CSCanada
Higher Education of Social Science
Vol. 6, No. 3, 2014, pp. 118-122
DOI:10.3968/4991

ISSN 1927-0232 [Print]
ISSN 1927-0240 [Online]
www.cscanada.net
www.cscanada.org

Analysis of Shanghai Composite Index Variation Based on Regression Analysis

CUI Yujie[a],*; XI Qipu[a]; HAO Junzhang[a]

[a]College of science, North China University of Technology, Beijing, China.
*Corresponding author.

Supported by College Student Research and Career-creation Program of Beijing (2013).

Received 12 January 2014; accepted 23 March 2014
Published online 26 May 2014

Abstract
In this paper, through collecting data of Shanghai Composite Index since 2007, we analyze overall trend of the Shanghai stock market after the financial crisis, and carry on the forecast to the future trend in order to provide a meaningful guidance for people's investment securities. Because the fitting results of simple regression is not good, we consider the long-term trend, seasonal fluctuations, cyclical fluctuations, irregular variables and other factors. We also add lagged variables and establish an ARIMA model through SPSS statistical analysis software. The fitting degree of model we built is good and the effect of prediction is significant improvement in the analysis.

Key words: Shanghai composite index; Time series; ARIMA model

Cui, Y. J., Xi, Q. P., & Hao, J. Z. (2014). Analysis of Shanghai Composite Index Variation Based on Regression Analysis. *Higher Education of Social Science*, *6*(3), 118-122. Available from: URL: http://www.cscanada.net/index.php/hess/article/view/4991
DOI: http://dx.doi.org/10.3968/4991

INTRODUCTION

From 2008 America financial crisis, the world financial conditions are always turmoil and by the flare-up of the European debt crisis is also effected the development of world economy. Although the economy in China is still doing well, but the Shanghai Composite Index is not as good as our economy. stock market also has a lot of risks so it will be much more important for us to study the variation of composite index. Dai wensheng and some other proficient use the support vector regression method, Mezali, H. and J.E. Beasley use the method of quantile regression to study composite index [2]. But we also found that use the method of linear regression model (LRM) to increase and lag factors can also get good effect on study composite index. Therefore, we choose the data of Shanghai Composite Index from 2007 to now as our object of study, use SPSS statistical analysis software and time series method to analysis and establish an ARIMA model. Through this way, we hope we can find out the change variation of Shanghai stock and look forward that our study can help people invest stock, service to the health development of economy.

1. ANALYSIS ON THE VARIATION OF THE SHANGHAI COMPOSITE INDEX

1.1 Data Collection and Description

We have collected the composite index information in Shanghai stock market from January 1, 2007 to June 30, 2013, 1574 trading days in total [3]. We calculate the monthly average Shanghai Composite Index to represent the overall level of the month and finally obtain 78 data, which we call time series according to the time order.

1.2 Influence Factors Analysis of Shanghai Composite Index

Using the multiplicative model: $Y=T\times S\times C\times I$ （1）

We consider the long-term trend (T), seasonal fluctuation (S), cyclical fluctuation (C), irregular change (I) and so on. According to the data, we draw the Shanghai Composite Index line chart (Figure 1) as follows (Cryer & Chan, 2008):

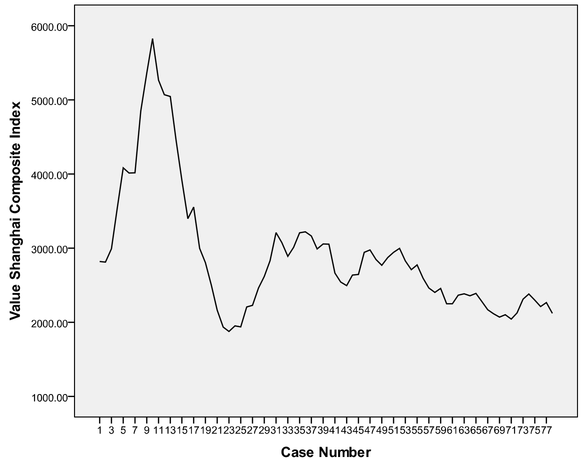

Figure 1
Shanghai Composite Index Line Chart

1.3 Factorization: The Long-Term Trend

The long-term trend is the time series in the long-term showing some continuous rising and falling or changes in the flat. For the time series we build a simple linear regression and a two order curve regression and get the trend line. For the model of simple linear regression, R^2=0.391; for the model of two order curve regression, R^2=0.413. So we can draw a conclusion that the fitting degree of two order curve is better than simple linear regression, but both of them are not fitting well enough. Therefore, in the latter part of this paper we introduce some lagged variables, and establish an ARIMA model. The fitting results of two order curve show as follows:

Table 1
Model Summary

R	R square	Adjusted R square	Std. error of the estimate
.643	.413	.398	672.797

The fitted curve equation: $\hat{y}_t = 4148.130 - 46.492t + 0.286t^2$ (2)

Table 2
Coefficients

	Unstandardized coefficients		Standardized coefficients	t	Sig.
	B	Std. error	Beta		
Case Sequence	-46.492	13.701	-1.215	-3.393	.001
Case Sequence ** 2	.286	.168	.609	1.700	.093
(Constant)	4148.130	234.526		17.687	.000

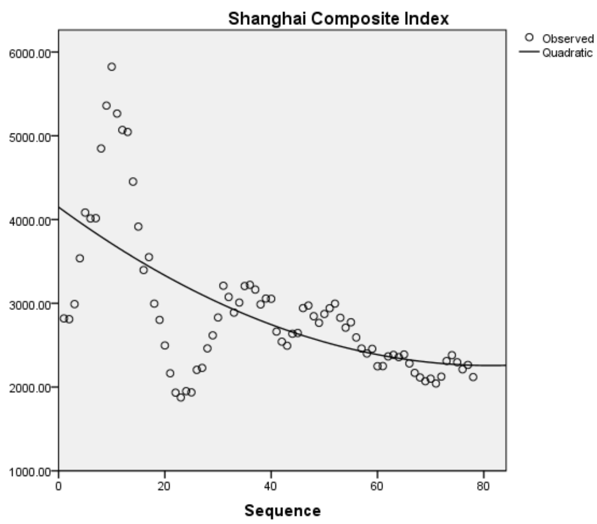

Figure 2
The Fitted Curve

The above results suggest that in recent years, the Shanghai Composite Index showed a downward trend which can be clearly seen from the figure. At the same time in the near future tends to be stable, and the future is likely to rise again.

1.4 Factorization: Seasonal Fluctuation

Seasonal fluctuation means the cyclical fluctuation of time series that repeat within one year. If the series changed in the season, we can use moving average method, exponential smoothing method and some other methods to determinate and eliminate it. Usually we'd like to use moving average method to deal with it.

After we establish the series year line chart, we can clearly find out from the chart that all the year series are basically consistent with season fluctuation except 2007 and 2008 these two special years. We use the moving average method to determine the season fluctuation. Because this season includes 12 data every year, so we perform at intervals of 12 moving average. Then the actual value of Shanghai Composite Index is divided by 12 periods moving average prediction value and we get a new time series constituted by the relative number. This time series compared to the original ones have missed 12 period data. Finally, a new series calculated and adjusted according to the average method, we get a monthly seasonal ratio. The following table (Table 3) shows the monthly seasonal ratio and the adjusted seasonal ratio, which can be clearly seen growth and decline periods of Shanghai Composite Index.

Table 3
Monthly Seasonal Ratio and the Adjusted Seasonal Ratio

Month	Seasonal ratio	Adjusted seasonal ratio
1	0.993229758	0.995781245
2	1.00503096	1.007612762
3	1.005232798	1.007815119
4	1.013157495	1.015760173
5	1.012757646	1.015359297
6	0.983591011	0.986117737
7	0.991446318	0.993993223
8	0.994459957	0.997014604
9	0.974583965	0.977087553
10	1.004405422	1.006985618
11	0.999801802	1.002370172
12	0.991555312	0.994102497

1.5 Factorization: Cyclical Fluctuation

Cyclical fluctuation is refers to the time series presenting around the long-term trend of a kind of wavy or shock type changes. It has no fixed pattern and cycle number is often more than one year and different lengths.

According to the multiplicative model, there is:

$$C \times I = \frac{T \times S \times C \times I}{T \times S} \tag{3}$$

Namely, with the original time series elimination long-term trend and seasonal changes, we can get the cycle fluctuation and the irregular changes. Then the irregular changes can be eliminated by moving average method and obtain the cycle fluctuation finally.

1.6 Factorization: Irregular Change

The irregular change refers to the phenomenon due to a variety of difficult to predict or accidental factors such as natural disasters, war, etc., with the development and change in a short period of time showing a sudden change or random variable. It can be used to understand the stochastic factors and the accidental factors impacting on the development and change the size of the data.

According to the multiplicative model, the irregular change is:

$$I = \frac{S \times T \times C \times I}{S \times T \times C} \tag{4}$$

Namely, the original time series has eliminated the above three items: the long-term trend, seasonal fluctuation and cyclical fluctuation.

2. ESTABLISH THE SHANGHAI COMPOSITE INDEX MODEL

2.1 The Stability of Shanghai Composite Index Series

Stability is one of the most important assumptions which are used to make a statistical inference for stochastic process. Therefore, if the sequence is non-stationary, we need to make a smooth processing for it.

The analysis of stability for time series by SPSS statistical analysis software, we get the following figures which are autocorrelation and partial autocorrelation plot. As can be seen from the figure that autocorrelation plot is trailing and sequence can be regarded as a stationary sequence. Therefore, we use the original series to build a time series model directly.

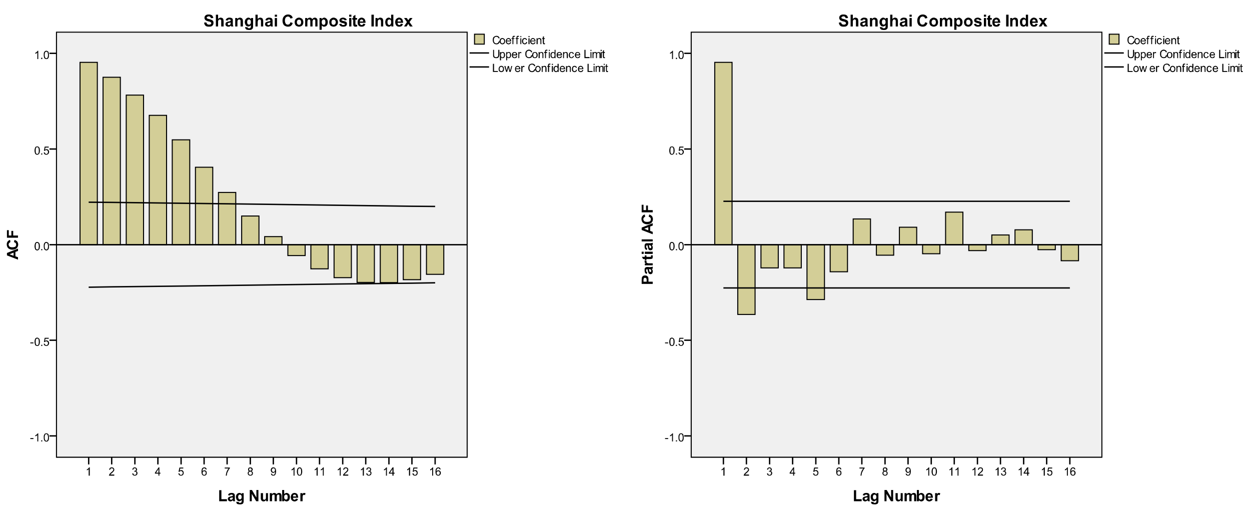

Figure 3
Autocorrelation and Partial Autocorrelation Plot

2.2 Identify and Test Shanghai Composite Index Model by Introducing the Lagged Variable

Due to the Shanghai Composite Index has seasonal fluctuations, so it is necessary to establish a seasonal ARIMA model. Because this is a stationary series, so we consider establishing a seasonal multiplicative model of the seasonal cycle of 12, which it is ARIMA (p，0，q)×(P，0，Q). Eventually the optimal model is ARIMA(1，0，1)×(0，0，1)$_{12}$, and the estimation and test of the parameters are as follows.

As you can see, with the R^2=0.932, the p-value of the model is 0.015, which has passed the test of goodness of fit and the model fitting effect is good. At the same time the parameters of the model also passed the significance test. Substituting the estimated values of the parameters into the model, we get as follows.

$$\phi(x)\Phi(x)\nabla^d\nabla_s^D Y_t = \theta(x)\Theta(x)e_t + \theta_0 \qquad (5)$$

Table 4
Model Statistics

Model	Number of Predictors	Model Fit statistics		Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	R-squared	Statistics	DF	Sig.	
The Shanghai Composite Index-Model_1	0	.932	.932	29.244	15	.015	0

Table 5
ARIMA Model Parameters

			Estimate	SE	t	Sig.
The Shanghai Composite Index-Model_1	Constant		2824.64	360.455	7.836	.000
	AR	Lag 1	.940	.042	22.372	.000
	MA	Lag 1	-.325	.115	-2.822	.006
	MA, Seasonal	Lag 1	.292	.122	2.397	.019

The final expression is obtained:

$$\hat{Y}_t = 0.94Y_{t-1} + Y_{t-12} - 0.94Y_{t-13} + e_t - 0.325e_{t-1} + 0.292e_{t-12} - 0.0949e_{t-13} + 2824.644 \qquad (6)$$

Line chart below shows the observed values and the fitting line (Figure 5).

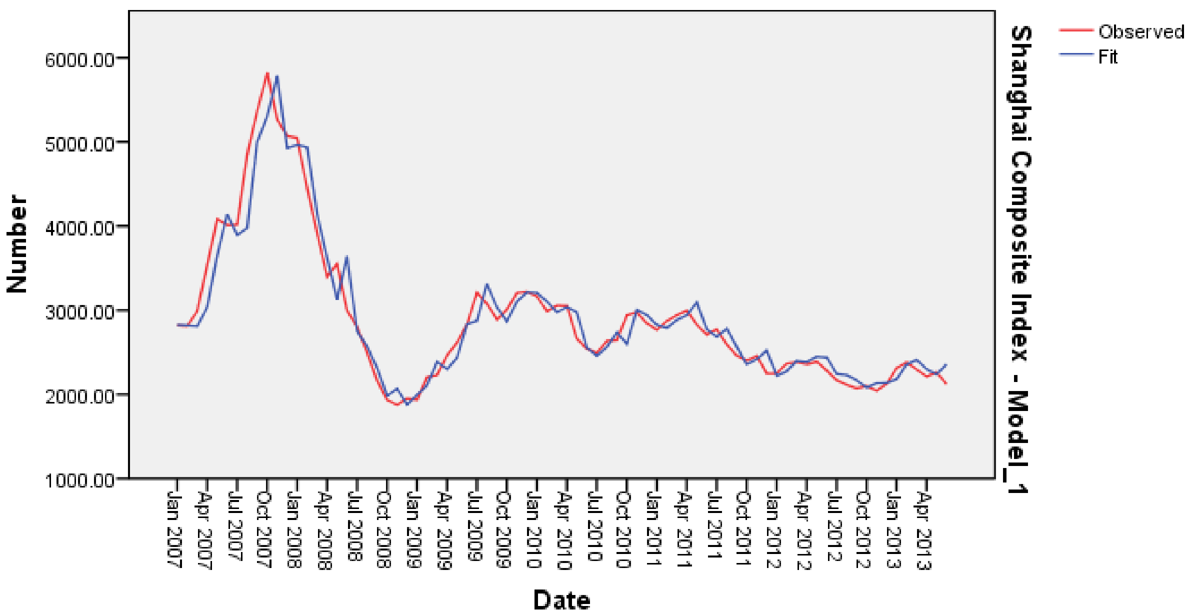

Figure 4
Observed Values and the Fitting Line

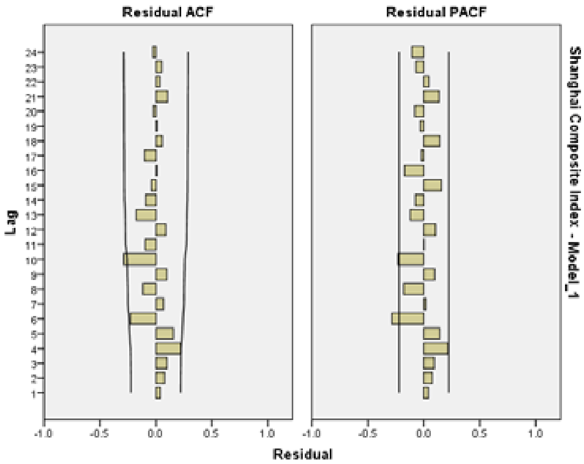

Figure 5
Residual Analysis Plot

As can be seen from the figure of residual analysis, the model also has passed the residual test. In general, the model fitting effect is improved obviously.

CONCLUSIONS AND SUGGESTIONS

The seasonal ARIMA model in this paper fits very well the changes of the Shanghai Composite Index in recent years. We can use this model make a short-term prediction reasonably.

Since 2007 to 2008 America financial crisis, the financial market of our country has been a great impact as well. After the Shanghai stock market crashed experience, although now tends to be stable, it cannot compare with a few years ago. Our government needs to take certain measures to stimulate China's financial market, so that China's financial market can be active again in the near future.

REFERENCES

Cryer, J. D., & Beasley, K. S. (2008). *Time Series Analysis With Applications in R*. Springer Texts in Statistics.

Dai, W. S., Shao,Y. J. E., & Lu, C. J. (2013). Incorporating feature selection method into support vector regression for stock index forecasting. *Neural Computing & Applications, 23*(6), 1551-1561.

Mezali, H., & Beasley, J. E. (2013). Quantile regression for index tracking and enhanced indexation. *Journal of the Operational Research Society, 64*(11), 1676-1692.