The 2022 International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/6/2022170

Research on the Influencing Factors of Stock Returns

Cheng Peng[1,a,*]

[1]*Statistic and Mathematic College, Yunnan University of Finance and Economy, Kunming, 650000, China*
a. 202005002386@stu.ynufe.edu.cn
**corresponding author*

Abstract: In consideration of the development of Chinese stock market, it is important to find out the factors that affect stock return. China A-share market is an immature capital market and difficult to change the nature of enclosure of money while foreign stock markets were established earlier and developed more mature. They have made fruitful achievements in the study of the causes of stock price fluctuations. At the same time, there are many successful experiences in guiding investment practice with investment analysis theory. Surveys shows that the average return of Chinese investors is negative. So, this is helpful for investors to have a better understanding of stock market. And it provides information which may reveal several problems of stock market at this stage. This paper mainly focuses on the influencing factors of stock return in Chinese stock market, using the most realistic data to try to figure out the relationship. The data of 282525 stocks which are from November 2020 to November 2022 to achieve multiple linear regression, and the result shows that PE (price-to-earnings ratio), PB (price-to-book ratio) and turnover are negatively correlated with stock return. Among them, turnover has the greatest influence on stock return, followed by PB and PE. Other two factors which are circulated market value and amount have little influence on stock market, so they can be ignored.

Keywords: Stock return, Chinese stock market, Multiple linear regression.

1. Introduction

With the development going fast in the field of Chinese securities market, figuring out the influencing factors of stock returns is important for those investors to analyze the situation of securities market as well as control the risk of loss. There are many factors to influence the securities returns, so it is initial to research the relation between influencing factors and securities returns. Stock market price fluctuation is the basis of stock market operation and the focus of stock investors. The fluctuation of stock price is affected by various economic factors and non-economic factors, macro factors and micro factors, analyze the basic influencing factors of stock price changes of listed companies, eliminate irrelevant factors, focus on relevant factors and combine the relevant factors with the analysis of growth of companies to dig out the potential intrinsic value of companies. It is useful for investors to make the correct decision on investments, at the same time to improve the qualities of listed companies as well as form a rational development in the stock market.

Wu applied ARMA-ARCH model and ARCH effect test as the research model and analysis method in stock return. After creating the model, correlation analysis and stationary test was done

DOI: 10.54254/2754-1169/6/2022170

to verify the accuracy and the feasibility of the model. In this paper, ARMA-ARCH model is used, showing the fluctuation of CSI 300 stock disobey normal distribution, and indicating that most of the Chinese investors are irrational [1]. Gong divided the influencing factors into 3 categories, they are macro factors, medium factors and micro factors, and applied panel regression model as the research method to modeling. From the result we can get that A-shares are mostly affected by macro factors. As for the micro factors, due to the difference of cost and model of operation in each industry, the profits also have rather fair-sized difference [2]. Gong applied the ARCH series model as a method in analysis of time series in stock returns, and single factor analysis with multi-factor cross-section analysis are brought out. Showing that the degree of fitting the equation cannot up to over 80%, that means there are some unknow factors are influencing the stock returns [3]. Wang divided the influencing factors into macro-economic factors and micro risk factors, and the conclusion shows that the relation between system risk and stock return is difficult to determine. When liquidity up to a certain point, liquidity is directly proportional to stock returns [4]. Zhang used the method of Fama-Macbeth regression analysis and based on the improved Fama-French three-factor model, empirical analysis has been conducted to obtain the result which indicates that in western developed country, the applicability of Fama-French three-factor model have been proved. According to the method of multiple linear Regression, descriptive statistical analysis and other factors, Fama-French three-factor model applies to Chinese A-Share market [5].

Xue used PCA and multiple linear regression in pharmaceutical industry and found out that indicators of profitability and Equity expansion ability index have a marked effect on the stock return of pharmaceutical industry listed companies [6]. Yi applied Fama-French five-factor model, factor Enhanced Quantile Regression and Panel Data Model. The conclusion of five-factor model indicates that Fama-French five-factors model have some explanatory ability on the stock return of New Third Board. Using panel data to carry on the regression analysis to predetermined variable and common factor variable and shows that the scale of companies, book-to-market and investment factors have significant negative effects on stock return [7]. Zhou applied valuation factor and Fama-French five-factor model to conclude that Fama-French five-factor model have preferable explanatory ability to the Stocks' Return in China's GEM, found out that China's GEM stock market size effect is the most significant [8]. Zhang applied cluster analysis and multiple linear regression between stock return and each indicator. The conclusion of correlation analysis shows that there exists strong correlation among value β, book-to-market, earnings per share and the growth rate of net profit. The multiple linear regression shows a significant linear correlation between stock return and selected indicators [9]. Li applied five-factor model and Fama-Macbeth cross section regression and indicated that there is a significant small market capitalization premium in the stock circulation value of each market segment, investing in small-cap stock can yield higher returns than investing in large-cap stocks [10]. For research on influencing factors of stock return, Fama-French model is used a lot and it is not easy to get a satisfying fitting equation.

This paper will use multiple linear regression as the initial method, and turnover rate, which affects earnings most directly, is used as the explanatory variable, use other five indicators including circulation market value, turnover, rise and fall, PB(price/book value ratio), P/E ratio as control variables for those are typical factors. Stock return is used as explained variable.

2. Method

2.1. Introduction of the Method and Problem Encountered

To determine the factors that affect stock return, predecessors tried a lot of things. In this paper, multiple linear regression is chosen because it is one of the most practical ways to study the relationship between variables and research trends, so it is simpler and easier to be understand

DOI: 10.54254/2754-1169/6/2022170

compare with other methods. In the beginning, there are 7 variables are chosen to refer to other literature, after conducting coefficient analysis, a variable that shows low significance is removed. Before carrying out linear regression, preliminary work including descriptive statistical analysis, correlation analysis and the significance of the coefficients for each variable are necessarily done.

A major problem with the experimental method is that the correlation analysis shows that there is no obvious relationship between explained variable and explanatory variables, so that the value of R square of regression equation is rather small. However, if the regression equation is not used as a prediction, it is still reliable.

2.2. Describing of the Data

The data of stock return is downloaded from CSMAR database and original 100,000 pieces of them is selected from November 2020 to November 2022 for the period of 2 years. In the sample, the value of each variable is partially missing, therefore, data cleaning is conducted, and null values are deleted in order to convenient for the following work. After data cleaning, 282,525 pieces of data is remained.

2.3. Descriptive Statistical Analysis

In order to process the data smoothly, and get an overall overview of the data, considering the indicator which is error of mean square is important to know whether the factors are suitable for being explained variable and explanatory variables.

2.4. Correlation Analysis

Correlation Analysis is also vital for knowing if there exist linearity between explanatory variables and to prevent the T-value is not significant with the high value of R square as well as the low explanatory power of regression equation.

2.5. Scatter Plot

Scatter Plot is for each explanatory variable against explained variable, and it shows the distribution and aggregation of data, and the trend line formula can also be obtained. Scatter plots can not only convey information about the type of relationship between variables, but also reflect the degree of clarity of the relationship between variables.

2.6. Coefficient Analysis

To see if the coefficient of each variable is suitable, according to the significance of each coefficient is very useful. T test can often be used to test the significance of each parameter in the regression equation, while F test can be used to test significance of the whole regression relationship.

2.7. Analysis of Variance

Analysis of variance in a comparable array, the total variation between the data is decomposed by each specified source of variation. For the measure of variation, the sum of squares of deviation is used. The analysis of variance method is to decompose the sum of the squares of partial deviation that can be traced back to the specified source from the total deviation squared.

DOI: 10.54254/2754-1169/6/2022170

2.8. Multiple Linear Regression

To show the relationship between explained variable and each explanatory variable, forming a regression equation.

3. Results and Discussion

3.1. Preparatory Work

In order to figure out the relation between stock return and other factors, partial stock data form November 2020 to November 2022 which was selected in CSMAR database is used. Using R studio as the initial tool and SPSS as an assistant to conduct data cleaning as well as a series of statistical analysis including descriptive statistical analysis, correlation analysis between variables, multiple linear regression as well as significance testing. In descriptive statistical analysis, the probability level to use to calculate the confidence interval on the mean is 0.95. For simplify the table 1 which are given followingly, assigning a different symbol to each variable.

Table 1: Symbols of Variables.

Variable	symbol
Stock Return	y
PE	x_1
PB	x_2
Turnover	x_3
Circulated Market Value	x_4
Change Ratio	x_5
Amount	x_6

Giving descriptive statistical analysis for each variable, and the result is given in the table 2.

Table 2: Descriptive Statistical Analysis for Each Variables.

	median	mean	S. E.	var	S. D.
y	1.09	1.74	0.00	3.53	1.88
x_1	33.47	54.43	0.25	18361.86	135.51
x_2	2.80	4.63	0.01	39.39	6.28
x_3	0.01	0.03	0.00	0.00	0.05
x_4	58.72	34.05	25.87	18.91	13.75
x_5	0.00	0.00	0.00	0.02	0.13
x_6	93.11	30.44	14.01	55.51	74.50

In Table 2, those 6 variables including error of mean square of stock return, P/E ratio, PB (price/book value ratio), turnover, change ratio and amount are all less than 1, that means the degree of deviation between a random variable and its estimated value is small enough. It can be considered that those variables are suitable in multiple linear regression as explanatory variables and explained variable.

DOI: 10.54254/2754-1169/6/2022170

3.2. Correlation Analysis

Correlation analysis between variables is to find out whether these variables are multicollinearity. It is obvious that the fitting equation can easily affected by strong linear, the variance of the estimated

Table 3: Correlation Analysis Between Variables.

	y	x_1	x_2	x_3	x_4	x_5	x_6
y	1.000	-0.173	-0.288	-0.184	0.136	-0.002	-0.025
x_1	-0.173	1.000	0.191	0.064	-0.017	0.003	0.100
x_2	-0.288	0.191	1.000	0.188	0.052	0.002	0.192
x_3	-0.184	0.064	0.188	1.000	-0.091	0.019	0.096
x_4	0.136	-0.017	0.052	-0.091	1.000	-0.001	0.522
x_5	-0.002	0.003	0.002	0.019	-0.001	1.000	0.011
x_6	-0.025	0.100	0.192	0.096	0.522	0.011	1.000

regression coefficient will get larger and the confidence interval for the regression coefficient will get wider with stronger linear.

In Table 3, the absolute value of correlation coefficients between variables are less than 0.6, representing the non-linear correlation between variables. In addition, partial correlation coefficient is introduced for it is more persuasive.

Table 4: Partial Correlation Coefficient.

	y	x_1	x_2	x_3	x_4	x_5	x_6
y	1.000	-0.117	-0.244	-0.116	0.140	0.002	-0.033
x_1	-0.117	1.000	0.133	0.000	-0.052	0.001	0.087
x_2	-0.244	0.133	1.000	0.128	0.018	-0.003	0.135
x_3	-0.116	0.000	0.128	1.000	-0.138	0.017	0.129
x_4	0.140	-0.052	0.018	-0.138	1.000	-0.005	0.534
x_5	0.002	0.001	-0.003	0.017	-0.005	1.000	0.011
x_6	-0.033	0.087	0.135	0.129	0.534	0.011	1.000

Partial correlation coefficient indicates the degree of correlation between two variables after eliminating the influence of other variables. The result of partial correlation coefficients indicates that between variables, there are no obvious correlation.

3.3. Scatter Plots

For each explanatory variable towards explained variable, conducting scatter plots and performing linear fitting.

In Fig.1, only the p-value of fitting equation of stock return to change ratio is 0.386, which greater than the given level of significance 0.05. That means the linear regression is not statistically significant while the linear regression of five other variables shows strong significance, due to their p-values are all less than 0.05. Giving these explanatory variables an analysis of coefficient by SPSS.

DOI: 10.54254/2754-1169/6/2022170

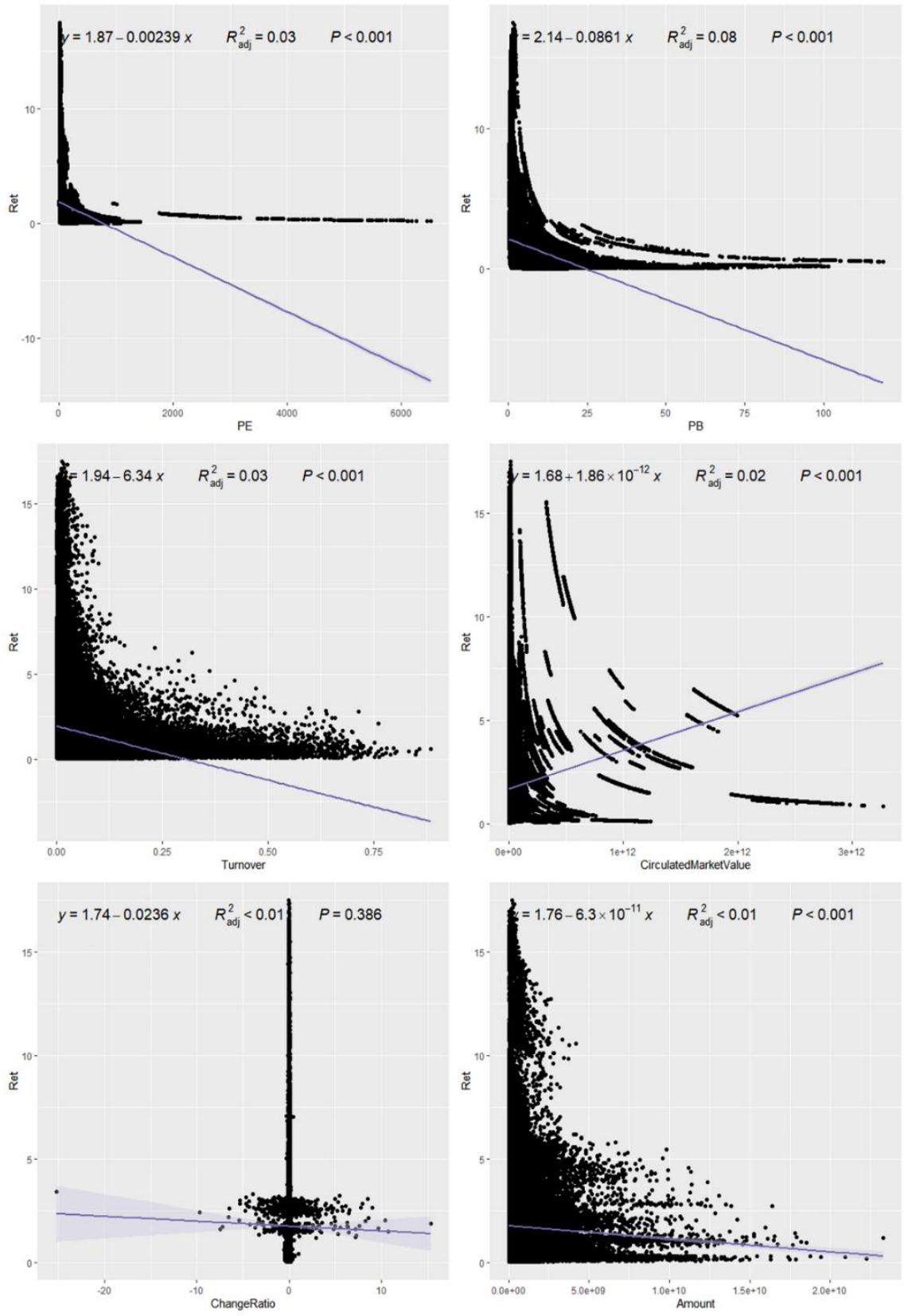

Figure 1: Scatter Plots of Each Explanatory Variable to Explained Variable.

DOI: 10.54254/2754-1169/6/2022170

3.4. Coefficient Analysis

Now consider coefficients of these six variables. In the table of coefficient, when the significance of the variable is bigger than given level of significance, the variable is not statistically significant in multiple linear regression model, and that corresponding variable should be deleted from the regression model. Not vice versa. The coefficient table is given below.

Table 5: Coefficients of six variable.

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
Model	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	2.239	0.004		501.906	0.000		
x_1	-0.002	0.000	-0.112	-63.093	0.000	0.954	1.049
x_2	-0.074	0.001	-0.246	-135.069	0.000	0.907	1.102
x_3	-3.814	0.061	-0.112	-62.355	0.000	0.936	1.068
x_4	0.000	0.000	0.161	77.925	0.000	0.706	1.417
x_5	0.000	0.000	-0.040	-19.180	0.000	0.684	1.462
x_6	0.027	0.025	0.002	1.065	0.287	1.000	1.000

In Table 5, the significance of change ratio is 0.287, obviously larger than the given level of significance 0.05, the variable is not statistically significant in multiple linear regression model. So, change ratio need to be removed because it is unsuitable as an explanatory variable in multiple linear regression.

3.5. Multiple Linear Regression

Multiple linear regression is performed by SPSS. The model summary is given in the table below.

Table 6: Model Summary.

Model	R	R^2	Adjusted R^2	Std. Error
1	0.367	0.134	0.134	1.741

The basic information is shown in Table 6, R square is used to describe the explanatory nature of the model to explanator variables. Despite R square value is only 0.134, the model is not used to make a prediction, so the value of R square can be taken off the table.

Analysis of variance compares whether multiple population means are equal or not, but its essence is to study the relationship between variables, which is the influence of classified independent variable and numerical dependent variable.

Analysis of variance. Using ANOVA to determine the significance of multiple linear regression populations, which means whether the input data can establish a valid equation. The analysis of variance is shown below.

The significance of F test is less than the given level of significance 0.05. Therefore, reject the null hypothesis, consider the regression equation is significantly effective.

DOI: 10.54254/2754-1169/6/2022170

Table 7: ANOVA.

	Sum of Squares	df	Mean Square	F	Sig.
Regression	135202.111	5	27040.422	8916.620	0.000b
Residual	871207.550	287282	3.033		
Total	1006409.661	287287			

Regression equation. In Table 5, the significance of each explanatory variable is less than the given level of significance 0.05, so these variables are statistically significant in the model and should be retained.

Therefore, the regression equation can be written as

$$y = -0.002x_1 - 0.074x_2 - 3.812x_3 + 2.206^{-12}x_4 - 1.013^{-10}x_5 \quad (1)$$

Although explanatory variables can explain 13.4% of the variation in stock return, which is not satisfying, it is still showing the relationship between stock return and other five variables. Four of explanatory variables including PE, PB turnover and amount are negatively correlated with stock return, only circulated market value has a slight positive correlation to stock return.

4. Conclusion

According to the regression equation (1), PE, PB and turnover rate have negative impact on stock return, and turnover rate influence stock return the most. A high turnover rate means active trading, strong willingness to trade in the market and good liquidity. At the same time, the stock with a higher turnover rate will have larger stock price fluctuations and greater risks. Among these five factors, PB has the second largest influence on stock return, followed by PE. As for the other two factors which are circulated market value and amount, their effect on equity yields is too small to be negligible.

References

[1] *Wu C 2014 Empirical Studies of the factors affecting the CSI 300 stock East China Institute of Technology pp 1-54.*

[2] *Gong C 2016 The Empirical Research of Factors Affecting Stock Return-Based on The Share Market Shanghai Jiaotong Uniersity pp 1-41.*

[3] *Gong J 2006 Research on the influencing factors of stock yield in Shanghai Securities Market Southwestern University of Finance and Economics pp 1-57.*

[4] *Wang F 2010 The Theoretical Analysis and Empirical Analysis of Influencing Factors of A Share's Stock Revenue in The Shanghai Stock Exchange Nanjing University of Finance and Economics pp 1-39.*

[5] *Zhang J 2014 Analysis of the influence factors of the Shanghai A-share stock returns-based on the perspective of the improved Fama-French Three-factor model Dongbei University of Finance and Economics pp 1-33.*

[6] *Cheng R 2018 Research on the Influencing Factors of Stock Returns in China's Pharmaceutical Industry Tianjin University of Finance and Economics pp 1-38.*

[7] *Yin H 2018 Empirical Study on The Impact of New Third Board Stock Returns-Based on The Five-Factor Model and The Panel Model Dongbei University of Finance and Economics pp 1-42.*

[8] *Zhou Q 2019 An Empirical Research on the Factors Affecting the Stocks' Return in China's GEM-Based on Fama-French Five-Factor Model Shangong University pp 1-47.*

[9] *Zhang J 2007 An Empirical Research on Factors Influencing Stock Returns Ratio of Chinese Listed Companies Dalian University of Technology pp 1-44.*

[10] *Li Z 2017 An Empirical Analysis on The Influence Factors of Chinese A-Share Market's Stock Returns-Based on Fama-French Five-Factor Model Dongbei University of Finance and Economics pp 1-43.*