Stock Market Regime Change prediction based on historical financial and macroeconomic indicators, and sentimental analysis of social media

Devansh Joshi Hardik Dudeja Mohit Bhar

Prakhar Maurya Shivam Arora Hemang Dahiya

April 23, 2023

Abstract

Hidden Markov Models are the de facto standard to predict regime changes in financial markets. Historically, Hidden Markov Models have been used in conjunction with realised volatility as input in order to model time series financial data. However, in this study directional changes have been used to model time series data. To improve the predictions of the Hidden Markov Model, in addition to directional change data, a feature vector corresponding to the sentiment analysis of historical social media posts has been used as an input to the Hidden Markov Model. Finally, the Chicago Board Options Exchange's CBOE Volatility index, 22 and 66 day returns of tesla share price in addition to the 22-day volatility of the tesla share price have also been used as inputs to the Hidden Markov Model.For the purpose of this study, the model has been applied to Tesla(TSLA) share price data to detect regime changes over the period (September 2021-September 2022).

Keywords: Sentiment Analysis, Hidden Markov Model, Directional Changes,Regime Changes

1 Introduction

Predicting regime changes in financial markets and individual stocks per se is a field of immense importance. Financial markets are prone to highly volatile swings in the wake of capricious changes and sudden demand-supply shocks. Regimes in financial markets represent different periods of an individual stock or the entire market based on certain key variables such as the mean and standard deviation of the share price. Being able to detect regime changes can be a potent tool to realise capital gains in financial markets. The method that we propose in this paper is to utilise a multivariate Hidden Markov Model. The paper utilises a Directional Change(DC) indicator originally proposed by [Tsang et al.] in addition to a sentiment analysis based feature vector based on historical twitter posts. Other inputs to the Hidden Markov Model are the Chicago Board Options Exchange's CBOE Volatility index, 22 and 66 day returns of the tesla share price in addition to the 22-day volatility of the tesla share price. The motivation to utilise directional changes stems from their ability to detect market swings at irregular time intervals which is advantageous over conventional realised volatility based approaches since realised volatility can only be measured at regular time intervals.

2 Problem Statement

Forecasting financial market trends and fluctuations in financial regimes is an important task in the financial sector. Our project aims to tackle this task by creating a model that unites traditional financial and macroeconomic indicators, such as index values, with innovative measures such as directional changes, well-established methods like time series analysis, and an examination of historical social media posts to predict stock market trends and regime changes based on their average and volatility. As per Tsang Et. al [8], in order to classify regimes, a plot between normalized average TMV and normalized average T is a reliable indicator of distinction between various regimes. We have tried to emulate their research and tried to classify regimes on this basis.

3 Motivation

The motivation behind this paper is to be able to accurately classify regimes in financial markets to be able to gauge when and when not to invest in equities based on the mean and standard deviation of the predicted regimes. Correctly classifying regimes can be extremely beneficial in realising stock market gains. Incorporating directional changes as well as sentiment analysis into one model allows us to enhance the fit of the proposed Hidden Markov Model.

4 Novelty

The novelty of the proposed approach comes from the amalgamation of sentiment analysis and directional changes into one consolidated model.To the best of our knowledge, existing literature has not fused these two potent approaches into one approach. The utilisation of directional changes stems from their ability to detect market swings at irregular time intervals which is advantageous over conventional realised volatility based approaches since realised volatility can only be measured at regular time intervals. The incorporation of sentiment analysis data adds a tacit layer of market sentiment on top of the inputs that have already been fed to the Hidden Markov Model.

5 Literature Review

Twitter serves as an important platform when it comes to expressing public opinions. [8]show how public opinions posted on Twitter can impact the stock price of a company. [8] have performed sentiment analysis using two different textual representations, Word2vec and N-gram for the inspection of public sentiments in tweets. Using sentiment analysis and machine learning principles, it proves that the positive news can lead to a gradual rise in the stock price of any company and vice versa. It concludes by proving that there is a strong correlation between the increase and decrease in the stock price of a company and public sentiments in twitter tweets. [8] have found a strong correlation between public sentiment on Twitter and a company's stock prices. It introduces a sentiment analyzer that categorizes tweets as positive, negative, or neutral. The study supports the claim that positive Twitter sentiment about a company can impact its stock price, with implications for future research.

[7]discuss the changes in stock values and predicts them based on recent economic news about companies, with focus on the news headlines. Various tools were used to analyze the sentiment of these headlines, including BERT, VADER, TextBlob, and a Recurrent Neural Network. The sentiment results were then compared to the stock changes during the same time period. BERT and RNN were found to be more accurate in determining emotional values, without neutral sections, as compared to VADER and TextBlob. By comparing the sentiment analysis results with the values of the stock market, it was possible to determine the moment when the stock values changed. Also, we can observe the difference between the models in terms of the effect of emotional values on the change in the stock market value, as visible in the correlation matrices. This study shows that economic news headlines can have an impact on stock market values, and different sentiment analysis tools can provide varied results. It also talks about how future research could expand the analyses, add new features, and include other tools to compare stock market predictions with different sentiment analysis tools, potentially incorporating these changes into a platform that is easy to use.

[6]In this paper by Tae Kyun Lee, Joon Hyung Cho, Deuk Sin Kwon, and So Young Sohnon discuss the different financial network indicators that might be used for investment strategies on the global stock market . Using a vector auto-regressive model, the authors propose undirected and directed volatility networks based on pair-wise correlations and system-wide connectedness. The network indicators are entered as inputs for computing strategies using ML approaches (logistic regression, support vector machine, and random forest), and thus their effect and usefulness are judged. Two strategies are designed based on the stock price indices: a global market prediction strategy and a regional allocation strategy for developed markets/emerging markets. The study results indicate that network indicators are important supplementary indicators that can aid in predicting the global stock market and the regional relative directions (up/down). This study aimed to explore the effectiveness of using a financial network based on global stock indices of 10 countries to build a global stock portfolio strategy and predict market crises and structural changes. Three machine learning techniques (LR, RF, and SVM) were utilized for prediction. Previous studies have lacked practical usefulness in this area.

[5]explore the use of machine learning techniques in predicting the stock market. Major focus of the review is on the markets and types of variables that are used as inputs in these techniques. The review analyzed 138 journal articles published between 2000 and 2019. The key contributions of this study are on two things: (1) a comprehensive examination of the data used, including the markets and stock indices analyzed in the predictions, as well as the 2173 unique variables utilized for stock market predictions, such as technical indicators, macro-economic variables, and fundamental indicators, and (2) a thorough evaluation of the machine learning techniques and their variations employed for the predictions. Additionally, a bibliometric analysis is presented, highlighting the most influential works and articles.

In their paper,[9]Edward Tsang and Jun Chen have used directional changes as opposed to the traditional method of using time series analysis to detect regime changes in financial markets. The advantage with directional changes is that they can capture market fluctuations at irregular intervals of time vis-a-vis conventional time series based approaches that only sample market movements at regular intervals such as seconds, minutes, hours etc. Regime changes capture mean and volatility of the stock market in order to gauge time periods where one can invest in equity and periods of high fluctuation where one should not be invested in equity. The authors make use of a hidden markov model in order to predict the hidden states/regimes using directional change as the observed states.

[3]have ascertained that investors are keenly interested in accurately predicting the stock market, but the volatile nature of the market driven by factors such as microblogs and news make it difficult to rely solely on historical data

for prediction. Therefore, it is important to consider external factors when predicting the stock market. Machine learning algorithms can be applied to social media and financial news data to predict stock markets as this information can influence investor behavior. In this study, they have used algorithms on social media and financial news data to determine the impact of this data on stock market prediction accuracy for a period of ten days. To improve the quality of predictions, they conducted feature selection and reduced spam tweets in the datasets. They have also performed experiments to identify stock markets that are difficult to predict and those that are more influenced by social media and financial news. They have compared the results of different algorithms to find a consistent classifier, and ultimately used deep learning and ensemble classifiers to achieve maximum prediction accuracy.

In their paper[2]have proposed a machine learning model for predicting stock market prices. Their proposed algorithm combines Particle Swarm Optimization (PSO) with Least Squares Support Vector Machines (LS-SVM) to predict daily stock prices. The model is developed based on the analysis of historical stock data and technical indicators. The PSO algorithm is used to optimize LS-SVM by selecting the best combination of free parameters. This avoids overfitting and local minima problems and improves prediction accuracy.They have applied and evaluated their proposed model using thirteen benchmark financial datasets and compared it with an artificial neural network using the Levenberg-Marquardt algorithm.

[1]'s paper builds on previous work to classify market regimes using a Hidden Markov Model based on a trend summarization approach. To compare and contrast market regimes across different financial markets, they can be positioned into an indicator space based on the two DC-based indicators. This allows for the observation of similarities and differences between market regimes. The DC-based indicator values are normalized using a feature scaling approach to allow for comparisons between markets. The authors applied this method to ten assets across four market periods and found that normal and abnormal regimes are clearly separable based on their positions in the trend summarization space. This has implications for monitoring regime changes and adjusting trading strategies.

[4]The authors of this paper introduced a novel approach that integrates part-of-speech tags into topic modeling techniques, which they named the "LDA-POS" method. The authors conclude that their hybrid deep learning model is a promising tool for predicting stock prices and could be useful for investors and financial analysts.

6 Database

The dataset consists of historical tweet dataset, historical S&P 500 index dataset,tesla share price data and CBOE's historical VIX values dataset. Historical S&P 500 index dataset,telsa share price data and CBOE's historical VIX values dataset were obtained using Python's inbuilt datareader. Tweet data was web scrapped using selenium and beautiful soup library.The time period is from September 30,2021 to September 29,2022.

7 Methodology

For implementing the Hidden Markov Model, python's hmmlearn library was used. Directional Changes were applied on historical data for the tesla share price. Since a trend i.e. direction change event followed by an overshoot event takes place over different time periods with a single value for R, to create daily data for R, the same value of R for a trend was taken for the entire duration of the days in the trend. The same was done for the TMV values and time taken for the completion of a trend value. The R values obtained were transformed using log transformation to model the data better and change the distribution to gaussian distribution.For the purpose of this research, the GaussianHMM from the hmmlearn python library has been used. GaussianHMM uses the expectation maximisation algorithm to learn the parameters of the Hidden Markov Model like the transition probability matrix, emissions probability matrix and the initial distribution probabilities. The expectation maximisation algorithm is an iterative algorithm and might get stuck at a local optima, for this purpose, whilst fitting, the random state of the GaussianHMM needs to be changed continuously to ensure that the algorithm converges for different values of the random state. Furthermore, another consideration with the Hidden Markov Model is that it is an unsupervised learning algorithm, as a result, the number of hidden states of the model need to be predetermined. To find the optimal number of regimes, the model is fit on the same dataset for different values for the number of hidden states. The score of each model is evaluated using model.score and appended in an array then the model with the best score is used for predictions. NLTK's sentiment analysis module includes an implementation of the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis tool. VADER is a rule-based approach to sentiment analysis that is designed specifically for analyzing the sentiment of social media text, where traditional sentiment analysis tools may struggle due to the informal nature of the language.

VADER uses a lexicon of words and phrases that are associated with positive or negative sentiment, as well as rules for how to combine these words and phrases to derive sentiment scores for text. It takes into account punctuation, capitalization, and other features of social media text, such as the use of emoticons and slang.

VADER produces sentiment scores for text on four dimensions: positive, negative, neutral, and compound. The compound score is a normalized score that combines the other three scores into a single value, ranging from -1 (most negative) to +1 (most positive).

NLTK's implementation of VADER is pre-trained and can be used out of the box for sentiment analysis tasks. It also includes a number of utility functions for preprocessing text and interpreting the sentiment scores produced by VADER. VADER is widely used for sentiment analysis of social media text and has been shown to perform well in this domain.

8 Evaluation

To evaluate the goodness of fit of our model we have made use of numerous evaluation metrics.

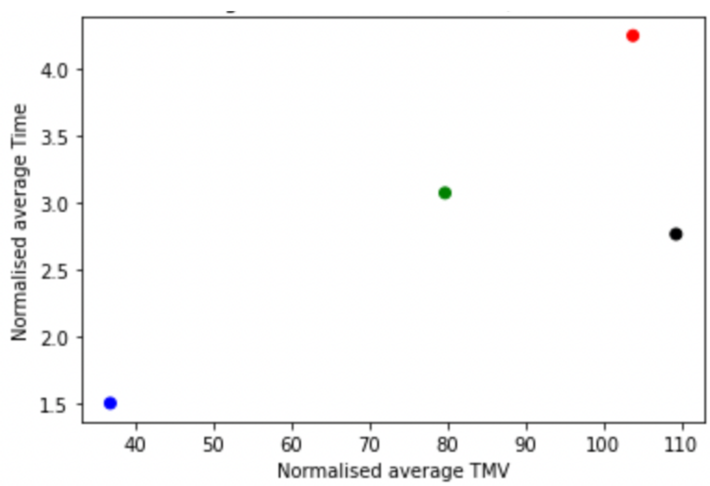

Figure 1: Normalised Average TMV vs Time

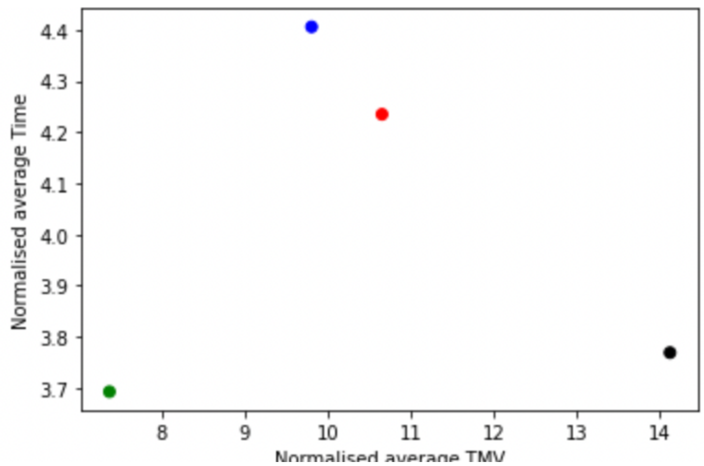

Figure 2: Normalised Average TMV vs Time

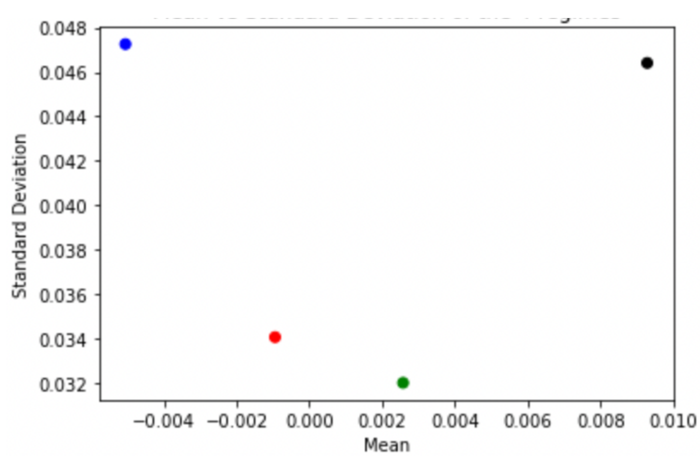

Figure 3: Mean vs Standard Deviation of the regimes

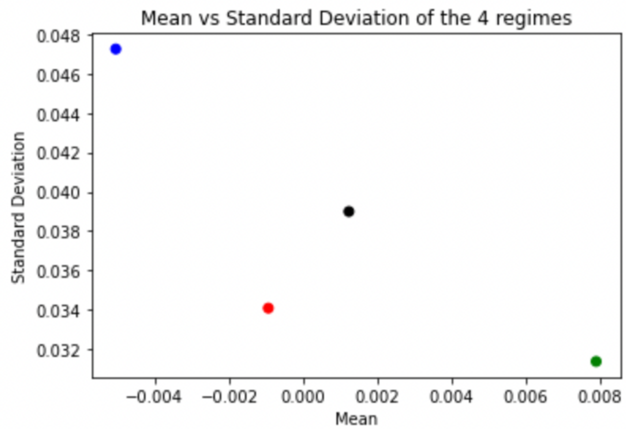

Figure 4: Mean vs Standard Deviation of the regimes

Regime	Daily
0.0	−0.000942
1.0	0.002576
2.0	−0.005054
3.0	0.009285

Regime	Daily
0.0	0.034082
1.0	0.032029
2.0	0.047255
3.0	0.046406

Figure 5: Mean vs Standard Deviation of the regimes

```
          Daily
Regime
0.0    -0.000942
1.0     0.007894
2.0    -0.005054
3.0     0.001229
          Daily
Regime
0.0     0.034082
1.0     0.031371
2.0     0.047255
3.0     0.038993
```

Figure 6: Mean vs Standard Deviation of the regimes

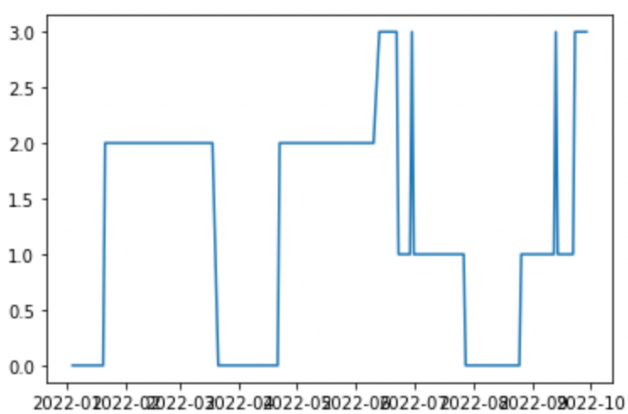

Figure 7: Regime Predictions

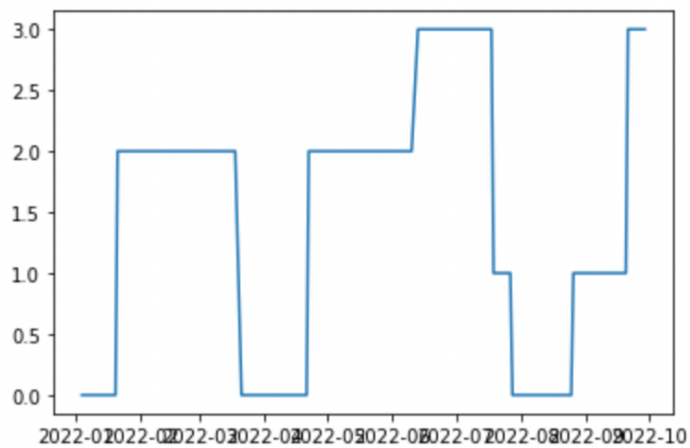

Figure 8: Regime Predictions

References

[1] Jun Chen and Edward Tsang. Classification of normal and abnormal regimes in financial markets. *Algorithms*, 11(12):202, Dec 2018.

[2] Ali Derakhshan and Hamid Beigy. Sentiment analysis on stock social media for stock price movement prediction. *Engineering Applications of Artificial Intelligence*, 85:569–578, 2019.

[3] Osman Hegazy, Omar S. Soliman, and Mustafa Abdul Salam. A machine learning model for stock market prediction. *International Journal of Computer Science and Telecommunications*, 4:17–23, 12 2013.

[4] Wasiat Khan, Mustansar ali Ghazanfar, Muhammad Awais Azam, Amin Karami, Khaled Alyoubi, and Ahmed Alfakeeh. Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, 13, 07 2022.

[5] Mahinda Mailagaha Kumbure, Christoph Lohrmann, Pasi Luukka, and Jari Porras. Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, 197:116659, 2022.

[6] Tae Kyun Lee, Joon Hyung Cho, Deuk Sin Kwon, and So Young Sohn. Global stock market investment strategies based on financial network indicators using machine learning techniques. *Expert Systems with Applications*, 117:228–242, 2019.

[7] László Nemes and Attila Kiss. Prediction of stock values changes using sentiment analysis of stock news headlines. *Journal of Information and Telecommunication*, 5(3):375–394, 2021.

[8] Venkata Sasank Pagolu, Kamal Nayan Reddy, Ganapati Panda, and Babita Majhi. Sentiment analysis of twitter data for predicting stock market movements. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pages 1345–1350, 2016.

[9] Edward Tsang and Jun Chen. Regime change detection using directional change indicators in the foreign exchange market to chart brexit. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2(3):185–193, 2018.