An Empirical Study of Improved Fama-French Five Factor Model Based on Stata in China's Pharmaceutical Manufacturing Industry under Big Data

Yan Zhang[1,a], Xinshuang Ma[1,b]

32218605@qq.com[a], 571941517@qq.com[b]

Hebei University of Science & Technology, No.26, Yuxiang Street, Yuhua District, Shijiazhuang, China[1]

Abstract: Since the occurrence of the new crown epidemic, all industries in China have received a huge impact, which has also continuously stimulated the rapid development of China's pharmaceutical manufacturing industry, in the context of the big data era this paper selected 110 pharmaceutical manufacturing companies in China, a total of 72 monthly data from 2015 to December 2021, and empirical analysis through the Fama-French five-factor model to study the excess return of impact factors. Then, a six-factor model is constructed by adding the sentiment factor to the Fama-French five-factor model and conducting regression tests to observe the effect of sentiment factor on the excess return of pharmaceutical manufacturing stocks, and the empirical process is implemented by Stata analysis software as well as Spss analysis software.

Keywords: Big Data, Fama-French five-factor model, pharmaceutical manufacturing, investor sentiment, linear regression

1. Introduction

The new crown epidemic, as a national event, affects each of us, which also brings opportunities for the development of pharmaceutical manufacturing industry, and some companies' performance has changed to different degrees, which makes it of great practical significance to revalue China's pharmaceutical manufacturing industry. This paper first tests the applicability of the Fama-French five-factor model in the sample interval scenario of pharmaceutical manufacturing industry in China's A-share market. For the purpose of the explanatory strength of the model on the excess return, this paper draws on the relevant research of domestic scholars on China's capital market and synthesizes the sentiment factors based on the sentiment indicators of pharmaceutical manufacturing industry stocks, such as the turnover ratio, price-earnings ratio, number of shareholders and institutional shareholding ratio, through principal component analysis, and adds them to the Fama-French five-factor model to form a six-factor model to investigate whether the six-factor model has an explanatory strength on China's pharmaceutical The six-factor model is used to investigate whether the explanatory power of the six-factor model on the stock market of China's pharmaceutical manufacturing industry will be enhanced.

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China
Copyright © 2023 EAI
DOI 10.4108/eai.19-5-2023.2334285

2. Literature Review

Zhibing Li found that there is no significant momentum effect and reversal effect in China's stock market, and the Fama-French five-factor model is better than the three-factor model as well as the CAPM model [1]. Xindong Zhang studied the monthly data of stocks in China's A-share market from 1999-2017 and found that there is a significant earnings effect in the market, but the investment factor does not have a significant effect on the excess return of stocks [2]. By verifying the validity of the Fama-French five-factor model in the steel industry in China's A-share market, Zhuxi Guo found that the earnings and investment factors failed the redundancy test and the applicability of the Fama-French five-factor model was poor [3]. Ziyi Li selected 85 listed companies in the 5G-related industry chain in China's market for the study, and found that the applicability of the Fama-French five-factor model was lower than that of the three-factor model [4].

Some scholars have added the investment sentiment factor as a variable in a multi-factor model to investigate the strength of the sentiment factor in explaining stock excess returns.Miller argues that under the constraints of heterogeneous beliefs and short selling, stock prices mainly reflect the expectations of optimistic investors and therefore stock prices will be higher than their intrinsic value [5].Festinger proposes a weakened cognitive imbalance, which means that when there are mutually contradictory views, people ignore perceptions that are less relevant to the prevailing view, which leads to investors being generally influenced by mainstream information [6]. lee uses new ways to obtain data and measure the magnitude of investment sentiment with them, which include data on the number of IPOs of new issues in the market, stock trading volume energy, etc [7]. The most influential measure of investment sentiment is the method proposed by Baker and Wurgler, who used principal component analysis to construct a new synthetic sentiment indicator and used it in a pricing model, which significantly improved the explanatory power [8]. There is a positive correlation between the excess return of the stock [9].

Qianqian Liu constructed the sentiment factor by extracting the principal components of six original indicators, and used the standard deviation of the sentiment factor plus or minus one times as the boundary to divide the sentiment factor into two parts: high-risk area and low-risk area, and considered it suitable for investment in the low-risk area. The portfolio of stock selection based on investment sentiment eventually performed well [10]. After constructing the investor opinion factor, Tian Faxiang used the double-similar ranking method to divide the sample into 5*5 portfolios according to market capitalization and investment sentiment, and conducted a multi-factor model regression, and the regression results showed that the more positive the investment sentiment, the higher the excess return [11].

3. Sample selection and model construction

3.1. Selection of the sample

Definition of the Fama-French five-factor model:

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{it} \quad (1)$$

where R_{it} denotes the return of the constructed portfolio i in period t, R_{Ft} denotes the market risk-free rate, R_{Mt} denotes the return of the composite market in period t weighted by market capitalization, (R_{Mt}-R_{Ft}) denotes the excess return of the composite market in period t weighted by market capitalization, SMB_t denotes the size factor, i.e., the difference between the return of a high size portfolio and a low size portfolio in period t. SMB_t is the size factor, which is the difference between the return of a high size portfolio and a low size portfolio in period t. HML_t is the book-to-market ratio factor, which is the difference between the return of a high book-to-market ratio portfolio and a low book-to-market ratio portfolio in period t.RMW_t is the profit factor, which is the difference between the return of the high profit level portfolio and the low profit level portfolio CMA_t is the investment factor, which is the difference between the return of a high investment level portfolio and a low investment level portfolio. In this paper, the above factors are constructed using the 2*3 grouping method adopted by Fama and French.

3.2. Emotional factor construction

Before constructing a sentiment factor, an indicator that can measure investor sentiment needs to be constructed. A popular method for quantifying sentiment in academia is the principal component analysis method used by Baker and Wurgler (2006). The idea of constructing a sentiment indicator by principal component analysis is to extract and synthesize a series of indicators that can reflect the sentiment of individual stocks into a new indicator, and use the synthesized new indicator to measure the sentiment of individual stocks. The data selected for this paper are the monthly turnover rate of individual stocks, the number of shareholders of individual stocks, the percentage of institutional holdings of individual stocks, and the price-to-earnings ratio of individual stocks.

The KMO of the principal component analysis method used in this paper took a value of 0.732, which indicates that the original variables are suitable for principal component analysis. We first processed the four original indicators and extracted the principal factors by the principal component method. In this paper, we consider the explanatory strength of individual stock sentiment and the monthly consistency of the indicators, and extract the indicator with the strongest explanatory strength of individual stock sentiment, the monthly turnover rate, as a sentiment indicator to measure the investment sentiment of individual stocks in each monthly period. Table 1 shows the explanatory strength of each original indicator for stock sentiment.

Table 1 Explanation table of KMO values

Indicator name	Explanatory strength	accumulative total
Monthly turnover rate	47.267%	47.267%
Number of shareholders	25.181%	72.448%
Institutional shareholding ratio	16.726%	89.174%
P/E ratio	10.826%	100%

The construction method of the sentiment factor SEN maintains consistency with the way other factors are constructed by using a 2*3 grouping and subtracting the weighted returns of portfolios with high investment sentiment from the weighted returns of portfolios with low investment sentiment.

4. Descriptive tests and correlation tests

Before conducting the empirical analysis, the factors were tested to determine whether each factor was suitable for the Fama-French five-factor model analysis. The following are the factor descriptive statistics.

Table 2 Descriptive Statistics

	average value	variance	min	median	max
MKT	0.321	5.345	-25.064	0.468	15.391
SMB	-1.052	3.469	-10.691	-1.094	7.765
HML	-2.187	4.190	-15.382	-1.367	8.359
RMW	1.274	3.963	-12.068	1.471	10.687
CMA	-1.239	3.293	-12.009	-1.199	6.625

Using the descriptive statistics between the factors in Table 2, we find that companies with large size, low book-to-market ratio and high earnings in China's pharmaceutical manufacturing industry also have higher excess returns. At the same time, when the market environment is good, the returns perform even better. The reason for this is because in our stock market, pharmaceutical manufacturing has been a popular industry, and the attention of institutional investors and general investors has been high, and institutional investors also select stocks with larger size and room for growth for allocation, which causes leading stocks in pharmaceutical manufacturing to be more popular.

After the descriptive tests, we performed inter-factor correlation tests as well as redundancy tests.

Table 3 Table of correlation coefficients between factors

	MKT	SMB	HML	RMW	CMA
MKT	1.000	-0.002	-0.023	-0.217*	0.380*
SMB	-0.002	1.000	0.074	-0.292*	0.360*
HML	-0.023	0.074	1.000	-0.167	0.164
RMW	-0.217*	-0.292*	-0.167	1.000	-0.438**
CMA	0.380*	0.360*	0.164	-0.438**	1.000

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Through Table 3, it can be found that the correlation coefficients between the factors are not high and the overall significance is not strong. The market factor is negatively correlated with the size factor, book-to-market ratio factor and earnings factor, and positively correlated with the investment factor. The market factor is significantly correlated with the earnings factor and the investment factor at the 10% level. The size factor is positively correlated with the book-to-market ratio factor and the investment factor, and negatively correlated with the earnings factor, where the size factor is significantly correlated with the earnings factor and the investment factor at the 10% level. The book-to-market ratio factor is negatively correlated with the earnings factor and positively correlated with the investment factor. The earnings factor is negatively correlated with the investment factor, and the two are significantly correlated at the 5% level.

Table 4 Five-factor redundancy test

	MKT	SMB	HML	RMW	CMA
Intercept term	0.752*	-0.424**	-1.793***	0.277*	-0.495**
t-value	(1.078)	(-0.910)	(-2.849)	(0.543)	(-1.363)
Adj R^2	0.126	0.117	0.013	0.267	0.389

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In Table 4, the redundancy is tested that the fifth factor is explained by the remaining four factors, the market factor, size factor, book-to-market ratio factor, earnings factor and investment factor are regressed with the remaining four factors respectively, the intercept terms are all significantly non-zero at different levels and the adjusted R2 are all small this indicates that the five factors cannot be explained by the remaining four factors and all pass the redundancy test.

5. Empirical Analysis

5.1. Results of multiple linear regression based on the Fama-French five-factor model

The following regression is based on the Stata software using the Fama-French five-factor model for China's pharmaceutical manufacturing industry. First, the 110 sample companies are divided into five groups according to company size, book-to-market ratio, profitability and investment level, and the regression analysis is conducted separately for 25 size-to-book-to-market ratio portfolios, 25 size-to-earnings portfolios and 25 size-to-investment for a total of 75 portfolios according to the 5*5 grouping method.

Table 5 Five-factor model linear regression results

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + {}_{\mathrm{Si}}SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{it}$$

Size	Small	2	3	4	Big
B/M			HML		
Low	-1.656***	0.042	-0.135	0.042	0.031
2	-1.753***	-0.343	-0.291**	0.173	0.036

3	-0.940**	-0.287	0.034	0.010	0.111
4	-0.881***	-0.055	-0.269	0.243	0.105
High	-0.498***	-0.136	0.036	-0.572*	0.080
Op			RMW		
Low	-0.166	0.173	0.487**	0.696**	0.392
2	0.030	-0.151	0.259	0.194	1.072***
3	0.212	-0.298	0.319*	1.026**	0.770**
4	0.067	0.163	-0.130	0.765***	0.558***
High	-0.267	-1.267	0.522***	0.552**	0.522***
Inv			CMA		
Low	-0.059	0.143	0.045	0.756	0.115
2	0.277	-0.357	0.023	0.090	0.719***
3	0.265	-0.310	-0.118	0.792	0.322**
4	0.969**	0.111	0.736	0.332	0.107*
High	0.050	-0.673	0.464	0.321	-0.514***

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In Table 5, there are 7 significant non-zero regression coefficients for the book-to-market ratio, 11 significant non-zero regression coefficients for the size-to-earnings portfolio, and 5 significant non-zero regression coefficients for the investment factors in the size-to-investment grouping regression results. the significance of the coefficients of the factors in the model needs further improvement.

5.2. Adding emotional factors to construct a six-factor model

The goodness of fit of the Fama-French five-factor model in China's pharmaceutical manufacturing industry still needs to be improved. In order to better explain the excess returns, a six-factor model is constructed by adding the sentiment factor to the Fama-French five-factor model.

$$R_{it}-R_{FT}=a_i+b_i（R_{MT}-R_{FT}）+s_iSMB_t+h_iHML_t+r_iRMW_t+c_iCMA_t+n_iSEN_t+e_{it} \quad (2)$$

Because of the new construction of the sentiment factor, it is necessary to add the sentiment factor to the original scale factor construction method for arithmetic averaging when calculating the scale factor.

In the following, the newly constructed six-factor model is first tested for redundancy to explore the validity of the newly constructed sentiment factor.

Table 6 Six-factor redundancy test

	MKT	SMB	HML	RMW	CMA	SEN
Intercept term	0.336*	-0.678*	-0.873**	0.328*	-0.116*	-0.701**
t-value	(0.536)	(-1.755)	(-2.037)	(0.884)	(-0.297)	(-1.188)

Adj R^2	0.124	0.282	0.170	0.309	0.314	0.106

* p < 0.1, ** p < 0.05, *** p < 0.01

Observing Table 6, in the redundancy test, it was found that the intercept terms after the regression of each factor were significantly non-zero and the goodness of fit was low, indicating that each factor could not be replaced by the other five factors, and the newly constructed six-factor model passed the redundancy test.

In the following, a six-factor model with the addition of sentiment factor is regressed on China's pharmaceutical manufacturing industry, and the sample is grouped according to the 5*5 grouping method above to obtain 25 size-to-book-to-market value portfolios, 25 size-to-earnings portfolios, 25 size-to-investment and 25 size-to-sentiment portfolios for a total of 100 portfolios and regress them separately.

Table 7 Linear regression results of the six-factor model

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + \mathrm{Si}\,SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + \mathrm{ni}SEN + e_{it}$$

Size	Small	2	3	4	Big
B/M			HML		
Low	-0.471**	0.048	0.108	-0.264*	-0.166
2	-1.264	-0.088	-0.236	-0.165	0.105
3	-1.524***	-0.437**	-0.359*	0.085	0.415
4	-0.372*	0.090	0.085	0.081	0.324**
High	-0.671***	-0.157*	0.090	0.114	0.199
Op			RMW		
Low	-0.088	-0.074	-0.137	0.056	0.448
2	-0.168	0.195	-0.063	0.514*	0.862**
3	-0.551	0.186*	0.271	0.077*	0.514**
4	-0.821	-0.414	-0.298*	0.414**	0.304**
High	-1.119***	0.132*	1.109***	0.406*	0.312*
Inv			CMA		
Low	-0.010	-0.094	0.060	0.261	-0.483
2	0.079	0.086	0.074	0.322	-0.125*
3	0.134	0.210	0.327	0.002*	0.081*
4	0.174	0.027	-0.216	0.228**	0.463*
High	-0.334	0.011	0.408	-0.048*	0.097*
Sen			SEN		
Low	-0.615	-1.059	-0.656*	-0.505	-0.829***
2	0.234	0.138	-0.887***	-0.494***	0.983
3	-0.687	-0.7990	0.150	0.005	1.046*
4	-0.453	-0.568	1.028	0.068***	1.620**
High	1.276***	0.943**	0.697*	-1.413***	1.90***

* p < 0.1, ** p < 0.05, *** p < 0.01

In Table 7, the regression coefficient of the book-to-market ratio factor is significantly non-zero in 9 of the 25 portfolios grouped by size-to-book, the coefficient of the earnings factor is significantly non-zero in 13 of the 25 portfolios grouped by size-to-earnings, the coefficient of the investment factor is significantly non-zero in 7 of the 25 portfolios grouped by size-to-

investment, and the coefficient of the sentiment factor is significantly non-zero in 12 of the 25 portfolios grouped by size-to-emotion. The coefficients of the sentiment factor are more significant in portfolios with high investment sentiment and take mostly positive values, implying that the excess returns of stocks with high investment sentiment are higher, while the sentiment factor in the high investment sentiment grouping explains the excess returns of stocks more strongly. The significance of the regression coefficient of the six-factor model is stronger than that of the five-factor model, indicating that it also explains the excess returns of China's pharmaceutical manufacturing stocks more strongly.

5.3. GRS test

This paper also conducts GRS tests on the two models to further illustrate the suitability of the two models for China's pharmaceutical manufacturing industry. The following table shows the GRS tests under different portfolios.

Table 8 GRS test

Combination	GRS	A|α_i|
Size-B/M		
HML、SMB、RMW、CMA	0.687**	0.494
HML、SMB、RMW、CMA、SEN	0.34***	0.652
Size-Op		
HML、SMB、RMW、CMA	1.712***	0.524
HML、SMB、RMW、CMA、SEN	1.534***	0.539
Size-Inv		
HML、SMB、RMW、CMA	2.223***	0.593
HML、SMB、RMW、CMA、SEN	1.986***	0.620

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

According to the GRS test results in Table 8, it can be seen that the GRS values of different portfolios decrease after adding the sentiment factor, and the mean value of the absolute value of α increases in different degrees, which indicates that the GRS test results of the six-factor model with the addition of the sentiment factor are superior, and the addition of the sentiment factor improves the explanatory power of the model.

6. Conclusion

This paper analyzes China's pharmaceutical manufacturing industry based on Stata analysis software in the context of big data and finds that there is a significant large-scale effect and low book-to-market ratio effect in the pharmaceutical manufacturing industry, i.e., the excess return of stocks of companies with large-scale and low book-to-market ratio is higher. This is inconsistent with the conclusion reached by Fama, and the reason for this analysis is that in China's stock market, pharmaceutical manufacturing has also been a popular industry, and the attention of institutional investors and general investors has been high, and the research on pharmaceutical manufacturing has been more in-depth, and institutional investors will also select some stocks with large scale and growth space for allocation, which causes the leading stocks in pharmaceutical manufacturing to be more popular. In this paper, the Fama-French

five-factor model and the six-factor model with emotion factor are used to analyze the pharmaceutical manufacturing industry in China, and the regression results show that there is an emotion effect in the pharmaceutical manufacturing industry, and the return of stocks with high emotion is higher. In the factor correlation test, all factors, including the sentiment factor, passed the test. In the regression of China's pharmaceutical manufacturing industry using the six-factor model with the addition of the sentiment factor, the sentiment factor is significantly significant, and the factor can well explain the excess return of the stocks of pharmaceutical manufacturing companies. With the addition of the sentiment factor, the goodness-of-fit of the model improved to varying degrees over the traditional Fama-French five-factor model, indicating that the model is suitable for empirical analysis of China's pharmaceutical manufacturing industry.

The innovation of this paper is to explain the excess return of stocks using the individual stock investment sentiment factor. The analysis shows that investors are not rational and the new construction of individual stock sentiment factor in the Fama-French five-factor model is more in line with the real-life situation. When investors invest in China's pharmaceutical manufacturing industry, they can look for some blue-chip stocks with larger market capitalization in circulation, high investment sentiment and higher profitability for layout, which have more stable excess returns.

References

[1] Li Z B, Yang G Y, Feng Y C, et al. An empirical test of the Fama-French five-factor model in the Chinese stock market[J].Journal of Financial Research, 2017(06): 191-206.

[2] Zhang X D, Li J Y. Do earnings factors and investment factors have pricing power? An empirical evidence from Chinese stock market[J].Finance and Economy, 2018(02): 10-18.

[3] Guo X Z. An empirical test of the Fama-French three-factor model and five-factor model for A-share steel firms[J].Hebei Enterprises, 2019(06): 33-36.

[4] Li Z Y, Li H. An empirical study of the Fama-French three-factor and five-factor models of corporate returns in the 5G sector in China[J]..China Market, 2020(10): 1-5.

[5] Miller E M. risk, uncertainty, and divergence of opinion[J].journal of Finance, 1977,32(4):1151-1168.

[6] Festinger L. congnitive dissonance.[J].scientific American, 1962,207(4):93-102.

[7] Lee W Y,Jiang C X,Indro D C. Stock market volatility,excess returns,and the role of investor sentiment.[J].Journal of Banking & Finance,2002,26(12):2277-2299.

[8] Baker M,Wurgler J. Investor Sentiment and the Cross-Section of Stock Returns[J].Management Journal, 2006,61(4):1645-1680.

[9] Yang C,Zhou L. Investor trading behavior,investor sentiment and asset prices[J].North American Journal of Economics & Finance, 2015,34:42-62.

[10] Liu Q Q. Research on quantitative investment strategies based on sentiment timing[D].Shandong University of Finance and Economics, 2015.

[11] Tian F X. A Study of Investor Sentiment and Stock Prices Based on Internet Public Opinion[D].Shanghai Jiao Tong University, 2019.