Models for Oil Price Prediction: A Case Study of Pakistan

By

Saheem Riaz Cheema

PIDE2017FMPHILETS14

Supervised by

Dr. Saud Ahmed Khan

A thesis submitted to

Pakistan Institute of Development Economics (PIDE)

In the partial fulfillment of requirements for the degree of

Master of Philosophy in Econometrics

Department of Economics & Econometrics

Pakistan Institute of Development Economics

Islamabad

2020

Pakistan Institute of Development Economics

CERTIFICATE

This is to certify that this thesis entitled: **"Models for Oil Price Prediction: A Case Study of Pakistan"** submitted by Mr. Saheem Riaz Cheema is accepted in its present form by the Department of Economics & Econometrics, Pakistan Institute of Development Economics (PIDE), Islamabad as satisfying the requirements for partial fulfillment of the degree of **Master of Philosophy in Econometrics**.

External Examiner:

Dr. Arshad Ali Bhaii
Assistant Professor
International Islamic University
Islamabad

Supervisor:

Dr. Saud Ahmed Khan
Assistant Professor
PIDE, Islamabad

Head, Department of Economics & Econometrics:

Dr. Karim Khan
Associate Professor/Head
Department of Economics &Econometrics
PIDE, Islamabad

DEDICATION

I would like to dedicate this thesis to my family, friends and teachers who always supported me throughout all the rights and wrongs of my life.

ACKNOWLEDGEMENT

All glories to thee Allah, the omniscient and omnipotent and his Benediction be upon his prophet. The savior of mankind from darkness of ignorance a symbol to be and to do right. My deepest thanks of Allah the Almighty, who made me able to do this work.

I express my heartiest obligation and appreciation to my worthy supervisor Dr. Saud Ahmed Khan for the valuable guidance, valuable suggestions, sincere and sympathetic attitude, especially for his politeness during the process of writing this dissertation.

I am really grateful to all the staff members, of Ph. D Section and the Library staff for their cooperation whenever I needed, I am also thankful to all my friends and classmates especially Mr. Rizwan Ahmad for their cooperation and encouragement.

Finally, I would like to express my deepest gratitude to all my family members, especially my mother and my wife for their encouragement and endless prayers. Special words of thanks for my late father for his continuous effort and support enabled me to achieve the peak of knowledge and career.

ABSTRACT

In the estimation of crude oil price prediction, parametric econometric and machine learning models are used, to predict the future price. The parametric econometric models include the conventional time series models by using the suitable(log-difference) transformation to fulfill the necessary assumptions according to the axioms of econometric modelling. The hybridization of ARIMA and GARCH is done to get the model with best predictions. The machine learning model (Recurrent Neural Network (Long Short-Term Memory) state of the art architect of neural network for sequential/time series data. The recent interest has been focused on developing the estimation technique to predict the future prices by using the series at level, rather than the return series. We find that the hybrid ARIMA-GARCH outperformed amongst all the models used in this study. But on theoretical basis and also the graph of predictions from RNN(LSTM) suggests that if we have to model high frequency data by estimating series at level rather than return series then one must go for the machine learning model RNN(LSTM).

Keywords: Crude Oil Price prediction, ARIMA, GARCH, ARIMA-GARCH, RNN(LSTM).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Oil is a commodity that influences everybody's daily life in a multitude of ways unlike any other. The industrial world depends on oil almost as much as the human body does on blood. Oil prices and supply impact transport, whether driving or flights on a regular basis, as well as economic development, as goods must be transported, and oil is used almost everywhere in the secondary sector. Machines need to be oiled, engines require fuel and in the automotive sector certain goods are completely oil dependent. For computers and circuit boards the insulation on the wires is made of palm oil. Popular products are manufactured from oil, such as shampoo, detergent, solvents, paint, ink, tires, lubricants, candle wax and hundreds of thousands of other products. Heating, and of course the military, depends heavily on oil prices and availability. The nature of oil, and its use in the global economy, makes finding any comparable resource difficult. It should also be borne in mind that oil is a non-renewable fossil fuel. Actually economic activity is fairly associated with oil consumption. Not only is the interest in oil prices and, in particular, in the ability to predict oil prices important for the aforementioned reasons.

For individuals such as traders, investors and risk managers, forecasting price volatility is crucial for understanding market dynamics. One of the highest demand and consumption in Pakistan is for Crude oil. Oil price is a dynamic time series data with massive periodic swings.

There are in fact two different modeling methods. The first is a structural model of the oil price, based on simple data such as demand and supply, which is applied using a linear regression. The second is an alternative through time series, which involves an analysis of linear and nonlinear time series. The nonlinear time series analysis uses an autoregressive model of a neural network, where the strength of the relationship is modelled using neural network regressions.

Due to the increasing fluctuations over time, forecasting the oil prices is difficult. Whenever demand for a commodity such as oil exceeds supply, the price can rise incredibly high, which is because demand and supply in the short run are quite inelastic. While people will be upset by higher oil prices, change of habits and consumption takes time. The broad range of cost of extracting oil results in increased volatility in the medium term. Consequently, market changes would trigger a much more significant price change than before.

Multi-period volatility forecasts are prominently featured in asset pricing, portfolio allocation, risk management, and most other finance areas where long-horizon risk measures are essential. These predictions can be made up in three very different ways. The first approach is to estimate a horizon-specific volatility model, such as a weekly or monthly GARCH, which can then be used to formulate direct volatility forecasts over the coming week, month, etc. The second method is to predict a daily prediction and then iterate the daily forecasts to obtain weekly or monthly forecasts. The literature regarding forecasting refers to the first approach as "direct" and the second as "iterated" by RNN (Recurrent Neural Network), a machine learning technique. The advantages of this approach is that, as in the direct approach, one focuses specifically on multi-period forecasts, while preserving the use of high-frequency data.

To forecast various types of time series, the long-term trend captured using autoregressive integrated moving average (ARIMA) models. ARCH based models have been used in the case of financial time series which have been observed to have volatility clustering where significant changes in the data appear to cluster together and result in persistence of the amplitudes of changes. Within the background of WTI (West Texas Intermediate) crude oil prices in US dollar ($) per barrel, the oil selling price will be modelled and forecast using the models ARIMA and GARCH. Although the models provided a good data fit with the GARCH being superior, it was shown that a combination of those two models (hybrid) could improve predictive accuracy.

In order to motivate the class of models, it is important to remember that a key ingredient of conditional volatility models is to add more weight to the most recent returns (i.e., information). When forecasting future (daily) conditional volatility, in the case of the original ARCH model that means the most recent (daily) squared returns have more weight. The theory of continuous time semi-martingale stochastic processes, more explicitly stochastic volatility continuous time jump-diffusions, is the basis of so-called realized volatility (RV) modeling. Whereas intra-daily data is used to create RV, prediction models put more weight on recent (daily) RV but do not differentiate between intra-daily returns, given the use of intra-daily data. If volatility is a constant process it would be normal, as Malliavin and Mancino (2005) pointed out, to weigh intra-daily data differently.

1.2 Crude oil prices and Forecast Failure

Considering the position of the Organization of Petroleum Exporting Countries (OPEC), another challenge in modeling the oil price emerges. OPEC can behave as a swing producer and so it only meets the remaining demand for oil after the non-OPEC supply has been depleted. OPEC acts as a price-taker in this scenario, but OPEC may also produce at total potential and take on a more competitive position in the global oil market, in which situation OPEC acts as a price maker. The effect of a dominant producer is easily observable, but more difficult to model for two reasons: the first is the inability to predict the producer 's behavior, the second is the inability to translate a particular behavior into the model. The forecasting based on this story is not appropriate in any ways and thus the investor is uncertain about the tomorrow's price which is main objective of this study. The forecasting is based on the appropriate modeling of the underlying process and identification of the true data generating process of the series. If one is ineligible of employing the true data model it becomes inevitable to incorporate the true forecasting methodology. This is assumed thus a true data generating process will be identified and then an appropriate estimation methodology will be employed, due to the limitations of the parametric approach the non-parametric method is also applied to meet the quest of the appropriate forecasting.

1.3 Problem Statement

Literature suggests that the investor/market player is very much interested in tomorrow's price. So, it's needed to estimate the series at level rather than return series. As the conventional time series models (ARIMA, GARCH, ARIMA-GARCH) are parametric and based on hard core assumptions which are the prerequisites, the most important is 'stationarity' of the series. To achieve stationarity different transformations are available, but when log-difference transformation applied on the series at level, it becomes return series. As discussed above we want to estimate the series at level for prediction of future price. To achieve this objective by bridging this gap through estimation of series at level rather than return series, we must switch on machine learning models (Recurrent Neural Network).

Oil prices have been modelled by ARCH techniques along its variants. However, one of the drawbacks of the ARCH specification, is that it looked more like a moving average specification than an auto regression and later it was improved as GARCH (p, q) to include the lagged conditional variance terms as autoregressive terms. The GARCH (p, q) depends both on past values of the shocks, which are captured by lagged squared residual terms, and on past values of itself, which are captured by lagged terms. The GARCH (1,1) model is a parsimonious alternative to an infinite ARCH(q) process. The investor is very much interested in tomorrow's price rather than the average price. this study comprises that which model gives better prediction of oil prices by using hybridized ARIMA-GARCH rather than simple ARIMA or GARCH and Machine learning model named as Recurrent Neural Network (Long Short-Term Memory). The addition to literature can also be justified on the basis that the GARCH family models are mainly based on the hard core assumptions of the classical econometrics and also the specific nature of the GARCH models is also questionable as described by the Robert Engle (1982). Thus, to mitigate this the study is assumed to switch to the assumption free methodology that is non-parametric models.

1.4 Study Objectives

- The Primary objective of this study is to suggest an appropriate model to forecast WTI crude Oil prices.
- The time series Models allows us to investigate some intriguing empirical modelling strategies and investigate the capability of another hybrid and machine learning RNN(LSTM) models in forecasting the volatility.

1.5 Hypothesis

The machine learning models (Recurrent Neural Network (Long Short-Term Memory)) provide better predictions than the conventional time series models.

1.6 Significance of the study

- Which one give the better predictions of WTI crude oil prices, conventional time series models or RNN (LSTM).
- Contrary to the existing estimation methods it is intended to estimate based on the Machine Learning techniques to improve the forecast.

1.7 Organization of the study

This study is structured into five sections or chapters, chapter one provides a short overview of the study covering the research problems, goals and hypothesis. Chapter two reviews the relevant literature. Chapter three discusses the methodology. Chapter four presents empirical results and discussion. The final section discusses the findings and suggestions.

CHAPTER 2

REVIEW OF LITERATURE

2.1 Introduction

Literature review gives fundamental, theoretical, and empirical background and effective information to comprehend depth and significance of a study problem. Reviewing past studies is therefore one of the first steps to understand, evaluate and solve a research issue. Previous studies on modeling of actual oil prices and their predictions have evaluated chronological order in the subsequent chapter.

Modelling and forecasting the changing aspects of WTI crude oil prices is not an easy task because the prices may possibly be fluctuating unpredictably from time to time and also depends on so many factors.

2.2 Previous studies for WTI oil price prediction

Tang and Hammoudeh (2002) used a nonlinear regression to estimate the price of an OPEC basket. Using stock rates of fuel from the OECD and relative market inventories. They examined the world oil price actions based on the goal zone model of the first decade. Using anecdotal data for the period 1988–1999, they found that OPEC had attempted to establish a poor target zone regime for the price of oil. The econometric tests showed that the oil price trend was not only influenced by real and significant OPEC actions but also balanced by the expectations of intervention from market participants. Consequently, the non-linear model based on the target zone principle has very strong predictive potential when the price of oil exceeds the band's upper or lower limit.

Zamani (2004) applied an econometrics approach for predicting short-term quarterly WTI crude spot oil prices. This paper provided a quarterly, short-term forecast model of WTI using OECD stocks, non-OECD demand and econometrics-based OPEC availability. Stock rates included SPR and industrial equilibriums in the OECD oil sector and OPEC supply is the most significant tool for controlling the price of oil often called non-OECD demand as a market indicator for that region. The econometrics relationships between certain variables were analyzed in this research. Because the commercial stocks and SPR seek different targets they were analyzed separately. Then price model forecasting evolved on the basis of lagged values of industrial production, non-OECD demand and OPEC supply. This model is mainly useful for the OPEC to analyze the effect of various supply limits on future oil prices. The model provided good dynamic forecasts for the post-Gulf War era, both in sample and out of sample. The model's in-sample and out-of-sample predictions are comparable to those obtained from other models. The model is intended for the realistic forecaster and is designed to be easy enough to quickly incorporate the variables in a spreadsheet or other software package. The flexibility and ease of updating makes this model appealing for exploring various scenarios to see the impacts that market shifts may have on monthly crude oil spot prices if inventories, supply, imports or demand change. Finally, it is easy to update the model structure regularly if there is a fundamental market change or a drop in the usual level of inventories.

Wang et al. (2005) discussed that the gains of ARIMA models. ARIMA models are essentially a group of distinctive linear models that are intended to be the best for linear time series data and captured the linear features in time series data. ARIMA models are thus best suited on a theoretical basis.

Similarly, Chinn et al. (2005) considered the predictive value of the energy futures and analyzed the relationship between the future prices and spot prices for the different energy commodities and the ARIMA model (1, 1 , 1) used for the prediction of crude oil prices.

Lanza et al. (2005) observed crude oil and commodity prices then analyzed using error correction models. They investigated the hypothesis that a shift in investment behavior among international oil companies (IOC) towards the end of the 1990s had long-lasting effects on OPEC strategies and the development of oil prices. In the aftermath of the Asian economic crisis, concerted investment restrictions were placed on the IOCs by financial-market pressures to increase short-term profitability. To compare the impact of those tacitly collusive capital restrictions on oil supply with an alternative defined by industrial stability, a partial equilibrium model for the global oil market was applied. The core findings indicated that even temporary economic and financial shocks

may have a long-term impact on the development of oil prices.

In a related research, Ye etal. (2006) included in the linear forecasting model proposed by Ye etal. (2002, 2005) nonlinear variables such as low- and high-inventory variables to predict short-run WTI crude oil prices. Using OECD stocks, demand from non-OECD countries and supply from OPEC. Since inventories have a lower or minimum level of activity, economic literature suggests a nonlinear relationship between inventory and commodity prices. In the short-run crude oil market, this was found to be the case. Two nonlinear inventory variables described and extracted from the usual monthly level and relative level of OECD crude oil inventories from the post-1991 Gulf War to October 2003 in order to explore this inventory-price relationship: one for the low inventory state and one for the high inventory state of the crude oil market. Incorporating low- and high-inventory variables into a single equation model to forecast short-run WTI crude oil prices improved the fit and predictive capability of the model.

On the other hand, Sadorsky (2006) suggested that it is easier for Vector Auto Regression (VAR) and for bivariate GARCH models out of sample predictions of a single GARCH model equation, and higher in forecasting the petroleum prices for future. He analyzed and the performance of four multivariate models of volatility, namely CCC, VARMA-GARCH, DCC and BEKK, to measure optimal portfolio weights and optimal hedge ratios for the crude oil spot and potential returns of two major international crude oil benchmarks, Brent and WTI, and to propose a strategy for crude oil hedges. The empirical findings indicated that Brent 's optimum portfolio weights of all multivariate volatility models suggest keeping futures in greater proportions than spot.However, for WTI, DCC and BEKK suggested that crude oil futures should be spotted, but CCC and VARMA-GARCH suggested that crude oil spot be kept for future. Additionally, the estimated optimal hedge ratios (OHRs) from each multivariate conditional volatility model give the time-varying hedge ratios, and suggested short crude oil futures with a high proportion of one dollar long in crude oil spot. Ultimately, the efficiency of hedging showed that DCC (BEKK) is the best (worst) model for OHR calculation in terms of lowering portfolio variance.

Slightly more recent, Dees et al. (2007) developed a linear world oil market model to predict oil demand, supply, and prices mainly focused on OPEC behavior. An econometric study of oil prices: they assessed arguments that the capacity of OPEC to control real oil prices has decreased and that the relationship between real oil prices and OPEC production can be used to check opposing OPEC activity hypotheses. An econometric study revealed that there is a statistically significant relationship between the real oil prices, the utilization of OPEC capacity, OPEC quotas, the degree to which

OPEC meets these production quotas, and crude oil stocks in the OECD. Such factors "Granger triggers" real oil prices but these factors are not "Granger affects" actual oil prices. The findings suggested that OPEC controls oil prices and that it is not possible to use previous models to check rival OPEC output behavior. The impact of OECD oil stocks on real oil prices suggested that private decisions on optimal crude oil stocks can have a major externality.

Similarly, Agnolucci (2009) used different types of GARCH models and pointed out unpredictable models to predict potential market uncertainty in the daily WTI, but the findings found were inconsistent and revealed their output with respect to statistical tests and various steps. He contrasted the predictive potential of two methods that can be used to forecast volatility: GARCH-type models in which predictions obtained after estimating time series variables, and an inferred volatility model in which predictions obtained by inverting one of the variables used for market options. The key focus of the work discussed here was to determine which model provide the best volatility forecast for the future WTI contract, evaluated according to statistical and regression-based criteria.

Murat and Tokat (2009) investigated the relationship between futures and spot crude oil prices and thus checked the ability of futures prices to predict spot price movements using random walk model. The crack spread in oil markets refers to the price relation of crude-products. Refiners are major oil market players, and are mainly exposed to crack spread. In other words, the purpose of protecting crack spread significantly drives refiner operation. In addition, oil consumers are active participants in the market for oil hedging, and are also vulnerable to crack spread. Hedge funds, from a different angle, use crack spread extensively to gamble on oil prices. The issue they want to answer is whether the crack spread futures can be a good indicator of oil price changes, based on the large amount of crack spread futures trading in oil markets. They investigated first if there is a causal link in a vector error correction system between the crack spread futures and the spot oil markets. They consider the causal effect of crack spread futures on both the long- and short-run spot oil market after April 2003 where we observed a systemic split in the model. They used the Random Walk Model (RWM) as a benchmark to analyze the forecasting efficiency, and they also tested the predictive power of crack spread futures against crude oil futures. The results showed that both the crack-spread futures and the crude oil futures outperformed the RWM and the crack spread futures are almost as strong as the crude oil futures in predicting spot oil market movements.

Cheong (2009) introduced ARCH models to predict the markets for crude oil. On another way, more recent studies have applied GARCH as well as various GARCH family models for forecasting the price of oil.

Kang etal. (2009) proposed CGARCH, FIGARCH, and IGARCH models for forecasting crude oil market volatility. This article explored the efficacy of a volatility model for three crude oil markets — Brent, Dubai, and West Texas Intermediate (WTI)—with respect to its ability to forecast and classify stylized facts of volatility, especially persistence of volatility or long memory. In that sense, using conditional volatility models, they assessed persistence in the volatility of the three crude oil prices. The Models CGARCH and FIGARCH are better suited to catch persistence as for the GARCH and IGARCH models.

Wang et al. (2009) implemented a GARCH method that makes volatility forecasts using sampled returns at a higher frequency than the horizon predicted. They called the model class High Frequency Data-Based Projection-Driven GARCH, or HYBRID-GARCH models, because the dynamics of variance are driven by what we consider HYBRID processes. The HYBRID processes may provide sampled data at any frequency.

In addition, Marzo etal. (2010) used various GARCH models to predict the potential volatility of the regular crude oil prices traded on NYMEX. The authors reached a conclusion that using the various statistical tests such as DM test, output indicators as modified heteroscedasticity of MSE, MSE and MAE, and success ratio, no model works well on a continuous basis.

In a related study, Mohammadi and Su (2010) contrasted the predictive results of various GARCH-type models to forecast the price of crude oil.

Wei et al. (2010) expanded the Kang et al. (2009) analysis with the implementation of linear and nonlinear GARCH-class models for the same function. Predicting volatility on the crude oil market: More proof using GARCH-class models expanded Kang et al. 's research (2009). For capturing the volatility characteristics of two crude oil markets — Brent and West Texas Intermediate (WTI), they used a greater number of linear and nonlinear generalized autoregressive conditional heteroskedasticity (GARCH) class models. The GARCH-class models' one-, five- and twenty-day out-of-sample volatility forecasts are tested using the superior predictive capability check. Unlike Kang et al. (2009), they found that no model can outperform all the other models for either the Brent or the WTI market over various loss functions. Overall, however, nonlinear GARCH-class models, capable of capturing long-memory and/or asymmetric volatility, are more predictive than linear models,

In 2010, however, Alquist and Kilian considered the issue of crude oil price forecasting and concluded that the models that provided lower MSPE in future values would be the better models for estimating future crude oil prices.

Wang et al. (2011) used the intra-day returns to forecast potential volatility at day horizons. The latter demands that they fixed periodic trends intra regular. Two methods were suggested and their relative merits compared. The first method used raw intra-daily data-recording the intra-daily periodic trends with the HYBRID procedure-while the second approach includes pre-adjusted intra-daily returns. They found that the former method dominates both in-sample and out-of-sample, albeit for different requirements of the HYBRID GARCH model.

Ahmad (2012) used the Box-Jenkins techniques to forecast Oman's average monthly crude oil prices at the end of which he proposed the use of the seasonal multiplicative model ARIMA (1 , 1, 5) x (1, 1, 1) in practice to estimate crude oil prices.

On top of that, Hou etal. (2012) used a new method using a non-parametric modeling methodology and forecasting crude oil price return uncertainty, the results showed that the GARCH non-parametric model's sample volatility forecast showed better performance from a GARCH parametric model range.

Lin et al. (2012) estimated that over the past five years, the global economy has undergone volatile instability due to significant rise in oil prices and attacks by terrorists. Although predicting oil prices accurately is necessary but extremely difficult, this study attempted to predict crude oil futures prices accurately by implementing three common neural networking methods, including the multilayer perceptron, the Elman recurrent neural network ERNN, and the recurrent neural fuse network RFNN. Experimental findings suggested that using neural networks to predict future prices of crude oil is acceptable, and consistent prediction is accomplished by the use of different training times. Further, the findings showed that learning efficiency can be increased in most cases by increasing the training time. This shows how the predictive capacity increases when through the training time under ERNNs and RFNNs BPNs were involved in the exceptional event, meaning the predictive capacity increases when the training time is shortened. To sum up, they concluded that in predicting crude oil futures prices, the RFNN outperformed the other two neural networks.

In addition, Ahmed et al. (2013) used the GARCH model to estimate frequent spot oil prices. This approach was used to illustrate non-linear models' advantages and efficiency over the liner models. The analysis consisted of fitting the three separate GARCH models like GARCH-G, GARCH-N and GARCH-T to the regular spot crude oil prices. The different models provided different results over the different data sets the GARCH-G model was considered to be the best model for WTI while the GARCH-N model was the best candidate model for forecasting Brent 's regular spot prices for crude oil.

Yusopc et al. (2013) submitted that rainfall dependency structure is typically very complex in time as well as in space. In this paper it was noted that the regular rainfall series of Ipoh and Alorsetar were affected by nonlinear variance characteristics sometimes referred to as variance clustering or volatility.

Ahmed and Shabri used the ARIMA, GARCH and SVM (Support Vector Mechanic) techniques in 2014 and concluded that the performance of the proposed vector mechanics technique is better than all other traditional methods based on RMSE and MAE's forecast accuracy measurement error.

Monfareda and Enke (2014) studied financial market volatility forecasting as well as the development of financial models, among others, is significant in the areas of risk management and asset pricing. Recent research found that asymmetric GARCH models outperform other GARCH family models when it comes to forecasting volatility. Three common Neural Network models (Feed-Forward with Back Propagation, Generalized Regression, and Radial Base Function) were implemented using this knowledge to help improve the performance of the GJR (1,1) method for estimating volatility over the next forty-four trading days. Around 1997-2011 four separate economic cycles were known to reflect real and contemporary phases of market calm and crisis during training and testing. In addition to stress testing for various neural network architectures to evaluate their performance under different turbulence and usual conditions on the U.S. market, their synergy was also accessed along with another econometric model.

Pahlavani and Roshan (2015) tried to compare the forecast performance of the ARIMA model and hybrid ARMA-GARCH models using regular exchange rate data from Iran against the U.S.Dollar (IRR / USD) from 20 March 2014 till 20 June 2015.The time from 20 March 2014 to 19 April 2015 was used to create the model while the remaining data were used to make sample forecasting and to test the model's predictive potential. Any of the data was obtained from Iran's central bank. Using the Box-Jenkins process, the correct ARIMA model was obtained and some hybrid models such as: ARIMA-GARCH, ARIMA-IGARCH, ARIMA-GJR and ARIMA-EGARCH were calculated for capturing volatilities of return sequence. The findings showed that ARIMA (7,2,12) – EGARCH (2,1) is the strongest model in terms of the lowest RMSE, MAE, and TIC parameters.

Osman et al. (2016) worked for GARCH family models are commonly used in the prediction of dynamic data from time series. In the current research, GARCH (1,1) model 's ability to forecast Malaysian gold, known as Kijang Emas, was enhanced by hybridizing it with the Artificial Neural Network (ANN). Estimates of volatility obtained using GARCH (1,1) model used as one of the input variables in ANN. Model efficiency evaluated by AIC, mean absolute error (MAE), and root mean square error (RMSE) computing.

Summary:

The literature cited above has shown that majority of them has implied the GARCH type modelling with higher order and ARCH type models with also higher orders. The literature on GARCH type models also confirms the estimation of lower order parsimonious models. Thus, a general approach to select the model was also ignored and that the literature has also confirmed the utilization of the parametric approaches which are highly restrictive based on highly restrictive assumptions. The limitations of the GARCH type models is also evitable thus using the mean and variance equations separately. From this, it is intended to utilize the less restrictive and parsimonious approach to tackle the issue of prices forecasting.

CHAPTER 3

METHODOLOGY

3.1 Introduction

The terms in *ARIMA* stands for Auto-Regressive-Integrated-Moving-Average. It is attributed to Box and Jenkins (1976). The ARIMA models are popular and effective tools for the time series forecasting. Thus, to incorporate the concepts of the ARIMA modelling developed by the Box and Jenkins (known as Box Jenkins methodology). This chapter starts with the usual AR and MA models describing the properties of the models and thus providing guidelines to identify the underlying process. The development of ARMA & ARIMA models along with the properties will be judged then a sequential based procedure will be opted to develop a more general GARCH model. In the end, ARIMA-GARCH and RNN (LSTM) models will be developed, and their forecasting ability will be judged.

3.2 The Autoregressive Models.

The Auto regression consists of modelling the time series, suppose $\mathrm{y_t}$ such that it consists of the past values of the evolving variable. If we let us assume the following AR Model;

$$y_t = \varphi y_{t-1} + e_t \qquad \text{--1}$$

it constitutes an AR model of first order denoted by AR(1), where the order of an AR model means the number of previous values that have been incorporated in the regression. Where $|\varphi| < 1$ it means the process is wide sense stationary and e_t is a Gaussian Error term. The logic of Autoregressive models indicates that what happens in time period 't' depends on what has happened in time period 't-1'. Now, if this time series y_t has a mean and variance that do not depend on time 't' then it means the time series is stationary and the covariance depends on the gaps between the series and do not on the actual time period for which this time series has been considered. Thus, to address these properties of time series the AR models have the following underlying statistical properties which are essential to develop a reasonable AR model. The unconditional mean and variance of an AR model is given by;

$$E(y_t) = \varphi E(y_{t-1}) + E(e_t) \quad ------------ \; 2$$

$$= 0$$

The variance is;

$$var(y_t) = var(\varphi y_{t-1}) + var(e_t)$$

$$= \frac{\sigma_u^2}{1 - \varphi^2 \sigma_t^2} \quad -------------- \; \mathbf{3}$$

The series is also characterized by the following Auto-covariance and Auto correlation Functions. Which are important results for the identification of the underlying AR process. These are given by the following functions.

$$\boldsymbol{cov(y_t - y_{t-1}) = E[(\varphi y_{t-1} + e_t) y_{t-1}]}$$

$$\boldsymbol{= E[(\varphi y_{t-1} . y_{t-1}) + E(e_t . y_{t-1})]}$$

$$= \varphi . \sigma_t^2 \quad ------------------- \; \mathbf{4}$$

this can easily be shown that

$$\boldsymbol{cov(y_t - y_{t-2}) = E[(\varphi y_{t-2} + e_t) y_{t-2}]}$$

$$\boldsymbol{= E[(\varphi y_{t-2} . y_{t-2}) + E(e_t . y_{t-2})]}$$

$$= \varphi^2 . \sigma_t^2 \quad ------------------- \; \mathbf{5}$$

in general we can show that

$$\boldsymbol{cov(y_t - y_{t-k}) = E[(\varphi y_{t-k} + e_t) y_{t-k}]}$$

$$\boldsymbol{= E[(\varphi y_{t-k} . y_{t-k}) + E(e_t . y_{t-k})]}$$

$$= \varphi^k.\sigma_t^2 \text{ ------------------ } \mathbf{6}$$

The Auto-correlation function is thus dented by:

$$cor(\boldsymbol{y_t} - \boldsymbol{y_{t-k}}) = \frac{\boldsymbol{cov(y_t - y_{t-k})}}{\sqrt{\boldsymbol{var(y_t).var(y_{t-k})}}}$$

$$= \frac{\varphi^k.\sigma_t^2}{\sigma_t^2}$$

$$= \varphi^k$$

Thus, an Autocorrelation function (ACF) of an AR(p) process dies out exponentially. The graph of an Auto correlation function against ‘k’ which is called the correlogram thus dies out exponentially and does not break at any lag. Thus, to identify the appropriate AR process it cannot be used as a tool. It involves the partial autocorrelation function which where the estimated coefficients by OLS of and AR(k) process are plotted against ‘k’. Now, if the observations are generated by an AR(p) process then theoretically these coefficients will be significant up to lag ‘p’ and zero beyond lag ‘p’. Thus, to sum up whenever to identify the AR process one should resort to the partial Auto correlation function.

So far there has been a discussion of only AR(1) process above but there is a generalization of AR(1) process which is an AR(p) process where ‘p’ denotes the number of lag values of the underlying variable. If we let us assume the AR(2) model in this case we have number of lagged values equal to 2. The AR(2) model can be written as:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + e_t \qquad \text{...7}$$

similarly, this can be generalized to incorporate ‘p’ lags as follows:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + e_t \qquad \ldots 8$$

using the summation notation this can be simply written as:

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + u_t$$

The other properties of the AR(p) process that mean, variance, covariance and correlation hold true as that for the AR(1) process. The stationarity condition for the AR(p) process is same as that for AR (1) process that is the root of the $\sum_{i=1}^{p} \varphi_i < 1$ holds true.

3.3 The Moving Average Models.

The moving average models state that the series depends on the past values of the error terms. That is the implication behind the MA models is that the sequence of series is generated in such a way that it incorporates the past errors while the constructing the data generating process of the series. The simplest form of the MA model is that of MA (1) model which constitutes that the series depends on the immediate past error. This can be written as:

$$y_t = \theta u_{t-1} + u_t \qquad \ldots 9$$

this can be translated into the general form by including the 'q' lagged error terms. This constitutes a model of order 'q'. Thus we have the following form of q order MA(q) model:

$$y_t = \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots \theta_q u_{t-q} + u_t$$

which can be written as by sing the summation notation :

$$y_t = \sum_{i=1}^{q} \varphi_i e_{t-i} + u_t$$

the moving average process by definition is the average of the ‘q’ white noise stationary processes. The moving average models are thus stationary. The moving average models’ possess the following properties which hold true for all the moving models from MA (1) to MA(q). The mean of MA models is clearly equal to zero as it is the average of the white noise error terms. The variance is thus given by the following function. Let for an MA (1) model we have:

$$var(y_t) = var(\theta u_{t-1} + u_t)$$

$$var(y_t) = \theta^2 \sigma_u^2 + \sigma_u^2$$

$$= \sigma_u^2 (1 + \theta^2) \ldots 10$$

The autocovariance would be:

$$cov(y_t, y_{t-1}) = E[(\theta u_{t-1} + u_t)(\theta u_{t-2} + u_{t-1})]$$

$$= E(u_t . u_{t-1}) + \theta E(u_{t-1}^2) + \theta^2 E(u_{t-1} . u_{t-2})$$

$$= \theta \sigma_u^2$$

since, u_t is serially uncorrelated therefore, it is easy to conclude that

$$cov(y_t, y_{t-k}) = 0, \quad \forall\, k > 1$$

from this conclusion it is clear that ACF of an MA process will be pf the form

$$corr(y_t, y_{t-k}) = \frac{cov(y_t, y_{t-k})}{var(y_t). var(y_{t-k})}$$

$$= \frac{\theta \sigma_u^2}{\sigma_u^2 (1 + \theta^2)} = \begin{cases} \dfrac{\theta}{1 + \theta^2} & \forall\, k = 1 \\ 0 & \forall\, k > 1 \end{cases} \qquad \ldots 11$$

Thus, the ACF of an MA(q) process will break off after the lag k=q the PACF of an Ma process thus decays slowly. To identify the order of an MA process therefore one needs to draw its ACF.

ACF and PACF of an MA model together with the AR models determine the appropriate order of the ARMA models. The PACF determines the order or lag length of the lagged values of the underlying series while ACF of an MA determine the order of the lagged errors to be included into the model. The AR and MA models together generate a third class of a modelling procedure known as the ARMA models which have an important implication for univariate time series prediction and Analysis.

3.4 ARMA Models

There is a combination of AR and MA models to give a new series of the model known as the ARMA models. If we let us assume that we have AR(p) and MA(q) models then together these both give a new class of the models known as ARMA(p,q) models and denoted by the following relation:

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{i=1}^{q} \theta_i e_{t-i} + u_t \qquad \ldots 12$$

This is known as the general form of the ARMA(p,q) model where p'q determines the order of the ARMA model. It is important to note that the stationarity concerned only with the AR part of the ARMA model. Therefore, for the ARMA model to be stationary it is mandatory for the root of φ_i to be remain less than 1 in absolute form. The stationarity concern makes the ARMA models to put up in more general form known as ARIMA models. Which incorporates the Stationarity of the series in a model? Therefore, a general stationary model of an ARMA an also be written as ARMA (p,0, q).

3.5 ARIMA Models

The ARMA models can only be made with the series which are stationary it means the series which have mean, variance and covariance all constant over time. However, most of the Economic and Financial time series show trend over time therefore the mean and variance do not remain constant over time as a result these are subjected to the non-stationarity problems. To avoid this and to model the non-stationary series another more general form of the model is developed known as the ARIMA models where “I’ stands for the integrated process or series. It means the order of differencing of the series to approach the stationarity of the economic time series. The General ARIMA (p, d, q) model is written as:

$$\nabla^d y_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \, \nabla^d y_{t-i} + \sum_{j=1}^{m} \beta_j \varepsilon_{t-j} + \varepsilon_t \ldots 13$$

this indicates that an integrated series of a general ARMA model must be differenced ‘d’ times to make it stationary and before it an be represented by an ARMA stationary and invertible process. Where, ‘d’ represents the order of differencing of the original time series that have been taken from the original time series to make stationary. If a process y_t has an ARIMA (p, d, q) representation then $\nabla^d y_t$ has an ARMA (p, q) representation. The appropriate order of ARIMA (p, d, q) model is usually determined by the Box-Jenkins Methodology. The modelling of general real-world time series phenomenon the ARIMA models are not usually may not be much effective due to the linearity constraint imposed on the working of the model. The Box-Jenkins model selection procedure consists of the following.

3.6 Box-Jenkins Methodology

The basic theory behind the ARIMA models is the principal of parsimony. The parsimony principal should come as a second nature to the economists and financial analysists. The principal states that adding additional coefficients to the regression will necessarily increase the fit of the model, but the cost may be reduction in the degrees of freedom. Thus, Box-Jenkins argued that “*parsimonious models produce better forecasts than do the overparametrized model*”. Therefore, to select a parsimonious model the Box-Jenkins have proposed three step procedure. Their main quest is to find the more appropriate ARIMA model. These are (a) identification (b) estimation and (c) Diagnostic checking.

We know that an AR(1) an be converted to an MA(∞) thus from a lower order AR process we can generate a higher order Moving Average models similarly, from an MA(1) we can generate infinite AR models. Thus, a lower order moving average process is invertible to the higher order AR models. This generates the problem while using the ARIMA models and give rise to many issues one of them is the Identification problem. The essence of this problem is that any model may give us more than one representation and in most cases many which are essentially equivalent. Therefore, how should we choose the best one and how should it be estimated? The trick is to define a parsimony model that is a model with the smallest number of possible parameters. One may think of defining a model with high order ARMA process and then reducing by discarding insignificant parameters, but this does not work. There may be many ways within the high order model to represent the same model and the estimation procedure is unable to choose between them. We therefore have to assume the form of the model before we estimate it. In this context it is known as the identification problem and Box-Jenkins methodology starts with this problem.

3.6.1 Identification

The identification starts with the visual examination of the ACF and PACF. The visual examination of time plot of the series gives useful information to decide about the possible order of the ARMA models. Plotting observations also provide information about the possible outliers, missing observations and structural breaks in the series. As we know that almost all the economic and financial time series data is non-stationary therefore it can also provide information about the stationarity of the data. Typically, the non-stationary variables have pronounced trend or do not have constant long run mean and variance. In theory if a series is non-stationary then its ACF will not die out or not shows any sign of decay at all. Now, if this is the case then series needs to be transformed to make it stationary. A common stationarity procedure is thus to transform the series into logarithmic form and then differencing.

One the stationarity has been achieved then the next step is to identify the possible order of ‘p’ and ‘q’. This is done by following ways:

For a pure MA (q) process the ACF will shows significant coefficients up to lag ‘q’ and then dies out immediately after the lag ‘q’. The PACF of an MA (q) process will tend to die out quickly by exponentially or by a damped sine wave. Similarly, the PACF of a pure AR(p) process will tend to show spikes up to lag ‘p’ and then dies out while the ACF of an AR (p) process will dies out quickly by exponentially or by damped sine wave. If neither ACF nor the PACF shows a clear cut off points, then in this case a mixed process is suggested.it is also possible to find the AR and MA orders of the model. One can think of ACF and PACF as being superimposed to one another.

3.6.2 Estimation

In the estimation stage each of the tentative model is estimated and the coefficients of the model are examined. The estimated models are then analyzed against each other based on the SBC and AIC criterion. To choose the parsimonious model one needs to choose with smallest AIC and SBC values. The SBC is preferable of these two criteria. The SBC is preferable. The Box-Jenkins methodology necessitates that the model be stationary and invertible.

3.6.3 Unit root testing. KPSS

A prominent test for the presence of a unit root is the KPSS test. [Kwiatkowski et al., 1992] Conversely to the Dickey-Fuller family of tests, the null hypothesis assumes stationarity around a mean or a linear trend, while the alternative is the presence of a unit root.

The test is based on linear regression, breaking up the series into three parts: a deterministic trend (*βt*), a rndom walk part (*rt*), and a stationary error (*εt*), with the regression equation:

$$y_t = r_t + \beta t + \varepsilon_t$$

$$r_t = r_{t-1} + \mu t$$

Where $\boldsymbol{\mu}$~$(0,\sigma^2)$ and are *iid* (independent and identically distributed). The null hypothesis is thus stated to be H_0: $\sigma^2=0$ while the alternative is H_a: $\sigma^2>0$. Whether the stationarity in the null hypothesis is around a mean or a trend is determined by setting $\beta=0$ (in which case $\boldsymbol{x}$ is stationary around the mean r_0) or $\beta\neq0$, respectively. The KPSS test is often used to complement Dickey-Fuller-type tests.

Ho: Non-existence of unit root (Series is stationary).

H1: Existence of unit root (Series is not stationary).

There is another issue that needs to be handle when discussing about the unit root problem of the series. That is fractionally integrated process when the stationarity of the series has been confirmed by the KPSS test that is the rejection of the null hypothesis at the level will lead to the integrated process that is the series is stationary at the first difference. But, the confirmation of the fractional process is necessary before application of the univariate model because it can change the modelling strategy that is instead of applying ARIMA we need to resort to the ARFIMA instead. Therefore, another test of unit root is applied to confirm with the null hypothesis which is reverse of the KPSS null. If the series is I(1) from one test then it must be I(1) from another test. The test which is performed is ADF which is explained below.

3.6.4 Augmented Dicky Fuller Test

Stationarity of time series can be determined using ADF test. The testing procedure is applied to the model.

$$\Delta y_t = \alpha_0 + \beta t + \theta y_{t-1} + \sum_{i=1}^{k} \alpha_i \Delta y_{t-1} + et$$

where yt is the tested time series, Δ indicates the first difference, k is the lag order of the autoregressive process. It has three different variants known ADF with intercept, ADF with trend only and with intercept and trend. The hypothesis testing of the ADF test is,

Ho: Existence of unit root (not stationary).

H1: Non-existence of unit root.

3.6.5 Diagnostic Checking

In the diagnostic stage the goodness of the fit of the model is examined. The standard practice is to plot the residuals of the model and to look for the outliers and evidence of the periods in which the data does not fit well. The statistic which are observed in this stage are the Box-Pierce Statistics and L-Jung Box Q-statistics (1979) which serve to detect the Autocorrelation in the residuals. The Box-Jenkins Approach can be summarized in the following diagram.

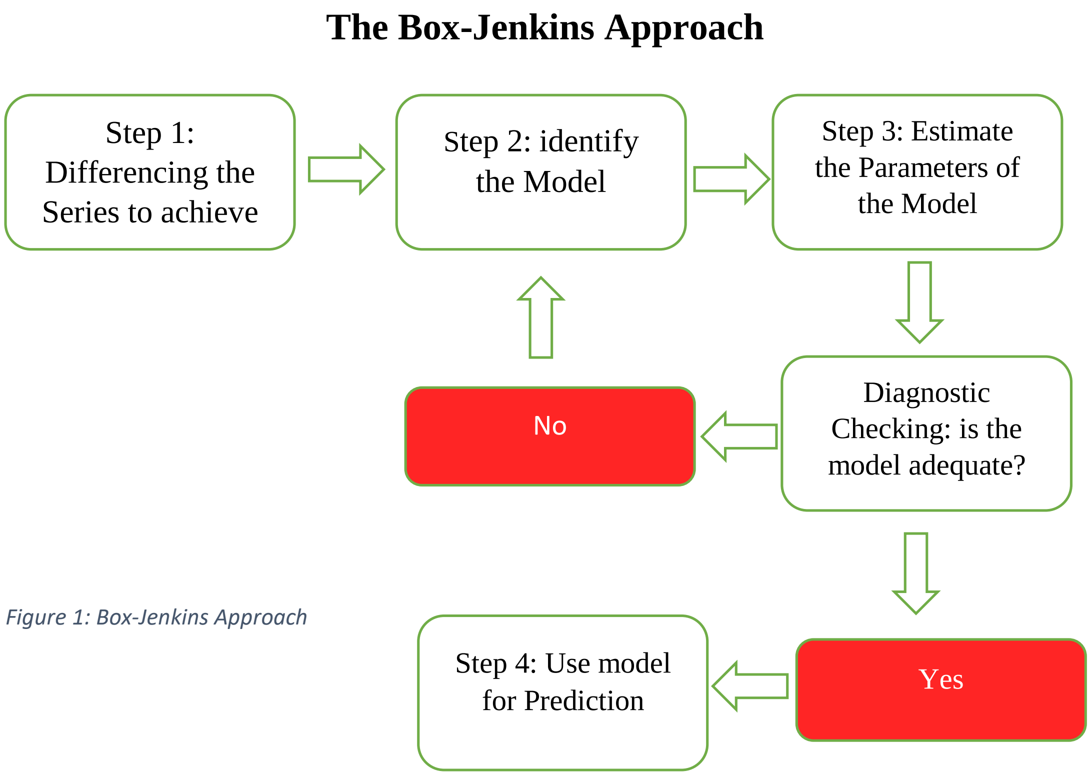


Figure 1: Box-Jenkins Approach

So far, the discussion involved the constant variance of the error terms as with the conventional econometric methodology it is assumed that the variance is constant that the assumption of the homoscedasticity of the error variance. But there are many financial and economic time series which do not involve the constancy of the variance of the error term. There are many time series which shows the period of high volatility clustering followed by the low period of clustering. Therefore, such time series modelling the conventional principal of constancy of the error terms is violated as a result the variance becomes heterogeneous. In such cases it is preferable to examine the patterns which allow the dependency of variance on its history. Suh models are known as the models with conditional heteroscedasticity. The univariate models which incorporate these conditional variances are known as Auto-regressive Conditional heteroscedastic models here the brief description of these models is presented for the purpose of the data analysis in the next chapter. These are as follows.

3.7 The ARCH Model

The ARCH model is attributed to the Robert F. Engle.[1] The model presented by the Engel suggests that the variance of residuals at time 't' depends on the square of the residuals from the past period. Thus, he suggested that it is better to model the mean and variance simultaneously when it is suspected that the conditional variance of the series is not constant. If we let us assume that

$$\boldsymbol{y_t = \alpha + \beta' X_t + e_t}$$

[1] The Robert F. Engle is attributed to the ARCH model who in his seminal paper entitled "Autoregressive Conditional Hetroskedasticity with the estimates of the variance of the United Kingdom inflation "published in Econometrica in 1982 open the new era foe financial modelling.

where,X_t is a K×1 vector of the explanatory variable and $\boldsymbol{\beta'}$ is a k×1 vector of the unknown coefficients. It is assumed normally that the residuals $\boldsymbol{e_t}$ distributed normally as zero mean and constant variance. In mathematical term this can be written as:

$$e_t \sim iidN(0, \sigma^2)$$

The Engels's idea starts by allowing for the heteroscedastic variance. One way to incorporate this idea is to allow the variance to depend on the one time lagged of the square of the error term that is:

$$\sigma_t^2 = \delta_o + \delta_1 \boldsymbol{e_{t-1}^2}$$

which is the basic ARCH(1) process.

The ides of the ARCH type model suggests that it will simultaneously model the mean and variance of the series with the following specification.

$$\boldsymbol{y_t = \alpha + \beta' X_t + e_t} \ \ldots \mathbf{14}$$

$$e_t \sim iidN(0, h_t)$$

$$h_t = \delta_o + \delta_1 \boldsymbol{e_{t-1}^2} \ \ \ldots \mathbf{15}$$

The equation (14) is the mean equation and (15) is the variance equation. Thus, the ARCH model simultaneously models the mean and variance equation. The estimated coefficient δ_1 must be positive for the positive variance. The model written above is a simple ARCH (1) model but it may be that the conditional variance may depend on the more than one lagged square error. Therefore, the idea of general ARCH(q) model is generated. If let us assume that the conditional variance depends on the 'q' lagged squares error term then the underlying model is written as:

$$h_t = \delta_o + \delta_1 \boldsymbol{e_{t-1}^2} + \delta_2 \boldsymbol{e_{t-2}^2} + \cdots + \delta_q \boldsymbol{e_{t-q}^2}$$

$$= \delta_o + \sum_{i=1}^{q} \delta_i \, \boldsymbol{e_{t-i}^2}$$

Thus, the ARCH(q) model will simultaneously model the mean and variance of the serries according to the following specification:

$$\boldsymbol{y_t = \alpha + \beta' X_t + e_t}$$

$$e_t \sim iidN(0, h_t)$$

$$h_t = \delta_o + \sum_{i=1}^{q} \delta_i \, \boldsymbol{e_{t-i}^2} \qquad \mathbf{16}$$

The coefficients $\delta_i's$ has to be positive for the variance to be positive. For the application purpose it is necessary to check for the possible presence of the ARCH effects in order to confirm which model requires the ARCH estimation instead the OLS. The simple test can be performed along the lines of Breusch-Pagan test which requires the estimation of mean equation.

$$\boldsymbol{y_t = \alpha + \beta' X_t + e_t}$$

obtain the residuals from the mean equation above by estimating it through OLS. an auxiliary regression is thus performed which consists of the squared residuals $\hat{u}^2$ on the lagged squared residual including the intercept. That is

$$\hat{u}_t^2 = \delta_0 + \delta_1 \hat{u}_{t-1}^2 + \delta_2 \hat{u}_{t-2}^2 + \ldots + \delta_q \hat{u}_{t-q}^2 + \epsilon_t$$

compute the ($T \times R^2$). Under the null hypothesis of no arch effects that the variances are all same or constant. The statistic ($T \times R^2$) has a χ^2 distribution with q degree of freedom. The rejection of null hypothesis suggests the evidence of the ARCH effects model. The ARCH(q) models are useful when the variability of series is expected to change more slowly than the ARCH (1) model. The ARCH models are quite difficult to estimate because they may yield the negative coefficients $\delta_i's$. To overcome this issue Tim-Bollerslev (1986) developed the idea of the GRACH type Modelling.

3.8 The GARCH Model

According to the Engel the drawback of the ARCH models was that it looks more like the moving average than the autoregressive, from this criticism a new idea was born to include the lagged conditional variance term as the Autoregressive term. This idea was worked out by the Bollerslev (1986), who published a paper entitled "Generalized Conditional Hetroskedasticity in the Journal of Econometrics" introducing a new family of GARCH type models. The general GARCH (p. q) model is presented as:

$$\boldsymbol{y_t = \alpha + \beta' X_t + e_t} \quad \ldots \mathbf{17}$$

$$e_t \sim iidN(0, h_t)$$

$$h_t = \delta_o + \sum_{i=1}^{p} \gamma_i \, \boldsymbol{h_{t-i}} + \sum_{i=1}^{q} \delta_i \, \boldsymbol{e_{t-i}^2} \quad \ldots \mathbf{18}$$

This model now state that the variance parameter that is h_t now depends on the lagged squared residuals and lagged values of itself. The former is captured by the lag error terms in the model and later is captured by the lagged h_t terms. When (p=0) the GARCH model reduces to the ARCH model. The simplest form of the GARCH (p, q) model is the GARCH (1,1) model whose variance equation is written as:

$$h_t = \delta_o + \gamma_1 \boldsymbol{h_{t-1}} + \gamma_2 \boldsymbol{e_{t-1}^2}$$

The GARCH (1,1) model is parsimonious alternative to the ARCH (q) process because with the GARCH (1,1) infinite ARCH process can be estimated. Therefore, it is theoretically established that instead estimating the higher order ARCH models it is essential to estimate the GARH (1,1) model and that few parameters to estimate and few degrees of freedom are lost. It is to note that the ARIMA (p, d, q) models capture the linear part of the data generating process of the series then GARCH (p, q) model is there to capture the non-linearity in the residuals. The best ARIMA (p, d, q) model together with the GARCH type models thus gives the complete picture of the data generating process of the underlying series thus in case of oil prices series the both version of linear and non-linear are combine together to find the best model for the purpose of improvement in the forecast. This is described as:

3.9 Hybridization (ARIMA-GARCH Model)

Let's assume an ARIMA (p, d, q) model of the form:

$$\nabla^d y_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \, \nabla^d y_{t-i} + \sum_{j=1}^{m} \beta_j \varepsilon_{t-j} + \varepsilon_t$$

Where, 'd' represents the order of differencing of the original time series that have been taken from the original time series to make stationary. If a process y_t has an ARIMA (p, d, q) representation then $\nabla^d y_t$ has an ARMA (p, q) representation. The rationale for the GARCH type model is that the errors from the ARIMA model above contains the ARCH effects thus the variance of the series conditions on the past realization that is captured by the GARCH (p, q) model. Thus, the linear part of the series is captures by the ARIMA model and then the non-linear part which is contained in the errors is captured by the GARCH model. Thus, it develops the rationale for ARIMA-GARH hybrid model given mathematically as.

$$\nabla^d y_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \, \nabla^d y_{t-i} + \sum_{j=1}^{m} \beta_j \varepsilon_{t-j} + \sum_{i=1}^{p} \gamma_i \, \boldsymbol{h_{t-i}} + \varepsilon_t \quad \dots \ 19$$

Where, $h_t = \delta_o + \sum_{i=1}^{p} \gamma_i \, \boldsymbol{h_{t-i}} + \sum_{i=1}^{q} \delta_i \, \boldsymbol{e_{t-i}^2}$

The ARIMA part = $\alpha_0 + \sum_{i=1}^{r} \alpha_i \, \nabla^d y_{t-i} + \sum_{j=1}^{m} \beta_j \varepsilon_{t-j}$

The GARCH part =$\sum_{i=1}^{p} \gamma_i \, \boldsymbol{h_{t-i}} + \sum_{i=1}^{q} \delta_i \, \boldsymbol{e_{t-i}^2}$ and errors ε_t thus assumed independent and identically distrusted. This models the residuals from the best *ARIMA* model using the GARCH method together with the ARIMA as the mean equation.

3.10 Machine Learning Models

Neural Networks is one of the most popular machine learning algorithms at present. It has been decisively proven over time that neural networks outperform other algorithms in accuracy and speed. With various variants like CNN (Convolutional Neural Networks), RNN (Recurrent Neural Networks), Deep Learning etc. neural networks are slowly becoming for data scientists or machine learning practitioners what regression analysis was one for statisticians. It is thus imperative to have a fundamental understanding of what a Neural Network is, how it is made up and what is its reach is described as under.

3.10.1 What is a Neuron?

As the name suggests, neural networks were inspired by the neural architecture of a human brain, and like in a human brain the basic building block is called a Neuron. Its functionality is similar to a human neuron, i.e. it takes in some inputs and fires an output. In purely mathematical terms, a neuron in the machine learning world is a placeholder for a mathematical function, and its only job is to provide an output by applying the function on the inputs provided.

3.10.2 Anatomy of a Neural Network:

Training a neural network revolves around the following objects:

- Layers, which are combined into a network or (model)
- The input data and corresponding targets
- The loss function, which defines the feedback signal used for learning
- The optimizer, which determines how learning proceeds

3.10.3 DATA and Sources:

The data on Crude oil prices in current US$ was used. The data was retrieved from West Texas intermediate (WTI). There are 2520 observations on 10 year daily crude oil prices. The link of data source is given as under: www.macrotrends.net/2516/wti-crude-oil-prices-10-year-daily-chart (URL). The data trends are shown as below:

Neural Network model of Machine Learning

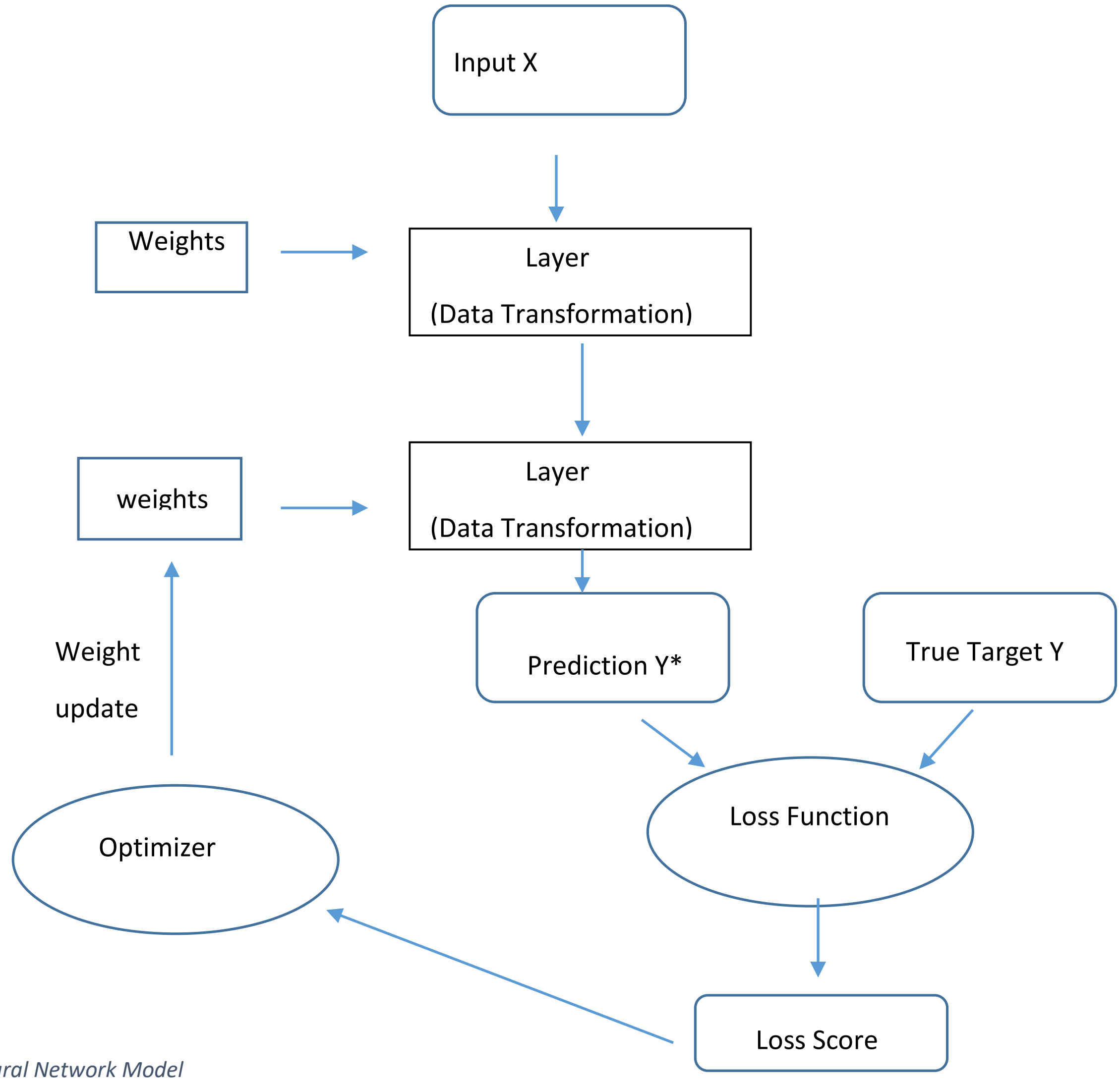


Figure 2 Neural Network Model

The network, composed of layers that are chained together, maps the input data to predictions. The loss function then compares the predictions to the targets, producing a loss value: a measure of how well the network's predictions match that was expected. The optimizer uses this loss value to update the network's weights.

3.11 Recurrent Neural Network (RNN)

Biological intelligence processes information incrementally while maintain an internal model of what it's processing, built from past information and constantly updated as new information comes in. A recurrent neural network (RNN) adopts the same principle, albeit in an extremely simplified version: it processes sequences by iterating through the sequence elements and maintaining a state containing information relative to what it has seen so far. In effect, an RNN is a type of neural network that has an internal loop.

A major characteristic of all neural networks has been observed as a densely connected networks and convents, is that they have no memory. Each input show to them is processed independently, with no state kept in between inputs. With such networks, in order to process a sequence or a temporal series of data points, you have to show the entire sequence to the network at once: turn into a single data point. The entire dataset transformed into a single large vector and processed in one go. Such networks are called 'feedforward networks.

The state of the RNN is reset between processing two different, independent sequences (such as two different reviews), so you still consider one sequence a single data point: a single input to the network. what changes is that this data point is no longer processed in a single step; the network internally loops over sequence elements. This RNN takes as input a sequence of vectors, which you will encode as a 2D tensor of size (timesteps, input features). It loops over timesteps, and each timestep, it considers its current state at 't' and the input at 't' (of shape (input_features,), and combines them to obtain the output at 't'. you'll then set the state for the next step to be this previous output. For the first timestep, the previous output isn't defined; hence, there is no current state. So, first initialize the state as an all-zero vector called the ‘initial state’ of the network.

Recurrent Neural Network

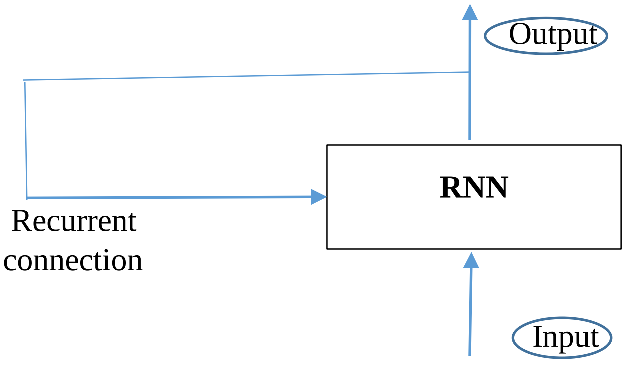


Figure 3: Recurrent Neural Network

Anatomy of Recurrent Neural Network

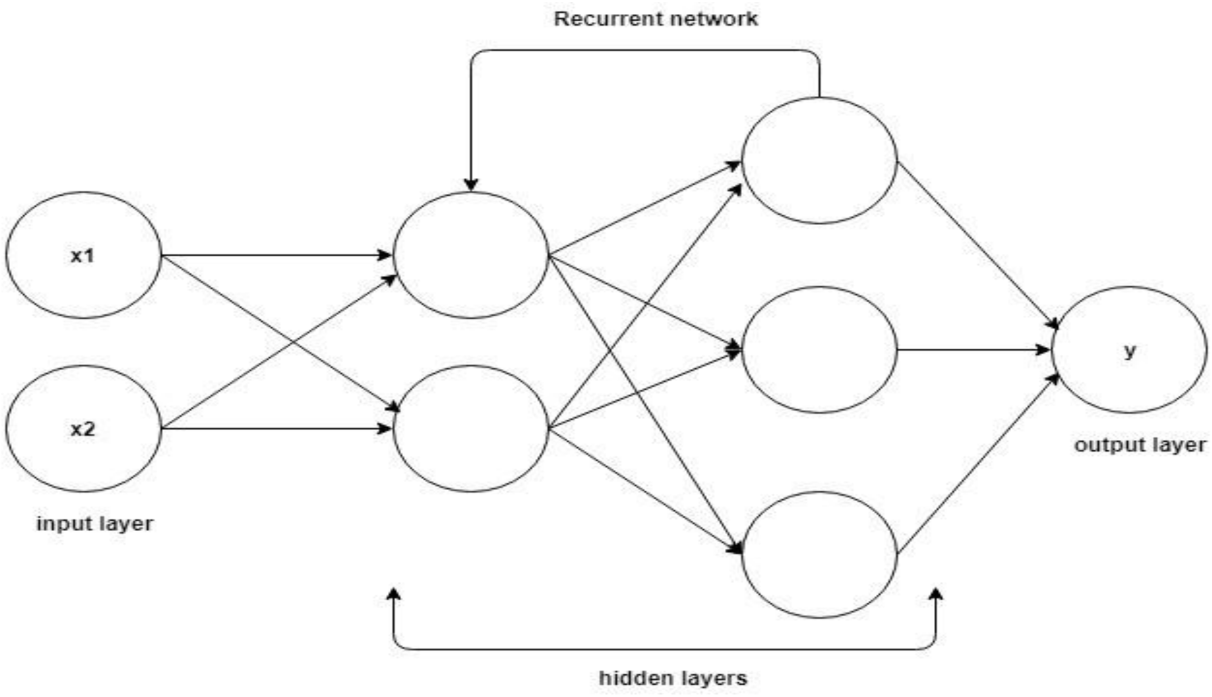


Figure 4: Anatomy of RNN

3.12 RNN (Long Short-Term Memory):

As the simple RNN is too simplistic to be of real use. Simple RNN has a major issue: although it should theoretically be able to retain at time ‘t’ information about inputs seen many timesteps before, in practice, such long-term dependencies are impossible to learn. This is due to ‘vanishing gradient problem’, an effect that is similar to what is observed with non-recurrent networks (feedforward networks) that are many layers deep: as if keep adding layers to a network, the network eventually becomes untrainable. The theoretical reason for this effect were studied by ‘Hochreiter, Schmidhuber, and Bengio’ in the early 1990’s. The LSTM and GRU layers are designed to solcve this problem.

The LSTM (Long Short-Term Memory) algorithm was developed by Hochreiter and Schmidhuber in 1997, it was the culmination of their research on the vanishing gradient problem. This layer is a variant of the simple RNN layer. RNN (LSTM) works better on long sequences than a simple or naïve RNN. So, in this way the advanced features of RNN (LSTM) can help to get most of deep-learning models.

It adds a way to carry information across many timesteps. Imagine a conveyor belt running parallel to the sequence that is processing. Information from the sequence can jump onto the conveyor belt at any point, be transported to a later timestep, and jump off, intact, when need it. This is essentially what LSTM does: it saves information for later, thus preventing older signals from gradually vanishing during processing.

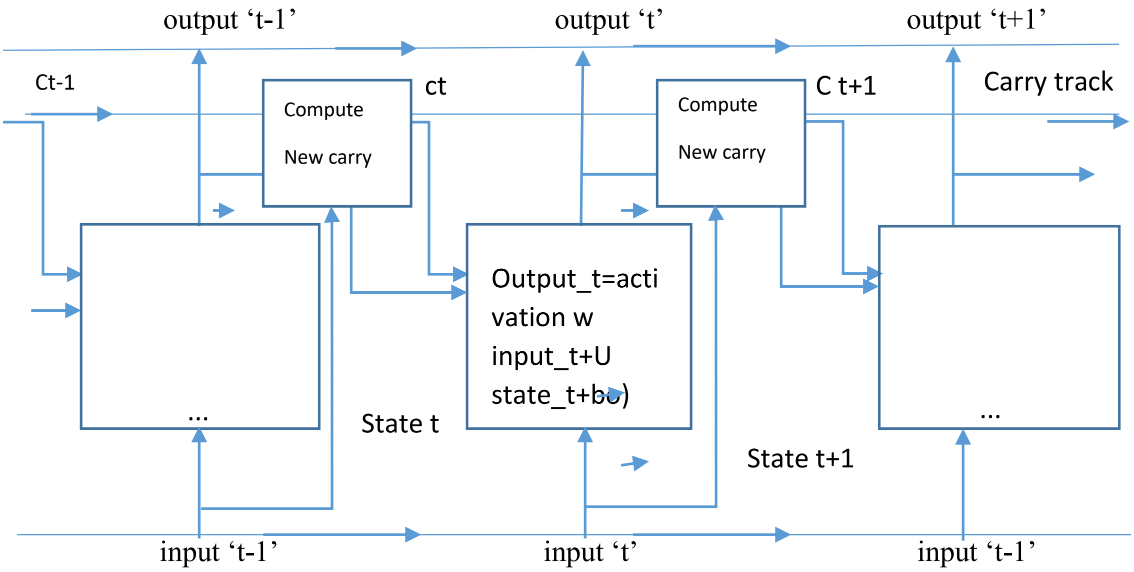


Figure 5: Anatomy of RNN(LSTM)

There are a lot of weight matrices, index the 'W' and 'U' matrices in the cell in the above figure with letter 'o' (W_o and U_o) for output. The data flow that carries information across timesteps. Call its values at different timesteps 'Ct' where 'C' stands for carry. This information will have the impact on the cell: it will be combined with the input connection and the recurrent connection (via a dense transformation: a dot product with a weight matrix followed by a bias add and the application of an activation function), and it will affect the state being sent to the next timestep (via an activation function and a multiplication operation). Conceptually, the carry dataflow is a way to modulate the next output and the next state simple so far.

The interpretation for what each of these operations is meant to do. For instance, the multiplication of c_t and f_t is a way to deliberately forget irrelevant information in the carry dataflow. Meanwhile, i_t and k_t provides information about the present, updating the carry track with new information. But at the end of the day, these interpretations don't mean much, because what these operations actually do is determined by the contents of the weights parameterizing them; and the weights are learned in an end-to-end fashion, starting over with each training round, making it impossible to credit this or that operation with a specific purpose. The specification of an RNN cell determines the hypothesis space- the space in which search for a good model configuration during training, but it doesn't determine what the cell does; that is up to the cell weights. The same cell with different weights can be doing very different things. So, the combination of operations making up an RNN cell is better interpreted as a set of constraints on search, not as a design in an engineering sense.

The choice of such constraints is based upon the implementation of RNN cells for the better optimization algorithm (like genetic algorithms or reinforcement learning processes) than to human engineers. There is no need to understand anything about the specific architecture of an LSTM cell. The LSTM cell is meant to do allow past information to be reinjected later, thus fighting the vanishing gradient problem.

The performance of RNN(LSTM) can be improved by tuning hyperparameters such as the embeddings dimensionality or the LSTM output dimensionality. Another may be lack of regularization. But honestly, the primary reason is that analysing the global, long-term structure of the reviews (what LSTM is good at) is not helpful for a sentiment-analysis problem.

3.13 Performances Evaluation:

- The performances of modelling and forecasting hybrid model and GARCH (1,1) model will be evaluated using Akaike's information criterion (AIC), mean absolute error (MAE) and root mean square error (RMSE). The formula are as follows:
- $AIC = -2\log\left[\sigma^{\wedge 2}\right] + 2K$
- MAE = $\frac{\sum_{i=0}^{n} ABS(y_t - \hat{y_t})}{n}$
- $RMSE = \sqrt{\frac{\sum_{i=0}^{n}(y_t - \hat{y_t})}{n}}$

- where $\sigma^{\wedge 2}$ = estimated model error variance;
- k = number of free parameters in the model,
- y_t= actual value,
- $y_t^{\wedge}$ = estimate.

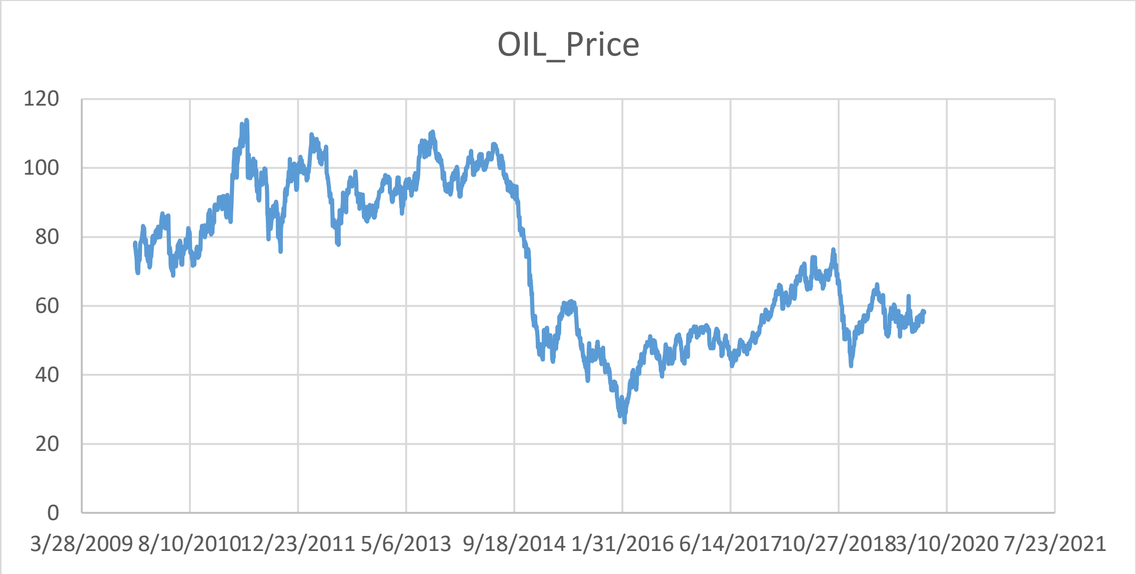
CHAPTER 4

DATA ANALYSIS

4.1 Introduction

This chapter includes the analysis of the time series data in two sections. First section explains all the results obtained by using conventional time series models and second section is all about the machine learning model RNN(LSTM). Now for the identification of the time series properties of the data and identification of the appropriate ARMA process. In the end the ARIMA-GARCH model will be estimated and forecast from this hybrid model will be analyzed through various evaluation statistics.to compare the RNN(LSTM) and ARIMA- GARCH this univariate analysis will be comparatively examined and finally a better model will be explained based on prediction capability.

4.2 Trends in Series over the Time



4.3 Identification of Appropriate order of AR&MA and Explanation of Results

The order '*p'* of an auto regressive time series is unknown empirically; it must be defined empirically it is known as the order determination of the AR process. There has been extensive literature on this issue. There are two approaches in general which are available for the determination of order of an AR process. These are Partial autocorrelation function and using some information criteria. The Partial autocorrelation function cuts-off beyond the lag 'p' therefore, it is customary to draw the PACF of the stationary data series. Let we have the following graph of the raw series of the oil prices.

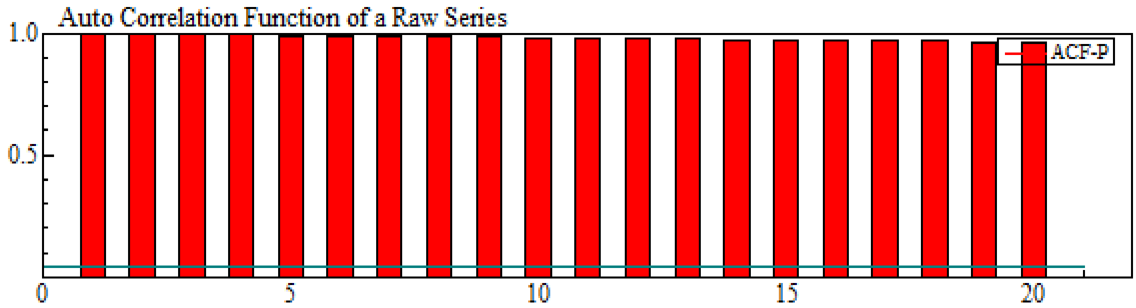


Figure 6: sample Auto correlation function (ACF) of raw data series (Oil prices US $) from 2009 to 2019 on daily basis for five working days

The ACF of a raw series does not break off at any lag thus it is clear that the series is not stationary and needs to be differenced to make it stationary for the forecasting and estimation purpose. Further it can be shown in the following diagram that the series is trending and there are changes in both level and slope.

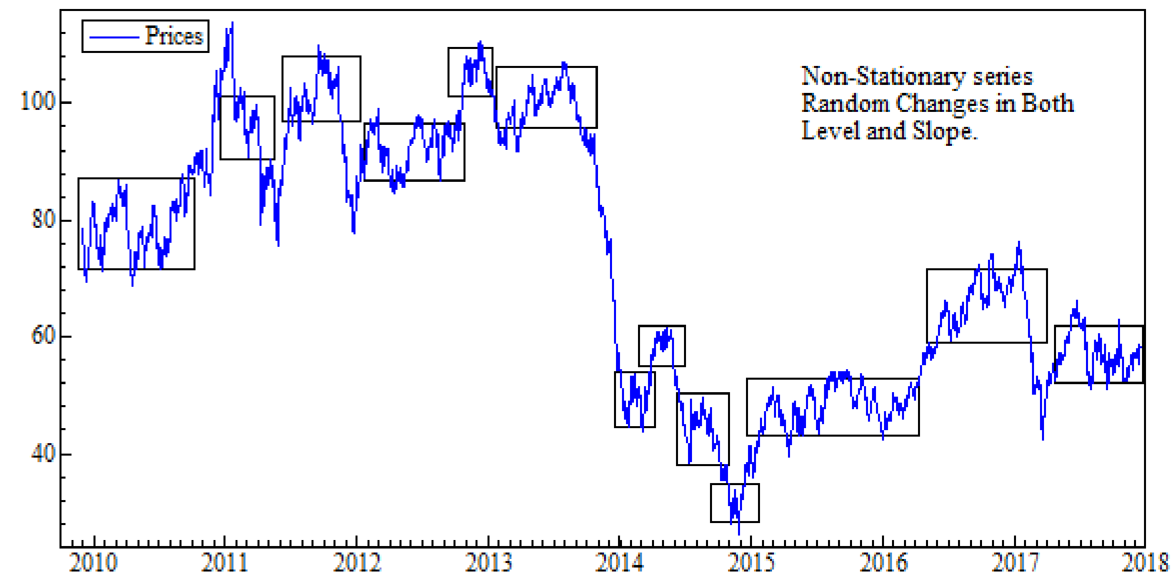


Figure 7: A non-stationary series with changes in both level and slope

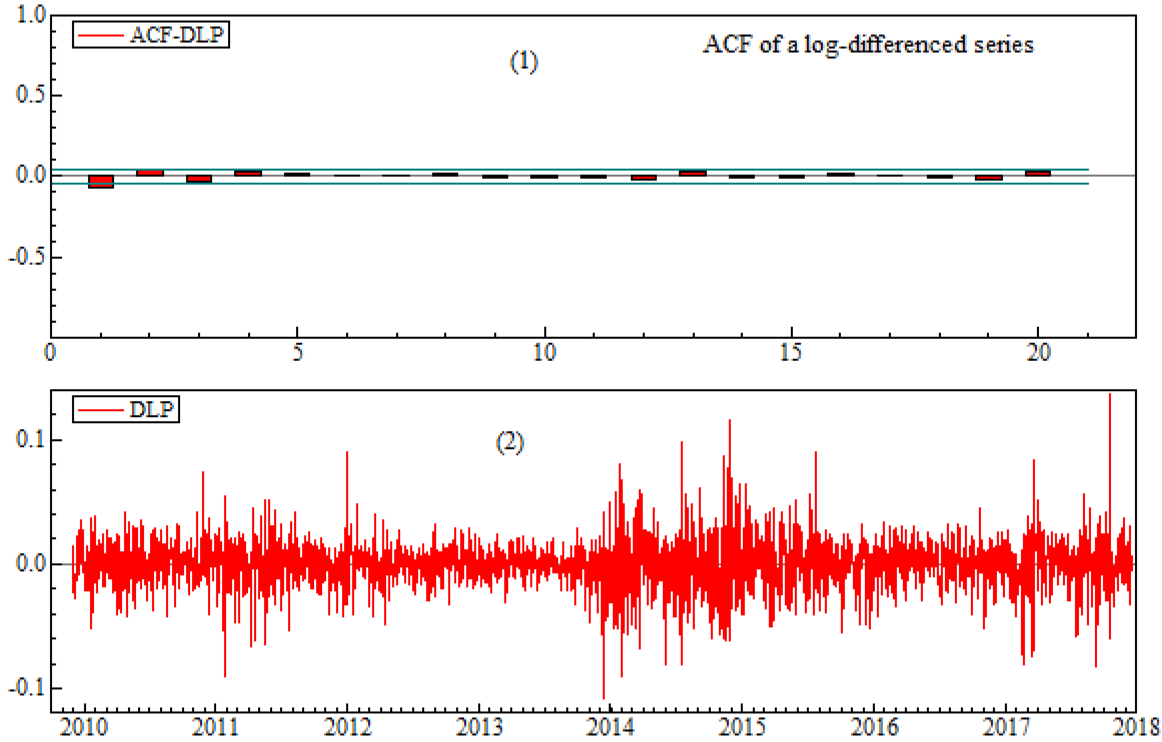
The logarithm and then first difference of the series was taken to make it stationary the resulting output is graphed below along with the ADF and KPSS results.

Table 4.1 ADF and KPSS Unit Root Tests

	ADF	Level	KPSS	
Variables	t-statistics	Critical value	t. Statistic	Critical value
lnP	-2.1891	-3.4115	0.7478	0.1460
		1st Difference		
Lnp	-52.4958	-1.9409	0.0760	0.1460

The table above shows that the data series is integrated of order one that is I (1) of differenced stationary. The null hypothesis of ADF is rejected at 5% level of significance as the estimated value of ADF is greater than the table value at level. Similarly, at first difference we can reject the null hypothesis of unit root. Similarly, the null hypothesis of KPSS that is the series is stationary is rejected at 5% level of significance as the estimated value of KPSS is greater than the critical at level thus the conclusion is that the series is subjected to unit root problem. The null of KPSS cannot be rejected at first difference thus it shows that the series is first differenced stationary. This also confirm that the series is not subjected to fractional integration. The stationarity of the series can also be confirmed from the following figure.

Figure 8: ACF of a log-diff. series:



The fig 4.3 above shows that the ACF of log-differenced series is decaying after lag 3 this shows that the time series is stationary further the (2) in the figure above also has shown that the series fluctuate around the mean value of zero. Therefore, it confirms that the series is stationary. As first difference of series makes the data stationary therefore, it follows an integrated path. The next task is to find the appropriate AR process for this the PACF of the series is drawn here.

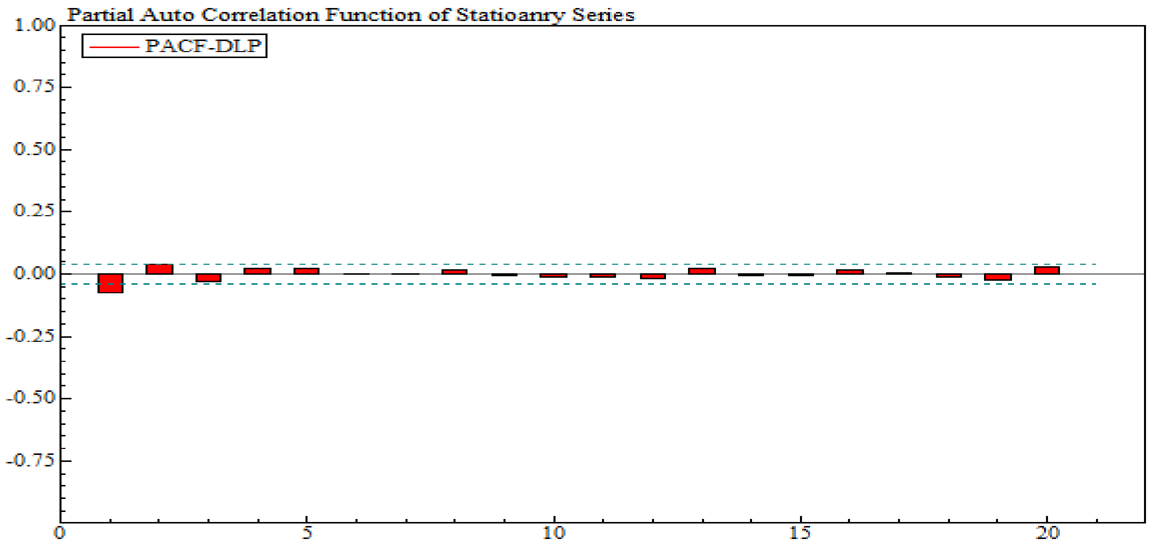


Figure 9: Partial Auto correlation function of an integrated series which confirms the possible AR(p) process of the oil prices.

The Dotted lines in the figure above indicate approximately two standard error limits (2S.E). the plot suggests maximum of an AR (4) model for the data because at 3rd lag the PACF of the series dies out. The criterion-based solution calls for the selection of AR terms with minimum of the criterion value. To determine the appropriate order for the moving average terms we need to draw the ACF of the integrated series which is given below.

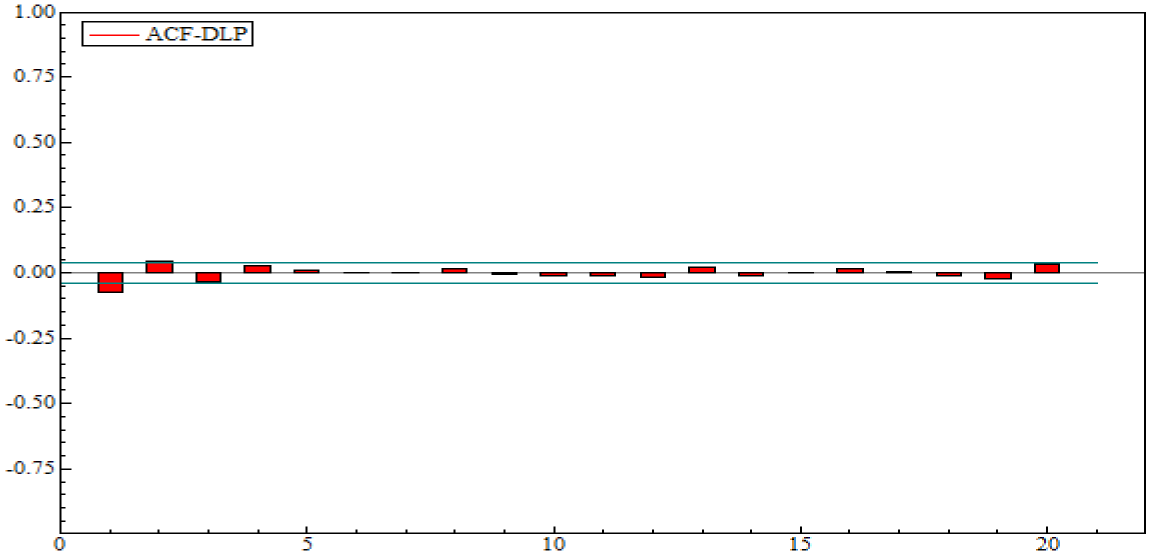


Figure 10: The ACF of an integrated series which shows the appropriate order of the MA model.

The diagram above shows the autocorrelation function of the integrated series the appropriate order of the MA model is thus MA (4) because after that the Autocorrelation function of the series breaks off. After identifying the appropriate order of the AR and MA terms the next step involves the identification of the ARIMA model. One thing that is to remember is that we need to estimate a parsimonious ARIMA model to be able to avoid the overfitting of the model. If the decision of the ARIMA order is kept on the visual examination of the ACF and PACF of the MA and AR models then it can be ARIMA (3,1,3) or ARIMA (4,1,4) or any combination of the these may be regarded as the appropriate order of the ARIMA (p, d, q) model but, the PAC and ACF are not appropriate to determine the appropriate order of the ARIMA models.

Tsay and Tiao (1984) proposed a new approach that uses the extended auto correlation function to determine the appropriate order of the ARMA models. The basic idea of this approach is relatively simple that is "if we have estimated the consistent estimates of the AR components of the ARMA models then we can derive the MA component and from this derived MA series the ACF can be used to identify the MA component (Tsay and Tiao, 1984). Therefore, the possible estimates of the ARIMA (p, d, q) models re given below along with the other fit statistics.

Table 4.2 Regression Results of An ARIMA (3,1,3) Model

Variables	Coefficient	Std.Error	t. Statistic	Probability
AR (1)	-0.756058	0.11907	-6.349717	0.0000
AR (2)	-1.091342	0.023598	-46.24654	0.0000
AR (3)	-0.566647	0.117354	-4.828531	0.0000
MA (1)	0.68876	0.125663	5.481014	0.0000
MA (2)	1.089924	0.024754	44.02937	0.0000
MA (3)	0.496563	0.125251	3.96456	0.0001
D	-0.60679	0.30681	-1.97481	0.0000
SE Regression	0.02038	Akaike info Criterion		-4.94544
SS residuals	1.0433	Schwartz Criterion		-4.92923
Log likelihood	6235.784	Hanan-Quin Criterion		-4.93956
D. Watson Stat	2.0011			

Table 4.2: the 'd' is the integrated part of the ARIMA model. It indicates that the series is stationary at first difference. The statistic is significant.

The results of the ARIMA (3, 1, 3) are given above. The table shows AR and MA coefficients are highly significant and that the tests on residuals also show that the model is stable, and residuals are independent and identically distributed. The Box-LJung statistics for the (Q-stat) is also insignificant for the ARIMA (3,1,3) model. The other possible ARIMA models have also been estimated but, they have higher standard errors of regression and that the SIC and AIC criterion is minimum only at the ARIMA (3,1,3) model. Therefore, the ARIMA (3,1,3) is regarded as the parsimonious model. The Box-Ljung Statistic for the $Q(1) = 0.012\ (0.972, Q(2) = 0.0405\ (0.980), Q(3) = 0.0625(0.996)$ which are all insignificant suggesting the same ARIMA model. Now this model can be used to predict the future values. The results of the forecasts are given below:

Table 4.3: Results of forecast statistic from An ARIMA (3,1,3) Model

RMSE	MAE	MAPE	Theil Inequality Coeff
0.02048	0.01479	100.2691	0.994141

If we analyses the series, there is volatility clustering in the data series. There are periods of high volatility followed by another period of high volatility and low volatility followed by another low volatility clustering. This suggests that there is an ARCH process in the data series which are estimated by the ARIMA-GRACH model the estimates are given below.

Table 4.4 Regression Results of GARCH (1,1)

Variables	Coefficient	Std.Error	t. Statistic	Probability
AR (1)	-0.734471	0.172492	-4.258006	0.0000
MA (1)	0.692786	0.184116	3.762769	0.0002
		Variance Equation		
C	0.00028	0.00078	3.5705	0.0004
ARCH (1)	0.08673	0.01844	4.7011	0.0000
ARCH (2)	0.03294	0.01876	1.7554	0.0792
GARCH (1)	0.9411	0.00631	148.737	0.0000
SE Regression	0.02041	Akaike info Criterion		-5.1280
SS residuals	1.4909	Schwartz Criterion		-5.1142
Log likelihood	6464.835	Hanan-Quin Criterion		-5.1230
D. Watson Stat	2.0542	RMSE(Forecast)		

The table shows the output of the GARCH (1,1) model. It is estimated for various orders and the decision is based on the minimum value of the AKAIKE information criterion. Thus, GARCH (1,1) gives the minimum value of the AIC. The output shows that the coefficients are highly significant except the ARCH (2) term that is also significant if we set the significance level at 10%. Thus, the model estimated is a valid model for prediction analysis. The prediction statistics are given below.

Table 4.5: Results of forecast statistic from GARCH (1,1) Model

RMSE	MAE	MAPE	Theil Inequality Coeff
0.02047	0.01479	99.9263	0.994141

The forecast from the GARCH (1,1) model is better than the ARIMA (3,1,3) model. The estimates from the ARIMA-GARCH model are given below. The hybrid model thus consists of the ARIMA as the mean equation and then non-linearity in the residuals captured by the GARCH model. Thus, a combined hybrid model is given below.

Table 4.6: Regression Results of an ARIMA-GARCH

Variables	Coefficient	Std.Error	t. Statistic	Probability
AR (1)	-0.741280	0.124248	-5.966127	0.0000
AR (2)	-1.089243	0.024445	-44.55928	0.0000
AR (3)	-0.552010	0.122388	-4.510339	0.0000
MA (1)	0.672377	0.130952	5.134518	0.0000
MA (2)	1.087152	0.025686	42.32431	0.0000
MA (3)	0.480417	0.130486	3.681761	0.0002
ARCH (1)	-5.86992	3.075348	-1.90870	0.0564
GARCH (1)	4.33704	2.17967	2.17934	0.0294
SE Regression	0.02037	Akaike info Criterion		-5.9557
SS residuals	1.03786	Schwartz Criterion		-5.5143
Log likelihood	6221.543	Hanan-Quin Criterion		-5.7237
D. Watson Stat	2.0022			

The ARIMA (3,1,3) and GARCH (1,1) model has been estimated. The results above shown that all the coefficients are significant except only one coefficient which is significant at 10%. The AIC has value of -5.9557 which minimum of all the previous models. Similarly, all other statistics of the ARIMA-GARCH model has been improved. Thus, it is a better model as compared to earlier estimated. The residuals diagnostics and other forecast statistics are given in the below page. These statistics also show that the ARIMA-GARCH model is better as compared to the ARIMA and GARCH alone. The fit statistics are given as follows;

Table 4.7: Tests for Fit based on ARIMA-GARCH

		Statistics	Probability Value
LM ARCH Test		179.5258	0.9997
Jarque Bera		1006.60	0.0000
Ljung-Box test	Q (1)	0.0912	0.972
Ljung-Box test	Q (2)	0.0405	0.980
Ljung-Box test	Q (3)	0.0625	0.996
Ljung-Box test	Q (4)	0.5256	0.971
Ljung-Box test	Q (5)	1.1445	0.950
Ljung-Box test	Q (10)	2.7057	0.988
Ljung-Box test	Q (15)	4.5981	0.995

Forecast Evaluation

RMSE	0.01935
MAE	0.0144
Theil Inequality	0.7124

All the test statistics are insignificant and there is no serial or auto correlation in the standardized residuals. The Ljung-Box statistics has confirmed that there is no ACF/PACF in the residuals obtained from the ARIMA-GARCH model at various lags order as given above in the table. The forecast valuation also has confirmed that the forecast errors obtained from the ARIM-GARCH model are minimum as compared to the other models

The Graph of the Actual and Fitted series and forecast standard errors are given below which also confirm the superiority of the ARIMA-GARCH model over the ARIMA and GARCH alone. Therefore, if we compare these three models the ARIMA-GARH has higher predictability ability but as the objective of the study is concerned about another machine learning model named as RNN(LSTM) then compare the predictions by using informal (graphical) methods and other performance evaluation criteria's. In the next section the following model will be estimated and then the comparative forecasting ability will be judged. The one which is more accurate and suitable based on theoretical background and its forecasting ability that is RNN(LSTM) or ARIMA-GARCH. This section will revolve around this issue.

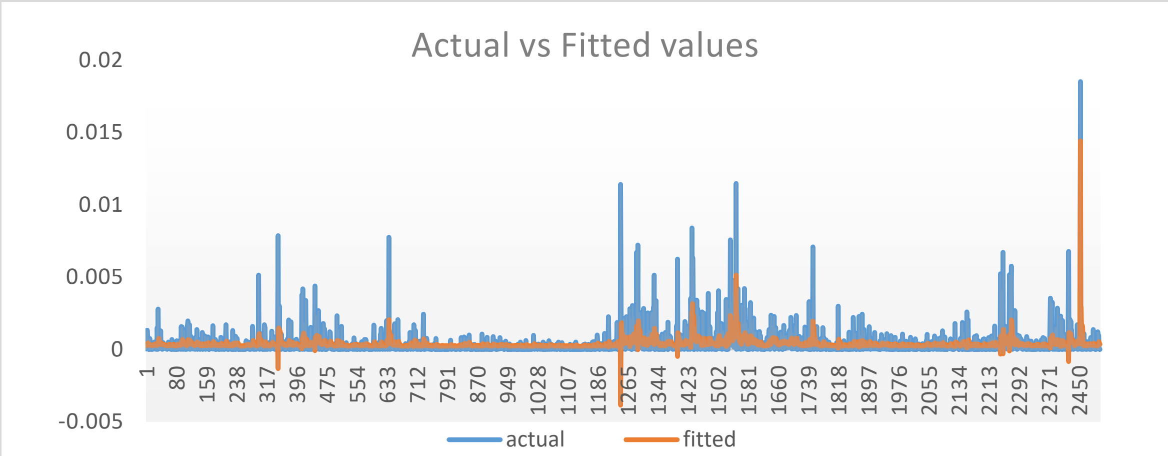


Figure 11: The actual series of Pakistan crude oil prices with the fitted values. The data is in the log-difference form

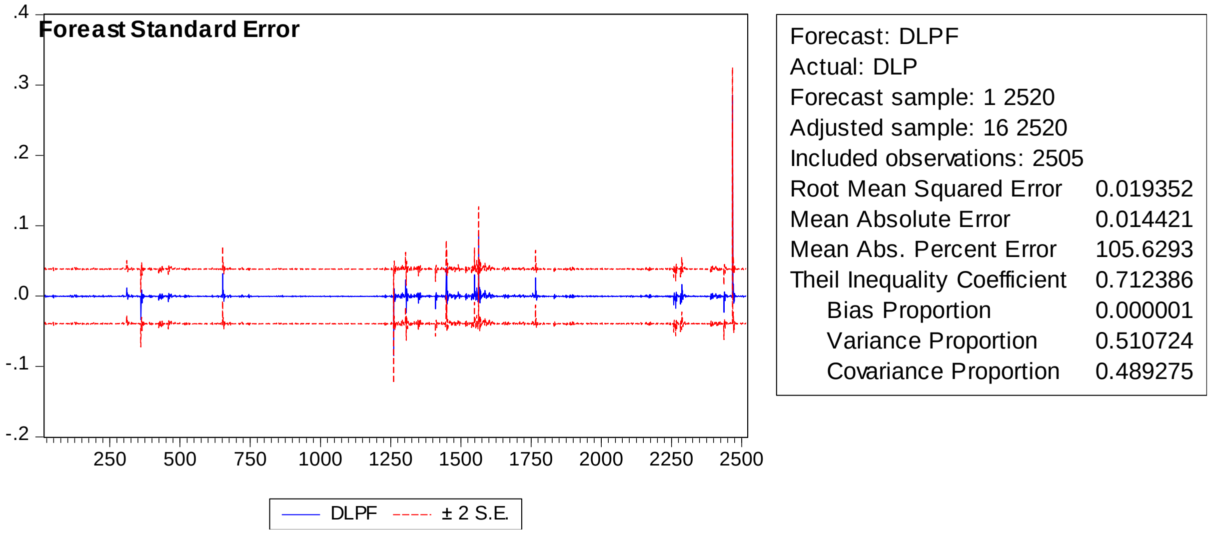


Figure 12: The figure shows the forecast errors of the fitted model. The bands with red colours indicate approximately 2SE of the data.

The forecast from this model ARIMA-GARCH is better than all other models the graphs of the actual and forecasted series also represents the same phenomenon in the above mentioned figure.

4.4 Results of RNN(LSTM)

The forecasts from the RNN (LSTM) model are given below. The RNN (LSTM) being more unrestrictive, flexible regarding hard core restrictions (assumptions) of conventional time series modelling like stationarity, normality etc. It is easy to design the architect, train the dataset on the designed algorithm of RNN (LSTM) to learn the pattern and tuning the hyperparameters then predict values and thus to implement. So, in this way as it is not based on hard and fast assumptions. Therefore, the forecasts statistics from this estimation technique are given as under:

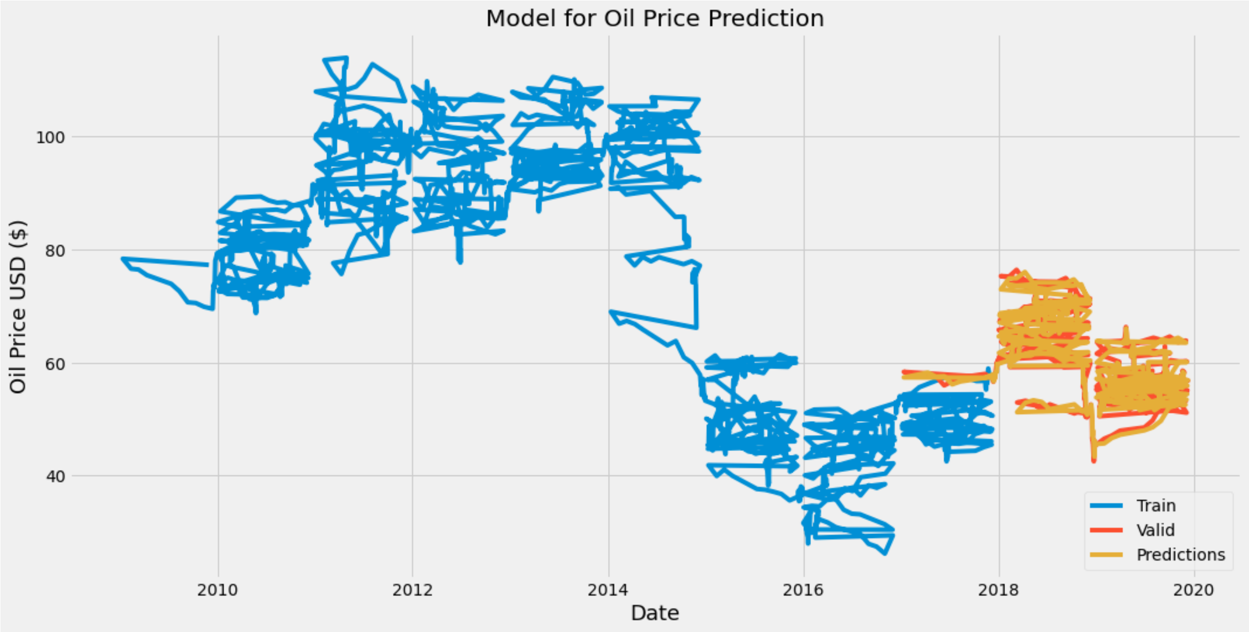


Figure 13: The figure shows the training dataset along with the predictions corresponding to their actual prices.

Table 4.8 Results of forecast statistic from RNN (LSTM) Model

RMSE	MAE	MAPE	D-Watson
0.02822	0.01450	0.0064%	1.9345

the machine learning model RNN(LSTM) has D-Watson value of 1.9345 which is significantly rejecting the null hypothesis and concluded that there is no autocorrelation/serial correlation. The residuals plot from this model given on the next page also clearly indicates that the residuals have no any regular pattern and became iid (independent and identically distributed) pattern of residual series which validates this model RNN(LSTM). Thus, the prediction from this model is valid and can be used for further analysis and forecasting. The graphs of actual and fitted series and residuals are given as follows:

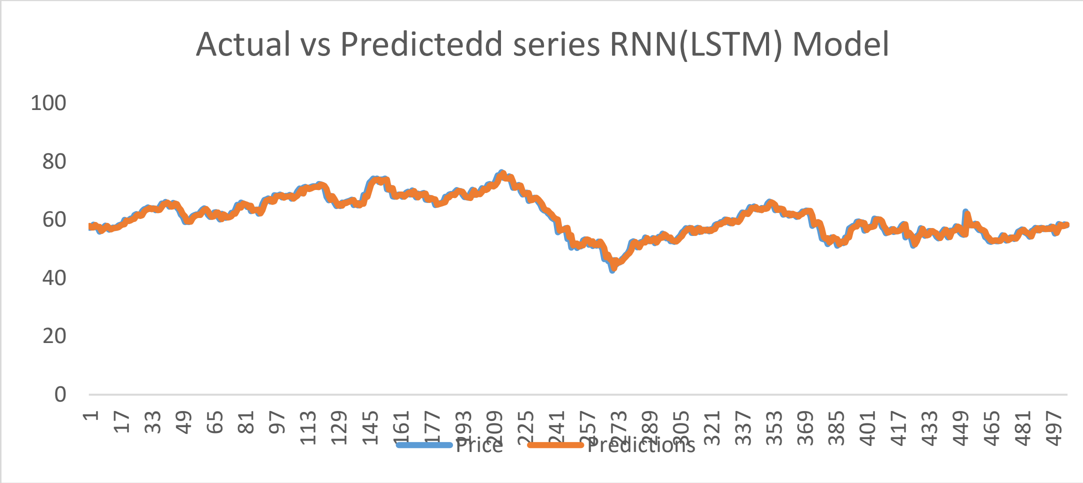


Figure 14: The actual and predicted Series Graph from the RNN (LSM) Model. The Series is Daily crude oil price of Pakistan

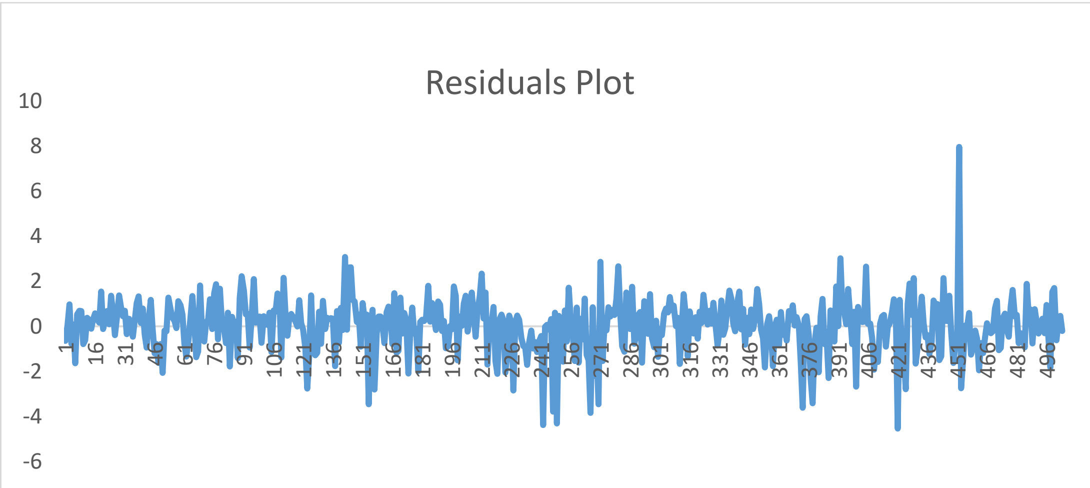


Figure 15: The residual plot from RNN (LSTM) showing the random behaviour.

The graph of actual and predicted series and residual plot show that the estimated model is valid because, the residual series has become iid and there is no conditional hetroskedasticity in the variance.

As the objective is to find the better model for the prediction of the oil prices in Pakistan therefore, the fitted models are analyzed based on their forecasted values. The ARIMA (3,1,3), GARCH (1,1), ARIMA(3,1,3)+GARCH(1,1) and then RNN(LSTM) were estimated. To compare these models and to choose the better among estimated models the following statistics were used (MAE and RMSE) to evaluate the forecast performance of various models. The comparison of the various models is given below:

Table 4.9 Forecast Accuracy Comparison

	RMSE	MAE
ARIMA (3, 1,3)	0.02048	0.01479
GARCH (1,1)	0.02047	0.01479
ARIMA(3,1,3)+GARCH(1,1)	0.01935	0.01442
RNN (LSTM)	0.02822	0.01450

based on the mean absolute errors and root mean square errors the better model for the prediction of the oil prices in Pakistan based on the variation explain ability and predictability, in the above table and figure 4.5 & 4.7. The hybrid model ARIMA(3,1,3)+GARCH(1,1) and RNN(LSTM) have minimum values of MAE therefore, in case of forecasting the future values based on the daily data of prices the Hybrid-GARCH model has performed better than the other estimated models. On the other hand, according to the figure 4.5 & 4.7 suggests that RNN(LSTM) is most appropriate model for the oil price predictions.

CHAPTER 5

CONCLUSION

5.1 Introduction

Th is about the core findings, recommendation of empirical analysis. The chapter is divided into two sections. Firstly, Section 5.2 deals with the core findings of this study and lastly the section 5.3 is about the limitations of the study.

5.2 Major findings of the study

The main objective focused on empirically testing of daily crude oil price of Pakistan and to suggest the most appropriate model for prediction for the daily oil price in Pakistan. As the investor/market player is very much interested in tomorrow's price. To achieve this objective, I estimated the series at level by using machine learning model (RNN(LSTM)) parallel to the estimation of conventional time series models on return series. It also includes the hybridization of the existing conventional time series model for volatility analysis for the oil price in Pakistan. The statistical treatments applied on the data and made it stationary by using the log differencing transformation and then ARIMA model was estimated by using the transformed data. As volatility clustering observed in the variance of the data series therefore, the GARRCH (1,1) model estimated to capture volatility. It can capture infinite ARCH processes the results suggested that the daily crude oil price for Pakistan was significantly showing the GARCH effects. The series of the forecasts were obtained then to compare this forecast with the other proposed hybrid models.

The Hybrid ARIMA-GARCH model was then estimated. It includes the series of variance from the residuals of ARIMA model which was showing the ARCH effect. Thus, to meet the quest for more appropriate model for the prediction of the oil prices another machine learning model RNN(LSTM) that is highly flexible and unrestricted method regarding parametric assumptions (stationarity, normality etc.) was also estimated and the series of forecast obtained. The obtained results have shown that hybrid ARIMA-GARRCH model performed the best amongst all estimated models but taking into account the informal procedure according to the graphical representation of predictions in figure 4.5 & 4.7 suggest that performance of RNN(LSTM) is best. The root mean squared error and the mean absolute errors of hybrid ARIMA-GARCH and RNN(LSTM) model are less than that amongst all the other estimated models and very close to each other. The conclusion of this empirical study thus supports the application of hybrid model to forecast the daily oil prices in Pakistan. As RNN(LSTM) is assumption free and it has the ability to deal with any sort of dataset, so, it should preferably be applied for the large dataset for good predictions. It also consumes a very small computational power to analyze the large datasets in one go.

The empirical results have shown that this type of hybridization that is capturing the linear part of the univariate analysis through the ARIMA type modelling then model the conditional variance through the GARH type modelling with ARIMA as the mean equation can perform better than to model them separately. Thus, study recommends the hybrid model to use for the prediction of future prices in case of oil prices in Pakistan.

The empirical results have shown that this type of hybridization that is capturing the linear part of the univariate analysis through the ARIMA type modelling then model the conditional variance through the GARH type modelling with ARIMA as the mean equation can perform better than to model them separately. Thus, study recommends

the hybrid model to use for the prediction of future prices in case of oil prices in Pakistan.

5.3 Limitations of the study

The actual prices may deviate far away from the predicted prices and the predictions become fallacious if the values predicted too early for the long run. So, any of the econometric or machine learning models which perform very well in the short run but may not perform well for predictions in the long run.

REFERENCES:

Abdullah, S. N. and Zeng, X. (2010). Machine learning approach for crude oil price prediction with Artificial Neural Networks-Quantitative (ANN-Q) model, Proceedings of the International Joint Conference on Neural Networks (IJCNN'2010), 1-8.

Agnolucci, P. (2009). Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models, Energy Economics 31, 316-321.

Ahmad, M., Modelling and forecasting Oman crude oil prices using Box–Jenkins techniques. International Journal of Trade and Global Markets, 2012. 5(1): p. 2430.

Ahmed, R.A. and A.B. Shabri, Daily crude oil price forecasting model using arima, generalized autoregressive conditional heteroscedastic and support vector machines. American Journal of Applied Sciences, 2014. 11(3): p. 425-432.

Alizadeh, A. and Mafinezhad, K (2010). Monthly Brent Oil Price Forecasting Using Artificial Neural Networks and A Crisis Index, Proceedings of the International Conference On Electronics and Information Engineering (ICEIE'2010), 2, 465-468.

Aloui, C. Hamdi, M., Mensi, W. and Nguyen, D. Y. (2012). Further evidence on the time varying efficiency of crude oil markets, Energy Studies Review 19(2), 38-51.

Alquist, R. and L. Kilian, What do we learn from the price of crude oil futures? Journal of Applied Econometrics, 2010. 25(4): p. 539-573.

Amano A. (1987). A Small Forecasting Model of the World Oil Market, Journal of Policy Modeling 9(4), 615-635.

Amin-Naseri, M. R. and Gharacheh, E. A. (2007). A hybrid artificial intelligence approach to monthly forecasting of crude oil price time series, The Proceedings of the 10th International Conference on Engineering Applications of Neural Networks (CEANN'2007), 160-167.

Barone-Adesi, G., Bourgoin, F. and Giannopoulos, K. (1998). Don't Look Back, Risk 11, 100-104.

Beidas-Strom, S. and Pescatori, A. (2014). Oil price volatility and the role of speculation.

Blanchard, O. J. and Gali, J. (2007). The macroeconomic effects of oil shocks: Why are the 2000s so different from 1970s? NBER working paper number 13368, 1-77.

Cheong, C.W. (2009). Modelling and forecasting crude oil markets using ARCH-type models Energy Policy 37, 2346-2355.

Chinn, M.D., M. LeBlanc, and O. Coibion, The predictive content of energy futures: an update on petroleum, natural gas, heating oil and gasoline, 2005, National Bureau of Economic Research.

Dees, S., Karadeloglou, P., Kaufmann, R. K. and Sanchez, M. (2007). Modelling the world oil market: assessment of a quarterly econometric model, Energy Policy 35, 178-191.

Fattouh, B. (2012). Speculation and oil price formation, Review of Environment, Energy and Economics (Re3), 1-5.

Ghaffari, A. and Zare, S. (2009). A novel algorithm for prediction of crude oil price variation based on soft computing, Energy Economics 31, 531-536.

Ghouri, S. S. (2006). Assessment of the relationship between oil prices and US oil stocks, Energy Policy 34, 3327-3333.

Gomes, G. S. D. S., Ludermir, T. B. and Lima, L. M. M. R. (2011). Comparison of new activation functions in neural network for forecasting financial time series‘‘‘‘ Neural Computing and Applications 20(3), 417-439.

Gori, F., Ludovisi, D. and Cerritelli, P.F. (2007). Forecast of oil price and consumption in the short term under three scenarios: Parabolic, linear and chaotic behavior, Energy 32, 12911296.

Gulen, S. G. (1998). Efficiency in the Crude Oil Futures Markets, Journal of Energy Finance and Development 3, 13-21.

Hagen, R. (1994). How is the international price of particular crude determined? OPEC Review 18, 145-158.

Haidar, I., Kulkarni, S. and Pan, H. (2008). Forecasting model for crude oil prices based on artificial neural networks, Proceedings of the International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2008), 103-108.

Hamilton, J. D. (1996). This is what happened to the Oil-Macro economy relationship, Journal of Monetary Economics 38(2), 215-220.

Hamilton, J. D. (2008). Understanding crude oil prices, The Energy Journal, International Association for Energy Economics 30(2), 179-206.

Hamilton, J. D. (2010). Nonlinearities and the macroeconomic effects of oil prices, NBER working paper number 16186, 1-22.

Haykin, S. (1999). Neural networks: A comprehensive foundation, Prentice Hall, New Jersey.

Hornik, K., Stinchcombe, M. and White, H. (1989). Multi-layer feedforward networks are universal approximators, Neural Networks 2, 359-366.

Hou, A. and S. Suardi, A nonparametric GARCH model of crude oil price return volatility. Energy Economics, 2012. 34(2): p. 618-626.

Huntington, H. G. (1994). Oil Price Forecasting in the 1980s: What Went Wrong? The Energy Journal 15(2), 1-22.

IMF working paper (WP/14/218). Bernabe, A., Martina, E., Alvarez-Ramirez, J. and Ibarra-Valdez, C. (2004). A multi-model approach for describing crude oil price dynamics, Physics A: Statistical Mechanics and its Applications 338(3), 567-584.

Jammazi, R. and Aloui, C. (2012). Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling, Energy Economics 34(3), 828-841.

Kaboudan, M. A. (2001). Compumetric forecasting of crude oil prices, Congress on Evolutionary Computation 1, 283-287.

Kang, S.H., Kang, S.M. and Yoon, S.M. (2009). Forecasting Volatility of Crude Oil Markets, Energy Economics 31, 119-125.

Kilian, L. (2008). The Economic effects of energy price shocks, Journal of Economic Literature 46(4), 871-909.

Kilian, L. and Murphy, D. P. (2011). The role of inventories and speculative trading in the global market for crude oil, Working paper. University of Michigan.

Kilian, L., Fattouh, B. and Mahadeva, L. (2013). The role of speculation in oil markets: What have we learned so far? The Energy Journal 34(3), 7-33.

Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps, Biological Cybernetics 43, 59-69.

Kulkarni, S. and Haidar, I. (2009). Forecasting model for crude oil price using artificial neural networks and commodity futures prices, International Journal of Computer Science & Information Security 2(1), 81-88.

Lackes, R., Börgermann, C. and Dirkmorfeld, M. (2009), Forecasting the Price Development of Crude Oil with Artificial Neural Networks, Lecture Notes in Computer Science 5518, 248-255.

Lanza, A., Manera, M. and Giovannini, M. (2005), Modeling and forecasting cointegrated relationships among heavy oil and product prices, Energy Economics 27, 831-848.

Liu, J., Bai, Y. and Li, B. (2007). A new approach to forecast crude oil price based on fuzzy neural network, Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), 273-277.

Marzo, M. and P. Zagaglia, Volatility forecasting for crude oil futures. Applied Economics Letters, 2010. 17(16): p. 1587-1599.

Sadorsky, P., Modeling and forecasting petroleum futures volatility. Energy Economics, 2006. 28(4): p. 467-488.

Wang, S.Y., L.A. Yu and K.K. Lai,, Crude oil price forecasting with TEI@ I methodology. Journal of Systematic Science. Complexity, 2009. 18(2): p. 145166.