International Journal of Multidisciplinary Research and Development
Online ISSN: 2349-4182, Print ISSN: 2349-5979
Impact Factor: RJIF 5.72
Received: 03-04-2018; Accepted: 08-05-2018
www.allsubjectjournal.com
Volume 5; Issue 5; May 2018; Page No. 44-49

Research review on time series forecasting of gold price movement

Wesley Goh Yi Yong, Suliadi Sufahani*
Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia

Abstract
This paper propose different techniques used in forecasting on time series analysis. Various techniques had been used in the field of time series forecasting from the traditional Box-Jenkins approach to the most popular neural network technique nowadays. However, there is no specifically the best method to deal with time series forecasting as the application of different time series forecasting methods has their own requirements and restrictions. In determining the movement of gold price, there are a lot of different methods being implemented by various authors to propose their models. Various time series forecasting method have been discussed in this paper which consists of several journal articles that related to gold price and some of the data mining techniques in time series forecasting retrieved from Google Scholar in this review.

Keywords: time series, box-jenkins, forecasting, gold price

1. Introduction
A precious metal is a rare naturally occurring, metallic chemical element of high economic value; precious metals were important as currency in the past but are now regarded mainly as investment and industrial commodities. Historically, gold remains resilient and performed better than the S&P 500 index during the 1973-1974 market crash, the oil crisis of the mid-1980s,the crisis of the new millennium combined with the tax bubble in the 1990s and the global financial crisis in 2008 (The Edge Malaysia, 2017). Hence, gold is known as a safe-haven asset as it has an inverse correlation with the US dollar. Investors tend to preserve the value of their asset when the economic is uncertain by investing in gold since it does not have heavy liability or unpredictability (Shafiee and Topal 2010) [2, 12]. Gold act as insurance for investors as it has low correlation with most assets thus can be used to reduce portfolio volatility and minimize losses during extreme market conditions.

The movement of gold price is concerned by investors, financial specialist, and government and to who gain their profit through gold trading. The impact to the movement of gold prices is influenced by a lot of factors which include macroeconomic variables and business cycles. However, related studies also suggested that there is high correlation between gold price and crude oil price (Chen and Fang 2013) [3, 18]. Besides, other author's result shows that there is a significant relationship between the precious metals and oil, which include gold, platinum and silver. Additional research also indicated that oil impacts gold more than gold impacts oil (Gabralla, Jammazi *et al*. 2013) [4].

The prices of precious metals varies from day to day, a lot of research had been done by researchers using different techniques. The traditional way of doing statistical analysis could be more complex and difficult as of today, the data storage is getting cheaper and lots of data can be collected for analysis. Big data could often be heard from various industries due to the readily available data that could be access by anyone to perform their business analysis which could reduce the cost and resources on data collection. Hence, data mining technique is getting more popular as it used to extract the needed information from a big pile of data for specific analysis.

Traditional way of doing time series forecasting is popular by using the Box-Jenkins method. However it can only deal with univariate time series analysis which includes only the dependent variable itself. Data mining allow the addition of various independent variables to help to forecast the movement of dependent variable and allow us to understand the indicators behind it. Hence, it is value added to any businesses.

Most research only concern about the movement of gold price with macroeconomic variables such as GDP growth rate, inflation rate, and consumer price index. There is no specific method in data mining used to forecast the gold price with other precious metals. In the study of volatilities of four precious metals which are gold, silver, platinum, and palladium prices, the authors claimed that precious metals are too distinct to be considered a single asset class, or represented by a single index (Batten, Ciner *et al*. 2010) [5].

2. Literature Review
The advancement of technology had boosted the collection and storing of data to a whole new level. Nowadays, data can be easily collected by various gadgets which connect to a computer system. Since data can be easily collected as the data storage is getting cheaper, data can also be retrieved and obtained with fingertips through internet by using computational software. In this situation, it brings up a whole

new challenge in the field of data analysis as traditional statistical analysis could not be able to perform efficiently as the data sets get larger because it will take a whole lot more of time and energy for a human being to do so. Hence, it is required to have both computational skills and analytical skills in order to improve the performance of data analysis.

Big data is being more and more often to be heard nowadays. Big data usually includes data sets that are so large and complex which the sizes are beyond the ability of standard statistical software tools to capture, curate, manage, and process data within a tolerable elapsed time (Snijders, Matzat *et al*. 2012) [6]. Big data requires a set of techniques and technologies with new forms of integration to reveal insights from datasets that are diverse, complex, and of a massive scale (Hashem, Yaqoob *et al*. 2015) [7].

Data mining is a process of extracting information from a huge data set and transform it into understandable structure for future use. Data mining is the interdepartmental field of study that are able to discover patterns and models from a large amount of information stored in data warehouse (Gaber, Zaslavsky *et al*. 2005) [8]. Data mining is also refers to the extracting or mining knowledge from large data stores or sets (Al-Radaideh, Assaf *et al*. 2013) [9]. However, data mining is not something new as various author had been using the technique in their research for past decade.

2.1 Box-Jenkins Approach

The Box-Jenkins Auto-Regression Integrated Moving Average (ARIMA) is one of the most sophisticated techniques in time series forecasting. A study of gold price forecasting in Indian was conducted by (Guha and Bandyopadhyay 2016) [10]. The study is based on the historical data of monthly gold price from November 2003 to January 2014 and ARIMA model was applied to forecast fluctuation of the gold price. The price of gold in this study is valued in Rs per 10 grams. Three model parameters which are Autoregressive (AR) indicating weighted moving average over past observations, integrated (I) indicating linear trends or polynomial trend, and Moving Average (MA) indicating weighted moving average over past errors are combined to form ARIMA (p, d, q) where:

p = order of autocorrelation
d = order of integration (differencing)
q = order of moving averages

The ARIMA model can also be expressed as follow:

$$X_t = \theta_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_q e_{t-q} \quad (1)$$

According to the researcher, ARIMA (1, 1, 1) is the best model chosen for prediction as all the significant values are less than 0.05. Furthermore, Root Mean Square Error (RMSE), Mean Absolute percentage Error (MAPE), Mean Absolute Error (MAE), Bayesian Information Criterion (BIC), and Lungs Box Q statistics are used as the statistical measure of the model. The authors also claimed that the gold price shows high first order auto-correlation through Durbin-Watson (DW) test.

Another Box-Jenkins analysis was conducted by (Faraway and Chatfield 1998) [11] which is the analysis of the well-known airline data. A total of 132 monthly data were used for the research and a SARIMA model of order (0, 1, 1) $(0, 1, 1)_{12}$ is the model chosen by the authors. The authors suggested that the Bayesian information criterion (BIC) is preferred to Akaike's information criterion (AIC) for comparing different models as BIC will penalize the addition of extra parameters more severely than AIC does.

A comparison has been made by (Shafiee and Topal 2010) [2, 12] to compare the accuracy of forecasting model of gold price by using reverting jump and dip diffusion model, and ARIMA model. The study focus on the gold price trends over the past 40 years from January 1968 to December 1988 and analysis was conducted by using monthly data to forecast the gold price for the next 10 years. The unit root test Augmented Dickey Fuller (ADF) test was conducted by the authors and the gold price was found non-stationary and the first differential transformation was applied. The reverting jump and dip diffusion model was claimed to perform better than ARIMA model through comparison of RMSE and MAE of both model. The new model proposed by the author is as follows which consider the drift, diffusion components, and jump and dip period of the gold price.

$$X_t = \alpha_1 + \alpha_2 t + \alpha_2 (1 \pm \gamma) + \alpha_3 D_1 + \alpha_4 D_2 + u_t \quad (2)$$

Where

$\alpha_2 t$ = First component or drift

$\alpha_2 (1 \pm \gamma)$ =Second component or the range of random movement

$\alpha_3 D_1 + \alpha_4 D_2$ = Third component or jump or dip

(Khan 2013) [13] proposed ARIMA model for accuracy checking on the prediction of gold price. The samples consist of gold price data in USD per ounce from 2 January 2003 to 1 March 2012. Most of the macro economy data are non-stationary are claimed by the authors. Augmented Dickey Fuller (ADF) test is constructed to determine the stationarity of the series, gold price is said to be non-stationary and only be found stationary after first difference is applied. Akaike Information Criteria (AIC) and Schwarz Information Criteria (SIC) are used to compare the ARIMA models. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) is used to evaluate the accuracy of the models. It was found that ARIMA (0, 1, 1) is the best model selected to predict the future gold price.

(Yaziz, Azizan *et al*. 2013) [14] proposed a hybrid method of ARIMA-GARCH modelling to forecast the gold price by using a total of 40 daily gold price. ARIMA is said to be powerful and flexible in forecasting but it is unable to deal with volatility and nonlinearity in a data set. GARCH is used to model the nonlinear patterns of the residuals. The authors claim that the gold price is non-stationary by conducting Augmented Dickey-Fuller (ADF) test which the first differencing is needed to make the series stationary. The results show that ARIMA-GARCH model outperform several models proposed by other authors.

2.2 Arimax model

The paper by (Wongdhamma 2016) [15] proposed a comparison of ARIMA and ARIMAX model on time series forecasting. The authors implement wavelet analysis (WA) to decompose the trend, seasonality, process variations, and noises of time series. The decomposed data is then used as exogenous variables on constructing an ARIMA model. The results show that WA are good regressors in ARIMAX which are able to capture nonlinear patterns well. ARIMAX-WD performs better than ARIMA in long term and nonlinear time-series forecasting application.

2.3 Regression

A linear regression approach is proposed by (Gharehchopogh, Bonab *et al*. 2013) [16] in the case study of predicting the S&P 500 index. The authors selected the volume as the dependent variables and average daily price as independent variables. The results shows that the accuracy is satisfied as there is 61.35% of similarities observed when compared to the out samples. However, the model has a low R-squared value, which indicates that the model does not contribute much for the prediction of volume. The model could be improved by considering other factors that could affect the daily volume trade via variable selection through data mining.

Multiple regression method is proposed by (Baker and Van Tassel 1985) [17] based on fundamentalist approach to determine the variables that affecting the gold price and model building for long term gold price prediction. The changes in commodity price index, changes in the value of US dollars, and future inflation rate is found to be affecting the monthly change of gold price. The results suggested that the price of gold can be expected to rise if there is a general increase in commodity prices.

Another multiple regression approach is conducted by (Chen and Fang 2013) [3, 18]. The return on investment in gold has higher correlation with other commodities (aluminium, petroleum, zinc, and etc.) than the business cycle and the main macroeconomic variables. The authors aimed to explore the long term determinants of gold price by considering the US money supply M2, CRB index, US dollar index, the Dow Jones Industrial Average, and the SPDR holdings as the exogenous variables in building the regression model. The study concludes that M2 of US money supply is the most affecting factor with gold price which they are highly positive correlated. CRB index of inflation levels also shows high positive correlation with the gold price. However, the model showed that gold price has low negative correlation with the Dow Jones Industrial Average, low positive correlation with the US dollar index and SPDR gold trust positions.

According to experts, there are a lot of economic factors which could affect the movement of gold price. Multiple linear regression method has been used to predict the gold price based on several economic factors such as inflation, currency price movements and others (Ismail, Yahya *et al*. 2009) [19]. Stepwise regression is used to remove the correlation between variables. At the same time, the authors proposed that the effects of significant lag in the cause-and-effect process in order to improve the performance of the model. Durbin-Watson statistics is used to check whether autocorrelation exist in the error terms. Error mean square MSE and mean squared prediction error MSPR are used to evaluate the model accuracy. From the study, the factors affecting the gold price include Bureau future (CRB) index, USD/Euro Foreign Exchange Rate EUROUSD, inflation rate and money supply (M1).

2.4 Support vector regression

Support Vector Regression (SVR) is used as a learning algorithm to predict the gold price by understanding the pattern of the historical gold price (Navin 2013) [20]. An assumption that made by the authors is that the historical data incorporate all the external effects behaviour. The data is transformed into generic dataset followed by cross validation process to feed them as inputs into support vector regression model. The type of kernel and special parameters are selected and applied to the model, the process was repeated until the accuracy of the model is good enough. The model is then used to test with the out of sample dataset for gold price prediction.

2.5 Neural network

Neural network and genetic algorithm were implemented by (Mirmirani and Li 2004) [21] to analyse the movement of gold price. Application of Genetic Algorithm along with neural networks is said to improve the learning power and robustness of the system. The analysis was done by using daily cash prices of gold from 12/31/1974 to 12/31/1998 which consist of 6008 data points. The results show that the historical prices of gold strongly affect the gold price in future. The authors also claimed that there is short-term time dependence of gold price movements with the time lag of 36 days.

Neural network has also be applied by (Grudnitski and Osburn 1993) [22] for forecasting S&P and gold futures price. Other than focusing on the price trend itself, the authors also consider general economic conditions and traders' expectations about the market in future. The authors conclude that neural network can be applied to forecast price changes of the markets with four factors considered which are:

1. Derivation of the parameters of the networks relies heavily on published commitments.
2. Selection of futures whose price are relatively insulated from natural phenomena.
3. Appropriate length for the training period.
4. Trading selectively rather than in every period.

In constructing the neural network model, the researchers may face a hard time on estimating the weights to be used in the model. It is not an easy task as the number of weights may be large and the objective function may have local minima (Faraway and Chatfield 1998) [11]. The application of neural network model is still a doubt for researchers as some study had proposed that the neural network model is no better than other methods. Neural networks has several drawbacks which include excessive training times, difficulty in obtaining and replicating a stable solution, overfitting of model, and the black box nature of the solutions. Although, neural network modelling is non parametric analysis which the process can be completely automated on a computer, Black boxes can sometimes give silly results.

2.6 Other Methods in Time Series Forecasting

Ensemble machine learning was proposed by (Gabralla, Jammazi *et al.* 2013) [4] to study the forecasting performance of daily WTI crude oil price with consideration of a number of influential features as inputs. Best-first search and genetic algorithm were used by the authors in selection of variables that influences the crude oil price. According to the authors, it is possible to use commercial material instead of gold price for crude oil price prediction. Their results show that the ensemble method performed better than SMOReg and IBL using only three attributes.

Another approach of using data mining classification in stock market prediction is proposed by (Paliyawan 2006) [23]. The author performs the analysis by using four different approaches which are neural network, decision tree, naïve Bayes, and k-nearest neighbours. The authors predict whether the market index will go up, go down or stay in the next 1, 6, and 21 days. In the mean times, the authors claimed that although NN models have high accuracy, many of them does not put into practice due to inability of neural networks to explain its reasoning. For the results, D-Tree with 3 classes is considered as the best model in the study. The authors applied chart patterns discovery to drill down into the model to find valuable patterns which could be easily used in decision making.

3. Discussion

This section highlights the key point that expressed by the authors of previous study.

Table 1: Summary of time series forecasting

Study	Technique	Research Problems
Qasem A. Al-Radaideh, Adel Abu Assaf & Eman Alnagi (2013) [9]	Decision tree classifier	To analyse the historical data of stocks using decision tree technique to help investors to decide when to buy or sell their stocks. The variables selected for the study involve only the historical price of the stock which only represents the market behaviour.
Banhi Guha & Gautam Bandyopadhyay (2016) [10] Khan (2013) [13]	Traditional Box-Jenkins methods (ARIMA and SARIMA)	To implement forecasting of gold price through time-series ARIMA model. The ARIMA model is selected through comparing AIC and SIC. The forecasting performance model selected is compared by using MAE, MAPE, RMSE, ARIMA model is suitable for short run only, it can only predict by using univariate data set. ARIMA is unable to deal with volatility and nonlinearity in a data set.
Julian Faraway & Chris Chatfield (1998) [11]	Neural network and ARIMA	To fit variety of neural network models to the well-known airline data and compares the resulting forecast with Box-Jenkins and Holt-Winters method. Neural network model were found to be possible that the fitting routine may not converge or may converge to a local minimum. Besides, NN models that fit too well may give poor out of sample forecast. The complicated black box structure of NN model is hard to understand and interpret.
Navin & Dr. G. Vadivu (2013) [20]	Decision tree and support vector regression (SVR)	Develop a forecasting model for predicting and forecasting gold prices based on past historical prices of gold. External factors are not being considered as it assume the historical prices incorporate all the external behaviour.
Shafiee and Topal (2010) [2, 12]	Reverting jump and dip diffusion model	To compare the accuracy of forecasting model of gold price by using reverting jump and dip diffusion model, and ARIMA model. The new model proposed by the author consider the drift, diffusion components, and jump and dip period of the gold price which performed better than ARIMA model.
Wongdhamma (2016) [15]	ARIMAX	Proposed a comparison of ARIMA and ARIMAX model on time series forecasting. Wavelet analysis (WA) is applied to decompose the trend, seasonality, process variations, and noises of time series. The decomposed data is then used as exogenous variables on constructing an ARIMA model.
Gharehchopogh, Bonab *et al.* (2013) [16]	Fuzzy Linear Programming	To predict the S&P 500 index using linear regression by using volume as dependent variable and average daily price as independent variable. The model shows low R-squared value, other factors could be identified to improve the model contribution.
Baker and Van Tassel (1985) [17] Ismail, Yahya *et al.* (2009) [19] Chen and Fang (2013) [3, 18]	Multiple linear regression	To build a model to predict gold price by considering exogenous factors. Several economic factors including commodity price index have been found to affect the gold price movement.
Navin (2013) [20]	Support vector regression	To predict the gold price by understanding the pattern of the historical gold price. Selection of kernel function is complicated. There is no benchmark for SVR approach. There is no optimal method on free parameter selection.
Mirmirani and Li (2004) [21]	Neural network and Genetic algorithm	To apply Genetic algorithm accompanied with neural networks to find an optimal net architecture in model selection in time series analysis. The results show that the historical prices of gold strongly affect the gold price in future. The authors also claimed that there is short-term time dependence of gold price movements.
Grudnitski and Osburn (1993) [22]	Neural network	To forecast S&P and gold futures price. The authors conclude that neural network can be applied to forecast price changes of the markets with four factors considered which are: Derivation of the parameters of the networks relies heavily on published commitments. Selection of futures whose price are relatively insulated from natural phenomena. Appropriate length for the training period. Trading selectively rather than in every period.
Gabralla, Jammazi *et al.* (2013) [4]	Ensemble machine learning	To study the forecasting performance of daily WTI crude oil price with consideration of a number of influential features as inputs.
Paliyawan (2006) [23]	Neural network,	To apply data mining classification in stock market prediction.

	decision tree, naïve Bayes, and k-nearest neighbours	The results show D-Tree with 3 classes is the best model in the study and chart patterns discovery can be applied to drill down into the model to find valuable patterns. Although NN models have high accuracy, many of them does not put into practice due to inability of neural networks to explain its reasoning.
Yaziz, Azizan *et al*. (2013) [14]	ARIMA-GARCH	To propose a hybrid method of ARIMA-GARCH modelling to forecast the gold price by using a total of 40 daily gold price. ARIMA and GARCH are able to complement each other to achieve a better forecasting model than traditional ARIMA model.

Previous studies showed that several approaches have been implemented in the time series forecasting in economic and financial time series analysis. However, there is no specifically the best method to deal with time series forecasting as the application of different time series forecasting methods has their own requirements and restrictions. Therefore, further improvement of traditional Box-Jenkins analysis in time series analysis could be improved. Table 2 describes the contributions that could be recognized in the future research.

Table 2: Contributions of the research

To forecast the movement of gold price with the consideration of the price of other precious metals.
To identify the relationship of gold to other precious metals.
To compare the accuracy of forecasting model using data mining forecasting technique and traditional Box-Jenkins approach.

4. Conclusion

The paper presents a review of literature concerned with the variety of methods that had been applied in the time series forecasting in the field of economics and finance. The growing of recorded data is faster than ever before nowadays which leads to huge data sets. Traditional statistical analysis methods may need to modify in order to improve the performances and speed of conducting analysis. In short, statistical knowledge and computational skills have to be combined in order boost the performance of statistical analysis. Further study could be done by considering the other industrial commodities as exogenous variables that that could have driven the gold price. Data mining techniques that could complement with the traditional statistical analysis should be considered in time series analysis and forecasting in future work.

5. Acknowledgement

We would like to say thank you to Universiti Tun Hussein Onn Malaysia (UTHM) and Office for Research, Innovation, Commercialization and Consultancy Management (ORICC), UTHM for kindly providing us with the internal funding TIER 1for proofreading (Vot U903).

6. References

1. The Edge Markets. Cover Story: Backed By Solid Gold. Retrieved on 2017-2018, from http://www.theedgemarkets.com/article/cover-story-backed-solid-gold
2. Shafiee S, Topal E. An overview of global gold market and gold price forecasting. Resources policy. 2010; 35(3):178-189.
3. Chen X, Fang Y. Enterprise systems in financial sector-an application in precious metal trading forecasting. Enterprise Information Systems. 2013; 7(4):558-568.
4. Gabralla LA, *et al*. Oil price prediction using ensemble machine learning. Computing, Electrical and Electronics Engineering (ICCEEE), International Conference on, IEEE, 2013.
5. Batten JA, *et al*. The macroeconomic determinants of volatility in precious metals markets. Resources policy. 2010; 35(2):65-71.
6. Snijders C, *et al*. Big Data: big gaps of knowledge in the field of internet science. International Journal of Internet Science. 2012; 7(1):1-5.
7. Hashem IAT, *et al*. The rise of big data on cloud computing: Review and open research issues. Information Systems. 2015; 47:98-115.
8. Gaber MM, *et al*. Mining data streams: a review. ACM Sigmod Record. 2005; 34(2):18-26.
9. Al-Radaideh QA, *et al*. Predicting stock prices using data mining techniques. The International Arab Conference on Information Technology (ACIT'2013), 2013.
10. Guha B, Bandyopadhyay G. Gold price forecasting using ARIMA model. Journal of Advanced Management Science, 2016, 4(2).
11. Faraway J, Chatfield C. Time series forecasting with neural networks: a comparative study using the air line data. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1998; 47(2):231-250.
12. Shafiee S, Topal E. An overview of global gold market and gold price forecasting. Resources policy. 2010; 35(3):178-189.
13. Khan MMA. Forecasting of gold prices (Box Jenkins approach). International Journal of Emerging Technology and Advanced Engineering. 2013; 3(3):662-670.
14. Yaziz S, *et al*. The performance of hybrid ARIMA-GARCH modeling in forecasting gold price. 20th International Congress on Modelling and Simulation, Adelaide, 2013.
15. Wongdhamma W. Upgrade from ARIMA to ARIMAX to Improve Forecasting Accuracy of Nonlinear Time-Series: Create Your Own Exogenous Variables Using Wavelet Analysis, 2016.
16. Gharehchopogh FS, *et al*. A linear regression approach to prediction of stock market trading volume: a case study. International Journal of Managing Value and Supply Chains. 2013; 4(3):25.
17. Baker SA, Van Tassel RC. Forecasting the price of gold: A fundamentalist approach. Atlantic Economic Journal. 1985; 13(4):43-51.
18. Chen X, Fang Y. Enterprise systems in financial sector–an application in precious metal trading forecasting. Enterprise Information Systems. 2013; 7(4):558-568.
19. Ismail Z, *et al*. Forecasting gold prices using multiple linear regression method. American Journal of Applied Sciences. 2009; 6(8):1509.

20. Navin DG. Big Data Analytics for Gold Price Forecasting Based on Decision Tree Algorithm and Support Vector Regression (SVR), 2013.
21. Mirmirani S, Li H. Gold price, neural networks and genetic algorithm. Computational Economics. 2004; 23(2):193-200.
22. Grudnitski G, Osburn L. Forecasting S&P and gold futures prices: An application of neural networks. Journal of Futures Markets. 1993; 13(6):631-643.
23. Paliyawan P. Stock Market Direction Prediction Using Data Mining Classification. 2006; 5:6.