

Analysis of Differences and Correlation between Different Industries based on Stock Price Information

Xinying Hu^{1, †} and Wenyang Ma^{2, *, †}

¹ Department of Statistics (International Class), South China Agricultural University, 510642, Guangdong, China

² Department of Data Science and Big Data, North China University of Technology, 100144, Beijing, China

* Corresponding Author Email: 2058513@brunel.ac.uk

[†]These authors contributed equally.

Abstract. Stock price fluctuation is the focus of people's life. Based on stock price information, this paper focuses on China's oil industry and its downstream industries (such as garment and textile industry, automobile industry and decoration and building materials industry) and makes a statistical analysis on their correlation. After the basic statistical description of the average performance of ten leading stocks in each industry in the past ten years, the factor analysis, linear regression analysis and weighted least square methods are adopted, and finally the conclusion is drawn: there is a strong correlation between the four selected industries, among which the correlation between petroleum energy and clothing and textile is the strongest; Perhaps due to the rise of new energy vehicles, the epidemic and other reasons, although the automobile industry and the decoration and building materials industry have always been able to better fit and have a strong correlation, but in recent years, the two and the oil industry to a certain extent. This result can provide reference for investors' investment direction and industry policy direction to some extent, and can also be used as a method and model to explore how other stock price information reflects industry information.

Keywords: Stock price; oil industry; linear regression; factor analysis; correlation.

1. Introduction

With the convenience of transportation and the advent of the information age, transaction costs and information asymmetry are relatively reduced, and the connection between all walks of life is getting closer and closer. And stock price information can better reflect a company's fundamentals, industry prospects, valuation comparison and other information to a certain extent. From this, it is necessary to select several leading stocks in some industries and make statistical analysis based on the performance of comprehensive stock price information of these stocks in a long period of time, so as to infer the differences and correlations among different industries, mine some valuable information and put forward some constructive views.

As one of the important financial products in the financial market, stock not only has important investment value and market value attribute, but also is the basis of other financial derivatives. So the change of stock price has been the focus of people's attention. Using statistical regression model as the main tool, Zhang pointed out that there was no statistical difference between the impact of changes in information surface on large-plate plate and small-plate plate [1]. Hu pointed out that there were significant differences in stock price fluctuations of listed companies in different industries by using the method of one-way analysis of variance and unitary analysis of variance, and that there was a linear positive correlation between enterprise growth and stock price fluctuations [2]. Zhao adopted vector autoregressive model (VAR) in a pioneering way and concluded that: The accounting profit of listed companies has the dynamic regularity of industry characteristics and random walk, the accounting profit information has obvious information content and market effect, there is a positive correlation between the accounting information disclosed in the financial accounting reports of listed companies and the stock price, and the explanatory power of accounting information on the stock price is increasing year by year [3].

Specifically, as for the impact of crude oil price changes on industries, Lee and Ni have clearly pointed out that almost all industries in the United States have to a large extent suffered from the impact of oil price shocks resulting in reduced demand [4]. Many studies in western countries show that the economic impact of crude oil price changes in energy is reflected in various aspects. Based on the efficiency wage model, Carruth et al. found that the fluctuation of crude oil price would affect the equilibrium unemployment rate and then lead to the change of output [5]. Eksi et al. analyzed seven manufacturing industries in Turkey using vector error correction model (VEC) and found that chemical, metal and plastic industries were highly sensitive to changes in oil prices [6]. Tan and Zhang, based on the TVP-FAVAR model, argued that the rise in oil prices caused by specific reasons would hinder production [7]. Li used empirical mode decomposition (EMD), vector autoregressive model (VAR) and other methods to study the "crude oil -PAT-PET" industrial chain, identified the impact and fluctuation impact of crude oil price changes on PAT and PET industries, and put forward corresponding policy and investment suggestions [8]. Similarly, Yang also studied the stock price indexes of electric power and public utilities and believed that there was a two-way relationship between them [9]. In addition, Wang found that there was no significant correlation between the overall Chinese stock market and energy, while crude oil prices and coal prices had a dramatic impact on the stock prices of listed companies in the chemical industry [10].

Based on this, this paper select a number of oil stocks in the Chinese market and some leading stocks in downstream industries (including textile and clothing, automobile manufacturing and building materials) to conduct statistical analysis on the monthly span stock price data from January 2013 to December 2022 in the past ten years, aiming to evaluate and judge the differences and relevance of these industries by descriptive statistical analysis and inferential statistical analysis.

2. Methods

2.1. Data Source and Basic Data Processing

The data of this study includes ten leading stocks in each of four Chinese industries, namely petroleum industry, clothing and textile industry, automobile industry and decoration and building materials industry. Each stock includes 120 stock price data in nearly ten years from January 2013 to December 2022 at monthly intervals. Each stock price data adopts the opening price of the forward adjusted option, and the specific stock code is shown in Table 1. The data source is Choice Financial Terminal.

Table 1. Stock list

Stock code(petroleum)	Stock code(textile)	Stock code(automobile)	Stock code(materials)
601857	600398	002594	000786
600028	002687	000625	000877
000554	002087	601633	002271
600688	601339	600104	600512
000096	000729	601238	601390
002221	002269	000951	600368
000059	600493	600006	600820
600256	600177	000800	601518
000637	002563	600166	601107
000819	600137	600066	002043

In terms of basic data processing, in order to better replace the corresponding industry with ten leading stocks, the average value of ten stocks in the same industry is taken according to the same time, and the missing data items are not included in the mean value formula, so as to obtain the development situation of each industry in the past ten years, make a curve with time as the horizontal

coordinate and stock price as the vertical coordinate, and then calculate the Basic descriptive statistics such as mean value variance, maximum value, etc.

2.2. Research Method

This research method is the main factor analysis and linear regression analysis, and the weighted least squares method. Firstly, correlation factor analysis is conducted on the four variables to judge the contribution rate of each factor. According to ten years after the average stock charts in a wide range of industries can see the direction of the general relationship between grouping study again. With linear regression analysis of the residual distribution first, or view the homogeneity of variance. Does not apply to homogeneity of variance of the linear weighted regression.

3. Results and Discussion

In order to study the correlation impact of the petroleum industry on its downstream industries, we select the petroleum industry and its representative three downstream industries, namely clothing and textile, automobile and decoration and building materials, and select three representative 10 leading stocks of each industry.

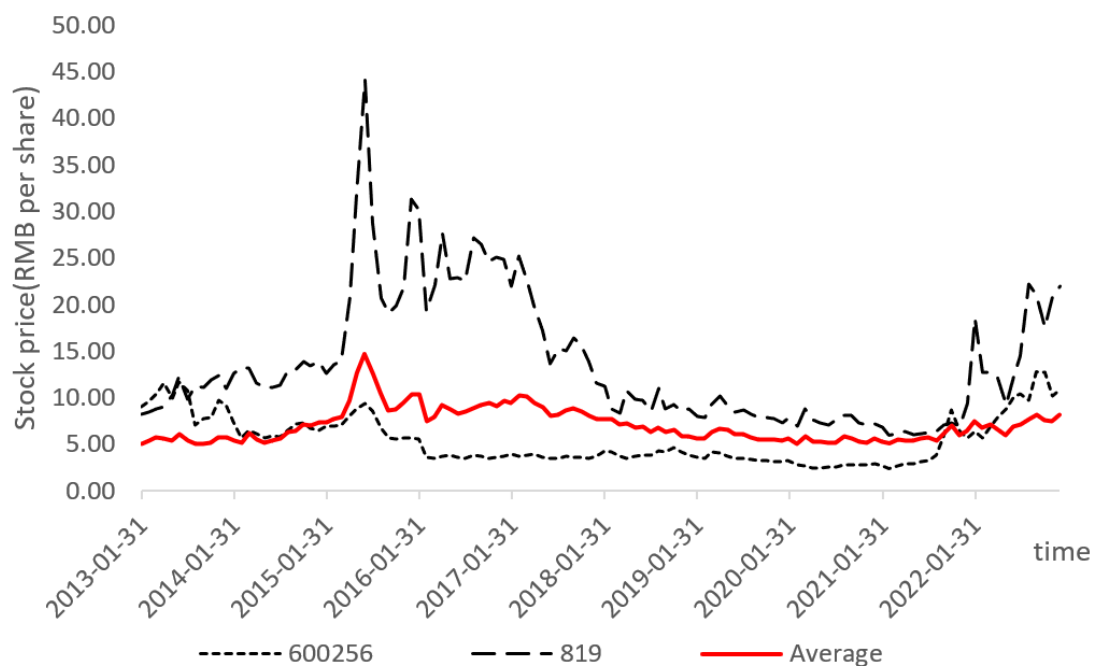


Fig. 1 Petroleum energy

(Note: select some stocks of the same type and the average value as example)

Fig. 1 shows the stock prices of the Petroleum energy industry from 2013-1-31 to 2022-1-31, from which we can see that the average trend of all stocks increased slightly in 2016, possibly because the US Petroleum energy stocks plummeted at that time, which led to the growth of Chinese Petroleum energy stocks.

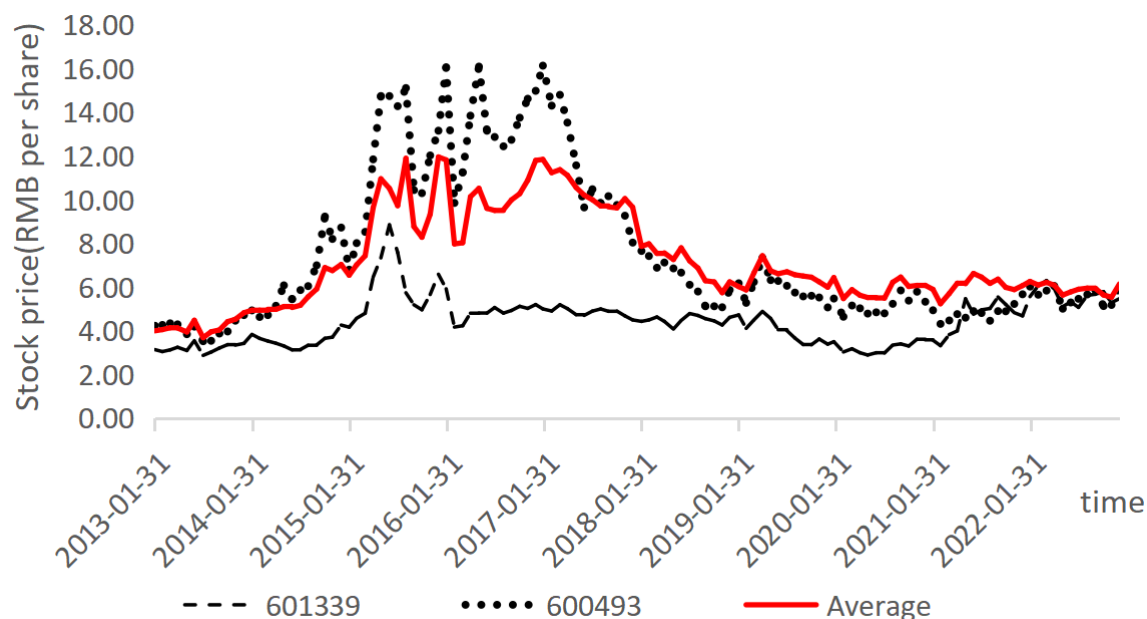


Fig. 2 Clothing and textile

Fig. 2 shows the stock prices of the Clothing and textile industry from 2013-1-31 to 2022-1-31. As can be seen from the graph, the overall trend of stock prices in all relevant industries has been stable. In 2019, the price of all stocks dropped slightly, possibly due to the influence of the rise of online shopping at that time.

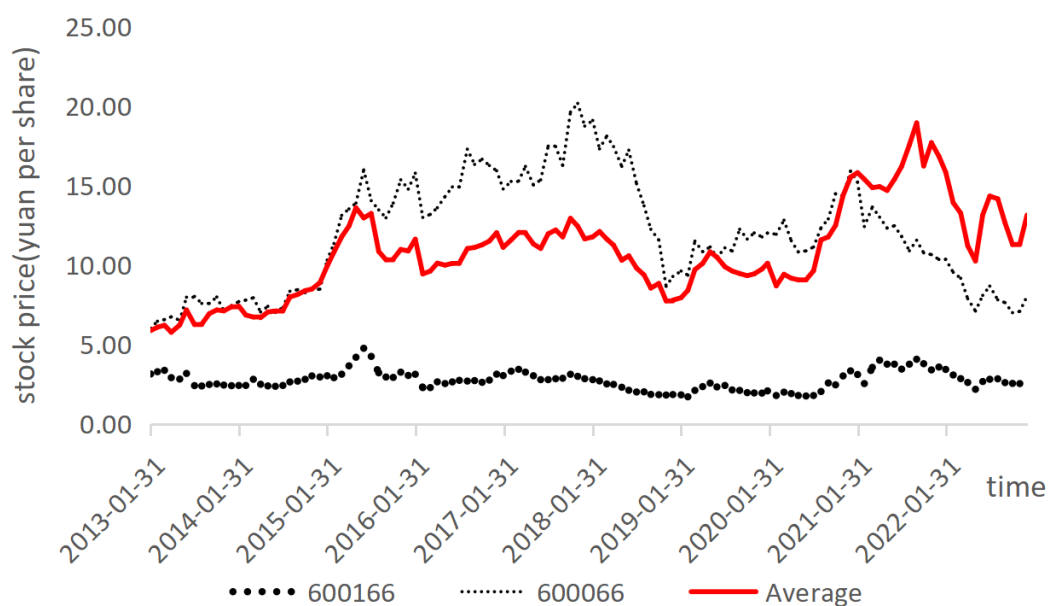


Fig. 3 Automobile industry

Fig. 3 shows the stock prices of the Automobile industry from 2013-1-31 to 2022-1-31, all stock prices rose in 2015 and 2022, especially in 2022, possibly because automation started to gain traction in 2015 and was largely realized in 2022, with prices remaining stable in between.

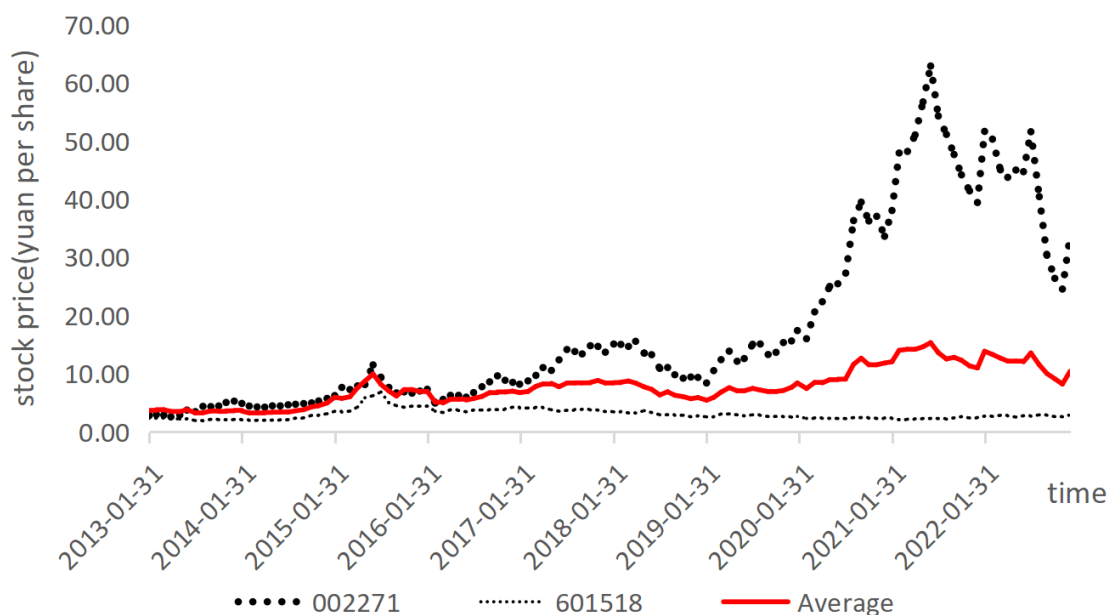


Fig. 4 Building materials

Fig. 4 shows the stock prices of the Building materials industry from 2013-1-31 to 2022-1-31, its stock price rose from 2021 to 2022, and then declined slightly in the later period, possibly because the physical construction industry did not take off due to the pandemic.

Table 2. Descriptive Statistics

Variable	Minimum	Maximum	Mean	Variance
x_1	5.04	14.72	7.0162	3.318
x_2	3.68	11.96	7.0982	4.656
x_4	5.77	18.94	10.7750	8.126
x_4	3.23	15.32	7.7305	9.907

As Table 2. shows, the average data of unit price per share of four industry leading stocks for ten years are respectively Oil and energy industry (x_1), Textile clothing industry (x_2), Automobile manufacturing industry (x_3), Building materials industry (x_4) for descriptive statistics.

As a result, the number of valid cases of the ten year's data of the leading stocks of the four industries is 120. Obviously, the variance of x_4 is about 9.9 at most, indicating that the industry has been unstable in the past ten years. But x_3 has the largest maximum, the largest range, the largest average, and the second variance, compared with the variance of other industries is also unstable. The most stable of the four sectors is oil and energy stocks.

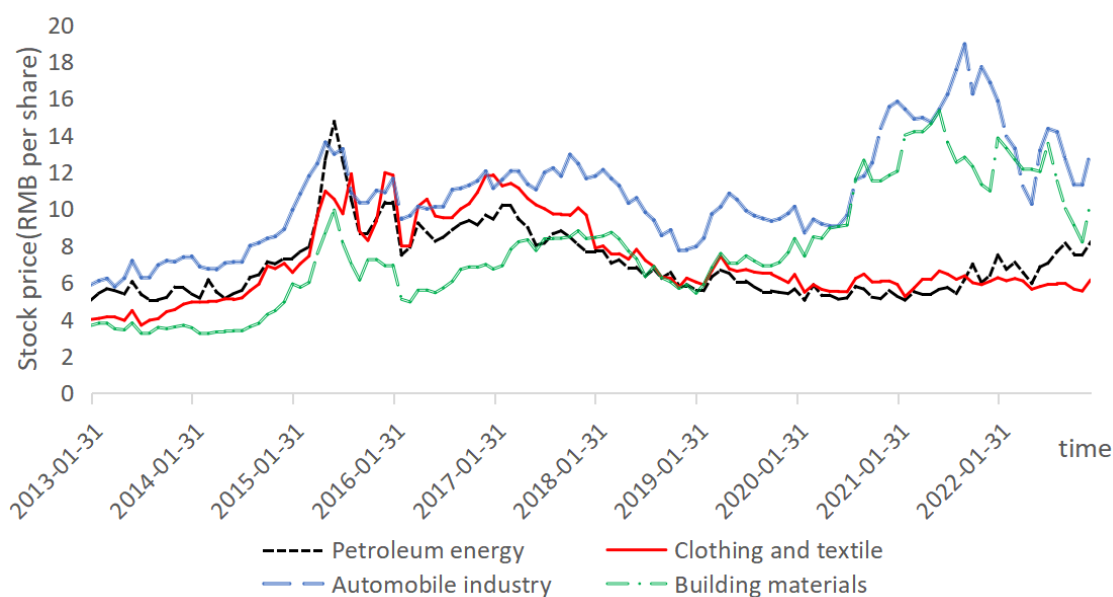


Fig. 5 Overall performance

Fig.5 shows the stock prices of all of industry from 2013-1-31 to 2022-1-31, Stock prices rose in 2015, while Clothing and textile started 2017, and the Petroleum and energy industries began to decline, while the Automobile and Building materials industries began to rise and then declined. Generally speaking, stock prices fluctuated throughout the decade.

First of all, the rising oil price has a great impact on the textile and garment industry chain that uses oil as raw materials. Nylon, polyester and cotton are all synthetic fibers. Synthetic fibers are high polymer polymers synthesized from raw materials obtained from petrochemical industry. Second, the price of oil also has an impact on private car owners. With the increasing demand for automobiles, the rising price of oil may make you more inclined to choose new energy vehicles, so it has a very strong directional impact on the automobile manufacturing industry. Finally, oil prices affect the building materials industry. Many common building materials are composed of petroleum and chemical processing, so the rise in oil prices is equal to the rising prices of raw materials, related to the market price will change.

According to the trend of stock industries in figure 5, the oil industry and textile industry, automobile manufacturing industry and building materials industries may be related. Therefore, linear regression analysis is conducted on the stocks of the two industries respectively.

To begin with, to the oil industry and the correlation of textile apparel industry is analyzed. The rising oil price has a great impact on the textile and garment industry chain which uses oil as raw materials. Nylon, polyester, and cotton are all synthetic fibers. Synthetic fibers are high polymer polymers synthesized from raw materials obtained from petrochemical industry. The average value of oil industry stock was used as independent variable and textile as dependent variable for linear regression. If the ordinary least squares analysis is to be carried out, the homogeneity of variance must be ensured first. Otherwise, when the model has heteroscedasticity problem, the variation degree of the parameter OLS estimates increases, which leads to the decrease of the model accuracy. Therefore, the residual graph is used to determine whether there is heteroscedasticity.

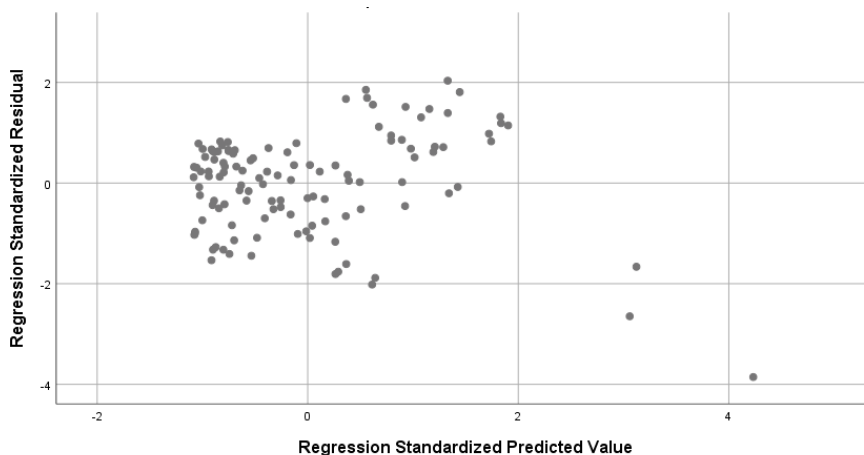


Fig. 6 Scatter plot dependent variable x_2

Heteroscedasticity is generally thought to be due to the magnitude of the residuals varying with the magnitude of the independent variable. Fig. 6 shows that the size of the independent variable has an effect on the residual, so there is a heteroscedasticity problem. The weighted least square method (WLS) is a mathematical optimization technique that weights the original model to make it a new model without heteroscedasticity, and then uses the ordinary least square method to estimate its parameters. The weighted least squares (WLS) method actually weights observations by the inverse of the measured variance in the regression, which reduces the impact of observations with large variance. The following chose to use the weighted least squares method for the analysis.

Assuming the regression equation is $y = ax + b$, the cost function becomes

$$J(a, b) = \frac{1}{N} \sum_{n=1}^N w_i (y_i - ax_i - b)^2 \quad (1)$$

Where w is the weight. The most important thing about WLS is to figure out the weight. By using SPSS software to set variables and scope of power. The independent variable x is weighted to calculate each weight variable within the power range. Set the power range from -1 to 5(as shown in Table 3.) to obtain $m=2.5$ with the maximum logarithmic likelihood value.

Table 3. Log-Likelihood Estimation

Power	Log-Likelihood Values ^a	Power	Log-Likelihood Values
-1.000	-198.939	2.500	-168.864 ^b
-0.500	-191.033	3.000	-169.426
0.000	-184.133	3.500	-170.915
0.500	-178.442	4.000	-173.200
1.000	-174.077	4.500	-176.166
1.500	-171.069	5.000	-179.716
2.000	-169.367	-	-

Note: a. Dependent variable: x_2 , source variable: x_1

b. The corresponding power is selected for further analysis because it maximizes the log-likelihood function.

Then take the inverse of the power of the weight variable, the formula is $\frac{m}{w}$, and calculate the best weight, and use the best weight to carry out weighted regression analysis to obtain Table 4.

Table 4. Coefficients^{a,b}

Model 1	B	Std. error	Beta	t	Sig.
Constant	-0.551	0.419	-	-1.317	0.191
x_1	1.092	0.066	0.838	16.546	0.000

Note: a. Dependent Variable: x_2

b. Weighted Least Squares Regression - Weighted by w

As can be seen from the coefficient estimation results (Table 4), the test sig value of x is far less than the significance level of 0.05. The model is established with the automobile manufacturing industry (x_3) as the dependent variable and the building materials industry (x_4) as the independent variable.

Table 5. Model Summary^b

Model	R	R ²	Adjusted R ²	Std. error
1	0.872 ^a	0.760	0.758	1.401

Note: a. Predictors: (Constant): x_4

b. Dependent Variable: x_3

Linear regression fitting effect as shown in Table 5, determine coefficient of Square (R) is 0.760. The closer the absolute value of the coefficient of determination is to 1, the better the model fits, so the model has a good fitting effect. According to the histogram Figure 7, it can be seen that the normal distribution is consistent with the homogeneity of variance, so this model can be continued.

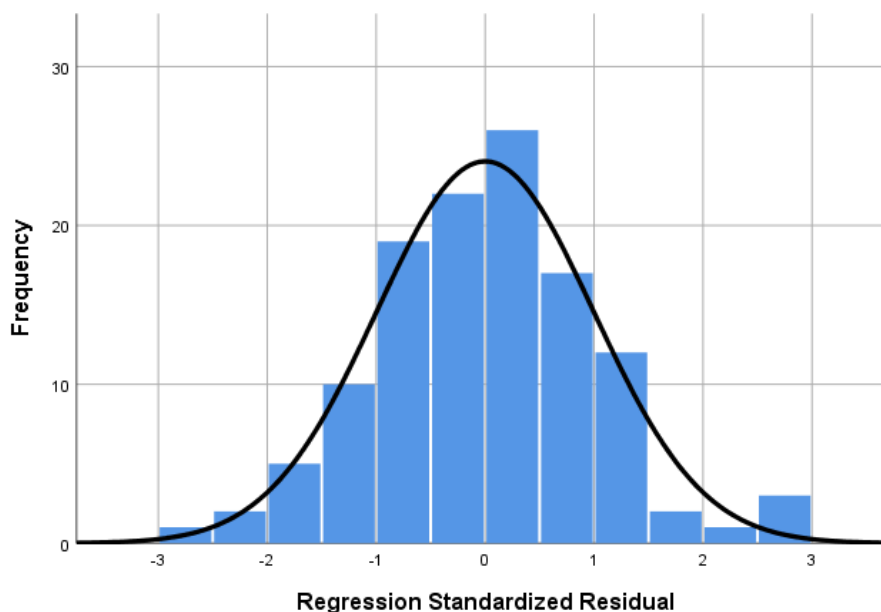


Fig. 7 Histogram Dependent Variable: x_3 , Mean=-1.83E-15, Std. Dev.=0.996

Table 6. Coefficients^{a,b}

Model 1	B	Std. error	Beta	t	Sig.
Constant	4.669	0.340	-	13.719	0.000
x_4	0.790	0.041	0.872	19.357	0.000

Note: a. Dependent Variable: x_3

b. Weighted Least Squares Regression - Weighted by w

Table 6 shows that both the independent variable and the constant term are significant. The basic idea of factor analysis is to find out a few random variables that can control the original variables $F_i (i = 1, 2, \dots, m)$ by analyzing the internal structure of the correlation coefficient matrix of variables m . The principle of selecting common factor is to include as much information in the original variable as possible, establish the model $X = A \cdot F + e$, ignore e , replace X with F , and use it to reproduce the correlation between the various components of the original variable X , so as to simplifying the variable and reduce the dimension.

In the following, factor analysis is carried out for the petroleum and energy industry (x_1), clothing and textile industry (x_2), automobile manufacturing industry (x_3), and building materials manufacturing industry (x_4). The mathematical model of factor analysis can be expressed as:

$$x_p = A_p \cdot m \cdot F_m + e_p \quad (2)$$

where X is a measurable dimensional random vector, each p component of which represents an index or variable. Here $p = 4$.

$F = (F_1, F_2, \dots, F_m)^T$ is an unobserved m -dimensional random vector, as its various components will appear in each variable, so called them public factor. Matrix A is called factor load matrix, and each element in the matrix is called factor load, representing the load of the i^{th} variable on the j^{th} common factor. They need to be estimated by the samples obtained from multiple observations of x .

Table 7. KMO and Bartlett's Test

Items	Value
KMO Measure of Sampling Adequacy	0.515
Bartlett's Test of Sphericity Approx. Chi-Square	358.259
Bartlett's Test of Sphericity df.	6
Bartlett's Test of Sphericity Sig.	0.000

The Kaiser-Meyer-Olkin Measure of Sampling Adequacy in Table 7 is to test whether the data is suitable for factor analysis. However, since stocks are selected as industry representatives for analysis in this study, there are missing values in the data. So the Kaiser-Meyer-Olkin Measure of Sampling Adequacy is 0.515.

According to the research results, only the number of characteristics of the first two factors is greater than 1, and the joint contribution degree reaches more than 90%.

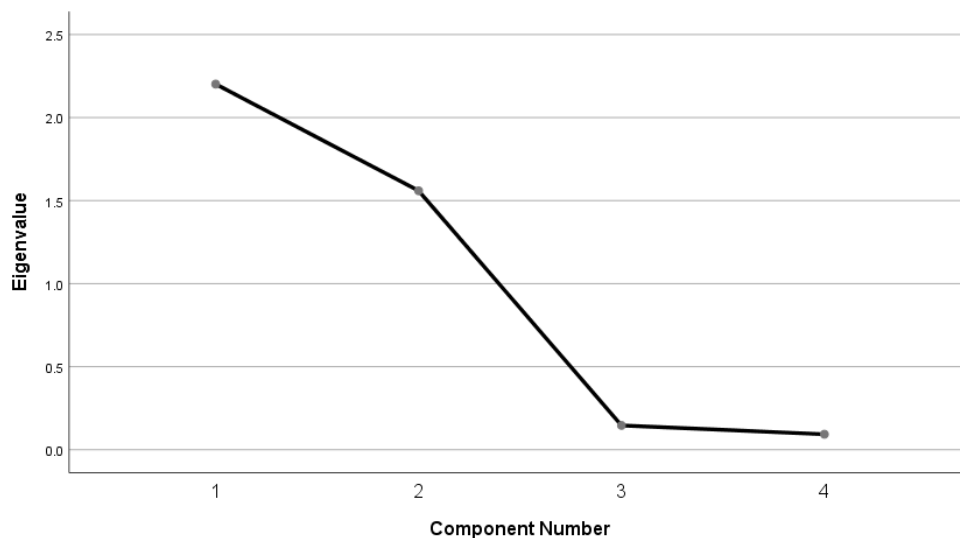


Fig. 8 Scree plot

Fig.8 shows that there are two components whose eigenvalues exceed 1. On the basis of Table 8 is a component transformation matrix, and as little change before rotating.

Table 8. Component Transformation Matrix

Component	1	2
1	0.718	0.696
2	-0.696	0.718

Note: Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

According to Table 8, directly write down the expressions for each common factor. The expression is as follows:

$$F_1 = 0.519x_1 + 0.511x_2 + 0.034x_3 - 0.116x_4 \quad (3)$$

$$F_2 = -0.061x_1 - 0.030x_2 + 0.502x_3 - 0.542x_4 \quad (4)$$

4. Conclusion

This paper studies the relationship between the four industries through the average of the stock prices of the representative leading stocks of the four industries from 2013 to 2022. The first is a brief description of the trends of the representative stocks of each industry. To the comprehensive analysis of the four variables, and then found that four variables, the oil industry and the correlation of textile manufacturing industry is stronger, more contribution in the model. However, due to the collection of stock data, there are many factors changing stock, so the effect of the model is not good. Perhaps because more and more cars use new energy as a power source, and the building materials industry also develops other materials as raw materials due to the rise in oil prices and the increase in raw material costs, so the correlation is not so strong. Then grouped according to the trend graph, points out the two groups of model fitting effect is good, findings may oil a greater effect on clothing industry. The relationship between the automobile industry and construction is also analyzed. Found two industry correlation is stronger, found that auto makers than building enterprise. Some abandoned car factories will be converted into construction factories.

References

- [1] Xiaolei Zhang. Some Statistical Studies on Stock Price Volatility of China's Stock Market -- Analysis of Plate Volatility. Yunnan Normal University, 2008.
- [2] Hu Jianren. Research on the Influence of Industry Characteristics on Enterprise growth and stock price Volatility. Fudan University, 2009.
- [3] Yinghong Zhao. Statistical Analysis of Accounting profit Status of Chinese Listed Companies and its relationship with stock price changes. Liaoning University, 2013.
- [4] Lee K, Ni S, Economics J O M, et al. On the dynamic effects of oil price shocks: a study using industry level data. Journal of Monetary Economics, 2002.
- [5] Alan, A, Carruth, et al. Unemployment Equilibria and Input Prices: Theory and Evidence from the United States. Review of Economics & Statistics, 1998.
- [6] Eksi H I, Senturk M & Yildirim S H. Sensitivity of Stock Market Indices to Oil Prices: Evidence from Manufacturing Sub-Section in Turkey. Panoeconomicus, 2012, 59(4): 463-474.
- [7] Xiaofen Tan, Junxiao Zhang, Yuejia Li. An analysis of international crude oil Price drivers from a broad perspective :2000-2015 -- An Empirical analysis based on the TVP-FAVAR model. China Soft Science, 2015(10): 47-59.
- [8] Wenlan Li. A study on the relationship between crude oil prices, PTA market and stock prices of listed companies in chemical fiber industry. Zhejiang University, 2021.
- [9] M H Yang. Research on the relationship between energy price fluctuation and stock price fluctuation. Chongqing University, 2017.
- [10] J B Wang. Research on Energy prices and stock price volatility of China's stock market and Chemical Industry. Chongqing University, 2017.