Heliyon 9 (2023) e22788

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Study on the sentimental influence on Indian stock price

Renju Rachel Varghese [a,b,*], Biju R. Mohan [a]

[a] National Institute of Technology Karnataka, Surathkal, Mangaluru, 575025, Karnataka, India
[b] Saintgits College of Engineering, Kottukulam Hills, Kottayam, 686532, Kerala, India

ARTICLE INFO

Dataset link: https://finance.yahoo.com/

Dataset link: https://github.com/renjuvarghese/Fintech-dataset

Dataset link: https://www.nseindia.com/

Dataset link: https://finance.yahoo.com/quote/tcs/history/

Keywords:
Parametric
Non parametric
Shannon transfer entropy
Renyi transfer entropy
Causality

ABSTRACT

The rapidly increasing scientific research on the stock market and the visible impact of media on equity prices are nowadays in limelight. To a greater extent, causal analysis can reckon the sentimental effect of the broadcasted content on stock valuation. We propose a four stage model to detect the direction of information flow between the news sentiment and stock price. Whilst web scraping explores and extracts the news datasets, the modified VADER algorithm finds the sentiments of the aired media. The associational causal analysis determines the cause effect between the news and stock price. The results suggest that the non parametric Shannon and Renyi's entropy approach supersedes the Granger test, a parametric study which is constrained to Gaussian time series with linear causation. Since Renyi's Entropy can perfectly identify the deluge of information during quick leaps, it is regarded as a beneficial formulation for investors when evaluating stocks with a fewer number of news mentions. The impact of news during the COVID-19 pandemic over the pharmaceutical sector was also done. The study infers an explicit information flow and direction of causality between news sentiment and stock price movement, which can be used to devise future investment and consumption strategies.

1. Introduction

The stock price of a firm is often the most noticeable aspect of its financial data. However, it is still debatable if it can make accurate predictions of the stock market. Random walk theory and the Efficient Market Hypothesis (EMH) were the foundations for the earlier studies that attempted to anticipate the market trend [[1] [2] [3] [4]]. The focus of academic discussion has recently changed due to behavioral finance to examine the relationship between investor sentiment and asset values. Financial choices are greatly influenced by emotion and mood, according to behavioral economics, which has a profound impact on individual behavior [5]. So, it is generally acknowledged that market forecasting relies heavily on investor opinion. The Efficient Market Hypothesis assumed that the primary factor that influenced stock market prices was news rather than current prices or past prices. Prices on the stock market will behave in accordance with a random walk pattern and more than fifty percent accuracy in predicting their behavior will be impossible [6]. According to the research of Chan Wesley and Robinson [[7] [8]], news events can have a substantial impact on the short term direction of stock prices. The impact of news on the mood of the market cannot be overstated [9]. An investor's perception of a company's position in the market may shift as a result of a series of unfavorable events [10]. When the market is optimistic about a certain company, the stock price of that company may rise; conversely, if the market is pessimistic, the value

* Corresponding author at: National Institute of Technology Karnataka, Surathkal, Mangaluru, 575025, Karnataka, India.
E-mail addresses: renjuvarghese.217it503@nitk.edu.in (R.R. Varghese), biju@nitk.edu.in (B.R. Mohan).

https://doi.org/10.1016/j.heliyon.2023.e22788
Received 8 May 2023; Received in revised form 30 October 2023; Accepted 20 November 2023

Available online 22 November 2023

of that company may vanish overnight. A company's future can be determined by its attitudes, regardless of whether or not those sentiments are firmly established in information and logic or are just a mere guess [[11] [12]].

The financial news pertaining to the stock market is continually updated, and investors are presented with a multitude of distinct news formats. This information has a tendency to disseminate feelings. When it comes to making financial decisions, traders and investors are susceptible to the emotions of the market. Thus, one of the most pressing issues in contemporary financial research is how to evaluate the impact of news on stock prices.

Within the scope of this research, we aim to develop a model to investigate the impact of financial news articles have on stock prices using associational causal analysis. In this study we use a unique method to determine the sentiment of financial news using the modified VADER algorithm. In quantitative economics, the examination of the interrelationships of possible causes is an essential component of both theoretical and empirical contributions. To quantify the relationship between the sentiment of the news and its effect on the stock price, we are using the concept of causality to find the cause effect. This paper address four research questions:

- Since correlation and causation are different, firstly this work finds whether there exists a correlation between stock market fluctuations and financial news sentiment. As a response to these issues, we need to compare market movement and sentiment data using Pearson's correlation coefficient.
- To analyze the interdependence between financial market sentiment and movements in stock prices, this work explores if the sentiment of the news drives the prices of the stock market or vice-versa. Quantification of this will be accomplished by the utilization of the associational causal analysis.
- Does an associational causal analysis of parametric and non-parametric approaches reveal any difference? This work intends to capture the directional flow of data between the stock market and financial news sentiment.
- Is there a difference between the outcomes in times of economic distress and times of normalcy? This could be clearly depicted by doing an analysis on a pharmaceutical company before and during COVID 19 period. This will be a great chance to repeat the experiment and see how results differ in calm and turbulent markets.

This research is vital for several reasons. Firstly, it fills a gap in current literature by examining how sentiment impacts Indian stock prices. The study also emphasizes the utilization of the Modified VADER algorithm (trained on financial domain) for determining news sentiment. Secondly, by utilizing both parametric and non-parametric causal analyses, this study offers a comprehensive view of the relationship. Parametric methods provide a structured approach, while non-parametric techniques capture more complex dynamics, including potential non-linearities. Moreover, by employing both types of analyses, the research overcomes limitations associated with relying on just one method. This dual-pronged approach enables a nuanced examination capable of unveiling causal relationships of news sentiment over stock prices in the Indian context that may have been previously overlooked, offering invaluable perspectives for both investors and researchers. Additionally, the study underscores the effectiveness of the Renyi transfer entropy approach, particularly in situations involving limited news mentions, further highlighting its practical relevance.

2. Related work

The section on related research is divided into three distinct segments. The first part delves into studies exploring the impact of news or social media on stock prices. The second part encompasses papers centered around sentiment analysis. Finally, the last segment addresses works focused on correlation and causal analysis.

Though, extensive literature has discussed the relationship among the news and stock prices in many ways, it neglected their causal relationship. Amidst the COVID-19 crisis, Das et al. [13] demonstrated that an augmenting mood with news headlines in conjunction with stock data significantly enhanced the precision of Nifty-50 predictions. Anbaee Farimani et al. (2022) [14] scrutinized the impact of temporal snapshots of information and sentiment trends on price prognostication across diverse currency pairs and Bitcoin. Figà (2022) [15] conducted a thorough examination into the repercussions of textual content. It was derived from newspapers and social media postings related to the returns and variances of components within the S&P500 index. It employed a method grounded in the GARCH family of models. Alqahtani (2020) [16] examined the effects of news on stock prices and discovered that any positive news typically prompts people to purchase stocks. Positive economic indicators and corporate acquisitions, too contribute to rising stock values. Instead of concentrating on the range of stock prices, Ren (2020) [17] chose the BIAS indicator based on the traits of investor behavior to examine how news impacts the pattern of stock prices over time. Yi [18] used a Mixed Frequency VAR model to investigate the impact of US federal spending news on the S&P 500 index. Their findings revealed the negative effect of spending news shocks on the index. According to Cui's study [19], stock prices tend to overreact to news in regard with governance, social, and the environment. Vega researched the impact of the media, and concluded that the stocks linked to confidential information drifted slightly and those linked to public news diverged significantly. These studies collectively emphasize the importance of news sentiment in relation to the stock market.

Vicari (2021) [20] claims that analysis of news sentiment using the techniques of NLP and deep learning can help to develop of an algorithmic trading strategy. The research was centered around scrutinizing 25 daily news headlines in relation to the Dow Jones industrial average. Yadhav (2020) [21] used an unsupervised method and the semantic orientation of text documents to conduct research on the sentiment of financial news. Nemes (2021) [22] conducted research in which he used stock news headlines to forecast stock price. In his job, he has used the BERT to determine the tone of the financial news and proved its reliability. Kim [23] employed a mathematical framework to study how human sentiment influenced stock movements, with a special focus on textual data. They utilized FinBERT, a specialized language model based on BERT, for sentiment analysis on financial texts. The study involved conducting a comprehensive sentiment analysis of BTC-related tweets using the Valence Aware Dictionary and sEntiment Reasoner method. This assessment encompassed the conversion of tweet text into a sentiment score. The research indicated that the performance of VADER enhanced by training it with various dictionaries.

The studies mentioned hereafter emphasize employing correlation analysis to examine the connection between news sentiment and stock prices. In earlier times, there was a tendency to misconstrue correlation as indication of a genuine cause-and-effect relationship. Consequently, numerous studies utilized Pearson correlation as a means to substantiate causality. This method allows researchers to characterize the correlation profile between financial time series. Typically, the Pearson correlation was employed to characterize the correlation profile and define the relationship between financial time series [24]. With the help of Pearson correlation analysis, Huang et al. (2019) [25] examined the correlation between national stock indices, both prior to and subsequent to the formation of the European Union. It is crucial to recognize that the conventional Pearson correlation analysis is constrained in its capacity. While it can discern a statistical link between variables, it lacks the capability to infer causality.

Advanced causality methodologies can provide comprehensive understanding of the causal association between news sentiment and stock prices. Many studies have been focusing Granger causality and its non linear variations to uncover the real cause-and-effect dynamics between these two variables [26]. According to Nogueria [27] the current tools for causality inference can be in two broad categories: process knowledge-based methods and data-driven methods. Whilst the former derived connection and causality from first-principle models and process topology, the latter converts the results into machine-readable representations (adjacency matrix and the signed directed graph). Common data-driven methods were cross-correlation analysis (CCA), granger causality analysis (GCA), and transfer entropy (TE). To describe the direction of causation between time series, the Granger-causality [28], [29] was introduced. Some research used the Granger-causality idea to look for such link in different financial time-series, since its discovery. Tran [30] Performed a Granger causality examination between energy consumption and Gross Domestic Product (GDP) for nine sets of developing economies and seven developed nations. Dong et al. [31] analyzed the connection between stock markets and internet-derived big data from search engines, public media, and social platforms. They found that a robust, bidirectional Granger causality, displaying both linear and nonlinear trends, linking stock markets and investors' web search behaviors due to shared trends and uncertain factors. Phoong et al. [32] examined predictive causality between stock prices and exchange rates of China, and the USA. Granger causality tests are applied, revealing unidirectional causal relationships in the US. Zhifang et al. [33] examined the nonlinear causality between Chinese investor sentiment and crude oil price using non linear granger causality test. On the contrary, Granger-causality was contingent on the chosen model and placed emphasis on the presence of a Granger-causal association rather than its magnitude. Since Granger causality fundamentally relied on the linear regression of the stochastic process, it is considered as one of the limitations. Even the non-linear adaptations of Granger causality estimations ultimately gauge linear interactions. Nevertheless, these methodologies tend to neglect the nonlinear intricacies that underlie interactions between variables, frequently resulting in an insufficient quantification of causal potency.

In order to accurately establish the cause-effect relationship between news sentiment and stock prices, there is a requirement for non-linear causal techniques. While some studies have begun employing these methods, the number of works focused on non-linear and non-parametric causal analyses in the equity market is relatively limited. As a method for the exchange of information, Transfer entropy was initially conceived and proposed by Schreiber [34], [35], [36]. Paulo Ferreira et al. [37] elucidated the influence of dynamics among stock market indices. Employing a nonlinear methodology centered on transfer entropy, the study identified notable influential relationship, with the US index that prominently stands out as key influencer, particularly in relation to certain CEEC indices. With a time delay and rolling window, Shaeowi [24] studied the Chinese stock market using linear Pearson and nonlinear transfer entropy. With the time delay, the linear correlation was favorably connected with transfer entropy at all time. Lim [38] Examined the transmission of information between the Credit Default Swap market and the stock market in the United States by utilizing Transfer Entropy (TE) while focusing on both inter- and intra-structural dimensions. They identified a substantial alteration in the information transmission dynamics amid the financial crisis. Yue [39] used transfer entropy to conduct an analysis of the information flows that occurred between the different industrial sectors of the Chinese stock market. Transfer entropy became an effective and widespread method to capture cause- and effect links in complex systems.

In conclusion, our extensive literature survey reveals a predominant focus on models designed to establish causal links within the equity market, with an emphasis on linear relationships. The existing literature underscores a notable gap in research, emphasizing the imperative to delve into alternative nonlinear transfer entropy methods, such as Renyi transfer entropy techniques and role of Shannon transfer entropy in the field of finance. This area remains underexplored, despite its potential to offer distinct advantages over conventional approaches. Investigating the application of Renyi transfer entropy methods could potentially yield enhanced insights into complex dynamic systems, allowing for a more nuanced understanding of information flow and causality in finance. Additionally, it is evident that these systems have predominantly been developed and tested using data from the US stock market, highlighting a consideration of nonlinear causal dynamics in future research endeavors. It is essential to undertake investigations to the utilization of various forms of transfer entropy (non linear causal dynamics), including Shannon and Rényi entropy, within the context of the equity market (see Table 1 for the overview of crucial studies on identifying causal relationships in the stock market).

3. Methods and data

This section furnishes an in-depth account of the dataset utilized in this study, followed by an elucidation of the theoretical and mathematical concepts underpinning the research in the subsequent portions.

Table 1
Overview of crucial studies on identifying causal relationships in the stock market.

Reference	Methodology	Summary	Limitation
Dong et al. [31]	Granger Causality	Examines stock market prices and data from social platform	Not suitable for a nonlinear system
Phoong et al. [32]	Granger Causality	Examines stock market prices and exchange rates	Validated for US dataset
Zhifang et al. [33]	Granger Causality	Examines investor sentiment and crude oil price	Not suitable for complex systems
Paulo et al. [37]	Shannon Transfer Entropy	Examines dynamics among stock prices	Fails to identify causality at the individual level.
Yue et al. [39]	Shannon Transfer Entropy	Sectoral Information flow	Considers the aggregate level impact
Lim et al. [38]	Shannon Transfer Entropy	Examines Information flow of CDS and Stockmarket	Does not test statistical significance.

Table 2
Company list.

Company Name	Stock Sector	Ticker Symbol
Tata Consultancy Services	IT industry	TCS
Tech Mahindra	IT industry	TECHM
Infosys Ltd	IT industry	INFY
Wipro Ltd	IT industry	WIPRO
HCL technologies	IT industry	HCLTECH
State Bank of India	Banking	SBIN
HDFC Bank Limited	Banking	HDFCBANK
Indusind Bank Ltd	Banking	INDUSINDBK
Tata Steel Limited	Iron and Steel	TATASTEEL
ITC Limited	FMCG	ITC
Tata Motors	Automobile	TATAMOTORS
Asian Paints Ltd	Paints	ASIANPAINT
Bharat Petroleum Corporation Ltd	Refineries	BPCL
Cipla Ltd	Pharmaceuticals and Drugs	CIPLA
Dr Reddy's Laboratories Ltd	Pharmaceuticals and Drugs	DRREDDY
Sun Pharmaceutical Industries Limited	Pharmaceuticals and Drugs	SUNPHARMA

3.1. Data

Two data sets which were used for this work are the historical stock prices (open, low, high, volume traded) and the financial news data, taken for each of 16 firms. The historical stock prices of the firm's are obtained from the Yahoo finance and the National Stock Exchange. Web scraping was the methodology used for fetching financial news for each firm. The news statistics of publicly listed top firms were collected for a period spanning between January 1, 2018 to December 22, 2022 (inclusive of both days). Stocks in information technology (IT) and banking sector were given greater importance because of its huge media attention when compared to other sectors in the stock market. To gauge the effect of the COVID 19 pandemic on stock values, we focused on the pharmaceutical industry. The list of companies selected is shown in Table 2.

3.2. Methods

A four-stage model architecture, as illustrated in Fig. 1, is employed for this study. The initial two stages entail data acquisition, where news articles and stock prices for various companies are gathered through web scraping. The collected financial news then undergo preprocessing to eliminate any extraneous or irrelevant information. Subsequently, sentiment analysis is performed using a modified version of the VADER algorithm from the field of Natural Language Processing. The subsequent step involves applying the Pearson correlation to ascertain the relationship between stock prices and news sentiment. Finally, the last stage encompasses an associational causal analysis, employing both parametric and non-parametric techniques, to quantitatively assess the relationship and trace the information flow between stock prices and news sentiment.

3.2.1. Web scraping

Web scraping involves automating the process of extracting extensive amounts of information from websites. This information primarily exists in the form of unorganized HTML data, which is then converted into a structured format within a spreadsheet or database. This structured data can subsequently be applied in various applications. Online scraping is a data mining technique that involves extracting information from specific web pages in order to produce vast pools of data that can subsequently be analyzed to discover new patterns. Two basic techniques exist for extracting data from a website: 1) Utilize the website API (if it exists). 2) Access the HTML of the webpage and extract useful information from it. This process is referred to as web harvesting, web scraping,

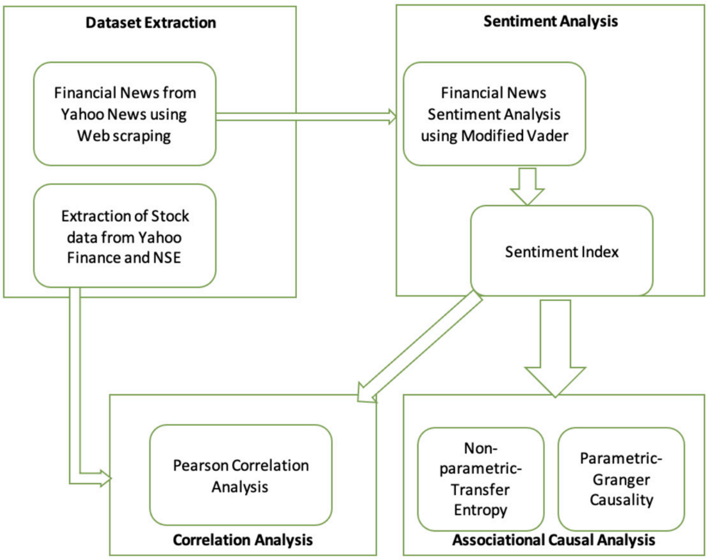

Fig. 1. Methodology.

or web data extraction. Web scraping was done for the news related to 16 companies (under study) which were published in the website of in Yahoo news and money between 2018 and 2022. Web scraping was implemented in Python with Beautiful Soup.

3.2.2. *Sentiment analysis of the financial news using modified VADER*

Sentiment Analysis, also referred to as Opinion Mining, is a specialized domain within Natural Language Processing (NLP) aimed at discerning and extracting opinions from textual data. It serves the purpose of ascertaining the author's sentiments, evaluations, attitudes, and emotions by analyzing the subjectivity inherent in the text. According to various research, changes in stock market prices reflect how the public feels about the companies. Using user sentiment in developing forecasts increased the accuracy of the prediction algorithms by 20%. As a result, information about the company from the media, business reports, user evaluations on social media, or investor sentiment can all shed light on how stock prices fluctuate. There are different NLTK libraries like Pattern, VADER, and textblob to analyze the sentiment of news articles. This work employed the VADER library to perform the sentiment analysis. The VADER (Valence Aware Dictionary and Sentiment Reasoner) is a rule-based sentiment analysis tool that helps determine the emotional tone of a piece of text. It employs a word dictionary with labeled sentiments and applies grammatical rules to gauge the overall sentiment score of the text. A sentiment lexicon comprises lexical properties, like words, which are generally classified as either positive or negative based on their impact on people's emotions. VADER has demonstrated proficiency in evaluating sentiment in social media posts, movie reviews, and product feedback. It not only discerns whether a sentiment is positive or negative, but also quantifies the intensity of that sentiment.

This research presents a refinement of the VADER algorithm for improved performance in analyzing financial news. We integrated the Loughran-McDonald Finance Dictionary to augment VADER's vocabulary, leading to a more tailored approach for financial sentiment analysis. The initial emotion lexicon of VADER comprised 7.5K words and 3.55K emojis. Some of these entries were replaced with terms sourced from the Loughran-McDonald Finance Dictionary. Among the vast dataset of 88K entries, we exclusively considered those clearly marked with a positive or negative sentiment. To enhance the accuracy of sentiment analysis, we implemented five heuristic rules. These rules encompassed checks for punctuation, capitalization, degree modifiers, contrastive conjunctions, and negated sentences. This step allowed us to refine any logical inferences that were deemed insufficient for a specific news scenario.

The Modified VADER algorithm conducted an in-depth analysis of signal polarity, resulting in positive, negative, and neutral scores. The neutral score was derived as the sum of the positive and negative scores. Subsequently, we calculated a compound score, representing a normalized total score. This was achieved by combining the positive and negative scores with a recommended normalization constant (alpha) of 1.2, as determined in this experiment. The compound score assigned to each headline provided a normalized value, ranging from -1 (indicating the most negative headline) to 1 (representing the most positive headline). Following this, we computed the daily score by averaging all the ratings assigned to the various news articles collected on that specific day.

3.2.3. *Sentiment index calculation*

Following the analysis of the datasets and computation of polarity 'P' for each news item, the results were aggregated for each observed day, regardless of the stock market's operational status. The polarity value 'Z' for day 't' was computed by summing the polarity 'P' of news item 'I' as outlined in equation (1).

$$Z_t = \sum_{k=1 \rightarrow i} P_i \tag{1}$$

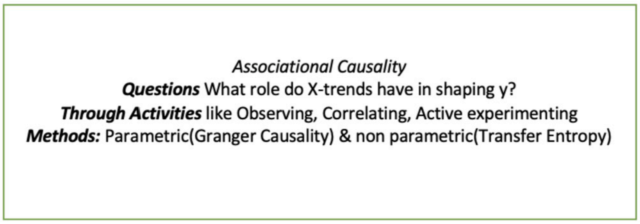

Fig. 2. Overview of Associational Causal Analysis.

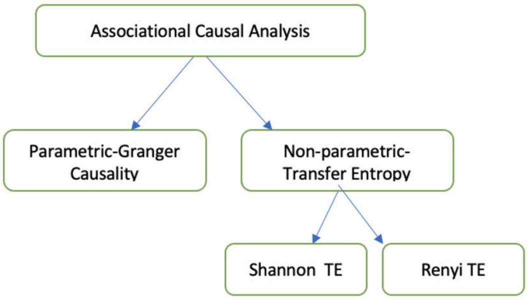

Fig. 3. Methods of Associational Causal Analysis.

$$Z = [Z_t, Z_{t+1}, Z_{t+2}, ...] \tag{2}$$

The change in stock price, or the overnight return 'Y', is determined by subtracting the previous day's closing price from the current day's closing price, as calculated using equation (3). The resulting time series 'Y', representing the stationary values of the stock prices, is defined by equation (4).

$$Y_t = SP_t - SP_{t-1} \tag{3}$$

$$Y = [Y_t, Y_{t+1}, Y_{t+2}, ...] \tag{4}$$

3.2.4. Correlation

The degree and strength between a pair of time series (Financial News price and Stock price change) can be determined using Pearson's correlation. The Pearson correlation is calculated as per equation (5),

$$\rho = \frac{\mathrm{cov}(X, Y)}{\sigma_x \sigma_y} \tag{5}$$

This yields a value between -1 and 1 that quantifies the linear correlation between the two time series, Y and Z. When the Pearson coefficient value is unity, it shows the highest correlation between the time series. The Pearson correlation coefficient is an inferential statistic, which means it may be utilized to evaluate statistical hypotheses. Specifically, we can examine the significance of any correlation between two time series.

3.2.5. Causality analysis

The concept of causality pertains to the interdependent factors that influence one another, and understanding this relationship holds significant importance. Presently, causation is a prominent area of research, particularly in the realm of causal analysis within the financial sphere. It aids in unveiling the cause-and-effect dynamics among various macroeconomic elements in the stock market. Both of these domains are presently witnessing a surge in interest.

In the domain of time series analysis, causality is represented along three levels, as elucidated by reference [40]. These levels range from the fundamental concept of association, progressing to intervention, and culminating in the notion of counterfactual. Association serves as a means to connect events from the past to future occurrences.

Moving beyond association, intervention involves purposefully modifying existing conditions and observing the ensuing outcomes. The highest rung on the Ladder of Causation is occupied by counterfactual reasoning. In this study, we endeavor to establish the causal association between financial news sentiment and stock prices. Figs. 2 and 3 provide an overview of the causal analysis process and the methods employed. At a higher level than association, intervention entails intentionally altering what is present and then watching the result. Counterfactual sit atop the Ladder of Causation's highest rung. In this work the associational causality between the financial news sentiment and the stock price are being tried to establish. The overview of association causal analysis and the methods are depicted in Fig. 2 and Fig. 3.

4. Theory

This section centers around the conceptual framework that measures the cause-effect relationship between financial news sentiment and stock market prices. Each subsection expounds on the theoretical and mathematical foundations of diverse approaches employed to establish associational causal analysis.

4.1. Parametric causality using Granger causality

The study utilized the Granger causality methodology to investigate the structural dynamics of causal relationships among the various variables. Granger causality is a statistical tool crafted to assess the feasibility of utilizing one time series for accurate forecasting of another time series. If the probability value falls below a designated threshold, the hypothesis would be rejected at that level. Different autoregressive models are taken into account to ascertain if process 'X' is a Granger cause of process 'Y', as defined in equations (6) and (7).

$$-Y(t) = \sum_{\infty\to\tau} (a_\tau Y(t-\tau) + \sum_{\infty\to\tau} (c_\tau Y(t-\tau) + \epsilon_c \tag{6}$$

$$Y(t) = \sum_{\infty\to\tau} (b_\tau Y(t-\tau) + \epsilon \tag{7}$$

where t denotes any time instance, a_τ, b_τ, c_τ are coefficients at a time lag of τ, ϵ_c, and ϵ are error terms in the two models.

The log ratio of the prediction error variances is calculated using F-statistic as per equation (8)

$$F_{X\to Y} = \ln \frac{var(\epsilon)}{var(\epsilon_c)} \tag{8}$$

If equation (7) provides a better explanation for Y(t) than equation (8), then $var(\epsilon_c) \leq var(\epsilon)$ and $F_{X\to Y}$ will be greater than zero, indicating that X Granger causes Y. Given the auto-regressive model's broad applicability and its minimal assumptions about the underlying process, it is extensively utilized for data-driven causality assessment in finance and stock market research. However, instead of quantifying the actual information flux between entities, it solely signifies the presence of information flow based on a linear relationship. In order to address the constraints of Granger Causality, the concept of transfer entropy (TE) was introduced by Schreiber [[34] [43]].

4.2. Non parametric causality using transfer entropy

During this phase, the study leveraged non-parametric information theory to evaluate the information flow and its magnitude between the sentiment index of financial news and stock market prices.

4.2.1. Shannon transfer entropy

The Shannon entropy forms the basis for the non parametric concept of transfer entropy, which quantifies the information exchanged between two independent variables.

The objective of transfer entropy (TE) is to quantify the amount of time-oriented information shared by two random processes. Given a coupled system (X,Y), where $P_Y(y)$ is the probability distribution function of the random variable Y and $P_{X,Y}$ is the joint probability distribution function between X and Y, the joint entropy between X and Y is given by equation (9).

$$H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} P_{X,Y}(x,y)\log P_{X,Y}(x,y). \tag{9}$$

The conditional entropy is defined as per the equation (10) and interpreted as the uncertainty of Y given a realization of X.

$$H(Y|X) = H(X,Y) - H(X) \tag{10}$$

The difference between the conditional entropy's is defined as transfer entropy as per equation (11).

$$TE(X\to Y|Z) = H\left(Y^F \middle| Y^P, Z^P\right) - H\left(Y^F \middle| X^P, Y^P, Z^P\right). \tag{11}$$

Due to the asymmetry of the measure, transfer-entropy, it is possible to quantify the directional coupling between systems. The net information flux can be defined as per equation (12)

$$\widehat{TE}_{X\to Y} = TE_{X\to Y} - TE_{Y\to X}. \tag{12}$$

This quantity is a measure of the dominant direction of the information flow. In other words, a good outcome suggests that information is predominantly flowing from X to Y, or that one system gives more predictive information about the other.

The estimations of transfer entropy presented above are typically flawed as a result of the consequences of a small sample size. The effective transfer entropy offers a solution to this problem. The effective transfer entropy can be calculated as equation (13).

$$ET_{X\to Y} = T_{X\to Y} - T_{X_{\text{shuffled}}\to Y} \tag{13}$$

Randomly selecting values from the time series of X and realigning them to form a new time series constitutes shuffling. Consequently, $T_{X_{\text{shuffled}} \rightarrow Y}$ converges to zero as the sample size increases, and any nonzero value is attributable to small sample effects. The estimations of transfer entropy derived from randomized data can therefore be utilized as a proxy for the bias caused by these small sample effects.

To evaluate the statistical significance of transfer entropy estimations, we employ a Markov block bootstrap technique as introduced by Dimpfl and Peter [[45] [46] [47]]. In contrast to randomization, the Markov block bootstrap maintains the interdependence between individual time series. Under the assumption that no information is being exchanged, it produces the distribution of transfer entropy estimates. Finally, the simulated time series are used to make an estimate of the Shannon or Renyi transfer entropy. The transfer entropy estimate distribution for the no-information-flow null model can be obtained by repeating this approach. Where $\hat{q}_{TE}$ is the quantile of the simulated distribution that corresponds to the original transfer entropy estimate, $1 - \hat{q}_{TE}$ is the p-value associated with the null hypothesis of no information transfer.

To ascertain the prevailing direction of information flow, one can validate it based on the bootstrapped distribution of the transfer entropy estimates. This validation involves computing p-values for the effective transfer entropy.

Failing to employ the Markov block bootstrap model for assessing transfer entropy estimations can lead to potential drawbacks in statistical inference. Traditional shuffling methods, for instance, may not adequately account for the dependencies inherent within each time series. This can result in a distorted estimation of the null distribution, potentially leading to erroneous conclusions regarding the presence or absence of information transfer. Moreover, without the Markov block bootstrap, there is a risk of overlooking critical nuances in the interplay between processes J and I. This may result in an underestimation of statistical dependencies, potentially masking genuine information flow between the variables. Therefore, not utilizing the Markov block bootstrap model may introduce a level of uncertainty and imprecision in the assessment of transfer entropy, potentially compromising the accuracy and reliability of the findings.

4.2.2. Renyi's transfer entropy

Alfréd Rényi, a Hungarian mathematician, devised Rényi entropy, which generalizes Shannon entropy. If a discrete random variable Y has m possible values, where the ith outcome has probability pi, then the Rényi entropy of order α is defined in equation (14).

$$H_J^{\alpha} = \frac{1}{1-\alpha} log \left(\sum_j p^{\alpha}(j) \right) \tag{14}$$

for $0 \leq \alpha \leq \infty$. In the case $\alpha = 1$ or this expression means the limit as ∞ approaches 1 or ∞ respectively. Since different parts of a distribution can be highlighted depending on order of α, Rényi entropy gives a more versatile tool for estimating uncertainty. The Rényi transfer entropy is a metric for gauging the amount of data passing from source J to destination I [41]. Rényi transfer entropy calculation may yield negative values, in contrast to the Shannon transfer entropy calculation, which always returns a positive number. Marchinski [42] established the notion of Effective Transfer Entropy to overcome the difficulty in calculating the right entropy from a small sample size is given in equation (15).

$$ETE_{X \rightarrow Y} = TE^R_{\alpha, X \rightarrow Y} - TE^R_{\alpha, X_S huffled \rightarrow Y} . \tag{15}$$

5. Results

This research introduces a robust metric for assessing and quantifying the influence of financial news on stock performance. Moreover, it provides an effective means of measuring the impact of financial news sentiment on stock performance. In this study, we conducted a statistical analysis to ascertain the linear relationship between stock price and news sentiment using Pearson correlation. To further quantify the relationship and the direction of information flow between the time series, we applied both parametric and non-parametric approaches, namely Granger Causality and Transfer Entropy.

Fig. 4 illustrates a comparison between the financial news sentiment of Cipla and the corresponding change in stock price. Leading up to the COVID-19 pandemic, both the frequency of mentions of Cipla in the news and the fluctuation in stock price remained relatively low. However, during the peak of the pandemic, there was a notable surge in news mentions of Cipla, particularly in connection with COVID-19 vaccines, as depicted in Fig. 4.

5.1. Pearson correlation

The statistical analysis to find the linear relationship between the stock price and the news sentiment using Pearson correlation was done and the results are being shown in Table 3.

We observed auto-correlation between the sentiment index and stock price changes, but found little to no correlation between stock price changes and financial news sentiment indexes. In the Indian stock market, a discernible linear correlation between stock price and news sentiment is not evident.

To assess the relationship, direction, and strength of information flow, we applied associational causal analysis and employed the methods of Granger causality and transfer entropy. The hypothesis under examination posits that in this scenario, stock price movements are influenced not only by previous stock prices but also by the sentiment in financial news. Conversely, the "null

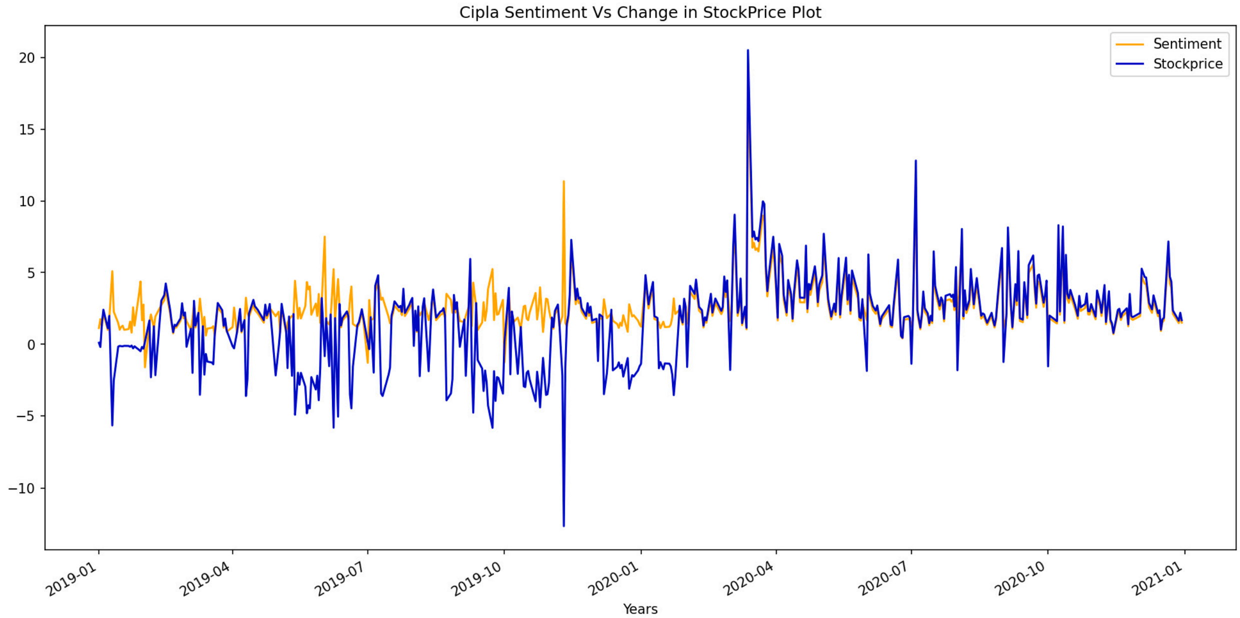

Fig. 4. Cipla News Sentiment vs Change in Stock Price.

Table 3
Correlation between News Sentiment and Stock Price.

Company Name	Correlation Value
Tata Consultancy Services	0.3728
Tech Mahindra	0.6542
Infosys Ltd	0.0692
Wipro Ltd	0.2669
HCL technologies	0.0896
State Bank of India	0.1311
HDFC Bank Limited	0.2700
Indusind Bank Ltd	0.0778
Tata Steel Limited	0.2211
ITC Limited	0.2719
Tata Motors	0.3090
Asian Paints Ltd	0.0271
Bharat Petroleum Corporation Ltd	0.0612
Cipla Ltd	0.0223
Dr Reddy's Laboratories Ltd	0.0752
Sun Pharmaceutical Industries Ltd	0.0485

hypothesis" suggests that changes in stock prices are not solely contingent on past stock values, but also on sentiment. Table 4 displays the results of our Granger causality assessments for stock price returns and news sentiment. It indicates, at a 5% level of significance, that shifts in stock prices precede the public's responses to news.

5.2. Shannon & Renyi transfer entropy results

The outcomes of the non-parametric approach employing Shannon Transfer Entropy and Renyi Transfer Entropy are presented in Table 5 and Table 6, respectively. To conduct an analysis of the information flow, we designated the variable 'X' to the Sentiment Index vector and 'Y' to the Stock Prices vector. The strength of the signal was then estimated as per equation (16).

Using this expression, if the information signal exhibited greater strength from $(X \rightarrow Y)$, it would assume a positive sign, indicating that information predominantly flows from financial news emotion to the stock market, which aligns with the primary focus of our investigation and the rationale for our study. Conversely, if the signal from $(Y \rightarrow X)$ held greater strength, the opposite sign would be anticipated, suggesting that activity in the stock market is influencing the activity in the news.

$$Strength_{InformationFlux} = \frac{ETE_{X\rightarrow Y}}{(ETE_{Y\rightarrow X} + ETE_{X\rightarrow Y})} \tag{16}$$

The null hypothesis posits that there exists no discernible information flow between news sentiment and the stock price. However, upon conducting the Markov bootstrapped transfer entropy analysis, we observed p-values less than 0.01, 0.05, and 0.1. These

Table 4
Granger Test Result.

Company Name	F value	P value
TCS	4.3302	0.03874
TECHM	0.2425	0.06375
INFY	1.875	0.02615
WIPRO	1.9849	0.08202
HCLTECH	3.5971	0.01061
SBIN	0.2355	0.639
HDFCBANK	2.977	0.09038
INDUSINDBANK	0.9704	0.3626
TATASTEEL	6.7183	0.01039
ITC	1.3668	0.3397
TATA MOTORS	18.519	0.0002866
ASIANPAINT	1.4541	0.2291
BPCL	2.62277	0.3519
CIPLA	0.1311	0.7208
DRREDDY	10.318	0.08478
SUNPHARMA	0.8552	0.4233

Table 5
Shannon Transfer Entropy Results.

Company	TE X → Y	Eff. TE X → Y	P values X → Y	TE Y → X	Eff. TE Y → X	P values Y → X	Strength
TCS	0.0562	0.0363	0.0167	0.0207	0.0019	0.3367	0.640
TECHM	0.2677	0.1564	0.0433	0.0838	0.1167	0.4300	0.6109
INFY	0.0517	0.0308	0.0467	0.0207	0.0004	0.3933	0.9871
WIPRO	0.0503	0.0211	0.011	0.0369	0.0161	0.001	0.5672
HCL	0.0562	0.0186	0.001	0.0458	0.012	0.04	0.6072
SBI	0.0251	0.0036	0.1033	0.0342	0.008	0.0367	0.6896
HDFC	0.1079	0.0747	0.0433	0.0093	0.0121	0.0404	0.8605
INDUSBK	0.3118	0.2685	0.0367	0.0117	0.11	0.5267	0.7093
TATASTEEL	0.0479	0.0182	0.0033	0.0366	0.0135	0.02	0.574
ITC	0.0278	0.077	0.209	0.00267	0.1722	0.061	0.31
TATAMOTORS	0.0331	0.0106	0.0067	0.317	0.0086	0.033	0.522
ASIAN	0.0078	0	0.0067	0.0047	0	0.67	NA
BPCL	0.0230	0	0.001	0.01	0	0.6033	NA
CIPLA	0.0547	0	0.293	0.0231	0	0.4867	NA
DRREDDY	0.104	0	0.5567	0.011	0	0.27	NA
SUNPHARMA	0.0388	0	0.2800	0.0207	0	0.5967	NA

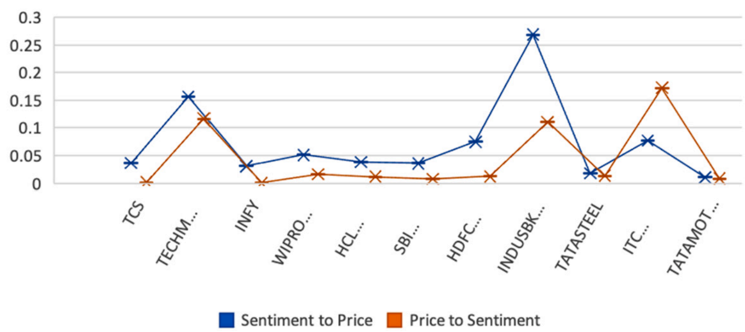

Fig. 5. Information Flux using Transfer Entropy.

thresholds signify a rejection of the null hypothesis, indicating the presence of significant information flows at the 1%, 5%, and 10% significance levels, respectively.

During the computation of Shannon transfer entropy between financial news sentiment and changes in stock prices, it was observed that all companies within the banking and technology sectors exhibited an information flow from the financial news index to the stock price changes. Notably, this information flow was most pronounced for Infosys. However, when employing Shannon's transfer entropy, certain corporations such as Asian Paints, Sunpharma, and BPCL did not demonstrate any discernible information flow. In other words, the p-value exceeded 0.10 in both directions. Subsequently, these companies underwent further analysis using Renyi transfer entropy, which revealed a significant information flow between these indices.

Table 6
Renyi Transfer Entropy Results.

Company	TE X → Y	Eff. TE X → Y	P values X → Y	TE Y → X	Eff. TE Y → X	P values Y → X	Strength
ASIAN	0.26622	0.0686	0.0200	0.1243	0.0261	0.7800	0.724
BPCL	0.1798	0.0629	0.730	0.2459	0.0472,	0.15,	0.5712
CIPLA	0.3662	0.0830	0.0033	0.2185	0.0465	0.6767	0.6409
DRREDDY	0.1759	0.0312	0.0700	0.1200	0.0308	0.633	0.5032
SUNPHARMA	0.1780	-0.0016	0.2667	0.1452	-0.0017	0.4900	0.4848

The findings from transfer entropy analysis reveal a bidirectional flow of information. As depicted in Fig. 5, for the majority of equities, the information flow from the sentiment of financial news to changes in stock prices is more pronounced than the reverse flow.

The Shannon transfer entropy algorithm proves particularly effective with extensive datasets, especially for corporations frequently mentioned in the media, such as financial institutions and technology-based businesses. On the other hand, for companies with less frequent media coverage, Renyi Transfer Entropy emerges as the preferred tool for uncovering information flow, as demonstrated in Table 5.

This experiment was conducted within the pharmaceutical industry to evaluate the impact of news during the COVID-19 pandemic. The datasets for Cipla and Dr. Reddy were partitioned into two segments: one preceding the onset of COVID-19, and the other comprising data from the pandemic period. Prior to the COVID-19 outbreak, the Shannon entropy results for both Cipla and Dr. Reddy were relatively weaker due to fewer mentions of these companies.

During the COVID-19 period, when there was a surge in mentions of the companies, Renyi transfer entropy was employed. This approach effectively captured the information flow between these indices despite the smaller dataset. Our findings suggest that Renyi transfer entropy demonstrates relative accuracy for small datasets and remains applicable for high-dimensional data. However, it is worth noting that it comes with the drawback of higher algorithmic processing costs.

6. Conclusions and future work

In this study, we delve into the causal connections between financial news sentiment and stock market values, aiming to unravel the reciprocal influence between these variables. Our investigation focuses on a diverse set of 16 companies spanning various sectors. Employing a combination of parametric and non-parametric approaches, we scrutinize the information flows between sentiment and market movements. Through the application of various information theory techniques, including a nonlinear methodology, we aim to enhance our capacity to capture causality, going beyond the conventional Granger approach. All of the companies and their respective Sentiment Indices were simulated by the usage of the Modified Vader Algorithm. Index creation model through a variety of statistical tests were also implemented and it outperformed conventional approaches.

Transfer entropy is a method for non parametric and associative causal analysis capable of quantifying the extent of information transmission within a time series, along with discerning its directional flow. The findings from the transfer entropy analysis demonstrate that there is an asymmetric information flow between sentiment and stock prices. Specifically, the information flows more strongly from sentiment to price, as quantified by the transfer entropy method. This indicates that changes in sentiment have a notable influence on stock prices, highlighting the significant impact of sentiment in financial markets. This approach supplants the Granger test, which is applicable solely to Gaussian time series exhibiting linear causation, and provides a superior tool for examining nonlinear causality. An advantageous feature of employing this specific formulation lies in the ability of Rényi's entropy, in conjunction with the specified parameter, to discern informational flow, particularly during rapid and substantial events such as sudden leaps. This is due to the fact that Rényi's entropy places more weight on the tails of the probability distribution than it does on its own center.

The use of modified VADER algorithm, specifically trained on financial data, offers several advantages over the classic VADER algorithm. Firstly, it allows effective sentiment analysis in the financial domain. This specialized training has tailored the algorithm to understand the unique language and context used in financial news and reports, which can be quite distinct from general text. Furthermore, the modified VADER algorithm reduces the number of cases where a neutral sentiment is assigned. This is significant in finance, where nuanced sentiment analysis is crucial for making informed investment decisions. By providing a sentiment score that leans more towards either positive or negative, rather than neutral, the modified algorithm offers a more actionable assessment of sentiment for financial professionals. Overall, the modified VADER algorithm, due to its domain-specific training, proves to be a more powerful tool for sentiment analysis in the financial sector, improves the accuracy and effectiveness of the analysis than the classic VADER algorithm used by Varghese [44].

The research highlights a key limitation in the context of associational causality, emphasizing the foundation of cause and effect separability among samples. When there is overlap between these samples, it introduces drawbacks that can potentially be resolved through the application of a non-parametric sliding window approach. Additionally, the study reveals an interesting observation regarding the interplay of Sentiment Indices across multiple companies, indicating that comments about one company can indirectly impact the stock prices of others. This suggests the potential for future research to explore the relationships between sectors in greater detail.

As a limitation of this research, it is worth noting that while the Modified Vader Algorithm enhances sentiment analysis in finance, there remains room for improvement. A forthcoming project is dedicated to the development of a sentiment algorithm tailored specifically to finance-related terminology. This endeavor holds the potential to yield even more compelling discoveries and practical advantages for participants in the stock market. It is duly acknowledged that stock prices are influenced by a multitude of factors, with news sentiment representing just one facet of this intricate puzzle. In our upcoming research, which centers on causal analysis within the equity market, we intend to conduct a comprehensive examination encompassing a diverse array of input variables, including macroeconomic elements. This will facilitate a comparative evaluation of the degree to which news sentiment contributes to the influence on share prices.

In summation, this research equips investors with invaluable instruments for making judicious decisions in the financial market. Through the incorporation of sentiment analysis into their strategies, investors stand to refine their market timing, fortify their risk management practices, and optimize the diversification of their portfolios. Furthermore, this comprehension of sentiment's influence aids investors in mitigating behavioral biases and aligning their risk tolerance with more efficacious investment methodologies. On the whole, this study augments our nuanced comprehension of the correlation between sentiment and Indian stock prices, providing invaluable insights that can redound to the benefit of investors, policymakers, and scholars engaged in the Indian financial market.

CRediT authorship contribution statement

Renju Rachel Varghese: Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Biju R. Mohan:** Writing – review & editing, Validation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

The financial news information regarding various equity firms was sourced from https://finance.yahoo.com/. The dataset utilized for this research is accessible in the repository at https://github.com/renjuvarghese/Fintech-dataset. Stock prices were obtained from both https://www.nseindia.com/ and https://finance.yahoo.com/quote/tcs/history/ for comprehensive coverage.

Acknowledgement

I extend my heartfelt gratitude to Dr. Jacob T. Varghese, Dr. Chacko Preno Koshy and Mr. Siju K S from Saintgits College of Engineering, Kottayam, for their unwavering support and invaluable guidance throughout the entire process of preparing and reviewing this manuscript. Their expertise and insights greatly enriched the quality of this work.

Appendix A

Table 7, provided in this appendix, shows exemplification of data pertaining to a company in the technological sector.

Table 7
Exemplification of Data Pertaining to a Company in the Technological Sector.

Date	News	Open	High	Low	Close	Volume	Sentiment
06/01/22	TCS Q3 PAT seen up 3.3% QoQ to Rs. 9941.2 cr: ICICI Direct	3812	3835	3772	3807.45	1810293	0.856
07/01/22	Tata Consultancy Services renews contract for Passport Seva Programme with government	3820	3864.90	3796.40	3853.50	2460591	0.251
10/01/22	What Kind Of Shareholders Hold The Majority In The Container Store(TCS)Group Inc.'s (NYSE:TCS) Shares?	3978	3978	3861	3879.85	393709	0.12
11/01/22	Tata Consultancy Services Q3 PAT to Rs. 9899 cr: Arihant Capital	3856	3925	3856	3915	1906106	0.781
12/01/22	Are These Consumer Discretionary Stocks Undervalued Right Now?	3925	3929	3836.55	3859.90	3203744	0.02
13/01/22	Buy TCS. TCS Consolidated December 2021 Net Sales at Rs 48885.00 crore up 16.35% Y-o-Y	3918	3923	3857	3897.90	6684507	0.891
14/01/22	TCS Standalone December 2021 Net Sales at Rs 40845.00 crore up 16.84% Y-o-Y. Accumulate Tata Consultancy Services	3877.85	3977	3860.05	3968.15	334812	0.752
18/01/22	Buy Tata Consultancy Services	4033.95	4041	3980	3990.60	2389041	0.673
25/01/22	The Container Store(TCS)Group Inc. Announces Third Quarter Fiscal 2021 Earnings Conference Call. Earnings Preview: Container Store(TCS)Group (TCS) Q3 Earnings Expected to Decline	3769.5	3809.40	3722.20	3769.90	3330501	-0.128

References

[1] D. Phan Tran Trung, H. Pham Quang, Adaptive market hypothesis: evidence from the Vietnamese stock market, J. Financ. Risk Manag. 12 (2) (2019) 81.
[2] T.O. Dibiah, O.R. Mojekwu, Test of random walk on selected stock markets in Africa, GPH Int. J. Bus. Manag. 6 (07) (2023) 01.
[3] B.G. Malkiel, Efficient market hypothesis, Finance (1989) 127–134.

[4] S.J. Brown, The efficient market hypothesis, the financial analysts journal, and the professional status of investment management, Financ. Anal. J. 76 (2) (2020) 5–14.
[5] A.K. Maan, A. Shiva, Mapping social media in the theoretical field of behavioural finance: a bibliographic analysis, Int. J. Electron. Finance 12 (1) (2023) 18–35.
[6] B. Pant, T.R. Bishnoi, Testing random walk hypothesis for Indian stock market indices, Research paper presented in: UTI Capital Market Conference Proceedings, 2001, pp. 1–15.
[7] W.S. Chan, Stock price reaction to news and no-news: drift and reversal after headlines, J. Financ. Econ. 70 (2) (2003) 223–260.
[8] Q. Li, T. Wang, P. Li, L. Liu, Q. Gong, Y. Chen, The effect of news and public mood on stock movements, Inf. Sci. 278 (2014) 826–840.
[9] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, J. Comput. Sci. 2 (1) (2011) 1–8.
[10] T.T. Souza, T. Aste, Predicting future stock market structure by combining social and financial network information, Phys. A, Stat. Mech. Appl. 535 (2019) 122343.
[11] Doron Kliger, Andrey Kudryavtsev, The availability heuristic and investors' reaction to company-specific events, J. Behav. Finance 11 (1) (2010) 50–65, https://doi.org/10.1080/15427561003591116.
[12] J. Kalyani, P. Bharathi, P. Jyothi, Stock trend prediction using news sentiment analysis, arXiv preprint, arXiv:1607.01958, 2016.
[13] N. Das, B. Sadhukhan, T. Chatterjee, S. Chakrabarti, Effect of public sentiment on stock market movement prediction during the COVID-19 outbreak, Soc. Netw. Anal. Min. 12 (1) (2022) 92.
[14] F. Fang, C. Ventre, M. Basios, L. Kanthan, D. Martinez-Rego, F. Wu, L. Li, Cryptocurrency trading: a comprehensive survey, Financ. Innov. 8 (1) (2022) 1–59.
[15] G. Figà-Talamanca, M. Patacca, An explorative analysis of sentiment impact on S&P 500 components returns, volatility and downside risk, Ann. Oper. Res. (2022) 1–23.
[16] A. Alqahtani, M.J. Wither, Z. Dong, K.R. Goodwin, Impact of news-based equity market volatility on international stock markets, J. Appl. Econ. 23 (1) (2020) 224–234.
[17] J. Ren, H. Dong, A. Popovic, G. Sabnis, J. Nickerson, Digital platforms in the news industry: how social media platforms impact traditional media news viewership, Eur. J. Inf. Syst. (2022) 1–18.
[18] B. Yemba, Y. Duan, N. Biswas, Government Spending News and Stock Price Index, 2023, available at SSRN 4538939.
[19] B. Cui, P. Docherty, Stock price overreaction to ESG controversies, 2020, available at SSRN 3559915.
[20] M. Vicari, M. Gaspari, Analysis of news sentiments using natural language processing and deep learning, AI Soc. 36 (2021) 931–937.
[21] A. Yadav, C.K. Jha, A. Sharan, V. Vaish, Sentiment analysis of financial news using unsupervised approach, Proc. Comput. Sci. 167 (2020) 589–598.
[22] L. Nemes, A. Kiss, Prediction of stock values changes using sentiment analysis of stock news headlines, J. Telecommun. Inf. Technol. 5 (3) (2021) 375–394.
[23] J. Kim, H.S. Kim, S.Y. Choi, Forecasting the S&P 500 index using mathematical-based sentiment analysis and deep learning models: a FinBERT transformer model and LSTM, Axioms 12 (9) (2023) 835.
[24] S. Peng, W. Han, G. Jia, Pearson correlation and transfer entropy in the Chinese stock market with time delay, Data Sci. Manag. 5 (3) (2022) 117–123.
[25] L.P. Huang, T. Qiu, G. Chen, L.X. Zhong, European Union effect on financial correlation dynamics, Phys. A, Stat. Mech. Appl. 528 (2019) 121457.
[26] A. Shojaie, E.B. Fox, Granger causality: a review and recent advances, Annu. Rev. Stat. Appl. 9 (2022) 289–319.
[27] A.R. Nogueira, A. Pugnana, S. Ruggieri, D. Pedreschi, J. Gama, Methods and tools for causal discovery and causal inference, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 12 (2) (2022) e1449.
[28] A. Seth, Granger causality, Scholarpedia 2 (7) (2007) 1667.
[29] A. Shojaie, E.B. Fox, Granger causality: a review and recent advances, Annu. Rev. Stat. Appl. 9 (2022) 289–319.
[30] B.L. Tran, C.C. Chen, W.C. Tseng, Causality between energy consumption and economic growth in the presence of GDP threshold effect: evidence from OECD countries, Energy 251 (2022) 123902.
[31] J. Dong, W. Dai, J. Li, Exploring the linear and nonlinear causality between internet big data and stock markets, J. Syst. Sci. Complex. 33 (3) (2020) 783–798.
[32] S.Y. Phoong, C.Z. Lim, S.W. Phoong, A Granger causality analysis between stock prices and exchange rates:: evidence from four countries, Int. J. Comput. Sci. Math. 17 (3) (2023) 284–294.
[33] Z. He, Dynamic impacts of crude oil price on Chinese investor sentiment: nonlinear causality and time-varying effect, Int. Rev. Econ. Finance 66 (2020) 131–153.
[34] T. Schreiber, Measuring information transfer, Phys. Rev. Lett. 85 (2) (2000) 461.
[35] A. Kaiser, T. Schreiber, Information transfer in continuous processes, Phys. D: Nonlinear Phenom. 166 (1–2) (2002) 43–62.
[36] S. Kim, S. Ku, W. Chang, J.W. Song, Predicting the direction of US stock prices using effective transfer entropy and machine learning techniques, IEEE Access 8 (2020) 111660–111682.
[37] P. Ferreira, A. Dionísio, D. Almeida, D. Quintino, F. Aslam, A new vision about the influence of major stock markets in CEEC indices: a bidirectional dynamic analysis using transfer entropy, Post-Communist Econ. 34 (2) (2022) 267–282.
[38] L. Qiu, H. Yang, Transfer entropy calculation for short time sequences with application to stock markets, Phys. A, Stat. Mech. Appl. 559 (2020) 125121.
[39] P. Yue, Q. Cai, W. Yan, W.X. Zhou, Information flow networks of Chinese stock market sectors, IEEE Access 8 (2020) 13066–13077.
[40] J. Pearl, The art and science of cause and effect, in: Shaping Entrepreneurship Research, Routledge, 2020, pp. 446–474.
[41] F.C. Thierrin, F. Alajaji, T. Linder, Rényi cross-entropy measures for common distributions and processes with memory, Entropy 24 (10) (2022) 1417.
[42] R. Marschinski, H. Kantz, Analysing the information flow between financial time series: an improved estimator for transfer entropy, Eur. Phys. J. B, Condens. Matter Complex Syst. 30 (2002) 275–281.
[43] S. Behrendt, T. Dimpfl, F.J. Peter, D.J. Zimmermann, RTransferEntropy—quantifying information flow between different time series using effective transfer entropy, SoftwareX 10 (2019) 100265.
[44] R.R. Varghese, B.R. Mohan, The causal effect of financial news on Indian stock market, in: TENCON 2022 - 2022 IEEE Region 10 Conference (TENCON), Hong Kong, 2022, pp. 1–5.
[45] T. Dimpfl, F.J. Peter, Using transfer entropy to measure information flows between financial markets, Stud. Nonlinear Dyn. Econom. 17 (1) (2013) 85–102.
[46] A. Bossman, Information flow from COVID-19 pandemic to Islamic and conventional equities: an ICEEMDAN-induced transfer entropy analysis, Complexity 2021 (2021) 1–20.
[47] T.L.D. Huynh, M.A. Nasir, X.V. Vo, T.T. Nguyen, "Small things matter most": the spillover effects in the cryptocurrency market and gold as a silver bullet, N. Am. J. Econ. Finance 54 (2020) 101277.