MODELLING OF CRUDE OIL PRICES USING HYBRID ARIMA-GARCH MODEL

NAPISHAH BINTI HASHIM

UNIVERSITI TEKNOLOGI MALAYSIA

MODELLING OF CRUDE OIL PRICES USING HYBRID ARIMA-GARCH MODEL

NAPISHAH BINTI HASHIM

A dissertation submitted in partial fulfillment of the

requirements for the award of the degree of

Master of Science (Mathematics)

Faculty of Science

Universiti Teknologi Malaysia

JUNE 2015

To my beloved parents, Hashim bin Jusoh and Maryam binti Jusoh, all my siblings, Mansor, Fadzilah, Fauziah and him, Muhamad Shaifful bin Zahari. Thank you for your support, love and encouragement.

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere thanks to my respected supervisor, Assoc. Prof. Dr. Maizah Hura binti Ahmad for her necessary guidance, patience support, helpful comments, corrections and suggestions. She inspired me greatly to complete this project. Her willingness to motivate me contributed tremendously to my thesis. I consider myself very fortunate for being able to work with a very considerate and encouraging supervisor like her. I really appreciate all the knowledge and ideas given to me. Without her offering to accomplish this work, I would not be able to implement this project.

Besides, I would like to express my gratitude toward the authority of Universiti Teknologi Malaysia in providing a good facility which is Perpustakaan Sultanah Zanariah (PSZ) that consists of useful book and journal. This project gives me an opportunity to learn about more topics in statistics related to Time Series.

I am also grateful to my respected internal examiner, Dr. Ani bin Shabri who gives me a chance to correct my thesis and presentation. His suggestions and comments has greatly improved the manuscript.

Moreover, an honourable mention goes to my parent, Hashim and Maryam and my family for their understandings, prayers and supports in completion of this wonderful project. They gave me motivation throughout the writing of my dissertation.

Finally, I would like to express my sincere appreciations to my friends, especially Muhamad Shaifful and those involved directly or indirectly in encouraging me towards the realization of this thesis. Without helps of the particular that mentioned above, I would face many difficulties while doing this project.

ABSTRACT

Modelling of volatile data has become the area of interest in financial time series recently. Volatility refers to the phenomenon where the conditional variance of the time series varies over time. The objective of this study is to compare the modelling performance of Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and hybrid ARIMA-GARCH model for the prices of crude oil. Eviews and Minitab software are used to analyze the data. The models investigated are GARCH and hybrid ARIMA-GARCH model. In parameter estimation, Maximum Likelihood Estimation (MLE) is the preferred technique for GARCH models while Ordinary Least Squares Estimation (OLS) and MLE will be used for hybrid ARIMA-GARCH models. The goodness of fit of the model is measured using Akaike's Information Criterion (AIC). The diagnostic checking is conducted to validate the goodness of fit of the model using Jarque-Bera test, Serial Correlation test and Heteroskedasticity test. Forecasting accuracies for both models are assessed using Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The model which gives the lowest measure of error is considered to be the most appropriate model. Empirical results indicate that modelling using hybrid model has smaller AIC, MAE and MAPE values compared to GARCH model. It can be concluded that hybrid ARIMA-GARCH model is better in modelling crude oil prices data compared to GARCH model.

ABSTRAK

Permodelan data tidak menentu merupakan bidang yang penting dalam siri masa kewangan sejak kebelakangan ini. Turun naik merupakan fenomena varians bersyarat siri masa yang berbeza pada setiap masa. Objektif kajian ini adalah untuk membandingkan prestasi permodelan di antara model Heteroskedastisiti Bersyarat Autoregresif Teritlak (GARCH) dengan model hybrid ARIMA-GARCH menggunakan data harga minyak mentah. Perisian Eviews dan Minitab digunakan untuk menganalisis data. Model yang digunakan adalah model GARCH dan hibrid ARIMA-GARCH. Untuk menganggar parameter, Anggaran Kemungkinan Maksimum (MLE) adalah kaedah yang dipilih untuk model GARCH manakala Anggaran Kuasa Dua Terkecil Biasa (OLS) dan MLE akan digunakan untuk model GARCH. Kesesuaian model dikira menggunakan Kriteria Maklumat Akaike (AIC). Semakan diagnostik dijalankan untuk menyemak kesesuaian model dengan menggunakan ujian Jarque-Bera, ujian Korelasi Bersiri dan ujian Heteroskedastisiti. Ketepatan ramalan untuk kedua-dua model ini dinilai menggunakan Min Ralat Mutlak (MAE) dan Min Peratus Ralat Mutlak (MAPE). Model yang mempunyai nilai ralat yang paling rendah dianggap sebagai model yang paling sesuai. Hasil kajian menunjukkan bahawa permodelan menggunakan model hibrid menghasilkan nilai AIC, MAE dan MAPE yang lebih rendah berbanding dengan model GARCH. Kesimpulannya, model hibrid ARIMA-GARCH lebih baik dalam permodelan harga minyak mentah berbanding dengan model GARCH.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.0 Introduction

Time series is a collection of data recorded over a period of time. Time series are recorded at equally spaced time intervals such as hourly, daily, weekly, monthly or yearly time separations. It can be discrete and continuous. The discrete time series is an observations measured at discrete points of time such as the number of cars in a country and number of heads when flipping three coins. Continuous time series is an observation measured at every time such as temperature reading and flow of a river. Time series is found in many fields, such as in economics, sociology, meteorology and astronomy. Examples include the sales by quarter of the Microsoft Corporation since 1985, the annual production of sulphuric acid since 1970, monthly salary, weekly employment rates and daily data on sales. Time series analysis provides tools for selecting a model that can be used to forecast future events.

Forecasting is the use of historical data to determine the direction of future trends. Forecasting involves the generation of a number, set of numbers or scenario that corresponds to a future occurrence. Forecasting is important in many fields including government, environmental sciences, medicine, politics, business, industry and finance. A forecasting exercise is usually carried out in order to provide an aid to decision-making and in planning the future. Typically, all such exercises work on the premise that if we can predict what the future will be like we can modify our behaviour now to be in a better position, than we otherwise would have been, when the future arrives. For example, forecasting is used by companies to determine how to allocate their budgets for an upcoming period of time. This is typically based on demand for the goods and services it offers. Investors utilize forecasting to determine if events affecting a company, such as sales expectations, will increase or decrease the price of shares in that company. Forecasting also provides an important benchmark for firms which have a long-term perspective of operations. Other common forecasts are traffic patterns and weather condition.

Forecasting can be classified as short-term, medium-term and long-term. Short-term forecast events involve only a few time periods up to 3 months. In certain business, daily forecast may be necessary but for other businesses, a short range forecast created in weekly or monthly time buckets may be adequate. It is used for planning purchases, job scheduling and job assignments. Medium-term forecasts extend from 1 to 2 years and it is useful in sales planning, production planning and cash budgeting. Long-term forecasts extend beyond 2 years. Long-term predictions are essential to allow sufficient time for the procurement, manufacturing, sales, finance and other departments of a company to develop plans for possible new plants, financing, development of new products and new methods of assembling (Bowerman, O'Connell and Murphree, 2013).

In addition, forecasting can be categorized into two categories which are quantitative and qualitative. Qualitative forecasting is using expert opinion, judgements and collective experience to predict the future events. There is little or no historical data. Delphi Method is widely used for this technique. Quantitative forecasting, on the other hand, is defined as the statistical technique that uses factual numbers to predict future events. This method formally discovers a pattern of historical data to identify an appropriate model and then use the model to extrapolate the pattern into the future. Examples of quantitative forecasting methods are Box-

Jenkins methods, Exponential Smoothing methods, stepwise autoregression and regression methods.

The current study is undertaken with the purpose of modelling time series data based on its historical pattern and the data's characteristic of being volatile in nature. Volatility clustering are evident when such data are plotted. Volatile clustering means large changes in the data tend to cluster together and resulting in persistence of the amplitudes of the changes.

1.1 Background of the Study

Over the last few years, modelling and forecasting volatility of a financial time series has become the area of interest. This is simply because volatility is considered as an important concept for many economic and financial applications such as portfolio optimization, risk management and asset pricing. Volatility means the conditional variance of the underlying asset return. The most well-known and frequently applied models for this volatility are the conditional heteroscedastic models. The main objective of building these models is to make a good forecast of future volatility which will be helpful in obtaining a more efficient portfolio allocation, having a better risk management and more accurate derivative prices of a certain financial instrument.

In the current study, modelling will be carried out using crude oil prices data. These data are chosen because apart of being volatile as it is the area of focus for this study, crude oil has a great importance to mankind. Crude oil is a naturally occurring, flammable liquid found in rock formations in the Earth. It is a complex mixture of several hydrocarbons which are compounds consisting of hydrogen and carbon. Crude oil may also include other organic compounds such as nitrogen, oxygen and sulphur. It is basically a fossil fuel that is obtained from the transformation of plant and animal remains over millions of years. It is most commonly found in place such as the sea bed. The physical properties of this oil also include a sticky consistency like tar. It may sometimes also have a thinner consistency. It may vary in appearance depending on its composition, but it is usually black or dark brown in colour. Crude oil is the most important natural resource of the industrialized nations. The various hydrocarbons found in crude oil can be separated through distillation and can be used to produce different types of refined petroleum products. It can generate heat, drive machinery and fuel vehicles and airplanes. Its components are used to manufacture almost all chemical products such as plastics, detergents, paints and even medicines.

The fluctuation of crude oil prices has affected many related sectors and stock market indices. The nature of the volatility of the crude oil prices is such that it is very sensitive to the international socioeconomic and geopolitical events. It is not, therefore, just the demand and supply or inventory and consumption that influence the crude oil prices, but there are many irregular factors such as weather, the marginal cost of oil production and technological changes. Crude oil prices will influence the cost of gasoline, home heating oil, manufacturing and electric power generation. The increase of oil prices will give impact to the cost of daily needs because our daily necessities depend on transportation.

The inconsistency of crude oil prices makes the modelling and forecasting of crude oil prices an important area of study. Forecasting crude oil prices are essential to avoid the future prices of the non-renewable natural resources to raise sky-rocket. Other than providing the information about the future oil prices to the public, crude oil forecasting is also crucial in determining the world economic movement.

Crude oil prices will be used as the case study in this research. Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and hybrid of Autoregressive Integrated Moving Average (ARIMA) and GARCH model will be developed to model the volatile data. Suitable time series model will be determined to obtain model that will be precise enough for modelling volatile data.

1.2 Statement of Problem

Volatility analysis of financial time series is a crucial aspect of many financial decisions. Volatility forecasts are important in order to either construct less risky portfolios, asset allocation or obtain higher profits. Hence, good analysis and forecasting of volatility become an important aspect in recent years.

In the current study, a set of volatile crude oil prices is modelled by using time series models. Eventhough ARIMA is a popular method for forecasting, it is not able to handle volatile data. To improve on the ARIMA model, it will be hybridized with GARCH model. The models proposed in this study are GARCH and hybrid ARIMA-GARCH model. ARIMA models have been used widely for forecasting different types of time series to capture the long term trend while in the case of financial time series that have been shown to have volatility clustering, ARCH based models have been used.

Parameter estimation is a need in developing the models. Common methods of estimation are Maximum Likelihood Estimation (MLE), Ordinary Least Square Estimation (OLS) and method of moments (MOM). In this study, the following question will be explored:

Between GARCH and hybrid ARIMA-GARCH model, which model is more accurate in modelling volatile data?

1.3 Objectives of the Study

The objectives of this study are:

1. To develop the best GARCH and hybrid ARIMA-GARCH models for crude oil prices.
2. To compare the modelling performances of GARCH and hybrid ARIMA-GARCH models for the prices of crude oil.

1.4 Scope of the Study

This study focuses on the hybrid ARIMA-GARCH model in modelling volatile data by using statistical tools in the EViews software. The data that are used in the current study is the prices of crude oil (Dollars per Barrel) obtained from United State Energy Information Administration (EIA) starting from 2nd January 1986 until 27th October 2014. Two time series models which are Box-Jenkins ARIMA model and GARCH model will be hybridized in this study. The performances of GARCH and hybrid ARIMA-GARCH model will be compared.

1.5 Significance of the Study

Through this study, we will forecast the prices of crude oil using hybrid ARIMA-GARCH method. The hybrid models inherit both qualities of Box-Jenkins and GARCH methods. Combining models can be an effective way to overcome the limitations of a component model and it is able to improve modelling accuracy. Generally, ARIMA models are able to handle nonstationary data while the GARCH model has the ability to capture the volatility by the nonconstant of conditional variance. It is hoped to produce the best model of prediction for volatile data by hybridizing both methods.

The precise prediction of crude oil prices is essential to avoid the future prices of the non-renewable natural resources to raise sky-rocket. This is because forecasting the prices of crude oil provides useful information which helps the government agencies or other policy makers to plan and manage their resources in a more efficient manner. This prediction will help government to take actions according to the situation.

Application of forecasting tools in financial areas will strengthen the multidisciplinary relationship between statisticians and economists. Better predictions can be obtained and would benefit both parties.

1.6 Limitation of the Study

Limitations of the study are on the data and models used. The data that are used in the current study are the daily prices of crude oil (Dollars per Barrel) for about 29 years, starting from 2nd January 1986 until 27th October 2014. In terms of models, the time series models used are Box-Jenkins ARIMA, GARCH and their hybrid. Since the data are volatile, only the performances of GARCH and hybrid ARIMA-GARCH model will be compared.

1.7 Organization of the Report

This study explores the potential of hybrid ARIMA-GARCH model in handling volatile data. The price of crude oil data will be used for this purpose. This report consists of five chapters. Chapter 1 presents the research framework. It starts with the introduction of time series and followed by the statement of problem, the objectives of the study, the scope of the study and the significance of the study.

The literature review is presented in Chapter 2. It consists of reviews on volatility, reviews on crude oil forecasting, reviews on ARIMA model, reviews on GARCH model and reviews on hybrid ARIMA-GARCH model.

Chapter 3 discusses the research methodology. It consists of stationarity testing, Box-Jenkins ARIMA methodology, GARCH methodology and hybrid ARIMA-GARCH methodology.

Analysis of the data is conducted in Chapter 4. The best hybrid model will be chosen based on Akaike's Information Criterion (AIC). The discussion of the results will be presented in the last section of this chapter.

The last chapter which is Chapter 5 presents the summary, conclusions and suggestions for further research.

REFERENCES

Ahmad, M.H., Ping, P.Y., Yazir, S.R., & Miswan, N.H. (2014). A Hybrid Model for Improving Malaysian Gold Forecast Accuracy. *International Journal of Math. Analysis*, 8, 1377-1387.

Ahmad, M.I. (2012). Modelling and Forecasting Oman Crude Oil Prices using Box-Jenkins Techniques. *International Journal of Trade and Global Markets*, 5, 24-30.

Ahmed, A.E.M., & Suliman, S.Z. (2011). Modeling Stock Market Volatility using GARCH Models Evidence from Sudan. *International Journal of Business and Social Science*, 2, 114-128.

Asadi, S., Tavakoli, A., & Hejazi, S.R. (2012). A New Hybrid for Improvement of Autoregressive Integrated Moving Average Models Applying Particle Swarm Optimization. *Expert Systems with Applications*, 39, 5332-5337.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics,* 31, 307-327.

Bowerman, B., O'Connell, R., & Murphree, E. (2013). *Business Statistics in Practice*. (5th ed.). United States: McGraw-Hill.

Chase, Charles (2013). *Demand-Driven Forecasting : A Structured Approach to Forecasting*. (2nd ed.).New Jersey: John Wiley & Sons, Inc.

Chong, C.W., Ahmad, M.I., & Abdullah, M.Y. (1999). Performance of GARCH Models in Forecasting Stock Market Volatility. *Journal of Forecasting*, 18, 333-343.

Dobre, I., & Alexandru, A.A. (2008). Modelling Unemployment Rate using Box-Jenkins Procedure. *Journal of Applied Quantitative Methods*, 3, 156-166.

Du, Y. (2012). *Modelling and Forecasting Volatility of Gold Price with Other Precious Metals Prices by Univariate Garch Models*. Master’s Degree. Uppsala Universitet.

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50, 987-1007.

Frimpong, J.M., & Oteng-Abayie, E.F. (2006). Modelling and Forecasting Volatility of Returns on the Ghana Stock Exchange using GARCH Models. *American Journal of Applied Sciences,* 3, 2042-2048.

Garcia, R.C., Contreras, J., Akkeren, M.V., & Garcia, J.B.C. (2005). A Garch Forecasting Model to Predict Day-Ahead Electricity Prices. *IEEE Transactions on Power Systems*, 20, 867-874.

Glosten, L., Jaganathan, R., & Runkle, D. (1993). Relationship between the Expected Value and Volatility of the Nominal Excess Returns on Stocks. *Journal of Finance*, 48, 1779-1802.

Hadavandi, E., Ghanbari, A. & Abbasian-Naghneh, S. (2010). Developing a Time Series Model Based On Particle Swarm Optimization for Gold Price Forecasting. *Third International Conference on Business Intelligence and Financial Engineering (BIFE 2010)*. 13-15 August. 2010. Hong Kong, IEEE Computer Society, 337-340.

Harper, A., Jin, Z., Sokunle, R., & Wadhwa, M. (2013). Price Volatility in the Silver Spot Market: An Empirical Study using GARCH. *Journal of Finance and Accountancy*, 13, 1-11.

Healy, A. (2007). *GARCH Modelling Applied to the Index of Industrial Production in Ireland*. Master’s Degree. National University of Ireland, Galway.

Khashei, M., Bijari, M., & Raissi, G.H.A. (2009). Improvement of Autoregressive Integrated Moving Average Models using Fuzzy Logic and Artificial Neural Networks (ANNs). *Neurocomputing*, 72, 956–967.

Khashei, M., Hejazi, S.R., & Bijari, M. (2008). A New Hybrid Artificial Neural Networks and Fuzzy Regression Model for Time Series Forecasting. *Fuzzy Sets and Systems*, 159, 769-786.

Kirchgassner, G., Wolters, J., & Hassler, U. (2013). *Introduction to Modern Time Series Analysis*. (2nd ed.). London: Springer.

Kun, H. (2011). *Modelling Volatility of S&P 500 Index Daily Returns: A Comparison Between Model Based Forecasts and Implied Volatility*. Master's Degree. Hanken School of Economics, Vasa.

Marzo, M., & Zagaglia, P. (2010). Volatility Forecasting for Crude Oil Futures. *Taylor & Francis Journals*, 17, 1587-1599.

Miswan, N.H. (2013). *Modelling and Forecasting Volatile Data by Using ARIMA and GARCH models*. Master's Degree. Universiti Teknologi Malaysia, Skudai.

Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, 59, 347-370.

Nihan, N.L., & Holmesland, K.O. (1980). Use of the Box and Jenkins Time Series Technique in Traffic Forecasting. *Elsevier Scientific Publishing Company*, 9, 125-143.

Nochai, R., & Nochai, T. (2006). Arima Model for Forecasting Oil Palm Price. *Regional Conference on Mathematics, Statistics and Applications*. 13-15 June. Universiti Sains Malaysia, Penang, 1-7.

Okyere, F., & Kyei, L. (2014). Temporal Modelling of Producer Price Inflation Rates of Ghana. *IOSR Journal of Mathematics*, 10, 70-77.

Padhan, P.C. (2012). Application of ARIMA Model for Forecasting Agricultural Productivity in India. *Journal of Agriculture and Social Sciences*, 8, 50-56.

Paul, R.K., Ghosh, H., & Prajneshu (2008). Development of Out-of-Sample Forecasts Formulae for ARIMAX-GARCH Model and their Application. *Journal of the Indian Society of Agricultural Statistics*, 1, 85-92.

Reider, R. (2009). Volatility Forecasting I: GARCH Models. 1-16.

Seneviratna, D.M.K.N., & Shuhua, M. (2013). Forecasting the Twelve Month Treasury Bill Rates in Sri Lanka: Box Jenkins Approach. *IOSR Journal of Economics and Finance*, 1, 42-47.

Shaari, M.S., Pei, T.L., & Abdul Rahim, H. (2013). Effects of Oil Price Shocks on the Economic Sectors in Malaysia. *International Journal of Energy Economics and Policy*, 3, 360-366.

Shabri, A., & Samsudin, R. (2014). Daily Crude Oil Price Forecasting using Hybridizing Wavelet and Artificial Neural Network Model. *Mathematical Problems in Engineering*, 2014, 1-10.

Sivakumar, P.B., & Mohandas, V.P. (2010). Performance Analysis of Hybrid Forecasting Models with Traditional ARIMA Models - A Case Study on Financial Time Series Data. *International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM)*, 2, 187-211.

Srinivasan, A. (1997). *Application of Information Technology and Statistical Process Control in Pharmaceutical Quality Assurance & Compliance*. Master's Degree. Massachusetts Institute of Technology, Gainesville.

Taylor, S. (1986). *Modelling Financial Time Series*. (2nd ed.). John Wiley & Sons: Chichester.

Wang, S., Yu, L., & Lai, K.K. (2004). A Novel Hybrid AI System Framework for Crude Oil Price Forecasting. *Lecture Notes in Computer Science*, 3327, 233-242.

Yaziz, S.R., Ahmad, M.H., Nian, L.C., & Muhammad, N. (2011). A Comparative Study on Box-Jenkins and GARCH Models in Forecasting Crude Oil Prices. *Journal of Applied Sciences*, 7, 1129-1135.

Yaziz, S.R., Azizan, N.A., Zakaria, R., & Ahmad, M.H. (2013). The Performance of Hybrid ARIMA-GARCH Modelling in Forecasting Gold Price. *20th International Congress on Modelling and Simulation*. 1-6 December. Adelaide, Australia, 1201-1207.

Yusof, F., Kane, I.L., & Yusop, Z. (2013). Hybrid of ARIMA-GARCH Modelling in Rainfall Time Series. *Jurnal Teknologi*, 63, 27-34.

Zhang, G.P. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*, 50, 159-175.