Journal of Policy Research, 9(2), 268-273.
https://jprpk.com
https://doi.org/10.5281/zenodo.7718731

Herd behavior in stock markets during COVID' 19 Pandemic: A machine learning approach

Fatima Iqbal[1], Dr. Sadia Farooq[2], Dr. Sajid Nazir[3]

Abstract

The COVID-19 pandemic brought unprecedented volatility and uncertainty to global financial markets. During this period, the concept of herd behavior emerged as a prominent factor influencing stock market dynamics. Understanding and quantifying herd behavior patterns during the pandemic is crucial for predicting market trends, detecting potential bubbles, and improving risk management strategies. Herd behavior is characterized by a sudden mimicry among investors which causes temporary deviation of stock market prices. This deviation exacerbates in the presence of extreme conditions or events such as the recent pandemic. Based on social media contextual data this study aims to investigate the presence of herd behavior during the global COVID'19 pandemic. For this purpose state of the art machine learning algorithms are employed as opposed to traditional methodologies that are being used in the past literature for quantifying herd behavior. The surprising results reveal how role of media sentiment during the pandemic shaped the stock markets and investor behavior.

Keywords: Herd behavior, COVID '19, Machine learning

1. Introduction

Advancements in big data has made stock price prediction an intricate task for financial engineers. This is due to the abundance of sources of information around us. One of them is social media which is shaping investor sentiment globally (Benferhat et al., 2017). This paper tests the hypothesis that investors mimic their financial decisions during convulsed periods which is to a great extent correlated to the public sentiment expressed on social media. For instance virtual investment communities actively discuss and make recommendations regarding daily trades. This influence investor sentiment by selecting and filtering information through volume and repetition of information (Alomari et al, 2021). Shiller (2016) explained further how financial news/media build interest by associating positive/negative news to positive/negative stock returns. Hence investors during bull market pile up positive theories arguing the economy will flourish driving prices even higher and vice versa. This behavioral disposition causing stock market prices to deviate from its fundamentals and is termed as 'herd behavior'. Traditional approaches to studying herd behavior in financial markets often rely on qualitative assessments, surveys, or expert opinions. However, with the recent advancements in machine learning techniques and the availability of vast amounts of financial data, researchers now have the opportunity to explore herd behavior from a quantitative perspective. By employing machine learning algorithms and data-driven models, we can uncover hidden patterns, identify key influencers, and gain insights into the dynamics of herd behavior during the COVID-19 pandemic.

Earlier studies have led the premise for the current study i.e. to investigate herding in the presence of big data sentiment. Given the fact that the recent COVID 19 pandemic transformed from a health crisis to an economic meltdown (Bouri et al., 2021). This study will utilize twitter data before and after the pandemic to see if their existed an anomaly. The findings of this research will be useful in establishing repercussions for individual investors, market regulators, and media organizations. Moreover the findings will aid individual investors in understanding the role of social media and its perilous effects on stock markets. Investors would be encouraged to avoid sentiment trading and adopt a more logical approach to understanding market mechanisms. Additionally, this study enlightens financial regulators on the role of social media in fostering extreme market conditions. These findings may be used by regulators to create a mechanism that would alert them when a speculative bubble is forming that would stop the market from free-falling following a market meltdown. To keep markets under check, a formal channel for information dissemination similar to social media may be developed.

2. Literature Review

Efficient market hypothesis by Fama (1965) suggests that all markets are efficient, and investors make financial decisions rationally provided stock market prices reflect all available information. Thus efficient markets cannot deviate from their fundamentals hence no possibility of formation of a speculative bubble. On the contrary, literature suggests investors are not always rational and mainly rely on aggregate market movement while optimizing their portfolios. Although EMH has been questioned on multiple empirical grounds but it has not been rejected. Instead a vast body of literature has highlighted alternative explanation to the market functioning referred to as behavioral finance.

[1] PhD Scholar, Hailey College of Commerce, University of the Punjab, Pakistan
[2] Assistant Professor, Hailey College of Commerce, University of the Punjab, Pakistan
[3] Associate Professor, Institute of Administrative sciences, University of the Punjab, Pakistan

Noise trading is one of the major explanation provided by proponents of behavioral finance. While explaining market irrationality, Black (1986) proposed that various investors have varied sentiments, and when players with incorrect beliefs trade, they introduce 'noise' into the market. These investors are referred as "noise traders." Investor sentiment is a belief about future pricing that is not supported by fundamentals such as future cash flows and discount rates. Initially, investor emotion was thought to be a disconnected phenomenon tied to a single asset or investment. However, Barber et al. (2008) shown that small individual investors - noise traders - are systematically connected and have the power to impact the entire market, especially when arbitrage is difficult. Recent studies like Garcia (2013) and Hanna et al. (2020) provide interesting long-run market level evidence on media sentiment and stock markets.

Herd behavior is one of the financial market anomalies that has been defined as a behavioral tendency characterized by a group of investors trading in the same direction over time. In such cases, investors discard their earlier ideas and emulate others unreasonably. As a result of such investor activity, asset prices deviate from their fair values, introducing volatility into the market (Fei & Liu, 2021). Herding in the context of digital platforms is a comparatively recent phenomena. Maqsood et al. (2020) investigated the impact of big local and global occurrences on stock exchange predictions using the Twitter dataset. On the basis of 11.42 million tweets, public mood was assessed. Findings demonstrated event-based sentiments increase stock market performance by employing deep learning algorithms. Dhall & Singh (2020) identified herd behavior in their research study at the industrial level during the pre- and post-COVID-19 pandemic, alongside bullish and bearish markets, using the model proposed by Chang et al. (2000). The daily stock closing prices of the national stock market (NSE) were used for experimentation purposes in this study, and the results suggest that there is no herd behavior prior to COVID-19.

There is a large body of literature on using Twitter data to forecast stock prices. Carvalho and plastino (2021) propose using sentiment analysis on tweets in conjunction with data on tweet frequency as input to an LSTM for stock price prediction. This analysis implies that publicly available Twitter data can be very beneficial for market prediction, based on promising outcomes. Chong et al. (2017) propose employing a BERT model for sentiment analysis and combining it with data on stock price changes. Many other publications have attempted to forecast stock prices using historical stock market data (without the use of social media data), such as Goldstein (2021), which combined SVM and Arima models to predict stock prices. Overall, most proposed systems for predicting stock prices rely on stock data, which is occasionally combined with social media context. We offer a novel approach that employs exclusively Twitter data as features, with the goal of verifying the premise that public mood expressed on Twitter can be utilized to forecast stock prices.

Moreover, a recent review of the literature reveals that combining artificial intelligence techniques such as deep learning with traditional methodologies can provide new metrics in the field of financial engineering and serve as motivation for future research ideas (Hasan et al.,2020). Because financial engineering is crucial to risk management practices, the benefits of deep learning models must be realized in the field of finance for improved forecasting accuracy. As a result, our study will fill this vacuum by investigating herd behavior in a social media context using machine learning technologies. In order to adequately study investor activity during this time, Twitter-based sentiment research will be undertaken in the context of unpredictable event such as COVID-19.

3. Methodology

The suggested methodology is divided into three parts: data gathering, sentiment computation of important events, and deep learning herd behavior prediction.

3.1. Data Acquisition

The sentiment analysis will be calculated for the news data extracted from Twitter especially related to COVID19 of Pakistan in 2018. Mega-events have a significant impact on the stock market performance (Maqsood et al., 2020). Furthermore, we focus on the most recent event because of the lack of availability of textual data set on Twitter.

We will use daily data of stock indices and individual stocks listed in the stock markets of Pakistan, and UK. Data covers the range from Jan 2006 to Dec 2021. Daily data of individual stocks and indices will be collected from yahoo finance. For testing herd anomaly, *CSSD* of asset returns is used as a measure of return dispersion in the literature (Christie & Huang, 1995a). We use a more robust measure called cross-sectional absolute deviation (CSAD) presented by Chang et al. (2000).

$$CSAD_t = \frac{1}{N}\sum_{i=1}^{N}\left|R_{i,t} - R_{m,t}\right|$$

3.2. Pre-processing/sentiment

The use of social media sentiment in business analytics is critical for increasing the effectiveness of prediction models. Researchers have effectively established a link between social media mood and business analytics, particularly for stock market predictions. The social media sentiment will be measured in this study for the period of COVID-19. First, tweets/textual data from the social media platform Twitter will be extracted. The textual data will next be preprocessed to remove any superfluous tokens or words. After tokenizing the textual data into a word list, the parsing algorithm will be used to separate whitespaces, URLs, emoticons, and punctuation. The sentiment analysis will then be carried out.

3.3. Deep learning herding prediction

Linear regression (LR) is a popular predictive analysis machine learning approach. These linear regression models describe and assess the relationship between dependent and independent variables. One of the simplest variants of the linear regression equation, with only one dependent and independent variable, is shown below.

$$y = mx + b$$

Where y represents the dependent variable, x represents the independent variable, and b represents the constant term, commonly known as the regression coefficient.

3.4. Support Vector Regression (SVR)

Support vector regression (SVR) is a regression model as well as a generalization of support vector machines and is able to solve both linear and non-linear regression problems. The SVR model builds a functional estimator using a fraction of the given dataset.

3.5. Single Layer Neural Network (SLP)

SLPs (single-layer neural networks) are the most basic type of artificial neural networks (ANNs). ANNs, or artificial neural networks, are computational models that mimic the human brain and are particularly useful for regression and predictive modelling problems. A simple single-layer neural network is comprised of only an input and output layer. In the subsequent step, the output of neurons from 1 to n is computed using the following equation:

$$\left(u_1 = \left(\sum iw_i x_i + b\right), u_2 = \left(\sum iw_i x_i + b\right) \dots \dots \dots . u_n \left(\sum iw_i x_i + b\right)\right)$$

In the above equation. $u_1 \ldots u_n$ is the total number of neurons in the input layers where w_i are the weights of each connection. More explicitly, every input gets multiplied with its associated weights.

4. Results & Discussion

4.1. Descriptive statistics

This section discusses the central tendency and dispersion measures for all variables. Furthermore, the Jarque-Bera test was used to determine normalcy. The results demonstrate that all of the variables are unconventional because the Jarque-Bera test statistics are highly significant at the 1% level.

Table 1: Descriptive Statistics

	ARMPK	ARMU	CSADPK	CSADU	SRMPK	SRMU
Mean	0.008	0.004	0.015	0.299	0.0015	0.0001
Std. Dev.	0.009	0.006	0.008	0.316	0.0038	0.0006
Jarque-Bera	12037	12144	56455	42234	820198	9391
Probability	0.000	0.000	0.000	0.000	0.000	0.000
Observations	3748	3748	3748	3748	3748	3748

The mean value of absolute market returns of all the three stock markets ranges from 0.007 to 0.010. Similarly CSAD of all the markets ranges from 0.015 to 0.499. In total we have 3748 daily observations.

4.2. Unit Root Test

We applied Augmented Dickey Fuller (ADF) test to investigate the unit root problem in our data sets. If data has a unit root problem then regression models do not provide reliable results. Following are the results of ADF test of unit root:

Table 2: Unit root test results

Variables	Level	1st Difference	Tabulated Test statistics at 5%
$\lvert R_{m,pk,t} \rvert$	-11.081***		-2.863
$\lvert R_{m,u,t} \rvert$	-9.214***		-2.863
CSADPK	-8.928***		-2.863
CSADU	-1.17	-64.155***	-2.863
$R^2_{m,pk,t}$	-10.956***		-2.863
$R^2_{m,u,t}$	-6.129***		-2.863

*** shows significance at 1%.

The results show that all the variables are stationary at level except CSADU which is stationary at 1st difference.

4.3. CSAD forecasting without Sentiment

In this part we present the results of volatility forecasting without considering any sentiment. The following Table 3 shows the detailed results on the basis of three different evaluation matrices i.e. MAE, MSE and RMSE.

Table 3: CSAD Forecasting for US and UK without sentiments

UK data			
	MAE	MSE	RMSE
LR	0.228	0.207	0.314
SVR	0.178	0.211	0.313
SLNN	0.204	0.102	0.300
PK data			
	MAE	MSE	RMSE
LR	0.038	0.002	0.045
SVR	0.056	0.004	0.063
SLNN	0.042	0.002	0.049

It is evident from the forecasted results of CSAD of UK that Support Vector Regression (SVR) and Single layer neural network (SLNN) model performs better than linear regression on the basis of Mean Absolute Error (MAE). However, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) of linear regression are more accurate as compared to SVR and Single Layer Neural Network (SLNN). In case of CSAD of Pakistan, SLNN outperforms linear regression and SVR on the basis of all the three evaluation matrices. Furthermore, linear regression provides better results of CSAD forecasting for Pakistan as compared to SVR and SLNN models.

Table 4: CSAD Forecasting of UK and Pakistan with sentiment of "COVID"

UK data			
	MAE	MSE	RMSE
LR	0.2959	0.1049	0.3240
SVR	0.2899	0.1127	0.3357
SLNN	0.2991	0.1054	0.3247
PK data			
	MAE	MSE	RMSE
LR	0.0371	0.0020	0.0447
SVR	0.0554	0.0039	0.0628
SLNN	0.0367	0.0021	0.0462

In this part we have incorporated the twitter sentiment of COVID'19 to forecast CSAD of UK and Pakistan. It is evident that linear regression performs better than SVR and SLNN models. In case of Pakistan, CSAD shows high spikes during the time of global financial crises of 2007-08 and COVID'19 period. Similarly market return also show similar negative patterns during the said periods. CSAD and market return of UK shows a similar trend as of Pakistan.

4.4. Estimation of Herding Coefficients

We estimated the regression coefficients in this part using linear regression, SVR, Single Layer Perceptron (SLP), and Multiple Layer Perceptron (MLP) models.

Our findings indicate that the overall outcomes for all four models, LR, SVR, SLP, and MLP, are as same as if there were no sentiment for the UK. However, when we included Twitter sentiment from COVID'19 in SLP and SVR models for Pakistan, herding coefficients became significantly significant. The herding coefficients for SVR and SLP are -0.454 and -0.089, respectively, and are significant at the 1% and 10% levels. P-values are given in parenthesis after the coefficients. The COVID'19 sentiment coefficient is also highly significant in the case of the UK indicating that the COVID'19 generated volatility in the stock markets. Furthermore, in the case of Pakistan, it worked as a spark for herd abnormality.

Table 5: Estimated Herding Coefficients for UK and PK with sentiment of "COVID'19"

	β_1	β_2	β_3	β_o
UK data				
LR	-0.326***	-0.368***	0.129***	0.199***
	(0.000)	(0.000)	(0.000)	(0.000)
SVR	-1.005***	-0.238***	0.662***	0.311***
	(0.000)	(0.000)	(0.000)	(0.000)
SLP	-0.305***	-0.430***	0.335***	0.329***
	(0.000)	(0.000)	(0.000)	(0.000)
MLP	-0.613***	-0.012	-0.222***	0.502***
	(0.000)	(0.712)	(0.000)	(0.000)
PK data				
LR	0.449***	-0.070	-0.010	0.121
	(0.000)	(0.169)	(0.168)	(0.000)
SVR	0.673***	-0.454***	0.007	0.101
	(0.000)	(0.000)	(0.392)	(0.000)
SLP	0.449***	-0.089*	-0.015	0.112
	(0.000)	(0.087)	(0.071)	(0.000)
MLP	0.414***	0.211***	-0.173***	0.126***
	(0.000)	(0.000)	(0.000)	(0.000)

5. Conclusion

Using machine learning techniques, the current study explores investors' mimic behaviour during convulsed times. Twitter sentiment analysis based on an uncertain event i.e. COVID 19 pandemic was carried out for this purpose. The findings of the study provide an evidence of herding behavior in the selected stock markets. In this perspective, the study offers credence to the behavioral theory of market functioning, in which investors are not rational and trade on noise to produce market speculations Adopting a machine learning approach allows us to explore intricate patterns and relationships that may have previously gone unnoticed. This research contributes to a comprehensive understanding of herd behavior and its implications for financial markets during the COVID-19 pandemic. Ultimately, the study aims to provide valuable insights into the mechanisms driving market dynamics, aiding in the development of informed decision-making strategies for market participants and policymakers alike.

References

Alomari, M., Al Rababa'a, A. R., El-Nader, G., Alkhataybeh, A., & Rehman, M. U. (2021). Examining the effects of news and media sentiments on volatility and correlation: Evidence from the UK. *The Quarterly Review of Economics and Finance*, 82, 280-297.

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, *21*(2), 785-818.

Benferhat, S., Tabia, K., & Ali, M. (2017). Advances In Artificial Intelligence: From Theory To Practice: 30th International Conference On Industrial Engineering And Other Applications Of Applied Intelligent Systems, IEA/AIE 2017, Arras, France, June 27-30, 2017, *Proceedings, Part II* (Vol. 10351). Springer.

Black, F. (1986). Noise. *The Journal of Finance*, *41*(3), 528-543.

Bouri, E., Demirer, R., Gupta, R., & Nel, J. (2021). COVID-19 Pandemic and Investor Herding in International Stock Markets. *Risks*, *9*(9), 168.

Chang, E. C., Cheng, J. W., & Khorana, A. (2000). An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, *24*(10), 1651-1679.

Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, *83*, 187-205.

Christie, W. G., & Huang, R. D. (1995a). Following the pied piper: Do individual returns herd around the market? *Financial Analysts Journal*, 31-37.

Christie, W. G., & Huang, R. D. (1995b). Following the pied piper: Do individual returns herd around the market? *Financial Analysts Journal*, *51*(4), 31-37.

Dhall, R., & Singh, B. (2020). The COVID-19 pandemic and herding behaviour: Evidence from India's stock market. *Millennial Asia*, *11*(3), 366-390.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, 38(1), 34-105.

Fei, T., & Liu, X. (2021). Herding and market volatility. *International Review of Financial Analysis*, *78*, 101880.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, *68*(3), 1267-1300.

Goldstein, I., Spatt, C. S., & Ye, M. (2021). Big data in finance. *The Review of Financial Studies, 34*(7), 3213-3225.

Hanna, A. J., Turner, J. D., & Walker, C. B. (2020). News media and investor sentiment during bull and bear markets. *The European Journal of Finance*, 26(14), 1377-1395.

Hasan, M. M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. *Journal of Big Data*, *7*(1), 1-17.

Maqsood, H., Mehmood, I., Maqsood, M., Yasir, M., Afzal, S., Aadil, F., Selim, M. M., & Muhammad, K. (2020). A local and global event sentiment based efficient stock exchange forecasting using deep learning. *International Journal of Information Management, 50*, 432-451.

Shiller, R. J. (2016). *Irrational exuberance: Revised and Expanded Third Edition.* Princeton University Press.