

A Modified ARIMA Model Based on Extreme Value for Time Series Modelling

Yaohui Bai

Cloud Computation and Big Data Research Center
Jiangxi University of Finance and Economics
Nanchang, China
e-mail: byhnpu@163.com

Benting Wan

School of Software & Communication Engineering
Jiangxi University of Finance and Economics
Nanchang, China
e-mail: ren_btw@163.com

Xixia Zong

Xi'an Aerospace Propulsion Test Institute
Xi'an, China
e-mail: zxx_0126@163.com

Wenyuan Rao

School of Software & Communication Engineering
Jiangxi University of Finance and Economics
Nanchang, China
e-mail: raowu2002@gmail.com

Abstract—The ARIMA model is an important method and is widely used in time series modelling. The model relies heavily on autocorrelation patterns in the data, and doesn't consider other factors. However, in most cases, the extreme value of series has an influence on the subsequent behaviour of series. But this information isn't considered in the original ARIMA model. To solve this problem, We proposes a modified ARIMA model based on the past maximum and minimum value of series to solve the modelling tasks which includes the factor of past extreme value in the model. The modified model is tested on USD-EUR exchange rate time series. The experimental results show that it is possible to improve the performance by considering extreme value for time series modelling compared to the original ARIMA model.

Keywords- Time Series; modelling; ARIMA; extreme value;

I. INTRODUCTION

A time series data is a sequence of numerical observations naturally ordered in time. Recently, the rapid development of information technologies have led to the situation in which huge amounts of information data accumulate in quick speed and in fact constitute various time series. The modelling of such time series is extremely important and vital, and has been attracting the attention of both practitioners and researchers. However, it is also considered a rather difficult problem, due to the many complex features frequently present in time series, such as irregularities, volatility, trends and noise, and so on. A number of techniques have been developed in an attempt to model time series based on their present and past behaviour.

Traditional time series modelling technologies, such as autoregressive integrated moving average (ARIMA)[1], exponential smoothing[2][3], decomposition[4], etc., have been widely and successfully used. More recently a number of machine learning techniques, such as neural networks[5~7], fuzzy systems[8], genetic algorithm[9], and SVM[10] are becoming promising directions in this

fields. Some showed improvement compared to traditional models. Although a number of modern intelligent technologies can be available in machine learning and pattern recognition, it has been rarely used in temporal aspects of time series modelling for the discrete values output.

Among them, ARIMA model is considered the most common choices for the time series modelling, which was popularized by George Box and Gwilym Jenkins in the early 1970s. The model only use the information in the series itself to model, and doesn't include other independent variables. However, the model is linear, and can be used to model stationary as well as non-stationary time series. With the widely application of the ARIMA models, there are a number of variations on the ARIMA model. For example, for the multiple time series, a VARIMA model may be appropriate, and for the seasonal effect, it is generally better to use a SARIMA.

Although the ARIMA model has been widely applied in time series modelling, the model has some shortcomings. The model rely heavily on autocorrelation patterns in the data, and doesn't consider other factors. However, in most cases, the extreme value of series has an influence on the subsequent behaviour of series. Therefore, in these cases, when we consider the model, we should include the extreme value of series into the model. This work proposes a modified ARIMA model based on the past maximum and minimum value of series to solve modelling tasks.

II. METHODOLOGY

A. Time Series

A time series is a set of data points, measured typically at successive times spaced at uniform time intervals, and defined by ,

$$X_t = \{x_t \in R | t = 1, 2, \dots, N\} \quad (1)$$

where t is the temporal index and N is the number of observations. Therefore X_t is a sequence of temporal observations orderly sequenced and equally spaced.

An intrinsic feature of a time series is that adjacent observations are dependent. With this dependence, future values of X_t can be forecasted based on known past values: to predict data points before they are measured. The future is given by X_{t+k} , $k=1,2,\dots$, where k represents the prediction horizon of k step ahead.

B. ARIMA

The acronym ARIMA, standing for "Auto-Regressive Integrated Moving Average", is a famous model of time series modelling proposed by Box & Jenkins in early 1970s. The model is generally referred to as an ARIMA(p,d,q) model, where parameters p, d, q are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. The basic idea of ARIMA models is that data sequences can be considered as stochastic sequences described by mathematics models approximately. Once the models is established, it can be used to better understand the data or to predict the future values of time series based on present values and past values. Nowadays, ARIMA models are the most common class of models for time series modelling. It is usually superior to common time-series analysis and multivariate regressions.

C. AR(p)

The AR(p) model is the first component, and has the form:

$$y_t = \delta + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t \quad (1)$$

where the $\varphi_1, \varphi_2, \dots, \varphi_p$ are the parameters, ε_t is a white noise error term, δ is constant.

AR(p) is equivalent to,

$$(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) y_t = \delta + \varepsilon_t \quad (2)$$

and its characteristic equation is,

$$\Phi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p = 0 \quad (3)$$

D. MA(q)

The MA(q) model is the second component of ARIMA model, and the notation MA(q) refers to the moving average model of order q. The model is written as :

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (4)$$

where the $\theta_1, \theta_2, \dots, \theta_q$ are the parameters, μ is the expectation, and the $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ are again, white noise error term.

MA(q) is equivalent to,

$$y_t - \mu = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) u_t = \Theta(L) u_t \quad (5)$$

E. ARMA(p,q)

The model ARMA(p, q) refers to the model with p autoregressive terms and q moving-average terms, and is written as,

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (6)$$

)

it is also equivalent to,

$$\Phi(L) y_t = (1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) y_t \quad (7)$$

$$\delta + \Theta(L) \varepsilon_t = \delta + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t \quad (8)$$

namely,

$$\Phi(L) y_t = \delta + \Theta(L) \varepsilon_t \quad (9)$$

F. ARIMA(p,d,q)

ARIMA(p,d,q) model is applied in some cases where data shows evidence of non-stationary, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to remove the non-stationary. The model is written as,

$$\Phi(L) \Delta^d y_t = \delta + \Theta(L) \varepsilon_t \quad (10)$$

G. Modified ARIMA(p,d,q) model

As traditional methods of time series modelling, the ARIMA model is an adaptive, extrapolative methods, can model seasonality, and in certain cases, can give some forecasts. The method has been widely applied in time series analysis. The underlying principle of the model is that ARIMA model is based on autocorrelations of series. However, in most cases, the extreme value of series has an influence on the subsequent behavior of series. For example, in stock trading behavior, people will consider the past maximum and minimum value of stock price, and in various futures time series, people will consider the same problem. The original ARIMA model does not take full account of this case.

To account for the influence of the past extreme value of time series, we can modify the model based on the past maximum and minimum value of the series. The modified model is given by,

$$\Phi(L) \Delta^d y_t + \beta_{\max} y_{\max} + \beta_{\min} y_{\min} = \delta + \Theta(L) \varepsilon_t \quad (11)$$

III. EXPERIMENTS AND RESULTS

In this study, we used the time series of USD - EUR exchange rate (1971Q1-2007Q4) to test the proposed method. Since the collapse of the Bretton Woods system in 1973, exchange rate fluctuations between the major countries become one of the most obvious characteristics in the world financial market. After the advent of the euro as currency in January, 1999, the USD - EUR exchange rate have had several ups and downs.

Fig .1 shows the changes in trend of USD - EUR (before 1999, it's German mark) exchange rate since 1973. Fig .2 shows the sample autocorrelation function and partial autocorrelation function of series. From the figure

and related ADF test, we can conclude that the series is non-stationary. To take the first difference of the series, the result is shown in Fig .3, the autocorrelation function and partial autocorrelation function shown in Fig .4. According to the figure and the SIC result, the differenced series can take AR (1) model, namely:

$$\Delta y_t = \delta + \beta_1 \Delta y_{t-1} + \varepsilon_t \quad (12)$$

The modified model is given by,

$$\Delta y_t = \delta + \beta_1 \Delta y_{t-1} + \beta_{\max} y_{\max} + \beta_{\min} y_{\min} + \varepsilon_t \quad (13)$$

The maximum value of the original series is 1.4232(1980Q3), and the minimum value is 0.6845(1985Q1). The calculation results of two models are shown in Fig .5-Fig .7, Table1 and Table2. According to the calculation results, the coefficients of are significantly, indicating that the extreme values have an impact on the development of series. At the same time, compared to the relevant residual plots, and error calculation results, it shows that the modified model has a better error performance than the original model.

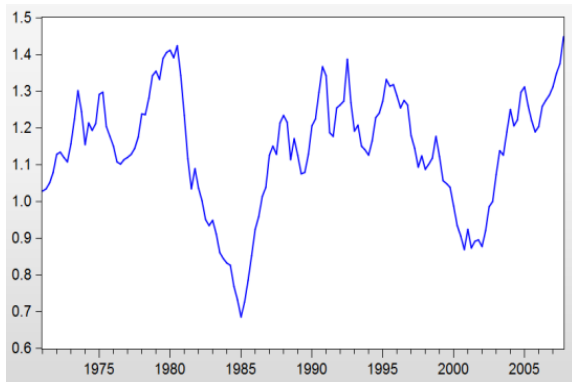


Figure 1. USD - EUR Exchange Rate (1971Q1-2007Q4)

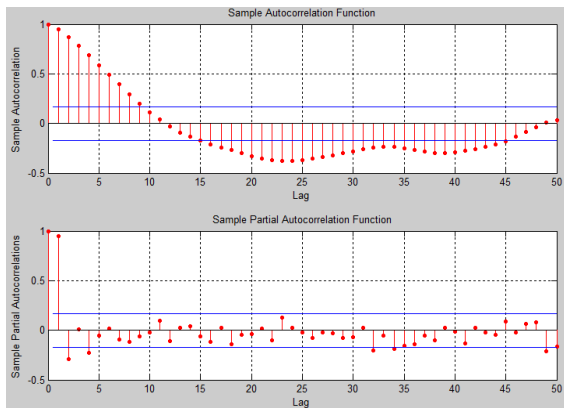


Figure 2. Sample Autocorrelation Function and Partial Autocorrelation Function

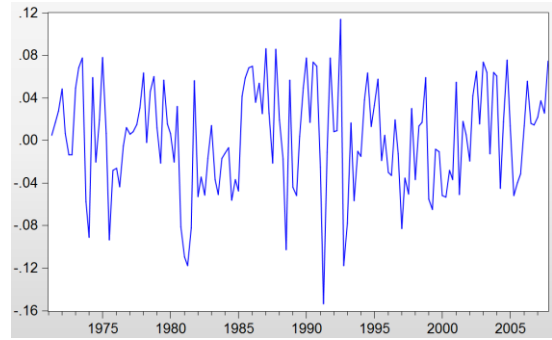


Figure 3. The 1st Difference Series of Exchange Rate

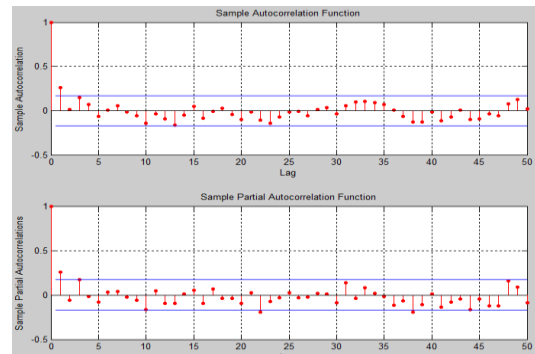
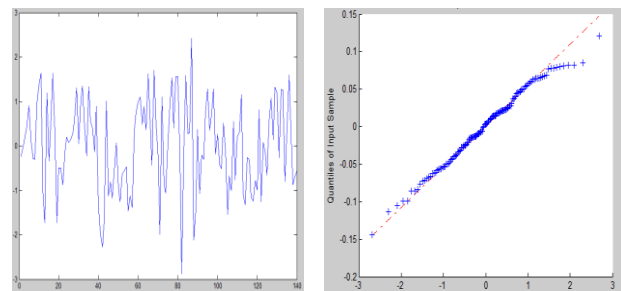
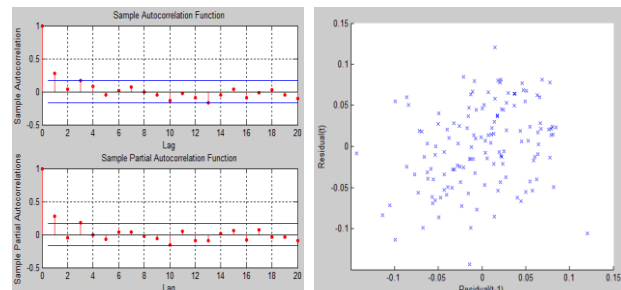


Figure 4. Sample Autocorrelation Function and Partial Autocorrelation Function of the 1st Difference Series



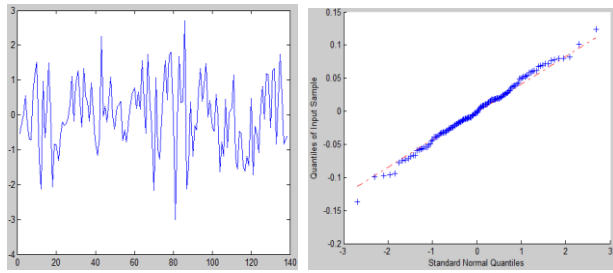
a) The Standardized Residuals

b) Residuals QQ plot

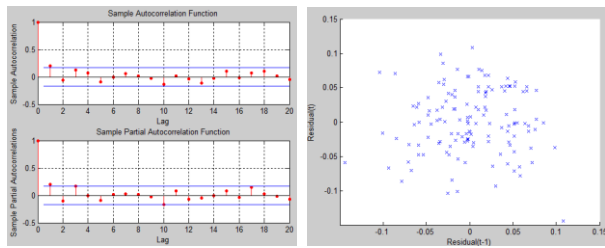


c) Residuals Autocorrelation Function d) Residuals vs. Lagged Residuals Scatter

Figure 5. The Computation Results of ARIMA Method



a) The Standardized Residuals plot b) Residuals QQ plot



c) Residuals Autocorrelation Function Lagged and Partial Autocorrelation Function Residuals Scatter d) Residuals vs. Lagged and Partial Autocorrelation Function Residuals Scatter

Figure 6. The Computation Results of the Modified ARIMA Method

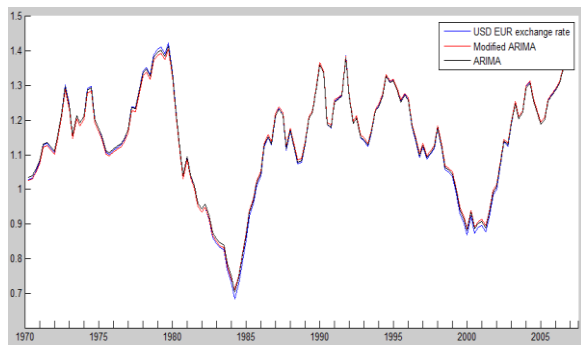


Figure 7. Residuals vs. Lagged Residuals Scatter of Modified ARIMA Method

TABLE I. Fitted results of ARIMA and modified ARIMA

Estimated Coefficients		Estimate	SE	tStat	pValue
ARIMA	AR(1)	0.2597	0.074	3.5081	0.0021
Modified ARIMA	AR(1)	0.1922	0.086	2.2226	0.0279
	β_{\max}	-0.0324	0.014	-2.122	0.0286
	β_{\min}	0.0173	0.0783	-2.209	0.0267

TABLE II. Error performances of ARIM and modified ARIMA

Error Performances	ARIMA	Modified ARIMA
Mean Error	8.1571e-06	1.8604e-18
Mean Absolute Error	5.4740	5.0908
Standard Deviation	0.0590	0.0485

IV. CONCLUSIONS

In this paper, We proposes a modified ARIMA model based on the past maximum and minimum value of series to solve time series modelling tasks. The premise of this problem is that the past extreme values(max/min values) will effects the behaviours of time series. The algorithm is tested on USD-EUR exchange rate time series. The experimental results show that it is possible to improve the performance by using the past max/min value.

ACKNOWLEDGEMENTS

This work is supported by the Humanities and Social Science Research Planning Project of College of Jiangxi Province, China(Project No.GL1115), and also supported by the National Natural Science Foundation of China (Grant No. 61363075). This article content has no conflict of interest.

REFERENCES

- [1] Box G, G Jenkins, and G C. Reinsel, Time series analysis: forecasting and control, 4th ed, John Wiley & Sons, Inc. 2008.
- [2] Gardner Jr E, "Exponential smoothing: The state of the art," International Journal of Forecasting, vol.22(4),2006, pp: 637-666, doi: 10.1002/for.3980040103.
- [3] Taylor, James W. Snyder, and Ralph D. "Forecasting intraday time series with multiple seasonal cycles using parsimonious seasonal exponential smoothing," Special Issue on Forecasting in Management Science, vol.40(6), 2009, pp: 748-757, doi:10.1016/j.omega.2010.03.004.
- [4] Theodosiou and Marina, "Forecasting monthly and quarterly time series using STL decomposition," International Journal of Forecasting, vol.27(4), 2011, pp: 1178-1195, doi:10.1016/j.ijforecast.2010.11.002.
- [5] White H., "Learning in artificial neural networks: A statistical perspective," Neural computation, vol.1(4), 1989,pp: 425-464, doi: 10.1162/neco.1989.1.4.425.
- [6] Charles Wong and Massimiliano Versace, "CARTMAP: a neural network method for automated feature selection in financial time series forecasting," Neural Computing and Applications, vol.21(5) 2012, pp: 969-977, doi:10.1007/s00521-012-0830-8.
- [7] Hussain A and A Knowles, "Financial time series prediction using polynomial pipelined neural networks," Expert Systems with Applications,vol. 35(3), 2008, pp: 1186-1199, doi: 10.1016/j.eswa.2007.08.038
- [8] Zhang J and H Chung, "Chaotic time series prediction using a neuro-fuzzy system with time-delay coordinates," IEEE Transactions on Knowledge and Data Engineering, vol.20(7) ,2008, pp: 956-964, doi:10.1109/TKDE.2008.35.
- [9] Szeto K, and L Fong, "How adaptive agents in stock market perform in the presence of random news: A genetic algorithm approach," Intelligent Data Engineering and Automated Learning-IDEAL 2000. Data Mining, Financial Engineering, and Intelligent Agents,2009, pp: 505-510, doi:10.1007/3-540-44491-2_74.
- [10] Guo, ZQ, Wang, HQ, and Liu, Q, " Financial time series forecasting using LPP and SVM optimized by PSO," SOFT COMPUTING, vol.17(5), 2013, pp: 805-818, doi:10.1007/s00500-012-0953-y.