ISSN 2616-7433 Vol. 4, Issue 5: 5-8, DOI: 10.25236/FSST.2022.040502

Data Analysis on Hushen 300 Index

Wengeng Cui

University of South Florida, Florida, FL33620, USA

Abstract: *Research purposes: to study the correlation between Hushen 300 index and its lagged index. The specific methods is as follows: establishing the auto-regressive model (AR (P)). And the conclusion is that the model results show that the index of four days and six days ago are of great reference for the prediction of Hushen 300 index on that day. There was a positive correlation between the four days ago and that day in terms of Hushen 300 index, and on the contrary, there is a negative correlation between six days ago and the same day of Hushen 300 index.*

Keywords: *Stock market prediction, Stock market volatility, Stock price forecasting*

1. Background Introduction

Stock market is a barometer of financial market, and it can directly reflect the volatility and uncertainty of market. Moreover, as a direct reflection of macroeconomic fluctuations, stock market volatility has a very close relation with the uncertainty of economic policies. While the uncertainty of economic policy will also change the expectation of economic entities on macroeconomic fundamentals, thus exacerbating the volatility of stock market[7]. Therefore, revealing the operation law of the stock market, exploring the influencing factors of the stock market, as well as stabilizing the stock market are very significant for the stability of the entire financial market[3]. And both regulators and stock market participants hope to maintain the stability of the stock market and maximize returns through quantitatively measuring and analyzing the return and volatility of the stock market, building a reasonable model to predict market and allocating assets efficiently[2][8]. Conventional models widely utilized to forecast stock markets include auto-regressive (AR), Auto-Regressive Moving Average (ARMA), and other models[1] . According to Narayan[5] and Rusu[6], yt depends on a weighted sum of the past values and the term of random perturbation,and auto-regressive model could improve the forecasting accuracy[4].

Hushen 300 index is the first unified index officially recognized in China's securities market, which was joint official release by Shanghai Stock Exchange and Shenzhen Stock Exchange on April 8, 2005 and used to show the overall trend of A-share market. It can represent the whole market, because the sample of selected stocks, including 121 Shenzhen market shares and 179 Shanghai sample stocks in the first batch of sample, covers up to 60% of the market value coming from the two stock exchanges. Since Hushen 300 index has a certain representativeness, so it can better price the assets and it can reflect the size and fluctuation of China's securities market to a large extent as a benchmark index. Not only in market, it is also a representative in industry, because the selected sample of stocks covers almost all industries. Such a representative index can be used as the evaluation standard of investment performance to provide basic conditions for index investment and index derivative product innovation. Therefore, it's worth analyzing and forecasting to such an important and representative stock index with at all costs.

2. Data Source and Index Design

Table 1: Data Variable Description Table

Variable Type	Variable Name	Value Range
Dependent Variable	y	-0.0924~0.0934
Explanatory Variables	y_1	-0.0924~0.0934
	y_2	-0.0924~0.0934
	y_3	-0.0924~0.0934
	y_4	-0.0924~0.0934
	y_5	-0.0924~0.0934
	y_6	-0.0924~0.0934

ISSN 2616-7433 Vol. 4, Issue 5: 5-8, DOI: 10.25236/FSST.2022.040502

The data of this case comes from Wind database, with a unit of seven days, a total of 4002 items about the Hushen 300 index and its lagged index. The time period is from January 12, 2005 to June 28, 2021. Because it is an autoregressive study, the dependent variable is the Hushen 300 index (y) of the day. Therefore, the explanatory variable is the historical data of the previous six days（y_1, y_2, y_3 , y_4 , y_5, y_6). Because the lag term is studied, the value ranges of the variables are quite different, because they represent the Hushen 300 index, as shown in Table 1.

3. Descriptive Analysis

The next step is to conduct descriptive analysis, check the data quality, and preliminarily judge the correlation between the Hushen 300 index and its lagged index on that day, so as to pave the way for later modeling research.

First of all, a simple histogram is made for each variable to observe the shape of data distribution (similar because it is a lag term), mainly to test whether there is obvious data anomaly. The results are shown in Figure 1: The Hushen 300 index is approximately normal distribution, and there is no obvious data anomaly.

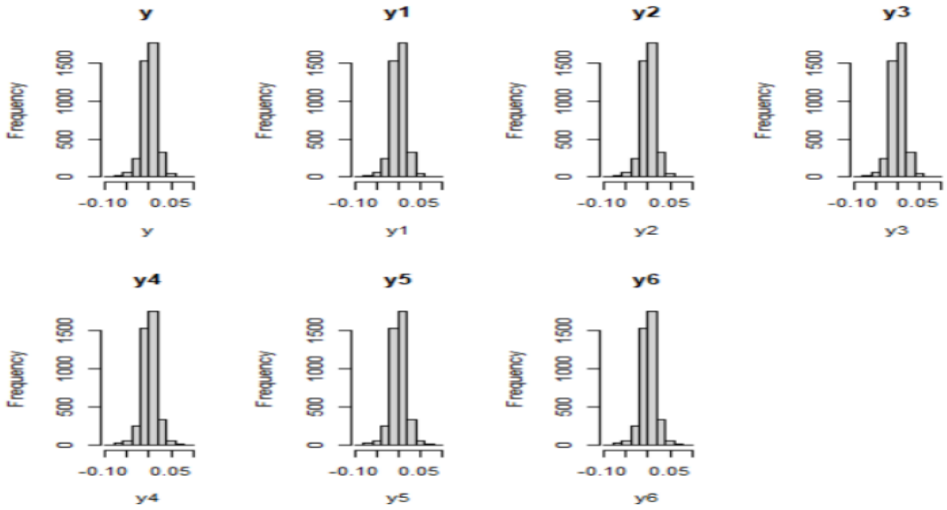

Figure 1: Histogram of Each Variable

Specific descriptive statistical indicators were calculated for each variable, and the results are shown in table 2. Because it is the research of its own lag item that the descriptive statistical indicators of each lag item are extremely the same, from which we can see that: The mean value (0.001) of each lagged index of Hushen 300 index is the same as that of this median (0.001), which indicates that the average Hushen 300 index is at 0.001. The results show that the data distribution of Hushen 300 index is basically symmetrical and uniform. At the end of the analysis by maximum and minimum value, there were no abnormal values (- 0924.0) and (- 0934.0).

Table 2: Description and Analysis of Each Variable

Variable Name	Sample Size	Mean Value	Standard Deviation	Maximum	Median	Minimum Value
y	4002	0.001	0.0169	0.0934	0.001	-0.0924
y_1	4002	0.001	0.0169	0.0934	0.001	-0.0924
y_2	4002	0.001	0.0169	0.0934	0.001	-0.0924
y_3	4002	0.001	0.0169	0.0934	0.001	-0.0924
y_4	4002	0.001	0.0169	0.0934	0.001	-0.0924
y_5	4002	0.001	0.0169	0.0934	0.001	-0.0924
y_6	4002	0.001	0.0169	0.0934	0.001	-0.0924

4. Modeling

On the basis of descriptive analysis, the relationship between the lagged index and the index of the

ISSN 2616-7433 Vol. 4, Issue 5: 5-8, DOI: 10.25236/FSST.2022.040502

day will be further analyzed.

Firstly, the autoregressive model is established $y = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3 y_3 + \beta_4 y_4 + \beta_5 y_5 + \beta_6 y_6 + E$, where,$\beta_0$, β_1 , ..., β_5, β_6 are the unknown regression coefficient, error term e's mean value is 0 and the variance is 0 ε^2.The related parameter estimation and test results are shown in table 1-3. The F-test of the whole model is highly significant (p-value < 0.001), which indicates that at least one explanatory variable is significantly correlated with Hushen 300 index. Decision coefficient after model adjustment$R_{adj}^2 = 0{,}006$. Secondly, the t-test results of each explanatory variable were examined and analyzed. The t-test results in table 3 show that except for$y_2(0.142)$ and $y_4(0.964)$ accounted for the p-value height was not significant, the other variables were significant y_6. The coefficient estimation of 1% is significantly not $0, y_4$. The estimation of the coefficient is highly significant at the level of 5%. At the same time, the VID value of the variable is very small, and there is no multicollinearity problem.Therefore, it is reasonable to exclude multicollinearity y_2 , y_5. The probability of not significant proportion.

Table 3: Regression Model Results

Variable Name	Regression Coefficient	Standard Error	P-Value	Variance Expansion Factor
Intercept Term	0.001	0.000	0.040	
y_1	0.026	0.016	0.094	1.004
y_2	-0.023	0.016	0.142	1.005
y_3	0.030	0.016	0.058	1.004
y_4	0.044	0.016	0.005	1.004
y_5	-0.001	0.016	0.964	1.005
y_6	-0.060	0.016	0.000	1.004
Global Model Checking	P-Value < 0.001		Adjusted R^2	0.006

Then, model selection is made according to AIC and BIC criteria, and the results are shown in table 4. This report found that AIC regression retained the y_5 except for other variables, BIC regression only retained the most significant $y_4$and y_6. This result is consistent with the previous t-test, $y_4$and y_6 are highly significant, and y_5 other variables also showed low significance. The following conclusions can be drawn from table 4 when other variables are kept unchanged.

In the case that other variables remain unchanged, the fourth-order hysteresis term y_4 the higher the index (that is, four days ago), the higher the Hushen 300 index on that day. This conclusion also seems to be in line with expectations, because stock volatility, which is often caused by the implementation of a policy or plan, will not appear until a few days later, so the day is more likely to be associated with the index of the previous days.

In the case that the other variables are controlled unchanged, the sixth order hysteresis termy_6The higher the index is, the lower the Hushen 300 Index index will be on that day, and it is extremely significant.In other words, the Hushen 300 Index index six days ago seems to have a high reference value, but it is not absolute. It may be that the current sample size is not large enough for the model to discover.

Table 4: AIC and BIC Regression Model Results

Variable Name	AIC Regression Coefficient	P-Value	BIC Regression Coefficient	P-Value
Intercept Term	0.001	0.040	0.001	0.034
y_1	0.026	0.094		
y_2	-0.023	0.141		
y_3	0.030	0.057		
y_4	0.044	0.005	0.047	0.003
y_5				
y_6	-0.060	0.000	-0.060	0.000
Global Model Checking	P-Value < 0.001		P-Value < 0.001	
Adjusted R^2	0.007		0.005	

Next, this report made a diagnosis for the model. For example, there are strong influence points by calculating Cook distance, and judging whether the model is abnormal by observing the residual graph.

ISSN 2616-7433 Vol. 4, Issue 5: 5-8, DOI: 10.25236/FSST.2022.040502

No abnormal problems were found in this report.

5. Conclusion and Prospect

This report is based on the Hushen 300 index data of Wind database. Through the establishment of autoregressive model (AR (P)), this paper studies the correlation between Hushen 300 index and its lagged index. The main conclusions of this report are summarized as follows: (1) the fourth order lagged index is positively correlated with the Hushen 300 index, while the sixth order lagged index is negatively correlated with the Hushen 300 index on that day; (2) The fifth order lagged index has little reference for the prediction of the Hushen 300 index on the same day; (3) this report lacks sufficient evidence to describe the impact of other lagged indexes on the Hushen 300 index of the same day. No other factors, such as trading volume, investor volume, etc. are considered in this report in terms of the impact on the Hushen 300 index, and linear regression model will be added to for in-depth study in future research work.

References

[1] Atsalakis, G. S., & Valavanis, K. P. (2010). Surveying stock market forecasting techniques-Part I: Conventional methods. Journal of Computational Optimization in Economics and Finance, 2(1), 45-92.
[2] Hamilton, J. D., & Lin, G. (1996). Stock market volatility and the business cycle. Journal of applied econometrics, 11(5), 573-593.
[3] Hu, Z. (1995). Stock market volatility and corporate investment. International Monetary Fund.
[4] Liu, H. Y., & Bai, Y. P. (2011). Analysis of AR model and neural network for forecasting stock price. Math. Pract. Theory, 41(4), 14-19.
[5] Narayan, P. K. (2006). The behaviour of US stock prices: Evidence from a threshold autoregressive model. Mathematics and computers in simulation, 71(2), 103-108.
[6] Rusu, V., & Rusu, C. (2003). Forecasting methods and stock market analysis. Creative Math, 12, 103-110.
[7] Schwert, G. W. (1989). Why does stock market volatility change over time? The journal of finance, 44(5), 1115-1153.
[8] Schwert, G. W. (1990). Stock market volatility. Financial analysts journal, 46(3), 23-34.