

Influencing Factors and Forecast of Stock Price Manipulation Based on Panel Data Logit Model

Zhang Jianfeng

Department of Finance

Xi'an University of Technology

Xi'an, China

527802384@qq.com

Hu Wexiu

Department of Finance

Xi'an University of Technology

Xi'an, China

hwxsxj@sina.com

Abstract—Stock price manipulation forecasting is to carry out judgment before the stock price being manipulated, and to identify indicators with which the stocks are likely to be manipulated. Based on the 2015 A-share market stock trading manipulation case panel data issued by China Securities Regulatory Commission, and on the basis of two facts that financial advantage is needed for a manipulator to implement manipulation, and that other investors are permitted to copy trading, the panel data Logit model is built to forecast the stock price manipulation by identifying the main influencing factors of stock price manipulation with the stepwise regression method. The results show that when the average stock holding ratio and the price volatility of the stock in the previous quarter are higher, and the beta value is lower, then it is more likely for the stock price to be manipulated. The overall prediction accuracy of the out-of-sample and the in-sample is 68.18% and 70.63% respectively, so the effectiveness is relatively good.

Keywords—Stock price manipulation; Panel data; Logit model; Manipulation forecasting; Manipulation management

I. INTRODUCTION

As the scale of China's capital market continues to expand, the number of stock manipulation cases is increasing year by year. According to China Securities Regulatory Commission statistics, 71 cases were placed on file in the year of 2015 with a year-on-year growth of 473%, and 88 cases were put on record at the first half of 2016 with a year-on-year growth of 84.75%. For stock market price manipulation, the regulatory measures for now are implemented in accordance with the process of investigating, filing, and punishing, which follows a post-punitive management model. Although the manipulators were punished and the attempted manipulators were warned in this regulatory mode, its effect is still not ideal because the manipulation had already been done and the investors' loss and the harm caused to the market were irreversible. Therefore, how to interpose the stock market price manipulation in advance and to identify the specific indicators with which the stock is likely to be manipulated so as to monitor and terminate the manipulation in time are the key processes in combating the manipulation from post-punishment to prevention and control.

John and Narayanan [1] are the first foreign scholars to study on how to identify manipulation. He uses financial statements to choose variables and determines the

manipulation of insider trading with logistic model, whose correct rate is about 60%. Huang Changqing, Chen Weizhong, Du Shaojian [2] take the manipulation cases occurred before 2002 as study sample, by analyzing the changes in indicators of the rate of return and the trading volume in the early, initial, mid and late stages of the manipulation, and they find that the average daily excess turnover and the average daily excess rate of return increase significantly during the manipulation. Lu Rong, Chen Xiaolin [3] test the accuracy of the neural network model, the decision tree model and the Logistic regression model in determining manipulations. The research shows results based on testing database that the neural network model works the best and the decision tree model is the second, and the Logistic regression model has the lowest accuracy. However, the Logistic regression model is the best when the self-selected data set is used, and the neural network is the second, and the decision tree is the worst. Fang Yuan, Hu Daiping [4] design a stock price manipulation determining system based on case reasoning by choosing indicators from two aspects of price/volume and risk. Li Mengyu [5] uses the difference-in-differences estimation to analyze the manipulation behavior of China's stock market, and the study finds that the daily return, effective spread, price effect and transaction size of manipulated stocks increases significantly during the operation period, and market manipulation early warning indexes are proposed based on Logit model in the study. Qian Shuitu, Chen Xinyun [6] also draw the conclusion that the Logit model has a good predictive effect.

There are still some shortcomings of existing studies mentioned above: (1) The main problem solved by existing researches is how to identify stocks that have been manipulated with high accuracy, and the results of the studies determine that whether the stock is manipulated or not, which makes contribution in offering technology in post-manipulation judgment, but no in-depth study is carried out on prior-manipulation judgment technology. (2) The data used in the existing studies are collected with the incident research method, which couldn't unify the samples into the same time series, so they are not suitable for the new feature of the manipulation - the cycle is shorter and shorter(usually one to two months). (3) Most of the indicators used to determine the manipulation are likely to show significant differences during the manipulation process, so when there are abnormal changes with these indicators, it means that the stock is already

Fund project: projects supported by National Social Science Foundation (16BGL066)

being manipulated, thus the existing methods and indicators are not suitable for early warning of the manipulation.

The innovation of this paper lies in the empirical demonstration with panel data, which can overcome the insufficiency of time series data multicollinearity. Selected indicators mainly reflect the characteristics of listed companies and the collection is done before the manipulation. The result of the study is to determine whether the stock is easy to be manipulated and it is used to predict the likelihood of a stock being manipulated before it occurs. The study conclusion provides technical support for management departments in focal regulation with limited resources.

II. DATA RESOURCE AND MODEL BUILDING

A. Data resource and description

According to the punished stock price manipulation cases recorded by China Securities Regulatory Commission in 2015, we find a total of 54 companies were involved in stock market manipulation disclosed by China Securities Regulatory Commission by manual collection. Excluding 4 companies whose data couldn't be acquired, a total of 50 effective basic samples are obtained and all the data of companies that are manipulated are derived from Wande Database. Among 50 basic samples, the Shanghai stock market accounts for 12, Shenzhen main-board 6, Shenzhen small and medium-sized board 12, and Shenzhen GEM 20. 20% of the basic samples were tested, and in order to ensure the reasonable representative and structure of test samples, each market plate is randomly selected in proportion and 11 samples are obtained. Finally, 39 cross-section samples are obtained, and each sample is filtered in accordance with the time series of four quarters of 2015 to build the panel data regression stock manipulation forecasting model.

B. Forecasting model building

Since the explained variables in this paper indicate whether the company's stock is easy to be manipulated and are binary variables pertaining to the discrete variable, so a discrete data model should be established for regression analysis. Since there isn't normal distribution assumption in the Logit regression model, and it's proved that the Logit regression model has a good judgment effect, so the Logit regression model is chosen in the study to construct the stock manipulation forecasting model.

Set the dependent variable y on behalf of the stock's likelihood of being manipulated, and take 2 values of 0, 1, $y=1$ stands for the fact that the company's stock is easy to be manipulated, $y=0$ stands for the fact that the company's stock is not easy to be manipulated. We use panel data in the study, so we need to add a parameter which reflects the individual heterogeneity. We add the unobservable individual effect parameter u_i in the model. If there is no individual effect, $u_i = 0$ and if there is fixed individual effect, then u_i is a fixed parameter, and if there is a random individual effect, u_i is a random variable. The panel Logit model is shown below.

$$y_i = \text{logit}(f(p)) = \ln\left(\frac{f(p)}{1-f(p)}\right) = \beta_0 + u_i + \beta_i x_i + \varepsilon \quad (1)$$

In the model, $P(y = 1) = p, P(y = 0) = 1 - p$, p represents the probability of $y = 1$ when the independent variable is x_i , $\ln\left(\frac{f(p)}{1-f(p)}\right)$ is the odds ratio of variable $y=1$.

C. Forecasting indicator selection

1) The basis of indicator selection

Stock market trading manipulation is defined in the securities law as "the behavior of manipulating security trading price or security trading volume alone or through conspiracy by accumulating capital advantage, stock holding advantage or continuous trading". It can be seen that the manipulator will take two aspects into account while choosing the target: first, the target could meet the capital advantage requirements so that manipulation cost could be saved while still being offered with manipulation ability; second, the target can easily attract other investors to follow the transaction to reach the manipulation goals of suppressing, fund raising, forcing up and selling. Based on this, 14 secondary forecasting indicators of 2 categories are selected in the study though text analysis.

The first category is the capital advantage demands indicators which are selected from the perspective of the manipulators in choosing the target stock, including the average stock holding ratio, the number of institutional holding, free float market capitalization and price-earnings ratio. The second category is indicators showing the targets' ability in attracting investors to follow the transaction which are selected from the perspective of the seduced ones in choosing target investment stock, including the average daily turnover sum, the average daily turnover volume, the average turnover rate, rise and fall amount compared with big board, net assets yield rate, current ratio, net profit year on year growth rate, Z value, volatility, and Beta coefficient. The connation and the algorithm of above indicators are from the Wande Database.

2) Sample data test analysis

First of all, the descriptive statistical analysis is carried out on the 14 indicators of 156 samples' panel data with software stata12.0 to check data dispersion, and modify the possible existence of abnormal data. The criterion for abnormal value is usually based on whether or not it exceeds the mean by two standard deviations. If the sample value exceeds the mean by three standard deviations, then it is called extreme anomaly, and the result of the first descriptive statistical analysis showed that abnormal value exists in most data. Due to excessive content, the result of the first descriptive statistical analysis is omitted here. In this paper, we use the method of winsorization to smooth the anomaly data. The principle is to find the quantile of the 1% and 99% of each variable. For example, for the variable X1, the quantile is a, b, and then replace the values less than a into a and replace the values higher than b into b, so the original data are directly altered into new data and abnormal data which are higher or lower than 1% are winsorized. According to the results of the first descriptive statistical analysis, we find that most of the indicators have a high anomaly value, so we use the corresponding quantile of 10% and 90% and winsorize each indicator in software stata12.0 and carry out the second descriptive statistical analysis whose result is omitted here due to massive content. It can be seen from the results that there is no abnormal value in each indicator, so they can be used for regression analysis.

Secondly, a relevant analysis is carried out on pre-selected indicators. When there are too many indicators, multicollinearity is likely to show up among them. In order to prevent multiple variables from simultaneously explaining one variable and reduce the accuracy of the forecasting model, we carry out a test to verify the interdependency among 14 indicators mentioned above, and the result is omitted here due to massive content. It can be seen from the result that only the absolute values of the correlation coefficient of x_3 and x_7 , x_{12} , x_6 and x_{14} , x_7 and x_8 , x_{12} , x_8 and x_{12} are higher than 0.5 and

the interdependency among the other majority is low, so the indicators could be used for empirical analysis.

3) Forecasting indicator selection

In this paper, the stepwise regression function of software stata12.0 is used to carry out stepwise regression on the firstly selected 14 indicators, and then key influencing indicators of the forecasting model are obtained. The results of stepwise regression are shown in Table I .

TABLE I. STEPWISE REGRESSION RESULT OF 14 INDICATORS

Ordered logistic regression				Number of obs = 156 LR chi2(3) = 26.1 Prob > chi2 = 0.0000 Pseudo R2 = 0.1255		
Log likelihood = -90.89						
Explained variable	Coefficient estimation	Std. Err.	z	P>z	[95% Conf.	Interval]
Average stock holding ratio	6.7276	3.4470	1.9500	0.0490	-0.0284	13.4835
Volatility	0.0540	0.0148	3.6500	0.0000	0.0250	0.0831
Beta coefficient /cut1	-0.2704	0.0905	-2.9900	0.0030	-0.4478	-0.0931
	1.5668	0.3726			0.8366	2.2971

After stepwise regression, the average stock holding ratio indicator (x_1) is selected among the original 4 capital advantage indicators, and two indicators of volatility (x_2) and Beta coefficient (x_3) are selected from 10 copy trading indicators. These 3 indicators will be included into the panel Logit regression model to build the forecasting model.

III. EMPIRICAL RESULT ANALYSIS

A. Empirical result

In empirical analysis, stata12.0 is used to conduct panel data binary logic regression and to check the effect. In order to

prevent the occurrence of dummy trap and constrain $u_i = 0$, we perform Hausman test on the panel Logit regression model. Test result shows Prob<chi2 = 0.0582, so the original hypothesis is accepted and the random effect model is select. Taking into account the readers' habit, the empirical result is showed in the form of regression coefficients, as shown in Table II .

TABLE II. LISTED COMPANIES STOCK TRADING MANIPULATION FORECASTING PANEL LOGIT REGRESSION RESULT

Random-effects GLS regression				Number of obs=156.00 Number of groups=39.00 Obs per group: min =4.00 avg=4.00 max=4.00 Wald chi2(3)=29.33 Prob > chi2=0.0000		
corr(u_i, X)=0(assumed)						
Coef.		Std. Err.	z	P>z	[95% conf. Interval]	
Average stock holding ratio	1.35	0.67	2.02	0.044	0.04	2.66
Volatility	0.01	0.00	4.10	0.000	0.01	0.02
Beta coefficient	-0.06	0.02	-3.31	0.001	-0.09	-0.02
Constant	0.16	0.07	2.23	0.026	0.02	0.30
sigma_u	0.00					
sigma_e	0.48					
rho	0.00					

$$Y = 0.16 + 1.35x_1 + 0.01x_2 - 0.06x_3 \quad (2)$$

The logic regression result showed in Table 2 is based on 156 sample panel data, and it can be seen from the result that average stock holding ratio, volatility and Beta coefficient all pass the Z statistic test under confidence ratio of 95%, indicating that these indicators have a significant effect on explained variables that they can be used as forecasting indicators in determining the stock's likelihood of being manipulated. According to the regression results, the stock trading manipulation forecasting model of listed companies can be obtained:

$$Y = 0.16 + 1.35x_1 + 0.01x_2 - 0.06x_3 \quad (3)$$

Since Y is a discrete variable, and it only takes two values of 0, 1, therefore, when $Y \geq 0.5$, Y is 1, and when $Y < 0.5$, Y is 0. Among 4 capital advantage indicators, only the average stock holding indicator is significant, and the other 3 indicators didn't show significant correlation with the stock's likelihood of being manipulated. Among 10 copy trading indicators, only volatility and Beta coefficient showed significant correlation with the explained variables, while the other 8 indicators including daily average turnover are not significantly related with them.

In order to verify the validity of the model, stata12.0 is used to test the forecasting accuracy both in and out of the sample, and 0.5 is set as the probability critical point to carry out forecasting. When $Y \geq 0.5$, it means that the stock of the company is easy to be manipulated, and when $Y < 0.5$, it

means that the stock is not likely to be chosen as the manipulation target. The in-sample test result is shown in Table III; the result of out-of-sample test is omitted here due to excessive contents.

TABLE III. LISTED COMPANIES STOCK TRADING MANIPULATION FORECASTING MODEL IN-SAMPLE TEST RESULT

Classified	----- True -----		Total
	D	~D	
+	36	18	54
-	24	78	102
Total	60	96	156
Classified + if predicted Pr(D) >= 0.5			
True D defined as V != 0			
Correctly classified			70.63%

It can be seen from the result of the in-sample test, the forecasting model can correctly judge for 36 times in the 60 stock manipulations, and correctly judge for 78 times in the 96 stock non-manipulations, showing an overall prediction accuracy of 70.63%, which is higher than the 60% accuracy of John and Narayanan [1]. The prediction accuracy is good. Similarly, the prediction accuracy is 68.18% with out-sample test result, indicating a relatively good effect.

B. Empirical result analysis

The regression result analysis of panel Logit in Table 2 is as follows: (1) Average stock holding ratio. The average stock holding ratio reflects the controllability of the manipulator on manipulating the stock price. When the manipulator disturbs the stock price with capital advantage, the higher the stock holding ratio is, the more advantage the manipulator has, and the fewer obstacles exist in disturbing the price, the more obvious the manipulator's capital advantage is, and the easier to force up or suppress the price. It can be seen from the regression coefficient that with the average stock holding ratio increasing by 1 percentage, the probability of the stock being chosen as the manipulation target will rise by 1.35%. (2) Volatility. The value of the volatility can be reflected in the technical analysis graphic on trading software interface. The higher the volatility is, the greater volatility of the stock price is, the more active the stock is, the more implicit speculation opportunities exist in stock trading, and the easier for the manipulator to induce the following traders [7]. It can be seen from the regression coefficient that with the volatility ratio increasing by 1 unit, the probability of the stock being manipulated increases by 1%. (3) Beta coefficient. Beta coefficient is the main reference indicator for the investors to evaluate the stock risk when choosing a stock. A low Beta coefficient indicates a relatively lower risk of the stock compared with the whole market. The lower the risk is, the more likely for it to be favored by the investors, and the manipulators who choose stocks with this characteristic indicator is easier to induce the investors to follow transaction. It can be seen from the regression coefficient that with Beta coefficient increasing by 1 unit, the probability of the stock being manipulated decreases by 6%.

IV. CONCLUSION

To improve the efficiency in managing stock market trading manipulation, management departments need to switch from post-punitive supervision into pre-preventive monitoring.

To figure out the features with which the company's stock is likely to be chosen as the target and whether we can effectively predict which kind of company stock is easy to be manipulated are the key points in reasonably allocating regulatory resource and carry out focal regulatory. In this paper, based on the two facts that capital advantage is needed for manipulation and other investors are permitted to follow transaction, 14 influencing factor are selected with the method of text analysis which can reflect the company stock's likelihood of being manipulated, and the 2015 stock manipulation cases issued by China Securities Regulatory Commission are used as sample to conduct empirical analysis with panel data Logit model. Average stock holding ratio, volatility, and Beta coefficient are key factors filtered from 14 indicators in forecasting the stock's likelihood of being manipulated with the method of stepwise regression method. A stock trading manipulation forecasting model is constructed based on random panel Logit model by Hausman test. By testing the in-sample and out-of-sample forecasting effect of the model, the forecasting accuracy reach 68.18% and 70.63% respectively, showing a relatively good effect.

REFERENCES

- [1] John, K. and Narayanan, R. Market Manipulation and the Role of Insider Trading Regulations [J]. *Journal of Business*, 1997, 70(2): 217-247.
- [2] Huang Changqing, Chen Weizhong. An Empirical Study on Stock Price Manipulation in China's Securities Market. [J]. *Journal of Tongji University (Natural Science)*, 2004, 32(9). (In Chinese)
- [3] Lu Rong, Chen Xiaolin. Research on Market Performance and Discrimination of Stock Manipulation Behavior, 2009, (4). (In Chinese)
- [4] Fang Yuan, Hu Daiping. Research on Judgment System of Stock Price Manipulation Based on Case-based Reasoning [J]. *Science Technology and Engineering*, 2011, 11(5). (In Chinese)
- [5] Li Mengyu. Research on China's Stock Market Manipulation and Early Warning Mechanism [J]. *Journal of Central University of Finance & Economics*, 2015, (10). (In Chinese)
- [6] Qian Shuitu, Chen Xinyun. Analysis on the Influencing Factors of Regional Risk of Rural Credit Cooperatives - Based on Panel Data Logit Model [J]. *Journal of Financial Research*, 2016, (9). (In Chinese)
- [7] Táliš J. Putniniš. Market Manipulation: A Survey [J]. *Journal of Economic Surveys*, 2012, 26(5):952-967.