THE UNIVERSITY OF CHICAGO

Analysis of Topic Modeling on the Stock Market

By

Wonje Yun

June 2024

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Preceptor: Sergio Salas

Sponsor: Nicholas Ross

Abstract

Earnings call is a presentation of the company that encompass not only the financial results but also the detailed explanation that may not appear on numbers. This paper investigates earnings call transcripts of listed companies and look for non-quantified signals that could explain the market better in terms of market trends and sector grouping with the use of topic modeling. With the use of topic modeling, we generate topics, which are grouped tokens that co-occurs the most frequently, that could possibly represent the market for each quarter. After evaluating the generated topics, this paper finds that the change in occurrence of tokens within the topic represents some market trend. Furthermore, the stocks, when grouped by the topic, showed comparable performance to the industry sector groups. We thus conclude that the topic modeling approach can identify implicit information that are not shown in the financial data.

Keywords: stock performance prediction; stock grouping; earnings call; topic modeling

Main Deliverable

1 Introduction

Earnings call are presentations of the company's executives towards shareholders and analysts after publishing accounting performances. Question and Answer (QnA) session in the earnings call is held between the executives and analysts to clarify or advocate the company's performance and ongoing projects/plans. After the general presentation of the company's financial performances, analysts question any unclear messages during the presentation or request an update on the previous projects and their goals. Therefore, earnings call's QnA session are highly thought of as an important textual content that supports the current and future financial performances.

Due to these characteristics, investigation in the relationship between the sentiment of the earnings call on the stock market has been conducted early on by previous researches (Doran et al., 2012; Price et al., 2012). Early research focused on the quantifying the tone of the presentation. After the technical advance of natural language processing (NLP) models, more attempts in analyzing large datasets and sentiment scores using machine learning techniques has taken place: techniques has been used such as sentiment score analysis, word embedding or neural network (NN) (Brennan, 2021).

More recent researches attempted to use NLP models to detect stock market structures (Harford et al., 2023; Kim & Nikolaev, 2023). The results of the previous research were able to imply that there is considerable result that NLP methods can identify stock market structures that are not explicit by the sock prices. Based on the prior research, we hypothesized that the earnings call data of companies would contain important information that reflect the market structure and information. Based on the hypothesis, this paper tries to answer two main research questions:

- Does QnA transcripts show any pattern of market trends?
- Can QnA transcript reflect companies' common information that are not in explicit form?

This study has attempted to answer the research questions by applying topic modeling approach to the earnings call text data. Among the topic modeling models, this study will be using hierarchical Dirichlet process (HDP) to generate the topics between the earnings call data for each quarter. With running the HDP, models are generated with the number of topics, the distribution of words within the topic, and the probability of each documents

Table 1: Change in occurrences of sampled tokens for each year

	Year								
Token	2014	2015	2016	2017	2018	2019	2020	2021	2022
hospital	1	1	1	1	1	0	3	1	2
drug	2	0	1	3	2	0	1	2	3
game	0	1	3	4	4	4	4	4	2
software	0	3	2	0	2	2	2	3	0
hotel	3	3	3	1	2	0	2	2	2
aircraft	0	0	0	2	2	1	0	1	0
oil	3	5	3	2	4	4	1	4	1
gas	5	3	2	4	4	3	3	6	4

being assigned to the topic being included. From the generated models, we attempted to answer the research questions. In section 2.1 we provide the analysis of whether market trends can be reflected in the change of topic words. In section 2.2 we provide the analysis of whether groups based on topic modeling can reflect implicit common information between the companies, by comparing stock performance in risk and return analysis with industrial sector groups.

2 Results

2.1 Market Trend

Similarly from Huang et al., 2018, this study aimed to identify change in topic keywords, but in a entire market sense. We assume the market trend to be defined as the tokens within each groups with the highest probability of appearance. We gathered the top 10 tokens for each group for every quarter and combined them into a representative token list. We then counted the number of occurrence of each tokens in the representative token list by quarter and summed them up by year. Some of the sample token results of changes in occurrences of tokens in the representative token list can be found in Table 1.

Although most of the tokens had similar appearances as in the tokens ”oil” and ”gas” or negligible signs of trend during the years as in ”drug” and ”software”, we could still find that there was a trend for some tokens for the past 9 years. For instance, the token ”hospital” appeared in small numbers prior to 2020, but saw an sudden increase in appearance at 2020. Also, the token ”aircraft” seemed a increase in appearance after 2017, but showed no sign of occurrence at year 2020. We could assume that this aspect may be due to the COVID-19 breakout staging from 2019. Furthermore, for the token ”game”, the appearance has been increasing and reached its peak in 2017, continuing up to 2021. We could say that the term

”game” has provoked interest among companies since 2017. This may be due to the rise of high performance general purpose graphics processing units and cloud gaming that has provoked interest during that era.

Thus we were not able to entirely conclude that topic token appearances can reflect market trends. However, we were still able to argue that under influential events, the occurrences of token can change between years.

2.2 Performance Evaluation

In representing the common information between companies, we attempted to cluster the companies based on their generated topic. The resulting number of topics from HDP are shown as in Table 2. We can infer that regardless of the quarter, the number of topics generated is between 9 to 11.

Table 2: Number of topics for each quarter generated with the HDP model

	Quarter			
Year	Q1	Q2	Q3	Q4
2014	10	10	10	10
2015	10	10	10	10
2016	10	10	10	10
2017	10	10	10	10
2018	10	10	10	10
2019	10	10	10	10
2020	10	10	10	10
2021	13	10	10	11
2022	9	11	10	10

We would use this number of topics to group the stocks. HDP model can represent each document as a weighted sum of topics, with the weight being the probability. Based on the topic with the largest weights, we allocated each company to the corresponding topic. For instance, if

$$COMP_s = \sum_{i=1}^{n} \alpha_i TOPIC_i \tag{1}$$

we allocate $\{COMP_s : TOPIC_t\}$, where t is such that $\alpha_t = \max \alpha_i$. Since the earnings call data was processed for one company to have one transcript, we were able to allocate each companies to the HDP topics. From this point, we would denote the company groups that were generated this way as HDP groups. We then generated a portfolio based on the HDP groups, where the portfolio price is calculated as an equally weighted portfolio for simplicity.

For comparison, we chose global industry classification standard (GICS) sector as the group represented by explicit common information. GICS is an industry classification method developed by S&P and MSCI, which has a hierarchical structure of sector, industry group, industry and sub-industry. Another reason for comparing with GICS sectors is due to the number of sectors in the GICS, which is 11 a close number to the topics our HDP model has generated. For comparison with the group portfolio, we used GICS sector ETF, derived using the `yfinance` library.

Applying the evaluation method discussed in Part 3, we have calculated the performances for both GICS and HDP groups and compared the result with the S&P500 index performance evaluation as a benchmark. Since the performance evaluation is to compare how would the asset would have performed in the future based on the past, the calculation for the evaluation methods are done such that the performance during quarter $t + 1$ is calculated for the topics grouped for quarter t. Moreover, since the HDP groups does not allocate to GICS groups one on one, we compared the ratio of groups over the total number of groups that perform better than the benchmark.

Normality Check Normality checks are needed to set the VaR calculation method. We applied Shapiro-Wilkins test for each group portfolios to investigate if there are non-normality in prices. Figure 1 displays the number of groups that fails the Shapiro-Wilkins test. From this result, we see that most of the times at least 2 groups fails the normality check. Due to this reason, we have used the non parametric method of historical simulation when calculating the VaR value.

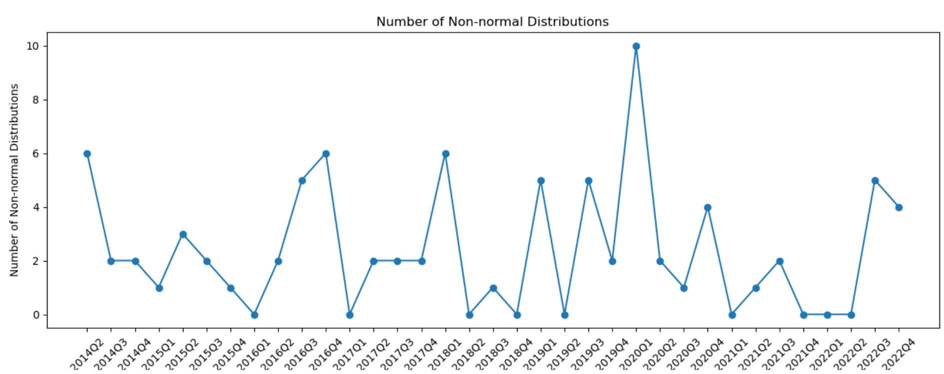

Figure 1: Number of topic groups that fails the Shapiro-Wilkins test.

Value at Risk We first compared the ratio of HDP groups and GICS groups that performed better than the market VaR. The comparison plot is shown as in Figure 2. According to the figure, we see that the outstanding performance ratio for the GICS groups dominate most of the case for HDP groups. Especially, this dominance was more evident during the 4th quarter of 2021 to the 4th quarter of 2022. During other periods, the dominance was not severe, or at least equivalent with the HDP groups. From this result, we were not able to conclude that topic modeling could capture any implicit connections between companies. However, the relatively small dominance of GISC groups except for 2022 can infer that a more risk optimized portfolio generation method, or using a more conservative risk performance measure can enhance the results.

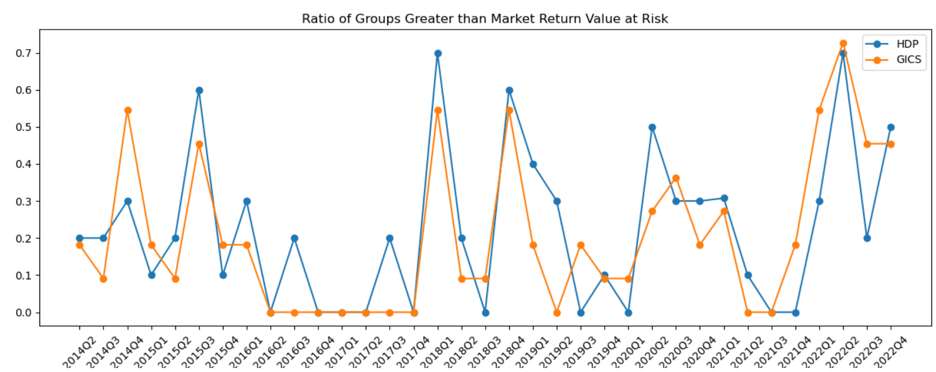

Figure 2: Ratio comparison of groups that has higher VaR than the market per quarter.

Conditional Value at Risk As an extension to the result from the VaR evaluation, we performed a more conservative approach of the risk measurement: the CVaR method. The comparison of the ratio that performed better than the market CVaR is shown in Figure 3. Although the ratio of GICS groups still mostly dominates the ratio of HDP groups, we were able to see that the dominance has become highly insignificant compared to the results for VaR. From this result we were able to imply that with a conservative risk evaluation method, HDP groups could perform at least comparable to the GICS groups.

Sharpe Ratio Our evaluation method for return measures were first carried out with the Sharpe ratio. Setting the 13 week treasury bill as the risk free rate, we have calculated the Sharpe ratio for GICS groups, HDP groups, and the market (S&P 500 index). Similar to the previous comparison method, we compared the ratio of groups that had better Sharpe ratio than the market, which could be seen in Figure 4. Unlike the previous analysis results, we were able to find that HDP groups slightly dominated the ratio of GICS groups. This result could imply that although topic modeling could not identify underlying information related to risks, the information related to returns could be distinguished.

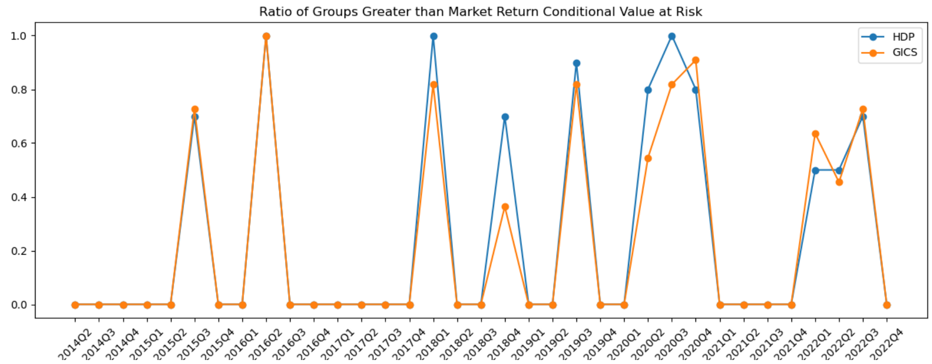

Figure 3: Ratio comparison of groups that has higher CVaR than the market per quarter.

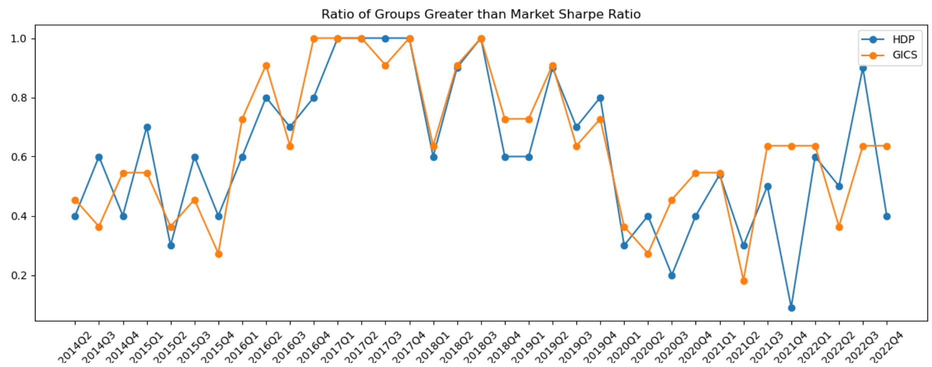

Figure 4: Ratio comparison of groups that has higher Sharpe Ratio than the market per quarter.

Information Ratio The IR is a form of extension of the Sharpe ratio that calculates the risk adjusted excess returns relative to the benchmark. With the use of S&P 500 index price as our benchmark, we compared the ratio of groups having better IR performance of 0.0, which is the value of IR for the benchmark (i.e the market). In this case, we find the result to be mixed with HDP groups and GICS groups dominating each other between quarters. This result infers that topic modeling may not reveal as much information for excess returns from the market. However, since the quarters with dominating HDP groups are more frequent, we can expect the result to improve with different portfolio optimization method.

3 Conclusion

In conclusion, we analyzed the topic modeling approach to the text information from the earnings call of the companies to identify implicit patterns. The QnA session text data

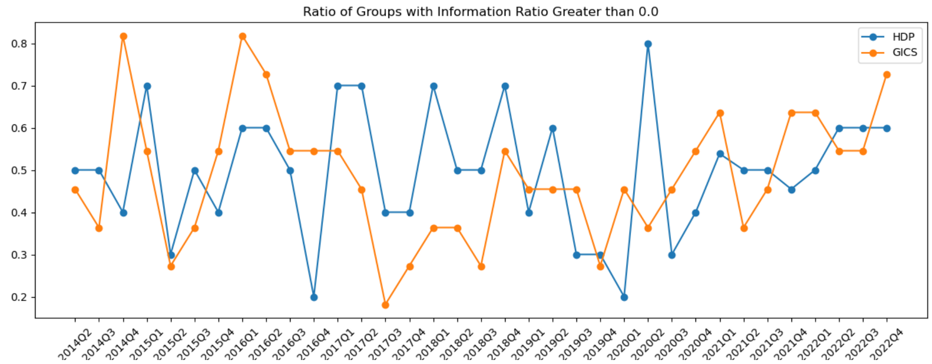

Figure 5: Ratio comparison of groups that has higher IR than the market per quarter.

are processed into tokens and the set of texts per each quarter is represented into multiple topics, that consist of tokens that are most likely to co-occur. Our analysis show that despite the top occurring tokens' appearance within the transcripts does not infer a trend in the market, the group of stocks gathered based on the topic model had comparable or better performance compared to the group of stocks by industry sector. Therefore, topic modeling can be said to infer implicit information that can is not represented explicitly in the market.

Literature review

Machine Learning (ML) has gained popularity in usage as hardware technology has improved. It has not only been popular in computer science and engineering field, but also in social science and economics field which deals with large data.

In the economics field, the use of ML has been mainly applied to devising new models in econometrics, or in simulating and predicting economic events. Athey et al., 2018; Athey and Imbens, 2019 introduces ML methodologies that are relevant to economic research in improving the econometric methods. The major methodologies involve regression and classification that are known to enhance performance when dealing with large data. Furthermore, the use of ML in causal inference has been introduced, where ML can enhance the prediction of treatment effect parameters. They both conclude that ML methodology can be applied in economics, and can benefit the researchers. Mullainathan and Spiess, 2017 also gives an approach where econometrics can benefit from ML methodologies, and explains the interpretation for the results ML models can generate. Research such as Ghoddusi et al., 2019; Storm et al., 2020 surveys through the application ML has made in the economic research.

In the field of finance, the use of ML has also been a key interest for researchers as new or regime changing algorithm came out. There has been plentiful work that aims to investigate patterned structures or trading methods that can be applied in the stock market using machine learning techniques or reinforcement learning techniques. There are already several books that introduce extensive range of machine learning tools and its applications to the data that can be found from the stock market (De Prado, 2018; Jansen, 2020). De Prado, 2018 introduces what is needed to comprehend machine learning models in the field of finance, and displays how certain machine learning models could solve some of the financial problems. Jansen, 2020 gives an introduction to a variety of machine learning models and reinforcement models that has possible use for algorithmic trading. There are also numerous journals that has applied machine learning models to finance related areas such as asset management (Y. Lee et al., 2023).

As ML techniques improved, algorithm to process languages and analyze patterns has come into interest as well. The concept of NLP was present as early as 1950s where researchers such as Chomsky, 2002 attempted to convert natural language to a computer understandable input. Since 1980s the NLP converted from handwritten rule to statistical approach, which led to the increase of machine-learning approach such as neural nets and long short-term memory. With the increase in hardware performance, NLP has become easier to execute and has become widely accessible to researchers in various fields.

Especially in the social sciences, where quantified and non-quantifiable data are both important, the use of NLP has been performed in many areas. Bail, 2016; Farzindar et al., 2015; Macanovic, 2022; Mullen and Malouf, 2006 are some examples of NLP being used in the social science field where text data are processed to analyze political and social behaviors. In economics NLP has not been of great interest, but research from Ash and Hansen, 2023; Roos and Reccius, 2024 suggest the use of narrative data in economic research may become useful.

In the financial sector, NLP has gained interest in that it could convert the non-quantitative data to quantitative form that could be applied to existing financial techniques. Mishev et al., 2020 suggest that NLP can be important in the financial market due to the efficient market hypothesis, that assumes all paset information being reflected in the price. The research compares multiple NLP algorithms in extracting text sentiments from financial headlines. A survey from Fisher et al., 2016; Xing et al., 2018 discusses the NLP application made in the financial sector, which ranges from market prediction to fraud detection and supply chain management. Their survey classifies the NLP application into semantic modeling, sentiment analysis and event extraction. Xing et al., 2018 further investigates possible financial datasource that NLP can be applied: corporate disclosures, financial reports, news and social media.

In the context of Xing et al., 2018's classification, earnings call can be considered a corporate disclosure. Furthermore, it is more associated with the market reaction compared to financial reports or media cover published later (Kimbrough, 2005), which makes it a important data source in financial NLP. There has been a wide study analyzing and using it for predicting the market movement (Doran et al., 2012; Frankel et al., 2017; Keith & Stent, 2019; Liang, 2016; Ma et al., 2020; Price et al., 2012; Solberg & Karlsen, 2018). Ma et al., 2020 used sentence embedding to earnings call transcripts and added as an input to a reinforcement model in predicting the future stock prices. More recently, Kantos et al., 2023 compared the performance of sentiment based trading strategy between different NLP models, which claim to generate alphas that are not explainable in a more traditional context. Furthermore, there are research that applies earnings call in a different approach than market prediction. Keith and Stent, 2019 investigated the difference or similarity between earnings call and pre/post analyst report based on the sentiment scores of the earnings call transcripts.

For many previous research involving NLP and financial text data, the main factor that was extracted from the text was sentiment data of the speaker or the text. However, NLP can also statistically reveal recurring patterns inside the input text. The method is called topic modeling, and is known to be able to process large collection of document and help organize and summarize the text data into representations (Abdelrazek et al., 2023; Barde & Bainwad, 2017).

According to the surveys of topic modeling, the topic modeling is aimed at uncovering the latent structure of the given documents to generate the topic, which is a collection of words that have high probability of co-occurrence. (Barde & Bainwad, 2017; Jelodar et al., 2019; Sandhiya et al., 2022). There has been various methods for topic modeling, from using the count of words to calculating the probability or co-ocurrence of the words (Barde & Bainwad, 2017). Among different methods, the most common and widely represented method for topic modeling is the latent Dirichelt allocation (LDA) method (Jelodar et al., 2019). This method assumes that documents are a mixture of topics, and the topics are characterized by the distribution of words (Barde & Bainwad, 2017; Jelodar et al., 2019).

The characteristic of topic modeling, that can process and analyze texts into of number of topics, enabled it to be applied to research that involves large text data. In the field of social science, topic modeling has been used to track the agenda of political speeches (Greene & Cross, 2015), or analyze the different positions of candidates over a given agenda (Valdez et al., 2018). Furthermore, topic modeling can be used in analyzing social media such as twitter to discover if social media covers information that traditional news media covers (Zhao et al., 2011).

In finance, topic modeling has been used in combination with sentiment analysis. Nguyen and Shirai, 2015 has performed topic modeling on twitter texts related to companies and performed sentiment analysis on the extracted topics. The sentiment and topic of the tweets were then used to predict stock price of the company. In terms of earnings call data, the application of topic modeling has been performed to compare common aspects between earnings call and analyst reports (Perico Ortiz et al., 2023). Also, Huang et al., 2018 generated topics within each industry sector and aimed to analyze whether analyst reports have any informative role by comparing the advent of topics between earnings call and reports. Furthermore, Hofmann et al., 2023 also analyzes the internal control quality of the company by detecting the existence of related keywords within the earnings call.

One feature of topic modeling is that it can be used to cluster the input documents based on their generated topics. Sun, 2014; Yau et al., 2014 attempted to cluster documents based on the topics generated by topic modeling and claimed that topic modeling can return clusters that well represents the relationship between the input documents. Grouping and clustering is important in analyzing the financial market, in that it could be used to generate portfolios or hedge risks between stocks that displays similar movement Farrell, 1974; Nanda et al., 2010. Furthermore, grouping and clustering stocks has shown to reveal underlying structures of the stock market (Aghabozorgi & Teh, 2014; Mantegna, 1999). The clustered structure can change with different economic situations and can be used as a indicator of crisis (Kocheturov et al., 2014).

We have seen that study of NLP in the finance field has been conducted thoroughly, and it does not limit to sentiment analysis. Other NLP analysis methods such as topic modeling has already been applied in the stock market text data and was covered by previous research. However, despite the importance of cluster analysis of the stock market, research that has attempted to cluster the stocks based on topic modeling is hard to be found. Thus, this research attempts in combining the topic modeling on earnings call in grouping stocks to analyze the underlying stock market structure.

Evaluation of Research Design

1 Data

1.1 Earnings Call Data

The main dataset we performed on for this research is the earnings call transcripts of S&P 500 companies. From the period of 2014 to 2022, we acquired the earnings call transcripts from Capital IQ Transcript data provided by S&P Global from Wharton Research Data Services (WRDS). There were total number of 478 companies with total of 16790 transcripts. An earnings call transcript of a company is divided into multiple rows based on the types of the talk, such as "Presentation Section", "Question", or "Answer". Following the analysis from Perico Ortiz et al., 2023, the types of the talk we've used are "Question" from analysts and "Answer" from executives, which they claim to have more related information with the stock prices.

Pre-Processing For most of the NLP models, pre-processing the text is needed to capture the context. Our pre-processing used `spacy` library with `python`, which has a fast computational speed and is widely used for NLP pre-processing. The texts were tokenized into separate grammatical parts (words, part of words, or punctuation) and lemmatized the tokens into its dictionary form to avoid redundancy in identifying the same token with different forms as separate tokens. During this process, we have selected the tokens that has noun as their part of speech. Since our research is aiming to investigate the market trends and common information among companies, we have thought of noun tokens as the more appropriate format of our research. Furthermore, our research used the bigram model to accommodate tokens that are more meaningful with two consecutive tokens combined. For example "North America" is a more appropriate token compared to "north" and "America". For the bigram model the `gensim` library from `python` has been used.

1.2 Stock Prices Data

For the topic modeling evaluation, we have used stock price data of S&P500 companies that matches the companies that appear in the earnings call transcript data. Since there are companies that has been delisted, our research derived the historical S&P500 component stock prices from WRDS. Moreover, for market price and risk free rate calculation, S&P500 index and 13 week US treasury bills were used. For these data, our study employed the Yahoo Finance API through the `yfinance` library from `python`.

2 Methods

2.1 Topic Modeling and Hierarchical Dirichlet Process

Topic modeling is a statistical algorithm for extracting the underlying structures of extensive documents first described by Papadimitriou et al., 1998. It is based on the idea that a text on a certain topic would be expected to contain the words related to the topic more frequently than others. The algorithm statistically infers commonly occurring words from the collection of documents and group the co-occurring words as a ”topic”.

Among various topic modeling methods, the most widely used model is the latent Dirichlet allocation (LDA) by Blei et al., 2003. Originally proposed in the field of biology, LDA uses a Bayesian method to uncover the probability distribution of the topics to appear in a certain document, and the probability distribution of the words within the topics. However, since it uses a parametric Bayesian model, it has to designate the number of topics to identify beforehand. Hierarchical Dirichlet process (HDP) is a non-parametric generalized method of LDA using the Dirichlet process, assuming that the topics share a common distribution. Unlike LDA, HDP does not require to designate the number of topics, but it assumes the number of topics and infers the number through the assumed common distribution of the topics (Teh et al., 2004). This non-parametric characteristic of HDP makes it suitable for finding group information that is not explicit, which suits our research into finding the underlying patterns of earnings call data. Also, study from Yau et al., 2014 claimed the HDP model to be better performing in document clustering compared to LDA. For implementing the HDP algorithm, this project used the `tomotopy` python library by M. Lee, 2022.

2.2 Performance Evaluation

To test whether there are latent information in the stock market, many researches use performance evaluation. For our research, we have evaluated the performance based on risks and returns. For risk evaluation, value at risk (VaR) and conditional value at risk (CVaR) were used. For return evaluation, the Sharpe ratio and information ratio (IR) were derived.

Value at Risk VaR is a concept used in risk management that quantifies risk as the loss of an investment given the probability that it will happen. The method assumes the distribution of the time series data and derive the lower α-percentile price. Then the value is assumed to be the loss of investment under $\alpha\%$ probability. Most widely used α is 1% or 5%, in which our research set $\alpha = 5\%$. The mathematical definition of VaR is given as follows:

$$VaR_\alpha(X) = -inf\{x \in \mathbf{R} | F_X(x) > \alpha\} \tag{2}$$

Conditional Value at Risk The CVaR, which is also called as expected shortfall, is a more conservative approach in evaluating the risks. It considers the period that it has made a loss under given probability, making the concept more apt for measuring risks in a time series data. Since the threshold for starting to consider the loss is the VaR value, it serves as a bound for VaR, making it a more conservative approach. The calculation is performed as the average VaR values between 0 to α probability:

$$CVaR_{1-\alpha}(X) = \frac{1}{\alpha} \int_0^\alpha VaR_{1-\gamma}(X) d\gamma \tag{3}$$

Sharpe Ratio Sharpe ratio is a widely used return evaluation method developed by Sharpe, 1966. It is defined as a risk-adjusted performance of an asset compared to the risk free rate, or the compensated return for the taken risk. Thus the higher Sharpe ratio, the more the asset returns given the same risk. The calculation for Sharpe ratio is as follows:

$$S = \frac{E[R - R_f]}{\sqrt{var(R - R_f)}} \tag{4}$$

Information Ratio IR is similar to the Sharpe ratio, but it differs in calculating the return of an investment compared to a risky benchmark index relative to the risk of the return. Compared to Sharpe ratio, IR is useful for comparing portfolio returns relative to the market. In many cases, S&P500 index returns are used as the benchmark index. The calculation of IR is performed as:

$$IR = \frac{E[R - R_b]}{\sqrt{var(R - R_b)}} \tag{5}$$

3 Data and Code Availability Statement

The code that I have used in this thesis is available in my github repository. Due to restrictions in data upload size in github, the data and models are not uploaded. However, the data that I have used in this thesis was extracted from WRDS, and the SQL query used is available in the github repository. The SQL query is written in `python` and can be run by using the WRDS cloud or WRDS database connection.

4 Evaluation

Earnings call transcripts are expected to imply non-quantified patterns that could explain the market better. Through our research, we were not able to answer the first research question that earnings call transcripts were able to identify some patterns related to market trends. However, for the second question regarding the existence of implicit common information among companies, we were able to find evidence that earnings call data contained information of companies that is not represented in its industry classification. Therefore, the use of topic modeling could enhance stock market analysis with the addition of patterns it could recognize.

Nevertheless, there are some improvements that needs to be done to make the results more valid and persuasive. First, the hyper-parameter used for training the HDP model was adopted from Sroka, 2020. Thus, with hyper-parameter tuning for the HDP model, we could optimize more of the generated results and could infer a better result. Second, our performance evaluation was done with the use of portfolios that were generated with equal weights per stocks. Although equally weighted portfolio is the simplest, it may not be the most optimized. Therefore, with application of portfolio methods that best optimizes within the HDP groups, we could expect an improvement in performance. Finally, we did not use the state-of-art NLP algorithms such as transformers and large language models (LLM). Since Vaswani et al., 2017, transformer based models have proved to better perform than other NLP methodology, and more recently LLM has dominated all other NLP methods. With better performing NLP models, we would be able to gain more evidence that could answer our research question.

References

Abdelrazek, A., Eid, Y., Gawish, E., Medhat, W., & Hassan, A. (2023). Topic modeling algorithms and applications: A survey. *Information Systems*, *112*, 102131.

Aghabozorgi, S., & Teh, Y. W. (2014). Stock market co-movement assessment using a three-phase clustering method. *Expert Systems with Applications*, *41*(4), 1301–1314.

Ash, E., & Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, *15*, 659–688.

Athey, S., et al. (2018). The impact of machine learning on economics. *The economics of artificial intelligence: An agenda*, 507–547.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*, 685–725.

Bail, C. A. (2016). Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media. *Proceedings of the National Academy of Sciences*, *113*(42), 11823–11828.

Barde, B. V., & Bainwad, A. M. (2017). An overview of topic modeling methods and tools. *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 745–750.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Brennan, M. (2021). *Predicting stock performance using quarterly analyst's call transcripts and natural language processing (nlp): An exploration* [Master's thesis, Utica College].

Chomsky, N. (2002). *Syntactic structures*. Mouton de Gruyter.

De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.

Doran, J. S., Peterson, D. R., & Price, S. M. (2012). Earnings conference call content and stock price: The case of reits. *The Journal of Real Estate Finance and Economics*, *45*, 402–434.

Farrell, J. L. (1974). Analyzing covariation of returns to determine homogeneous stock groupings. *The Journal of Business*, *47*(2), 186–207.

Farzindar, A., Inkpen, D., & Hirst, G. (2015). *Natural language processing for social media*. Springer.

Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management*, *23*(3), 157–214.

Frankel, R. M., Jennings, J. N., & Lee, J. A. (2017). Using natural language processing to assess text usefulness to readers: The case of conference calls and earnings prediction. *Available at SSRN 3095754*.

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, *81*, 709–727.

Greene, D., & Cross, J. P. (2015). Unveiling the political agenda of the european parliament plenary: A topical analysis. *Proceedings of the ACM web science conference*, 1–10.

Harford, J., He, Q., & Qiu, B. (2023). Firm-level labor-shortage exposure. *Available at SSRN 4410126*.

Hofmann, C., Kuhn, S. B., & Schwaiger, N. (2023). Using earnings conference call discussions to assess internal control quality. *FAR Working Papers*.

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management science*, *64*(6), 2833–2855.

Jansen, S. (2020). *Machine learning for algorithmic trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with python*. Packt Publishing Ltd.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent dirichlet allocation (lda) and topic modeling: Models, applications, a survey. *Multimedia tools and applications*, *78*, 15169–15211.

Kantos, C., Joldzic, D., Mitra, G., & Hoang, K. T. (2023). Comparative analysis of nlp approaches for earnings calls. In *Handbook of alternative data in finance, volume i* (pp. 215–228). Chapman; Hall/CRC.

Keith, K. A., & Stent, A. (2019). Modeling financial analysts' decision making via the pragmatics and semantics of earnings calls. *arXiv preprint arXiv:1906.02868*.

Kim, A. G., & Nikolaev, V. V. (2023). Profitability context and the cross-section of stock returns. *Chicago Booth Research Paper*, *11*(23).

Kimbrough, M. D. (2005). The effect of conference calls on analyst and market underreaction to earnings announcements. *The Accounting Review*, *80*(1), 189–219.

Kocheturov, A., Batsyn, M., & Pardalos, P. M. (2014). Dynamics of cluster structures in a financial market network. *Physica A: Statistical Mechanics and its Applications*, *413*, 523–533.

Lee, M. (2022, July). *Bab2min/tomotopy: 0.12.3* (Version v0.12.3). Zenodo. https://doi.org/10.5281/zenodo.6868418

Lee, Y., Thompson, J. R., Kim, J. H., Kim, W. C., Fabozzi, F. A., Fabozzi, F. J., Musumeci, J., Feibel, B., Cornell, B., Nagy, Z., et al. (2023). An overview of machine learning for asset management. *The Journal of Portfolio Management*, *49*(9), 31–63.

Liang, D. (2016). *Predicting stock price changes with earnings call transcripts* [Master's thesis, University of North Carolina at Chapel Hill].

Ma, Z., Bang, G., Wang, C., & Liu, X. (2020). Towards earnings call and stock price movement. *arXiv preprint arXiv:2009.01317*.

Macanovic, A. (2022). Text mining for social science–the state and the future of computational text analysis in sociology. *Social Science Research*, *108*, 102784.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, *11*, 193–197.

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of sentiment analysis in finance: From lexicons to transformers. *IEEE access*, *8*, 131662–131682.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Mullen, T., & Malouf, R. (2006). A preliminary investigation into sentiment analysis of informal political discourse. *AAAI spring symposium: computational approaches to analyzing weblogs*, 159–162.

Nanda, S., Mahanty, B., & Tiwari, M. (2010). Clustering indian stock market data for portfolio management. *Expert Systems with Applications*, *37*(12), 8793–8798.

Nguyen, T. H., & Shirai, K. (2015). Topic modeling based sentiment analysis on social media for stock market prediction. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1354–1364.

Papadimitriou, C. H., Tamaki, H., Raghavan, P., & Vempala, S. (1998). Latent semantic indexing: A probabilistic analysis. *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, 159–168.

Perico Ortiz, D., Schnaubelt, M., & Seifert, O. (2023). A topic modeling perspective on investor uncertainty. *FAU Discussion Papers in Economics*.

Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, *36*(4), 992–1011.

Roos, M., & Reccius, M. (2024). Narratives in economics. *Journal of Economic Surveys*, *38*(2), 303–341.

Sandhiya, R., Boopika, A., Akshatha, M., Swetha, S., & Hariharan, N. (2022). A review of topic modeling and its application. *Handbook of Intelligent Computing and Optimization for Sustainable Development*, 305–322.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of business*, *39*(1), 119–138.

Solberg, L. E., & Karlsen, J. (2018). *The predictive power of earnings conference calls: Predicting stock price movement with earnings call transcripts* [Master's thesis, Norwegian School of Economics].

Sroka, E. C. (2020). Don't be Afraid of Nonparametric Topic Models (Part 2: Python). Retrieved March 11, 2024, from https://towardsdatascience.com/dont-be-afraid-of-nonparametric-topic-models-part-2-python-e5666db347a

Storm, H., Baylis, K., & Heckelei, T. (2020). Machine learning in agricultural and applied economics. *European Review of Agricultural Economics*, *47*(3), 849–892.

Sun, X. (2014). Textual document clustering using topic models. *2014 10th International Conference on Semantics, Knowledge and Grids*, 1–4.

Teh, Y., Jordan, M., Beal, M., & Blei, D. (2004). Sharing clusters among related groups: Hierarchical dirichlet processes. *Advances in neural information processing systems*, *17*.

Valdez, D., Pickett, A. C., & Goodson, P. (2018). Topic modeling: Latent semantic analysis for the social sciences. *Social Science Quarterly*, *99*(5), 1665–1679.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Natural language based financial forecasting: A survey. *Artificial Intelligence Review*, *50*(1), 49–73.

Yau, C.-K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics*, *100*, 767–786.

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. *Advances in Information Retrieval: 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings 33*, 338–349.