ISSN 2616-5902 Vol. 5, Issue 14: 1-8, DOI: 10.25236/AJBM.2023.051401

Study on the Correlation between Investor Sentiment and Stock Price Co-movement

Siqiang Gong[1,a], Hua Fang[1,*]

[1]Business School, University of Shanghai for Science and Technology, Shanghai, 200093, China
[a]gongqianze@163.com
[*]Corresponding author

Abstract: *This paper empirically investigates the correlation between investor sentiment and stock price co-movement through a panel vector autoregressive model. The study divides investor sentiment into market sentiment, which represents the overall market attitude, and retail sentiment, which represents the attitude of retail investors, while the studied stock price co-movement refers to the linkage between individual stocks and the market. It is found that the relationship between investor sentiment and stock price co-movement shows a unidirectional effect. Investor sentiment can have an impact on stock price co-movement, however, stock price co-movement cannot significantly affect investor sentiment. Investor sentiment has a negative impact on stock price co-movement. Market sentiment has a significant impact on stock price co-movement, but retail sentiment has a non-significant impact on stock price linkage. This study is a further expansion of the research field on investor sentiment and stock price co-movement, providing a new regulatory perspective for regulator authorities and a new reference for investors to make rational investments.*

Keywords: *Investor Sentiment; Stock Price Co-movement; PVAR Model*

1. Introduction

Since the 1980s, financial anomalies have occurred frequently, and stock price co-movement as a financial anomaly has attracted attention in recent years. Traditional financial theories have struggled to explain financial anomalies such as stock price co-movement. The emergence of behavioral finance theory has provided new explanations for these anomalies and has received increasing recognition [1].

China's capital market started late compared to the developed Western countries and lacks experience in its development, and there are still more problems. As of the end of January 2023, the number of Chinese investors has reached 212,981,000, of which the number of natural persons is 212,469,100, accounting for more than 99.7%. The number of Chinese investors is huge and the percentage of retail investors is high. Therefore, financial anomalies are more significant in China.

2. The Construction of Indicators and Data Description

In this paper, we select CSI 300 constituent stocks as the research object, firstly construct investor sentiment indicators and stock price co-movement indicators, and then study the correlation between them.

From the perspective of the whole and individual ,this paper divide investor sentiment into market sentiment, which represents the overall attitude, and retail investor sentiment, which represents the individual attitude. According to the characteristics of the two sentiments, different data and methods are adopted to synthesize and construct the corresponding stock price co-movement indicators respectively. On this basis, the correlation between market sentiment and stock price co-movement and the correlation between retail investor sentiment and stock price co-movement are studied separately.

Referring to the method of Yu[2] , we select three indicators, MTV (market trading volume), MTR (market turnover rate), CCI (consumer confidence index), and their lagged one-period data to synthesize Market Sentiment Indicator MIS by Principal Component Analysis. The monthly data of the above indicators from January 2020 to December 2022 were obtained from the wind database, and three principal components were selected using principal component analysis, with a cumulative contribution of nearly 95%. Since the market sentiment is identical for all stocks, the time series can be expanded into panel data in the subsequent processing.

Retail investor sentiment mainly represents the attitudes of retail investors, while stock bar posts better reflect the attitudes of retail investors [3]. Therefore, text data on stock bars are chosen to construct retail investor sentiment. In this paper, CSI 300 constituent stocks are selected as the research object, and the comments of the Eastern Fortune stock bar are used as metadata to construct the proxy indicators of retail investor sentiment. The scrapy crawler framework was used to obtain the stock bar data of the CSI 300 constituent stocks in December 2022. Subsequently, the sentiment value of each comment is obtained using the snownlp package, and drawing on Bolin et al [4], the retailer sentiment indicator Sentiment is constructed. The retailer sentiment of stock i in period t (here the unit of t is days) is expressed as:

$$Sentiment_{i,t} = \left(\frac{SentimentMark^{pos}{}_{i,t} - SentimentMark^{neg}{}_{i,t}}{SentimentMark^{pos}{}_{i,t} + SentimentMark^{neg}{}_{i,t}}\right)\ln(1 + SentimentNum_{i,t}) \quad (1)$$

Where $SentimentNum_{i,t}$ denotes the total number of posts for stock i in period t. Its value consists of two components: the number of positive posts and the number of negative posts for that stock in period t. Here, the post sentiment determined by snownlp is considered as positive posts if it is greater than or equal to 0.5, and negative posts if it is less than 0.5.

$$SentimentNum_{i,t} = SentimentNum^{pos}{}_{i,t} + SentimentNum^{neg}{}_{i,t} \quad (2)$$

$SentimentMark_{i,t}^{pos}$ denotes the total positive sentiment score of stock i in period t. Its value is obtained by accumulating the positive post scores in period t. where the score of a post j of stock i in period t is the sentiment value of post j obtained by snownlp. $SentimentMark_{i,t}^{neg}$ denotes the total negative sentiment score of stock i in period t, which is calculated in a similar manner.

$$SentimentMark^{pos}{}_{i,t} = \sum_{j=1}^{SentimentNum^{pos}} Mark^{pos}{}_{i,j} \quad (3)$$

$$SentimentMark^{neg}{}_{i,t} = \sum_{j=1}^{SentimentNum^{neg}} Mark^{neg}{}_{i,j} \quad (4)$$

In the solution will encounter a certain day the number of comments is too small, only negative or positive these cases, leading to errors in the program, for such problems, this paper adopts the method of excluding this day, so the results will exist some missing values, so the obtained data set for the unbalanced panel data, but considering the time of a year individual data missing impact is small, within an acceptable range.

The stock price co-movement studied in this paper is to the linkage between individual stocks and the broader market. To ensure the consistency of the study, monthly stock price co-movement indicators and daily stock price co-movement indicators need to be constructed separately. The approach used here is to use the decidable coefficient R^2 of the CAPM model regression equation as an indicator of the co-movement between individual stocks and the market [5]. The coefficient of determination, R^2, is the value of the current month's linkage between the stock and the market. The 5-minute high-frequency data of individual stocks in one day are regressed, and the coefficient of determination R^2 represents the co-movement value of individual stocks with the market in that day. The high-frequency data are obtained from BAOSTOCK open source data platform, and the other data are obtained from RESSET database and CSMAR database, and the sample period is consistent with the corresponding sentiment indicators. Finally, the panel data sets of monthly co-movement indicator *comovem* and daily co-movement indicator *comoved* of CSI 300 constituents are obtained respectively.

Table 1 shows the description of the variables in this paper.

Table 1: Descriptive Statistics.

Variable	Description
MIS	Market Sentiment
sentiment	Retail Sentiment
comovem	Monthly stock price co-movement value
comoved	Daily stock price co-movement value

3. Model Construction and Solutions

The PVAR model can take individual differences into account and also solve the endogeneity problem, which can well reflect the dynamic relationship between stock price co-movement and investor sentiment. So the correlation between investor sentiment and stock price co-movement is studied by constructing a PVAR model. According to the previous index construction, PVAR models of market sentiment and stock price co-movement and PVAR models of retail investor sentiment and stock price co-movement will be built here respectively.

3.1. Correlation between Market Sentiment and Stock Price Co-movement

3.1.1. Stationarity Test

The comovem variable passes the IPS Stationarity test the MIS variable is a time series expanded panel data, which is not applicable to the IPS test method, and here it is assumed that MIS is a stable data. the IPS test is a unit root test for panel data, whose original hypothesis indicates that the panel data is a non-stable process, and the alternative hypothesis indicates that the panel data is a stable process. The results are shown in Table 2, corresponding to a p-value close to 0. The original hypothesis is strongly rejected, indicating that comove is a stable panel.

Table 2: Results of IPS test for Comove variables.

Im-Pesaran-Shin Unit-Root Test	
Wtbar = -103.25	p-value < 2.2e-16

3.1.2. Model Construction

$$\begin{cases} comovem_{i,t} = \alpha 1_i + \sum_{l=1}^{p} \beta 1_l comovem_{i,t-l} + \sum_{l=1}^{p} \gamma 1_l MIS_{i,t-l} + \varepsilon 1_{i,t} \\ MIS_{i,t} = \alpha 2_i + \sum_{l=1}^{p} \gamma 2_l MIS_{i,t-l} + \sum_{l=1}^{p} \beta 2_l comovem_{i,t-l} + \varepsilon 2_{i,t} \end{cases} \tag{5}$$

Eq. (5) is a PVAR model of market sentiment and stock price co-movement, where i denotes individual (i = 1,2,...300) and t denotes period (t=1,2,...,36). Here, comovem and MIS are treated as two endogenous variables, l denotes the lag order, p denotes the optimal lag order,εis the residual term obeying normal distribution,αdenotes the constant term,βandγis the corresponding coefficient matrix.

3.1.3. Determination of the Optimal Lag Order

The optimal lag order is determined by comparing these information criteria, BIC, AIC and HQIC, together. The solution is performed using the panelvar package [6] in R language. The models with lags of order 1 to 5 are solved separately according to the rule of thumb, and then the information criteria of different orders are compared. It can be seen from Table 3 that the values of BIC, AIC and HQIC for lag order 1 are the minimum values, so the optimal lag order of equation (5) is finally determined to be lag order 1, i.e., PVAR (1).

Table 3: Monthly PVAR model information criterion results.

	BIC	AIC	HQIC
1	-21577.9	-4458.52	-10784.3
2	-21448.7	-4445.79	-10738.8
3	-21283.3	-4427.48	-10676.7
4	-21076.7	-4397.88	-10592.3
5	-20833.8	-4361.92	-10490.6

3.1.4. Model Solving

The model is solved by the panelvar toolkit in R language, and the results obtained are shown in Table 4. As can be seen, for stock price co-movement, both lagged one-period stock price co-movement and market sentiment indicators pass the significance test; while for market sentiment, lagged one-period stock price co-movement value is not a significant variable affecting market sentiment, while lagged one-period market sentiment passes the significance test, indicating that market sentiment is significantly influenced by lagged one-period sentiment. In terms of the direction of influence, both variables are

ISSN 2616-5902 Vol. 5, Issue 14: 1-8, DOI: 10.25236/AJBM.2023.051401

positively influenced by their own lagged period variables, market sentiment is also positively but insignificantly influenced by lagged period stock price co-movement, while stock price co-movement is negatively influenced by lagged period market sentiment.

Table 4: PVAR model results (market sentiment and stock price co-movement model).

Dynamic Panel VAR estimation, two-step GMM		
	comovem	MIS
lag1_comovem	0.1160 ***	0.074
	(0.0159)	(0.0813)
lag1_MIS	-0.0255 ***	0.5207 ***
	(0.0048)	(0.0234)

(*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$)

3.1.5. Model Stability Test

The stability of the model is further examined and it is found that the modulus corresponding to both eigenvalues are less than 1 and all eigenvalues fall within the unit circle, indicating that the model is stable. Table 5 shows the specific values of the eigenvalues and the corresponding modulus.

Table 5: Monthly PVAR (1) model eigenvalues and modulus.

Eigenvalue	Modulus
0.51602	0.51602
0.120746	0.120746

3.1.6. Impulse Response Analysis

In order to further research examine the dynamic relationship between market sentiment and stock price co-movement, the responses to shocks are studied separately and the corresponding impulse response analysis is performed. The results are shown in Figure 1.

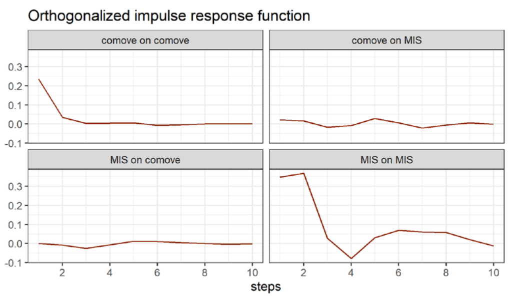

Figure 1: Impulse Response Chart (Market Sentiment and Stock Price Co-movement Model).[1]

When affected by shocks from the comove variable, the response made by the comove variable largely shows a positive decreasing trend. The response reaches its maximum in the first period, then decreases sharply in the second period, continues to decrease slowly in the third period, and then remains largely stable around 0 from the fourth period. This indicates that the stock price co-movement is affected by its own shocks quickly and briefly, and this effect decays quickly to 0. When affected by shocks from the comove variable, the MIS variable makes a smaller response, showing alternating positive and negative changes. The response in each period is small, indicating that market sentiment is barely affected by the stock price co-movement shocks. When receiving shocks from the MIS variables, the comove variables also make a response of relatively small magnitude, but with a certain regularity. The first period response is 0, the second period response is the most intense and reaches a peak, and then the response gradually flattens out and remains stable around 0. The response direction of the stock price co-movement variable to the market sentiment shock is negative. It indicates that the impact of market sentiment on stock price co-movement is negative, the degree of impact is relatively small, and has a lag, in that it has no impact on the current stock price co-movement and has the greatest impact on the next period stock price co-movement. When receiving shocks from the MIS variables, the MIS variables make a more pronounced and volatile response. This impact shows a slow rise first, a sharp decline from the second period, a minimum in the fourth period, and then a circular trend of impact. This illustrates that the impact of market sentiment itself is significant and the direction of the impact is basically positive.

[1] In this figure comove equals to comovem.

ISSN 2616-5902 Vol. 5, Issue 14: 1-8, DOI: 10.25236/AJBM.2023.051401

Therefore regulators can grasp the expected changes in market sentiment accordingly and prepare in advance.

3.1.7. Variance Decomposition

To further examine the overall sources of variation in the variables and the long-term effects among the variables, a variance decomposition of the prediction errors was performed, and the corresponding results are shown in Table 6 and Table 7.

Table 6: Comove variance decomposition.

	comovem	MIS
1	1	0
2	0.998017	0.001983
3	0.997216	0.002784
4	0.996983	0.003017
5	0.99692	0.00308
6	0.996903	0.003097
7	0.996898	0.003102
8	0.996897	0.003103
9	0.996897	0.003103
10	0.996897	0.003103

As can be seen, stock price co-movement is explained by itself in a large proportion and by market sentiment in a small proportion. In the first period, the stock price co-movement explains 100% of the variance and the market sentiment explains 0% of the variance, and the explanatory power of the stock price co-movement remains above 99% in the following periods, while the explanatory power of the market sentiment remains a low level. This indicates that the stock price co-movement is mainly influenced by itself, while it is minimally influenced by market sentiment and starts from the second period.

The MIS variable is also mainly influenced by itself, and is essentially stable at over 97% by itself and below 3% by comove within ten period. Although the contribution of comove to the variance of MIS is small, it is not difficult to find that its contribution is gradually increasing as the period progresses.

Table 7: MIS variance decomposition.

	comovem	MIS
1	0.016628	0.983372
2	0.022288	0.977712
3	0.023832	0.976168
4	0.024242	0.975758
5	0.024351	0.975649
6	0.02438	0.97562
7	0.024387	0.975613
8	0.02439	0.97561
9	0.02439	0.97561
10	0.02439	0.97561

3.2. Correlation between Retailer Sentiment and Stock Price Co-movement

3.2.1. Stationarity Test

IPS tests were performed using the plm package in R language. The results in Table 8 show that both variables, comoved and sentiment, pass the IPS test.

Table 8: Variable IPS test.

	Im-Pesaran-Shin Unit-Root Test	
sentiment	Wtbar = -85.49	p-value < 2.2e-16
comoved	Wtbar = -86.11	p-value < 2.2e-16

3.2.2. Model Construction

Here again, a PVAR model is developed to examine the correlation between sentiment and comoved.

ISSN 2616-5902 Vol. 5, Issue 14: 1-8, DOI: 10.25236/AJBM.2023.051401

The specific form of the model is shown in equation (3.2).

$$\begin{cases} comoved_{i,t} = \alpha 3_i + \sum_{l=1}^{p} \beta 3_l comoved_{i,t-l} + \sum_{l=1}^{p} \gamma 3_l sentiment_{i,t-l} + \varepsilon 3_{i,t} \\ sentiment_{i,t} = \alpha 4_i + \sum_{l=1}^{p} \gamma 4_l sentiment_{i,t-l} + \sum_{l=1}^{p} \beta 4_l comoved_{i,t-l} + \varepsilon 4_{i,t} \end{cases} \tag{6}$$

Eq. (6) is a PVAR model of the correlation between retail investor sentiment and stock price co-movement, where i denotes individual (i =1,2,...300) and t denotes period (t =1,2,...,36). l denotes lag order, p denotes optimal lag order,εis a residual term that follows a normal distribution,αdenotes a constant term,βandγis the corresponding coefficient matrix.

3.2.3. Determination of the Optimal Lag Order

The optimal lag order is also determined by comparing BIC, AIC and HQIC. The results in Table 9 show that the values of BIC, AIC and HQIC with one lag order are the smallest, so the optimal lag order of PVAR model in equation (6) is finally determined to be one lag order, i.e. PVAR (1).

Table 9: Results of daily PVAR model information criteria.

	BIC	AIC	HQIC
1	-6899.513	-1381.563	-3489.094
2	-6794.105	-1369.64	-3448.827
3	-6661.822	-1357.827	-3398.346
4	-6498.79	-1340.975	-3332.809
5	-6289.427	-1304.394	-3237.293

3.2.4. Model Solving

The results of model solving for retail sentiment and stock price co-movement are shown in Table 10. In terms of the significance of the variables, for stock price co-movement, both lagged one-period retail sentiment and stock price co-movement indicators fail the significance test; for retail sentiment, the lagged one-period stock price co-movement variable fails the significance test, while the lagged one-period retail sentiment variable passes the significance test. This indicates that retail sentiment is significantly influenced by lagged one-period retail sentiment. In terms of the direction of influence, both retail sentiment and stock price co-movement variables are positively influenced by their own lagged one-period variables, and retail sentiment is positively influenced by lagged one-period stock price co-movement, while stock price co-movement is negatively influenced by lagged one-period retail sentiment.

Table 10: PVAR model results (retail investor sentiment and stock price co-movement model).

Dynamic Panel VAR estimation, two-step GMM		
	comoved	sentiment
lag1_comoved	0.0034	0.1128
	(0.0476)	(0.1196)
lag1_sentiment	-0.0036	0.1206 ***
	(0.0098)	(0.0277)

(*** p < 0.001; ** p < 0.01; * p < 0.05)

3.2.5. Model Stability Test

The stability of the model is further examined and it is found that the modulus corresponding to both eigenvalues are less than 1 and all eigenvalues fall within the unit circle, indicating that the model is stable. Table 11 shows the specific values of the eigenvalues and the corresponding modulus.

Table 11: Daily PVAR (1) model eigenvalues and modulus.

Eigenvalue	Modulus
0.117032	0.117032
0.006924	0.006924

3.2.6. Impulse Response Analysis

To further examine the dynamic relationship between retail investors' sentiment and stock price co-movement, their responses to shocks are studied separately and the corresponding impulse response analysis is performed. The results are shown in Figure 2.

ISSN 2616-5902 Vol. 5, Issue 14: 1-8, DOI: 10.25236/AJBM.2023.051401

When exposed to shocks from the comove variable, the response from the comove variable tends to first decrease and then remain constant. This response peaks in the first period, then decreases rapidly, and from the second period onwards it decreases rapidly to near zero and remains stable. When subjected to shocks from the comove variable, the response of the retailer sentiment variable sentiment tends to rise, then fall and then remain constant. This response is maximized in the second period and then decreases rapidly to zero from the third period onwards. When subjected to shocks from retail sentiment *sentiment*, the response of the comove variable is almost zero, indicating that the stock price co-movement is minimally affected by shocks from retail sentiment. In terms of direction, the responses are all less than or equal to zero when subjected to shocks from sentiment variables, so the stock price co-movement is negatively affected by investor sentiment shocks. The response of the sentiment variable is strong when it is hit by the sentiment of retail investors, and the overall trend is decreasing and then tends to zero and remains stable.

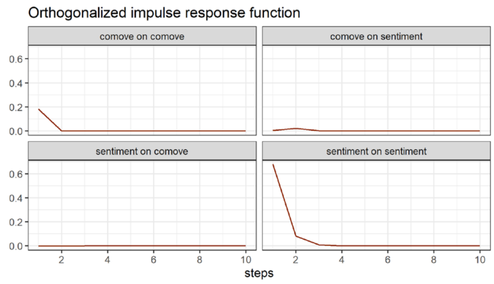

Figure 2: Impulse Response Chart (Retailer Sentiment and Stock Price Co-movement Model).[2]

3.2.7. Variance Decomposition

To further examine the overall sources of variation in the variables and the long-term effects among the variables, a variance decomposition of the model prediction errors was performed, and the corresponding results are shown in Table 12 and Table 13.

Table 12: Decomposition of comoved variance.

	comoved	sentiment
1	1	0
2	0.999821466	0.000179
3	0.999818724	0.000181
4	0.999818686	0.000181
5	0.999818685	0.000181
6	0.999818685	0.000181
7	0.999818685	0.000181
8	0.999818685	0.000181
9	0.999818685	0.000181
10	0.999818685	0.000181

Table 13: Sentiment variance decomposition.

	comoved	sentiment
1	3.70E-05	0.999963
2	0.000970106	0.99903
3	0.000984207	0.999016
4	0.000984402	0.999016
5	0.000984405	0.999016
6	0.000984405	0.999016
7	0.000984405	0.999016
8	0.000984405	0.999016
9	0.000984405	0.999016
10	0.000984405	0.999016

As can be seen, the stock price co-movement is explained by a large percentage of the stock price co-movement itself and by a small percentage of the retail sentiment. In the first period, the stock price co-

[2] In this figure comove equals to comoved.

ISSN 2616-5902 Vol. 5, Issue 14: 1-8, DOI: 10.25236/AJBM.2023.051401

movement explains 100% of the variance and the market sentiment explains 0% of the variance, and the explanatory power of the stock price co-movement remains above 99% in the following periods, while the explanatory power of the retail sentiment remains a low level. This indicates that the stock price co-movement is mainly influenced by itself, while the influence of retail sentiment is very small and starts from the second period. Sentiment variable is also mainly influenced by itself and basically stays above 99% influenced by itself and below 1% influenced by comove in the ten periods. Although the contribution of comove to the variance of sentiment is small, it is not difficult to find that its contribution is gradually increasing as the period progresses.

4. Conclusion

By building a PVAR model, it is found that investor sentiment and stock price co-movement show a one-way influence relationship. Investor sentiment has an impact on stock price co-movement, but stock price co-movement does not have a significant impact on investor sentiment. By dividing investor sentiment into retail sentiment and market sentiment, it is further found that there is a significant difference between the two types of sentiment on stock price co-movement, with market sentiment having a significant effect and retail sentiment having a non-significant effect.

In response to the above results, this paper provides further explanation. The stock price co-movement studied in this paper is the linkage between individual stocks and the market, and the selected sample is the CSI 300 constituent stocks. Market sentiment represents the attitude of the market as a whole, and in terms of volume, market sentiment can have an impact on stock price co-movement, while the opposite is not true. The fact that retail investor sentiment does not have a significant impact on stock price co-movement shows that although the total size of retail investors in China is large, their power in the stock market is not strong, and this large but not strong feature deserves attention, and also reflects that most of the large group of retail investors may play the role of sheep.

Based on the above empirical study, this paper suggests that we should strengthen retail investor education, encourage the development of investor online communication platforms, optimize the regulatory layout, and further improve the market system.

References

[1] Dong ZY, Kang ZP. Behavioral finance and the debate on the efficient market hypothesis[J]. Ningxia Social Science, 2006(4): 43-48.
[2] Yu QL. Research on the impact of investor sentiment on stock price linkage [D]. Southwest Jiaotong University, 2018.
[3] Duan JJ, Liu HZ, Zeng JP. Analysis of information content of Chinese stock online forums[J]. Financial Research, 2017(10): 178-192.
[4] Lei B, Liu Z, Song Y. On stock volatility forecasting based on text mining and deep learning under high-frequency data [J]. Journal of Forecasting, 2021, 40(8): 1596-1610.
[5] Roll R. R-squared [J]. Journal of finance, 1988, 43(3): 541-566.
[6] Sigmund M, Ferstl R. Panel vector autoregression in R with the package panelvar[J]. The Quarterly Review of Economics and Finance, 2021, 80: 693-720.