Noname manuscript No.
(will be inserted by the editor)

Forecasting Conditional Volatility of Indian Stock Market using Public Emotions

Priyanka Verma · Rajendra N. Paramanik

Received: date / Accepted: date

Abstract This paper investigates the nexus between public emotions and Indian stock market risks. The emotional intelligence of the average investor plays an important role in the overall behaviour of the market. Eight emotions namely Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise and Trust have been extracted from news articles from financial news in India. The impact of these emotions is studied on time varying conditional volatility of NIFTY and SENSEX. Our results indicate that crowd emotions of joy, anticipation, and sadness, derived from stock markets related news, are significant predictors of stock market risks for Indian markets.

Keywords LSTM · Human Emotions · Indian Stock Market Risk · Conditional Volatility

1 Introduction

Stock market behavior is mostly complex (non-linear) and volatile and has been a field of interest to many researchers. The Efficient Market Hypothesis (EMH) states that stock market prices are largely driven by new information and follow a random walk pattern. However, several attempts have been made to extract patterns in the way stock markets behave and respond to external stimuli.

Financial risk in any market can be analysed from financial volatility. The concept of financial volatility is also a required parameter for pricing many kinds of financial assets and derivatives and is critical. Financial volatility has been well studied for developed markets, however, lesser attention has been paid towards an extensive study of the volatility in the emerging stock market of India.

Through past studies [1], it has been shown that the reasons for stock market anomalies included the behaviour of investors like overreaction, overconfidence, over estimation, less sophistication level of the investor and the biased behaviour of the investors. The emotional intelligence of the average investor plays a very significant role in the overall behaviour of the crowd[1].

In this paper, we test this hypothesis based on behavioral economics that emotions of investors affect stock markets behaviour. Through our study we investigate the dependence of time-varying risks in Indian stock market and human emotions, which has not yet been studied to the best of our knowledge. We believe that emotions would be able to map any policy shocks or macroeconomic uncertainty to give good forecast of risks in financial markets. Conditional volatility has been used as a measure of time-varying risks in financial markets. This study also presents a rich database of financial news headlines in India from April 2007 to March 2020, for future research work.

The rest of the paper is organized as follows. Section 2 presents a literature review of the related research. Section 3 presents data and methodology for empirical analysis and section 4 compares the proposed models and concludes the paper.

Priyanka Verma
Tel.: +91-7888738908
E-mail: vpriyanka0492@gmail.com

Rajendra N. Paramanik
IIT Patna : rajendra@iitp.ac.in

2 Related Work

Various noteworthy past studies have analysed relation between emotions and stock market behaviour in developed countries. Through study [4], it was found that large-scale collective emotions (representing public moods) on Twitter are correlated with the volatility of Dow Jones Industrial Average (DJIA). A number of studies have analyzed texts from social network services (SNS), blogs and news to analyze correlations between stock prices and public emotion as reaction to social events and news [4, 5, 6, 7, 8]. Most of these works analyzed by categorizing public emotions are only as positive or negative.

It has been shown [2, 3] that news titles are more useful for prediction compared to news contents. Analysing news-trend patterns using headlines alone can significantly improve classification performance compared to results using the whole news body or sentence-level analysis [2]. Headlines are one of the earliest forms in which information arrives into the markets [3].

Most of the studies consider stock market behavior in developed countries and there are few noteworthy studies for Indian financial markets. The research [9] studied the effects of good and bad news on volatility of the BSE 500 index using EGARCH and TGARCH models. They concluded that arrival of bad news in the market would result in the volatility to increase more than good news. Another study [10], extracted positive and negative sentiments from news headlines for Stock Price Prediction of two Indian companies Infosys and Wipro. In a similar study [11], showed that Calmness and Happiness are the two Granger Causality parameters affecting the BSE values over a period of 3–4 days.

We find that within the studies done for Indian markets, a minimal number of papers analyse time varying conditional volatility. In study [12], forecasting models for conditional volatility in Indian capital markets were studied. However, the authors did not take into account the impact of financial news, human emotions and neural networks.

Study [13] presented a novel approach to determine public emotions as a numeric score by applying emotion analysis to daily news articles. This work considered all types of news articles to study changes in the trading volume and the closing price of a KOrea composite Stock Price Index(KOSPI). However, this study was done for the developed market of and did not measure conditional volatility.

Considering the above review of literature, we find that none of the studies analysed the impact of different types of emotions on composite stock market index of NIFTY and SENSEX for forecasting conditional volatility. Second, previous studies have investigated stock market returns, whereas the emphasis of our study is on financial risks, studied through time-varying conditional volatility. Also, the majority of work considers sentiment analysis only for twitter data, without taking into account daily financial news. Most of the previous work is done for developed countries. Very few studies for developing country of India have used the recent neural networks. Our paper takes into account the daily financial news of India from 2007 to 2020 for a period of 13 years and from 8 different sources including twitter. Additionally, our study deal with two major structural breaks in stock index series because of 2008 Financial Crisis and COVID-19. Such a study is novel in Indian markets context to the best of our knowledge.

3 Data and Methodology

3.1 Data

Daily news headlines have been collected from financial sections of ‘The Hindu Business Line’, ’MoneyControl’, ‘Twitter’, ‘Bloomberg’, ‘Livemint.com’, ‘in.reuters.com’, and ‘moneyworks4me.com’ from 2007-04-19 to 2020-03-16. Also, for the same period NIFTY and SENSEX daily closing prices have been taken. Through these closing prices conditional volatility has been calculated using a simple GARCH(p,q) model where p and q are each 1.

3.2 Methodology

Sentiment Analysis We combined news headlines and twitter data for each day and cleaned them remove trivial words. We used NRC Emotion Lexicon [16, 17] to extract public emotions information from the headlines of the new articles. The lexicon associates each word to the relevant emotions amongst the 8 standard emotions of Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise and Trust. E.g., ’accident’ is associated to ’Fear’, ’Sadness’ and ’Surprise’. For the daily combined news, scores values for each of the eight standard emotions is calculated by lexicon by multiplying the number of words mapped to that emotion and weight of that emotion in lexicon.

Conditional Volatility Estimation Simple GARCH(1,1) model has been used to calculate conditional volatility from daily returns of BSE and SENSEX. GARCH(1,1) models are favored over other stochastic volatility models by many economists due to their relatively simple implementation: since they are given by stochastic difference equations in discrete time, the likelihood function is easier to handle than continuous-time models, and since financial data is generally gathered at discrete intervals.

GARCH(p,q) model can be represented as

$$X_t = \sigma_t Z_t \tag{1}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i X_{t-1}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{2}$$

Where $\alpha_0 > 0$ and $\alpha_i \geq 0, i = 1, ..., q$ and $\beta_j \geq 0, j = 1, ..., p$.

Pre-processing Data rows with missing values due to absence of BSE and SENSEX data for weekends and other holidays was dropped off. Then remaining data was split into training and test set. The test set contained 30 days data from 1 Feb,2020 to 16 March, 2020. Training data was scaled in the range of [0,1] using *MinMaxScaler()* from *Scikit-learn* library in Python. Scaling will make sure the variance of the features are in the same range.

Compiled data was then used to train various learning algorithms for forecasting conditional volatility for next day as shown in the figure 1. Comparison is carried out between deep learning based LSTM model and traditional time series models using Root Mean Square Error(RMSE). RMSE scores were calculated using forecasted values and actual values from test data for a period of 30-days. RMSE is used for error calculation as it punishes large errors and results in a score that is in the same units as the forecast data, namely conditional volatility.

$$RMSE = \sqrt{\frac{1}{n} \Sigma_{i=1}^{n} \Big(\hat{y}_i - y_i\Big)^2} \tag{3}$$

4 Causal Analysis

In order to ascertain whether the emotions scores can be used to predict the future stock movements, we computed the p-values using Granger Causality analysis. Granger Causality analysis finds how much predictive information one signal has about another over a given lag period. The p-value measures the statistical significance of our result i.e. how likely we could obtain the causality value by random chance; therefore, lower the p-value, higher the predictive ability.

The individual time series were made stationary before computing the values. The analysis was made from lag of 0 to 12 days for each emotion Granger causing conditional volatility. The minimum p-values from all lags are reported in Table 2. It is clear from Table 1 that Anticipation, Joy and Sadness are most helpful in predicting the closing values of NIFTY as per Granger causality analysis.

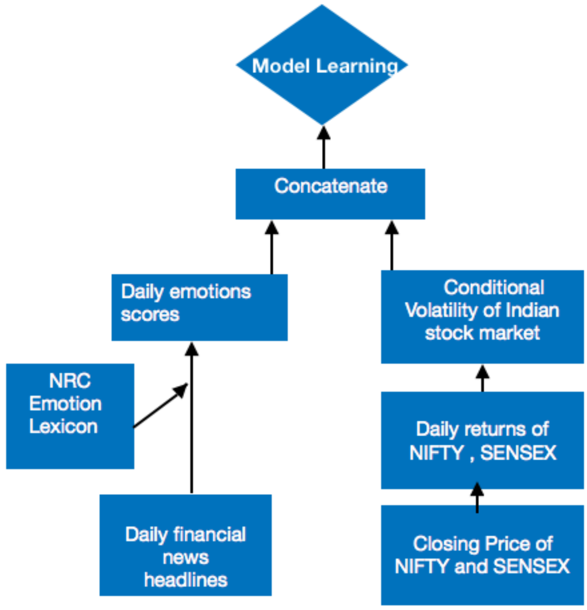

Fig. 1 Our technique

Table 1 p-values for Granger causality analysis

Emotion	p-value
Anger	0.8890
Anticipation	0.2028
Disgust	0.8374
Fear	0.4444
Joy	0.1436
Sadness	0.3432
Surprise	0.5218
Trust	0.5635

Granger causality analysis helps us in understanding that certain emotions do cause changes in stock market behaviour. However, it is based on linear regression, whereas the correlation between stocks and emotions is certainly non linear. Therefore, other learning algorithms are used for further analysis.

5 Model Learning and Prediction

Neural networks have been considered to be a very effective learning algorithm for decoding nonlinear time series data, and financial markets often follow nonlinear trends [18]. Long Short Term Memory networks– LSTM is a special kind of Recurrent Neural Networks, capable of learning long-term dependencies. We tried variations of LSTM to learn and study the actual dependencies.

In order to measure training accuracy, a rolling-window validation was used. The direct k-fold cross validation method is not applicable for temporal data because sequence of observations need to be maintained.

Table 2 RMSE Scores using different Algorithms

Model	30-day window	15-day window	7-day window
(A) Univariate vanilla LSTM model [No emotions]	1.16140	1.35121	1.29592
(B) Multi-variate vanilla LSTM model [Sadness, Joy, Anticipation]	1.20496	1.19208	1.26801
(C) Multi-variate vanilla LSTM model [All 8 Emotions]	1.70928	1.50326	1.61543
(D) A-LSTM [All 8 Emotions]	2.19849	2.06860	2.1378

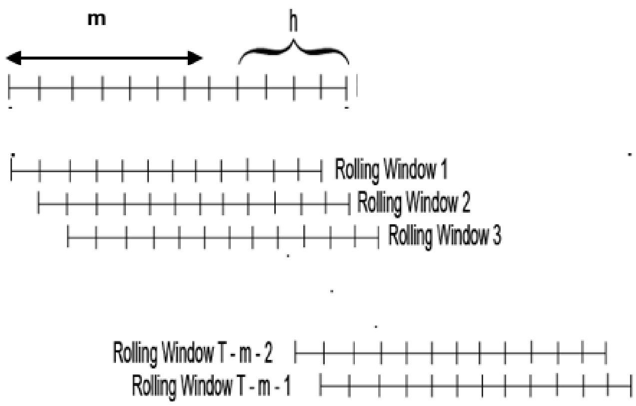

Fig. 2 Rolling window Analysis[15]

Model is trained for rolling window of size of 'm' days and a forecast horizon of 'h' days. The window is incremented by 1 day and the actual values of next days is made available to the model to train on next window as shown in Fig[2].

Our analysis of Granger Causality indicates that 'Sadness', 'Anticipation' and 'Joy' would be most causative of conditional volatility. However, all other emotions also have been considered as features. This is because certain emotions, which may not directly Granger cause conditional volatility, Granger cause other emotions. E.g., through our computation, we noticed that 'Disgust' and 'Fear' Granger cause 'Sadness' with p-values of 0.16 and 0.0007, respectively and 'Disgust' Granger causes 'Anger' with value 0.0002.

Table 2 denotes the RMSE scores for various algorithms when 'h'(forecast window)= 1 day and 'm'(rolling window of training data)= 30 days, 15 days or 7 days. The reported RMSE score has been calculated for each day forecast on test data of 30 days. Model (A) is a simple uni-variate vanilla LSTM model. It serves as our base model as no emotions are used as features for forecasting. Model (B) uses ['Sadness', 'Joy', 'Anticipation'] as feature set in a simple vanilla LSTM model. Model (C) uses all 8 emotions as features to simple vanilla LSTM architecture. Model (D) adds to the architecture of Model (C) a simplified Attention layer as proposed in study [19].

We find that Model (B) gives the best result with a rolling window of past 15 days. Our result is in conjunction with the Granger Causality analysis as the three emotions 'Sadness', 'Joy' and 'Anticipation' improve the explanatory power of the base model. We also notice that all the 8 emotions do not help in improving the forecast as Model (C) and Model (D) perform poorer. This shows that by adding more features we would essentially be overfitting the data.

It can also be concluded that a rolling window of 15 days gives the best results for all the models. This can be attributed to the fact that adding 30 days of past information make the models to overfit, whereas a smaller window of 7 days causes models to underfit.

The attention mechanism in Model (D) gives it the ability to integrate information over long-term and was expected to perform well. However, as mentioned by the authors in [19], the models fails to preserve temporal order because computing an average over time discards order information. Hence, it can be observed that long-term past values of public emotions do not significantly impact conditional volatility.

6 Conclusion and Future Work

We have investigated the causative relation between public emotions as measured from daily collection of financial news articles and the Indian stock market risk.

Our research shows that firstly public emotions can be captured from English financial news headlines for Indian markets through the help of lexicon and natural language processing. Second, amongst the eight standard human emotions 'Joy', 'Sadness' and 'Anticipation' are the most causative of Indian stock markets volatility. The addition of these 3 emotions as feature set to a univariate forecasting model improves its explanatory power. The performance measure we have used is rolling window cross validation, which is more indicative of the market movements for financial data. Finally, the performance of vanilla LSTM model worsens due to the addition of simplified attention mechanism layer.

Finally, our analysis doesn't take into account some factors. Our study makes use of only English news headlines and tweets, because of which we can not capture actual public sentiments. Additionally, NRC Emotion Lexicon is based out of English, hence would not be able to capture emotions associated with Indian events, festivals and Indian local language words in English e.g.,'Muhurat trading' that is special trading session that takes place usually on the festival of Diwali in India.

An important research direction for future work is to identify the sectors of the economy and different shares that get most affected with emotions. Also, investigating the magnitude of reaction and persistence on volatility due to emotions can be an interesting line of research.

7 Implications

Movements in financial markets are linked to the overall health of the economy. The study about the relation between Conditional Volatility of Domestic Macroeconomic Factors and Conditional Stock Market Volatility [14], showed the increasing interdependence of financial markets and macroeconomic fundamentals in India. Hence, the study is important for policy makers to understand overall macroeconomic risks. Rising volatility can trigger financial instability and further apprehensions by mean of huge capital outflow.

The various forms of volatility in financial markets play important roles in portfolio management, especially in asset allocation, asset pricing, portfolio selection, portfolio diversification, and risk management. So, forecasting of conditional volatility of Indian stock market would be beneficial for investors and financial market experts. A forecast of increasing volatility, would indicate a need towards diversification of portfolio investments between stable and risky sectors.

Conflict of interest

Authors declare that they have no conflict of interest.

References

1. Dr. Renu Isidore. R et al., A Behavioral Finance Perspective of the Stock Market Anomalies, International Journal of Research in Management, Economics and Commerce,ISSN 2250-057X, Impact Factor: 6.384, Volume 08 Issue 04, April 2018, Page 5-9
2. Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Using structured events to predict stock price movement: An empirical investigation. In Proc. of EMNLP, pages 1415–1425, Doha, Qatar, October 2014. Association for Computational Linguistics.
3. Kira Radinsky, Sagie Davidovich, and Shaul Markovitch. Learning causality for news events prediction. In Proc. of WWW, pages 909–918. ACM, 2012.
4. Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. Journal of Computational Science, 2(1):1–8, 2011.
5. Wesley S Chan. 2003. Stock price reaction to news and no-news: drift and reversal after headlines. Journal of Financial Economics 70, 2 (2003), 223–260.
6. John R Nofsinger. 2005. Social mood and financial economics. The Journal of Behavioral Finance 6, 3 (2005), 144–160.
7. Gabriele Ranco, Darko Aleksovski, Guido Caldarelli, Miha Grčar, and Igor Mozetič.2015. The effects of Twitter sentiment on stock price returns. PloS one 10, 9(2015), e0138441.
8. Nadine Strauß, Rens Vliegenthart, and Piet Verhoeven. 2016. Lagging behind? Emotions in newspaper articles and stock market prices in the Netherlands. Public Relations Review 42, 4 (2016), 548–555.
9. Goudarzi, Hojatallah, and C. S. Ramanarayanan. "Modeling asymmetric volatility in the Indian stock market." International Journal of Business and Management 6.3 (2011): 221.
10. Kirange, D. K., Deshmukh, R. R. (2016). Sentiment Analysis of News Headlines for Stock Price Prediction. Composoft, An International Journal of Advanced Computer Technology, 5(3), 2080-2084.
11. R. Ahuja, H. Rastogi, A. Choudhuri and B. Garg, "Stock market forecast using sentiment analysis," 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2015, pp. 1008-1010.
12. Pandey, A. Volatility models and their performance in Indian capital markets. Vikalpa, 30(2), 27-46. (2005).
13. Chayanin Wong and In-Young Ko. Predictive Power of Public Emotions as Extracted from Daily News Articles on the Movements of Stock Market Indices. In Web Intelligence (WI), 2016 IEEE/WIC/ACM International Conference on. IEEE,705–708.(2016)
14. Kumari, and Mahakud, Jitendra. Relationship Between Conditional Volatility of Domestic Macroeconomic Factors and Conditional Stock Market Volatility: Some Further Evidence from India. Asia-Pacific Financial Markets. 10.1007/s10690-014-9194-7?sa$_{c}ampaign = email/event/articleAuthor/onlineFir.$(2014)

15. https://in.mathworks.com/help/econ/rolling-window-estimation-of-state-space-models.html
16. Crowdsourcing a Word-Emotion Association Lexicon, Saif Mohammad and Peter Turney, Computational Intelligence, 29 (3), 436-465, 2013.
17. Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon, Saif Mohammad and Peter Turney, In Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text,June 2010, LA, California https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm
18. A. Lapedes and R. Farber. Nonlinear signal processing using neural network: Prediction and system modeling. In Los Alamos National Lab Technical Report.
19. Raffel, Colin, and Daniel PW Ellis. "Feed-forward networks with attention can solve some long-term memory problems." arXiv preprint arXiv:1512.08756 (2015).