APPLICATION OF ARIMA AND GARCH MODELS IN FORECASTING CRUDE OIL PRICES

LEE CHEE NIAN

UNIVERSITI TEKNOLOGI MALAYSIA

APPLICATION OF ARIMA AND GARCH MODELS IN FORECASTING CRUDE OIL PRICES

LEE CHEE NIAN

A dissertation submitted in partial fulfillment of the
requirement for the award of the degree of
Master of Science (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

NOVEMBER 2009

Specially dedicated to

my beloved parents, brother, sisters

and

those people who have guided and inspired me throughout my journey of education

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my heartiest gratitude to everyone who involved directly or indirectly in contributing to this completed study. I am grateful to all the supports given to me.

This particular research would never be able to accomplish without the support of my beloved supervisor, Associate Professor Dr. Maizah Hura Ahmad who is extremely knowledgeable about the time series. She has spent her valuable time giving advice, shared her experience with me and assisted me all the way long. I am truly grateful for having such a wonderful supervisor.

Lastly, I would like to thank to my family who always supports me and also thank to all my friends for their assistance to enable the completion of this study.

ABSTRACT

Crude oil is an important energy commodity to mankind. Several causes have made crude oil prices to be volatile. The fluctuation of crude oil prices has affected many related sectors and stock market indices. Hence, forecasting the crude oil prices is essential to avoid the future prices of the non-renewable natural resources to raise sky-rocket. In this study, daily WTI crude oil prices data is obtained from Energy Information Administration (EIA) from 2nd January 1986 to 30th September 2009. We use the Box-Jenkins methodology and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) approach in forecasting the crude oil prices. An Autoregressive Integrated Moving Average (ARIMA) model is set as the benchmark model. We found ARIMA(1,2,1) and GARCH(1,1) are the appropriate models under model identification, parameter estimation, diagnostic checking and forecasting future prices. In this study, the analyses are done with the aid of EViews software where the potential of this software in forecasting daily crude oil prices time series data is explored. Finally, using several measures, comparison performances between ARIMA(1,2,1) and GARCH(1,1) models are made. GARCH(1,1) is found to be a better model than ARIMA(1,2,1) model. Based on the study, we conclude that ARIMA(1,2,1) model is able to produce accurate forecast based on a description of history patterns in crude oil prices. However, the GARCH(1,1) is the better model for daily crude oil prices due to its ability to capture the volatility by the non-constant of conditional variance.

ABSTRAK

Minyak mentah merupakan komoditi tenaga yang penting untuk umat manusia. Beberapa penyebab telah memjadikan harga minyak mentah akan berubah-ubah. Fluktuasi harga minyak mentah telah mempengaruhi pelbagai sektor berkaitan serta indeks pasaran saham. Oleh sebab itu, ramalan kepada harga minyak mentah adalah agak penting untuk mengelakkan harga masa depan sumber alam yang tidak diperbaharui daripada meningkatkan mendedak. Dalam kajian ini, harga minyak mentah harian WTI data yang diperolehi daripada Energy Information Administration (EIA) dari 2 Januari 1986 sampai ke 30 September 2009. Kami menggunakan metodologi Box-Jenkins dan pendekatan Generalized Autoregressive Conditional Heteroscedasticity (GARCH) dalam meramalkan harga minyak mentah. Sebuah model Autoregressive Integrated Moving Average (ARIMA) ditetapkan sebagai model patokan. Kami menemukan ARIMA(1,2,1) dan GARCH(1,1) adalah model yang sesuai di bawah pengenalan model, estimasi parameter, diagnostik pemeriksaan dan peramalan harga masa depan. Dalam kajian ini, analisis yang dilakukan dengan bantuan perisian EViews di mana potensi perisian ini akan dieksplorasi dalam memprediksi harga minyak mentah harian data siri masa. Akhirnya, dengan menggunakan beberapa ukuran, perbandingan prestasi di antara ARIMA(1,2,1) dan GARCH(1,1) model diuji. GARCH(1,1) ditemui untuk menjadi model yang lebih baik daripada model ARIMA(1,2,1). Mengikuti kajian ini, kami membuat kesimpulan bahawa model ARIMA(1,2,1) boleh menghasilkan perkiraan yang tepat berdasarkan keterangan pola-pola dalam sejarah harga minyak mentah. Namun, GARCH(1,1) adalah model yang lebih baik untuk harga minyak mentah harian kerana kemampuannya untuk menangkap volatilitas oleh pemalar bukan varians bersyarat.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACF	-	Autocorrelation functions
ADF	-	Augmented Dickey-Fuller
AIC	-	Akaike Information Criterion
ANFIS	-	Adaptive Network-based Fuzzy Inference System
ANN	-	Artificial Neural Networks
API	-	American Petroleum Institute
AR	-	Autoregression
ARCH	-	Autoregressive Conditional Heteroscedasticity
ARIMA	-	Autoregressive Integrated Moving Average
ARMA	-	Autoregressive Moving Average
CBP	-	Correlated Bivariate Poisson
CGARCH	-	Component GARCH
DW	-	Durbin-Watson
EGARCH	-	Exponential GARCH
EIA	-	Energy Information Administration
EViews	-	Econometric Views
EVT	-	Extreme Value Theory
FIAPARCH	-	Fractional Integrated Asymmetric Power ARCH
FIGARCH	-	Fractionally Integrated GARCH
GARCH	-	Generalized Autoregressive Conditional Heteroscedasticity
GED	-	Generalized Exponential distribution
GUI	-	Graphical User Interface
HSAF	-	Historical Simulation with ARMA Forecasts
HT	-	Heavy-tailed
IGARCH	-	Integrated GARCH
ILS	-	Interval Least Square

IPE	-	International Petroleum Exchange
IV	-	Implied Volatility
JB	-	Jarque-Bera
LM	-	Lagrange Multiplier
MA	-	Moving Average
MAE	-	Mean Absolute Error
MAPE	-	Mean Absolute Percentage Error
MRS	-	Markov Regime Switching
MSFE	-	Mean Squared Forecast Error
NYMEX	-	New York Mercantile Exchange
OPEC	-	Organization of the Petroleum Exporting Countries
PACF	-	Partial Autocorrelation Functions
PP	-	Phillips-Perron
QMS	-	Quantitative Micro Software
RMSE	-	Root Mean Squared Error
SIC	-	Schwarz Information Criterion
SVM	-	Support Vector Machine
TAR	-	Asymmetric Threshold Autoregressive
Theil-U	-	Theil Inequality Coefficient
US	-	United State
VaR	-	Value at Risk
VECM	-	Vector Error Correction Model
WTI	-	West Texas Intermediate
2SLS	-	Two-stage Least Squares

LIST OF SYMBOLS

$\bar{R}^2$	-	adjusted R-squared
$\hat{s}_t$	-	standardized residuals
$\hat{\varepsilon}$	-	estimated residual
$\hat{\varepsilon}^2$	-	sum-of-squared residuals
ε_t	-	residuals
ε_t^2	-	residuals squared
ξ_t	-	white noise process
Ω_{t-1}	-	measurable function of time $t-1$ information set
H_0	-	null hypothesis
R^2	-	R-squared
f_0	-	estimator of the residual spectrum at frequency zero
l_t	-	likelihood of ε_t
v_t	-	residual of time series
x_t	-	optional exogenous regressors
$\bar{y}$	-	mean of the dependent variable
y_t	-	differenced of crude oil prices time series
z_t	-	time series of crude oil prices
α_i	-	coefficients for ARCH
γ_0	-	consistent estimate of the error variance
ρ_k	-	autocorrelation
$\hat{\sigma}$	-	estimator for the standard deviation
σ^2	-	unconditional variance
σ_t^2	-	conditional variance
χ^2	-	Chi-squared
ϕ_k	-	partial autocorrelation
Δ	-	difference linear operator

B	-	backshift operator
F	-	F-statistic
L	-	likelihood of the joint realizations
Q	-	Q-statistic
X	-	$n \times k$ matrix of independent variables
d	-	amount of differencing
k	-	number of regressors
l	-	log likelihood
n	-	number of observations
p	-	order of the autoregressive part
q	-	order of the moving average part
s	-	standard error of the regression
t	-	time
y	-	n-dimensional vector of dependent variable
β	-	k-vector of coefficients
ε	-	n-vector of disturbances

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Crude oil or petroleum is a naturally occurring and flammable liquid found in rock formations in the earth. It has consisting of a complex mixture of hydrocarbons of various molecular weights plus other organic compounds.

The main characteristics of crude oil are generally classifies according to its sulphur content and its density which the petroleum industry measured by its American Petroleum Institute (API) gravity. Obviously, crude oil may be considered light if it has low density with API gravity less than about 40. Typically, heavy crude has high density with API gravity 20 or less. In other words, the higher the API gravity, the lower in its density. Brent crude is important benchmark crude which has an API gravity of 38 to 39. Crude oil may be referred to as sweet if it contains less than 0.5% sulphur or sour if it contains substantial amounts of sulphur. Sweet crude is preferable to sour one because it is more suited to the production of the most valuable refined products.

Moreover, the geographical location of crude oil production is another main count. In the crude oil market, the two current references or pricing markers are West Texas Intermediate (WTI) and Europe Brent. The former is the base grade traded, as 'light sweet crude', on the New York Mercantile Exchange (NYMEX) for delivery at Cushing, Oklahoma. While the latter is traded on London's International Petroleum

Exchange (IPE) for delivery at Sullom Voe and is also one of the grades acceptable for delivery of the NYMEX contract (Lin and Tamvakis, 2001).

The price of a barrel of oil is highly dependent on both its grade, determined by factors such as its specific API gravity, sulphur content and also location of production. The vast majority of oil is not traded on an exchange but on an over-the-counter basis. Some other important benchmarks include Dubai, Tapis (Malaysia), Minas (Indonesia) and Organization of the Petroleum Exporting Countries (OPEC) basket. The Energy Information Administration (EIA) uses the imported refiner acquisition cost where the weighted average cost of all oil imported into the United State (US) known as "world oil price".

Look back into the past, the increasing oil prices has affected certain benchmark indices widely followed and traded. On the other hand, the scientific community is confused over the absolute quantities of oil reserves. In fact, crude oil is a limited and non-renewable natural reserve. The on going demand of crude oil and its refined products will consequently in oil supply scarcity. In the end, this energy commodity is most likely to keep an upward trend in the future if without any alternative replacements for crude oil.

There are ample studies addressing the accuracy of crude oil volatility modelling and forecasting. These include Autoregressive Conditional Heteroscedasticity, ARCH-type models (Fong and See, 2002; Giot and Laurent, 2003), Asymmetric threshold autoregressive (TAR) model (Godby *et al.*, 2000), and artificial based forecast methods (Fan *et al.*, 2008a), Interval Least Square (ILS) (Xu *et al.*, 2008), Support Vector Machine (SVM) (Xie *et al.*, 2006), Artificial Neural Networks (ANN) (Kulkarni and Haidar, 2009), Adaptive Network-based Fuzzy Inference System (ANFIS) (Ghaffari and Zare, 2009), Fuzzy Neural Network (Liu *et al.*, 2007), Autoregressive Moving Average (ARMA) (Cabedo and Moya, 2003), and etc. However, the complexity of the model specification does not guarantee high performance on out-performed out-of-sample forecasts.

One of the model that has gained enormous popularity in many areas and forecasting research practice is Box-Jenkins method. Thus, the purpose of this study is to forecast crude oil prices using Box-Jenkins method. However, despite the fact that the Box-Jenkins method is powerful and flexible, it is not able to handle the volatility that is present in the data series. To handle the volatility in the crude oil data, the current study proposes the use of the Generalized ARCH (GARCH) model. Using the forecasts obtained from the Box-Jenkins model as a benchmark, the forecasts obtained from the GARCH will be evaluated.

1.2 Background of the Study

In statistics, a sequence of random variables is heteroscedastic if the random variables have different variances. The term means "differing variance" and comes from the Greek "hetero" ('different') and "skedasis" ('dispersion'). In contrast, a sequence of random variables is called homoscedastic if it has constant variance.

In particular, we consider crude oil prices data as heteroscedastic time series models where the conditional variance given in the past is no longer constant(Palma, 2007). In a financial analysis, forecast of future volatility of a series under consideration are often of interest to assess the risk associated with certain assets. In that case, variance forecasts are of direct interest, of course (Lütkepohl, 2005).

One of the most prominent stylized facts of returns on crude oil prices is that their volatility changes over time. In particular, periods of large movements in crude oil prices alternate with periods during which prices hardly change. This characteristic feature commonly is referred to as volatility clustering.

It was first observed by Nobel Prize winner, Robert Engle (1982) that although many financial time series, such as, stock returns and exchange rates are unpredictable, there is apparent clustering in the variability or volatility. This is often referred to as conditional heteroscedasticity since it is assumed that overall the series is stationary but the conditional expected value of the variance may be time-dependent.

Later, Bollerslev (1986) had modified Engle's ARCH model into a more generalized model called GARCH model with is a simplified model to ARCH model but more powerful. Currently, this model has been widely used in many financial time series data. The simple GARCH model is able to detect the financial volatility in a time trend.

1.3 Statement of the Problem

The price of the energy commodity is highly volatile throughout the time. Since crude oil prices variability does affect other sectors and stock market, the prediction of future crude oil prices becomes crucial.

This study will explore the following question :

"Which method between the Box-Jenkins and GARCH performs better in forecasting crude oil prices, which is of high volatility?"

1.4 Objective of the Study

The objectives of this study are as follows:

1.4.1 To estimate suitable Box-Jenkins and GARCH models for forecasting crude oil prices.

1.4.2 To evaluate the performance of the GARCH and Box-Jenkins models in forecasting crude oil prices.

1.4.3 To forecast using EViews software.

1.5 Scope of the Study

This study focuses on the Box-Jenkins and GARCH models to forecast crude oil prices. Since the oil price volatility is the main concern, the study uses only daily data. The data were obtained from EIA from 2nd January 1986 to 30th September 2009.

1.6 Significance of the Study

Since crude oil market is highly volatile, the estimation of the time series model must be able to detect its volatility. We have to determine the precisely Box-Jenkins and GARCH models when forecasting the volatility of crude oil prices. The process will be done with the aid of software. As a result of this study, a model and software that can be used to forecast volatile time series can be proposed.

1.7 Summary and Outline of the Study

This dissertation is organized into 5 chapters. Chapter 1 discusses the research framework. It begins with the introduction to crude oil and the background of the study. The objectives, scope and the significance of this study are also presented.

Chapter 2 reviews crude oil prices in forecasting. First, crude oil prices will be reviewed. Then, the volatility in crude oil prices will be discussed. The discussions start on the past researchers' work in Box-Jenkins methodology and GARCH-type models are also presented. Finally, conditional heteroscedasticity are explained.

Chapter 3 begins methodology. In this chapter analysis of data sets using the Autoregressive Integrated Moving Average (ARIMA) and GARCH models are carried out.

In chapter 4, a detail present on the analysis of the same data sets using the ARIMA and GARCH models. Also, comparison between the ARIMA and GARCH models are made.

Chapter 5 summarizes and concludes the whole study and makes some suggestions for future investigation.