

Prediction and Empirical Study of Stock Yield Volatility Based on Event Study

Shen Mingcai^{1a} Liu Xin^{1b} Huang Xi^{1c} Cao Zhaohuan^{1d} Su Ganya^{*1e}

¹Statistics institute, Chengdu University of Information Technology Chengdu, Sichuan, China

Abstract: Stock market event is an important source of information for investment decision, and it is of practical significance to quantify the event and predict the fluctuation range of future return under such event. Most researchers study stock market events horizontally, that is, to study the impact of a current event on the stock price of a certain sector or industry, while the paper attempts to study vertically the impact of a certain event of a single listed company on the return. Based on the internal relations between public announcement and stock yield of listed companies, the paper deduced the daily yield prediction model of event window by VAR(p) to exclude subjective "estimation" in the past and verifies the feasibility of the model.

1 Estimation Period and Window Period of Definite Event

Event Study is a quantitative analysis that collects statistical data before and after the occurrence of an event and conduct in-depth technical analysis of the event's impact [1]. It studies whether the stock price will fluctuate when a certain event takes place in the market, and whether "abnormal return" will occur [2].

The paper studied the timeliness of stock market event's impact on individual stock yield by "impulse response analysis" to analyze the cycle in which messages from Sina microblog act on the stock market [3]. The trading volume is used to measure the impact of the event, and the yield is used to measure the change of the yield. The VAR(p) model (vector autoregression model) of event impact and stock yield is given by:

where F_t refers to the stock's daily turnover sequence, R_t refers to the stock's daily return sequence, C is a constant, $C_{i,j}$ is the corresponding coefficient.

A. Data selection and processing

In terms of earnings, the daily return is adopted to test market efficiency because compared with monthly and annual data, daily data can reflect the impact of event on earnings in a timely manner and reduce problems of joint inspection, thus ensuring the accuracy of event study in this type of researches [4].

The daily closing price (P) and daily trading volume (F_0) of a certain stock in a certain period of time were collected, and events in the number of n during the same period were included. Calculate the daily return $\left(\frac{P_2 - P_1}{P_1}\right)$ based on the closing price, which is denoted as R_t , and

denote the average number of total trading volume \bar{F} , and $F_t = F_0 - \bar{F}$.

A certain kind of event is selected as the target event. In order to avoid the impact of non-target event on the experiment, the sample is taken from the occurrence of this target event to the day before the occurrence of the next target event. Data collected after the first occurrence of a non-target event in that period are excluded, including data of the target event if there is non-target event in the same day. The renewed period of sample is set as the observation period, with different duration of each period. The period with effective impact on the stock price after the occurrence of the event is defined as the window period, and the cycle of observation period minus the window period estimated by the impulse response analysis is defined as the estimation period. The estimation period is an observation period not affected by the event, during which the return is regarded as normal. The arithmetic average of the return in the estimated period is taken to get the normal mean return in the window period without considering the market trend. The real-time return minus the normal mean return in the estimation period is the abnormal return brought by the event in the observation period. The diagram is shown in Figure 1.

B. Establishment of VAR model and causality test

The EViews software is used to make line chart for preliminary observation and test of the stability of two sequences. Once the sequences pass the test, the confidence level and lag phase will be determined to establish the VAR model, which will be then verified for stability and Granger causality between the two variants.

^a925394476@qq.com ^blx001201@163.com ^c1129858063@qq.com ^d708869423@qq.com ^e1176322265@qq.com

C. *The window period of estimated event*

he established VAR model was used for impulse response analysis. By analyzing tables and graphs of the impulse response analysis, the estimated window period was determined, that is, the number of days during which the certain event acts on the stock yield.

2 Model Establishment

A. *Normal mean return in the estimation period*

As China's capital market is in a weak position and it is highly subjective in the selection of the estimation

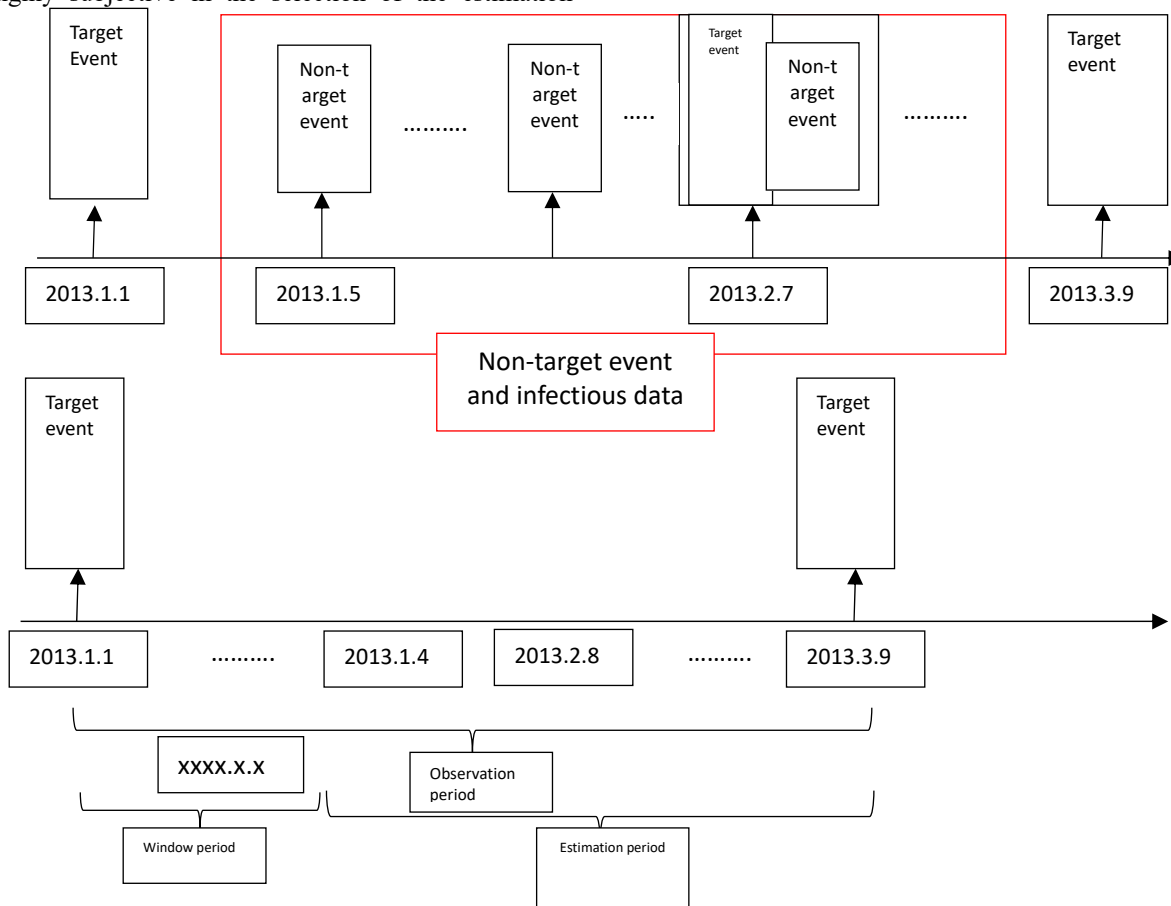


Figure 1 Diagram of Data processing

a trend to predict the normal return in the future. If there is no trend in the normal return, the historical normal return could be used to represent the normal return in the future based on the theory of history repeating.

2) *Statistical description of normal return*

The statistical distribution diagram presents a general distribution of the subject, thus determining the possible distribution model, based on which the specific distribution could be determined. The description and diagram of statistical distribution of the return in the estimation period could be used to determine possible distribution model and statistics of normal return, offering references for the distribution model of mean return.

period, the period beyond the window period in the observation period is used as the estimation period rather than the subjective selection of an estimation period.

1) *Trend test*

If there is a time trend, the historical normal return cannot serve as a direct indicator of the future return. If we want to express the future return by normal return under a trend, we need to remove the trend and establish the model of normal return under

3) *Random sampling of normal return and standard normalization of sample mean*

As the return is volatile between positive and negative, normally it is not a constant. Therefore, the overall mean value is not an indicator of general return. The paper, based on the independent identical distribution of normal daily return, randomly sampled returns in the estimation period under the central-limit theorem and established a distribution model of normal mean returns.

Assume the normal mean return is \bar{R}_n , then the distribution model of \bar{R}_n is given by:

$$\bar{R}_n \sim N(\mu, \sigma) \tag{1}$$

The standardized approximate distribution:

$$Z = \frac{\bar{R}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \quad (2)$$

where μ is the population mean of normal return, and σ is the population variance of normal return.

4) *The range of normal mean return volatility*

The value-at-risk (VAR) is used to predict the maximum return and maximum loss of a certain day in the future with a certain confidence level. If the distribution parameters of a sample's mean value is known, the estimated VAR parameters could be used to calculate the maximum and minimum normal mean return of a certain day in the future with a confidence level of $1 - \alpha$ in the equation of $\mu + \sigma * Z_{\frac{\alpha}{2}}$ and $-\sigma * Z_{\frac{\alpha}{2}}$ respectively, that is, there is an $1 - \alpha$ confidence that the normal return will average within a range on a certain day in the future.

$$[\mu - \sigma * Z_{\frac{\alpha}{2}}, \mu + \sigma * Z_{\frac{\alpha}{2}}] \quad (3)$$

where $Z_{\frac{\alpha}{2}}$ is the quantile at a given confidence level.

B. *Abnormal mean return in the window period*

The impulse response analysis is used to evaluate the time-validity of certain event's impact on stock price, thus obtaining the duration of such event's window period. The abnormal return in the windows period is obtained by real-time yield minus mean normal return in the estimation period. The impulse response analysis measured only the windows period of the event without verifying the significance of abnormal return in the same period. Further variation is required to determine the authenticity of an event's impact on return in the windows period.

1) *Significance test of abnormal return*

Significance test determines the authenticity of impulse response analysis. If the abnormal return is significantly zero in the window period, it means that the event has no effect on the return, and there is no time-validity. The hypothetical test is used to test the significance of a given distribution model. However, as the distribution of abnormal return is not known and the return could be positive or negative, the hypothetical test is not applicable to the significance test.

As a number relative to the normal return, the abnormal return could be significant if the ratio of the mean absolute value of the abnormal return to that of the normal return is greater than 1, which means that the event has an impact on the return. Assume the ratio is ρ , then the expression for ρ is:

$$\rho = \frac{\frac{1}{N} \sum_i^N \sqrt{(R_i^*)^2}}{\frac{1}{N} \sum_i^N \sqrt{(R_{ni})^2}} \quad (4)$$

where N means the number of days of return, R_{ni} is the normal daily return, and R_i^* is the abnormal daily return.

2) *Statistical description of abnormal return*

The study of abnormal return used the same approach processing normal return, that is, all window periods were integrated for statistical description and plotting of the distribution diagram to observe the general

distribution of abnormal return and calculate the parameters of abnormal return. In the statistical description, the possible distribution model and statistical parameters of abnormal return were determined, which provided foundation for the establishment of sample mean distribution model of abnormal return. In the statistical description, μ^* means the population mean of abnormal return, and σ^{*2} is the population variance of abnormal return.

3) *Random sampling of abnormal return and standardized normalization of sample mean*

The modeling of abnormal return sample was the same as that of normal return, and the center-limit theorem was used for approximate estimation. Random sampling of the abnormal return had a size of $n \geq 30$, which was large enough. The sample mean was statistically described coupled with a distribution diagram. The prediction was done at a certain confidence level, with the abnormal mean return fluctuating within a certain range on the premise that the distribution of the abnormal mean return was known. The known statistical distribution is used to predict the fluctuation range of the abnormal return's sample mean before standardizing it. The abnormal mean return was standardized in the same way as the normal return.

Assume the abnormal mean return is \bar{R}^* , then the distribution model of \bar{R}^* could be expressed as:

$$\bar{R}^* \sim N(\mu, \sigma) \quad (5)$$

Standardized approximate distribution:

$$Z = \frac{\bar{R}^* - \mu^*}{\frac{\sigma^*}{\sqrt{n}}} \sim N(0,1) \quad (6)$$

where μ^* is the population mean of abnormal return, and σ^{*2} is the population variance of abnormal return.

4) *The range of abnormal mean return volatility*

Assume the distribution of abnormal mean returns is known, and apply VAR to calculate the maximum and minimum abnormal mean return on a certain day in the future with a confidence level of $1 - \alpha$: $\mu^* + \sigma^* * Z_{\frac{\alpha}{2}}$ and $\mu^* - \sigma^* * Z_{\frac{\alpha}{2}}$ respectively, that is, there is a confidence that the abnormal mean return on certain day in the future will be in the range of:

$$[\mu^* - \sigma^* * Z_{\frac{\alpha}{2}}, \mu^* + \sigma^* * Z_{\frac{\alpha}{2}}] \quad (7)$$

C. *Determine the quantile of return*

As the return could be positive or negative, the sample mean could only represent the population mean rather than the real-time return. To estimate the range of fluctuation for real-time return, it is necessary to know its distribution. The indefinite distribution of real-time return coupled with the characteristics of leptokurtosis and fat tail could be used to simulate the distribution of profit and loss by sorting the position gains and losses of financial portfolio by size[5]. Given percentile, the quantile under the confidence degree is solved, and then the non-parametric estimation of VaR is used to estimate the fluctuation range of return. Rank the normal return in the estimation period in ascending order to find the $\frac{\alpha-1}{2}$

quantile. To avoid errors caused by data insufficiency, select the two numbers before and after the median with the percentile to calculate the mean of the selected data. Divide the difference between the mean and the population mean by the population standard deviation to obtain the quantile $X_{\frac{\alpha_1}{2}}$ at the confidence level of $\frac{\alpha_1}{2}$. Find the quantile at the $1 - \frac{\alpha_1}{2}$ confidence level in the same way. For the quantile of abnormal return at the confidence level of the $\frac{\alpha_1}{2}$ and $1 - \frac{\alpha_1}{2}$, the method still applies.

Assume the percentile of historical normal return $\frac{\alpha_1}{2}$ corresponds to R_{n0} , and the two return rates before and after it are R_{n-2} , R_{n-1} , R_{n1} and R_{n2} respectively, the mean is:

$$\frac{1}{5} \sum_{i=-2}^2 R_{ni} \quad (8)$$

The population mean is μ and variance σ^2 , with a quantile of $X_{\frac{\alpha}{2}}$ at the corresponding confidence level.

Then, the quantile at the confidence level of $\frac{\alpha}{2}$ is:

$$X_{\frac{\alpha_1}{2}} = \frac{\frac{1}{5} \sum_{i=-2}^2 R_{ni} - \mu}{\sigma} \quad (9)$$

D. Estimating the range of returns

Technical analysts believe that investors' trading behavior will follow a certain pattern, so that the past pattern of price changes may take place repeatedly in the future, that is, "history repeating" [6].

1) Interval estimation of normal return

The sample mean obtained by random sampling could somewhat represent the population mean. For those within the fluctuation range of normal mean return between $[\mu - \sigma * Z_{\frac{\alpha}{2}}, \mu + \sigma * Z_{\frac{\alpha}{2}}]$, the range also represents the distribution of population mean. In the distribution of actual normal return at the $1 - \alpha_1$ confidence level, the upper and lower limit of the normal mean return were taken for calculation of population mean, and the estimated intervals of normal return included:

Estimated interval of the lower limit:

$$[\mu - \sigma * Z_{\frac{\alpha}{2}} + \sigma * X_{\frac{\alpha_1}{2}}, \mu - \sigma * Z_{\frac{\alpha}{2}} + \sigma * X_{1-\frac{\alpha_1}{2}}] \quad (10)$$

Estimated interval of the upper limit:

$$[\mu + \sigma * Z_{\frac{\alpha}{2}} + \sigma * X_{\frac{\alpha_1}{2}}, \mu + \sigma * Z_{\frac{\alpha}{2}} + \sigma * X_{1-\frac{\alpha_1}{2}}] \quad (11)$$

To ensure accuracy of the interval estimation, the bottom of the lower limit estimated interval and the top of the upper limit estimated interval were selected and integrated into a new estimation interval of normal return:

$$[\mu - \sigma * Z_{\frac{\alpha}{2}} + \sigma * X_{\frac{\alpha_1}{2}}, \mu + \sigma * Z_{\frac{\alpha}{2}} + \sigma * X_{1-\frac{\alpha_1}{2}}] \quad (12)$$

where $X_{\frac{\alpha_1}{2}}$ is the quantile of normal return at the confidence level of $\frac{\alpha_1}{2}$, and $X_{1-\frac{\alpha_1}{2}}$ is the quantile of normal return at the confidence level of $1 - \frac{\alpha_1}{2}$.

2) Interval estimation of abnormal return

The interval estimation of abnormal return was conducted in the same way as that of normal return. In the distribution of abnormal return, the upper and lower limit of abnormal mean return were taken for calculation of population mean, and the estimated intervals included:

Estimated interval of the lower limit:

$$[\mu^* - \sigma^* * Z_{\frac{\alpha}{2}} + \sigma^* * X_{\frac{\alpha_1}{2}}^*, \mu^* - \sigma^* * Z_{\frac{\alpha}{2}} + \sigma^* * X_{1-\frac{\alpha_1}{2}}^*] \quad (13)$$

Estimated interval of the upper limit:

$$[\mu^* + \sigma^* * Z_{\frac{\alpha}{2}} + \sigma^* * X_{\frac{\alpha_1}{2}}^*, \mu^* + \sigma^* * Z_{\frac{\alpha}{2}} + \sigma^* * X_{1-\frac{\alpha_1}{2}}^*] \quad (14)$$

where $X_{\frac{\alpha_1}{2}}^*$ is the quantile of abnormal return at the confidence level of $\frac{\alpha_1}{2}$, and $X_{1-\frac{\alpha_1}{2}}^*$ is the quantile of abnormal return at the confidence level of $1 - \frac{\alpha_1}{2}$.

To ensure accuracy of the interval estimation, the bottom of the lower limit estimated interval and the top of the upper limit estimated interval were selected and integrated into a new estimation interval of normal return:

$$[\mu^* - \sigma^* * Z_{\frac{\alpha}{2}} + \sigma^* * X_{\frac{\alpha_1}{2}}^*, \mu^* + \sigma^* * Z_{\frac{\alpha}{2}} + \sigma^* * X_{1-\frac{\alpha_1}{2}}^*] \quad (15)$$

3) Interval estimation of Return

Based on event study methodology, the actual return in the window period can be divided into normal and abnormal return. Through statistical analysis and modelling of data, the estimation interval of normal and abnormal return was added up to estimate the possible interval of daily return in the window period brought by such event. According to the theory of "history repeating", the interval estimation could be used to predict possible fluctuations of stock yield brought by this type of events. The estimated interval is given by:

$$[\mu - \sigma * Z_{\frac{\alpha}{2}} + \sigma * X_{\frac{\alpha_1}{2}} + \mu^* - \sigma^* * Z_{\frac{\alpha}{2}} + \sigma^* * X_{\frac{\alpha_1}{2}}^*, \mu + \sigma * Z_{\frac{\alpha}{2}} + \sigma * X_{1-\frac{\alpha_1}{2}} + \mu^* + \sigma^* * Z_{\frac{\alpha}{2}} + \sigma^* * X_{1-\frac{\alpha_1}{2}}^*] \quad (16)$$

3 Model Demonstration and Validation

A. Determination of the window period and estimation period

1) Data selection and processing

The paper took the event "senior manager turnover" of ICBC as an example to demonstrate and verify the model. 1521 sets of data of ICBC (601398), including daily closing price (P) and trading volume (F0) from November 13, 2013 to February 10, 2020 were downloaded from the software Eastmoney. A total of 352 sets of data sequenced as R and F were obtained after data processing.

2) Establishing VAR model

Draw the ling graph using EViews to observe the stability of the return and sequence R and F. The two

sequences were found unaffected by any trend or intercept. The two sequences went through stability test, which were stable at the confidence level of 0.05. The VAR model was built with eight lag phases. The model also went through stability test, with no roots left outside the unit circle, proving its stability.

The model was given by: $F = 0.174574652133 * F(-1) - 0.00618740961454 * F(-2) - 0.00907446721901 * F(-3) + 0.0553881167541 * F(-4) + 0.240026721114 * F(-5) + 0.0398096552736 * F(-6) + 0.134071848108 * F(-7) + 0.197505977814 * F(-8) - 1554280128.93 * R(-1) - 3474340148.21 * R(-2) + 1166460240.49 * R(-3) + 1057417667.09 * R(-4) - 1174218809.91 * R(-5) + 530010935.344 * R(-6) - 533763362.914 * R(-7) + 453587856.525 * R(-8) + 17290159.4856$

$R = 6.18168501725e-12 * F(-1) - 5.7518859456e-13 * F(-2)$

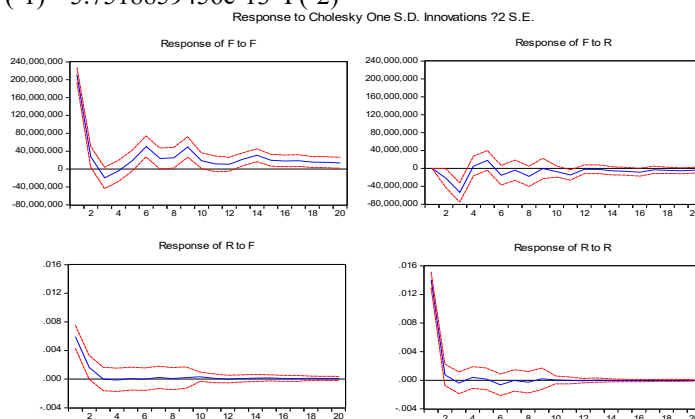


Figure 2 Chart of Pulse Response

B. Normal mean return in the estimation period

1) Trend test

There are many methods of trend test. This paper uses image method to describe the return in the estimation period, which generally reflects whether the normal return has a certain trend.

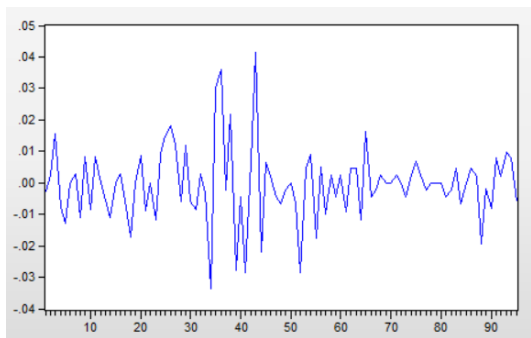


Figure 3 Tendency Chart of Normal Return

Figure 3 shows that the normal return in the estimation period has always been mean reversion around the X-axis without trend. The test results show that in the long run, there is no obvious fluctuation in the daily return of a single stock in the absence of

The established VAR model was used for impulse response analysis, and the result is shown in Figure 2. The estimated window period is 2, that is, the target event's impact on stock yield lasts for 2 days.

$- 1.17172010328e-12 * F(-3) + 1.0175508474e-13 * F(-4) + 8.32199996824e-13 * F(-5) - 2.73212986532e-13 * F(-6) + 3.84760771023e-13 * F(-7) - 1.05358372626e-13 * F(-8) + 0.0543985812308 * R(-1) - 0.0209788406972 * R(-2) + 0.0527876441846 * R(-3) + 0.00163161812697 * R(-4) - 0.0567887836227 * R(-5) + 0.0116548534863 * R(-6) - 0.0193400680934 * R(-7) + 0.0265944824699 * R(-8) - 0.000286146487417$

3) Causality testing and estimation of event window period

The Granger causality test of F sequence and R sequence was carried out, and the results are shown in the following graph. At the confidence level of 0.05, the null hypothesis "R does not Granger Cause F" was rejected, which means "R does Granger Cause F".

events, which always fluctuates within a certain range. The historical normal daily return can represent the future normal daily return.

2) Statistical description of normal return

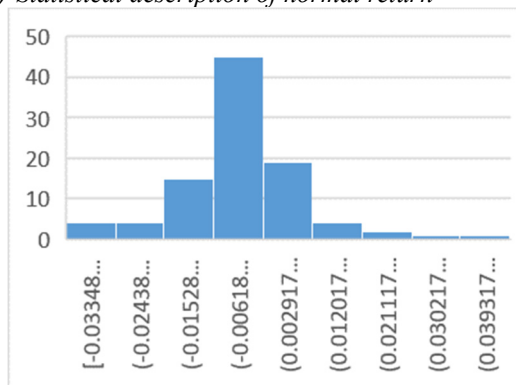


Figure 4 Statistical Distribution of Normal Return

All the daily returns in the estimation period were selected for statistical description and graph drawing. Figure 4 is a statistical distribution diagram, indicating that the normal return in the estimation period generally follows a normal distribution. Table 1 is the result of statistical description, showing the mean normal return in the estimation period $\mu = -0.00043$, and the standard deviation $\sigma = 0.0119109$, which means the normal return does not follow normal distribution.

Table 1 Statistical Description of Normal Return

Statistical Description of Normal Return	
Mean	-0.00043
Standard Error	0.001222
Median	0
Mode	0
Standard Deviation	0.0119109
Variance	0.000142
Kurtosis	2.534397
Skewness	0.327641
Minimum Value	-0.03348
Maximum Value	0.041475
Number of Observation	95

3) *Random sampling of normal return and standardized normalization of sample mean*

The paper randomly sampled the normal return at the size of 30 using the central-limit theorem with a total of 200 samples, which were used for statistical description and standardization.

The standardized normal mean return \bar{R}_n follows the distribution below:

$$\bar{R}_n \sim N(-0.00043, 0.0119109)$$

The standardized approximate distribution:

$$Z = \frac{\bar{R}_n + 0.00043}{\frac{0.0119109}{\sqrt{30}}} \sim N(0,1)$$

4) *Range of normal mean return volatility*

Given the normal return distribution at a certain confidence level, the range of normal return volatility could be obtained. The paper set the confidence level at 90%, and calculated the range of normal return volatility:

$$\left[-0.00043 - \frac{0.0119109}{\sqrt{30}} * 1.96, -0.00043 + \frac{0.0119109}{\sqrt{30}} * 1.96\right]$$

C. *Abnormal mean return in the window period*

1) *Significance test of abnormal return*

By use of Excel, calculate the average absolute value of normal returns $\frac{1}{N} \sum_i \sqrt{(R_{ni})^2} = 0.001818$, and the average absolute value of abnormal returns $\frac{1}{N} \sum_i \sqrt{(R_{ni}^*)^2} = 0.001818$, and we have:

$$\rho = \frac{\frac{1}{N} \sum_i \sqrt{(R_{ni}^*)^2}}{\frac{1}{N} \sum_i \sqrt{(R_{ni})^2}} = 5.838035 > 1$$

$\rho > 1$ means the abnormal return is significantly higher than the normal return, which to some extent indicates that the "senior manager turnover" ICBC had an impact on the stock return

2) *Statistical description of abnormal return*

The study of abnormal return is the same as that of normal return. The abnormal return of all window periods was integrated, and the statistical description and distribution were carried out using Excel to observe the statistical characteristics and general distribution of abnormal return. The results are as follows:

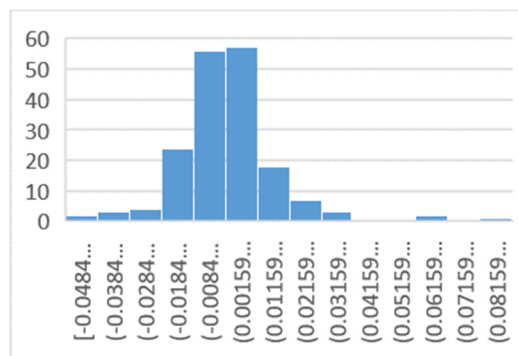


Figure 5 Statistical Distribution of Abnormal Return

Table 2 Statistical Description of Abnormal Return

Statistical Distribution of Abnormal Return	
Mean	0.001892
Standard Error	0.001241
Median	0.00043
Mode	0.00043
Standard Deviation	0.016508
Variance	0.000273
Kurtosis	7.179909
Skewness	1.354693
Minimum Value	-0.0484
Maximum Value	0.090826
Number of Observation	177

The statistical results in Table 2 show that the abnormal mean return is $\mu^* = 0.001892$, the standard deviation is $\sigma^* = 0.016508$, the kurtosis is 7.179909, and the skewness is 1.354693. This shows that the company's "senior manager turnover" will generate positive stock yield in most cases. In Figure. 5, the statistical distribution shows that the abnormal return roughly follows a normal distribution while the specific distribution does not follow a normal distribution based on the results of statistical description.

3) *Random sampling of abnormal return and standardized normalization of sample mean*

A random sampling of 177 abnormal returns was conducted with a sample size of 30. A total of 200 samples were selected, and the sample mean was statistically described and standardized.

The statistical results show that after standardization, the abnormal mean return is $2.51E-16$, the variance is 1, the kurtosis is -0.25213, and the skewness is 0.27248, indicating that the standardized sample mean has the characteristics of leptokurtosis and fat tail as well as left skewness.

In the paper, \bar{R}^* means abnormal mean return, whose standardized approximate distribution is given by:

$$\frac{\bar{R}^* - 0.001892}{\frac{0.016508}{\sqrt{30}}} \sim N(0,1)$$

4) *Range of abnormal mean return volatility*

Given the distribution of the abnormal mean return, the paper used the method to estimate the volatility range of normal mean return to process the abnormal mean

return at a given confidence level of 90%, which is given by:

$$\left[0.001892 - \frac{0.016508}{\sqrt{30}} * 1.96, 0.001892 + \frac{0.016508}{\sqrt{30}} * 1.96\right]$$

D. Determination of quantile and interval estimation

The paper set the confidence level at 90% and ranked the normal returns in the estimation periods in ascending order, followed by the selection of normal returns with the quantile of 5% and 95% as well as returns two days before and after the normal returns. The returns averaged $\frac{1}{5} \sum_{i=-2}^2 R_{ni} = -0.023004$ and $\frac{1}{5} \sum_{i=-2}^2 R_{ni} = 0.019202$ with a corresponding quantile of $X_{0.05} = -1.895262$, $X_{0.95} = 1.648283$ in the estimated interval of $[-0.019427, 0.022780]$.

The abnormal returns in the window periods were ranked in ascending order, followed by the selection of abnormal returns with the quantile of 5% and 95% as well as returns two days before and after the abnormal returns. The returns averaged $\frac{1}{5} \sum_{i=-2}^2 R_i^* = -0.021678$ and $\frac{1}{5} \sum_{i=-2}^2 R_i^* = 0.025959$ with a corresponding quantile of $X_{0.05}^* = -1.401721$, $X_{0.95}^* = 1.483963$ in the estimated interval of $[-0.027155, 0.032296]$.

The estimated intervals of normal and abnormal returns were added up to obtain the estimated interval of real-time return $[-0.046582, 0.055076]$, indicating a confidence level of 90%. The real-time return rate in the window period of future events is within the interval $[-0.046582, 0.055076]$.

E. Model validation

The daily returns of ICBC from October 17, 2007 to August 7, 2009 were selected as the data basis to calculate the real-time return in the window periods based on data processing introduced above. The real-time yield in the window period was statistically described, with the maximum value of 0.042222 and the minimum value of -0.044776, indicating that the real-time return in the window period was within the interval $[-0.044776, 0.042222]$. At the confidence level of 90%,

the return was predicted by the model to be within the interval $[-0.046582, 0.055076]$. By comparing the two intervals, the real-time return interval is within the model prediction, which to some extent indicated that the predicted results of the model were feasible.

4 Conclusion

Individual's preferences and the given confidence level could cause variation in the predicted intervals, but pose no differences to the prediction model. The prediction model is not only limited to firm event, but also applicable to other companies and events.

Acknowledgement

This paper is a phased achievement of "Exploration of Investment Strategy of Scientific Innovation Board Based on AI Quantitative Transaction", a national innovation training program for college students in 2019.

References

- [1] Dou W W, Ji Y. External financing and customer capital: a financial theory of markups[J]. Social Science Electronic Publishing, 2015, 4(2): 21-45.
- [2] Zhou Wenmin. The Research the performance of Company Back-door listing based on Fama French Three-Factor model [D]. Sichuan: Southwestern University of Finance and Economics, 2009:2
- [3] Yang Yufan. Research on the Relationship between Sina Weibo Message and Stock Market-Based on Text Data Mining Technology [D]. Beijing : Beijing Jiaotong University, 2018:3.
- [4] Gomber P, Schweickert U, Theissen E. Liquidity dynamics in an electronic open limit order book: an event study approach [J]. European Financial Management, 2015, 21(1):52-78.
- [5] Wang Chao. Based on the Shanghai composite index of three historical simulation method of empirical comparative study [D]. Beijing: University of International Business and Economics, 2017:13
- [6] Huang Tengfei. Research on the Investment Decision of Chinese Financial Market Based on Chaos Theory [D]. Jiangsu: Nanjing University of Aeronautics and Astronautics, 2013:100-101