

where G_t, H_t and C_t are prices per barrel of gasoline, heating and crude oil at time t. After putting a hedge in place, the refiner does not have to worry about any movements in the absolute futures prices. The demand for crude oil derives from that of refined products such as gasoline or heating oil. If the value of a crack spread is positive, then the refined products or outputs have higher prices than the raw materials or inputs, and the spread is profitable and vice versa. Refiners trade on the spreads petroleum futures so as to hedge their anticipated cash positions or enhance profitability (Girma, Paulson, et al., 1999).

Crack spreads are also mean reverting just like most energy price series are, since the long term pricing relationship between the input prices and the output revenue will tend to converge to the long term average (Girma, Paulson, et al., 1999). This study uses co-integration theory to investigate the underlying long term equilibrium relationships among the prices of crude oil and the distillate fuels. Economic theory suggests a long term relationship between input prices and output revenues and spreads will tend to converge to a long term average.

The remainder of this chapter is organised as follows. In section 5.1, we briefly review the concept of dynamic hedging in energy markets, specifically using crack spreads on oil futures prices. In section 5.2 We discuss various optimal hedge ratios and hedge effectiveness.

5.1 Hedging in energy markets

In financial risk management, the idea to reduce uncertainty, where possible, has always seemed good. Hedging describes a risk management strategy applied in order to limit or offset the probability of loss that is as a result of fluctuations in commodity prices, currencies, or securities by transferring risk without actually buying an insurance policy. To hedge an investor would invest in two securities with negative correlations, or simultaneously take equal and opposite positions in both the spot and the derivatives market. When considering a portfolio of stocks, management of risk becomes even more challenging for two reasons: (1) the conditional variances of stock returns will co-evolve; (2) the correlations between the returns may also vary. MGARCH models have the ability accommodate these effects.

Chen et al. (2003) discuss that in hedging, we construct a portfolio that combines investments both in the spot and futures market, geared at reducing the fluctuations in it's value. The level of performance of the hedging strategy relies on the construction of an optimal portfolio. According to portfolio theory, a rational decision maker is able to sketch an efficient frontier, given a set of risky assets, so that every portfolio lying on the efficient frontier is efficient (Ding et al., 2009; Markowitz, 1952).

A futures contract is an agreement between two parties to trade a given asset at an agreed upon date, an agreed price and a a specified location in the future. Futures are designed in a way that minimizes exposure to risk so that, a change in the value of a cash position offsets a change in the value of an opposite futures position and vice versa. In addition to this, futures contracts are also preferred for hedging because they are liquid,

and the ttract lower transaction costs (Chang et al., 2011). Hedging strategies reduce price volatility substantially without reducing returns significantly. Firms usually only hedge the downside risk since they are only concerned with exposure to unwanted risk that arises from price fluctuations. Hedging involves the determination of the optimal hedge ratios which can be classified as either risk-minimizing or utility maximizing (Cecchetti et al., 1988; Chen et al., 2001, 2003). For a detailed discussion on hedging using futures, see Chen et al. (2003).

5.1.1 Hedging the 3:2:1 crack spread

One of the most widely used hedging strategies is based on portfolio variance minimization of the . However, this method has been largely criticized for ignoring the expected return on the hedged portfolio, which then makes it inconsistent with the mean-variance framework, unless investors are infinitely risk averse. The hedge ratio can be calculated using the returns of the hedged portfolio or the profit of the hedged portfolio. Using the 3:2:1 ratio, the gross cracking refiner margin is calculated as follows:

$$\frac{(2 \times G \times 42) + (1 \times H \times 42) - (3 \times C)}{3} \tag{5.2}$$

where G and H are the prices of gasoline and heating oil per gallon and C is the price of crude oil per barrel. Consider a refiner who has an obligation to purchase crude oil and sell the refined products and he wishes to hedge his commitments since he is exposed to the risk of crude oil prices increasing while at the same time refined product prices are falling, thereby resulting in a considerable reduction his refiner margins. He decides to use the 3:2:1 crack spread strategy to secure the current favourable cracking profits, so that his portfolio consists of a long position on 3 units of crude futures and short positions on 2 units gasoline futures and 1 unit of heating oil futures respectively. At maturity, the refiner buys crude oil in the spot market, cracks it and sells the gasoline and heating oil in the spot market. The refiner concurrently closes the long position on the crude and the corresponding short positions in the gasoline and heating contracts.

Suppose the spot prices in dollars per barrel for the crude, gasoline and heating contracts at time t are $C_{s,t}$, $G_{s,t}$ and $H_{s,t}$ and the futures prices are $C_{f,t}$, $G_{f,t}$ and $H_{f,t}$ respectively, and that the transaction cost at time t per barrel are given by y_t, and the refiner's profit at time t is given by π_t. Suppose further that at time $t-1$, the refiner makes all the decisions about the futures positions and at time t, all futures positions are liquidated and all the spot transactions occur. Then, since the futures market closely reflects the spot market and assuming an optimal hedge position of γ_i taken on each commodity i, it means the refiner will make a profit in dollars per barrel of:

$$\begin{aligned}\pi_t = &\frac{2}{3}G_{s,t} + \frac{1}{3}H_{s,t} - C_{s,t} + \frac{2}{3}\gamma_1\left(G_{f,t-1} - G_{f,t}\right) + \frac{1}{3}\gamma_2\left(H_{f,t-1} - H_{f,t}\right) \\ &+ \gamma_3\left(C_{f,t} - C_{f,t-1}\right) - y_t\end{aligned} \tag{5.3}$$

with a variance of

$$\begin{aligned}\sigma^2_{\pi_t} = &\frac{4}{9}\sigma^2_{G_{s,t}} + \frac{1}{9}\sigma^2_{H_{s,t}} + \sigma^2_{C_{s,t}} + \frac{4}{9}\gamma_1^2\sigma^2_{G_{f,t}} + \frac{1}{9}\gamma_2^2\sigma^2_{H_{f,t}} + \gamma_3^2\sigma^2_{C_{f,t}} + \frac{4}{9}\sigma_{G_{s,t}H_{s,t}} \\ &- \frac{4}{3}\sigma_{G_{s,t}C_{s,t}} - \frac{8}{9}\gamma_1\sigma_{G_{s,t}G_{f,t}} - \frac{4}{9}\gamma_2\sigma_{G_{s,t}H_{f,t}} + \frac{4}{3}\gamma_3\sigma_{G_{s,t}C_{f,t}} - \frac{2}{3}\sigma_{H_{s,t}C_{s,t}} \\ &- \frac{4}{9}\gamma_1\sigma_{H_{s,t}G_{f,t}} - \frac{2}{9}\gamma_2\sigma_{H_{s,t}H_{f,t}} + \frac{2}{3}\gamma_3\sigma_{H_{s,t}C_{f,t}} + \frac{4}{3}\gamma_1\sigma_{C_{s,t}G_{f,t}} + \frac{2}{3}\gamma_2\sigma_{C_{s,t}H_{f,t}} \\ &- 2\gamma_3\sigma_{C_{s,t}C_{f,t}} + \frac{4}{9}\gamma_1\gamma_2\sigma_{G_{f,t}H_{f,t}} - \frac{4}{3}\gamma_1\gamma_3\sigma_{G_{f,t}C_{f,t}} - \frac{2}{3}\gamma_2\gamma_3\sigma_{H_{f,t}C_{f,t}}\end{aligned} \tag{5.4}$$

obtained using the formula

$$\sigma^2_\pi = \sum_{i=1}^{n}\omega_i\sigma_i^2 + \sum_{i=1}^{n}\sum_{j=i+1}^{n} 2\omega_i\omega_j\rho_{ij}\sigma_i\sigma_j \tag{5.5}$$

where ω_i represents the weight of investment in security i and $\rho_{ij}\sigma_i\sigma_j$ is the covariance between securities i and j where we have n securities in total.

At time t, the refiner liquidates the crude futures contracts at $C_{f,t}$ and purchases the crude spot position at $C_{s,t}$. For the gasoline and heating contracts, the refiner goes short at time $t-1$ and offsets (buys back) these contracts at time t. The scenarios depicted in table 5.1 could play out depending on which direction the market moves.

Table 5.1: Possible Outcomes at time t

Position	Scenario	Result
Long on crude	$C_{f,t-1} > C_{f,t}$	loss on futures
	$C_{f,t-1} \leq C_{f,t}$	break even or profit on futures
Short on gasoline	$G_{f,t-1} < G_{f,t}$	gain on futures
	$G_{f,t-1} \geq G_{f,t}$	break even or loss on futures
Short on heating oil	$H_{f,t-1} < H_{f,t}$	gain on futures
	$H_{f,t-1} \geq H_{f,t}$	break even or loss on futures

5.2 Optimal hedge ratios

Hedging involves determining the optimal hedge ratio which depends on a particular objective function to be optimized (Chang et al., 2011; Chen et al., 2003). From existing literature, it is suggested that various hedge ratios can be classified as either being utility-maximizing or risk-minimizing (Chen et al., 2001). The hedge ratios that are risk-minimizing, proposed by Ederington (1979), involve a specific risk measure such as the variance, the semi-variance or the mean-Gini coefficient of return being minimized. The utility-maximizing hedge ratios proposed by Cecchetti et al. (1988) incorporate both the expected return and the variance of the hedged portfolio.

However, for these strategies to be consistent with the principle of expected utility maximization, either the utility function needs to be quadratic or the returns should be jointly normally distributed, otherwise, either specific utility function or a specific returns distribution must be used (Chen et al., 2003). For the estimation of hedge ratios, researchers have used various methods from the simple OLS technique to more complex methods such as conditional heteroscedastic or cointegration heteroscedastic methods. Hedge ratios can also be classified as being static or dynamic (where the hedge ratio is adjusted based on new information received).

Once these ratios are established, it's important to check if they are effective. According to De Jong et al. (1997), hedge effectiveness is defined as the percentage reduction in the portfolio variance. It is the percentage reduction in risk criterion with hedging versus without hedging (Ederington, 1979). Specifically, the hedging effectiveness for minimum

variance is determined as

$$HE = 1 - \frac{\text{Variance of hedged portfolio}}{\text{variance of unhedged portfolio}} \tag{5.6}$$

5.2.1 The full or naive hedge

This is the simplest and oldest method used to determining the number of futures contracts to be used in a specific hedging strategy. It simply measures the position taken in the underlying asset and takes an equal but opposite position in futures contracts (De Jong et al., 1997).

Consider a portfolio consisting of a long spot position on C_s units and a short futures position on C_f units, and let S_t and F_t denote spot and futures prices at time t respectively, then the return on the resulting hedged portfolio, denoted by r_h can be given by

$$r_h = \frac{C_s S_t r_s - C_f F_t rf}{C_s S_t} = r_s - \gamma r_f \tag{5.7}$$

where $r_s = \dfrac{S_t - S_{t-1}}{S_{t-1}}$ and $r_f = \dfrac{F_t - F_{t-1}}{F_{t-1}}$ are one period simple returns on the spot and futures positions respectively and $\gamma = \dfrac{C_f F_t}{C_s S_t}$ is the hedge ratio. In the situation where $\gamma = 1$, meaning $C_f F_t = C_s S_t$, then, we have a naive or a full hedge. If the value of the hedged portfolio is an exact match with the value of the underlying contract, then, a hedger who wants to close out his position at maturity of the contract would simply buy contracts covering the entire position.

The hedge ratio can also be considered in terms of profits in which case, the profit on the hedged portfolio π_h at time t is given by

$$\pi_{h,t} = C_s \Delta S_t - \gamma C_f \Delta F_t \tag{5.8}$$

where $\Delta S_t = S_t - S_{t-1}$, $\Delta F_t = F_t - F_{t-1}$ are profits on one period spot and futures contracts respectively and $\gamma = \dfrac{C_f}{C_s}$.

If the investors hedged position has to be closed out at time t, before to maturity, then the hedger would be exposed to the basis risk b_t, which describes the difference between the spot and futures price at time t and is defined as

$$b_t = S_t - F_t \tag{5.9}$$

5.2.2 Minimum-Variance Criterion

A hedge has one main objective, to minimise the risk of a given position. Hedging effectiveness is the reduction of the variance of the value of a position hedged with futures (Ederington, 1979). From equation (5.7), if the investor hedges a fraction γ of his cash position in the futures market and leaves a fraction $(1-\gamma)$ unhedged, then the expected return of the hedged position will be given by

$$E[r_h] = E[r_s] - \gamma E[r_f] \tag{5.10}$$

and the variance of the return on the hedged portfolio would be given by

$$\begin{aligned}
\sigma^2_{r_h} &= E\left[r_h^2\right] - E[r_h]^2 \\
&= E\left[r_s^2 - 2\gamma r_s r_f + \gamma^2 r_f^2\right] - E[r_s]^2 - \gamma E[r_f]^2 \\
&= E\left[r_s^2\right] - E[r_s]^2 + \gamma^2 E\left[r_f^2\right] - \gamma^2 E[r_f]^2 - 2\gamma E[r_s r_f] \\
&= \sigma^2_{r_s} + \gamma^2 \sigma^2_{r_f} - 2\gamma \sigma_{r_s r_f}
\end{aligned} \tag{5.11}$$

We can see therefore that the expected return as well as the variance will vary with the hedge position. The return on the hedged position will always be exposed the basis risk, and hence, no hedging strategy can completely eliminate risk. We can however get the risk minimizing hedge position by differentiating equation (5.11) with respect to γ and

equating the resulting function to zero and solving for γ as shown in equation (5.12).

$$\begin{aligned} \sigma^2_{r_h} &= \sigma^2_{r_s} + \gamma^2\sigma^2_{r_f} - 2\gamma\sigma_{r_s r_f} \\ \frac{\partial\sigma^2_{r_h}}{\partial\gamma} &= 2\gamma\sigma^2_{r_f} - 2\sigma_{r_s r_f} = 0 \\ \gamma &= \frac{\sigma_{r_s r_f}}{\sigma^2_{r_f}} = \rho\frac{\sigma_{r_s}}{\sigma_{r_f}} \end{aligned} \tag{5.12}$$

where $-1 < \rho < 1$ measures the correlation between spot and futures market returns. From this we can see that an estimate for γ can be derived from historical data by simply regressing r_s on r_f. If however the hedge position is time varying, so that there is time variation in the joint distribution of r_s and r_f, then, regression cannot accurately estimate the hedge ratio.

For the dynamic case where the hedge ratio is changed based on the current or conditional information, the hedge ratio is represented by

$$\gamma_t|\mathcal{F}_{t-1} = \frac{\sigma_{r_{s,t} r_{f,t}}|\mathcal{F}_{t-1}}{\sigma^2_{r_{f,t}}|\mathcal{F}_{t-1}} \tag{5.13}$$

where $\mathcal{F}_{t-1}$ represents the set of information available at time $t-1$

For the 3:2:1 crack hedge described by equation (5.2), the first order conditions for maximization are then given as equations (5.14) to (5.16)

$$\begin{aligned} \frac{\partial\sigma^2_{\pi_t}}{\partial\gamma_1} = &\frac{8}{9}\gamma_1\sigma^2_{G_{f,t}} - \frac{8}{9}\sigma_{G_{s,t}G_{f,t}} - \frac{4}{9}\sigma_{H_{s,t}G_{f,t}} + \frac{4}{3}\sigma_{C_{s,t}G_{f,t}} + \frac{4}{9}\gamma_2\sigma_{G_{f,t}H_{f,t}} \\ &- \frac{4}{3}\gamma_3\sigma_{G_{f,t}C_{f,t}} \end{aligned} \tag{5.14}$$

$$\begin{aligned} \frac{\partial\sigma^2_{\pi_t}}{\partial\gamma_2} = &\frac{2}{9}\gamma_2\sigma^2_{H_{f,t}} - \frac{4}{9}\sigma_{G_{s,t}H_{f,t}} - \frac{2}{9}\sigma_{H_{s,t}H_{f,t}} + \frac{2}{3}\sigma_{C_{s,t}H_{f,t}} + \frac{4}{9}\gamma_1\sigma_{G_{f,t}H_{f,t}} \\ &- \frac{2}{3}\gamma_3\sigma_{H_{f,t}C_{f,t}} \end{aligned} \tag{5.15}$$

$$\begin{aligned}\frac{\partial \sigma^2_{\pi_t}}{\partial \gamma_3} &= -2\gamma_3\sigma^2_{C_{f,t}} + \frac{4}{3}\sigma_{G_{s,t}C_{f,t}} + \frac{2}{3}\sigma_{H_{s,t}C_{f,t}} - 2\sigma_{C_{s,t}C_{f,t}} - \frac{4}{3}\gamma_1\sigma_{G_{f,t}C_{f,t}} \\ &\quad - \frac{2}{3}\gamma_2\sigma_{H_{f,t}C_{f,t}}\end{aligned} \tag{5.16}$$

Equating these to zero and rearranging these equations gives

$$\tfrac{8}{9}\gamma_1\sigma^2_{G_{f,t}} + \tfrac{4}{9}\gamma_2\sigma_{G_{f,t}H_{f,t}} - \tfrac{4}{3}\gamma_3\sigma_{G_{f,t}C_{f,t}} = \tfrac{8}{9}\sigma_{G_{s,t}G_{f,t}} + \tfrac{4}{9}\sigma_{H_{s,t}G_{f,t}} - \tfrac{4}{3}\sigma_{C_{s,t}G_{f,t}}$$

$$\tfrac{4}{9}\gamma_1\sigma_{G_{f,t}H_{f,t}} + \tfrac{2}{9}\gamma_2\sigma^2_{H_{f,t}} - \tfrac{2}{3}\gamma_3\sigma_{H_{f,t}C_{f,t}} = \tfrac{4}{9}\sigma_{G_{s,t}H_{f,t}} + \tfrac{2}{9}\sigma_{H_{s,t}H_{f,t}} - \tfrac{2}{3}\sigma_{C_{s,t}H_{f,t}}$$

$$\tfrac{4}{3}\gamma_1\sigma_{G_{f,t}C_{f,t}} + \tfrac{2}{3}\gamma_2\sigma_{H_{f,t}C_{f,t}} + 2\gamma_3\sigma^2_{C_{f,t}} = \tfrac{4}{3}\sigma_{G_{s,t}C_{f,t}} + \tfrac{2}{3}\sigma_{H_{s,t}C_{f,t}} - 2\sigma_{C_{s,t}C_{f,t}}$$

which can be expressed in matrix form as shown in equation (5.17)

$$\begin{aligned}&\begin{bmatrix} \frac{8}{9}\sigma^2_{G_{f,t}} & \frac{4}{9}\sigma_{G_{f,t}H_{f,t}} & -\frac{4}{3}\sigma_{G_{f,t}C_{f,t}} \\ \frac{4}{9}\sigma_{G_{f,t}H_{f,t}} & \frac{2}{9}\sigma^2_{H_{f,t}} & -\frac{2}{3}\sigma_{H_{f,t}C_{f,t}} \\ \frac{4}{3}\sigma_{G_{f,t}C_{f,t}} & \frac{2}{3}\sigma_{H_{f,t}C_{f,t}} & 2\sigma^2_{C_{f,t}} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} \\ &= \begin{bmatrix} \frac{8}{9}\sigma_{G_{s,t}G_{f,t}} + \frac{4}{9}\sigma_{H_{s,t}G_{f,t}} - \frac{4}{3}\sigma_{C_{s,t}G_{f,t}} \\ \frac{4}{9}\sigma_{G_{s,t}H_{f,t}} + \frac{2}{9}\sigma_{H_{s,t}H_{f,t}} - \frac{2}{3}\sigma_{C_{s,t}H_{f,t}} \\ \frac{4}{3}\sigma_{G_{s,t}C_{f,t}} + \frac{2}{3}\sigma_{H_{s,t}C_{f,t}} - 2\sigma_{C_{s,t}C_{f,t}} \end{bmatrix}\end{aligned} \tag{5.17}$$

Solving for the $\gamma_i's$ yields the optimal hedge positions that minimise the variance. Usually these $\gamma_i's$ are less that the full hedge. For $\gamma_1 = \gamma_2 = \gamma_3 = 0$ there is no hedge in place. For $\gamma_1 = \gamma_2 = \gamma_3 = 1$, we have a full hedge or a naive hedge.

Equation (5.17) can compactly be expressed in the inform

$$\boldsymbol{A\gamma} = \boldsymbol{B} \tag{5.18}$$

so that

$$\boldsymbol{\gamma} = \boldsymbol{A}^{-1}\boldsymbol{B} \tag{5.19}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \frac{8}{9}\sigma^2_{G_{f,t}} & \frac{4}{9}\sigma_{G_{f,t}H_{f,t}} & -\frac{4}{3}\sigma_{G_{f,t}C_{f,t}} \\ \frac{4}{9}\sigma_{G_{f,t}H_{f,t}} & \frac{2}{9}\sigma^2_{H_{f,t}} & -\frac{2}{3}\sigma_{H_{f,t}C_{f,t}} \\ \frac{4}{3}\sigma_{G_{f,t}C_{f,t}} & \frac{2}{3}\sigma_{H_{f,t}C_{f,t}} & 2\sigma^2_{C_{f,t}} \end{bmatrix},$$

$$\boldsymbol{B} = \begin{bmatrix} \frac{8}{9}\sigma_{G_{s,t}G_{f,t}} + \frac{4}{9}\sigma_{H_{s,t}G_{f,t}} - \frac{4}{3}\sigma_{C_{s,t}G_{f,t}} \\ \frac{4}{9}\sigma_{G_{s,t}H_{f,t}} + \frac{2}{9}\sigma_{H_{s,t}H_{f,t}} - \frac{2}{3}\sigma_{C_{s,t}H_{f,t}} \\ \frac{4}{3}\sigma_{G_{s,t}C_{f,t}} + \frac{2}{3}\sigma_{H_{s,t}C_{f,t}} - 2\sigma_{C_{s,t}C_{f,t}} \end{bmatrix}$$

and

$$\boldsymbol{\gamma}' = [\gamma_1 \quad \gamma_2 \quad \gamma_3]$$

5.2.3 Certainty Equivalence and the Mean-Variance Criterion

If a hedger minimises risk without any regard to the effects of expected returns, the hedge cannot be optimal unless the investor is totally risk averse. This is due to the fact that high risk is usually associated with high returns and therefore at equilibrium, the risky assets are priced such that they earn an expected return which is over and above the risk-free rate of return. This implies that hedging away the risk also must in the same way hedge away the expected return of bearing that risk (Cecchetti et al., 1988). The mean-variance criterion for the optimal hedge incorporates both risk and return in deriving the hedge ratio (Chen et al., 2003). It is related the expected utility hypothesis under the assumptions that the investor is risk averse (Ding et al., 2009) and that investment returns are normally distributed (Featherstone and Moss, 1990). The utility functions that represent risk aversion are assumed to be differentiable, increasing and concave, such as exponential and quadratic functions. Under this approach, the optimal hedge is the one that maximises the expected utility. We can use these when deriving the certainty equivalent of an investment opportunity, so that different hedge positions can be compa-

red differently by examining their certainty equivalent returns.

The certainty equivalent describes a certain wealth level for which a decision maker is indifferent as concerns the risky alternatives. It is the minimum sum of money an investor would accept, to forego the opportunity to participate in an event for which the outcome is uncertain. To compute the certainty equivalent for a risky opportunity, an expenditure function or an inverse utility function is equated to the expected utility.

Under the mean-variance criterion, we use the optimum mean variance hedge ratio to establish the number of contracts that should be trade in each market. The investor maximises his expected profits adjusted for risk. In this situation, the investor seeks to maximize select γ_i in order to maximize the utility at the end of the period. The objective function to be optimised is thus given by:

$$\underset{\gamma_{i,(i=1,2,3)}}{\text{Max}} \left[E\left(\pi_t | \mathcal{F}_{t-1}\right) - \frac{1}{2}\lambda \text{ Var}\left(\pi_t | \mathcal{F}_{t-1}\right) - y_t \right] \tag{5.20}$$

where π_t is defined in equation (5.3), λ is the risk aversion coefficient of the refiner or trader in question and $\mathcal{F}_{t-1}$ denotes the set of information that is available at time $t-1$, which is important if the hedge is to be dynamic. $\lambda < 1$ denotes risk seeking attributes, $\lambda = 1$ denotes risk neutral attributes and $\lambda > 1$ denotes risk aversion.

Suppose we have a negative exponential utility function $u\left[w(x)\right] = -\exp\left[-\lambda w(x)\right]$ where the wealth w is a function of an investment bundle x and λ is the Arrow-Pratt measure of risk aversion. Suppose the returns on x follow a multivariate normal distribution, i.e. $w(x) \sim N\left(\mu, \sigma^2\right)$ and the density function corresponding $w(x)$ is $f(w(x)) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{\left(w(x)-\mu\right)^2}{2\sigma^2}\right)$, then, the expected utility function would be given

by

$$\begin{aligned} E\left[u(w)\right] &= -E\left[\exp(-\lambda w)\right] \\ &= -\int_{-\infty}^{\infty} \exp(-\lambda w) f(w) dw \\ &= -\int_{-\infty}^{\infty} \exp(-\lambda w) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(w-\mu)^2}{2\sigma^2}\right) dw \\ &= -\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\lambda w - \frac{(w-\mu)^2}{2\sigma^2}\right) dw \end{aligned} \tag{5.21}$$

Notice from the last part of equation (5.21) that

$$\begin{aligned} -\lambda w - \frac{(w-\mu)^2}{2\sigma^2} &= -\lambda w - \frac{(w-\mu)^2}{2\sigma^2} + \lambda\mu - \lambda\mu + \frac{\lambda^2\sigma^2}{2} - \lambda\mu + \frac{\lambda^2\sigma^2}{2} \\ &= -\left(\lambda\mu + \frac{(w-\mu)^2}{2\sigma^2} - \lambda\mu + \frac{\lambda^2\sigma^2}{2}\right) - \lambda\mu + \frac{\lambda^2\sigma^2}{2} \\ &= -\frac{1}{2}\left(\frac{(w-\mu)^2}{2\sigma^2} + 2\lambda(w-\mu) + \lambda^2\sigma^2\right) - \lambda\mu + \frac{\lambda^2\sigma^2}{2} \\ &= -\frac{1}{2\sigma^2}\left((w-\mu) + \lambda\sigma^2\right)^2 - \lambda\mu + \frac{\lambda^2\sigma^2}{2} \end{aligned}$$

Substituting this in equation (5.21), we have

$$\begin{aligned} E\left[u(w)\right] &= -\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\lambda w - \frac{(w-\mu)^2}{2\sigma^2}\right) dw \\ &= -\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left((w-\mu) + \lambda\sigma^2\right)^2 - \lambda\mu + \frac{\lambda^2\sigma^2}{2}\right) dw \\ &= -\exp\left(-\lambda\mu + \frac{\lambda^2\sigma^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left((w-\mu) + \lambda\sigma^2\right)^2\right) \end{aligned} \tag{5.22}$$

Notice from equation (5.22), we can define a density function of a normally distributed random variable with a mean of $(\mu - \lambda\sigma^2)$ and variance σ^2 as $g(w) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left((w-\mu) + \lambda\sigma^2\right)^2\right.$

This means

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left((w-\mu) + \lambda\sigma^2\right)^2\right) dw = 1$$

and therefore we have

$$\begin{aligned} E\left[u(w)\right] &= -\exp\left(-\lambda\mu + \frac{\lambda^2\sigma^2}{2}\right) \\ &= -\exp\left(-\lambda\left(\mu - \frac{\lambda\sigma^2}{2}\right)\right) \\ &= -\exp\left(-\lambda CE\right) \end{aligned} \tag{5.23}$$

where $CE = \mu - \frac{\lambda\sigma^2}{2}$ is the mean-variance certainty equivalent function based on the exponential utility function. From equation (5.23), we notice that CE is a monotonic transformation of $E\left[u(w)\right]$, and hence CE represents the same preference as $E\left[u(w)\right]$. Maximizing the CE translate to maximizing expected utility of wealth. Comparing this with equation (5.20), we see the similarity except we are considering the profit at time t, π_t instead of the return μ_t. We can further note that in the mean-variance coordinate plane, the certainty equivalent function is a straight line and as such, this gives an effective selection procedure when constructing the preference orderings for the combinations of portfolios on the efficient frontier.

From the objective function given in equation (5.20), we want to maximize

$$\underset{\gamma_{i,(i=1,2,3)}}{\text{Max}}\ CEI_t = \underset{\gamma_{i,(i=1,2,3)}}{\text{Max}}\left[E\left(\pi_t|\mathcal{F}_{t-1}\right) - \frac{1}{2}\lambda\left(\sigma^2_{\pi_t}|\mathcal{F}_{t-1}\right) - y_t\right] \tag{5.24}$$

where π_t is given by equation (5.3) and $\sigma^2_{\pi_t}$ is given by equation (5.4). This objective function in equation (5.24) is similar to a quadratic utility function to be maximized. The first order conditions for maximization are then given as equations (5.25) to (5.27)

$$\begin{aligned} \frac{\partial CEI_t}{\partial\gamma_1} = \frac{2}{3}E\left[G_{f,t-1} - G_{f,t}\right] - \lambda\Big\{&\frac{4}{9}\gamma_1\sigma^2_{G_{f,t}} - \frac{4}{9}\sigma_{G_{s,t}G_{f,t}} - \frac{2}{9}\sigma_{H_{s,t}G_{f,t}} \\ &+\frac{2}{3}\sigma_{C_{s,t}G_{f,t}} + \frac{2}{9}\gamma_2\sigma_{G_{f,t}H_{f,t}} - \frac{2}{3}\gamma_3\sigma_{G_{f,t}C_{f,t}}\Big\} \end{aligned} \tag{5.25}$$

$$\begin{aligned}\frac{\partial CEI_t}{\partial \gamma_2} = \frac{1}{3}E\left[H_{f,t-1} - H_{f,t}\right] - \lambda\Big\{&\frac{1}{9}\gamma_2\sigma^2_{H_{f,t}} - \frac{2}{9}\sigma_{G_{s,t}H_{f,t}} - \frac{1}{9}\sigma_{H_{s,t}H_{f,t}} \\ &+\frac{1}{3}\sigma_{C_{s,t}H_{f,t}} + \frac{2}{9}\gamma_1\sigma_{G_{f,t}H_{f,t}} - \frac{1}{3}\gamma_3\sigma_{H_{f,t}C_{f,t}}\Big\}\end{aligned} \tag{5.26}$$

$$\begin{aligned}\frac{\partial CEI_t}{\partial \gamma_3} = E\left[C_{f,t} - C_{f,t-1}\right] - \lambda\Big\{&\gamma_3\sigma^2_{C_{f,t}} + \frac{2}{3}\sigma_{G_{s,t}C_{f,t}} + \frac{1}{3}\sigma_{H_{s,t}C_{f,t}} \\ &-\sigma_{C_{s,t}C_{f,t}} - \frac{2}{3}\gamma_1\sigma_{G_{f,t}C_{f,t}} - \frac{1}{3}\gamma_2\sigma_{H_{f,t}C_{f,t}}\Big\}\end{aligned} \tag{5.27}$$

These, are then equated to zero, divided by λ and equation (5.25) and (5.26) multiplied by $\frac{3}{2}$ and 3 respectively, giving

$$\begin{aligned}\frac{E\left[G_{f,t-1} - G_{f,t}\right]}{\lambda} &- \frac{2}{3}\gamma_1\sigma^2_{G_{f,t}} + \frac{2}{3}\sigma_{G_{s,t}G_{f,t}} + \frac{1}{3}\sigma_{H_{s,t}G_{f,t}} - \sigma_{C_{s,t}G_{f,t}} \\ &- \frac{1}{3}\gamma_2\sigma_{G_{f,t}H_{f,t}} + \gamma_3\sigma_{G_{f,t}C_{f,t}} = 0\end{aligned} \tag{5.28}$$

$$\begin{aligned}\frac{E\left[H_{f,t-1} - H_{f,t}\right]}{\lambda} &- \frac{1}{3}\gamma_2\sigma^2_{H_{f,t}} + \frac{2}{3}\sigma_{G_{s,t}H_{f,t}} + \frac{1}{3}\sigma_{H_{s,t}H_{f,t}} - \sigma_{C_{s,t}H_{f,t}} \\ &- \frac{2}{3}\gamma_1\sigma_{G_{f,t}H_{f,t}} + \gamma_3\sigma_{H_{f,t}C_{f,t}} = 0\end{aligned} \tag{5.29}$$

$$\begin{aligned}\frac{E\left[C_{f,t} - C_{f,t-1}\right]}{\lambda} &- \gamma_3\sigma^2_{C_{f,t}} - \frac{2}{3}\sigma_{G_{s,t}C_{f,t}} - \frac{1}{3}\sigma_{H_{s,t}C_{f,t}} + \sigma_{C_{s,t}C_{f,t}} \\ &+ \frac{2}{3}\gamma_1\sigma_{G_{f,t}C_{f,t}} + \frac{1}{3}\gamma_2\sigma_{H_{f,t}C_{f,t}} = 0\end{aligned} \tag{5.30}$$

and since we are interested in the $\gamma_i's$, we rearrange equations (5.28) to (5.30) as follows

$$\begin{aligned}&\frac{2}{3}\gamma_1\sigma^2_{G_{f,t}} + \frac{1}{3}\gamma_2\sigma_{G_{f,t}H_{f,t}} - \gamma_3\sigma_{G_{f,t}C_{f,t}} \\ &= \frac{E\left[G_{f,t-1} - G_{f,t}\right]}{\lambda} + \frac{2}{3}\sigma_{G_{s,t}G_{f,t}} + \frac{1}{3}\sigma_{H_{s,t}G_{f,t}} - \sigma_{C_{s,t}G_{f,t}}\end{aligned} \tag{5.31}$$

$$\begin{aligned}&\frac{2}{3}\gamma_1\sigma_{G_{f,t}H_{f,t}} + \frac{1}{3}\gamma_2\sigma^2_{H_{f,t}} - \gamma_3\sigma_{H_{f,t}C_{f,t}} \\ &= \frac{E\left[H_{f,t-1} - H_{f,t}\right]}{\lambda} + \frac{2}{3}\sigma_{G_{s,t}H_{f,t}} + \frac{1}{3}\sigma_{H_{s,t}H_{f,t}} - \sigma_{C_{s,t}H_{f,t}}\end{aligned} \tag{5.32}$$

$$\begin{aligned}&\frac{2}{3}\gamma_1\sigma_{G_{f,t}C_{f,t}}+\frac{1}{3}\gamma_2\sigma_{H_{f,t}C_{f,t}}-\gamma_3\sigma^2_{C_{f,t}}\\&=-\frac{E\left[C_{f,t}-C_{f,t-1}\right]}{\lambda}+\frac{2}{3}\sigma_{G_{s,t}C_{f,t}}+\frac{1}{3}\sigma_{H_{s,t}C_{f,t}}-\sigma_{C_{s,t}C_{f,t}}\end{aligned}\tag{5.33}$$

Equation (5.31) to (5.33) can be expressed in matrix form as shown in equation (5.34)

$$\begin{aligned}&\begin{bmatrix}\frac{2}{3}\sigma^2_{G_{f,t}} & \frac{1}{3}\sigma_{G_{f,t}H_{f,t}} & -\sigma_{G_{f,t}C_{f,t}}\\ \frac{2}{3}\sigma_{G_{f,t}H_{f,t}} & \frac{1}{3}\sigma^2_{H_{f,t}} & -\sigma_{H_{f,t}C_{f,t}}\\ \frac{2}{3}\sigma_{G_{f,t}C_{f,t}} & \frac{1}{3}\sigma_{H_{f,t}C_{f,t}} & -\sigma^2_{C_{f,t}}\end{bmatrix}\begin{bmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\end{bmatrix}\\&=\begin{bmatrix}\frac{E\left[G_{f,t-1}-G_{f,t}\right]}{\lambda}+\frac{2}{3}\sigma_{G_{s,t}G_{f,t}}+\frac{1}{3}\sigma_{H_{s,t}G_{f,t}}-\sigma_{C_{s,t}G_{f,t}}\\ \frac{E\left[H_{f,t-1}-H_{f,t}\right]}{\lambda}+\frac{2}{3}\sigma_{G_{s,t}H_{f,t}}+\frac{1}{3}\sigma_{H_{s,t}H_{f,t}}-\sigma_{C_{s,t}H_{f,t}}\\ -\frac{E\left[C_{f,t}-C_{f,t-1}\right]}{\lambda}+\frac{2}{3}\sigma_{G_{s,t}C_{f,t}}+\frac{1}{3}\sigma_{H_{s,t}C_{f,t}}-\sigma_{C_{s,t}C_{f,t}}\end{bmatrix}\end{aligned}\tag{5.34}$$

Solving for the $\gamma_i's$ yields the optimal hedge positions for the optimal hedge.

5.2.4 Dynamic hedge ratios

We consider monthly spot and futures contract 1 prices for Cushing OK WTI, RBOB gasoline and No. 1 heating oil as well as the spot prices since energy contracts for the period running from 2^{nd} of January 2006 to 22^{nd} of May 2015. Energy contracts generally mature after one month. The static OLS hedge ratios obtained from the formula $\gamma=\frac{\sigma_{r_s r_f}}{\sigma^2_{r_f}}$ are $\gamma'=[0.9173771 \quad 0.903419 \quad 1.004232]$, and the dynamic OLS ratios obtained from the dynamic conditional covariances after fitting the DCC-GARCH are shown in Figure 5.1. In as much as these hedge ratios are dynamic, they do not account for the interdependence within markets and as such they do not take care of any volatility spillovers in the energy markets.

Given the refining profits shown below, we seek to explore various hedge ratios, both

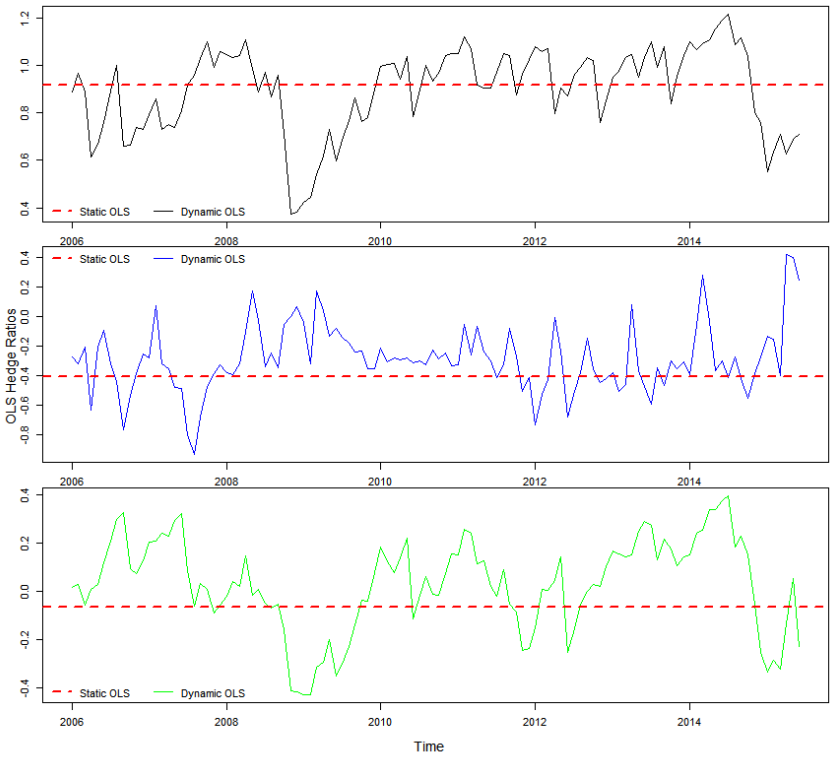

Figure 5.1: Dynamic OLS hedge ratios

static and dynamic.

$$\pi_t = \frac{2}{3}G_{s,t} + \frac{1}{3}H_{s,t} - C_{s,t} + \frac{2}{3}\gamma_1\left(G_{f,t-1} - G_{f,t}\right) + \frac{1}{3}\gamma_2\left(H_{f,t-1} - H_{f,t}\right) + \gamma_3\left(C_{f,t} - C_{f,t-1}\right) - y_t,$$

For the static case, with no assumption on the agents utility preferences, the respective values of $\boldsymbol{A}$ and $\boldsymbol{B}$ defined in equation (5.19) are obtained from the variance covariance matrix represented in table 5.2.

Table 5.2: Variance covariance matrix for log returns of energy spot and futures contracts

	$\Delta\sigma_{C_f}$	$\Delta\sigma_{C_s}$	$\Delta\sigma_{G_f}$	$\Delta\sigma_{G_s}$	$\Delta\sigma_{H_f}$	$\Delta\sigma_{H_s}$
$\Delta\sigma_{C_f}$	8.753	8.79	8.498	8.423	6.777	6.725
$\Delta\sigma_{C_s}$	8.79	8.856	8.532	8.426	6.867	6.816
$\Delta\sigma_{G_f}$	8.498	8.532	13.554	12.434	7.948	7.754
$\Delta\sigma_{G_s}$	8.423	8.427	12.434	14.938	8.122	7.840
$\Delta\sigma_{H_f}$	6.777	6.867	7.948	8.122	8.178	7.388
$\Delta\sigma_{H_s}$	6.725	6.816	7.754	7.840	7.388	7.338

From table 5.2 and equations (5.17) to (5.19), we find that

$$A = \begin{bmatrix} 0.012048262 & 0.003532519 & -0.011330396 \\ 0.003532519 & 0.001817262 & -0.004518203 \\ 0.003532519 & 0.004518203 & 0.017505868 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.0031242763 \\ 0.0006740192 \\ -0.0018662056 \end{bmatrix} \text{ and } \gamma = \begin{bmatrix} 0.31807891 \\ -0.40935395 \\ -0.06513709 \end{bmatrix}.$$

and

Figure 5.2 shows plots of dynamic hedge ratios under the mean variance criterion where the returns are fitted using the DCC-GARCH model. Here, utiliy preferences are not considered, but volatility spillovers are accounted for.

For the dynamic hedge which takes care of utility maximisation as well as volatility spill overs, we have

$$A_t = \begin{bmatrix} \frac{2}{3}\sigma^2_{G_{f,t}} & \frac{1}{3}\sigma_{G_{f,t}H_{f,t}} & -\sigma_{G_{f,t}C_{f,t}} \\ \frac{2}{3}\sigma_{G_{f,t}H_{f,t}} & \frac{1}{3}\sigma^2_{H_{f,t}} & -\sigma_{H_{f,t}C_{f,t}} \\ \frac{2}{3}\sigma_{G_{f,t}C_{f,t}} & \frac{1}{3}\sigma_{H_{f,t}C_{f,t}} & -\sigma^2_{C_{f,t}} \end{bmatrix},$$

$$B_t = \begin{bmatrix} \frac{E\left[G_{f,t-1}-G_{f,t}\right]}{\lambda} + \frac{2}{3}\sigma_{G_{s,t}G_{f,t}} + \frac{1}{3}\sigma_{H_{s,t}G_{f,t}} - \sigma_{C_{s,t}G_{f,t}} \\ \frac{E\left[H_{f,t-1}-H_{f,t}\right]}{\lambda} + \frac{2}{3}\sigma_{G_{s,t}H_{f,t}} + \frac{1}{3}\sigma_{H_{s,t}H_{f,t}} - \sigma_{C_{s,t}H_{f,t}} \\ -\frac{E\left[C_{f,t}-C_{f,t-1}\right]}{\lambda} + \frac{2}{3}\sigma_{G_{s,t}C_{f,t}} + \frac{1}{3}\sigma_{H_{s,t}C_{f,t}} - \sigma_{C_{s,t}C_{f,t}} \end{bmatrix}$$

and

$$\gamma_t' = \begin{bmatrix} \gamma_{1,t} & \gamma_{2,t} & \gamma_{3,t} \end{bmatrix},$$

and the dynamic values of A and B are obtained from the dynamic correlation models fitted in section 4.6.5.

Figure 5.3 and 5.4, we consider two cases, a risk seeker and a risk averse agent or investor respectively. For Figures 5.2, 5.3 and 5.4, the first panel gives the values of dynamic γ_1

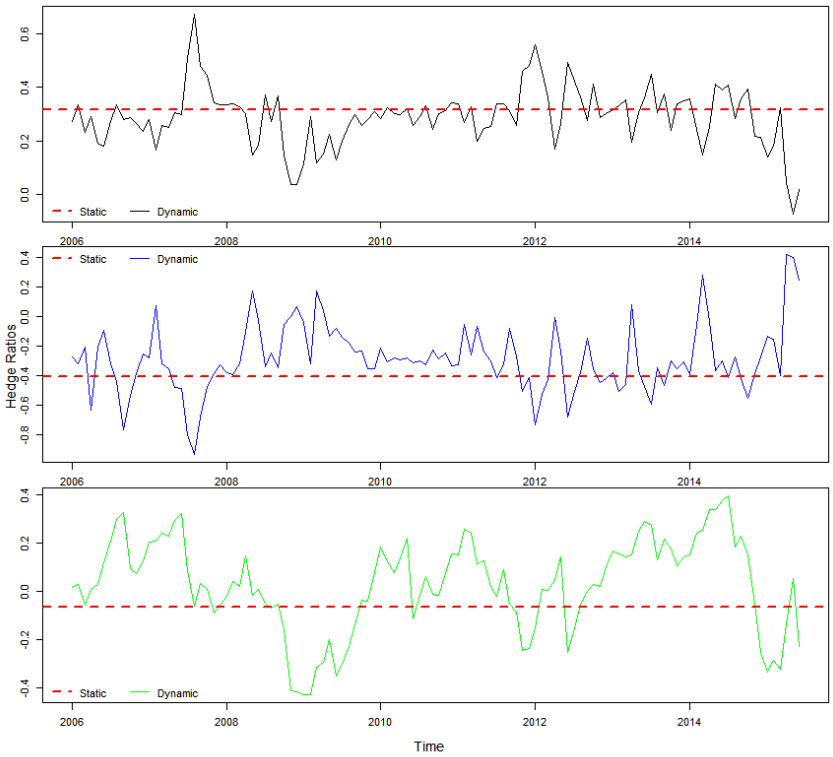

Figure 5.2: Dynamic hedge ratios for 3:2:1 crack spread with no utility

which represents the positions held on the gasoline futures, the second panel represents the values of γ_2 representing the positions held on the heating oil futures and the last panel displays the values of γ_3 which are the dynamic hedge ratios for the crude oil futures. From figures 5.2, 5.3 and 5.4, it is clear that a dynamic hedge does give a representation of what actually happens through time in terms of price fluctuations. The hedge ratios are rebalanced every time with respect to various occurrences in the market and this would give more realistic hedge scenarios. The dynamic ratios reflect the volatilities experienced in the market through time. The difficulty with application of dynamic ratios would be with respect to the cost of rebalancing the portfolio each time to reflect the dynamic hedge ratios.

The respective profits are as shown in Figure 5.5 where panel one represents profits from spot trading, panel two shows profits arising from a naive or full hedge where $\gamma_1 = \gamma_2 = \gamma_3 = 1$. This kind od scenario is very volatile and records some of the highest profits and losses. Panel three gives profits from a dynamic OLS hedge ratio, and

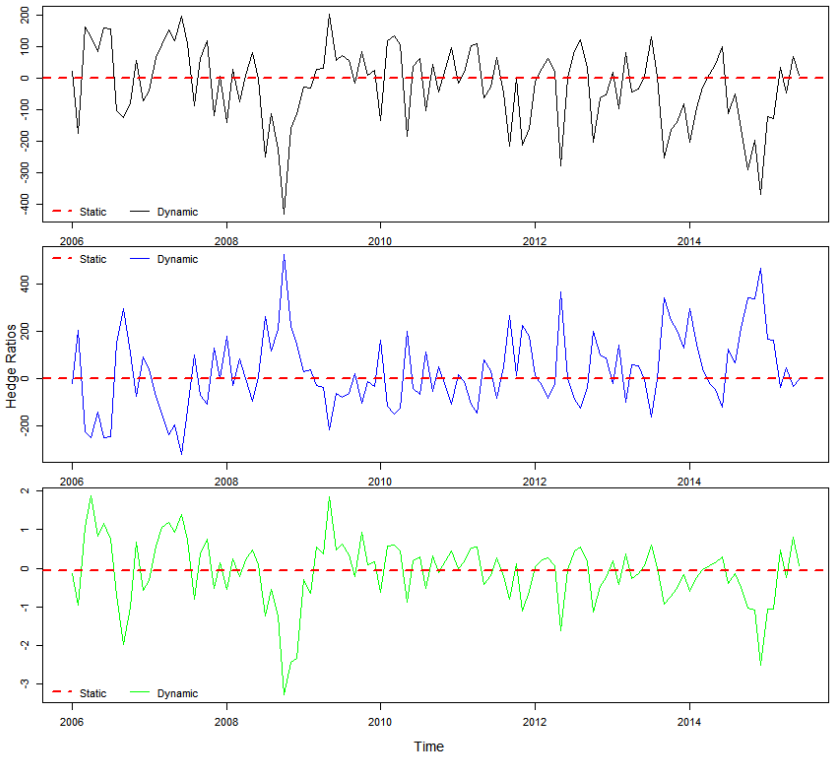

Figure 5.3: 3:2:1 crack dynamic hedge ratios with $\lambda = 2$

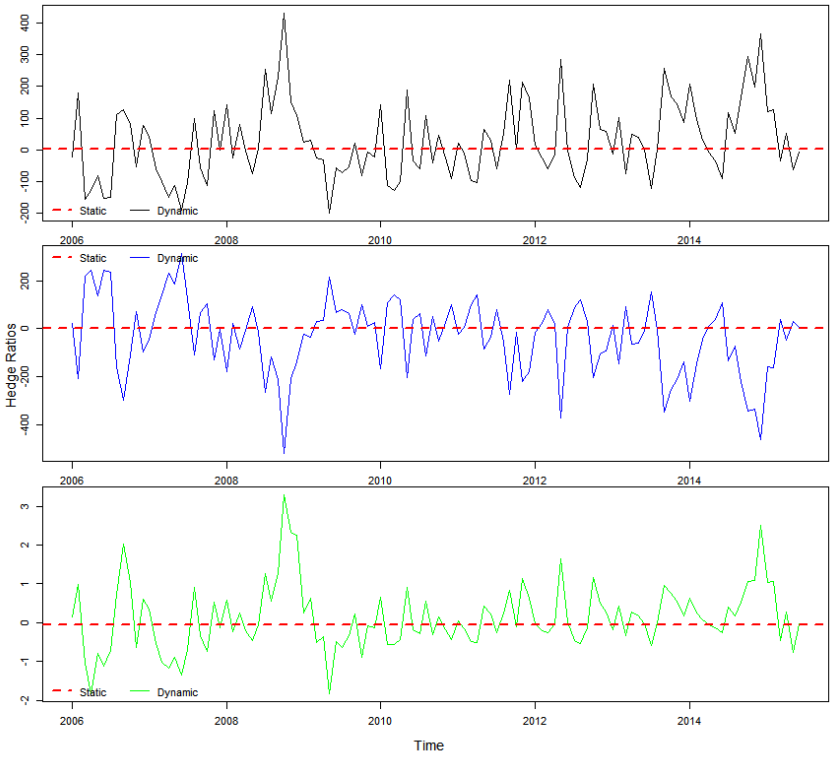

Figure 5.4: 3:2:1 crack dynamic hedge ratios with $\lambda = -2$

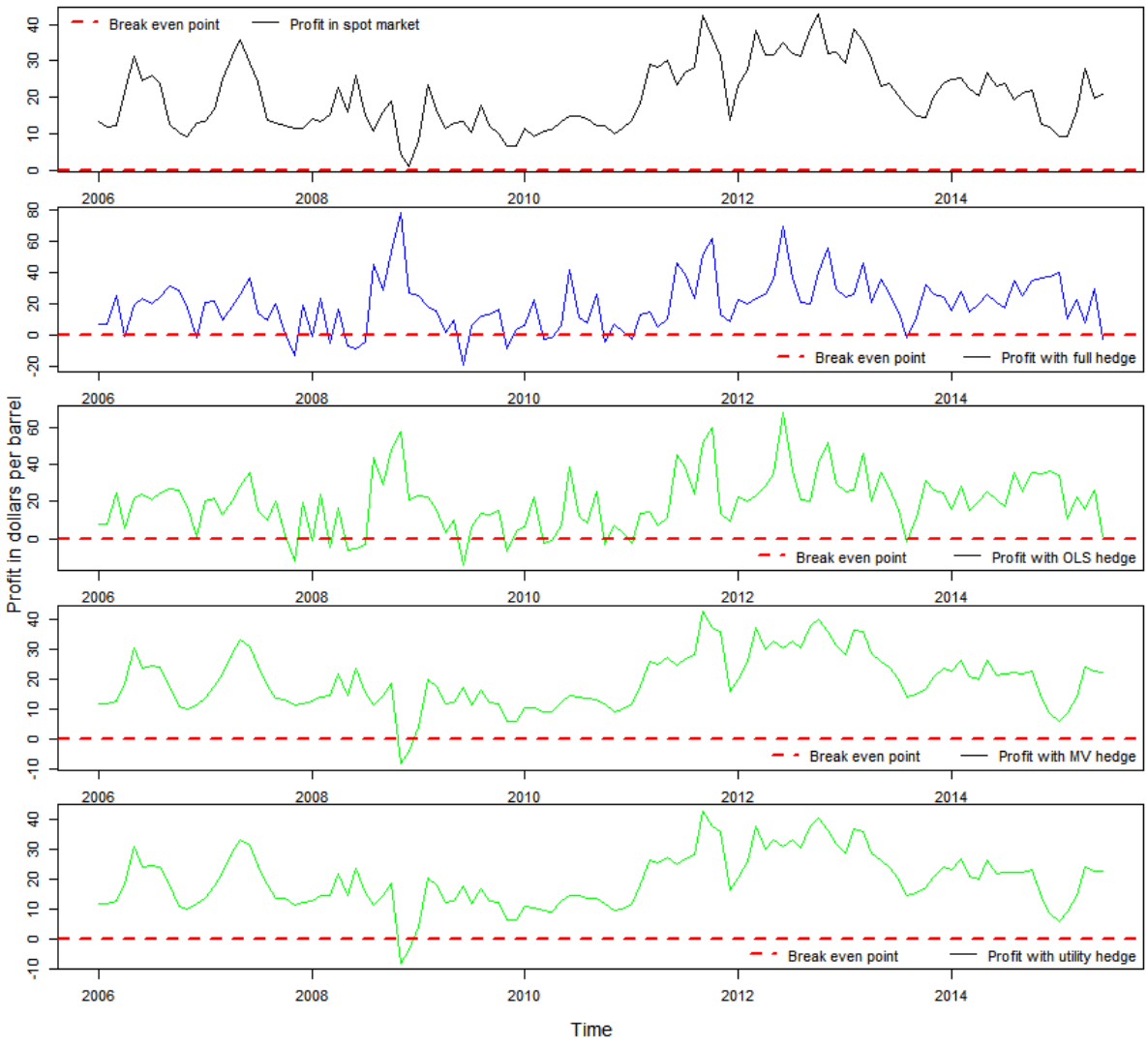

Figure 5.5: Profits across various hedging strategies

panel four and five give the profits from variances minimising hedge ratio under minimum variance. For panel five , we assume the investor is maximising utility, subject to a negative exponential utility function. This gives the exact same results are those obtained by applying mean-variance portfolio theory.

Table 5.3 gives a summary of the hedge ratios, portfolio variances and hedge effectiveness measures.

From this Table, we see the naive hedge gives the weakest hedge effectiveness and largest portfolio variance after the case with no hedge. However, the static mean variance

Table 5.3: Average hedge ratios and hedge effectiveness

Hedge Method	γ_1	γ_2	γ_3	Portfolio variance	Hedge effectiveness
No hedge	0	0	0	380.99	0
Naive hedge	1	1	1	285.65	0.2502
Static OLS	0.9174	0.9034	1.0042	265.55	0.303
Dynamic OLS	0.8982	0.9175	1.0043	240.01	0.37
Static MV	0.3181	-0.4094	-0.0651	80.018	0.79
Dynamic MV	0.2898	-0.286	0.0302	87.111	0.7714
Dynamic utility maximizing	26.123	-34.048	0.1223	87.111	0.7714
Spot market	0	0	0	**82.776**	**0.7827**

hedge still looks stronger than any other strategy though the dynamic hedge is a very clear improvement from the cases where we use the naive hedge or the OLS hedge ratio. Both these two do not account for interdependencies or spillovers present across the energy markets. The dynamic minimum variance hedge and the dynamic utility maximising hedge both give the exact same results with respect to the portfolio variance and hedge effectiveness implying that incorporating the risk aversion coefficient for the trader has no effect on the portfolio variance as well as the hedge effectiveness, it only affects the positions taken by the trader. This means incorporating the spillover effects is very effective. Introducing multivariate heteroscedasticity may not seem to produce better results compared to the static hedge where we already improve the trading strategy by trading a crack spread, however, considering that it gives a different hedge ratio at every time point, it definitely improves the hedging strategy for the refiner over time.. It seems that having taken care of the interdependencies, all we need is to take a position on the input and output products and the trader wil be adequately covered.

Considering the profits shown in table 5.5, there are more opportunities for loss under the full hedge, followed by the OLS hedge ratios. For the last two panels on the graph where we consider the mean variance hedge and the mean variance hedge with risk aversion, profits present in exactly the same way. This shows that with the minimum variance hedge, the refiner is adequately covered even if the risk aversion parameter is not considered.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This work employs statistical and econometric techniques to investigate and model financial time series trends in energy markets. To do this, daily closing prices for a period of about 10 years for Cushing OK WTI, RBOB and number 1 heating oil spot and futures contracts traded in the NYMEX are considered. The work also investigates the existence of stylised facts in these series, in order to fit an appropriate model that adequately describes the market dynamics.

Price data are generally heteroscedastic and co-evolve with occasional spikes and jumps. The price series have a strong positive correlation. The series are tested for stationarity using the DF test and results show the existence of unit roots. Normality is tested for using the JB test and non-normality is established. Return series are then generated from the price series through differencing, and then tested for stationarity using the DF test. The results reveal that return series are indeed mean stationary, but are definitely not variance stationary. Several ARMA models are fit to the return series and the standardised residuals analysed. For the crude futures return series, the best model turns out to be an $ARMA(6, 11)$ and for the crude spot it is an $ARMA(7, 11)$ among others, based on their AICs. These resultant models however contravene the Gaussian innovation assumption.

We then propose a combined $ARMA(p, q) - GARCH(P, Q)$ model to capture the ARCH effects in the variance, then find that the best model under these circumstances is the $ARMA(0, 0) - GARCH(1, 1)$, implying a constant mean conditional variance equation. We finally find that for example, for the crude futures return series, the $GARCH(16, 19)$

is the best model based on the AIC, although a $GARCH(1,1)$ does quite well for this and all the other six series with the residual analysis conforming to the assumption of Gaussian Innovations. GARCH models can therefore adequately model the trends and patterns in the energy markets. The trends also depict time varying variability and high persistence of oil price shocks. These shocks therefore have a significant impact on the prices of these energy prices.

There is evidence of co-integration and co-evolution after carrying out tests. All the series are integrated of order one and both the E-G and Johansen's cointegration test reveal co-integration relations. From the Johansen's test for co-integration, we obtain a rank of 4, making this the optimal number of lags. The trace test also reveals 4 co-integrating relations. After fitting a VECM, the residuals are not multivariate normal, and this is confirmed by the results of the multivariate extension of the JB residuals normality test. The null hypothesis of no ARCH effects is also rejected, meaning the noise is MGARCH. This means we end up with a vector error correction (VEC) model with GARCH noise which can be fitted as a seemingly unrelated regression (SUR)-GARCH model. However, given that in Chapter 3, the most parsimonious model for the univariate analysis had a zero mean and GARCH variance, i.e. ARMA(0,0)-GARCH(1,1), we proceed to fit MGARCH models to price data to capture the source and magnitude of volatility spill overs. We find that the volatility spillover between crude oil, gasoline and heating oil price series is obvious. Specifically, we fit the BEKK, CCC and DCC variants of multivariate GARCH. For the BEKK model, convergence is achieved only for the bivariate case due to the curse of high dimensionality.

The DCC-GARCH(1,1) turns out to be the most parsimonious since the joint ARCH and GARCH parameters are significant. Conditional variance and conditional correlations across crude oil and distillate fuel markets are time varying and dynamic in conformity with the arguments in most literature. The parameters are estimated using MLE and the dynamic conditional correlations and covariances are obtained. These parameters are

then used in the estimation of the dynamic hedge ratios. A comparison of the naive hedge, OLS hedge, both static and dynamic and mean variance hedge, both static and dynamic is carried out and the static minimum variance using the crack spread emerges the best with the highest hedge effectiveness and lowest portfolio variance. It should be noted that on the crack, taking care of interdependencies is just enough to achieve better hedge ratios.

Bringing in dynamic hedges improves the hedge ratios, but not above what is achieved by the static model in the long run, but given the fact that volatility is time varying, the dynamic hedge ratios are by far more realistic as they give a true representation of the variance dynamics of the day. For shorter maturities, they would most definitely emerge superior. It should be further noted that for the dynamic hedge, there is loss of simplicity and an additional cost implication brought about by the need to constantly rebalance the hedge ratios. With respect to OLS hedge ratios, it should be noted that they perform very poorly on the crack spread and in cases where volatility spillover has to be considered. In any hedge scenario however, the spill over effects are important to consider as they give a significant improvement to the hedge ratios.

6.2 Recommendations

Given the evidence of the existence of co–integration, and hence the implication of reversion to a long term mean, mean reverting factor models can be employed to investigate the spill over effect in the volatilities of energy prices and other stochastic volatility models can be used to capture and explain the interdependencies.

6.3 Areas for Further Research

In this work, the investor's utility preferences are assumed to follow a negative exponential function and this implies a mean variance framework with constant risk aversion. Other utility functions and portfolio optimization techniques such as generalized semi-variance (GSV) can also be employed to investigate how different the dynamic hedge ratios would

be. In the same respect, other measures of risk such as the semi-variance or the mean-Gini coefficient of return can also be minimized to determine the optimal hedge ratio.

Portfolios constructed from other cracking ratios or other hedging strategies can also be considered. The same analysis can also be carried out for emerging markets. To study dynamic hedge ratios, Flexible Dynamic Conditional Correlation (FDCC) models which generalize the DCC and relax the assumption of common dynamics among all assets can be considered. Similarly, SUR models can be used to fit VECM models with MGARCH errors. VECM with GARCH errors can be represented by fitting a VEC-GARCH model by rewriting the VEC-GARCH model as a seemingly unrelated regression (SUR)-GARCH model.

REFERENCES

Alexander, C. (2008). *Practical financial econometrics*. Wiley.

Alexander, C., I. Giblin, and W. Weddington (2002). "Cointegration and asset allocation: A new active hedge fund strategy". *ISMA Centre Discussion Papers in Finance Series*.

Alizadeh, A. H., M. G. Kavussanos, and D. A. Menachof (2004). "Hedging against bunker price fluctuations using petroleum futures contracts: constant versus time-varying hedge ratios". *Applied Economics* 36.12, pp. 1337–1353.

Baillie, R. T. and T. Bollerslev (1990). "A Multivariate Generalized ARCH Approach to Modeling Risk Premia in Forward Foreign Exchange Rate Markets". *Journal of International Money and Finance* 9, pp. 309–324.

Bauwens, L., C. M. Hafner, and S. Laurent (2012). *Handbook of volatility models and their applications*. Vol. 3. John Wiley & Sons.

Bauwens, L., S. Laurent, and J. V. Rombouts (2006). "Multivariate GARCH Models: A Survey". *Journal of Applied Econometrics* 21, pp. 79–109.

Beck, A. et al. (2013). "Empirical analysis of ARMA-GARCH models in market risk estimation on high-frequency US data". *Studies in Nonlinear Dynamics and Econometrics* 17.2, pp. 167–177.

Benavides, G. (2004). "Price volatility forecasts for agricultural commodities: an application of historical volatility models, option implieds and composite approaches for futures prices of corn and wheat". *Option Implieds and Composite Approaches for Futures Prices of Corn and Wheat (January 2004)*.

Black, F. (1976). "The pricing of commodity Contracts". *Journal of Financial Economics* 3, pp. 167–179.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroscedasticity". *Journal of Economics* 31, pp. 307–327.

Bollerslev, T. (1987). "A Conditionally Heteroscedastic Time Series Model for Speculative Prices and Rates of Return". *The review of Economics and statistics* 69(3), pp. 542–547.

Bollerslev, T. (1990). "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model". *The review of economics and statistics*, pp. 498–505.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992). "ARCH modeling in finance : A review of the theory and empirical evidence". *Journal of Econometrics* 52.1-2, pp. 5–59.

Box, G. E. and G. M. Jenkins (1976). *Time series analysis: forecasting and control, revised ed.* Holden-Day.

Brockwell, P. J. (2002). *Introduction to time series and forecasting*. Vol. 1. Taylor & Francis.

Byun, S. J. and B. Min (2011). "Conditional Volatility and the GARCH Option Pricing Model with Non-normal Innovations". *The Joural of Futures Markets* 00, pp. 1–28.

Campbell, J. Y., A. W.-C. Lo, and A. C. MacKinlay (1997). *The econometrics of financial markets*. Vol. 2. princeton University press Princeton, NJ.

Campbell, J. and L. Hentschel (1992). "No news is good news: An assymetric model of changing volatility in stock returns". *Journal of Financial Econometrics* 31, pp. 281–318.

Carmona, R. and V. Durrleman (Oct. 2003). "Pricing and Hedging Spread Options". *SIAM Review* 45.4, pp. 627–685.

Cecchetti, S. G., R. E. Cumby, and S. Figlewski (1988). "Estimation of the optimal futures hedge". *The Review of Economics and Statistics*, pp. 623–630.

Chan, N. H. (2011). *Time Series: Applications to Finance with R and S-Plus*. Vol. 837. John Wiley & Sons.

Chang, C.-L., M. McAleer, and R. Tansuchat (2010). "Analyzing and forecasting volatility spillovers, asymmetries and hedging in major oil markets". *Energy Economics* 32.6, pp. 1445–1455.

Chang, C.-L., M. McAleer, and R. Tansuchat (2011). "Crude oil hedging strategies using dynamic multivariate GARCH". *Energy Economics* 33.5, pp. 912–923.

Chen, S.-S., C.-F. Lee, and K. Shrestha (2001). "On a Mean-Generalized Semivariance Approach to Determining the Hedge Ratio". *Journal of Futures Markets* 21.6, pp. 581–598. ISSN: 1096-9934.

Chen, S.-S., C.-f. Lee, and K. Shrestha (2003). "Futures hedge ratios: a review". *The Quarterly Review of Economics and Finance* 43.3, pp. 433–465.

Clewlow, L. and C. Strickland (2000). *Energy Derivatives: Pricing and Risk Management*. Lacima.

Cont, R. (2001). "Empirical properties of asset returns: stylized facts and statistical issues". *Quantitative Finance* 1, pp. 223–236.

Cross, P., P. Desrochers, and H. Shimizu (Dec. 2013). *The Economics of Petroleum Refining: Understanding the business of processing crude oil into fuels and other value added products*. Tech. rep. Canadian Fuels Association.

Danielsson, J. (2011). *Financial risk forecasting: The theory and practice of forecasting market risk with implementation in R and Matlab*. Vol. 588. John Wiley & Sons.

Dash, S. and E. Skyba (Nov. 2011). *Inside the Oil Markets: Crack Spreads*. Tech. rep. TradeStation Labs.

De Jong, A., F. De Roon, and C. Veld (1997). "Out-of-sample hedging effectiveness of currency futures for alternative models and hedging strategies". *Journal of Futures Markets* 17.7, pp. 817–837. ISSN: 1096-9934.

Dickey, D. A. and W. A. Fuller (June 1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root". *Journal of the American Statistical Association* 74 (366a).

Ding, X.-S., X. Chen, and J.-H. Zhang (Sept. 2009). "Certainty Equivalent in Portfolio Management". *The eighth International Symposium on Operations Research and it's Applications*, pp. 115–123.

Dooley, J. J. (1998). "Unintended consequences: energy R&D in a deregulated energy market". *Energy Policy* 26.7, pp. 547–555.

Dorsman, A., J. L. Simpson, and W. Westerman (2013). *Energy economics and financial markets*. Springer.

Duan, J. C. (1995). "The GARCH Option Pricing Model". *Journal of Mathematical Finance* 5(1), pp. 13–32.

Duan, J.-C. and S. R. Pliska (2004). "Option Valuation with Co-integrated Asset Prices." *Journal of Economic Dynamics and Control* 28, N0. 4, pp. 727–754.

Durbin, J. and G. S. Watson (1950). "Testing for serial correlation in least squares regression: I". *Biometrika* 37.3/4, pp. 409–428.

Durbin, J. and G. S. Watson (1951). "Testing for serial correlation in least squares regression. II". *Biometrika* 38.1/2, pp. 159–177.

Ederington, L. H. (1979). "The hedging performance of the new futures markets". *The Journal of Finance* 34.1, pp. 157–170.

EIA (June 2015). *U.S. Energy Information Administration*. URL: http://www.eia.gov/about/.

Enders, W. (2008). *Applied econometric time series*. John Wiley & Sons.

Engle, R. F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the variance of United Kingdom Inflation". *Econometrica* 50(4), pp. 987–1007.

Engle, R. F. (1984). "Wald, likelihood ratio, and Lagrange multiplier tests in econometrics". *Handbook of econometrics* 2, pp. 775–826.

Engle, R. F. and C. W. Granger (1987). "Co-integration and error correction: representation, estimation, and testing". *Econometrica: journal of the Econometric Society*, pp. 251–276.

Engle, R. F. and K. F. Kroner (1995). "Multivariate Simultaneous Generalized ARCH". *Econometric Theory* 11, No. 1, pp. 122–150.

Engle, R. F. and V. Ng (1993). "Measuring and testing the impact of news on volatility". *Journal of Finance* 48, pp. 1749–1778.

Engle, R. F. and A. J. Patton (2001). "What good is a volatility model". *Quantitative finance* 1.2, pp. 237–245.

Engle, R. F. and K. Sheppard (2001). *Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH*. Tech. rep. National Bureau of Economic Research.

Fabozzi, F. J., S. M. Focardi, and P. N. Kolm (2006). *Financial Modeling of the Equity Market: From CAPM to Cointegration*. Frank J. Fabozzi Series. Wiley.

Featherstone, A. M. and C. B. Moss (1990). "Quantifying Gains to Risk Diversification using Certainty Equivalence in a Mean-Variance model: An Application to Florida Citrus". *Southern Journal of Agricultural Economics*.

Fishburn, P. C. (Mar. 1977). "Mean-risk Analysis with risk Associated with Below-Target Returns". *The American Economic Review* 67.2, pp. 116–126.

Fleming, J. and B. Ostdiek (1999). "The impact of energy derivatives on the crude oil market". *Energy Economics* 21.2, pp. 135–167.

Foresti, P. (2006). "Testing for Granger causality between stock prices and economic growth". *MPRA*.

Francq, C. and J.-M. Zakoian (2011). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.

French, K., G. Schwert, and R. Stambaugh (1987). "Expected stock returns and volatility". *Journal of Financial Econometrics* 19, pp. 3–29.

GARP (2009). *Foundation of Energy Risk Management*. John Wiley & Sons.

Girma, P. B., A. S. Paulson, et al. (1999). "Risk arbitrage opportunities in petroleum futures spreads". *Journal of Futures Markets* 19.8, pp. 931–955.

Glosten, L., R. Jagannathan, and D. Runkle (1993). "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return of Stocks". *Journal of Finance* 48, pp. 1779–1801.

Granger, C. W. (1969). "Investigating causal relations by econometric models and cross-spectral methods". *Econometrica: Journal of the Econometric Society*, pp. 424–438.

Granger, C. W. and P. Newbold (1974). "Spurious regressions in econometrics". *Journal of econometrics* 2.2, pp. 111–120.

Haigh, M. S. and M. T. Holt (Mar. 2002). "Crack Spread Hedging: Accounting for Time-Varying Volatility Spill Overs in the Energy Futures Markets". *Journal of Applied Econometrics* 17, pp. 269–289.

Harris, R. I. and R. Sollis (2003). *Applied time series modelling and forecasting*. J. Wiley.

Herwartz, H. and H.-E. Reimers (2002). "Empirical modelling of the DEM/USD and DEM/JPY foreign exchange rate: Structural shifts in GARCH-models and their implications". *Applied Stochastic Models in Business and Industry* 18, pp. 3–22.

Hunjra, A. I. et al. (2011). "Risk and Return Relationship in Stock Market and Commodity Prices: A Comprehensive Study of Pakistani Markets". *World Applied Sciences Journal* 13.3, pp. 470–481.

Hyndman, R. J. and A. B. Koehler (2006). "Another look at measures of forecast accuracy". *International journal of forecasting* 22.4, pp. 679–688.

Jarque, C. M. and A. K. Bera (1987). "A Test for Normality of Observations and Regression Residuals". *International Statistical Review / Revue Internationale de Statistique* 55.2, pp. 163–172.

Johansen, S. (1988). "Statistical analysis of cointegration vectors". *Journal of economic dynamics and control* 12.2, pp. 231–254.

Johansen, S. (1991). "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models". *Econometrica: Journal of the Econometric Society*, pp. 1551–1580.

Johansen, S. (1995). "Identifying restrictions of linear equations with applications to simultaneous equations and cointegration". *Journal of econometrics* 69.1, pp. 111–132.

Johansen, S. and K. Juselius (1990). "Maximum likelihood estimation and inference on cointegration-with applications to the demand for money". *Oxford Bulletin of Economics and statistics* 52.2, pp. 169–210.

Jondeau, E., S.-H. Poon, and M. Rockinger (2006). *Financial modeling under non-gaussian distributions*. 1st ed. Springer Finance. Springer.

Kanamura, T., S. T. Rachev, and F. J. Fabozzi (2011). *A profit model for spread trading with an application to energy futures*. Tech. rep. Working paper series in economics.

Liu, H., E. Erdem, and J. Shi (2011). "Comprehensive evaluation of ARMA–GARCH (-M) approaches for modeling the mean and volatility of wind speed". *Applied Energy* 88.3, pp. 724–732.

Ljung, G. M. and G. E. Box (1978). "On a measure of lack of fit in time series models". *Biometrika* 65.2, pp. 297–303.

Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Lütkepohl, H. and M. Krätzig (2004). *Applied time series econometrics*. Cambridge University Press.

Mandelbrot, B. (1963). "The Variation of Certain Speculative Prices". *The Journal of Business* 36.4, pp. 394–419.

Markowitz, H. (1952). "Portfolio Selection". *The Journal of finance* 7.1, pp. 77–91.

Mills, T. C. and R. N. Markellos (2008). *The econometric modelling of financial time series*. Cambridge University Press.

Montgomery, D. C., C. L. Jennings, and M. Kulahci (2011). *Introduction to time series analysis and forecasting*. Vol. 526. John Wiley & Sons.

Moshiri, S. and F. Foroutan (2006). "Forecasting Nonlinear Crude Oil Futures Prices". *The Energy Journal* Volume 27.Number 4, pp. 81–96.

Osterwald-Lenum, M. (1992). "A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics1". *Oxford Bulletin of Economics and statistics* 54.3, pp. 461–472.

Pagan A.R.and Schwert, G. (1990). "Alternative models for conditional stock volatility". *Journal of Econometrics* 45, pp. 267–290.

Pfaff, B. (2012). *Financial risk modelling and portfolio optimization with R*. John Wiley & Sons.

Pindyck, R. S. (2004). "Volatility and Commodity Price Dynamics". *The Journal of futures Markets* 24.11, pp. 1029–1047.

Rachev, S. T. et al. (2007). *Financial econometrics: from basics to advanced modeling techniques*. Vol. 150. John Wiley & Sons.

Ramos, S. B. and H. Veiga (2014). *The Interrelationship Between Financial and Energy Markets*. Springer.

Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*. Springer Series in Statistics. Springer US.

Salman, D. M. and E. M. Atya (2014). "What is the role of Financial Development and Energy Consumption on Economic Growth? New Evidence from North African Countries". *International Journal of Finance & Banking Studies* 3.1, pp. 137–149.

Schwert, G. (1989). "Why does stock market volatility change over time?" *Journal of Finance* 44, pp. 1115–1154.

Shumway, R. H. and D. S. Stoffer (2010). *Time series analysis and its applications: with R examples*. Springer Science & Business Media.

Sjö, B. (2008). "Testing for unit roots and cointegration". *Web adresi*.

Taylor, S. J. (2007). "Modelling financial time series".

Tsay, R. S. (2010). *Analysis of Financial Time Series*. 3rd ed. John Wiley & Sons, Inc. Hoboken, New Jersey.

Tsay, R. S. (2013). *Multivariate Time Series Analysis: With R and Financial Applications*. 1st ed. Wiley.

Von Eye, A. and G. A. Bogat (2004). "Testing the assumption of multivariate normality". *Psychology Science* 46, pp. 243–258.

Wei, W. and S. Wei (1990). *Time series analysis: univariate and multivariate methods*.

Winston, C. (1993). "Economic Deregulation: Days of Reckoning for Microeconomists". English. *Journal of Economic Literature* 31.3, pp. 1263–1289.

Wurtz, D., Y. Chalabi, and L. Luksan (2009). "Parameter estimation of ARMA models with GARCH/APARCH errors an R and SPlus software implementation". *Journal of Statistical Software, forthcoming*.

Zivot, E. and J. Wang (2007). *Modeling Financial Time Series with S-PLUS®*. Vol. 191. Springer Science & Business Media.

Appendices

Appendix A

ARMA MODEL SELECTION MATLAB CODE

```
%% Estimation
%This code estimates all ARMA models with p=0,1,...,15 and q=0,1,...15,
%and then looks at AICs.
MAX_AR = 15;
MAX_MA = 15;
% Use a cell array to hold results from the models.  Cell arrays allow
%arbitrary data to be stored.
models = cell(MAX_AR+1,MAX_MA+1);
% Options to suppress display.  ARMA models are estimated using lsqnonlin
options =  optimset('lsqnonlin');
options.MaxIter = 1000;
options.Display='none';

% Setup arrays to hold values
AICs = zeros(MAX_AR,MAX_MA);

% Get the effective T since MAX_AR will be held back
T = length(Ddata(:,1)) - MAX_AR;
% Loop over AR and MA order
for i=0:MAX_AR
for j=0:MAX_MA
disp(['AR: ' num2str(i) ' MA: ' num2str(j)])
```

```
% The MAX_AR term below enforces the holdbac, which is requires
% when comparing AR models with different lag lengths.  Failing to
% use this will produce log-likelihods base don different number of
% observations.
[p,ll,ser] = armaxfilter(Ddata(:,1),1,1:i,1:j,[],[],options,MAX_AR);
% Use a struct to store results
model = struct();
model.parameters = p;
model.LL = ll; % log likelihoods
model.SER = ser;
models{i+1,j+1} = model;

% Store other values for use later
AICs(i+1,j+1) = ll - 2*(1+i+j);
end
end
```

Appendix B

AIC VALUES FOR THE ARMA MODELS FITTED TO DIFFERENCED CRUDE FUTURES DATA

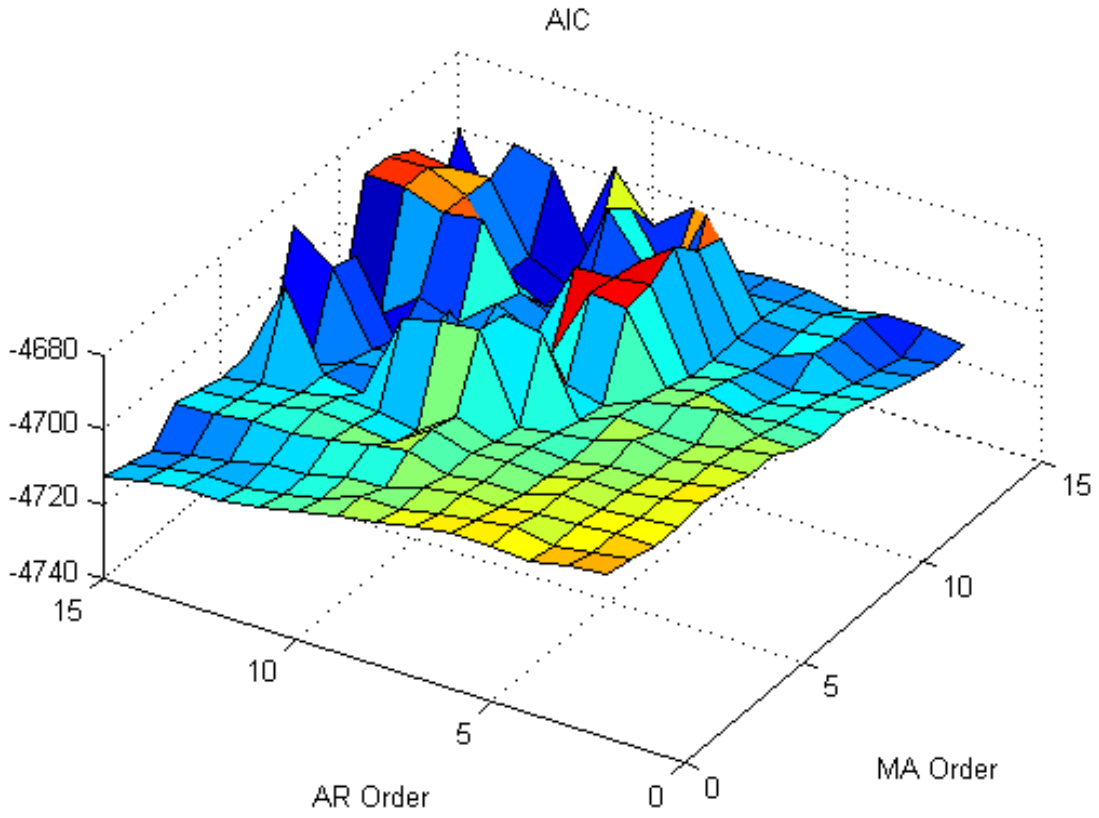

Figure B.1: Plot for the AICs for the various ARMA models fitted for the differenced crude futures price data

Table B.1: Results from the ARIMA Model fitted to the differenced crude futures price series:

	MA															
AR	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	-4.6971	-4.6955	-4.6966	-4.6984	-4.6985	-4.6980	-4.6995	-4.7014	-4.7033	-4.7053	-4.7072	-4.7092	-4.7090	-4.7110	-4.7129	-4.7147
1	-4.6957	-4.6970	-4.6985	-4.6992	-4.6982	-4.6996	-4.7014	-4.7017	-4.7053	-4.7073	-4.7069	-4.7087	-4.7110	-4.7129	-4.7149	-4.7151
2	-4.6964	-4.6984	-4.6984	-4.6998	-4.7001	-4.7016	-4.7017	-4.7033	-4.7052	-4.7067	-4.7083	-4.7102	-4.7127	-4.7147	-4.7166	-4.7141
3	-4.6984	-4.7003	-4.6996	-4.7010	-4.7020	-4.7033	-4.7041	-4.7052	-4.7015	-4.7087	-4.7104	-4.7123	-4.7079	-4.7092	-4.7095	-4.7138
4	-4.6980	-4.6980	-4.6999	-4.7000	-4.7014	-4.7033	-4.7018	-4.7050	-4.7034	-4.7054	-4.7072	-4.7105	-4.7118	-4.7107	-4.7115	-4.7157
5	-4.6979	-4.6995	-4.7015	-4.7033	-4.7045	-4.7055	-4.7051	-4.7045	-4.7055	-4.7075	-4.7102	-4.7110	-4.7104	-4.7107	-4.7121	-4.7142
6	-4.6994	-4.7014	-4.7017	-4.7035	-4.7038	-4.7049	-4.7068	-4.7085	-4.7103	-4.6891	-4.6885	**-4.6841**	-4.6901	-4.6926	-4.7123	-4.7140
7	-4.7014	-4.7034	-4.7031	-4.7050	-4.7048	-4.7068	-4.7086	-4.6888	-4.7077	-4.6887	-4.6899	-4.6914	-4.7126	-4.7146	-4.6950	-4.7148
8	-4.7032	-4.7035	-4.7066	-4.7017	-4.7015	-4.7035	-4.6852	-4.6871	-4.7065	-4.7085	-4.6903	-4.7128	-4.7148	-4.7075	-4.7158	-4.7016
9	-4.7052	-4.7072	-4.7069	-4.7035	-4.7086	-4.7106	-4.6864	-4.7130	-4.7083	-4.7113	-4.7120	-4.7139	-4.7149	-4.6940	-4.6999	-4.7096
10	-4.7072	-4.7092	-4.7093	-4.7055	-4.7071	-4.7091	-4.6890	-4.7111	-4.6980	-4.7121	-4.7072	-4.7095	-4.7185	-4.7185	-4.7173	-4.6973
11	-4.7092	-4.7111	-4.7101	-4.7070	-4.7065	-4.7083	-4.7126	-4.7113	-4.7130	-4.7140	-4.7154	-4.6927	-4.7128	-4.7141	-4.6954	-4.7221
12	-4.7088	-4.7106	-4.7125	-4.7069	-4.7085	-4.7103	-4.7113	-4.7128	-4.7150	-4.7183	-4.7119	-4.6946	-4.6956	-4.6964	-4.6920	-4.7210
13	-4.7105	-4.7125	-4.7143	-4.7087	-4.7099	-4.7113	-4.6907	-4.7173	-4.6962	-4.6986	-4.7197	-4.6907	-4.6909	-4.6963	-4.7175	-4.7212
14	-4.7125	-4.7145	-4.7163	-4.7088	-4.7108	-4.7127	-4.7125	-4.7144	-4.6883	-4.7101	-4.7191	-4.6908	-4.6916	-4.6950	-4.7149	-4.7003
15	-4.7142	-4.7162	-4.7136	-4.7155	-4.7173	-4.7187	-4.6950	-4.7152	-4.7171	-4.7154	-4.7196	-4.7215	-4.7010	-4.7191	-4.7180	-4.6963

all values multiplied by 10^3

Appendix C

AICS AND BEST MODEL FIT FOR CRUDE SPOT PRICES

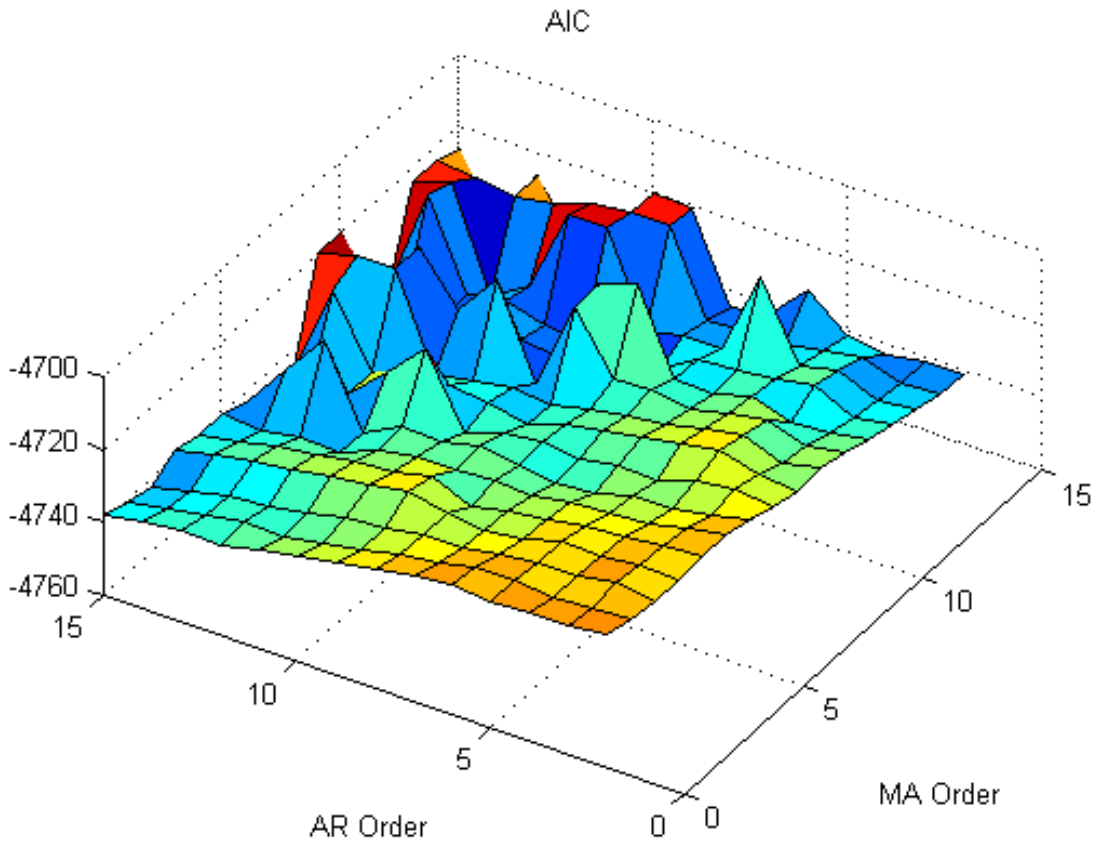

Figure C.1: Plot for the AICs for the various ARMA models fitted for the differenced crude spot price data

Table C.1: Results from the ARIMA Model fitted to the differenced crude spot price series:

	MA															
AR	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	-4.7246	-4.7237	-4.7250	-4.7249	-4.7262	-4.7244	-4.7259	-4.7278	-4.7297	-4.7317	-4.7334	-4.7354	-4.7350	-4.7367	-4.7385	-4.7402
1	-4.7238	-4.7255	-4.7265	-4.7268	-4.7256	-4.7259	-4.7279	-4.7298	-4.7317	-4.7313	-4.7330	-4.7350	-4.7369	-4.7386	-4.7405	-4.7408
2	-4.7246	-4.7262	-4.7246	-4.7258	-4.7266	-4.7279	-4.7299	-4.7289	-4.7306	-4.7336	-4.7349	-4.7368	-4.7386	-4.7402	-4.7413	-4.7411
3	-4.7248	-4.7265	-4.7267	-4.7273	-4.7281	-4.7299	-4.7319	-4.7301	-4.7275	-4.7294	-4.7312	-4.7375	-4.7335	-4.7355	-4.7375	-4.7427
4	-4.7257	-4.7253	-4.7267	-4.7270	-4.7285	-4.7306	-4.7317	-4.7335	-4.7294	-4.7314	-4.7331	-4.7344	-4.7355	-4.7375	-4.7395	-4.7414
5	-4.7244	-4.7259	-4.7279	-4.7299	-4.7305	-4.7315	-4.7348	-4.7332	-4.7313	-4.7335	-4.7352	-4.7372	-4.7364	-4.7159	-4.7416	-4.7325
6	-4.7259	-4.7278	-4.7298	-4.7310	-4.7317	-4.7335	-4.7354	-4.7350	-4.7332	-4.7352	-4.7340	-4.7364	-4.7375	-4.7402	-4.7415	-4.7427
7	-4.7276	-4.7296	-4.7298	-4.7300	-4.7342	-4.7326	-4.7343	-4.7340	-4.7351	-4.7371	-4.7345	**-4.7143**	-4.7431	-4.7403	-4.7422	-4.7441
8	-4.7296	-4.7315	-4.7318	-4.7275	-4.7292	-4.7337	-4.7330	-4.7350	-4.7386	-4.7393	-4.7179	-4.7172	-4.7405	-4.7424	-4.7204	-4.7221
9	-4.7314	-4.7334	-4.7324	-4.7293	-4.7315	-4.7332	-4.7350	-4.7386	-4.7362	-4.7381	-4.7426	-4.7417	-4.7434	-4.7177	-4.7197	-4.7216
10	-4.7333	-4.7353	-4.7347	-4.7309	-4.7326	-4.7350	-4.7346	-4.7200	-4.7350	-4.7401	-4.7183	-4.7402	-4.7425	-4.7172	-4.7214	-4.7470
11	-4.7353	-4.7350	-4.7367	-4.7324	-4.7341	-4.7393	-4.7400	-4.7289	-4.7397	-4.7420	-4.7298	-4.7420	-4.7394	-4.7412	-4.7242	-4.7398
12	-4.7348	-4.7367	-4.7371	-4.7338	-4.7354	-4.7389	-4.7180	-4.7388	-4.7389	-4.7143	-4.7163	-4.7421	-4.7411	-4.7474	-4.7270	-4.7264
13	-4.7365	-4.7384	-4.7400	-4.7358	-4.7377	-4.7406	-4.7300	-4.7406	-4.7203	-4.7174	-4.7443	-4.7405	-4.7177	-4.7200	-4.7258	-4.7497
14	-4.7383	-4.7402	-4.7420	-4.7379	-4.7400	-4.7396	-4.7425	-4.7430	-4.7448	-4.7184	-4.7197	-4.7427	-4.7459	-4.7234	-4.7236	-4.7259
15	-4.7399	-4.7408	-4.7411	-4.7412	-4.7406	-4.7426	-4.7426	-4.7452	-4.7473	-4.7211	-4.7204	-4.7405	-4.7417	-4.7499	-4.7424	-4.7292

all values multiplied by 10^3

Table C.2: Results from the $ARIMA(7, 0, 11)$ Model fitted to the differenced crude spot price series:

Parameter	Value	Standard Error	$t-$ Statistic
Constant	-0.0002	0.0034	-0.0725
AR(1)	0.7286	0.28254	2.5787
AR(2)	0.1615	0.2815	0.5737
AR(3)	0.0541	0.1576	0.3433
AR(4)	-0.1605	0.1569	-1.0225
AR(5)	-0.7468	0.1711	-4.3644
AR(6)	0.7983	0.213	3.7487
AR(7)	0.1019	0.2376	0.428
MA(1)	-0.7793	0.2834	-2.7501
MA(2)	-0.1498	0.2961	-0.506
MA(3)	0.0075	0.1645	0.0458
MA(4)	0.1658	0.1577	1.052
MA(5)	0.6659	0.1718	3.876
MA(6)	-0.7881	0.2099	-3.755
MA(7)	-0.1036	0.2461	-0.4209
MA(8)	0.0726	0.0226	3.212
MA(9)	0.0108	0.0316	0.3408
MA(10)	-0.0711	0.0227	-3.1339
MA(11)	0.0563	0.0186	3.0357
Variance	3.1344	0.049	63.9692

Appendix D

AICS AND BEST MODEL FOR GASOLINE FUTURES PRICES

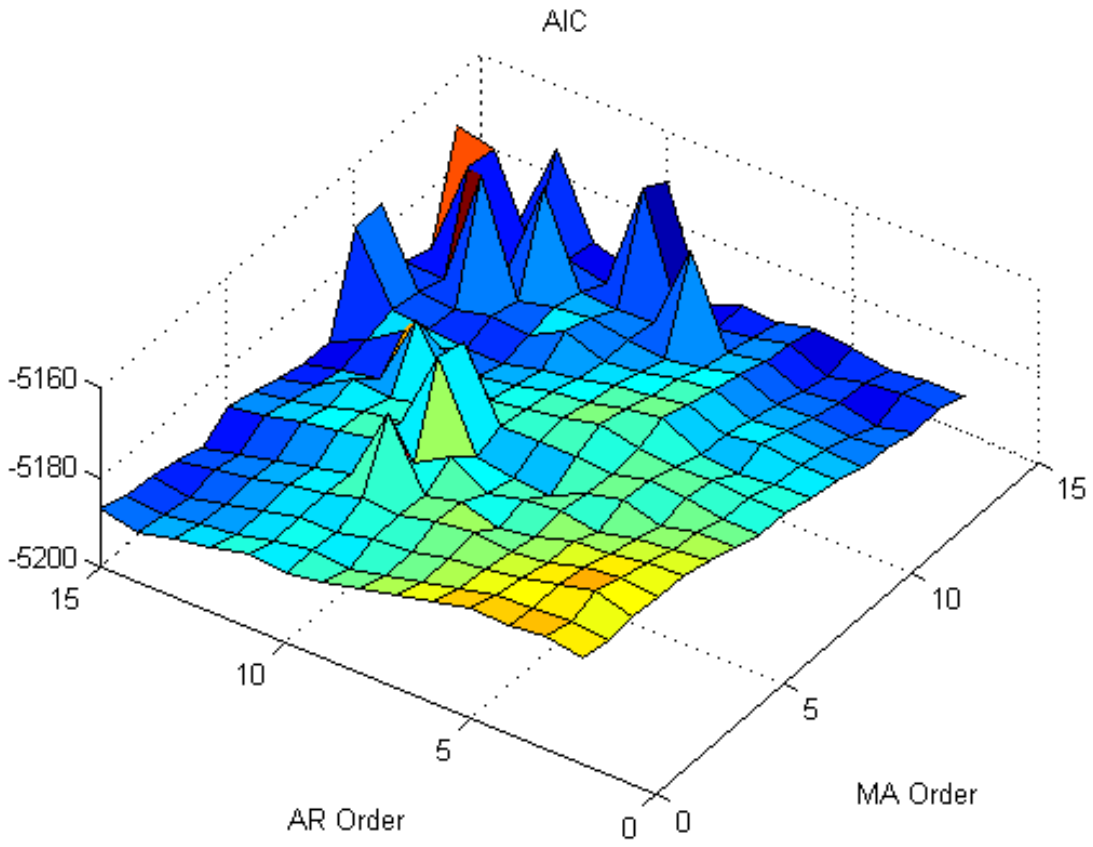

Figure D.1: Plot for the AICs for the various ARMA models fitted for the differenced gasoline futures price data

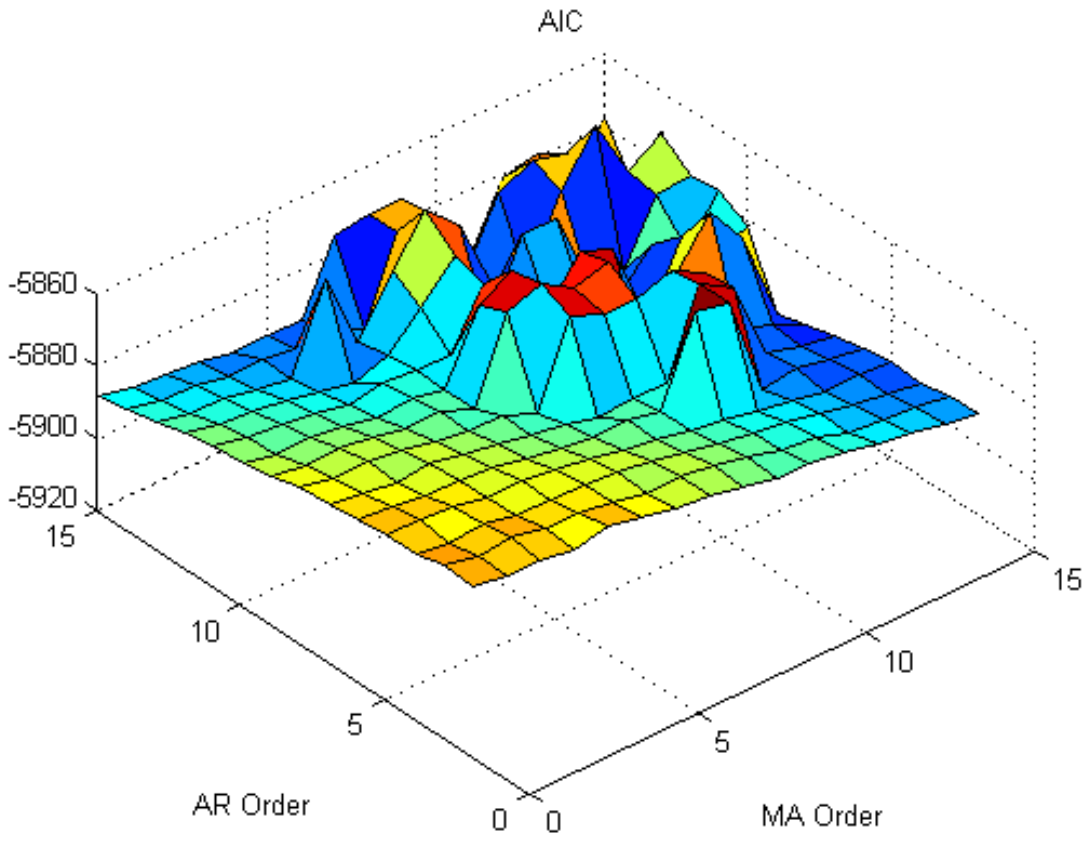

Figure D.2: Plot for the AICs for the various ARMA models fitted for the differenced gasoline spot price data

Table D.1: Results from the $ARIMA(12, 0, 12)$ Model fitted to the differenced crude spot price series:

Parameter	Value	Standard Error	$t-$ Statistic
Constant	0.0024	0.0239	0.0997
AR(1)	-0.1342	0.1512	-0.8873
AR(2)	-0.0482	0.1187	-0.4064
AR(3)	-0.0493	0.1054	-0.4675
AR(4)	0.1852	0.0716	2.588
AR(5)	-0.2060	0.0714	-2.8842
AR(6)	-0.0047	0.0724	-0.0656
AR(7)	0.0564	0.0755	0.747
AR(8)	-0.199	0.0722	-2.757
AR(9)	0.4822	0.0658	7.3252
AR(10)	0.5027	0.1059	4.7471
AR(11)	0.2889	0.1232	2.3454
AR(12)	-0.2534	0.1264	-2.0057
MA(1)	0.1055	0.1482	0.7119
MA(2)	-0.0072	0.1172	-0.061
MA(3)	0.0208	0.105	0.1979
MA(3)	-0.152	0.0721	-2.109
MA(5)	0.2303	0.0698	3.2991
MA(6)	0.0136	0.0751	0.1813
MA(7)	-0.0666	0.0771	-0.8648
MA(8)	0.198	0.0746	2.6551
MA(9)	-0.4718	0.0693	-6.8054
MA(10)	-0.4476	0.1075	-4.1639
MA(11)	-0.2526	0.1207	-2.0931
MA(12)	0.3474	0.1232	2.8196
Variance	4.5267	0.0869	52.1162

Appendix E

AIC VALUES FOR THE GARCH MODELS FITTED TO DIFFERENCED CRUDE FUTURES DATA

Table E.1: Results from the GARCH Models fitted to the differenced crude futures price series:

	ARCH						
GARCH	1	2	3	4	5	6	7
1.	8.8102	8.8102	8.8102	8.8102	8.8102	8.8102	8.8102
2.	8.8080	8.8080	8.8080	8.8089	8.8080	8.8080	8.8101
3.	8.8020	8.8020	8.8020	8.8020	8.8020	8.8020	8.8020
4.	8.7925	8.7922	8.7922	8.7913	8.7913	8.7927	8.7927
5.	8.7923	8.7912	8.7912	8.7912	8.7912	8.7912	8.7912
6.	8.7923	8.7912	8.7912	8.7912	8.7912	8.7877	8.7877
7.	8.7923	8.7891	8.7891	8.7863	8.7863	8.7874	8.7874
8.	8.7962	8.7874	8.7874	8.7874	8.7874	8.7901	8.7901
9.	8.7908	8.7910	8.7910	8.7930	8.7926	8.7896	8.7896
10.	8.7908	8.7909	8.7909	8.7929	8.7946	8.7856	8.7915
11.	8.7908	8.7909	8.7909	8.7855	8.7946	8.7856	8.7856
12.	8.7908	8.7836	8.7836	8.7877	8.7946	8.7856	8.7856
13.	8.7908	8.7836	8.7836	8.7855	8.7946	8.7856	8.7856
14.	8.7908	8.7854	8.7854	8.7874	8.7874	8.7849	8.7849
15.	8.7921	8.7837	8.7837	8.7837	8.7837	8.7801	8.7801
16.	8.7921	8.7837	8.7837	8.7837	8.7837	8.7801	8.7801
17.	8.7921	8.7837	8.7837	8.7837	8.7837	8.7801	8.7801
18.	8.7921	8.7837	8.7837	8.7837	8.7837	8.7849	8.7849
19.	8.7921	8.7837	8.7837	8.7837	8.7837	8.7801	8.7801
20.	8.7921	8.7837	8.7837	8.7837	8.7837	8.7801	8.7801

all values multiplied by 10^3

Table E.2: Results from the GARCH Models fitted to the differenced crude futures price series:

	ARCH					
GARCH	8	9	10	11	12	13
1.	8.8102	8.8102	8.8122	8.8122	8.8122	8.8122
2.	8.8101	8.8080	8.8105	8.8105	8.8105	8.8105
3.	8.8020	8.8020	8.8029	8.8029	8.8029	8.8029
4.	8.7941	8.7941	8.7959	8.7959	8.7911	8.7911
5.	8.7887	8.7887	8.7919	8.7919	8.7919	8.7919
6.	8.8009	8.7887	8.7866	8.7866	8.7866	8.7866
7.	8.7874	8.7874	8.7866	8.7866	8.7866	8.7866
8.	8.7917	8.7917	8.7901	8.7901	8.7901	8.7882
9.	8.7922	8.7922	8.7884	8.7884	8.7884	8.7901
10.	8.7848	8.7848	8.7855	8.7855	8.7855	8.7855
11.	8.7848	8.7848	8.7855	8.7855	8.7855	8.7855
12.	8.7848	8.7848	8.7871	8.7871	8.7871	8.7871
13.	8.7848	8.7848	8.7871	8.7871	8.7871	8.7871
14.	8.7835	8.7835	8.7854	8.7854	8.7854	8.7854
15.	8.7808	8.7808	8.7801	8.7801	8.7821	8.7800
16.	8.7808	8.7808	8.7801	8.7801	8.7821	8.7800
17.	8.7808	8.7808	8.7801	8.7801	8.7821	8.7800
18.	8.7835	8.7808	8.7801	8.7801	8.7821	8.7800
19.	8.7808	8.7808	8.7835	8.7801	8.7835	8.7855
20.	8.7855	8.7808	8.7835	8.7835	8.7835	8.7855

all values multiplied by 10^3

Table E.3: Results from the GARCH Models fitted to the differenced crude futures price series:

	ARCH						
GARCH	14	15	16	17	18	19	20
1.	8.8117	8.8198	8.8231	8.8265	8.8301	8.8202	8.8098
2.	8.8111	8.8111	8.8145	8.8141	8.8176	8.8191	8.8098
3.	8.8029	8.8049	8.8103	8.8143	8.8163	8.8163	8.8097
4.	8.7911	8.7931	8.7882	8.7882	8.7882	8.7882	8.7951
5.	8.7919	8.7919	8.7827	8.7827	8.7827	8.7827	8.7853
6.	8.7866	8.7866	8.7842	8.7842	8.7842	8.7842	8.7890
7.	8.7866	8.7866	8.7831	8.7831	8.7831	8.7831	8.7879
8.	8.7901	8.7901	8.7851	8.7851	8.7851	8.7851	8.7877
9.	8.7901	8.7901	8.7839	8.7839	8.7839	8.7839	8.7869
10.	8.7855	8.7855	8.7823	8.7823	8.7823	8.7823	8.7869
11.	8.7855	8.7855	8.7813	8.7813	8.7813	8.7813	8.7853
12.	8.7889	8.7909	8.7801	8.7801	8.7801	8.7801	8.7818
13.	8.7889	8.7909	8.7829	8.7829	8.7829	8.7829	8.7858
14.	8.7892	8.7892	8.7776	8.7776	8.7776	8.7776	8.7794
15.	8.7829	8.7847	8.7842	8.7862	8.7862	8.7862	8.7794
16.	8.7829	8.7847	8.7862	8.7862	8.7862	8.7862	8.7821
17.	8.7829	8.7847	8.7789	8.7789	8.7789	8.7789	8.7821
18.	8.7829	8.7847	8.7785	8.7785	8.7785	8.7785	8.7807
19.	8.7837	8.7852	**8.7764**	8.7764	8.7764	8.7764	8.7799
20.	8.7837	8.7852	8.7764	8.7764	8.7764	8.7764	8.7789

all values multiplied by 10^3

Appendix F

GARCH MODEL SELECTION CODE IN MATLAB

```
Ddata=diff(data,1);
% Lag values to use in model selection
n=20;
options = optimset('fmincon');
options.MaxIter = 1000;
options.Display='none';
AICs = zeros(n,n);
for i=1:n;
for j=1:n;
Mdl = garch(i,j);
[EstMdl,EstParamCov,logL,info] = estimate(Mdl,Ddata(:,1));
params=length(info.X); %number of parameters
[aic] = aicbic(logL,params);
AICs(i,j)=aic;
end
AICs(i,j)= AICs(i,j)+1;
end
```

```
%% Fit the specific GARCH model
Model = garch(19,16)
[EstModel,EstParamCov,logL,info] = estimate(Model,Ddata(:,1));
params=length(info.X); %number of parameters
[aic] = aicbic(logL,params);

%% Residual Analysis
V=infer(EstModel,Ddata(:,1));% infer conditional variances
T=length(Ddata(:,1));
figure
subplot(2,1,1)
plot(V)
grid on
xlim([0,T])
title('Infered Conditional variances')
% computing standardized residuals
res = (Ddata(:,1)-EstModel.Offset)./sqrt(V);
title('Standardized Residuals')
subplot(2,1,2)
histfit(res,40)
title('Standardized Residuals')
figure
subplot(2,1,1)
qqplot(res) % testing normality of residuals
subplot(2,1,2)
boxplot(res)
figure
subplot(2,1,1)
autocorr(res,40)
```

```
ylim([-0.1 0.1])

subplot(2,1,2)

parcorr(res,40)

ylim([-0.1 0.1])
```

Appendix G

$GARCH(1,1)$ RESIDUAL ANALYSIS

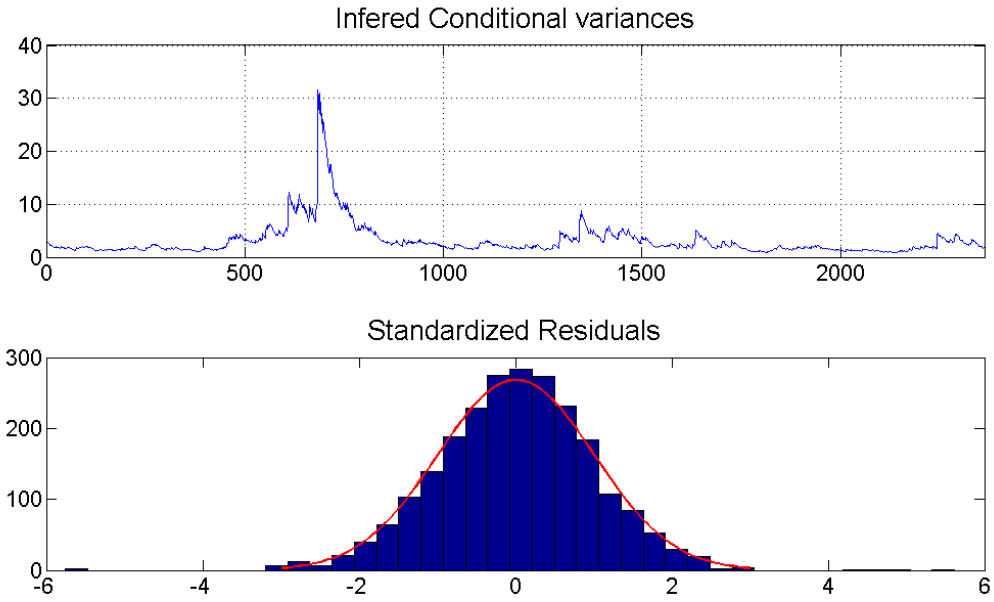

Figure G.1: $GARCH(1,1)$ Conditional variances plot and standardized residuals histogram

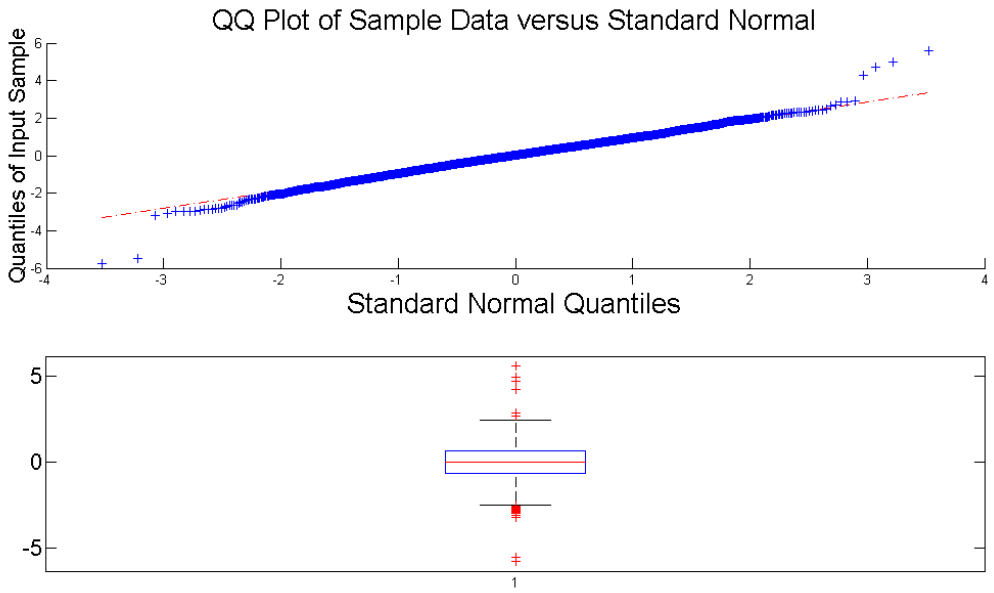

Figure G.2: GARCH(1,1) Standardized residuals QQ-plot and boxplot

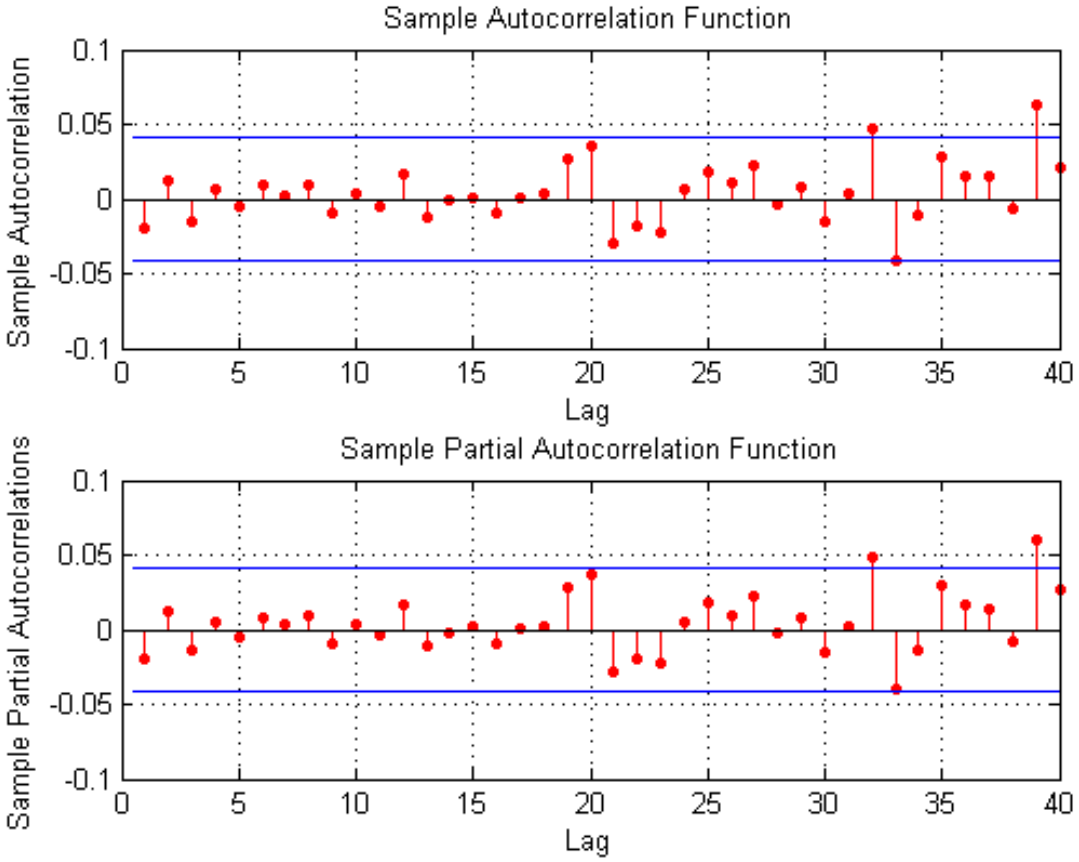

Figure G.3: GARCH(1,1) ACF and PACF of standardized residuals

Appendix H

GARCH MODEL FORECASTING MATLAB CODE

```
%% Fit the GARCH(19,16) model
Model = garch(19,16)
[EstModel,EstParamCov,logL,info] = estimate(Model,Ddata(:,1));
params=length(info.X); %number of parameters
[aic] = aicbic(logL,params);
%% extract alphas
alphas=[];
 for i=1:1:length(EstModel.ARCH);
     alpha=cell2mat(EstModel.ARCH(i));
     alpha(i)=alpha;
     alphas=[alphas alpha(i)]
 end
%% extract betas
 betas=[];
 for j=1:1:length(EstModel.GARCH);
     beta=cell2mat(EstModel.GARCH(j));
     beta(j)=beta;
     betas=[betas beta(j)]
 end
%% Fitting model to forecast differenced series
fut_ret=zeros(300,1);% vector for storing recursive values of returns
fut_var=zeros(300,1)% vector for storing recursive values of variance
```

```
fut_at=zeros(300,1)% vector for storing recursive values of innovations
ret=Ddata(1:end-300,1);
F_ret=[ret;fut_ret];
var=v;
F_var=[var;fut_var]% vector storing forcasted variances
Inn=Ddata(1:end-300,1);
F_at=[Inn;fut_at] % vector storing forecasted innovations
for t=2063:2362
    F_var(t)=EstModel.Constant+alphas(1)*F_at(t-1)^2+alphas(2)*F_at(t-2)^2+...
        alphas(5)*F_at(t-5)^2+alphas(6)*F_at(t-6)^2+alphas(8)*F_at(t-8)^2+...
        alphas(9)*F_at(t-9)^2+alphas(10)*F_at(t-10)^2+alphas(13)*F_at(t-13)^2+...
        alphas(14)*F_at(t-14)^2+alphas(15)*F_at(t-15)^2+alphas(16)*F_at(t-16)^2+...
        betas(16)*F_var(t-16)+ betas(19)*F_var(t-19);
    F_at(t)=sqrt(F_var(t))*randn;
    F_ret(t)=F_at(t);
end
F_ret
F_var
%% plot forecasted differenced series and variances from the GARCH(19,16)
 figure
 subplot(2,1,1)
 plot(dates(2:end-300),Ddata(1:end-300,1))
 hold on
 plot(dates(end-300:end-1),F_ret(end-300:end-1),'r')
 hold off
 datetick('x','yyyy','keeplimits')
 grid on
subplot(2,1,2)
 plot(dates(2:end-300),F_var(1:end-300,1))
```

```
 hold on
 plot(dates(end-300:end-1),F_var(end-300:end-1),'r')
 hold off
 datetick('x','yyyy','keeplimits')
 grid on
 axis tight
 %% forecast crude futures prices using GARCH (19,16)
 fut_price=zeros(length(Ddata(end-300:end)),1)
 CF_F=[data(1:2062,1);fut_price]
 for t=2063:2362
     CF_F(t)=F_ret(t)+CF_F(t-1);
 end
 figure
 plot(dates,data(:,1))
 hold on
 plot(dates(end-300:end),CF_F(end-300:end),'r')
 hold off
 datetick('x','yyyy','keeplimits')
 grid on
 axis tight
 %% Measuring forecasting ability
new_F_var=F_var(2063:end);
Sigma_Rt=sqrt(1/length(new_F_var)*sum(new_F_var)); % ralised volatility
dev=zeros(300,1);
for i=1:length(new_F_var);
    dev(i)=sqrt(new_F_var(i))-Sigma_Rt;
end
RMSFE=sqrt(1/length(new_F_var)*sum(dev.^2));
```

Appendix I

ENGLE-GRANGER TEST FOR CO-INTEGRATION MATLAB CODE

```
%% Import data from spreadsheet
% Script for importing data from the following spreadsheet:
%
%    Workbook: C:\Users\aduda\Dropbox\2015\Main\analysis\DATA.xlsx
%    Worksheet: DATA
%
% To extend the code for use with different selected data or a different
% spreadsheet, generate a function instead of a script.


% Auto-generated by MATLAB on 2015/02/03 19:47:16

%% Import the data, extracting spreadsheet dates in MATLAB serial date number format
[~, ~, raw, dateNums] = xlsread('C:\Users\user.IDEA-PC\Dropbox\2015\Main\analysis\time

%% Replace date strings by MATLAB serial date numbers (datenum)
R = ~cellfun(@isequalwithequalnans,dateNums,raw) & cellfun('isclass',raw,'char'); % Fi
raw(R) = dateNums(R);

%% Create output variable
DATA = reshape([raw{:}],size(raw));
```

```
%% Clear temporary variables
clearvars raw dateNums R;

%% create a financial time series object and plot a series for the entire
%data set
format short
dates = DATA(:,1);
data = DATA(:,[2:15]);
freq = 1;
tsobj= fints(dates, data,{'CF','CS','GF','GS','HF','HS','A_crkF','B_crkF',...
'C_crkF','D_crkF','A_crkS','B_crkS','C_crkS','D_crkS'}, freq);

figure(1)
plot(tsobj)
%% create a financial time series object and plot a series of the raw data
%for the spot and futures markets
format short
dates = DATA(:,1);
raw = DATA(:,[2:7]);
freq = 1;
raw_tsobj = fints(dates, raw,{'CF','CS','GF','GS','HF','HS'}, freq);
figure(1)
plot(raw_tsobj)
grid on
%% create a financial time series object and plot a series of the crack
%spread for the spot and futures markets
format short
dates = DATA(:,1);
spread = DATA(:,[8:15]);
```

```
freq = 1;
spread_tsobj = fints(dates, spread,{'A_crkF','B_crkF','C_crkF','D_crkF',...
'A_crkS','B_crkS','C_crkS','D_crkS'}, freq);
figure(1)
plot(spread_tsobj)
grid on
%% Convert object to matrix and extract each series. This is necesarry for
%the individual plots. This is for the raw series
CF=fts2mat(raw_tsobj.CF);
CS=fts2mat(raw_tsobj.CS);
GF=fts2mat(raw_tsobj.GF);
GS=fts2mat(raw_tsobj.GS);
HF=fts2mat(raw_tsobj.HF);
HS=fts2mat(raw_tsobj.HS);
%%Convert object to matrix and extract each series. This is necesarry for
%the individual plots. This is for the the 3:2:1
%crack spread series
crkF=fts2mat(spread_tsobj.C_crkF);
crkS=fts2mat(spread_tsobj.C_crkS);
%% choosing optimal lag length for cointegrated model
%Selection of lag length is challenging to automate.  Instead, this code
% estimates all cointegration regression models with p=0,1,...,4 and then
% looks at AICs.
% P = 15;
% AICs = zeros;
% for i=0:P;
%         [~,~,~,~,reg1,~] = egcitest(spread(:,[3 7]),'lags',i);
%         AICs(i+1) = reg1.AIC
% end
```

```
%% Engle-Granger test for cointegration on the raw price series
[h,pValue,stat,cValue,reg1,reg2] = egcitest(raw,'test','t2');
h
pValue
%%
a = reg1.coeff(1);
b = reg1.coeff(2:6);
beta=[1;-b];
%dates = datestr(dates)
plot(dates,raw*beta-a)
datetick('x','yyyy','keeplimits')
grid on
axis tight
%% Test if the residuals are stationary
[h,pValue] = adftest(raw*beta-a);

%% Multiple Cointegrating Relations
% Permutations of the data variables:
P0 = perms([1 2 3 4 5 6]);
[~,idx] = unique(P0(:,1)); % Rows of P0 with unique regressand y1
P = P0(idx,:); % Unique regressions
numPerms = size(P,1);
Y = raw;
% Preallocate:
T0 = size(Y,1);
HEG = zeros(1,numPerms);
PValEG = zeros(1,numPerms);
CIR = zeros(T0,numPerms);
```

```
%% Run all tests:
for i = 1:numPerms

YPerm = Y(:,P(i,:));
[h,pVal,~,~,reg] = egcitest(YPerm,'test','t2');
HEG(i) = h;
PValEG(i) = pVal;
c0i = reg.coeff(1);
bi = reg.coeff(2:6);
CIR(:,i) = YPerm*[1;-bi]-c0i;

end

fprintf('\n=== Different Engle-Granger tests, same data ===\n\n')
HEG,PValEG

% Plot the cointegrating relations:
figure
C = get(gca,'ColorOrder');
set(gca,'NextPlot','ReplaceChildren','ColorOrder',circshift(C,3))
plot(dates,CIR,'LineWidth',1)
title('{\bf Multiple Cointegrating Relations}')
legend(strcat({'Cointegrating relation  '}, ...
num2str((1:numPerms)')),'location','NW');
datetick('x','yyyy','keeplimits')
axis tight
grid on
```

Appendix J

JOHANSEN'S TEST FOR CO-INTEGRATION MATLAB CODE

```
clear all
close all
clc
%% Import data from spreadsheet
% Script for importing data from the following spreadsheet:
%
%    Workbook: C:\Users\aduda\Dropbox\2015\Main\analysis\DATA.xlsx
%    Worksheet: DATA
%
% To extend the code for use with different selected data or a different
% spreadsheet, generate a function instead of a script.


% Auto-generated by MATLAB on 2015/02/03 19:47:16


%% Import the data, extracting spreadsheet dates in MATLAB serial date number format
[~, ~, raw, dateNums] = xlsread('C:\Users\user.IDEA-PC\Dropbox\2015\Main\analysis\time


%% Replace date strings by MATLAB serial date numbers (datenum)
R = ~cellfun(@isequalwithequalnans,dateNums,raw) & cellfun('isclass',raw,'char'); % Fi
raw(R) = dateNums(R);
```

```
%% Create output variable
DATA = reshape([raw{:}],size(raw));

%% Clear temporary variables
clearvars raw dateNums R;

%% create a financial time series object and plot a series for the entire
% data set
format short
dates = DATA(:,1);
data = DATA(:,[2:15]);
freq = 1;
tsobj= fints(dates, data,{'CF','CS','GF','GS','HF','HS','A_crkF','B_crkF',...
'C_crkF','D_crkF','A_crkS','B_crkS','C_crkS','D_crkS'}, freq);

figure(1)
plot(tsobj)
%% create a financial time series object and plot a series of the raw data
% for the spot and futures markets
format short
dates = DATA(:,1);
raw = DATA(:,[2:7]);
freq = 1;
raw_tsobj = fints(dates, raw,{'CF','CS','GF','GS','HF','HS'}, freq);
figure(1)
plot(raw_tsobj)
grid on
%% create a financial time series object and plot a series of the crack
%spread for the spot and futures markets
```

```
format short
dates = DATA(:,1);
spread = DATA(:,[8:15]);
freq = 1;
spread_tsobj = fints(dates, spread,{'A_crkF','B_crkF','C_crkF','D_crkF',...
'A_crkS','B_crkS','C_crkS','D_crkS'}, freq);
figure(1)
plot(spread_tsobj)
grid on
%% Convert object to matrix and extract each series. This is necesarry for
%the individual plots. This is for the raw series
CF=fts2mat(raw_tsobj.CF);
CS=fts2mat(raw_tsobj.CS);
GF=fts2mat(raw_tsobj.GF);
GS=fts2mat(raw_tsobj.GS);
HF=fts2mat(raw_tsobj.HF);
HS=fts2mat(raw_tsobj.HS);
%%Convert object to matrix and extract each series. This is necesarry for
%the individual plots. This is for the the 3:2:1
%crack spread series
crkF=fts2mat(spread_tsobj.C_crkF);
crkS=fts2mat(spread_tsobj.C_crkS);
%% Johansen trace test for cointegration on the raw price series
[h,pV,stat,cV,mles] = jcitest(raw,'model','H1','lags',4);
h
pV
%%
Y=raw;
YLag = Y(4:end,:);
```

```
T = size(YLag,1);

B = mles.r4.paramVals.B;

c0 = mles.r4.paramVals.c0;

plot(dates(4:end),YLag*B+repmat(c0',T,1))

title('{\bf Cointegrating Relations from Johansens co-integration test}')

legend(strcat({'Cointegrating relation  '}, ...

num2str((1:4)')),'location','NE');

datetick('x','yyyy','keeplimits')

grid on

axis tight
```