From TSA To A Modified ODE

Meiyu XUE[1], Choi-Hong LAI[2]

1 College of Mathematics and Computer Science, Fuzhou University, Fuzhou, Fujian, China, 350316

2 Department of Mathematical Sciences, University of Greenwich, London, UK, SE10 9LS

Abstract

In understanding Big Data, people are interested to obtain the trend and dynamics of a given set of temporal data which in turn can be used to predict possible futures.

This paper examines a time series analysis (TSA) method and an ordinary differential equation (ODE) approach in modelling the price movements of petroleum price and of three different bank stock prices over a time frame of three years. Computational tests consist of a range of data fitting models in order to understand the advantages and disadvantages of these two approaches. A modified ODE model, with different forms of polynomials and periodic functions, is proposed. Numerical tests demonstrated the advantage of the modified ODE approach. Computational properties of the modified ODE are studied.

Keywords: TSA, ARIMA, ODE, MAPE

1 Introduction

Observing the trend and forecasting the future are always required in all kinds of market. In understanding big data, people are more interested to obtain the trend and dynamics of a given set of temporal data which in turn can be used to predict possible futures.

Classic statistical methods are usually used to perform the task, such as regression analysis, cluster analysis and so on. As a branch of statistics, time series analysis (TSA) is very popular for modelling temporal data [1]. People did great efforts in applying TSA in temporal market analysis. In 1970, Box and Jenkins proposed autoregressive integrated moving average (ARIMA) model [2]. In order to handle time-varying property of variance, Engle (1982) derived autoregressive conditional heteroscedasticity (ARCH) model [3]. Next, Bollerslev (1986), Glosten et al. (1991) and Nelson (1991) derived Generalized ARCH (GARCH) model, Threshold ARCH (TARCH) model and Exponential ARCH (EARCH) model respectively.

One of the disadvantages of these statistical methods is that people need a large amount of market data. In such cases, numerical methods, Ordinary Differential Equations (ODE) [4], Partial Differential Equations (PDE) or Stochastic Differential Equations (SDE), would be taken into account.

This paper examines a modified ODE approach and compares it with TSA in modelling the price movements of petroleum price and of three different bank stock prices over a time frame of three years. The market data were obtained from the official web page [5]. Computational tests consist of a range of data fitting models in order to understand the advantages and disadvantages of these two approaches. Then, a modified ODE model, with different forms of polynomials and periodic functions, is proposed. Numerical tests demonstrate the advantages of such modification. Computational properties of the modified ODE are studied.

The rest of this article is organized as follows. In section 2, ARIMA model and an ODE method are introduced and then results of them are compared. Section 3 presents the modification of the ODE model. The empirical analysis is shown in section 4. Finally, the article is concluded in section 5.

2 The ARIMA and the existing ODE models

2.1 Fundamental methods

Time Series Analysis (TSA) comprises methods for analysing temporal data. Models for time series data contain many forms representing different stochastic processes. In statistics and econometrics, and in particular in time series analysis, the autoregressive integrated moving average (ARIMA) models are often applied in some cases where data show evidence of nonstationary. Wan J. and Wen Z. found ARCH model didn't always show better compared to ARIMA model [6]. For simplicity, attention was only given to ARIMA model in this section.

ARIMA models are generally denoted ARIMA(p,d,q) where parameters $p,d,$ and q are nonnegative intergers, p is the order of the autoregressive model, d is the degree of differencing, and q is the order of the moving average model [7].

Given a time series of data y_t where t is an integer index and y_t is a real number, then an ARIMA(p,d,q) model is given by:

$$\left(1-\sum_{i=1}^{p}\varphi_i B^i\right)\Delta^d y_t=\left(1-\sum_{i=1}^{q}\theta_i B^i\right)\varepsilon_t \qquad (1)$$

Where B is the lag operator such that

$$B^k y_t = y_{t-k}, \qquad k=0,1,2,\dots$$

And the symbol Δ is the differencing operator such that

$$\Delta^d y_t=(1-B)^d y_t, \quad d=0,1,2,\dots$$

The φ_i are the parameters of the autoregressive part, the θ_i are the parameters of the moving average part and the ε_t are white noise error terms.

Case $d = 0$ corresponds to the ARMA(p, q) model. What's worth mentioning is that the ARMA(p, q) models are used for stationary data. If data is nonstationary, one should try ARIMA(p, d, q) models. Finally, one could determine the best model according to the Akaike information criterion (AIC) or the Bayesian information criterion (BIC).

Next, consider the Cauchy initial value problem,

$$y' = ay, \quad y(t_0) = y_0 \qquad (2)$$

One can solve (2) by means of numerical integration or obtain an analytic solution if a is given. It is also possible to calibrate a at different time intervals.

One approach for solving (2) is given by Lascsáková [8]. The particular solution of problem (2) is

$$y = y_0 e^{a(t-t_0)}$$

Substituting the point (t_1, y_1) to this particular solution, we have

$$y_1 = y_0 e^{a(t_1-t_0)} \qquad (3)$$

From equation (3), a is obtained as below,

$$a = \frac{1}{t_1 - t_0} \ln\left(\frac{y_1}{y_0}\right) \qquad (4)$$

At the next time t_2 , one has

$$y_2 = y_1 e^{a(t_2-t_1)} \qquad (5)$$

From equation (5), a is obtained again,

$$a = \frac{1}{t_2 - t_1} \ln\left(\frac{y_2}{y_1}\right) \qquad (6)$$

Generalizing the previous principle, one can get the solution of problem (2) in the following form

$$y_{i+1} = y_i e^{a(t_{i+1}-t_i)} \qquad (7)$$

Here, $$a = \frac{1}{t_i - t_{i-1}} \ln\left(\frac{y_i}{y_{i-1}}\right) \qquad (8)$$

2.2 Comparing TSA and the existing ODE model

This section compares the time-domain TSA method given in equation (1) and the ODE approach given in equation (2) in modelling the price movements of petroleum price and of two bank stock prices over a time frame of three years.

For the observed data $(t_0, y_0), (t_1, y_1), (t_2, y_2), \ldots, (t_n, y_n)$ of a time series of data y_t, the absolute percentage error (APE) and mean of APE (MAPE) are chosen applied as the criterion to evaluate the models in this paper. They are defined as:

$$APE_i = \frac{|\hat{y}_i - y_i|}{y_i}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} APE_i$$

Let the symbol $\hat{y}_i$ denote the approximated value at the time (day/month/year) t_i, the symbol y_i denotes the observed value at the time (day/month/year) t_i. Although the MAPE is less often used the mean square error (MSE) and the mean absolute error (MAE), it is a more natural error measure, and has several advantages [9].

For petroleum data in 2013, an appropriate model is ARIMA(0, 1, 13)

$$x_t = x_{t-1} + \varepsilon_t - 0.1338\varepsilon_{t-5} - 0.1226\varepsilon_{t-13}$$

The calculated results are shown in table 1. Table 1 indicates that all the APEs of TSA are less than 5%. There are 249 APEs of ODE and only one APE of ODE is not less than 5% but less than 7.5%. It seems that there is almost no difference between these two approaches in this sense. But, the MAPE of TSA is less than that of ODE. It's well known that APE and MAPE are the smaller the better. AS a consequence, people would prefer the TSA method in this market.

Table 1. Comparing ODE and TSA of petroleum price(2013)

APE	ODE	TSA
[0, 5%)	249	250
[5%, 7.5%)	1	0
[7.5%, 10%)	0	0
[10%, 1)	0	0
MAPE	1.2597%	0.8817%

For petroleum data in 2014, an appropriate model is ARIMA(6, 1, 0)

$$x_t = -0.1662 + 0.8299x_{t-1} + 0.1701x_{t-2} + 0.1727x_{t-6} - 0.1727x_{t-7} + \varepsilon_t$$

For petroleum data in 2015, an appropriate model is ARIMA(1, 1, 0)

$$x_t = 0.8695x_{t-1} + 0.1305x_{t-2} + \varepsilon_t$$

The results of comparing ODE and TSA of petroleum price (2014, 2015) are shown in table 2 and table 3.

Table 2. Comparing ODE and TSA of petroleum price(2014)

APE	ODE	TSA
[0, 5%)	235	248
[5%, 7.5%)	9	1
[7.5%, 10%)	4	0
[10%, 1)	2	1
MAPE	1.7445%	1.0670%

Table 3. Comparing ODE and TSA of petroleum price(2015)

APE	ODE	TSA
[0, 5%)	198	227
[5%, 7.5%)	30	18
[7.5%, 10%)	15	5
[10%,1)	7	0
MAPE	3.4100%	2.2922%

Similarly, this paper also worked on the share values of two banks over a period of about 750 days. The results were obtained in table 4.

Table 4. Comparing ODE and TSA of bank share values

	Barclays bank		Lloyds bank	
APE	ODE	TSA	ODE	TSA
[0, 5%)	623	694	626	706
[5%, 7.5%)	90	54	82	39
[7.5%, 10%)	39	14	41	19
[10%, 1)	20	10	22	8

From the above examples, TSA seems to show better results compared to ODE. However, it is possible to modify the form of the derivative given in (2).

3 Modification of the ODE model

There are different ways of modifying the ODE model given in (2). For example the form of the derivative given in (2) may be changed. This section introduces several alternatives in such modification.

If the data y_t is not an exponential function of time variable t, equation (2) may be modified. Equation (8), which in fact defined the parameter a as a piecewise function, may also be modified. Hence, the modification consists of the derivative itself and the parameter a. After modification, problem (2) can be transformed into:

$$y' = f(y), \qquad y(t_0) = y_0 \tag{9}$$

or

$$y' = f(t)y, \qquad y(t_0) = y_0 \tag{10}$$

Several different forms of $f(.)$ as listed in table 5 have been tested.

Table 5. Possible forms of $f(x)$

$\alpha \sin x + \beta$	$\alpha e^x + \beta$
$\alpha x + \beta$	$\frac{\alpha x}{\beta + x}$
$\alpha x^2 + \beta x + \gamma$	$\alpha \cdot 2^{-x/\beta}$
$\alpha \ln x + \beta$	$\alpha \beta^x$
$\alpha \sin x + \beta x + \gamma$	$\alpha + \beta \gamma^x$
$\alpha \sin x + \beta \ln x + \gamma$	$e^{\alpha + \beta \gamma^x}$
$\alpha \ln x + \beta x + \gamma$	$\frac{1}{\alpha + \beta \gamma^x}$
$\alpha \sin x + \beta x + \gamma \ln x + \delta$	$\alpha e^{\beta x} + \gamma$
……	……

Problems (9) and (10) are actually separable differential equations. A general form which leads to a non-separable differential equation is given as below:

$$y' = a(t)y + s(y) \tag{11}$$

It should be noted that $s(y)$ is itself a function of y. The forms of $a(t)$ and $s(y)$ could be primary functions, such as exponential function, trigonometric function, logarithmic function, and power function. Primary functions could be expanded to power series under special conditions. Furthermore, sometimes the data y_t might be periodic. Henceforth, the derivative y' might consist of a polynomial and a periodic function. A generalized model is given as below:

$$y' = g(t, y) = \left(\sum_{i=0}^{M} a_i t^i\right) y + b_0 + \sum_{j=1}^{N} b_j \sin\left(\frac{2\pi j y}{\theta} + c_j\right) \tag{12}$$

The unknown parameters $a_0, a_1, \dots, a_M, b_0, b_1, \dots, b_N, c_1, c_2, \dots, c_N, \theta$ are estimated according to the approach for inverse problem[10]. The numerical solution is obtained by 4th order Runge-Kutta one-step method, which is the most widely known member of the Runge-Kutta family [11].

$$y_{n+1} = y_n + \frac{1}{6}h(k_1 + 4k_2 + k_3),$$

where

$$h = t_{n+1} - t_n,$$

$$k_1 = g(t_n, y_n),$$

$$k_2 = g\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_1\right),$$

$$k_3 = g\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_2\right),$$

$$k_4 = g(t_n + h, y_n + hk_3)$$

The bigger the values of M or N is, the more parameters one should estimate. Further more the bigger the values of M or N is, the more likely the Jacobian matrix is singular. It should be noted that the bigger the values of M or N is, the more difficult the computational work. From all these points of view, one would usually take M or N to be less than four.

4 Empirical analysis

Applying the above ODEs (9), (10), (11) and (12) to the petroleum data and three bank share prices, some improved results are obtained. In practice, one would prefer equation (12),which consists of a polynomial and a periodic function. The results are shown in tables 6 - 11.

Table 6. MAPE of petroleum (2013) according to equation (12)

MAPE		N		
		1	2	3
M	0	0.89524%	0.88784%	0.88896%
	1	Singular	0.89059%	0.88722%
	2	Singular	singular	singular
	3	Singular	singular	singular

Table 7. MAPE of petroleum (2014) according to equation (12)

MAPE		N		
		1	2	3
M	0	1.08986%	1.11213%	1.10441%
	1	Singular	singular	1.09418%
	2	Singular	singular	singular
	3	Singular	singular	singular

Table 8. MAPE of petroleum (2015) according to equation (12)

APE		N		
		1	2	3
M	0	2.27096%	2.26665%	2.30960%
	1	2.25025%	2.26416%	2.26109%
	2	Singular	singular	singular
	3	Singular	singular	singular

Table 9. MAPE of Barclays bank according to equation (12)

APE		N		
		1	2	3
M	0	2.23898%	singular	singular
	1	2.25212%	singular	singular
	2	Singular	singular	singular
	3	Singular	singular	singular

Table 10. MAPE of Lloyds bank according to equation (12)

APE		N		
		1	2	3
M	0	2.29275%	2.30361%	2.31223%
	1	2.29840%	2.29030%	2.31427%
	2	Singular	singular	singular
	3	Singular	singular	singular

Table 11. MAPE of RBS bank according to equation (12)

APE		N		
		1	2	3
M	0	2.21889%	2.21824%	2.21455%
	1	2.21350%	2.21267%	singular
	2	Singular	singular	singular
	3	Singular	singular	singular

In the above five tables, one could choose the best model with the smallest MAPE. For example, for petroleum data in 2013, the smallest MAPE occurs when $M = 1$ and $N = 3$. For Barclays bank, the smallest MAPE occurs when $M = 0$ and $N = 1$. The parameters are estimated according to the approach for inverse problem. Results are shown as table 12.

Table 12. The estimated parameters

Parameters	Petroleum			bank		
	2013	2014	2015	Barclays	Lloyds	RBS
a_0	-0.03031	0.01215	-0.02185	-0.00947	-0.00638	-0.00914
a_1	0.00001		-0.00004		-0.00001	-0.00002

b_0	2.90619	-1.30755	1.26420	2.36196	0.43602	4.88964
b_1	0.24353	-0.22307	0.16355	1.31137	-0.07560	0.82936
θ	0.49755	0.74135	0.46282	0.50023	0.49999	0.49879
c_1	-4.50499	398.30817	12.81797	1.56126	49.13539	-8.37850
b_2	-0.17368				0.18431	0.84405
c_2	-11.26577				1.52255	-5.84867
b_3	0.11570					
c_3	-15.31979					

Now, one can compare the results of equation (2) and equation (12). As one can see in the table 13 that the modified ODE given in (12) does improve the results in the sense of MAPE.

Table13. MAPE compared with different $\boldsymbol{y'}$

	Model given in (2)	Model given in (12)
Petroleum 2013	1.2597%	0.8872%
Petroleum 2014	1.7445%	1.0809%
Petroleum 2015	3.4100%	2.2471%
Barclays	3.2526%	2.2390%
Lloyds	3.1877%	2.2855%
RBS	3.0739%	2.2127%

5 Conclusions

In order to obtain the trend and forecast the future with higher accuracy, the idea of modifying the ODE model is proposed and the form as equation (12) seems to be the best modification. Based on the obtained result, it can be stated that, such modification provides good understanding of the trend and the dynamics of the price movement. This provides a good way forward in forecasting. Furthermore, in comparing with statistical methods, numerical methods for ODEs show the advantage that fewer historical market data is required.

Finally, recall the following problem which involves a deterministic function $\mu(t,y)$:

$$y' = \mu(t,y) \qquad (13)$$

This paper provides an insight on various forms of the right-hand side of problem (13). The authors anticipate that this work will lead to a systematic and an accessible way of forecasting the dynamic market, particularly some of the price movements in the financial market.

References

[1] Tsay, R. S. (2005). Analysis of Financial Time Series, 2nd ed. New York: John Wiley & Sons.

[2] Engle R F. Autoregressive conditional heteroskedasticity with estimates of the variances of UK inflation [J]. Econometrica, 1982, 31: 987-1008

[3] Bolleslev T. Generalized autoregressive conditional heteroskedasticity [J]. Journal of Econometrics, 1986, 31: 307-327

[4] Gear, C. W. (1971). Numerical Initial Value Problems In Ordinary Differential Equations. New Jersey: Prentice-Hall.

[5] http://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm

[6] Wan J, Wen Z. The comparison of ARIMA model and ARCH model application in Hong Kong stock price index prediction [J]. Journal of Applied Statistics and Management, 2001, 20: 1-4

[7] https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average

[8] Lascsáková M. the Numerical Model of Forecasting Aluminium Prices by Using Two Initial Values, Studies of the University of Žilina, Mathematical Series, 2009, 23:85-92

[9] A De Myttenaere et al. Mean Absolute Percentage Error for regression models [J]. Neurocomputing, 2016, 192:38-48

[10] Richard C. Aster et al. (2012). Parameter Estimation and Inverse Problems, 2nd ed. Amsterdam: Elsevier.

[11] Delin Tan, Zheng Chen. On A General Formula of Fourth Order Runge-Kutta Method [J]. Journal of Mathematical Science & Mathematics Education, 2012, 7(2): 1-10

[12] Cryer, J. D. and Chan, K. S. (2008). Time Series Analysis with Applications in R , 2nd ed. New York: Springer-Verlag.