Modern Applied Science

www.ccsenet.org/journal.html

Vol. 3, No. 5
May 2009

Applying Time Series Analysis Builds Stock Price Forecast Model

Jun Zhang (Corresponding author)
Department of Science, Yanshan University, Hebei 066004, China
Tel: 86-335-865-7691 E-mail:zhangjunmath@163.com
Rui Shan &Wenfang Su
Department of Science, Yanshan University, Hebei 066004, China

Abstract

Time series analysis is a theory that used random process and mathematical statistics theory to analyze time .It is apply comprehensive to national economy macroeconomic adjustment and control, area complex development plan, enterprise operating management, market potential forecasting, weather hydrology prediction. It is an important means for estimation and forecast. The stock price has very deep effect to the economic benefits of the nation and the macro-economy policy. So people pay close attention to it. In this article, SSE composite index of one year is fitted two kinds of time series models, then forecast in short-time. Comparing the estimated valve with the true valve, the result is the relative error is small. So I think the model is suited to the data. At last, compare the two models.

Keywords: SSE composite index, Time series analysis, ARIMA model, ARCH model

In the production and scientific study, through carrying on observation and measurement to some group of or a variable, obtained a series of discrete digit in a series of times, the series is called time series. Put the series of discrete digitals form a series set. The series set is called time series. The time series analysis is a mathematical technique that through curve fitting and parameter estimation establish mathematical model with the system observation data and forecast the future trend. Stock price fluctuations has reflected the state economy change to a certain degree and affected the national macroeconomic policy. Therefore through economic indicator's change forecast stock price trend is the topic which throughout the people explore. Taken of time series analysis method not only is the cost low but also accuracy. But the stock price fluctuations is influenced by many kinds of the economic agent and the non-economic agent factors, for example the economic agent of market rate, fiscal levy, price level, domestic international political situation, war disaster, the non-economic agent factor of the investors mood and the transaction limits stipulation. This paper uses the time series analysis two kind of model fitting stock price trend.

1. ARIMA model

ARIMA (p, d, q) the model has the following structure:

$$\begin{cases} \varphi(B)\nabla^d x_t = \theta(B)\varepsilon_t \\ E(\varepsilon_t)=0 \quad \operatorname{var}(\varepsilon_t)=\sigma_\varepsilon^2 \quad E(\varepsilon_t\varepsilon_s)=0 \quad s\neq t \\ E(x_s\varepsilon_t)=0 \quad \forall s>t \end{cases}$$

The abbreviated formula is $\nabla^d x_t = \dfrac{\theta(B)}{\varphi(B)}\varepsilon_t$, $\{\varepsilon_t\}$ zero average value white noise series.

$\varphi(B)=1-\varphi_1 B-\varphi_2 B^2-\cdots-\varphi_p B^p$ is the autoregressive coefficient polynomial.

$\theta(B)=1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q$ is the moving average coefficient polynomial.

$\nabla^d=(1-B)^d$, ∇ is the difference operator, d is the difference order, B is backward shift operator $Bx_t = x_{t-1}$.

2. ARIMA model modeling step

2.1 Data processing

First judge the series whether to be steady or not by observing the autocorrelation coefficient figure and the partial autocorrelation coefficient figure. Second if the series is not steady, carries on the series difference or the season difference for eliminating the tendency fluctuation and the season fluctuation. If the difference can not make the series to be steady, taking the logarithm to the series to eliminates the different variance, then make the series difference. Third the steady time series is carried on the white noise check to judge whether it to be related.

2 Fix the step of the model

First observe the autocorrelation coefficient figure and the partial autocorrelation coefficient figure of steady non-white noise series to fix model autoregressive step p and moving average step q. But the model is not necessarily only. Second selects the suitable estimation method to estimate the unknown parameter the value.

2.3 Optimize model

Carries on the residuals white noise check and the parameter significance check to the fitting model, selects superior model in the through the check models.

2.4 Model prediction

Using fitting model forecast the series short-term trend.

3. ARCH model

The ARCH model full title is auto regressive conditional different variance model.

Its complete structure is

$$\begin{cases} x_t = f\left(t, x_{t-1}, x_{t-2}\cdots\right) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t}\, e_t \\ h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_i \varepsilon_{t-j}^2 \end{cases}$$

$f\left(t, x_{t-1}, x_{t-2}\cdots\right) + \varepsilon_t$ is the $\{x_t\}$ Auto-Regressive model; $e_t \overset{i.i.d}{\sim} N(0,1)$.Series$\{\varepsilon_t\}$ is zero average value random residuals series and have different variance $\operatorname{var}(\varepsilon_t) = h_t$,different variance equal to average value of residual square series: $E(\varepsilon_t^2) = h_t$.Under normal distribution supposition $\varepsilon_t \big/ \sqrt{h_t} \sim N(0,1)$,With residual square series autocorrelation coefficient inspects the different variance function the autocorrelation. If there is some autocorrelation coefficient is not zero, the residual square series autocorrelation coefficient is impermanent zero and different variance function exist autocorrelation. Then it is possible that through constructing residual square series autoregressive model to fit the different variance function $h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_i \varepsilon_{t-j}^2$.

4. Example analyses

The figure 1 shows the Shanghai Composite Index closing price data from January 18th, 2007 to January 13, 2008.

4.1 Establish the ARIMA model

The figure demonstrates that the series is not steady. The series has a certain trend, but does not have the season effect. Because of that reason carries the first order difference on this sequence. The following is autocorrelation and the partial autocorrelation coefficient chart and ADF Unit Root tests diagram. Because figure 2 and figure 3 shows autocorrelations and the partial autocorrelation coefficient of the series after first order difference are first order truncation, the first order difference sequence is steady. Figure 4 show all of the P value is smaller than 0.001.It is also explained that the sequence is steady. Both the two checks have confirmed the first order difference series is steady.

Figure 5 is the white noise check figure. It shows that under the level of =0.05, P value is all smaller than α .The result indicated that first order difference series is non-white noise sequence. Because the first order difference sequence is steady non-white noise sequence, it is can be used to established the ARIMA model.

Because the autocorrelation and the partial autocorrelation coefficient of the difference series is first order truncation try to establish ARIMA(1,1,0) model and ARIMA(0,1,1) model. Make comparison the two models

Table 1 shows the AIC value and the SBC value and erroneous estimated value of the AR (1) model are smaller than those of MA (1) model. It can be determined that ARIMA (1,1,0) model is better than ARIMA (0,1,1) model. Therefore designate ARIMA (1,1,0) model to fit the sequence. The chart 6 is the model parameter estimated value and the significance check and the residual white noise check result. Because the P value is bigger than 0.05 in the residual white noise check result, the residual is the white noise series.

The establishment model is $\left(1+0.29161B\right)\left(1-B\right)x_t = 11.08637\varepsilon_t$.

Figure 7 shows fitting figure of ARIMA (1,1,0) model. The figure shows that the fitting figure and the original figure superpose basically. So we can decide the model build successfully.

The modeling goal is the forecast. Table 2 shows the model short-term predicted results. Comparing the predicted values with the true values, the relative errors are minor. Relative error's average value is 0.04072.

4.2 Establish the ARCH model

Check the residual series which have rejected the tendency item to judge whether it to be the autocorrelation or the different variance. Figure 8 shows residual series DW check amount P value is smaller than 0.001; it explained that the residual series is the autocorrelation. And the PortmanteaQ statistics and Lagrange number multiplication LM check amount P value is smaller than 0.001; it explained that residual series is different variance

Set up the model after screening:

$$x_t = 2833 + 13.4309t + u_t$$

$$u_t = 0.8257u_{t-1} + 0.1262u_{t-2} + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t}e_t$$

$$h_t = 14732 + 0.1966\varepsilon_{t-1}^2$$

After testing the parameters of the model are significant. The whole model's R^2 value is 0.9842. It is to meet the requirements. Figure 9 shows ARCH model fitting figure. The red is the fitted curve, and the black is the original curve.

Table 3 is the short-term predicted of model results. Comparing the predicted values with the true values, the relative errors are minor. Relative error's average value is 0.01902.

5. Conclusion

Comparing the two models, judging form the predicted outcome, the ARCH model relative error is smaller, so ARCH model fitting better than the ARIMA model. It is mainly because of the stock price change misalignment behavior; ARIMA model for time series prediction only considers the characteristics of the time series, without taking into account the stock price itself is affected by many unpredictable and complex factors; these factors can only indicated by the stochastic disturbing term in the ARIMA model that is actually unable to display in the forecast expectation value. In addition, ARIMA models generally assume that the model residuals are zero average value and same variance, but in fact stock index series of china often are different variance.

References

Brockwell P, J, & Davis. R, A. (2001).*Time series: theory and method* .Springer Press.

GAO, H., X. (1998) .*SAS system SAS/ETS the software user manual*. China Statistics Press.

Hen, B., L. (2008).Stock price prediction model for the optimal selection, *Statistics and Decision,* (6), 135-137.

KITAGAWA ,G,A (1981).nonstationary time series model and its fitting by recursive technique. *J Time Ser Anal.*2(2):103-116.

Wang, Y. (2005).*Application time series analysis*. People's University of China publishing house.

WONG, C, S. (2000) .On a mixture autoregressive model, *J of the Royal Statistical Society* (62) part1:95-115.

Table 1. model comparison form

	AIC	SBC	Std Error
ARIMA(1,1,0)	3014.754	3021.732	122.1921
ARIMA(0,1,1)	3026.719	3033.705	122.0717

Table 2. ARIMA (1,1,0) model predicted outcome

Date	True values	Predicted values	Relative errors
20080114	5497.90	5490.87	0.0012
20080115	5443.79	5503.47	0.0190
20080116	5290.61	5514.19	0.0422
20080117	5151.63	5525.47	0.0725
20080118	5180.51	5525.47	0.0687

Table 3. ARCH model predicted outcome

Date	True values	Predicted values	Relative errors
20080114	5497.90	5518.28	0.0037
20080115	5443.79	5531.48	0.0161
20080116	5290.61	5480.21	0.0358
20080117	5151.63	5334.09	0.0354
20080118	5180.51	5201.56	0.0041

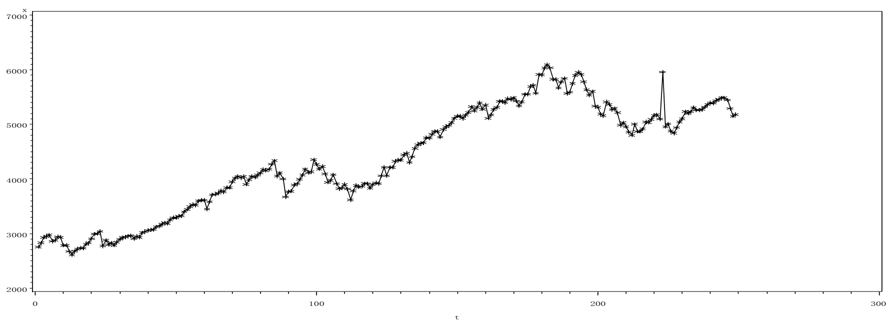

Figure 1. Shanghai Composite Index series figure

```
                                   Autocorrelations
Lag   Covariance    Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      STD Error
 0    16003.489     1.00000     |                    |********************|          0
 1    -4474.896     -.27962     |              ******|   .                |      0.063500
 2    660.600       0.04128     |                  . |* .                 |      0.068285
 3    964.252       0.06025     |                  . |* .                 |      0.068385
 4    -717.344      -.04482     |                  .*|   .                |      0.068599
 5    -917.712      -.05734     |                  .*|   .                |      0.068717
 6    719.487       0.04496     |                  . |* .                 |      0.068910
 7    -1074.005     -.06711     |                  .*|   .                |      0.069028
 8    184.910       0.01155     |                  . |   .                |      0.069290
 9    -1411.894     -.08822     |                 .**|   .                |      0.069298
10    542.795       0.03392     |                  . |* .                 |      0.069750
```

Figure 2. Autocorrelation coefficient figure

```
                Partial Autocorrelations
Lag   Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
 1    -0.27962      |              ******|   .               |
 2    -0.04004      |                  . *|   .              |
 3     0.06635      |                  .   |* .              |
 4    -0.00879      |                  .   |   .             |
 5    -0.08084      |                  .**|   .              |
 6     0.00296      |                  .   |   .             |
 7    -0.05079      |                  . *|   .              |
 8    -0.01494      |                  .   |   .             |
 9    -0.10502      |                  .**|   .              |
```

Figure 3. Partial autocorrelations coefficient figure

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < T au	F	Pr > F
Zero Mean	0	-314.262	0.0001	-20.75	<.0001		
	1	-332.927	0.0001	-12.88	<.0001		
Single Mean	0	-316.073	0.0001	-20.87	<.0001	217.85	0.0010
	1	-341.894	0.0001	-13.01	<.0001	84.62	0.0010
Trend	0	-316.663	0.0001	-20.88	<.0001	217.99	0.0010
	1	-344.224	0.0001	-13.02	<.0001	84.81	0.0010

Figure 4. ADF Unit Root Tests figure

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > Chi Sq	--------------------Autocorrelations--------------------					
6	22.84	6	0.0009	-0.280	0.041	0.060	-0.045	-0.057	0.045
12	28.59	12	0.0045	-0.067	0.012	-0.088	0.034	0.092	0.011
18	30.11	18	0.0364	-0.001	0.074	0.005	0.003	0.015	-0.008
24	35.26	24	0.0646	0.025	-0.084	-0.047	0.052	0.023	-0.075

Figure 5. white noise inspection figures

Unconditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	11.08637	6.08708	1.82	0.0698	0
AR1, 1	-0.29161	0.06174	-4.72	<.0001	1

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > Chi Sq	--------------------Autocorrelations--------------------					
6	3.28	5	0.6566	-0.014	-0.025	0.072	-0.050	-0.066	0.015
12	11.09	11	0.4356	-0.061	-0.032	-0.087	0.043	0.123	0.040
18	14.20	17	0.6532	0.026	0.093	0.033	0.017	0.030	0.016
24	20.24	23	0.6276	0.011	-0.106	-0.087	0.023	-0.020	-0.053

Figure 6. ARIMA (1,1,0) model output result and residual check

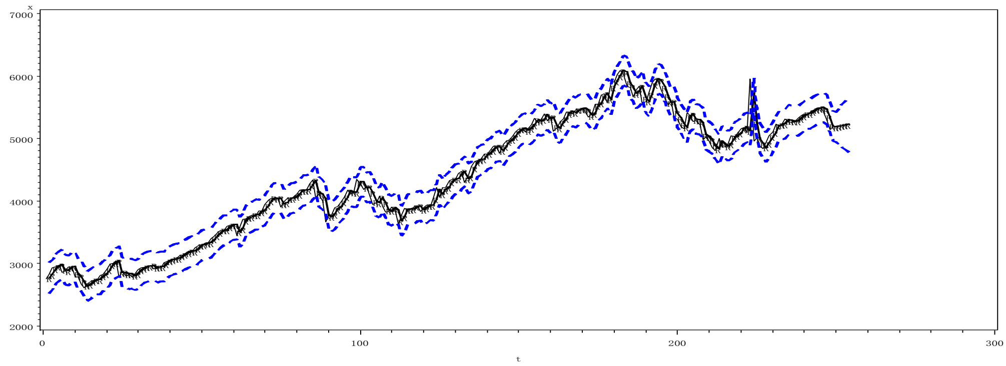

Figure 7. ARIMA (1,1,0) model fitting figure

Q and LM Tests for ARCH Disturbances

Order	Q	Pr > Q	LM	Pr > LM
1	196.6829	<.0001	198.7025	<.0001
2	351.0688	<.0001	198.7249	<.0001
3	470.0340	<.0001	198.9361	<.0001
4	560.3033	<.0001	199.0966	<.0001
5	634.5792	<.0001	199.6533	<.0001
6	697.4607	<.0001	199.7831	<.0001
7	745.4301	<.0001	200.5298	<.0001
8	788.5607	<.0001	201.6425	<.0001
9	829.3971	<.0001	202.1698	<.0001
10	871.2981	<.0001	202.5459	<.0001
11	916.2612	<.0001	202.6061	<.0001
12	957.3173	<.0001	203.2968	<.0001

Figure 8. Different variance check figure

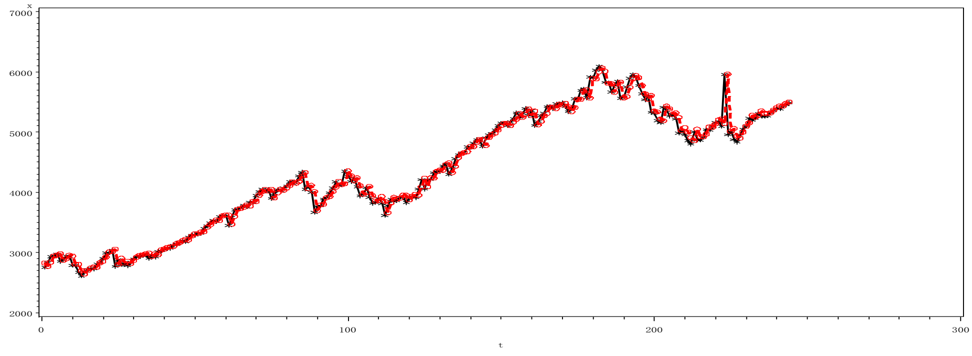

Figure 9. ARCH model rendering