

Bigdata Analysis of the Validity of Capital Asset Pricing Model (CAPM) Under COVID-19

Chenxi Jiang^{1, a, †}, Xiang Li^{2, *, †}, Jiahao Pan^{3, b, †} and Yuao Qiao^{4, c, †}

¹Finance, Nanjing University of Finance & Economics, Nanjing, China

²Finance, Shanghai Lixin University of Accounting and Finance, Shanghai China

³Computer Science and Technology, Tianjin University of Technology, Tianjin, China

⁴Mathematics and Applied Mathematics, Shanghai University, Shanghai China

*Corresponding author: 191901221@stu.lixin.edu.cn, ^a285429762@qq.com,

^bpanjiahao2020@163.com, ^c2834316593@qq.com

[†]These authors contributed equally.

Abstract. In 2020, the Chinese stock market has undergone severe turmoil as the COVID-19 swept the world. On this basis, the traditional CAPM model might be unable to explain some of the phenomena in the capital market and there are few empirical studies have focused on exploring relationships between the epidemic and validity of CAPM model. With this in mind, this study aims to investigate the degree to which the COVID-19 impacted on the validity of the CAPM model in different industries. Specifically, this paper selects the monthly closing price data of stocks of 15 listed companies in cyclical and defensive industry in Chinese stock market from January 2018 to December 2021. Based on the least square method and big data techniques, we analyze sample data and construct a regression equation between market return and the daily return of individual stocks or portfolios. According to the analysis, the effectiveness of the CAPM model for enterprises in different industries varies greatly, and the combination of investment can effectively resist risks. These results shed light on quantitative assets pricing model implementation for special issues similar to COVID-19.

Keywords: CAPM; Bigdata analysis; COVID-19.

1. Introduction

Contemporarily, with the rapid development of big data, cloud computing and artificial intelligence information technology, we are in the Internet era of big data. In addition to the traditional news reports more and more popular in the media, we also have been generated data in the online interactive activities. Therefore, human beings are faced with the data size is bigger and bigger. The core of the application of big data is the production, collection, storage, processing and analysis of massive data. Data is essentially a kind of information, and the accumulation and analysis of large-scale information can help us better deal with risks and challenges [1]. The Capital Asset Pricing Model (CAPM) describes the relationship between systemic risk and the expected return of assets especially stocks, it proposed by the economist William Sharpe, who won the Nobel Prize in 1990, in his 1970 book named Portfolio Theory and Capital Markets. As one of the most important parts of modern finance theory, asset pricing model has attracted the close attention of many experts, scholars and market investors since it was put forward. Investors in the stock market hope to predict stock prices more accurately through this method, so as to obtain greater returns. The traditional CAPM model points out that the returns of individual stocks are only related to the systemic risk (also known as market risk) of the whole market. However, as more and more scholars through empirical analysis show that there are a large number of phenomena in the market that the actual return and CAPM pricing are different, and a single systemic risk cannot fully measure the average return of stocks.

As the COVID-19 swept the world, China's stock market has undergone extremely violent turmoil [2], and there are more and more phenomena in the capital market that cannot be explained by traditional CAPM models. In this case, it is urgent to find a research method to verify the effectiveness of traditional CAPM models in the context of the global epidemic. This paper attempts to apply the

Internet big data technology to verify the effectiveness of the CAPM model by analyzing the monthly closing price data of 15 representative stocks of cyclical and defensive stocks and providing suggestions for investors to construct their own reasonable investment strategies and evaluation systems [3]. Investors can consider other factors based on the classic CAPM model in the context of the epidemic to better respond to major public health events such as COVID-19.

The capital asset pricing model was proposed by Sharpe (1964), which divides the expected rate of return of an asset into the sum of two parts. One is the risk-free rate, and the other is the part that represents the market factors that are connected, (the risk premium). According to the theoretical model, the capital market line and the securities market line are constructed, which reveal the connection between the effective asset portfolio and risk, as well as measure the linear relationship between systemic risk and the expected return.

Because the CAPM model is established on the basis of the MarkWitz model, the assumptions of the MarkWitz model should be referred to when studying the CAPM model:

- When investors invest in two stocks with the same risk level, choose the one with a large investment return, and when investing in two stocks with the same return, choose the one with a small risk level.

- The more assets an investor has.

- Investors can analyze that the probability distribution of investors' returns is normally distributed.

- Choose the variance (or standard deviation) of the rate of return as a representation when measuring the magnitude of the investment risk.

- Expected rate of return and investment risk are the two indicators that are prioritized in the investment strategy.

The premise of the CAPM model is given as follows:

- There is no inflation in the capital market, and the discount rate remains unchanged.

- Investors can borrow funds without restriction.

- Other expenses incurred in the investment are negligible (including tax charges).

- The information on the capital market can be shared with every participant timely and freely.

- Investors have the same investment expectations.

- All investors participating in the market have the same investment period [4].

Black argued that if the market portfolio was efficient, there would be a linear positive correlation between the β value and the expected rate of return [5]. Roll stated that there was a loophole in the tests conducted at that time: none of them could prove the validity of the market portfolio, thus the CAPM was tested on an unreliable basis [6]. Banz added the factor of company size to the CAPM test, and criticized the CAPM. It is found that after a certain degree of risk adjustment, those small market capitalization companies have a higher average stock yield than large market capitalization companies, and the company's equity scale effect is also discovered [7].

As a matter of fact, the impact of the epidemic on China's stock market is significant, which also affects investors and participants in the capital market. On this basis, the effectiveness of the CAPM model in some industries in COVID-19 provides a reference for enterprises to reduce risks and improve competitiveness in the post-epidemic era.

The rest part of the paper is organized as follows. The Sec. II will introduce the selection of samples and the analysis methods. Subsequently, the Sec. III, will present the empirical results. Afterwards, the Sec. IV, will analyze the reason for the results as well as demonstrate the limitations. Eventually, a brief summary will be given in Sec. V.

2. Methodology

2.1 Data

The sample data of this essay all come from the JoinQuant platform, and the monthly closing price data of 15 enterprises in China's stock market are selected as our research object. Due to the outbreak of COVID-19 in year 2020, China's stock market has been affected seriously. This essay aims at

studying the change of the validity of CAPM model in enterprises before and after the COVID-19 as well as the financial anomalies caused by the epidemic. Therefore, the sample interval selected is 720 samples in all from January 2018 to December 2019 and January 2020 to December 2021. The detail information of the selection is given in Table. 1. Monthly closing prices of 15 cyclical and defensive industries are summarized in Table. 2.

Table 1. Industry code and number of enterprises

Industry type	Industry code	Industry	Number of enterprises
Cyclical industry	A	Financial sector	3
	B	Tourism	3
	C	Non-ferrous Metal	3
	D	pharmaceutical industry	3
Defensive industry	E	Utilities	3

Table 2. Related information about the stocks of 15 enterprises

Industry code	Stock code	Individual stock	Industry code	Stock code	Individual stock
A	600036	China Merchants Bank	D	600196	Fosun Pharmaceutical
	601211	Guotai Junan Securities		600276	Hengrui Medicine
	601318	PingAn Insurance		600380	Health Yuan
	600138	China Youth Travel Service			
B	600754	Jin Jiang Hotels	E	600681	Baichuan Energy
	603136	Tianmu Lake		600886	SDIC Power Holdings
	600362	Jiangxi Copper Corporation		601199	Jiangnan Water Affairs
C	600392	Shenghe Resources			
	603799	Huayou Cobalt Co.,ltd			

$E(R_i)$ was obtained by two methods separately. The former one was simple arithmetic average, which meant that the rate of return of portfolio was calculated by averaging the monthly return rate of all the 15 stocks, while the latter one was obtained by fixed weight, which meant that the rate of return of portfolio was calculated by weighted average method [8]. According to this method, the weight of individual stocks in the monthly rate of return of the portfolio was the ratio of the arithmetic average of the total market capitalization of each stock in 24 months to the arithmetic average of the total market capitalization of the portfolio in 24 months.

We used the calculation of the simple rate of return. Assuming that the investor owns a certain asset in a given cycle, from day $t-1$ to day t , the rate of return is calculated as shown following:

$$E(R_{i,t}) = r_t = \ln \frac{P_t}{P_{t-1}} \quad (1)$$

Where P_{t-1} is the closing price of this asset in month $t-1$, P_t is the closing price of this asset in month t and $E(R_{i,t})$ is the rate of return of the i^{th} stock in the t^{th} month [3]. In addition, we chose the

one-year Treasury bond rate as the risk-free rate R_f and selected the monthly closing price of the Shanghai (securities) composite index (SSEC) to calculate the total market return $E(R_m)$.

2.2 Evaluation methods

A linear regression equation is constructed between market returns and individual stock return or portfolio returns. The data were analyzed using ordinary least squares. The model equation is shown below:

$$E(R_i) - R_f = \beta_i[E(R_m) - R_f] + \varepsilon_i \quad (2)$$

Where $E(R_i)$ is the expected rate of return of stock i or portfolio, $E(R_m)$ is the rate of return of the market portfolio, R_f is the risk-free interest rate, and β_i depicts the sensitivity of stock i or portfolio to the market risk and ε_i is a constant including residuals and random disturbance terms.

Table 3. Model validity testing of cyclical industry

Industry	Individual stock	COVID-19	R^2	P_0	β	P_1
Financial sector	China Merchants Bank	0	0.722	2.85E-07	1.0818	0.000
		1	0.334	0.00312	1.098	0.003
	Guotai Junan Securities	0	0.79	1.44E-08	1.2949	0.000
		1	0.512	8.52E-05	1.1338	0.000
	PingAn Insurance	0	0.583	2.25E-05	0.7939	0.000
		1	0.341	0.00272	1.1533	0.003
Non-ferrous Metal	Jiangxi Copper Corporation	0	0.816	3.58E-09	1.1574	0.000
		1	0.143	0.0686	1.5718	0.069
	Shenghe Resources	0	0.289	0.00819	0.9662	0.008
		1	0.017	0.54	0.7324	0.540
	Huayou Cobalt Co.,ltd	0	0.416	0.000891	1.8475	0.001
		1	0.150	0.0619	2.0148	0.062
Tourism	China Youth Travel Service	0	0.656	2.87E-0.6	1.2606	0.000
		1	0.306	0.00509	1.4188	0.005
	Jin Jiang Hotels	0	0.381	0.00169	0.9655	0.002
		1	0.524	6.44E-05	2.6064	0.000
	Tianmu Lake	0	0.520	0.000103	1.2202	0.000
		1	0.138	0.00489	2.0919	0.005

Table 4. Model validity testing of defensive industry

Industry	Individual stock	COVID-19	R^2	P_0	β	P_1
Pharmaceutical Industry	Fosun Pharmaceutical	0	0.345	0.0032	0.9928	0.003
		1	0.077	0.189	1.4591	0.189
	Hengrui Medicine	0	0.248	0.0155	0.7324	0.015
		1	0.086	0.163	0.7994	0.163
	Health Yuan	0	0.317	0.00515	0.9162	0.005
		1	0.263	0.0104	1.42	0.010
Utilities	Baichuan Energy	0	0.489	0.000203	0.8801	0.000
		1	0.209	0.0247	1.0209	0.025
	SDIC Power Holdings	0	0.406	0.00107	0.8215	0.001
		1	0.077	0.19	0.7423	0.190
	Jiangnan Water Affairs	0	0.889	1.68E-11	0.9823	0.000
		1	0.146	0.0651	1.0775	0.065

3. Results

3.1 Empirical analysis

As listed in Tables III and IV, the goodness of fit of the stocks of 15 enterprises randomly selected are totally different, and the goodness of fit also show great changes before and after the COVID-19. Among them, 46.67% had goodness of fit greater than 0.5 before the COVID-19, 13.33% had goodness of fit greater than 0.3, while 13.33% had goodness of fit less than 0.3. After the COVID-19, 13.33% had goodness of fit greater than 0.5, 26.67% had goodness of fit greater than 0.3 and 60.00% had goodness of fit less than 0.3.

Table 5. Model validity testing of portfolio

Method	COVID-19	R^2	P_0	β	P_1
Simple Arithmetic	0	0.826	1.95E-09	1.0608	0.000
Average	1	0.817	1.46E-09	1.356	1.000
Fixed	0	0.85	4.04E-10	0.9842	0.000
Weight	1	0.513	8.24E-05	1.2643	0.000

3.2 Data analysis

According to the Eq. (2), $E(R_i)$ is the expected monthly rate of return of the i^{th} stock, $E(R_m)$ is the rate of return of portfolio, and R_f is the rate of return of risk-free. Through regressive analysis of the above formula, we can get the β_i , α_i , ε_i and R_i^2 of 15 stocks.

Above all, it could be seen from the data in Table III. and Table IV., R_i^2 of individual stocks in the industry had significant differences before and after the epidemic. The three greatest R_i^2 enterprises were Jiangnan Water Affairs, Jiangxi Copper Corporation, and Guotai Junan Securities before the COVID-19, while the three greatest R_i^2 enterprises were Jinjiang Hotels, Guotai Junan Securities and PingAn Insurance after the COVID-19. Besides, R_i^2 of representative enterprises in cyclical industry were generally greater than the R_i^2 of representative enterprises in defensive industry whether before or after the COVID-19. As vividly shown in Figs 1. and 2, Jiangxi Copper Corporation in non-ferrous metal industry and Jiangnan Water Affairs in utilities showed extreme differences before and after the COVID-19, with Jiangxi Copper Corporation as high as 0.673, and Jiangnan Water Affairs as high as 0.743.

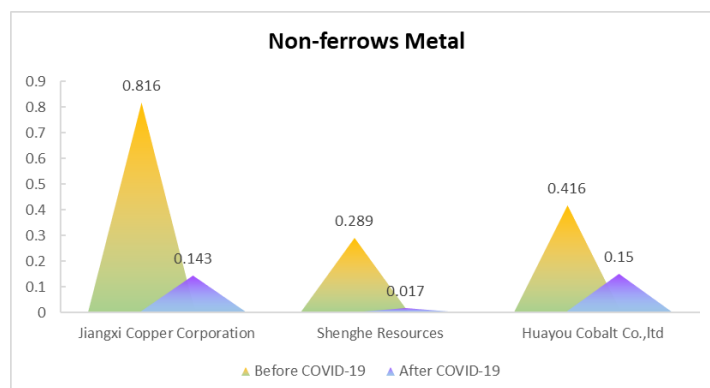


Fig. 1 Changes in the coefficient of determination for the three representative companies in the non-ferrous metals sector before and after the COVID-19.

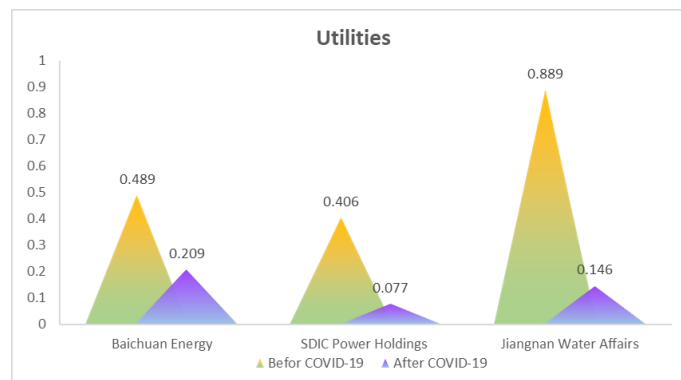


Fig. 2 Changes in the coefficient of determination for the three representative companies in utility sector before and after the COVID-19.

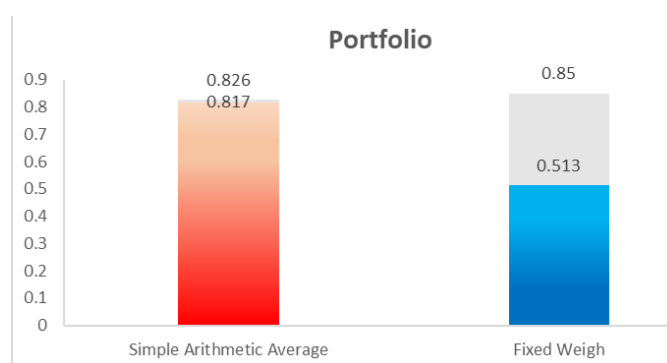


Fig. 3 Changes in coefficient of determination of portfolio before and after the COVID-19.

For the portfolio, as can be seen from the Fig. 3, the R_i^2 obtained from the regression analysis reached above 0.8 both before and after COVID-19 when the returns were calculated using the simple arithmetic mean method. In contrast, the R_i^2 obtained by the regression analysis of the rate of return calculated by the fixed weight method differed somewhat before and after the epidemic, with R_i^2 after the epidemic being smaller than R_i^2 before the COVID-19. In addition, the β values in the linear equations derived from both methods increased after the epidemic.

3.3 Significance

In the third column of the Tables. III-V, 0 means that the industry was not affected by the epidemic and 1 means that it was affected by the epidemic. The larger the coefficient of determination R_i^2 , the better the goodness of fit of the sample. Additionally, under the 95% confidence level, the smaller the p-value (<0.05) corresponding to the F-statistic for each group, the more significant the equation. Besides, the beta coefficient (β) portrays the volatility of expected returns and should normally be greater than 1 (i.e., the beta of the market a

To sum up, the regression statistics of the above values are important in drawing conclusions, which serve as the main reference tool for comparing the validity of the CAPM model in different industries in the context of the epidemic.

4. Discussion

4.1 Interpretation of the results

After regressive analysis, the goodness of fitting showed significant differences among stocks in various industries. For individual industry, the financial sector in the cyclical industry had a seemingly better goodness of fit compared with other industries whether before or after the epidemic, and simultaneously, the utilities in the defensive industry had a seemingly better goodness of fit compared

with other industries whether before or after the epidemic. For individual stocks, Guotai Junan Securities had a high goodness of fit before and after the epidemic, while Shenghe Resources, Fosun Pharmaceutical and Hengrui Medicine had a relatively low goodness of fit. R_i^2 of regression equation before the epidemic was greater than R_i^2 of regression equation after the epidemic, indicating that the goodness of fit of the equation before the epidemic was better than that after the epidemic. In addition, the value of β of some cyclical industries increased dramatically after the epidemic, e.g., Jinjiang Hotels and Tianmu Lake, indicating that the epidemic had increased the sensitivity of these enterprises to the system risk.

The determination coefficient R_i^2 of the defensive enterprise regression equation is generally less than the determination coefficient R_i^2 of the periodic regression equation. It indicates that the goodness of fit of the equation in the periodic industry is better than that in the defensive industry. The determination coefficient R_i^2 of the same portfolio (whether the data obtained by regression analysis based on the simple arithmetic average method or the data obtained by regression analysis based on the fixed weight method) is greater than that of most single enterprise regression equations before and after the epidemic. It shows that the goodness of fit of investment portfolio is obviously higher than that of single enterprise. The β value in the regression equation is generally greater than 1, which indicates that the market risk of assets is greater than the systemic risk of the entire stock market.

4.2 Analysis of causes

Non-ferrous metal industry such as Jiangxi Copper Group, is a large-scale joint enterprise integrating copper mining, selection, smelting and addition in China's non-ferrous metal industry, and is China's largest copper product production base and an important sulfur chemical raw material and gold, silver and rare metal production area [9]. Jiangxi Copper industry before the epidemic is 0.816, after the epidemic is 0.143, the company due to the impact of the new crown epidemic caused by large-scale shutdown in 2020, the company's performance declined seriously. In 2021, under the superposition of triple positives such as the supply cycle of copper mines, the over-issuance of US dollar currency, and the stimulation of the 1.9 trillion US dollars, copper prices rose sharply, and the self-sufficiency rate of copper mines in Jiangxi Copper was high (up to 15.3%), and the performance also highly benefited from the rise in copper prices, but the current valuation of the company's price-to-book ratio was still below the historical median level [10].

Utilities such as Jiangnan Water, the company mainly engaged in tap water production and sales, water supply engineering design. The impact of the COVID-19 on the sewage treatment industry is more negative, although the country urgently issued a series of support policies, but the impact of the COVID-19 on water affairs is not only short-term but also long-term impact, relevant agencies on water quality requirements will be higher and higher, before the epidemic R_i^2 is 0.889, after the epidemic is 0.146.

4.3 Limitation

In reality, this paper does have some drawbacks and limitations, where the primary one is the data quantity since we only selected 15 stocks in the analysis. Besides, the time period of the data only spans for four years, which might lead to overfitting, as the test time span may be short and there are occasional fluctuations of individual stocks. In addition, the risk-free interest rate chose the interest rate of one-year Treasury bonds, but China's interest rate has not been fully marketized, and the majority of long-term Treasury bonds may lead to errors in the test results. Finally, the test method is relatively simple, mainly adopted the least square method, regression analysis.

5. Conclusion

In summary, we set out to measure the ability of firms to withstand epidemic shocks by analyzing the change in the effectiveness of the CAPM model for different enterprises after the epidemic. Based

on the analysis, the goodness of fit varies across the 15 firms indicates that the validity of the CAPM model varies considerably across firms in different industries. On the contrary to expectations, the CAPM model was significantly more effective in cyclical sectors than defensive sectors both before and after COVID-19. Additionally, from the perspective of model validity, Jiangxi Copper Corporation and Jiangnan Water Affairs were the most affected by the epidemic, as evidenced by the significant reduction in model validity. As for the portfolio consisting of the 15 stocks we selected, the regression results for its sample data show that the pre-epidemic β is smaller than the post-epidemic β , indicating an increase in market risk for the portfolio after the epidemic. Finally, compared to the goodness of fit of a single firm, the fit of the portfolio is much better, demonstrating that investing in a portfolio can be effective in protecting against risk to some extent. Nevertheless, with a small sample size, caution must be applied, as the findings might not be transferable to the researches based on daily data. The increasingly significant background of the normalization of the epidemic has posed greater challenges to investors and to our stock market. Overall, these results will offer a guideline for future research on China's stock market in the context of the epidemic, and may also provide some insights for relevant policy makers and investors.

References

- [1] S.Y. Pan, "Big data, machine learning and asset pricing," *Modern management science*, vol.1, 2019, pp. 6-8.
- [2] Y.N. Li, "The impact of media information disclosure of COVID-19 on the stock prices of China's pharmaceutical sector," *Dissertation of Harbin Institute of Technology*, 2021.
- [3] W. J. Wang, S. Tao, J. Y. Li, W. B. Hou, "An empirical test of the validity of Capital asset pricing model for Chinese enterprises," *Journal of Huaibei Normal University (Natural Science edition)*, vol. 42, 2021, pp.23-31.
- [4] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *The Journal of Finance*, vol. 19(3), 1964, pp. 425-442.
- [5] F. Black, "Capital market equilibrium with restricted borrowing," *The Journal of business*, vol. 45(3), 1972, pp. 444-455.
- [6] R. Roll, "A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory," *Journal of financial economics*, vol. 4(2), 1977, pp. 129-176.
- [7] R. W. Banz, "The relationship between return and market value of common stocks," *Journal of financial economics* vol. 9(1), 1981, pp. 3-18.
- [8] D. Chen, W. Min and Z. Fang, "An empirical test of CAPM model on China's capital market," *Investment and Entrepreneurship*, 2021, vol. (31), pp.36-38.
- [9] X. C. Wu, "Evolution of global copper industry chain pattern and Strategic choice of Jiangxi province," *Dissertation of Jiangxi University of Finance and Economics*, 2021.
- [10] S. Xu "Sheng He Resources: overseas rare earth smelting contribution increment," *Daily Business News*, vol.34, 2021.