DECLARATION OF AUTHORSHIP

Candidate number:

Title of thesis: Using Categorical Data to Facilitate Learning for Financial Disclosures; Applications of the Attribution Dictionary

(The thesis falls under the school of management)

Word count: 14,994

Please tick to confirm the following:

I have read and understood the University's disciplinary regulations concerning conduct in examinations and, in particular, the regulations on plagiarism (*The University Student Handbook* Section 8.8; available at https://www.ox.ac.uk/students/academic/student-handbook).

I have read and understood the Education Committee's information and guidance on academic good practice and plagiarism at https://www.ox.ac.uk/students/academic/guidance/skills?wssl=1 .

The thesis I am submitting is entirely my own work except where otherwise indicated.

It has not been submitted, either partially or in full, either for this Honour School or qualification or for another Honour School or qualification of this University (except where the Special Regulations for the subject permit this), or for a qualification at any other institution.

I have clearly indicated the presence of all material I have quoted from other sources, including any diagrams, charts, tables or graphs.

I have clearly indicated the presence of all paraphrased material with appropriate references.

I have acknowledged appropriately any assistance I have received in addition to that provided by my [tutor/supervisor/adviser].

I have not copied from the work of any other candidate.

I have not used the services of any agency providing specimen, model or ghostwritten work in the preparation of this thesis/dissertation/extended essay/assignment/project/other submitted work. (See also section 2.4 of Statute XI on University Discipline under which members of the University are prohibited from providing material of this nature for candidates in examinations at this University or elsewhere: http://www.admin.ox.ac.uk/statutes/352-051a.shtml).

[*Insert subject-specific statements as necessary*]

I agree that this electronic copy may be used should it be necessary to confirm my word count or to check for plagiarism.

1 http://www.ox.ac.uk/students/academic/guidance/skills

Candidate number: 1173690

Using Categorical Data to Facilitate Learning for Financial Disclosures
Applications of the Attribution Dictionary

Thesis in Management, 2021

Abstract

This paper aims to develop natural language understanding over corporate annual reports (87,834 MD&A sections in 10-Ks), combining theories on corporate strategy and finance with machine learning and key words lists to predict corporate performance. In particular, the techniques used captured how organizations describe their strategies (eg. actions, competencies), their external environment (eg. competitors, economic environment) and the performance results attributed to them. I find that the ways in which firms have disclosed their strategies are strongly associated with their performance returns during the financial crisis and the Dot-Com-Bubble. I use disclosures as a proxy for firms' strategy and I find that various strategic content discussed have different precedent causal effects on firms that are performing well and firms performing badly. There is little published research in this field. The dataset utilized in this paper is the largest ever since for the purpose of returns prediction. Overall, I hope this paper can elicit future research in the area of topic modelling within finance.

TABLE OF CONTENTS

1 Introduction

Advances in computing have made the analysis of textual data increasingly tractable. In finance, recent studies have addressed how financial markets respond to the language in newspaper articles (Tetlock, 2007; Tetlock, Saar-Tsechansky & Macskassy, 2008), earnings reports (Loughran & McDonald, 2011), and various types of regulatory disclosures (Hanley & Hoberg, 2012). However, the NLP technology used in firm performance/stock price prediction is nascent. The most influential studies in this area are Tetlock (2007) and Loughran and McDonald (2011). They evaluate sentiment by weighting terms based on a pre-specified sentiment dictionary and summing sentiment scores. A smaller volume of research also uses textual data to infer information on strategic management or organizational behaviour perspective and observe the relation between current disclosure and future firm performance. The textual data used for these studies were often obtained through manually encoded texts.

Nevertheless, there are various issues associated with each of these methods. Existing sentiment dictionaries capture polarity (for instance, how positive/negative a firm's filing is, by counting the number of positive/negative words in the document) but not context (the subject matter the firm is feeling optimistic/pessimistic about). Manual labelling has the downside of human errors and sample size restriction. Papers using this method usually cover fewer than 100 companies and across a maximum time span of five years, as researchers must manually search through the texts. No research thus far has looked into creating a standard workflow that can be replicated in future studies.

On a high level, the common issue lies in the fact that each fails to account for different unobserved features of financial text. I argue that caveats of previous analysis methods lies in the following:

1. NLP toolkits focus on predictions of stock returns, but they do not provide the means to understand the underlying company fundamental value drivers reflected in stock price reactions. There are no current techniques that enable evaluation of discussions of firm decision making, firm actions and thought process.

2. Past research controlled for insufficient characteristics within firm's textual disclosures. For example, only the sector of the firm may be accounted for, but not characteristics of the economy or the institution in which the firm is situated, and the macroeconomic events that the firm is affected by.

This paper serves to bridge this gap by combining financial dictionaries and unsupervised machine learning to create an automated text analysis workflow that future researchers can replicate to form a predictive model for future firm performance. I account for the following factors to arrive at a predictive model:

1. Disclosure on features internal or external to the firm, understanding the implications of specific organizational efforts (eg. "strategic acquisition", "superior management") on financial performance. Most literature relating to this analysis involves manual encoding. There is currently no academic literature that works on quantifying these features. I make this possible by leveraging a relational dictionary.
2. In addition to counting sentiment words (eg. "poor", "positive") in the reports, I developed a system to identify financial performance vocabulary (eg. "increase in revenue") which are more likely correlated with actual superior performance for a firm and predictive of how the firm will perform in the future.
3. I control for the high level textual characteristics composition (eg. accounting for the topic the document most likely belongs to, be it regulatory/compliance discussion, investments, securities and derivatives), as well as sector/industry of the business.
4. To instrument the above discussion, I study the macroeconomic events surrounding the business.

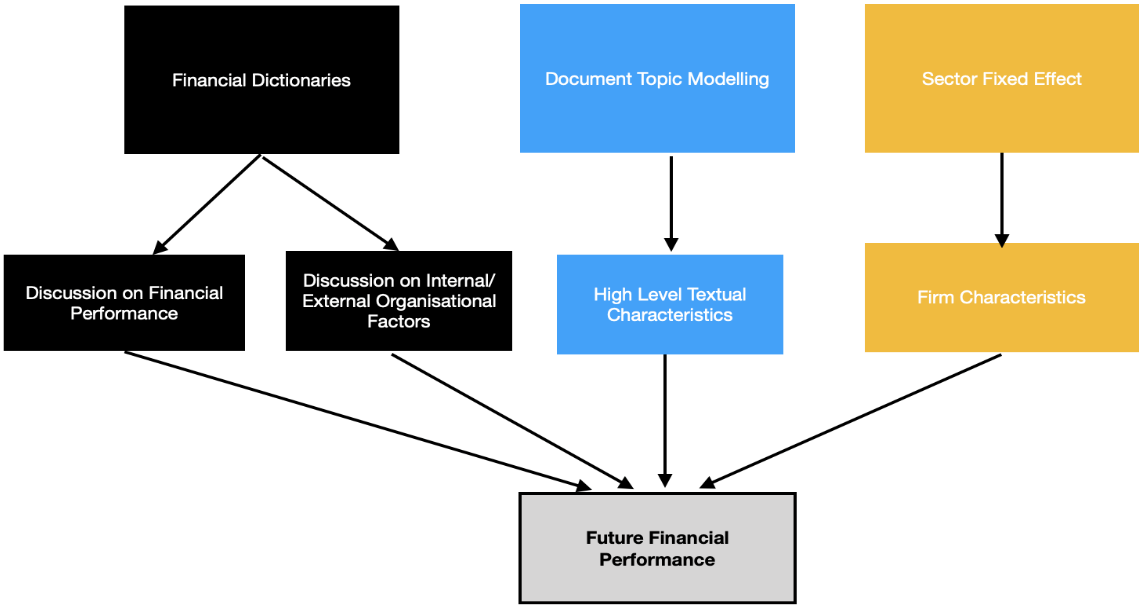


Figure 1. Proposed Prediction Model

The textual data used in this paper comes from 80,000 MD&A (management discussion and analysis) sections of firms' disclosure, and each is about 1000-6000 words. Located in companies' 10-Ks, the MD&A gives the investor an opportunity to look at the company through the eyes of management by providing both a short and long-term analysis of the business of the company. Yet, given that the very purpose of the MD&A is focused on justifying performance with strategy/actions taken by the firm, it is important for a researcher to associate rationales behind performance with the firm's actual performance.

This paper begins by evaluating current methods of NLP in the industry and related background research. The empirical part of this paper is divided into two sections. First, I deploy topic modelling techniques to study the evolution of MD&A disclosure content over time and the reasons behind this. Similar to Dyer et al. (2016), I make the observation that readily observable firm characteristics or non-textual characteristics do not sufficiently explain the trends in MD&A vocabulary, thereby supporting my construction of the prediction model. I subsequently explain how I arrive at the prediction model. I explain the strategic management literature that inspired the web of dictionaries used to proxy for financial performance and organizational factors. I examine the correlation between the correlation between textual disclosures and future returns.

2 Theoretical background

Research on predicting firm operating performance is rather limited compared to research on predicting stock returns. Past NLP research utilized pre-determined dictionaries, or supervised machine learning to assess disclosure characteristics (e.g., uncertainty, positive/negative tone) based on human defined classification themes. Others studies used unsupervised learning methods to associate measures of readability, similarity, deception, or length with firm fundamentals.

In this section I introduce tools for textual analysis and their applications to modern NLP financial applications. I describe the use of machine learning methods and dictionary methods deployed in past literature to perform sentiment analysis on financial text.

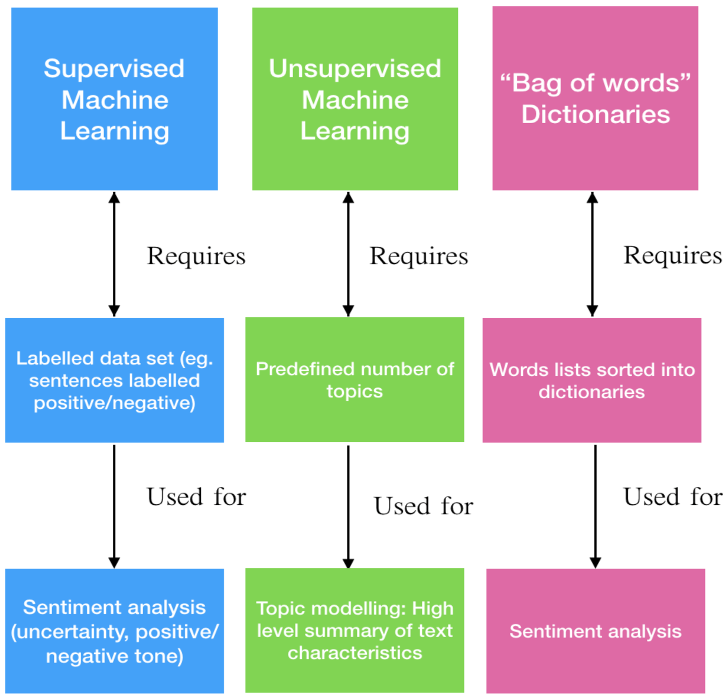


Figure 2. Common NLP methods at present

2.1 Machine Learning Methods

Machine learning methods in the financial NLP field can be grossly categorized into two categories: supervised and unsupervised approaches. In the context of financial NLP, supervised learning requires a labeled dataset, whilst unsupervised machine learning looks for previously undetected patterns in data with no existing labels. Supervised machine learning is used for text classification (ie. deciphering whether a sentence/document conveys positive/negative sentiment), whereas unsupervised machine learning is used for topic modelling, categorizing documents with respect to the topics discussed within them or extracting (latent) topics from texts.

2.1.1 Supervised Sentiment Analysis in stock price prediction

Natural language processing is most commonly applied to study stock performance. The efficient market hypothesis stipulates that investors consider all available information in their decision making process (Fama, Fisher, Jensen & Roll,1969; Fama 1970). Market efficiency implies that any news, once released into the market, will be immediately assimilated into prices. Yet, empirical evidence from natural language processing (NLP) uncontroversially supports the contrary argument that information contained in financial market events is predictive of future asset price paths. The application of NLP in stock price prediction is still nascent compared to other fields, and remains predominantly focused on analyzing sentiments in financial disclosures at the word-level.

Supervised sentiment analysis uses a set of training documents, classified into a set of predefined categories, to generate a statistical model that can be used to classify any number of new unclassified documents. Sentiment scores are then used in a secondary statistical model for investigating phenomena such as stock returns in financial markets (Tetlock, 2005). A critical issue in this field is the lack of classified textual data. A few research agendas seek to bridge this gap: Malo et al. (2014) trained classifiers to conduct sentence-level semantic analysis for financial news and provided a Financial Phrase Bank consisting of a set of 5,000 sentences, manually annotated by 16 subject experts. This resource was updated by Sinha et al (2019), who also released an entity-annotated news dataset containing over 12,000 headlines and their related financial sentiment. Oliveira et al. (2016) produced a stock market sentiment lexicon, which includes 20,551 items extracted automatically from microblogs (StockTwits and Twitter).

However, building a model based solely on these existing data sets is inadequate for two reasons. First, all existing labelled datasets are classified based on polarity, this makes supervised sentiment analysis more applicable to stock price forecasting than to predicting operating performance. Stock prices are highly sensitive to public sentiment but firm performance is down to miscellaneous factors (eg.

management). Polarity of news/financial filings may be an effective proxy for public sentiment but less so for the components leading to firm performance because it is more about how decision making of the firm leads to long run impact. Second, in the process of producing a labeled dataset, annotators reviewing financial text would assign a positive/negative tag to isolated sentences. Yet, in an actual news event, some positive/negative textual description may already assimilated into prices and others might not, thus amplifying the problem of ambiguous causality. For example, prices of stock X may already reflect the fact that numerous articles hypothesize that stock X might fall. The ambiguity in the textual context reduces the predictive ability of the model. As a means of disambiguation, it would be preferable if documents are labelled based on multiple features instead (eg. certainty, financial content, etc). Gathering labelled data for the task is strenuous and requires the expertise of financial analysts. At present, there is no available tagged dataset for feature engineering.

Additionally, machine learning approaches to sentiment analysis are subject to criticism due to their lack of transparency: Most supervised machine learning utilize methods such as Naïve Bayes (Antweiler and Frank, 2004), LSTM networks (Maia et al, 2018) and neural networks (Kraus and Feuerriegel, 2017). All of these methods use unpublished rules and filters to measure the context of documents, and hence are opaque and difficult to replicate.

An alternative solution researchers developed is to use stock returns to screen for sentiment charged words, and use those as the labelled dataset. Ke et al (2019) designed a workflow that could screen for sentiment charged words based on their cooccurrences with stocks of high/low returns. However, this method would still fail to capture the nuance of the financial language, as it is still a “bag of words” model that does not take into account the importance of syntax. It does not account for negation, hence it is unable to distinguish between “decrease in debt” and “increase in debt”. Additionally, the classification algorithm would not pick up phrases, as it only uses single word tokens.

In this paper, I choose neither to adopt a pre labelled dataset nor to utilize the aforementioned forms of supervised machine learning. It would not be possible to rectify the problems inherent in the techniques and the datasets. I also recognize that new patterns would emerge in future texts that lessen the predictability of currently available pre labelled texts.

2.1.2 Latent Dirichlet Allocation in Topic Modelling

Topic modelling can be described as a method for finding a group of words from a collection of documents that best represent the information in the collection. Before the advent of topic modelling, researchers relied on manual classifications to control for the differences between documents. Boudoukh et al. (2013) use an ex ante list of 14 predefined categories (such as "acquisition, deal, legal or award") to differentiate between relevant news for companies to study the impact of news on abnormal stock return. Gooding and Briscoe (2019) used paid data from All Street Research, containing 3097 instances, with categories defined by analysts which they narrowed down to 1824 examples and 11 categories. This study suffered from a small sample size: several categories containing less than 100 examples which meant that they were not enough to train and test.

Later research shows further integration of topic modelling with machine learning methods, especially Latent Dirichlet Allocation (LDA), an unsupervised machine learning algorithm aiding researchers to extract a number of predefined topics from a collection of financial text. Researchers define the number of topics they seek to extract, and the algorithm finds the most likely choices for these topics and outputs them in terms of word vectors. Past research focused on using LDA to predict currency fluctuations, equity returns and examining validity and truthfulness of corporate disclosures. LDA has also allowed topic modelling to be scaled to large samples.

Author	Text studied	Purpose
Jin et al (2013)	Bloomberg news articles	Currency fluctuations
Bao and Datta (2014)	10-K risk disclosure section (section 1A)	Summarize risk-related topics
Hoberg, and Maksimovic (2014)	10-K MD&As	Corporate disclosure quality assessment
Dyer, Lang and Lawrence (2016)	10-K	Financial text evolution with respect to new FASB and SEC requirements
Feuerriegel et al (2016)	German ad-hoc press releases	Abnormal returns of stocks
Hanley and Hoberg (2016)	Bank 10-Ks	Potential systematic risk identification
Huang, Lehavy, Zang & Zheng (2014)	Analyst reports and the text narrative of conference calls	Thematic Content Comparison

Figure 3. Summary of LDA literature and their contributions at present

One of the downsides to LDA is the occasional lack of human interpretability and spurious results, often due to the complex nature of financial language. Computationally distinguishing between topics referring to "firm inherent competencies" and "changes enabled by management" using a statistical model is far more difficult than distinguishing between documents referring to "music" and "animals", because, in the case of the former, there are far more words that co-occur in both contexts. Hence, LDA would fail to precisely account for all existing topics. However, this problem would be rectified if these potential topics are further categorized with human efforts. This is what I will attempt to do in the empirical analysis of this paper: I describe how using a relational dictionary may be used to capture this information.

2.2 Dictionary Methods

As opposed to machine learning methods, dictionary based methods have the benefits of being friendly to the exercise of human backtesting. Dictionaries are word lists grouped into categories, with each category defined by its associated attributes, used to assign tags to financial texts, as a result of the tagged textual document predictions on stock returns can be made.

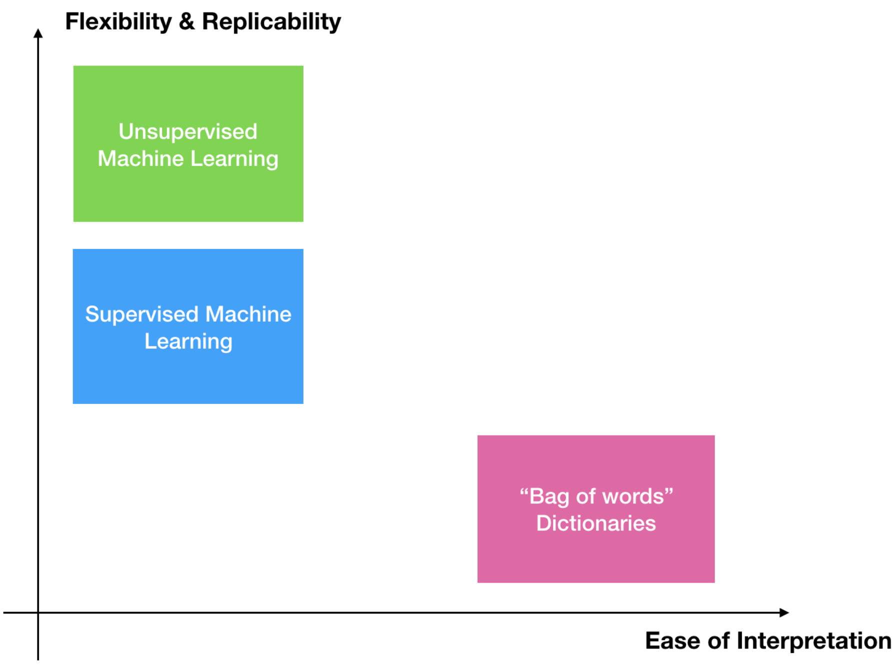


Figure 4 .

Past dictionaries were created to classify the tone of the documents. The studies pre-dating the paper have almost exclusively relied on generic dictionaries. Tetlock (2007) was among the first to demonstrate the benefits of using the Harvard General Inquirer (a general purpose dictionary) in a financial context. There are also other variants which use a combination of heuristics, WORDS (e.g. Frazier et al., 1984), Diction or Wordstat to perform dictionary-based searches for sentiment cues (Demers & Vega, 2010; Davis, Piger, & Sedor, 2006).

The most widely used finance specific sentiment dictionary to date is created by Loughran and McDonald (2011). Their main contribution is to point out that many words that have a negative connotation in other contexts, like tax, cost, crude (oil) or cancer, may have a positive connotation in earning reports. For example, a healthcare company may mention cancer often and oil companies are likely to discuss crude extensively. Loughran and McDonald (2011) conclude that as much as 73.8 percent of Harvard General Enquirer’s negative words do not have a negative sense in financial documents. To this date, Loughran and McDonald dictionary is the most popular dictionary used in a financial context due to its simplicity and clarity. The dictionary consists of words grouped into multiple categories, such as ‘neg’, ‘pos’, ‘uncertain’, ‘litigious’ and ‘constraining’. Entries under each category are in the form of a single word and their possible inflections (eg. loss, losses, lossed).

However, the LM dictionary may still yield biased results. Firstly, it covers only unigrams (single-word dictionary entries). I seek to build on the LM dictionary by developing a more sophisticated lexicon consisting of both unigrams and bigrams (single-word and two-word phrases). Using the LM dictionary, we would classify an improvement in economic condition or refinement of company strategy as positive factors. The former is entirely out of the control of the firm whereas the latter is the result of actions taken deliberately by the firm. It may well be that the improvement in economic condition is well known and priced in whereas opinions on actions taken by the firm is not. Hence, only the latter may be of material use to analysts. The LM dictionary also fails to capture descriptions of performance outcomes such as “increased revenue” (positive) and “decreased cost” (negative) because “increased” and “decreased” are classified as neutral.

3 Initial Vocabulary Modelling

Before formulating the research hypothesis, I seek to examine trends underlying textual MD&As using LDA and create interpretable results of the broad topics discussed, in an objective manner.

3.1 Sample Selection and Construction

The management discussion and analysis section (MD&A) in corporate 10-Ks is one of the most read and important components of the financial statements. Most literature finds a significant correlation between current fundamentals and market reactions and textual disclosures.

I webscrape 10-K filings filed electronically from SEC EDGAR from 1993 to 2018. This yields 149,139 10-K and 10-K405 filings from which I was able to parse the Management’s Discussion and Analysis. Management’s Discussion and Analysis comprises Item 7, which usually describes the results of operations, internal and external factors relevant to the business’ performance, and Item 7A, Qualitative and Quantitative disclosures about Market Risks. I have excluded 10-K-A from our sample and eliminated disclosures that contain fewer than 1000 characters, which are disclosures without material information or disclosures that have MD&A section incorporated by reference to the annual report (usually incorporated into exhibit 13, which is kept as a separate file to the main 10-K filing on SEC Edgar). In the latter case, similar to Loughran and McDonald (2011), I find that the beginning and ending positions of the MD&A document when filed in an exhibit are not demarcated in a manner that facilitates accurate parsing. Aside from the MD&A section, exhibit 13 often contains financial statements, notes, as well as the auditor's report, all of which is irrelevant for the purpose of my exercise.

I designed the matching algorithm to capture the position of “item 7. management’s discussion and analysis” and “item 8. consolidated /audited financial statements” and extract the text in between when it satisfies certain heuristics. I remove all numbers and numerical tables, keeping only the text. I identified 87,834 MD&A sections that fulfill this requirement.

I subsequently perform several text preprocessing steps that are common in text mining (Manning and Schütze, 1999). I transform the running text into a matrix notation that would allow for further calculations with the “Scikit learn” package in Python . First of all, I remove stop words that frequently occur in the English language. I use the default list of English stop words in Python’s NLTK package which consists of common, short function words that do not add additional meaning to our text - examples of these are conjunctions, prepositions and pronouns such as “ourselves”, “her”, and “between”.

3.2 Topic Modelling

First, I would like to extract high level features and interpret what these statistical results mean in a managerial context with unsupervised machine learning. I analyze the evolution of topics using LDA (Blei, 2003). LDA is a robust method that relies on statistical correlations among words in a large set of documents, it is a dimensionality-reduction technique, similar to principal components analysis, which transforms high dimensional textual data to low dimensional data. The fundamental challenge with any NLP procedure is that raw text suffers from the curse of dimensionality, which makes it computationally intractable. LDA would allow me to explicitly identify and empirically quantify the low-dimensional representation so that it retains meaningful properties of the texts. More information on the statistical definition of the method and the code can be found in the appendix..

3.2.1 Topic length evolution

The first set of empirical results aims to study the changes in topic distributions over time. Dyer (2017) studied the evolution of the changes in length of topics on whole 10-Ks. However, no up to date research has been done on the MD&A section. Given that the content in the MD&A confers the most amount of information on a company’s strategy and operations, I hypothesize that results from running LDA over the MD&A will differ meaningfully from Dyer (2017). I replicate Dyer (2017) to study the evolution of length of disclosure for MD&As instead.

I first used LDA to output 25 clusters from the set of MD&A documents. I then assigned each document to their most probable cluster and collected the length of MD&A documents for each topic across the time frame of 1993-2018. I compute the median length of all the documents that are categorized into each topic in each year and plot the observations graphically (see figure 6.). I show in a tabular form the broad clusters in which I group the computer generated clusters from LDA (see table 1). In general, clusters fall under two major categories: sector-specific (upper table) and business/operations specific (lower table). For sector-specific clusters, only the most prominent sectors by US GDP (eg. finance, real estate, oil and gas, healthcare) having the most distinctive vocabulary (eg. "futures", "mortgage", "oil", "drug") are clearly demarcated by the algorithm. For business/operations specific clusters, the output of LDA insufficiently distinguishes between the aforementioned strategies dimensions I previously mentioned. The boundaries between several clusters appear murky (eg. Operations Financials) with many clusters having the same vocabulary.

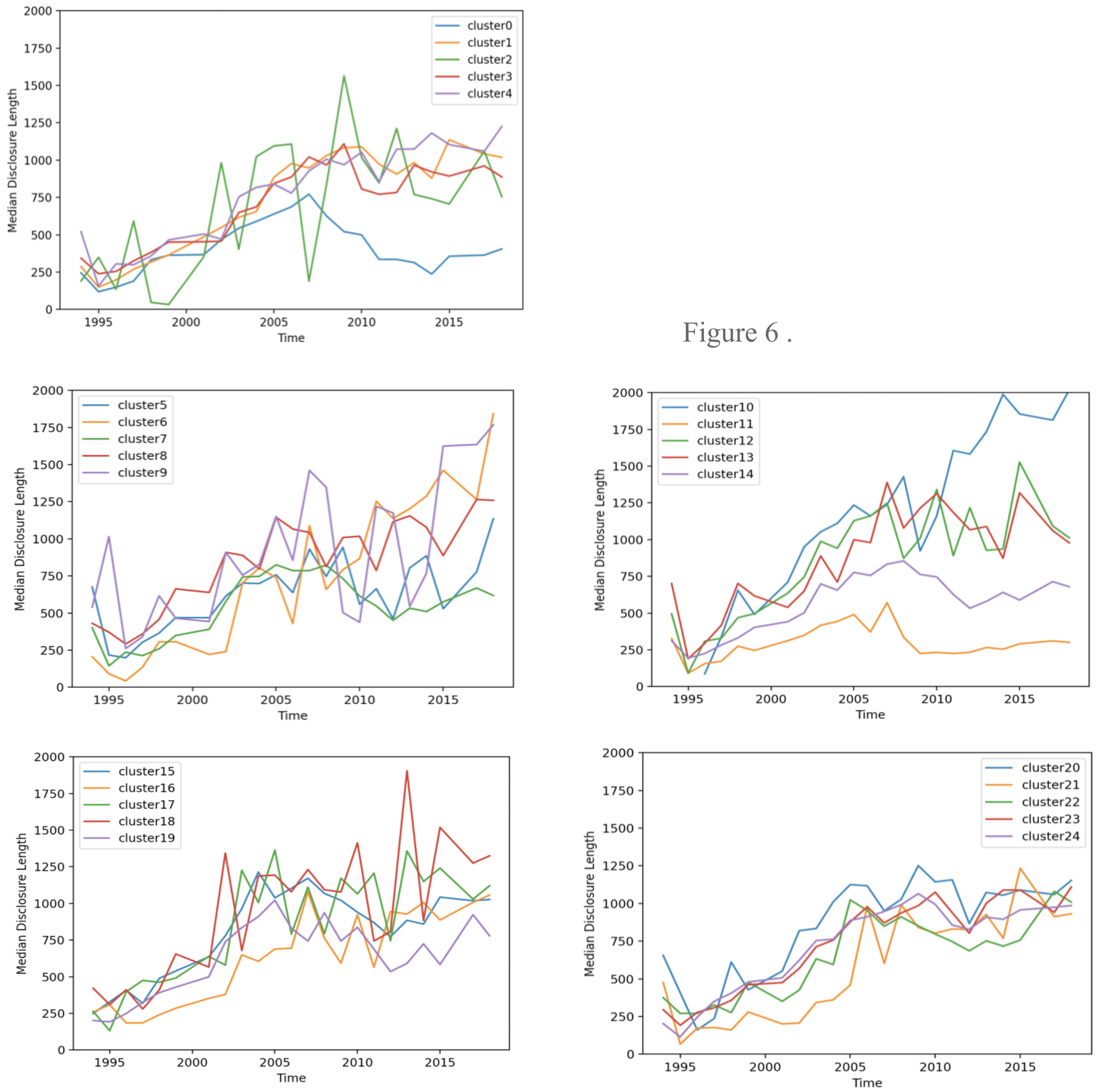


Figure 6 .

Cluster title	Corresponding Description	Examples of highly ranked words
22	Oil and Gas Production; Power and Energy	“gas”, “oil”, “natural”, “power”, ”production”, “prices”
21	R&D; Drug development	“product”, “development”, “research”, “regulatory”, “drug”
6,8	Real Estates	“loans”, “bank”, “mortgage”, “rate”, “assets”, “estate”
20	Derivatives; Trading	“energy”, “trading”, “price”, “futures”
5	Retail; Warehouse	“warehouse”, “rental”, “retail”
Cluster title	**Corresponding Description**	**Examples of highly ranked words**
0,7,9,11,12,14,	Operations financials (Revenue, Cash, etc)	“million”, “increase”, “revenue”, “cash”, “income”
16	Valuation	“rate”, “futures”, “operations”,”value”,”revenue”
18, 23	Products; Services; Customers	“products”, “financial”, “sales”,”customers”,”market”
2	Securities	“stock”, “credit”, “securities”, “rate”
17	Management	“financial”,“income”, “cash”, “management”, “operating”
13	Decision making; Evaluation	“may”, “could”, “ability”, “subject”
10	Environmental	“environmental”, “may”, “products”, “costs”
1,24	Supply Chain	“partnership”, “distributions”, “net”, “development”
15	Insurance	“insurance”, “loss”, “costs”, “year”
3,4,19	Corporate Borrowing	“interest”, “expenses”, “tax”

Table 1.

Looking at the first stage results obtained, a general trend is evident: disclosures have become longer over time across all topics (though not necessarily increasing across all years). This observation is different from that of Dyer et al (2016), who, after conducting the same analysis on the whole sample of 10-Ks, found that only documents pertaining to compliance with SEC & accounting standards increased in length markedly in the sample period. The finding aligns with Dechow et al.(2010), who observed that with increasing MD&A length, managers increasingly use boilerplate disclosure (ie. standard disclosure that uses many words with little firm-specific content). As the lack of concision of MD&A may reduce the value of the information MD&As provide, the need to arrive at an automated way to identify essential information becomes much more essential.

Additionally, most of the clusters that experience slow growth in length over time (eg. cluster 0, cluster 11, cluster 14) relate to revenue, cash, and operating financials (although cluster 9 is an outlier). These clusters are less analytical and focus on describing material performance. On the contrary, cluster 18 (products, services, customers) and cluster 13 (decision making, evaluation) experience more fluctuations in length. This tells us that much of the increase in overall MD&A length can be explained by the addition of more strategic content, rather than "boilerplate" content that solely describes the organization's performance.

Thirdly, amongst all 25 clusters, the cluster relating to environmental concerns experienced the most dramatic increase. This reflects that ESG has become a quintessential part of corporate disclosure over time. The Sustainability Accounting Standard Board (SASB) was established in 2011 to develop standards for companies to make comparable, consistent, comparable and reliable disclosure about sustainability or ESG matters. The increase in disclosure length on environmental concerns certainly demonstrates progress in a regulatory sense. However, information from LDA is insufficient to tell us if firms that disclose more ESG content are likely to witness improved performance. In parallel with O'Donovan (2002), firms may be using specific micro-tactics dependent on whether the purpose of the response is designed to gain, maintain or repair a firm's environmental legitimacy (that is, to act within the bounds of what society identifies as socially acceptable behaviour). In the subsequent part of this paper, I discuss the correlation between more environmental disclosure and firms' performance.

Lastly, discussion clusters exhibit different degrees of cyclicality. In the event of a financial crisis, the length of topics relating to the external economy and debt increases whilst those relating to internal operations decline. For example, cluster 2 (securities), 3 (corporate borrowing) , 6 (real estate) , 20 (derivatives, trading) have seen the most increase in length post financial crisis (2018-2019). On the other hand, cluster 9 (operations financials) and 16 (valuations) have seen the most significant decline in length. An explanation can be made on the basis of the attribution theory: firms tend to attribute good news to own superior management and bad news to external reasons. Given that the financial crisis is a systems wide event, firms would likely shorten the discussion of poor operational performance and describe the crisis in great detail.

3.2.2 Emergence of New Topics

The study on topic length evolution reveals interesting cross sectional characteristics of MD&A texts. Nevertheless, as observed, some topics have become immensely more popular across time. Hence, I subsequently analyze the emergence of new topics over time. I instead use LDA to select the topic with the most different content with respect to the 25 topics in previous year. In doing so I iteratively find the topic that is an "outlier" with respect to the content of the other topics in the previous years.

As the topics calculated by the LDA model are a vectorised representation of words, it is possible to calculate how similar they are with the use of cosine similarities, defined as the degree of similarity between two sparse vectors. The algorithm first extracts the top 1000 words that are most probably assigned to the respective topic, outputs these words in vector form. Subsequently, the sum of the dot products of these vectors is computed and the result is normalized by the multiple of the magnitudes of these vectors (see Figure 7). For each year, I compute a rolling set of cosine similarity calculations by comparing each topic with every topic in the previous year. I take the mean of the former and find the topic with the smallest mean (least similar to all the topics in the previous year).

$$argmin_{\ a} \sum_{b=0}^{25} cosinesim(C_{\ at}, C_{\ bt-1})$$

$$= argmin_{\ a} \sum_{b=0}^{25} \frac{C_{\ at} C_{\ bt-1}}{|C_{\ at}|\ |C_{\ bt-1}|}$$

$$= argmin_{\ a} \sum_{b=0}^{25} \frac{\sum_{i=0}^{1000} C_{\ ati} C_{\ b(t-1)i}}{\sqrt{\sum_{i=0}^{1000} C_{\ ati}{}^2}\ \sqrt{\sum_{i=0}^{1000} C_{\ b(t-1)i}{}^2}}$$

Figure 7 .

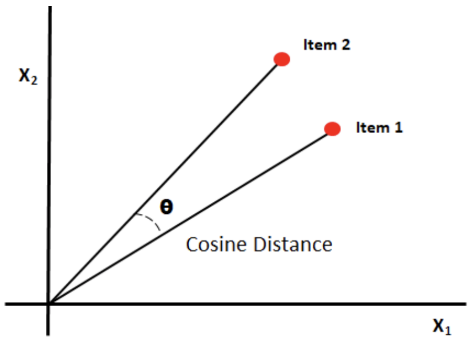


Figure 8 .

For a pictorial illustration of cosine similarities (see Figure 8), for each $t \in [1993 - 2018]$, I choose the Cat to minimize $cos\, \theta$ in the diagram (or similarly, to maximize θgiven its bounds: 0<θ<180) , with 'item 1' and 'item 2' being word vectors respectively of 1000 length.

Table 2. shows the resulting topics each year from the analysis with the lowest cosine similarity with all previous years. I broadly categorized topics into the following categories:

- **Orange**: global macroeconomic event
- **Yellow**: regional macroeconomic event
- **Green**: macroeconomic trend
- **Blue**: management related

1995	1996	1997	1998	1999	2000	2001
“mexico”	“gas”	“electric”	“medicare”	“software”	“media”	“fuel”
“peso”	“oil”	“power”	“patients”	“computers”	“ventures”	“plant”
“fiscal”	“depletion”	“energy”	“therapy”	“merchandise”	“cable”	“nuclear”
“bankruptcy”	“agreement”	“new”	“occupancy”	“hardware”	“operating”	“generation”

2002	2003	2004	2005	2006	2007	2008
“gas”	“PCs”	“yankee”	**“sales”**	**“management”**	“prices”	“fund”
“power”	“network”	“nuclear”	**“products”**	**“credit”**	“trading”	“credit”
“electric”	“telephone”	“environmental”	**“income”**	**“growth”**	“energy”	“futures”
“energy”	“ended”	“decommissioning”	**“tax”**	**“advertising”**	“long”	“markets”

2009	2010	2011	2012	2013	2014	2015
“value”	“futures”	**“debt”**	“store”	“power”	“etfs”	“oil”
“ended”	“dollar”	**“capital”**	“credit”	“energy”	“assets”	“gas”
“fell”	“lower”	**“management”**	“flows”	“idaho”	“nav”	“reserves”
“global”	“investments”	**“accounting”**	“retail”	“environmental”	“commodity”	“development”

2016	2017	2018
“acquisitions”	“recovery”	“regulatory”
“legacy”	“offset”	“changes”
“fell”	“variance”	“credit”
“global”	“coal”	“federal”

Table 2.

First, it is apparent that the time series study captures mostly macroeconomic events, because they are the most surprising. The 1995 reference to “peso” and “bankruptcy” relates to the Mexican peso crisis where the Mexican government was forced to devalue the peso against the US dollar. The period 2007-2010 see financial descriptions that relate to the global financial crisis, with trading ideas being prominent in 2007 and sentiment becoming increasingly bleak in 2009. The clusters also capture macroeconomic factors that were not as internationally known: the reduction of supply and peak in oil price in 1996, Enron’s collapse in 2001, as well as the decommissioning of the Yankee power station in 2004. Several of the clusters also exhibit the emergence of new macroeconomic trends: advancement of new energy in 1997, the dot com bubble in the early 2000s, and the advent of environmentalism in the early 2010s.

Comparatively few topics relate to changes in management practice, highlighted in blue. It may be possible to explain the different timings of topic emergence with the ways in which management seek to attribute performance. Management may seek to discuss external factors when negative shocks occur to defer blame (eg. during the financial crisis). During a boom, management would discuss how the firm is able to capitalize upon the favorable macro trend (eg. during the Dot-Com). When there are no significant shocks to talk about, management discussion focuses more on internal, strategy-related topics. Whilst LDA adequately captures system wide shock and the sectors that experience the most dramatic change over time (eg. software, new energy), it is rather difficult to use LDA to capture changes concerning management related disclosure (eg. strategy-relevant content). Furthermore, results from LDA may be difficult to interpret: in the mid 2010-late 2010s, there are few major macroeconomic events, and results from LDA are difficult to interpret objectively.

4 A model for document information

Although the use of LDA provided insightful results into the content and evolution of MD&A disclosures, it is subject to important caveats. First, the clusters produced require interpretation by the researcher, which are made ex-post based on human intuition, hence they may not be scientifically or statistically reasonable. Second, information from LDA would still be unable to capture a complete picture of textual disclosure to sufficiently explain corporate performance. LDA can characterize external factors that are clearly demarcated from the others (eg. macroeconomic factors, sector specific discussion), but it is largely insufficient in capturing corporate discussion that relates to strategy. Following an initial examination of document topics with LDA, I want to come up with a proxy for the information that disclosures contain and arrive at a way to measure these proxies.

4.1 Document information proxies

I want a proxy for all the information that can be inferred from textual data and how this might have implications on corporate performance.

Firstly, I need to account for a firm's current performance, as it may be correlated with future performance. It may be that some form of mean reverting behaviour exists, and firms having filings with positive performance may experience a decline in performance in the longer run.

Second, I find multiple dimensions of the firm's strategies that are critical to a firm's building of competitive advantage. In prominent strategic management literature, there are three complementary views: positional, resource based and value system.

- **The positional view**

Porter (1979)'s theories are at the forefront of the positional approach, describing strategy as "building defenses against competitive forces or finding positions in the industry where the forces are the weakest". The prescriptive value in understanding and applying the "Five Forces Analysis" lies in positioning a firm in a way such that it is less vulnerable to attack within the industry. The use of Porters' Five Forces analysis allows one to identify the potential for a firm in making profit in the industry and allows a firm to determine competitive intensity or industry attractiveness.

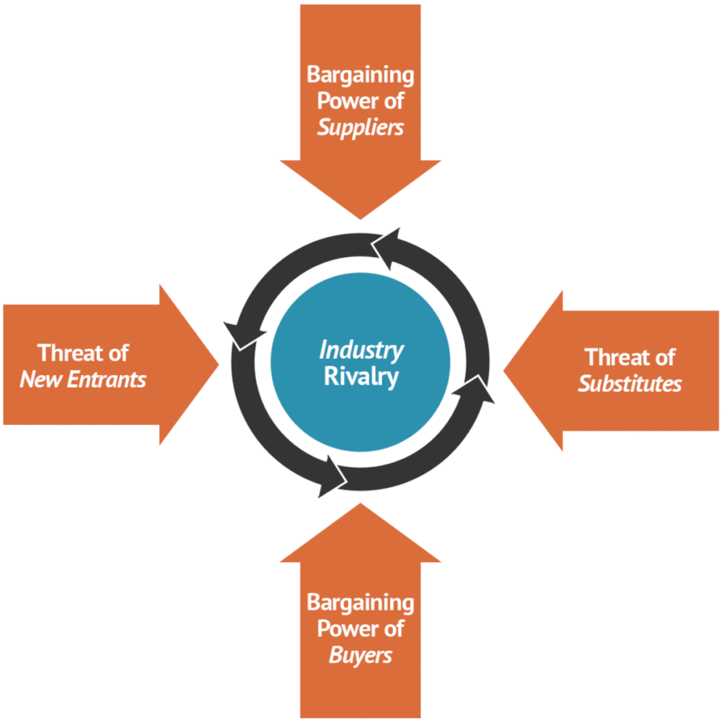


Figure 9 .

- **The resource based view**

The resource based view has a much stronger focus on the internal resources of the firm, which can be defined as the " tangible and intangible assets a firm uses to choose and implement its strategies" (Barney, 2001). These assets come together to shape the key capabilities of a company, through which the company archives a sustained, competitive advantage by leveraging unique firm resources that highly impact its strategy. Companies hold resources that differ across four parameters: value, rareness, imitability, and substitutability, and if a company "discovers" a particular set of unique resources that directly impact its strategy, it is capable of realizing competitive advantage that cannot be replicated by competitors.

- **Value system**

Porter (1985)'s value chain focuses on systems, and how inputs are changed into outputs purchased by consumers. He describes a value chain common to all businesses, that he divides into primary (relating to the primary, sale and support of a service) and supporting activities. Primary activities relate directly to the physical creation, sale, maintenance and support of a product or service (eg. inbound/outbound logistics, operations, marketing/sales). Support activities relate to technological development, infrastructure, human resource management and procurement. Understanding of the value chain could be used by firms to find opportunities to increase value.

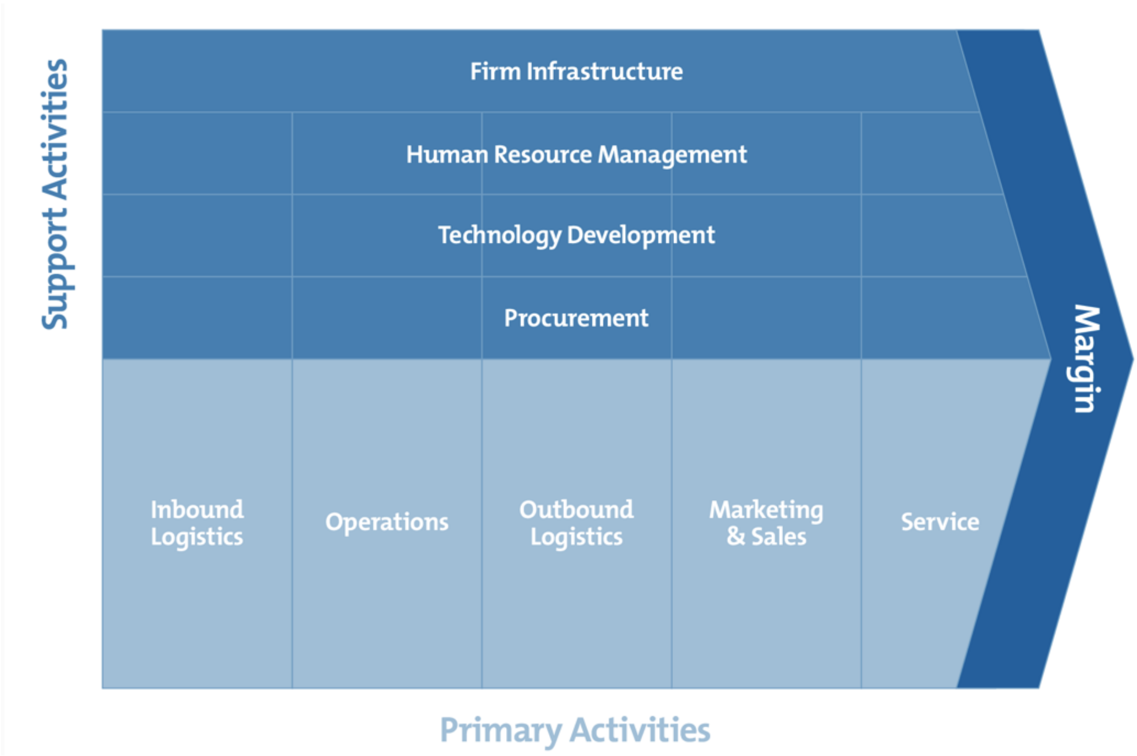


Figure 10 .

Nevertheless, on a textual basis, it may be difficult to differentiate between the strategy dimensions (especially between the resource based view and Porter's value system) at a high level just because there are multiple areas of overlap in terms of vocabulary. For example, superior technology may be discussed as a resource in RBV and as part of the support activities in Porter's value chain. Additionally, it is rather difficult to distinguish between activities in similar categories (eg. between inbound logistics and outbound logistics). In order to incorporate all characteristics of the aforementioned, but at the same time sufficiently differentiate between distinct categories, I choose to regroup strategic discussions into the following categories:

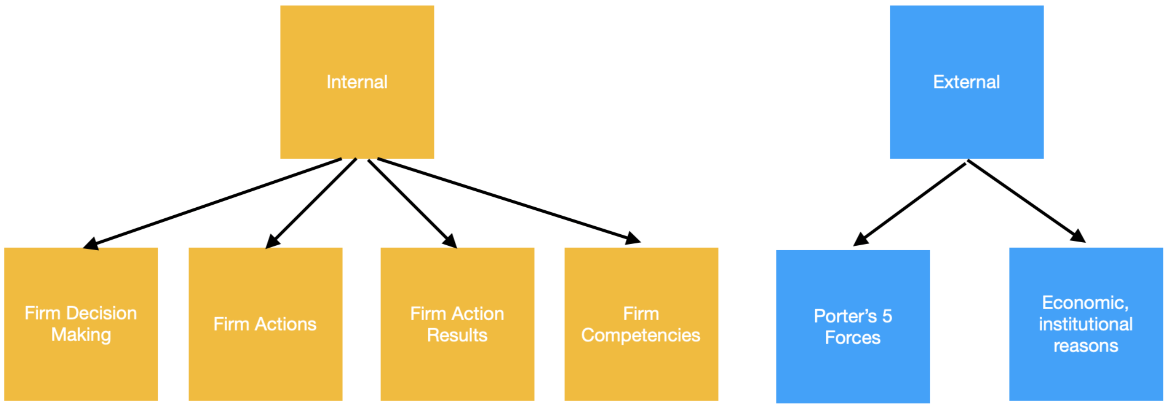


Figure 11 .

The internal dictionary is constructed to mirror the stages of a firm's decision making process and the firm's competencies, and the external dictionary accounts for the forces external to the business. The internal category consists of textual measures that account for firm competency (in alignment with the resource-based-view), and various stages at a firm's strategic management process (Barney, 2001): from decision making to actions to results realization. Each of the steps in the process would correspond to certain firm activities in Porter's value system. The external category consists of Porter's Five Forces and economic, institutional forces that firms discuss.

Aside from strategic discussions, I would also like to collect data on the associated qualifications in terms of sentiment or performance results as a consequence of these strategies. I measure the degree of firms' responsibility taking (if any). Firms may wish to attribute performance to internal/external reasons. In past literature, firms tend to attribute positive results to internal causes while attributing negative performance to the external environment to take credit for successes and avoid blame for failures (Bettman and Weitz, 1983). The directionality of attribution is indicative of future performance. It is widely posited that attributions are likely to be "self-serving". Previous finance research used textual data in corporate reports to analyse the relationship between textual attribution and performance. Staw et al.

(1983) and Salancik and Meindl (1984) find that positive sentiment expressed in corporate annual reports is usually correlated with poor future performance. Bowman (1976) finds that more successful firms place emphasis on their own strategies and less successful firms place blame on external excuses (ie. weather) in their annual reports. Firms that tend to be more optimistic in their disclosure (disclosure of positive revenue increase, etc), may experience worse performance in the future.

5 Quantifying Topics

My research question then focuses on how internal and external disclosure may be determined. Instead of relying solely on LDA, I propose to use an ex-ante method of financial dictionaries to classify the content of filings which outputs a document metric with respect to the categories detailed in Figure 12.

5.1 The Internal/External and Performance Dictionaries

To account for both performance polarity and internal/external attributes, the dictionary needs to be classified into 2 sections.

First, I require a means to quantify textual disclosure of corporate performance. All existing dictionaries are focused on quantifying sentiment (eg. identify cues such as "bad" or "positive") and they are also unigram dictionaries (containing single word tokens). As in this study I am only interested in corporate performance (ie. textual disclosure on financials, such as "revenue" and "cost"), existing dictionaries would not be effective in accomplishing this task.

Second, I need a method of quantifying strategic factors, according to the categories defined in the econometrics model. Past research has relied almost exclusively on using LDA to do so. This is hardly satisfactory as I have shown that LDA does not sufficiently capture several dimensions of strategic content. I instead utilize a set of dictionaries to compute this.

The dictionaries I have constructed contain unigrams and bigrams and they are used to classify performance outcomes and internal/external features. They are constructed with both manual and statistical methods. First, the vocabulary of the available sample is tokenized with each unigram (single word token) and bigram (two words token). The tokenized vocabulary reflects all possible one word, or two-words combinations that can occur in the sample. The tokenized vocabulary is subsequently narrowed down using term weighting with TFIDF, leaving a sample of ~480,000 unigrams and bigrams to be categorised into dictionaries. This removes word combinations that appear infrequently across documents and frequently in single documents, and are thus unlikely to refer to performance outcomes, events common to all firms in the external environment, or typical management actions. Further labelling is conducted by research assistants to classify the remaining vocabulary further into categories in Figure 12.

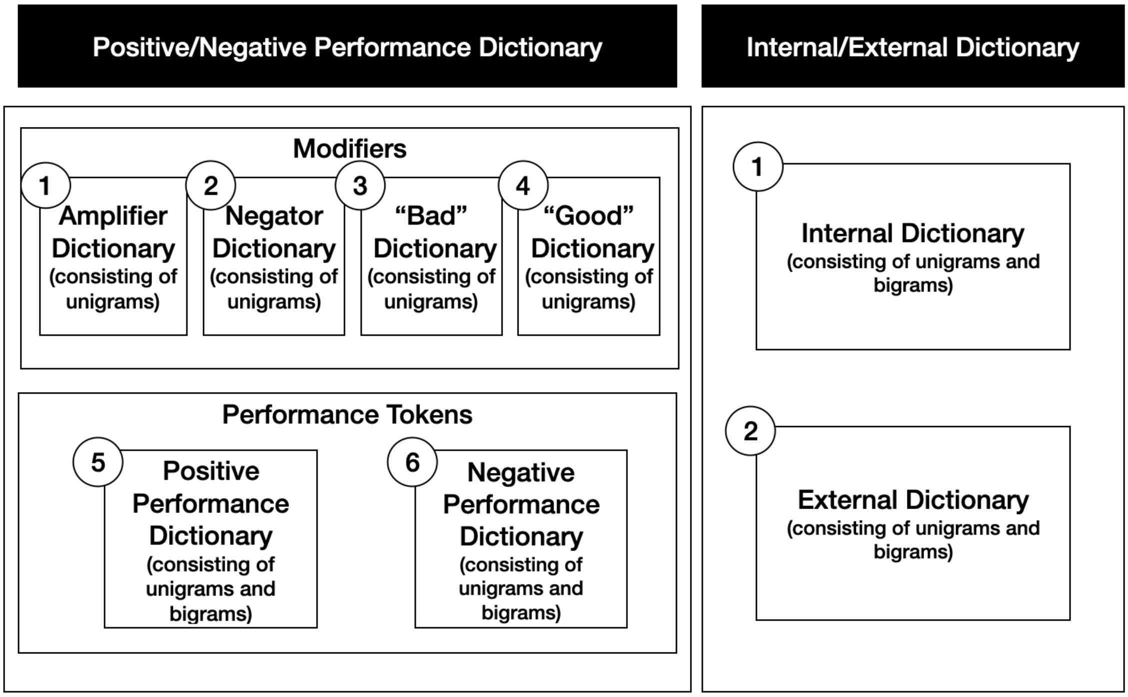


Figure 12 . Dictionary Composition

The performance dictionary consists of 6 sub dictionaries, which are divided into amplifiers/negators/bads/goods (eg. "increase", "decrease", "disaster") and performance financial words (eg. "revenue", "income") (see Figure 12 and 13). Amplifiers are defined as words which enhance the meaning of a performance outcome (e.g. "increase") and negators are those which reverse the sentiment attached (e.g. "decrease").

An amplifier/negator/bad/good and a performance financial word form a phrase group ("increase income"), the polarity of this phrase group is then calculated by multiplying the polarity of its two components. Let "income" to be of polarity 1, whilst "debt" to be of polarity -1. Consider "increase income" (polarity +1), and "increase debt" (polarity -1), "decrease income" (polarity -1), and "decrease debt" (polarity +1). Whilst the polarity of "income" or "debt" is unchanged in the case of the former, it is reversed in the case of the latter because increase is of polarity +1 and decrease polarity -1. The polarities of the performance phrases are multiplied by the polarity of the amplifiers/negators to obtain the resulting polarity for the overall phrase. Aside from amplifiers and negators, the performance dictionary also consists of "goods" and "bads". A "good" (eg. increase, rises) converts any performance phrase associated with it to a positive performance vocabulary; conversely a "bad" (eg. decreases, falls) converts any performance phrase associated with it to a negative performance vocabulary.

The internal and external dictionary consists of unigrams and bigrams that refer to factors affecting performance which are either internal to the firm or, conversely, a product of the firm's external environment. Such internal factors include strategic decisions made by the firm, operational improvements, or performance enhancing organisational strengths such as proprietary technologies or strong management (RBV). External causes of performance largely consist of the competitive landscape (Porter's Five Forces), environmental threats over which management has little control, such as industry competition, legislation, lawsuits, or wider macroeconomic conditions.

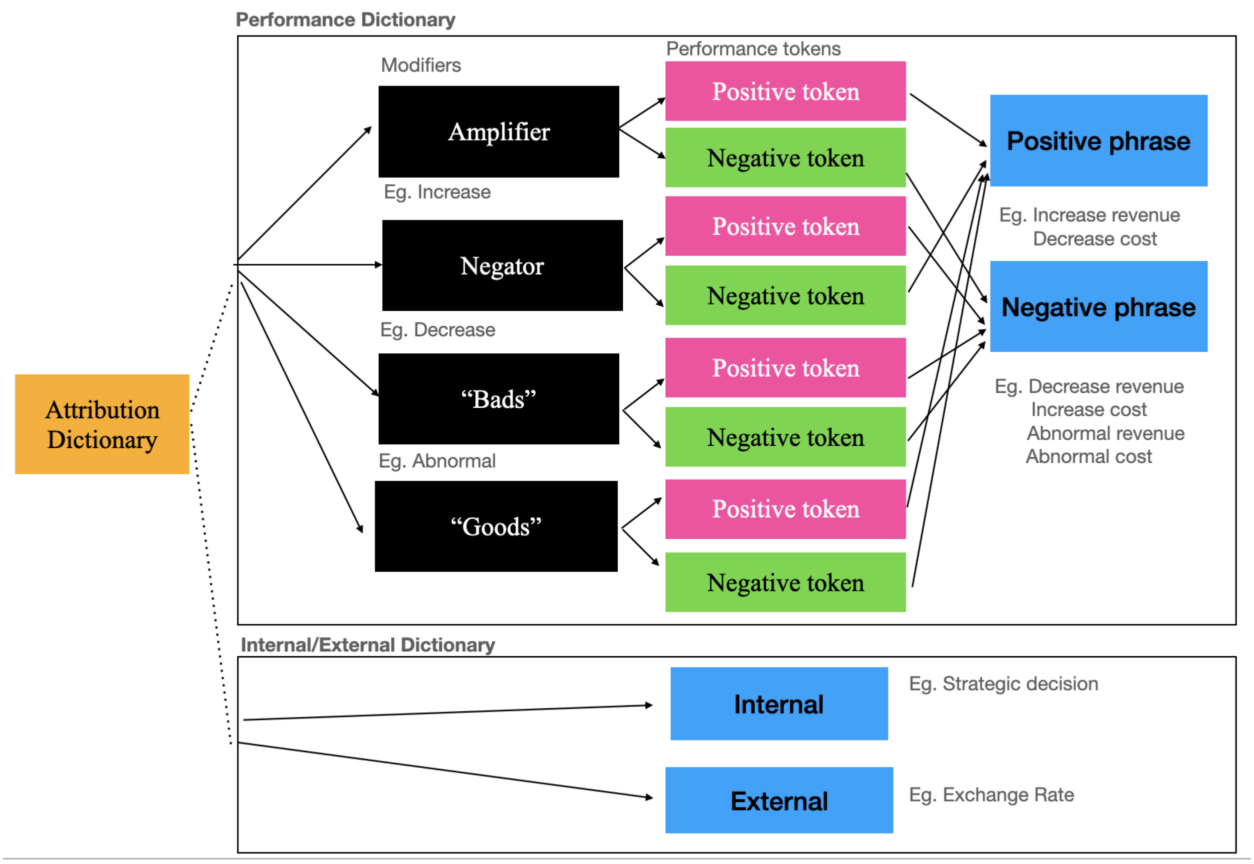


Fig 13: Relationship between individual dictionaries

Below is an example paragraph that reveals how this system of classification differs between the Loughran and McDonald dictionary.

A number of factors may **decrease the income** generated by the centres including the national **economic climate,** the regional and local **economy** which may be **negatively** impacted by **rising unemployment, industry slowdowns,** adverse **weather conditions,** natural **disasters** and other factors, local real estate conditions such as an **oversupply** of or a reduction in **demand for retail** space or retail goods, availability and **creditworthiness of current and prospective tenants,** decreased levels of **consumer spending consumer confidence** and **seasonal spending,** especially during the holiday season when many retailers generate a disproportionate amount of their annual sales, **negative** perceptions by retailers or shoppers of the safety convenience and attractiveness of a center, **acts of violence** including **terrorist activities** and **increased costs** of maintenance insurance and operations including real estate **taxes.**

Fig 14: Comparison between Phrases Identified

Whilst the Loughran and McDonald dictionary identifies the blue tokens that either connote positive or negative meaning, the system of classification I use identifies tokens that fall under the performance vocabulary categories (in red) and internal/external categories (in green).

5.11 The Performance Dictionaries

Further to the general description of the relevant dictionaries, I further define the subcategories in each set of dictionaries.

- An “amplifier” enhances the polarity of the finance specific vocabulary that it is attached to (eg. increased, enhanced, booms, surges, acquire, retain, etc).
- A “negator” negates the polarity of the finance specific vocabulary that it is attached to (eg. decreased, reduces, etc).
- Words in the “bad” category directly makes the polarity of the finance specific vocabulary negative (eg. adverse, constrains, etc).

- Words in the "good" category directly makes the polarity of the finance specific vocabulary positive (eg. positive, amazing, etc).

The positive performance dictionary consists of financial performance tokens and business activities that when amplified, benefits the business (eg. revenue, sales, income, acquisition asset). Conversely, the negative performance dictionary consists of financial performance tokens and business activities that when amplified, negatively affects the business (eg. costs, risks)

Tokens in the dictionaries are used to conduct lookup in the text, so phrases are assigned heavier weights if they are longer (see algorithm design). To ensure the relevance of each entry in the dictionary, only tokens that have a unique meaning are incorporated.

For instance, "accounting cost" is **not** incorporated as "costs" is in our dictionary and "accounting" does not add an additional layer of meaning to "cost". For the same reasons, keeping "certain amortization" is not meaningful because "certain" does not add to "amortization". "Advertising budget" is not included as "advertising" does not add to the polarity of "budget".

However, "Debt maturities" is **relevant** because both debt and maturity are significant to the polarity of the phrase "Expense reimbursements" is relevant because reimbursement alters the meaning of expense.

5. 12 The Internal and External Dictionaries

Further to the initial round of classification, I classify tokens in the "internal" and "external" dictionaries to subcategories as illustrated in Figure 11. Phrases in the internal dictionary are defined to fall under any one or more of the following categories. In the below description, I clarify the relationship between the respective dictionaries and how I constructed the empirical model.

5. 12.1 The Internal Dictionaries

Category I: Firm decision making:

Tokens under this category relate to maintaining, evaluating, altering management decisions and discussion about opportunities.

Porter (1996) states that the role of strategy is to define position, determine trade-offs and forge fit among activities. “Designing” strategy is a prevalent school of thought in the field of strategic management. Mintzberg (1990)’s design school of strategy describes strategy as a process of design to achieve an essential fit between external threat, opportunities and internal distinctive competence.

I hypothesise that firms disclosing more of this type of content are likely to experience future growth in performance. Yet, the converse may also be argued: a company’s choice to enter a new position makes sense only if it has the ability to turn a system of complementary activities into a sustainable advantage.

Table 1. Example Tokens and Associated Bigrams identifying phrases in the “Firm decision making” segment of the dictionaries.

Unigram Token and Associated Bigrams	**Dictionary Unigram/Bigram Frequency**
Competitive (eg. Competitive Strength)	252
Achieve (eg. Transaction Achieved)	287
Solution (eg. Solutions Provided)	631
Discover (eg. Product Discovery)	44

Category II: Firm competencies:

Tokens under this category may include physical capital resources, human capital resources, organizational capital resources, production/maintenance resources, administrative resources, or organizational learning resources and strategic vision resources, examples being innovations, innovative skills, technology, license, intellectual property, rights, rights to patents, copyrights, trademarks, brands, hallmarks, service marks, technical competence, other forms of abilities, etc.

This category serves as a proxy for firms resources, aligning with the resource based view (RBV), profits for firms within one industry differs from profits from another due to differing internal capabilities and barriers to resource acquisition and imitation. Profits for firms within one industry differs from another due to heterogeneity (Barney, 1991) and isolating mechanisms (Rumelt, 1984). Tokens under this category either fall under innovation or unique resources as innovation allows firms to be equipped with resources that are non-imitable or non-substitutable.

Table 2. Key Tokens and Associated Bigrams identifying phrases in the "Firm competencies" segment of the dictionaries.

Unigram Token and Associated Bigrams	**Dictionary Token Frequency**
Software (eg. Software Developed)	736
License (eg. User Licenced)	3120
Property (eg. Security Properties)	767
Develop (eg. Innovation Developed)	2850

Category III: Firm actions:

Tokens under this category relate to M&A activities, partnerships, and investment undertaken. They are a proxy for how organizations take actions to improve their value chain or utilize their resources.

There has been related research examining the implications of M&A news announcements on share prices. Eckbo (2013) shows that merger announcements typically involve a large premium over existing price of the acquired company (between 30%-40% on average), and lead to a large and rapid change in market prices, suggesting that the announcement is news to the market. Routledge et al. (2013) uses a large sample and a regularized logistics regression model to predict merger targets and acquirers' performance from MD&As. Yet few studies looked into the implications of M&A on long run performance.

Table 3. Key Tokens and Associated Bigrams identifying phrases in the "Firm actions" segment of the dictionaries.

Unigram Token and Associated Bigrams	**Dictionary Token Frequency**
Partner (Trading Partnership)	4932
Merge (Acquisition Merger)	56
Consolidate (Subsidiaries Consolidated)	601

Investment (Value Investment)	8164

(* Note that irrelevant phrases containing the unigram will be removed from the dictionary, for example, "consolidate statement" may refer to the "consolidation of statement", which is irrelevant)

Category IV: Results/Inference from firm actions:

Some actions/strategic elements of firms may manifest implicitly, such as fees/costs revenue originating from firms' actions (eg. distribution fees, margin ratios). Tokens in this category are proxies for specific changes made to improve components in firms' value chain.

There may be overlaps between tokens that fall under this category and tokens in the performance dictionary: "advertising budget" is an example that is both internal and falls under our positive performance dictionary.

Table 4. Key Tokens and Associated Bigrams identifying phrases in the "Inference from firm actions" segment of the dictionaries.

Unigram Token and Associated Bigrams	Dictionary Token Frequency
Repayment (eg. Service Repayment)	1271
Collaboration (eg. Collaboration Agreement)	32
Operate (eg. Suspended Operations)	2951
Expand (eg. Successful Implementation)	369

5. 12.2 The External Dictionaries

Phrases in our **external dictionary** are defined to fall under any one or more of the following categories.

Category I: Porters' 5 Forces:

Analysis of the competition faced by the business, such as competitive rivalry, supplier power, buyer power, threat of substitution and threat of new entry, etc.

Key Unigram Token and Associated Bigrams	Dictionary Token Frequency
Supplier (eg. Supplier Interruptions)	3587
Competition (eg. Aggressive Competition)	2105
Demand (eg. Customer Demand)	2409
Aggressive (eg. Aggressive Advertising)	39

(* Note that irrelevant phrases containing the unigram will be removed from the dictionary, for example, "aggressive" may refer to actions of own firm, which is irrelevant)

Category II: Institutional or Regulatory Factors and Economic factors:

Geopolitical tensions in recent elections, governmental legislations, lawsuits, exchange rates, taxes, risk, foreign currency, fluctuations, interest rates, foreign currency, forward, option, forward position, option position, other shocks such as natural disasters, adverse weather conditions, cyber security threats, terrorism, etc.

Key Unigram Token and Associated Bigrams	Dictionary Token Frequency
Fiscal (eg. Fiscal Debt)	79
Treasury (eg. Treasury Yields)	361
Legislations (eg. Abilities Legislations)	727
Catastrophe (eg. Weather Catastrophe)	27

6 Search Algorithm Design

Similar to Loughran and McDonald (2011), the dictionary is used to compute a count. Each sentence in the corpus will be classified according to whether it contains a positive/negative performance token and an internal/external token. I perform the linear scan by iterating through each word in the sentence and for a match in the both types of dictionaries. Since the sample size is large, there is a need to reduce the time complexity of the search algorithm, and the final algorithm I designed works in O(N).

For example, in the sentence "Firm X experienced an increase in sales revenue due to the economic boom in China last week" , a score of 4 will be added to the "*positive-Institutional or Regulatory Factor*s" category. This is because the phrase "*increase in sales revenue*" has a length of 4, and "*economic boom"* is an external economic event.

If the sentence containing a performance phrase group does not contain an internal/external phrase, then the sentence is classified as neutral. For example, the sentence "Firm X experienced an economic boom in China last week" adds 1 to the "*neutral-Institutional or Regulatory Factor*s*"* category.

At last, the scores in each category, such as "*neutral-Institutional or Regulatory Factor*s*"* and "*positive-decision making*" is recorded by summing the scores across all sentences and dividing by the total length of the document.

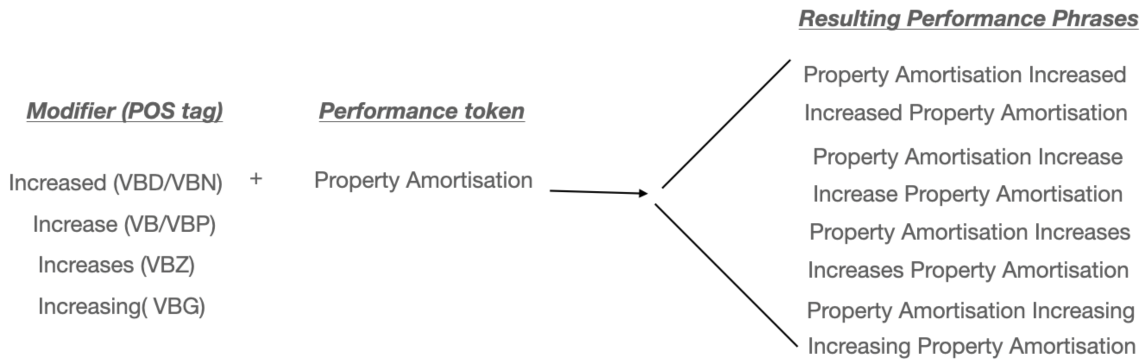
<u>Search Algorithm</u>

As I am required to take into account the length of the performance phrase group in the sentence, I designed a greedy search algorithm to do so.

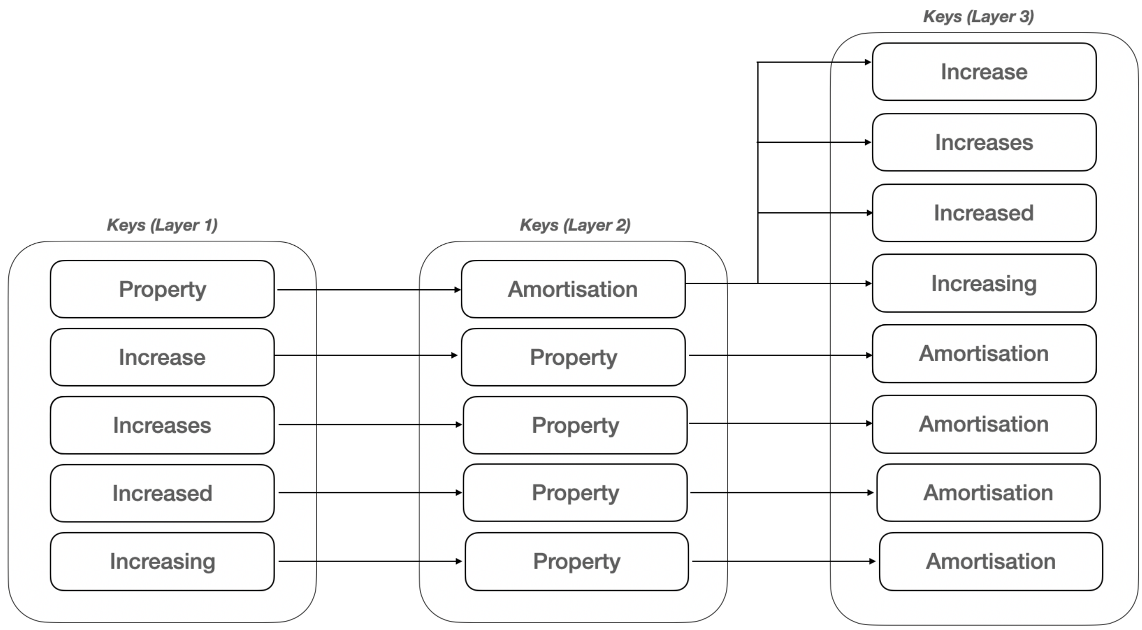
A positive/negative performance token is formed by:

1. An amplifier/negator/good/bad (eg. increase)
2. A unigram or a bigram in either of the pos/neg category in our performance dictionaries (eg. property amortisation)

A resulting performance phrase is constructed from permuting a positive/negative performance token and a performance token in any order. All possible inflection of the words in the phrase will be accounted for (eg. "increase" becomes "increases" or "increased"). For example, consider the resulting performance phrase from variations of the forms of the word "increase" and the performance token "property amortisation".



Consider the arrangements of these permutations in the form of a nested hashmap (or alternatively, a trie tree with a maximum height of 3) : which enables us to compare word by word.



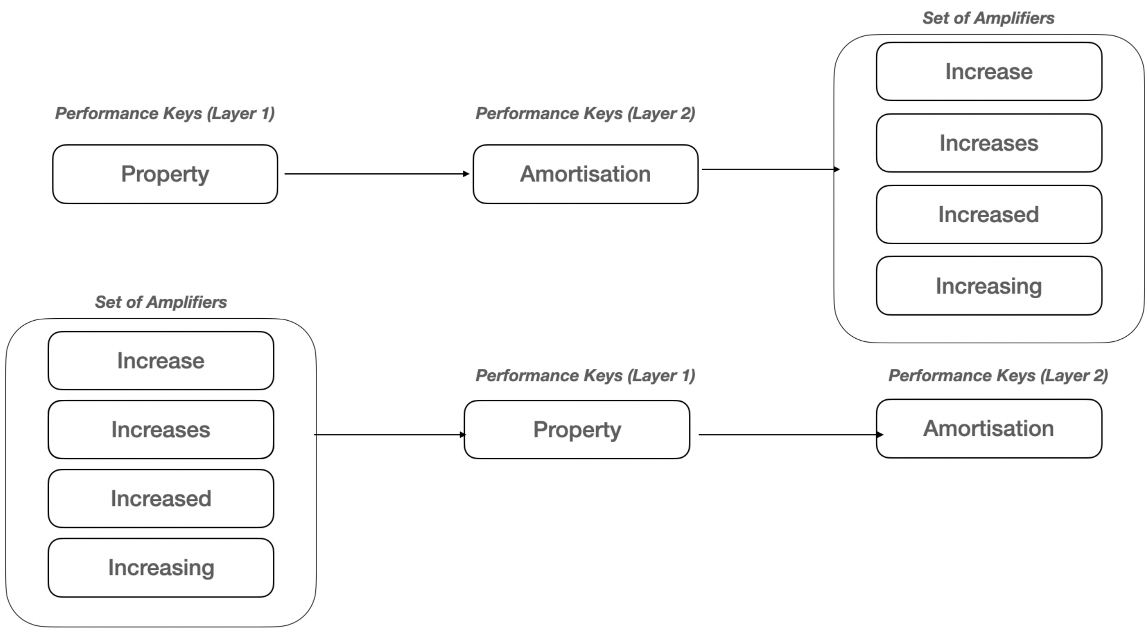
The word identification process works in the following manner. We first tokenize the document by sentence, and convert the tokenized sentence into a type list. We loop through the token entries in our sentence, and if the token appears in the keys (layer 1), we check if the restricted window of the next 1 to 3 words contains a word in keys (layer 2). If so, we check if the restricted window of the next 1 to 3 words contains a word in keys (layer 3). If all layers are matched or there is a blank string in the final key layer, we terminate the search and return the full length of the words list matched.

The following texts will be identified by our algorithm

Example matched phrase	Word Count produced
As a result, property amortisation increased	3
There is an increase in property amortisation	4
An increase in our property amortisation	5

Optimisation

To reduce the space complexity of our algorithm, we instead store amplifiers, negators and bads as 3 separate sets, and tokens under our performance phrase list in a hashmap. Key phrases we are trying to match all take the form “amplifier/negator” + “performance token” or “performance token”+ “amplifier/negator”.



Let the length of the sentence be w. We initiate an index at position 0 and increment the index procedurally until w-1 is reached. For each index we check whether it is in the set of amplifiers/negators/bads or in the layer 1 of the positive or negative performance dictionaries. Then we proceed to examine whether any tokens in the next nested layer of the dictionary appears within the subsequent 1 to 3 words in the sentence, until we finish checking all 3 layers.

In the case of computing the performance word count, if we were to identify the whole phrase, we move to the next word.

Our internal/external dictionary consists of mainly bigrams and unigrams, and whilst iterating through every word in the sentence, lookup is conducted with reference to the internal/external dictionary as well.

The pseudocode below is modified and implemented for each category in our amplifier/negator/bad, pos/neg performance and int/ext dictionaries (For each word, we conduct lookup in a maximum of 7 dictionaries).

Algorithm 1 Pseudocode for polarity prediction

Input: inputText, financial text sentence in type list
Dictionaries
Output: Document count

```
1.  increment_layer_1=0
2.  increment_layer_ 2=0
3.  index=0
4.  WHILE increment_layer_1+increment_layer_ 2<length of inputText:
5.      FOR increment_layer_1 in range(1,4):
6.        FOR increment_layer_ 2 in range(1,4):
7.          IF increment+index<length of inputText:
8.            IF word at index position in inputText is in dictionaries:
9.                IF word at (index+increment_layer_1) position in inputText is in the next nested
    layer of dictionaries:
10.               IF word at (index+increment_layer_1+increment_layer_ 2) position in inputText
    is in the next nested layer of dictionaries:
11.                   DO accumulate the score for each class in result,  conditionally terminate
    search
12.               ELIF the next nested layer of dictionaries is empty:
13.                   DO accumulate the score for each class in result, conditionally terminate
    search
14.             ELIF the next nested layer of dictionaries is empty:
15.                 DO line 11-12, conditionally terminate search
```

7 Empirical Results

7.1 Econometric model

My overall econometric specification estimates the overall effect of disclosure measures on financial performance, which I proxy with ROA.

The empirical model may be formulated in the following form:

$$\triangle ROAit = \alpha + \eta\boldsymbol{\beta}_{p\, for\, p\in\{1,2,3,4\}} neutralint_{pit\, for\, p\in\{1,2,3,4\}} + \lambda\boldsymbol{\beta}_{m\, for\, m\in\{1,2\}} neutralext_{mit\, for\, n\in\{1,2\}}$$
$$+ \eta n_{p\, for\, p\in\{1,2,3,4\}} negintattri_{pit\, for\, p\in\{1,2,3,4\}} + \lambda n_{m\, for\, m\in\{1,2\}} negextattri_{mit\, for\, n\in\{1,2\}}$$
$$+ \eta p_{p\, for\, p\in\{1,2,3,4\}} posintattri_{pit\, for\, p\in\{1,2,3,4\}} + \lambda p_{m\, for\, m\in\{1,2\}} posextattri_{mit\, for\, n\in\{1,2\}}$$
$$+ \delta_1 ROA_{it-1} + \delta_2 log(numseg)_{it} + \delta_3 length_{it} + \delta_4 lev_{it} + \delta_5 mcap_{it} + u_{it}$$

Where change in ROA is calculated as:

$$\frac{ROAt+5-ROAt}{ROAt}$$

A time lag of 5 years is taken to capture the full effect of changes made external or internal to the firm on the business. In the empirical model, int, ext, $intattri$ and $extattri$ are independent variables that describe the characteristics of the textual disclosures and the rest are control variables that describe the characteristics of the firm.

The full sample used is from 1993-2018, consisting of 88,154 10-Ks MD&A documents.

Variable	*Variable Description*
neutralint	Metric accounts for texts that convey strategic content classified as "internal" in fig. 7, without any attribution to performance. It intends to measure the quantity of discussion, and not the favorability. *Example sentence:* *"We pursue a strategy of supplementing internal growth by acquiring other financial companies or their assets and liabilities."* Sentence falls under the "firm action" category, and illustrates the means by which the firm seeks to achieve its strategy.
neutralext	Metric accounts for texts that convey strategic content classified as "external" in fig. 7, without any attribution to performance. It intends to measure the quantity of discussion, and not the favorability. *Example sentence:* *"Under the authority of eesa, treasury instituted the tarp capital purchase program to encourage u.s. financial institutions to build capital to increase the flow of financing to u.s. businesses and consumers and to support the u.s. economy. "* Sentence falls under the "economic, institutional reasons" category, as it is an example of how the firm reacts to external forces in its environment.
negintattri and posintattri	Metric accounts for texts that convey strategic content classified as "internal" in fig. 7, with some to performance. It intends to measure the polarity/favorability of the discussion (whether a positive, or negative performance outcome is attributed to strategic discussion). *Example sentence:* *"The decrease in gross profit as a percentage of sales for 2014 as compared with 2013 and for 2013 as compared with 2012 was primarily due to increases in promotional activity and product cost increases, some of which were not passed on to customers."* Sentence falls under the "firm action" category and attributes negative performance "a decrease in gross profit" to firm action, hence it would increment negintattri by 5/(documentlength).
negextattri and posextattri	Metric accounts for texts that convey strategic content classified as "external" in fig. 7, with some to performance. It intends to measure the polarity/favorability of the discussion (whether a positive, or negative performance outcome is attributed to strategic discussion). *Example sentence in sample:* *"These loans may be adversely affected by conditions in real estate markets or in the economy in general."* Sentence falls under the "economic, institutional reasons" category, and attributes worse performance as a result of "loans may be adversely" to the economic conditions.
ROA	Return on assets (Net Income/Revenue) x (Revenue/Average Total Assets) is used as a common indicator of firm performance, commonly used as a proxy for for evaluating firm performance (Edward et al., 1976; Aerts, 2001)

log(numseg)	Number of business segments, used as a proxy for the size of the firm. Empirical evidence suggests different evidence on how firm size affects growth in firms' financial performance: Gibrat's Law states that the expected increase in firm growth is proportional to its size, Bentzen et al. (2011) finds that firm's growth rates are more likely to be positively related to firm size.
lengthit	Disclosure length of the MD&A section, controlled for as it may influence the proportion of internal/external discussion.
levit	Leverage (total debt to total asset ratio) is indicative of a firm's ability to meet its financial obligations. The majority of conducted empirical studies find a negative relationship between company returns and leverage. Baker (1973) examined the effects of financial leverage on industry profitability and concluded that firms who earned systematically higher returns had a relatively low degree of leverage. There are numerous different theories on the optimisation of firms' capital structure. A well known theory is the pecking-order theory (Myers & Majluf, 1984), which states that firms prefer internal financing to fund their operations. The trade-off theory, in contrast to the pecking-order theory, suggests that firms can reach an optimal level of leverage, in which the benefits of tax shields are directly offset by costs from financing distress (Kraus & Litzenberger, 1973; Myers, 1984). The theories mentioned also imply that certain relationships between leverage and profitability are expected, endorsing a non-zero coefficient on leverage.
mcap	Market capitalization is an important market indicator of the value of shares and the value of companies in general (Toramane et al., 2009; Dias 2013). Empirically, Donaldson (2015) finds a significant positive correlation between firms' ROA and market capitalisation.

Based on the nature of the variables included in the regression, and from my previous hypothesis, I posit the following hypotheses.

1. Attributing negative performance to internal reasons likely correlates with improved future performance
2. A display of more strategic awareness focusing on the internal of the organization likely correlates with with improved future performance
3. A display of more (positioning) strategic awareness likely correlates with improved future performance
4. Attributing negative performance to external reasons likely correlates with improved future performance

5. Higher leverage likely affects future performance in a neutral or negative way.
6. By empirical evidence, market capitalization is positively correlated with ROA.

7.2 Baseline Results

I first apply the regression model specified in part 7.1 in a simplified form, where the regressors $negintattri$ and $posintattri$ into one regressor and $negextattri$ and $posextattri$ into one regressor.

Standard errors are clustered at the GIC industry level to allow for correlation of errors within each group (Hansen, 2017). The results are shown below in the table below. Note that the regressors "Decision Making, Firm Competencies, etc" would be regressors accounting for proportion of unattributed internal/external discussion. The regressors such as "Attribution Decision Making" and "Attribution Firm Competencies" account additionally for polarity, a higher metric means that the firm attributes more positive performance to decision making.

Table . Future ΔROA. This table shows the results of OLS regressions of Future ΔROA (dependent variable) on disclosure and firm characteristics. The regressions use industry (Ken French's 48 industry classification) and fiscal year fixed effects. Errors are clustered by industry. Variables are defined in Table 2, and Section 3 of the text. t -statistics are shown in brackets. *, **, and * denote significance at the 1%, 5% and 10% levels respectively.**

Regression results: **** stands for 1 percent significance level, ** stands for 5 percents significance level and * stands for 10 percent significance level (t-statistics in brackets)*

	(1)
Decision Making	0.3422 **
	(0.1722)
Firm Competencies	0.2319
	(0.3052)
Firm Actions	0.0544
	(0.5956)
Results from Actions	0.7272 *
	(0.4276)
Porter's Five Forces	-0.3204
	(0.9007)
Institutional/Economic	-0.1600

	(0.8875)
Attribution Decision Making	-0.5251
	(0.6783)
Attribution Firm Competencies	-0.4132
	(2.279)
Attribution Firm Actions	3.293
	(6.866)
Attribution Results from Actions	-2.894
	(6.497)
Attribution Porter's Five Forces	-19.75 *
	(11.31)
Attribution Institutional/Economic	26.23 **
	(10.85)
$log\ (sentcount)$	-3.788e-08
	(4.950e-07)
ROA_{t-1}	5.532e-05
	(1.269e-03)
$log\ (marketcap)$	-0.003198 *
	(0.001866)
leverage	-0.031773 ***
	(0.01248)
log(Segments)	9.0291e-03
	(4.8910e-03)
N	39670
Multiple R-squared:	0.01809
Adjusted R-squared:	0.01386
Fixed Effect	Year and industry
Error clustering	Industry

First, I find that disclosure on decision making and results from actions to be positively correlated with $\triangle ROA$. One basis point increase in decision making disclosure (no. of decision making associated vocabulary divided by total length of disclosure) is correlated with 0.34 basis points higher $\triangle ROA$. Correspondingly, one basis point increase in results from actions (no. of firm action related vocabulary divided by total length of disclosure) is correlated with 0.72 basis points higher. This coincides with our hypothesis that strategic awareness and its communication has a strong degree of influence on firm's financial performance. It is also reasonable that the coefficient associated with results from actions is more significant than decision making or actions. Given that firms can selectively disclose information and are rewarded for the amount of information that leads to tangible improvements, firms would prefer to disclose more positive results from actions than the other categories, as it embodies more certainty. In contrast, decision making and firm actions embody more risk than tangible outcomes, and yield less expected growth in performance. To summarize, results show that firms may have a tendency to disclose information that they have more certainty will lead to a positive performance improvement.

Second, I observe that the coefficients on Porter's Five forces and Institutional/Economic factors are correlated with performance change into the future. One basis point increase in net positive performance attributed to Porter's Five forces is correlated with 19.75 basis points of lower $\triangle ROA$. A reasonable interpretation of this result is that Porters' Five forces are more focused on identifying threats to performance. Firms that are daring to attribute negative performance to Porter's Five Forces are willing to reflect upon their vulnerabilities and seek ways to circumvent them. It is interesting to note that the corresponding coefficient on Porter's Five Forces without attribution is not statistically significant. This tells us that reflecting on strategies is insufficient to yield a positive performance outcome: successful firms not only discuss Porter's Five Forces on a strategic level, but they also dare to pinpoint operational weaknesses (leverage, capital structure, profitability, etc) that relate to these forces of competition. Perhaps they are successful because they recognise and act on these weaknesses while less successful firms do not.

Our result also shows an interesting dichotomy: whilst both Porter's Five Forces and Institutional/Economic are categories that relate to the external environment of the firm, their relationship with future $\triangle ROA$ are polar opposites. One basis point increase in net positive performance attributed to Institutional/Economic discussion is correlated with 26.23 basis points of higher $\triangle ROA$. Compared with discussion on Porter's Five Forces, which is proactive, discussion on Institutional/Economic forces is reactive. Firms are more likely to cast the blame of poor performance on Institutional/Economic forces, and in doing so conceal internal weaknesses (eg. blaming poor revenue on a lack of economic demand instead of management). By contrast, firms attributing poor performance to Porter's Five Forces are conscious of where their internal weaknesses are in relation to the industrial landscape.

However, the observations did not coincide with our hypotheses for “firm competencies” and “firm actions”, as the coefficients for each variable, both with and without attribution, are not statistically significant. This could be due to our dictionaries insufficiently proxying for textual measures, misspecifications in the statistical model, or the fact that small changes in performance resulting from small changes in textual attributes may be too hard to capture statistically . Given the first and second cases, there is room for improving the model. First, there is the possibility that counts produced from our dictionary method introduce multicollinearity. For instance firms in the medical or pharmaceutical industry would disproportionately use “patent”, and firms in the technology industry would disproportionately use “develop”, violating assumptions of independence of x variables.

To test for robustness, I calculate variance inflation factors (VIF) that detects multicollinearity in regressions. Mathematically, the VIF of a regression model variable is equivalent to the ratio of the overall model variance to the variance of a model that includes only the single independent variable. A high VIF indicates that the associated independent variable is highly collinear. However, as can be observed from the table below, all VIF quantities fall within the desired range (<5), invalidating the concerns for multicollinearity.

Categories	*GVIF*
as.double(cat1neutral)	*1.351274*
as.double(cat2neutral)	*1.234637*
as.double(cat3neutral)	*1.376913*
as.double(cat4neutral)	*1.228152*
as.double(cat5neutral)	*1.178656*
as.double(cat6neutral)	*1.232957*
as.double(cat1p - cat1n)	*1.741510*
as.double(cat2p - cat2n)	*2.448715*
as.double(cat3p - cat3n)	*1.741776*
as.double(cat4p - cat4n)	*1.564963*

as.double(cat5p - cat5n)	*1.380680*
as.double(cat6p - cat6n)	*1.476128*
log(as.double(marketval)	*1.268695*
as.double(lev)	*1.297984*
as.double(`roat-1`)	*1.012223*
gic	*3.404887*
year	*1.745910*

Second, I suspect that the relationship between $\triangle ROA$ and the textual measures are more intricate than one regression can capture. It may be highly likely that the coefficients on the textual variables differ with respect to year. This leads me to conduct regression on an annual basis, collect and plot the coefficients.

7.3 Time Series Regression

In total, I run 21 separate regressions (one each during 1994-2014 because variable change in ROA cannot be computed from 2015 onwards as requires a 5 years lag) using the exact specifications described in section 7.1. Below plot the coefficients over the years. I decided to split the 6 attribution categories into neutral, positive and negative attribution (each plotted on a different diagram), as I believe that they have different degrees of impact on $\triangle ROA$, because a well performing firm would tend to describe attribution differently from a poorly performing firm.

Scattered dots with **dark borders** and **coloured centers** are statistically significant. Dots with no asterisk are significant at the 5% level, dots with * are significant at the 1% level, dots with ** are significant at the 0.1% level, and dots with *** are significant at the <0.01% level.

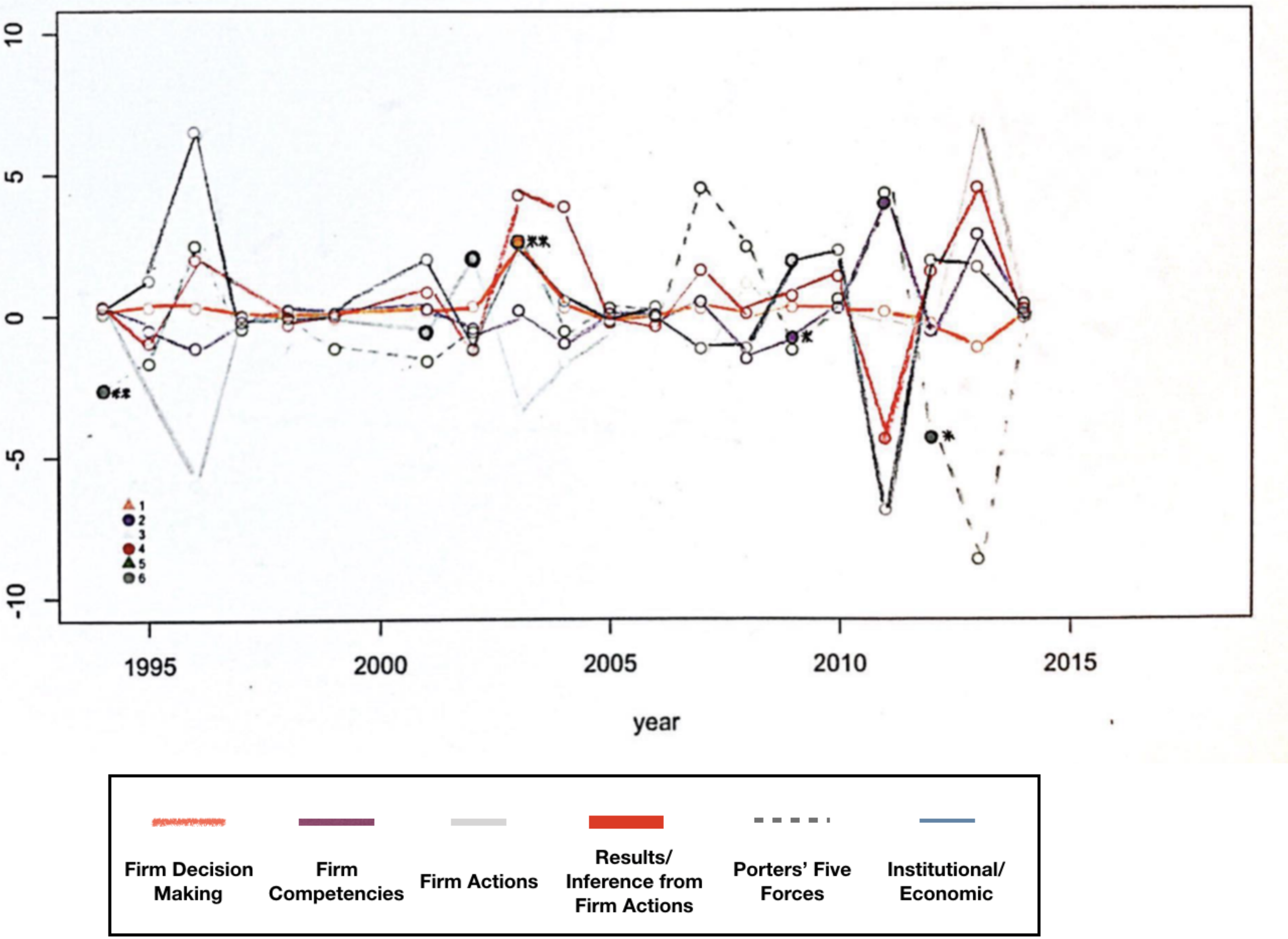


Fig 16. Neutral

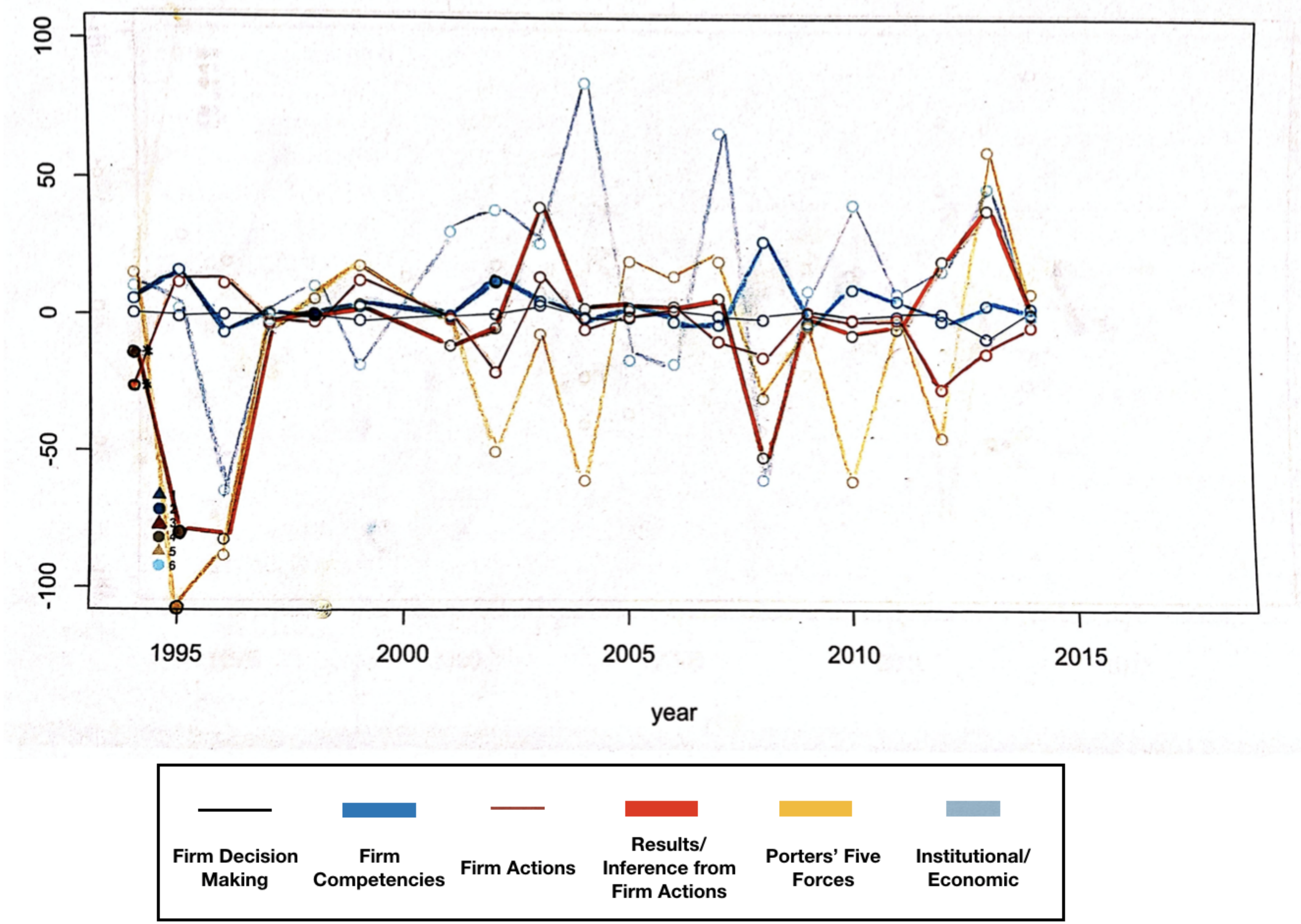


Fig 17. Positive Attribution

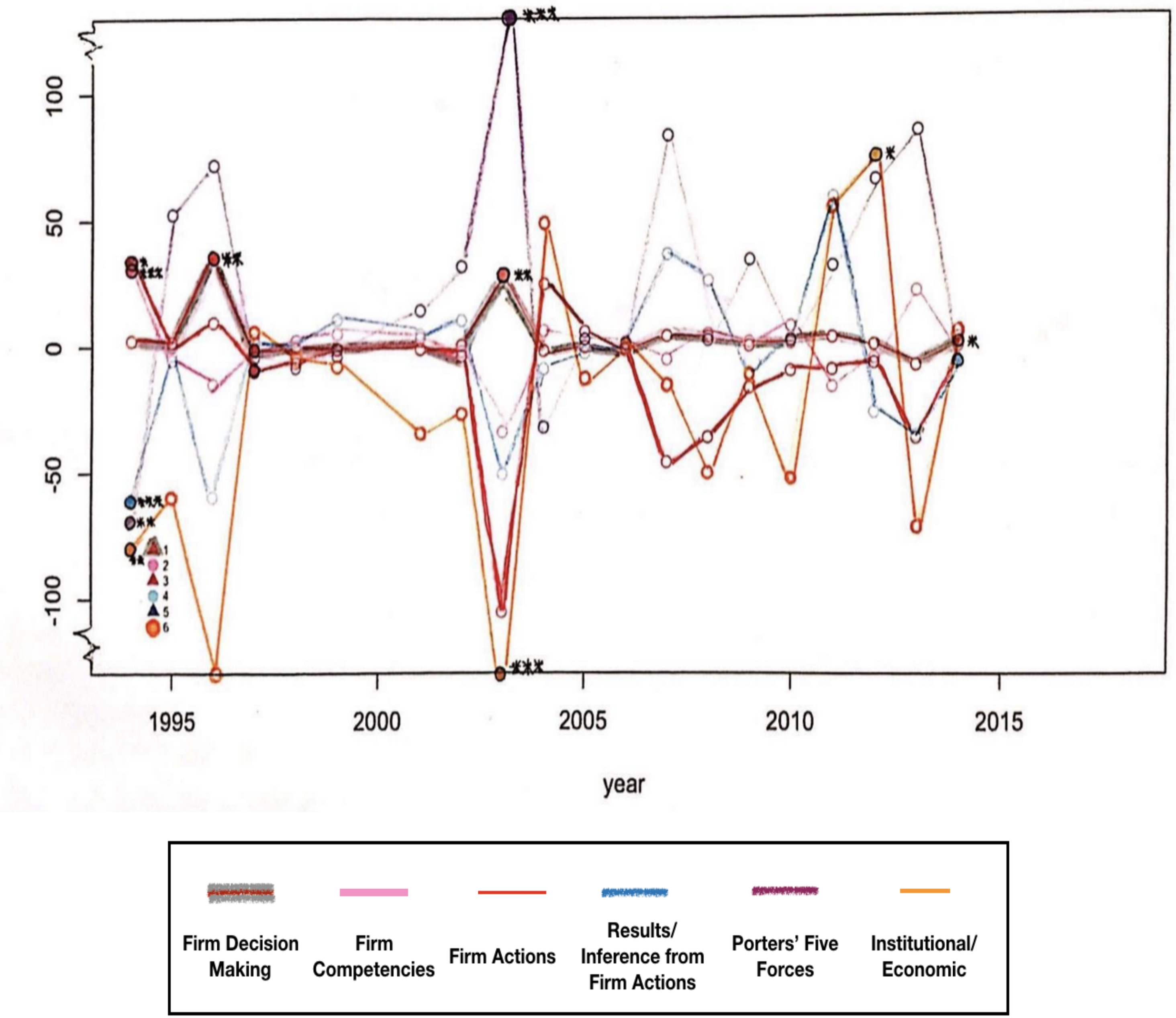


Fig 18. Negative Attribution

It is indeed the case that coefficients differ wildly across years. Following these sets of regressions, we are able to make more interesting conclusions.

First, examining Figure.18, I find 2003 to be of special significance, as multiple coefficients have high magnitudes and are statistically significant at the <0.01% level. Because I chose to model the change in ROA as the percentage change in ROA from the present years to five years into the future, the coefficient in 2003 reflects the percentage change in ROA from 2003 to 2008. My findings are as follows:

1. Companies which attributed negative performance to ***Porter's Five Forces*** (the competitive landscape) in 2003 experienced **<u>a stark increase</u>** in ROA from 2003 to 2008, significant at the <0.01% level.
2. Companies that attributed negative performance to ***institutional/economic*** forces in 2003 experienced **<u>a stark decrease</u>** in ROA from 2003 to 2008, significant at the <0.01% level.
3. Companies that attributed negative performance to ***firm actions*** in 2003 experienced **<u>an increase</u>** in ROA from 2003 to 2008, significant at the 0.1% level.

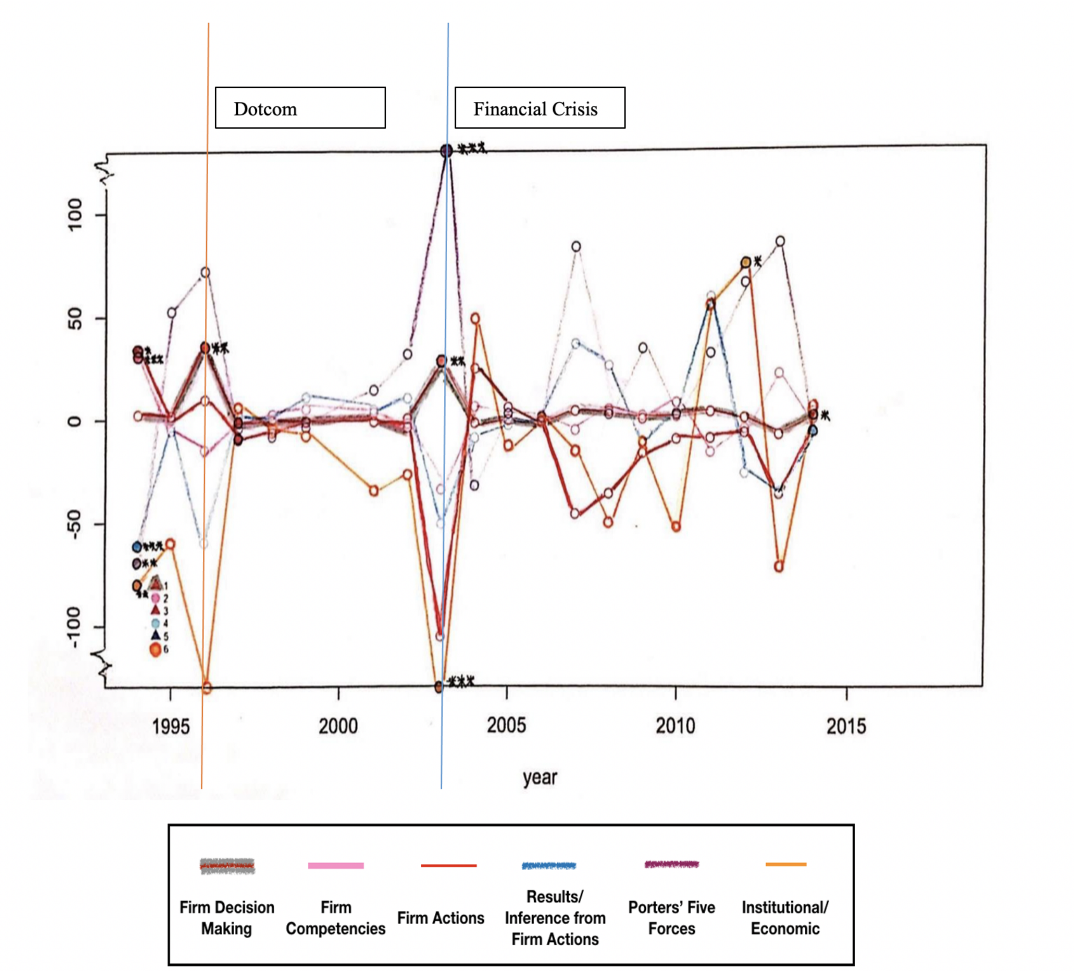
The coefficients on both "Porter's Five Forces" and "institutional/economic" are extremely statistically significant and are scaled up 10 fold when compared with previous years. It is particularly interesting that the statistical relationship between attribution and performance is only pronounced in the event of the financial crisis. A plausible explanation of the findings is that firms attributing negative performance to Porter's Five Forces demonstrate active management and awareness of their competitive landscape in advance of economic crisis, and thus are more effective at crisis management than their peers. Contrary to this, firms that used to attribute negative performance to external economic or institutional factors were poor at preempting crisis, hence experiencing inferior results. This coincides with Salancik and Meindl's (1984) findings on the future performance of firms which have low locus of control in their disclosures attribute performance to factors beyond the control of management.

Likewise, the coefficient on decision making is also significant and positive. Firms that attribute negative performance to their concrete actions show a strong degree of responsibility taking, and have the audacity to take risks (when faced with blame from stakeholders). Compared to attributing negative performance to "competencies", "actions" and "results from actions", management assumes more responsibility in attributing poor performance to decision making (because decision making comes directly from the managers themselves). The ability of management to trace the root cause to decision making also demonstrates more thorough problem analysis. This coincides with Salancik and Meindl(1984)'s finding on the relationship between impression management and future performance.

To summarize briefly the results from the discussion above, I arrived at three insights:

1. Unlike past literature on attribution theory, which shows that attributing performance to internal/external causes is correlated with good/bad future performance, I find that the predictability of attributing negative performance is most pronounced before systemic macroeconomic events. The attribution of negative performance may be used as a proxy for firms' crisis management skills. Alternatively it may just be the case that firms which recognise poor results and thus seek to act on them will be more likely to reverse poor performance before it gets worse, leading them out of crises or meaning that crises are prevented because poor performance is corrected before it gets worse.
2. A history of disclosing root cause analysis or competitive positioning analysis has a positive impact on performance outcome. Management taking responsibility also has a positive impact on performance outcome (managers that are more daring to disclose bad decision making are associated with improved future firm performance). Perhaps due to the fact that the recognize that prior decision making has been suboptimal.
3. Firms that cast blame of poor performance on external economic factors experience poorer returns at times of economic crisis.

Again, we can see other evidence that supports our insights from above. Consider 1996 in Figure.18. The coefficients correspond to the % change in ROA from 1996 to 2001 (capturing the duration of the Dot-Com-Bubble). It is clear that the insights deduced from the financial crisis holds true for the burst of the internet bubble in 2001. In 1996, the coefficients on Porter's Five Forces, Institutional, Economic Factors, as well as decision making are in the same direction as during the financial crisis, but since the magnitude of the stock market shock is smaller in 2001 compared to 2008, the coefficients are not statistically significant. If we compare the magnitude of the coefficients between the financial crisis (blue vertical line) and the Dot-Com-Bubble (orange vertical line) in a further labelled diagram of Figure 18 (upper diagram) with the magnitude of their shock to the economy and duration (lower diagram) in figure 19, we find close alignments.



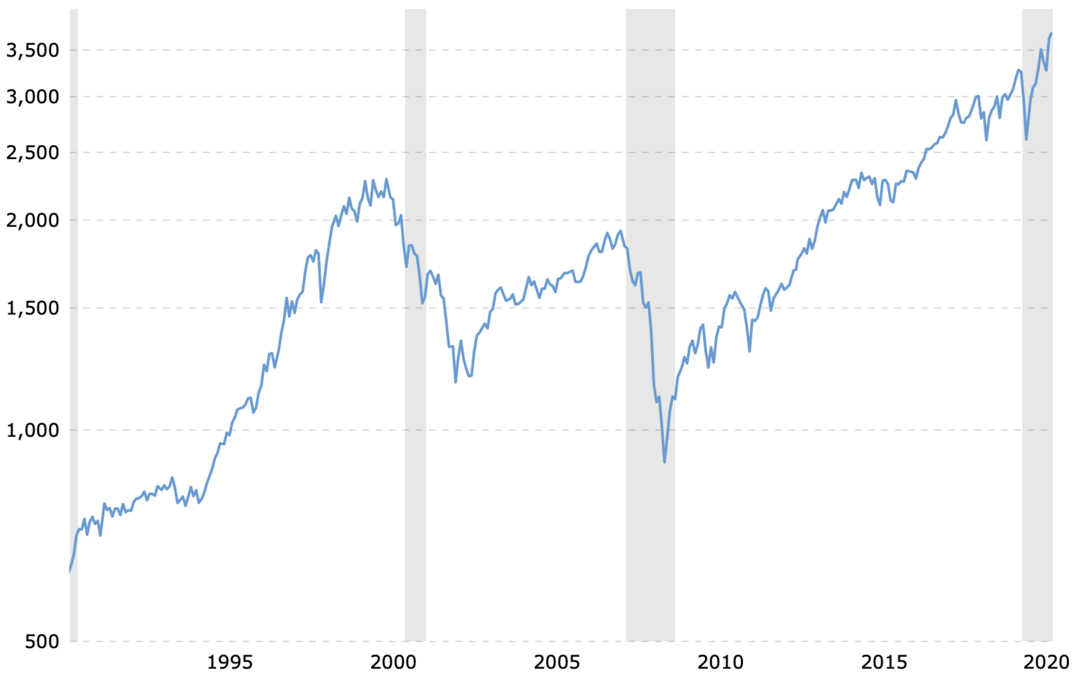


Figure 19. S&P index across time

Now, looking at Figure 17 on the attribution of positive performance, I arrive at the following findings:

1. Most of the significant coefficients on the internal categories are negative across the years. This aligns with most attribution theory literature: attributing positive performance to internal competencies is correlated with worse performance, associated with firms trying to over-glorify their results.
2. However, the relationship between positive performance attribution and future change in ROA is not as consistent as with negative performance attribution,
    a. First, the coefficient on attributing positive performance to ***results from actions*** is negative in 1994 and 1995, at the 5% and 1% levels respectively. This means that firms attributing positive performance to results in 1994 and 1995 experience a **decline** in ROA from 1994-1999 and 1995-2000, respectively.
    b. Second, the coefficient on attributing positive performance to ***actions*** was negative in 1994, at the 1% level, hence, firms that attribute positive performance to results to actions experienced a **decline** in ROA from 1994-1999.
    c. The coefficient on attributing positive performance to ***Porter's Five Forces*** is negative in 1995, at the 5% level, firms that attribute positive performance to the external environment experienced a **stark decline** in ROA from 1995-2000.

It is evident that 2000-2001 sees the evolution and bursting of the Dot-Com-Bubble. However, surprisingly, in contrast to observations with respect to the Dot-Com-Bubble, no insightful correlation can be found between future performance and textual disclosure for the time frame relevant to the financial crisis. All of the coefficients in 2003 (which correspond to ROA change from 2003 to 2008), are not significant. Thus, it is possible that the relationship between means of attributing positive results and future returns depends more on the reason and nature of the recession, and less on the magnitude of the recession.

To arrive at possible reasons behind this observation, I would want to study the differing causes behind the Dot-Com-Bubble and the financial crisis. First, the Dot-Com-Bubble stemmed from the firms acting themselves, due to irrational decisions made by a category of firms. The financial crisis was caused by excessive risk taking by banks and the bursting of the housing bubble, which was discussed to a lesser extent before it happened. Second, the Dot-Com-Bubble was led by stakeholders-wide fads and decisions made within organizations (internal), whereas the financial crisis was led by information asymmetry and regulatory failure (external). Actions in the lead up to the Dot-Com-Bubble were publicly disclosed: many internet companies needed justification for their actions as they incurred immense net operating losses by aggressively spending on advertising and promotions. Greed and excessive optimism are openly revealed: from the significant coefficients on ***Porter's Five Forces, actions,*** and ***results to actions,*** we understand that attributing positive performance to the favorable industrial landscape and fads and fashions is negatively correlated with future performance.

However, as compared to the Dot-Com-Bubble, events in the build up to the financial crisis were less to do with the firms themselves than they were to do with the financial system. Firms' eventual bankruptcy and decline in performance were caused by the surge in real interest rates. As compared to the Dot-Com Bubble, the financial crisis is more related to the behaviour of financial institutions, which was not featured in comparable length as strategic disclosure amongst firms in the lead up to the Dot-Com-Bubble.

An additional point to note is that that all of the significant observations refer to performance time frames ending in 2000 at the latest, before the dot com bubble burst - they do not refer to the recession itself and the consequent negative performance it brought about - this likely explains why the same results are not replicated for observations in 2003 - the 5 year time frame after 2003 includes some of the 2007-10 recession while these 1994-2000 time frames do not. Hence, it can be reasoned that positive attribution may be useful to detect the evolution of fads and fashions in the lead up to a bubble, but do not correlate strongly with a recession itself.

A less explainable relationship can be found between neutral (unattributed discussion) and future performance. This shows that, again, strategic justifications should be evidenced to be able to lead to material impact on future performance.

Briefly summarizing the results obtained on attributions of positive performance and how this compares with results obtained for attributions of negative performance (it is important to note that correlation does not imply causation and the interpretations offered are speculative) :

1. Attribution of negative performance to decision making and the industrial landscape tells us about the firm's degree of responsibility taking and strategic reflection.This relates positively with future performance during economic recessions. However, attribution of positive performance to actions and industrial competition may be interpreted as excessive optimism and a "red flag" that is causal of a crisis.

2. Unlike what is suggested by all previous authors, it would be worth investigating attribution from an events study perspective. Analysis of the attribution of negative performance would aid researchers to predict how a firm might perform during a time of crisis, whereas the attribution of positive performance may be used to study the unfolding of bubbles and excessive optimism.

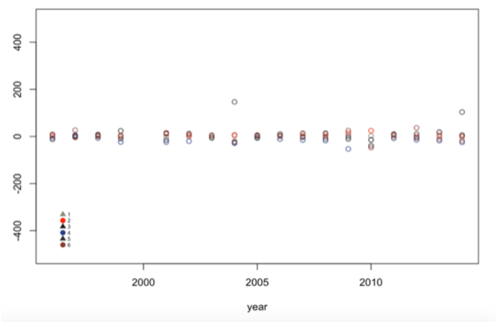
7.4 Combining LDA with Attribution

Subsequently, I want to make cross-sectional observations on how the correlation between measures of textual disclosure vary together with disclosure content. Hence, I have conducted separate regressions with respect to the LDA sectors specified in section 3.

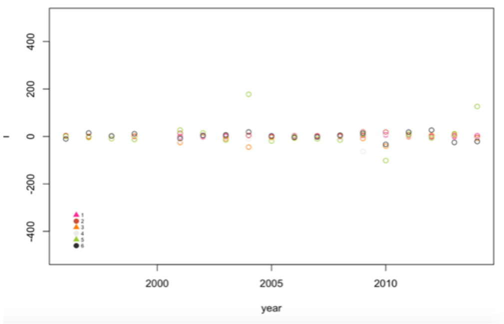
If one were to define information in disclosure that has tangible impact on performance as relevant, the econometric model is subject to underlying imperfections. This is primarily because identical information in documents may not need to have the same impact on performance, even controlling for observable firm characteristics. In previous procedures, we have controlled for firm characteristics, yet we did not account for disclosure characteristics. Disclosure characteristics proxy for firm priorities and industry positions that we cannot otherwise observe. For example, consider firms A and B of identical size, leverage and in the same industry, with firm A an innovator and firm B an incumbent. Firm A would disclose more information on R&D and product development whilst firm B would not. Thus, the textual measures would be proxies for different subjects, for instance, the resource based view variable would proxy for different organizational resources and perspectives (eg. superior technologies for firm A and superior management in firm B). Hence, it is highly likely that the disclosure content of Firm A and B would be correlated with performance in different ways. Additionally, as a result of the nature of their business, some firms tend to disclose more material and relevant information about their decision making.

To investigate this issue further, I conduct a set of regressions for every category for negative attribution, as the magnitudes of coefficients obtained in this category are largest out of the three categories examined. The diagrams below show results for categories relating to negative attributions. It is apparent that documents with content that is more exposed to financial market cyclicality have returns patterns that are further dispersed. For instance, clusters are more dispersed for real estate, products/services/customers, insurance, supply chain and corporate borrowing. The impact of textual disclosure on performance is more variational because there is less certainty to organizational actions.

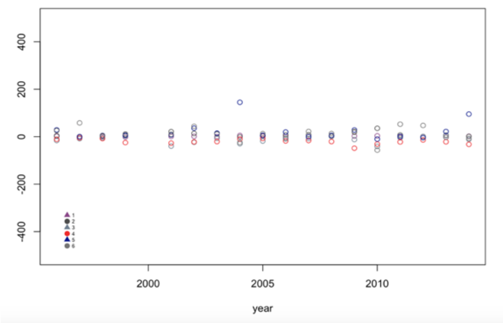
The categories where coefficients vary the least across time are securities and valuation. Both categories are less relevant to strategy. It would seem that as disclosure on any strategic content becomes more minimal, it would be harder to find a statistically significant relationship between them and performance.



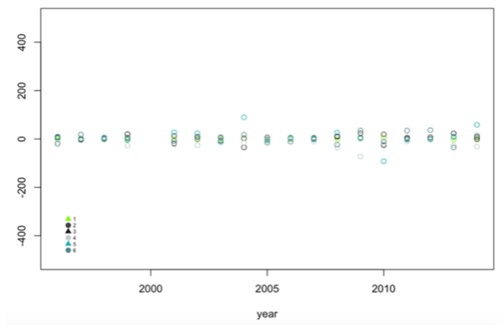
Oil and Gas



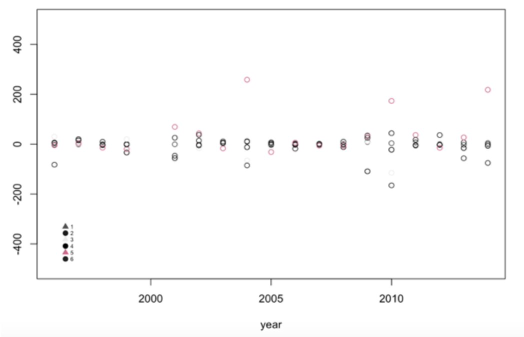
Derivatives; Trading



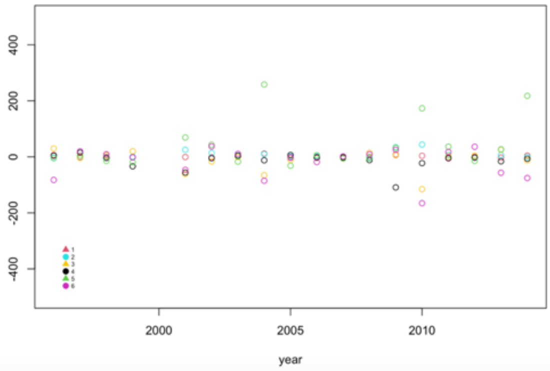
R&D



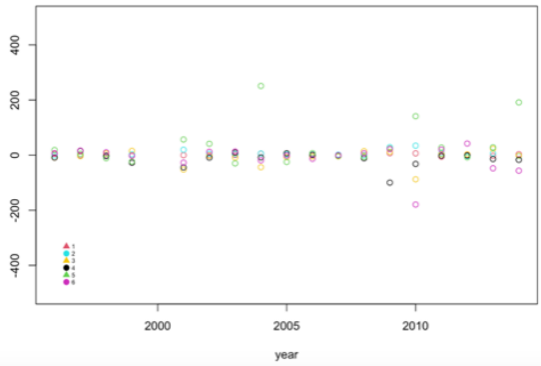
Retail, warehouse



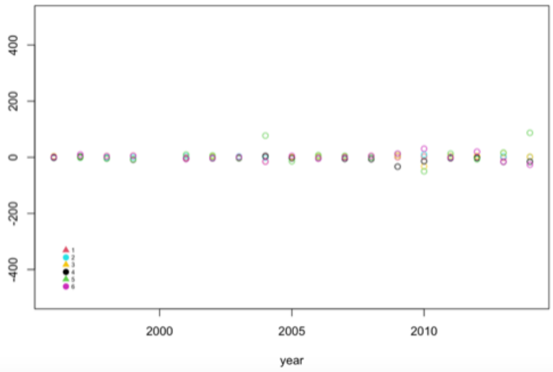
Real Estate



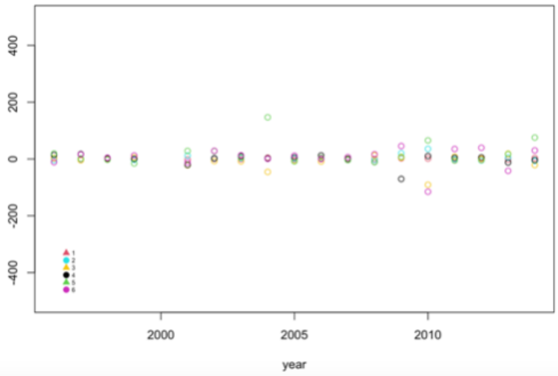
Operations



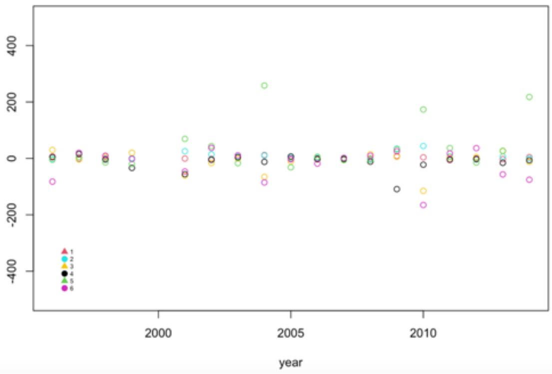
Management



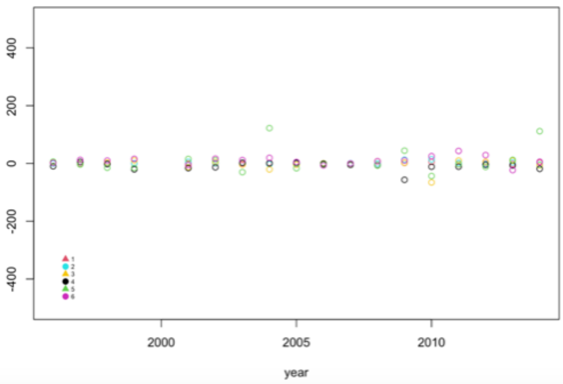
Valuation



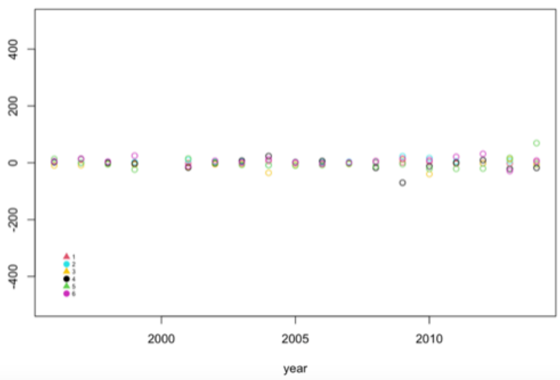
Decision Making



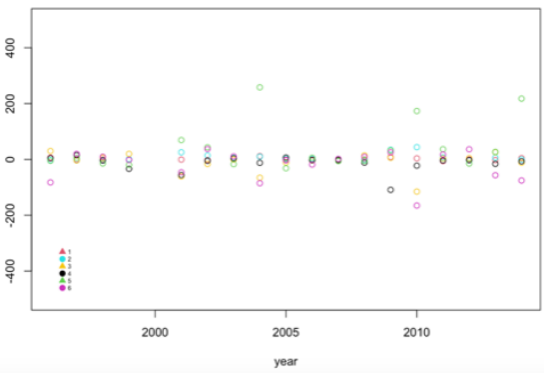
Products/Services/Customers



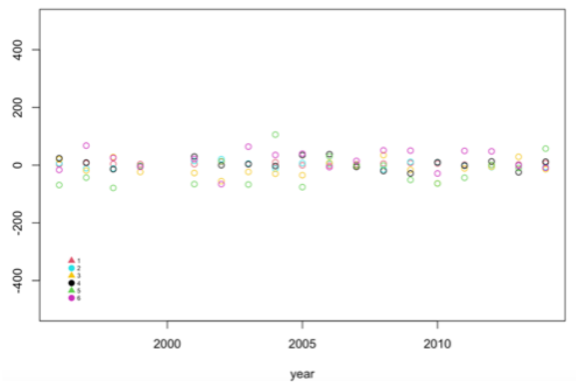
Environmental



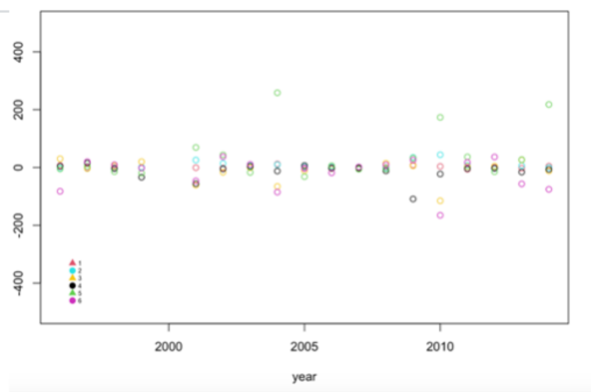
Securities



Supply Chain



Insurance



Corporate Borrowing

7.5 Examining the implication of newly emerging topics

An additional research agenda is to look into how corporate performance is affected by disclosures relating to the emergence of new topics as suggested in section 3.22. One would expect some trends to have a permanent impact on firms' performance (eg. global events),but others less so (eg. regional events). I select one arbitrary topic across each category of “global macroeconomic event”, “regional macroeconomic event” and “macroeconomic trend”. These are topic categories “mexico peso crisis” (global macroeconomic event), “decommissioning of the Yankee power plant” (regional macroeconomic event), and “dot-com bubble” (global macroeconomic trend). I account for all companies that have disclosed related topics in the associated years (those companies would fall into the respective clusters), and extract the ROA evolution of those companies in the coming years.

To enable direct comparison between firms' performance, I used time-series clustering to group the time series into ones that exhibit similar patterns of evolution, using the k-means algorithm. For each of the 3 data sets, I adopt the k-means algorithm to partition the time series into 3 clusters in which each observation belongs to the cluster with the nearest mean (minimizes within cluster variances). The metric used to measure the distance between time series is chosen to be the dynamic time warping (DTW) distance metric. Given series:

$$X = (X_0, X_1, ..., X_n) \quad \text{and} \quad Y = (Y_0, Y_1, ..., Y_n),$$

$$DTW(X,Y) = \min_{\pi} \sqrt{\sum_{(i,j)\in\pi} d(X_i, Y_j)^2},$$

where π is the path of index pairs of elements in each time series, the DTW metric is calculated as the square root of the sum of the squared distances between each element in X and its nearest point in Y. The DTW metric is more suitable than the standard Euclidean metric, as it is more sensitive to time shifts between series. Additionally, the DTW metric is better for comparing two series that may not be precisely aligned in time or length, as is the case with our datasets.

In figures 20, 21 and 22, I provide visualizations of my result that consists of a coloured figure with all ROA time series and results post clustering, with time series assigned to their most likely cluster.

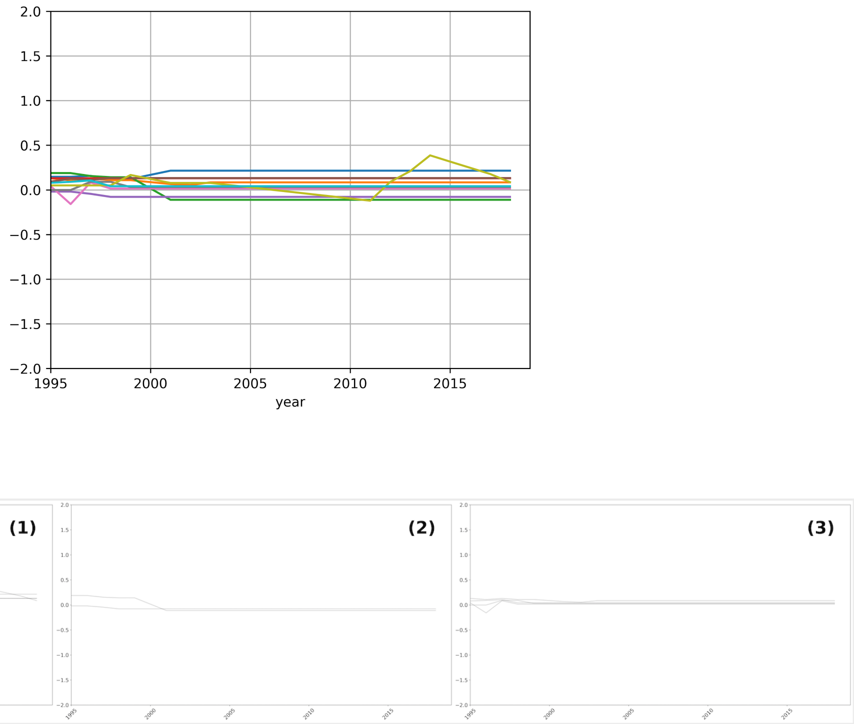


Figure 20 (Above). Mexico Pesos Crisis (Global Macroeconomic Event)

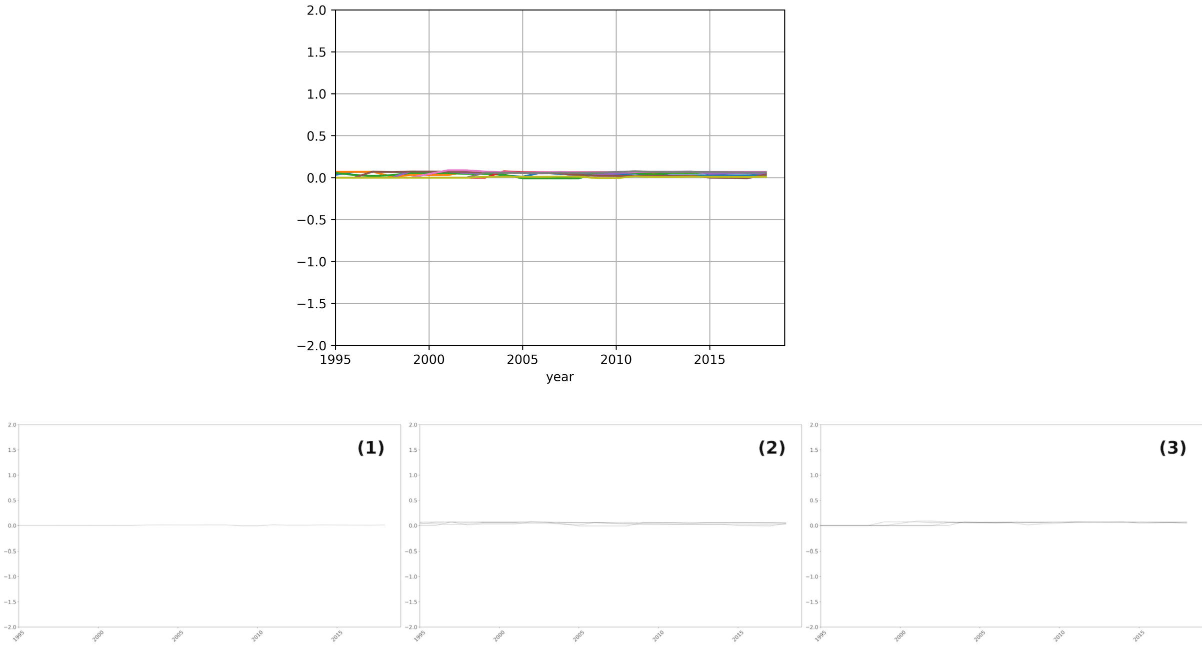


Figure 21 (Above). Yankee Power Station Abandonment (Regional Macroeconomic Event)

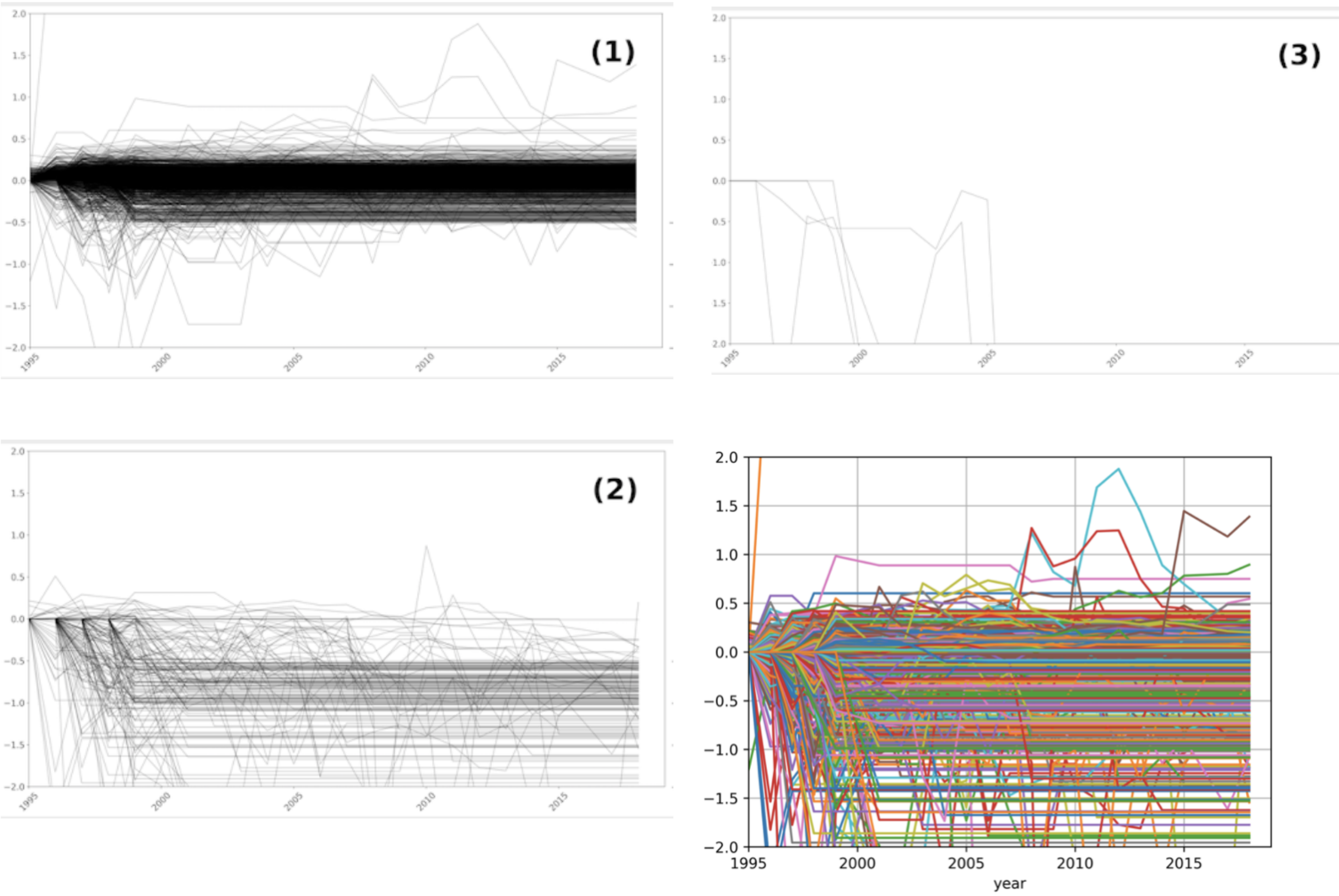


Figure 22 (Above). The Dot Com Bubble (Global Macroeconomic Trend)

First, it is observed that the firms that have disclosed content in relation to the Dot-Com-Bubble experienced more drastic ROA fluctuations than both firms that have disclosed content on the Mexico Pesos Crisis and Yankee Power Station abandonment. Compared to the other two cases, firms involved in Yankee Power Station abandonment experienced the least degree of ROA fluctuations (notice all graphs are plotted on the same scale). A potential explanation to this is that the evolutionary patterns of ROA are reflective of the severity and systemic impact of the economic events: that is, the more systemic and widely penetrating the subject disclosed, the more influence the subject will have on the performance of the firms.

Second, looking closely at Figure 22, it is apparent that firms that were related to the Dot-Com-Bubble can be roughly categorized into those experiencing (1) mild ROA fluctuations/slight increase, (2) moderate ROA decrease and (3) severe ROA decrease, with the exception of one firm (see orange colour line) which has its ROA surge up drastically. Amongst the firms that experienced a decline in ROA, the shock is rather persistent, and rarely do firms experience a positive ROA in years following experiencing negative ROA. The result raises a plausible possibility, that is, amongst firms that are affected by fads and fashions, most firms that blindly follow the trends fail. Viewed together with the revealing results section 7.2., I believe that it would be beneficial to particularly study the causal effect of strategy (using information provided in corporate disclosures as a proxy) on performance of the specific set of firms which were at the core of the Dot-Com-Bubble.

7.6 Testing the Causality of Attribution (A Case Study on the Dot-Com Bubble)

Previously, I had arrived at conjectures of using attribution disclosure to predict performance, by examining correlation. A logical step that follows the above analysis is to examine which (if any) of the manners of firm decision making have a causal effect on the Dot-Com-Bubble, using textual measures as a proxy.

I believe it will be uninformative to study the causal impact of attribution disclosure on the mean value of all ROA time series. It is likely that the causal effect of attribution disclosure may not be homogenous across all firms. First, firms may be inclined to disclose information differently when they experience different performance: poor performing firms may have a tendency to cast blame on the economy and successful firms to praise their own accomplishments. Second, if there truly exists variations as such, results obtained would be skewed to reflect the poor performing firms. As seen in Figure 22, there are very few firms that reaped benefits from the Dot-Com-Bubble and significantly more firms that suffered.

To provide a solution, I run the Causality test on three different time series representing the means of the three clusters obtained from the k-means analysis. As seen in Figure 23, taking the mean of the time series results in the following three clusters. The method I use to test for causality is the Granger Causality test. The Granger Causality test is used to determine whether one time series is useful in forecasting another. Regressions test correlations, but Granger (1969) argued that causality could be examined by measuring the ability to predict the future values of a time series using prior values of another time series - that is, the Granger Causality test measures predictive causality, or precedence causality.

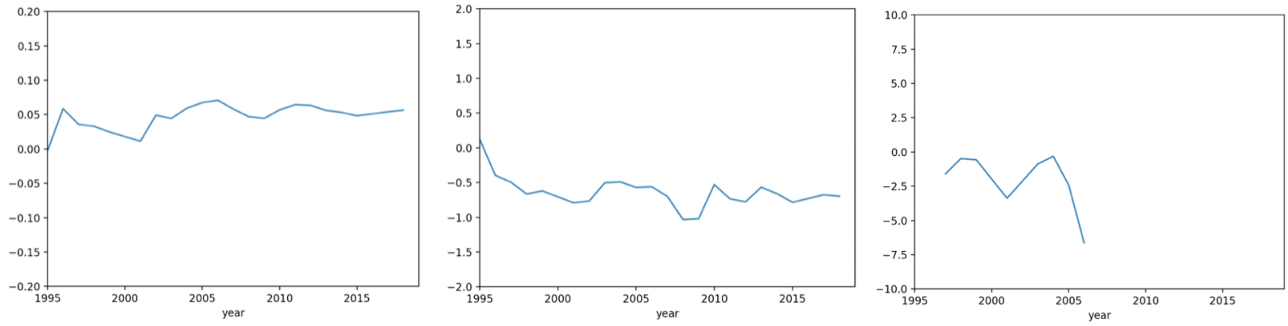


Figure 23 (Above). Means of time series in clusters 1 to 3 (from left to right, respectively)

The general form of the Granger Causality test is as follows:

Unrestricted model:

$$ROA_t = c_1 + \sum_{i=1}^{p} a_i ROA_{t-i} + \sum_{i=1}^{p} b_j x_{t-j} + u_t$$

Restricted model:

$$ROA_t = c_1 + \sum_{i=1}^{p} a_i ROA_{t-i} + u_t$$

$$H_0: b_1 = b_2 = b_3 = \cdots = b_p = 0$$
$$H_1: at\ least\ one\ of\ the\ coefficients\ b_i\ is\ non-0$$

I first estimate the regression, based on the first expression given, and subsequently the second expression, calculating RSS in both cases. The F-test statistic is obtained from the RSS in the first and second regressions, such that:

F = [{(n-k) /p }.{(RSSrestricted-RSSfull) / RSSfull}]

Let x represent the respective attribution metric variable obtained from the texts (we have 6 categories,see legends of Fig 16, 17 and 18), and binary negative/positive performance classifications associated with each category. Hence, in total there are 12 options for x. We also have 3 sets of ROA time series to choose from. Hence I run 36 sets of the Granger Causality test.

Polarity on performance	Attribution Topic	Cluster Number	Pearson Correlation Coefficient	F-Score	p-Value	LAG
Positive	Porter's Five Forces	1 (well/average performing firms)	0.437649	8.02839	0.01062**	1
Negative	Decision Making	1 (well/average performing firms)	-0.624924	7.500814	0.013049**	1
Negative	Institutional/Economic	1 (well/average performing firms)	0.47543	4.578218	0.045574**	1
Positive	Actions	2 (negatively performing firms)	-0.234482	5.233959	0.033787**	1
Negative	Actions	2 (negatively performing firms)	0.505132	2.283736	0.04719**	1
Negative	Competencies	2 (negatively performing firms)	-0.427385	4.637141	0.034616**	5
Negative	Porter's Five Forces	2 (negatively performing firms)	-0.40133	5.095697	0.035951**	1
Negative	Institutional/Economic	2 (negatively performing firms)	-0.465766	4.866507	0.039894**	1

Figure 24 (Above). Results to the Granger Causality Test on Dot-Com-Bubble data

For ease of interpretation, I only included the test results that were statistically significant at the 5% level. I discuss the following results and how they compare with the results in section 7.2:

1. In this section, it is seen that for firms involved in the Dot-Com-Bubble, the attribution of positive performance to Porter's Five Forces yields even better returns for the average performing firms, and negative attribution yield even worse performance for the badly performing firms.
    - The results tell us that firms entrenched in the fads and fashions of the bubble would experience gradually diverging performance upon studying how their performance depends on competitive positioning (strategizing about industrial competition induce firms that experienced positive performance continue to perform strongly, and firms that experienced negative performance perform negatively). Firms with good returns continue to build on their competitive advantage, whereas firms with poor returns blame their luck on the already unfavourable industrial landscape, and thus perform even more poorly.

2. The attribution of negative performance to institutional/economic reasons yields even better returns for the average performing firms, and yields even worse performance for the badly performing firms.
    - This observation is rather unorthodox, as most past researchers who investigated the topic found that attribution of negative performance to institutional/economic reasons yield poorer results to all firms alike. However, I find that the tendency is only pronounced when firms suffer poor returns. It may well be that well-performing firms discuss institutional/economic events for strategic reasons. Overall, this clearly means that the same discussions reveal different things about two different groups of firms. It may also be that low-performing firms are doing badly because of institutional/economic factors, and attributing negative performance to this indicates a lack of control. High performing firms may be doing well in spite of poor economic conditions - highlighting this in the MD&A indicates management are able to identify challenges in the environment but are also able to successfully steer the firm through them.

3. The attribution of positive performance to actions yields worse results and the attribution of negative performance to actions yields better results for badly performing firms.

    - As suggested in section 7.2, firms that dare to admit mistakes and take responsibility tend to have a stronger future performance. The result in this section reinforces this idea. In addition, the attribution of positive performance to firm actions may be an impression management technique to pretend management is in control of the situation. At the same time, management is ignoring the actual drivers of positive performance in the external environment, so will not be able to respond when the external environment later turns unfavourable because all their attention is directed inwards towards internal attributions for performance.

8 Conclusion

In this paper, I investigate the predictability of a firm's performance using information in textual disclosure. Whilst previous publications relied on manual labelling, I seek to engineer a workflow to perform this analysis.

First, I studied the content composition of 10-K MD&A disclosures. Contrary to Dyer (2016), who performed the same analysis on the whole sample of 10-Ks and found that only documents pertaining to compliance with SEC & accounting standards increased markedly in length in the sample period, I find that most of the clusters that experience slow growth in length over time relate to revenue, cash and financials, and focus on describing material performance. Topics that experience high growth are strategic. Discussion topics also exhibit different degrees of cyclicality, and those relating to financial markets experience the most fluctuation in length over time. Secondly, I investigated the emergence of new topics over time using pooled cosine similarities. I find that new topics reflect global and regional macroeconomic events, such as the financial crisis, and evolving fads and trends, such as the technology boom in the early 2000s.

Using the information provided by topic modelling and referring to strategic management literature, I create an econometric model to measure the effect on performance. I leverage a set of dictionaries and design a search algorithm to quantify strategic information in corporate disclosures. Through the empirical results, I find that the development of strategy and its communication are key to favourable future firm performance. Disclosure on decision making and results from firm actions are positively correlated with future performance. Furthermore, the attribution of net positive performance outcome to Porter's Five Forces is negatively correlated with future performance, whilst the attribution of net positive performance outcomes to institutional/economic forces is positively correlated with future performance. These results show that assuming responsibility for decision making is important and the concealment of internal weaknesses is negatively correlated with operational results.

Furthermore, I conduct time series regression to observe the variations of the coefficients on attribution with respect to time. Importantly, I find that firms taking responsibility for their decision making are better at crisis management. First, I find that firms which are aware of their competitive landscape, showing evidence of active decision-making, experience positive returns. On the other hand, firms that attribute negative performance to external events experience negative returns. This observation holds across both the financial crisis in 2008 and the Dot-Com-Bubble in the early 2000s. While past literature has found that there is a negative correlation between future performance and attribution of negative performance to external events, no study thus far has focused on time series differences of attribution and sought to provide explanation with respect to crisis management. Second, I find that the relationship between attribution and performance relates to the nature of the economic crisis. When the nature of the crisis stems from firms' irrationality, studying overconfidence and organizational greed may help to preempt the crisis. Conversely, when the crisis stems from the external economy, studying strategic disclosure and firm responsibility-taking helps to predict which firms can sustain the pressure of the crisis. In the last section of the paper, I provide a case study of the Dot-Com-Bubble from an event-study perspective. I find that firms experiencing negative returns continue performing badly if they cast blame on the external environment; however, firms that experience positive returns are able to improve their performance by publicly disclosing their weaknesses.

It is hoped that the tools and methods developed for this paper can be used to inspire more sophisticated analysis in performance analysis, such as replacing the dictionaries with entities drawn from automatic entity recognition and replacing the nested dictionaries data structure used with knowledge graphs. Similar studies can be done on other forms of corporate texts, such as call transcripts. It should be noted that an inevitable drawback in using using 10-Ks disclosure as a proxy for strategy is that they suffer from the use of impression management techniques, which is found by many authors to be prevalent (e.g. Salancik and Meindl, 1984). Fiol (1995) finds that positive and negative attributions found in annual reports and internal planning documents are not significantly correlated. Finally, it may also be interesting to study the relationship between degree of compliance and ESG related disclosure and firm performance.

9 Appendix

- **Statistical Document Model**

First consider a document model of the following form:

- Words are represented using unit-based vectors that have a single component equal to 1 and all other components equal to 0. As a result, the vth word in the vocabulary is represented by $w^{v} = 1$ and $w^{u} = 0$ for $u \neq v$.
- A document is a sequence of N words denoted by $w = (w_1, w_2, \ldots, w_N)$
- A corpus is formed by a collection of M documents: $D = \{w_1, w_2, \ldots, w_M\}$

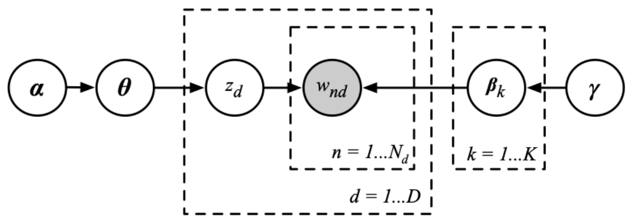
Latent Dirichlet Allocation

LDA is a generative probabilistic model (Blei, 2012). The underlying assumption of LDA is that every document contains a mixture of hidden (latent) topics. Over a distribution of topics, we can infer a topic (and assign a document to it) by choosing the topic that has the highest probability given by a set of words (Blei, Ng, & Jordan, 2003; Blei, 2012).

(i) High Level Summary

LDA assumes that every document will share the same set of topics but exhibit topics in different proportions. For a document consisting of N word positions, we first populate the word positions with the topic from which the word will come from. We then examine the probability mass function from which the word is drawn, and select a word from the topic we picked.

(ii) Model of LDA



$$\begin{aligned} \theta &\sim \mathrm{Dir}(\alpha) \\ \beta_k &\sim \mathrm{Dir}(\gamma) \\ z_d|\theta &\sim \mathrm{Cat}(\theta) \\ w_{nd}|z_d, \beta &\sim \mathrm{Cat}(\beta_{z_d}) \end{aligned}$$

Fig. Description of each stage of the sampling process

- **α** is the proportion parameter.
- w_{nd} is the observed variable of words
- π_d is the per-document topic proportions
- z_d is the per word topic assignment
- β_k is topics
- Gamma is topic distribution

A topic, k, denoted by β_k , is a probability mass function over the entire vocabulary. A topic proportion for document d, denoted by π_d, is a probability function (mixture) over topics for document d.

π is a k-dimensional Dirichlet variable (where the dimension is the predefined number of topics we extract from the documents). π lies in the (k-1)-simplex, if

$if\ \pi_i \geq 0, \quad \sum_{i=1}^{k} \pi_i = 1$, we say that the density function of π~Dir(**α**). The Dirichlet distribution is given by:

$$\mathrm{Dir}(\boldsymbol{\pi}|\alpha_1,\ldots,\alpha_m) = \frac{\Gamma(\sum_{i=1}^m \alpha_i)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^m \pi_i^{\alpha_i - 1} = \frac{1}{\mathrm{B}(\boldsymbol{\alpha})} \prod_{i=1}^m \pi_i^{\alpha_i - 1}$$

The Dirichlet distribution is used because it is conjugate to the multinomial distribution and has finite dimensional sufficient statistics.

For each topic k $\in$ *{1,...,K},* we draw a topic proportion β_k from a Dirichlet Distribution Dir($\boldsymbol{\alpha}$). The K-th dimensional probability vector π, from the Dirichlet distribution goes into a multinomial distribution. P (π$|\alpha$) varies with alpha, when alpha is 1, the probability of each outcome is equally likely. As we vary the parameter alpha, we arrive at different points where the multinomial will land.

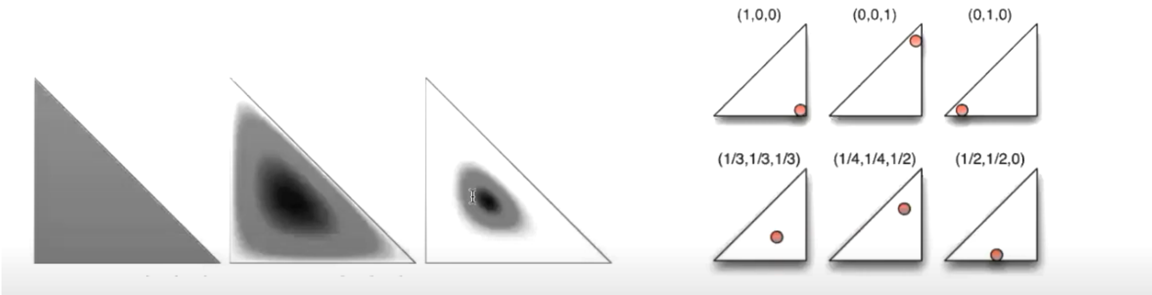


Fig. Dirichlet distribution visualization on a 3-Simplex versus multinomial distributions

The multinomial distribution is a generalization of the binomial function. That is, for n independent trials, each trial is placed into exactly 1 of k categories. The multinomial distribution models the probability of n independent trials each of which leads to a success for exactly one of k categories.

$$p(\mathbf{k}|\boldsymbol{\pi}, n) = p(k_1,\ldots,k_m|\pi_1,\ldots,\pi_m, n) = \frac{n!}{k_1!k_2!\ldots k_m!} \prod_{i=1} \pi_i^{k_i}$$

Then, for each document d $\in$ *{1,...,M},* we draw a multinomial distribution from a dirichlet distribution with parameter α.

Subsequently, for each word position, n $\in$ *{1,...,N},* we select a hidden topic z_n from the topic proportion for the document using the multinomial distribution from the previous step.

Then, for the word position n, we select a word from the corresponding topic β_{z_n}, using the topic selected in the previous step.

Hence, the joint likelihood expression is given by:

$$P(\theta, \beta, w, z) = \prod_{d=1}^{D} P(\theta_d \mid \alpha) \prod_{k=1}^{K} P(\beta_k \mid \eta) \prod_{n=1}^{N} P(z_{d,n} \mid \theta_d) P(w_{d,n} \mid z_{d,n}).$$

Bibliography

Alfano, S. J., Feuerriegel, S., & Neumann, D. (2015). Do pessimists move asset prices? Evidence from applying prospect theory to news sentiment. *SSRN Electronic Journal*. doi:10.2139/ssrn.2602353

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance, 59*(3), 1259-1294. doi:10.1111/j.1540-6261.2004.00662.x

Bao, Y., & Datta, A. (2014). Simultaneously discovering and quantifying Risk types from TEXTUAL RISK DISCLOSURES. *Management Science, 60*(6), 1371-1391. doi:10.1287/mnsc.2014.1930

Barney, J. B. (2001). Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view. *Journal of Management, 27*(6), 643-650. doi:10.1177/014920630102700602

Blei, D.M., et AL. (2003) Latent Dirichlet Allocation. Journal of machine learning Research, 3, 993-1022. - references - scientific research publishing. (n.d.). Retrieved March 12, 2021, from https://www.scirp.org/reference/ReferencesPapers.aspx?ReferenceID=2234902

Boudoukh, J., Feldman, R., Kogan, S., & Richardson, M. (2013). Which news moves stock prices? A textual analysis. doi:10.3386/w18725

Bowman, J. S. (1976). Managerial ethics in business and government. *Business Horizons, 19*(5), 48-54. doi:10.1016/0007-6813(76)90035-5

Davis, A. K., Piger, J. M., & Sedor, L. M. (2006). Beyond the numbers: An analysis of optimistic and pessimistic language in earnings press releases. doi:10.20955/wp.2006.005

Demers, E. A., & Vega, C. (2010). Soft information in earnings announcements: News or noise? *SSRN Electronic Journal*. doi:10.2139/ssrn.1153450

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K TEXTUAL DISCLOSURE: Evidence FROM Latent Dirichlet Allocation. *Journal of Accounting and Economics, 64*(2-3), 221-245. doi:10.1016/j.jacceco.2017.07.002

Eckbo, B. E. (2013). Corporate takeovers and economic efficiency. *SSRN Electronic Journal*. doi:10.2139/ssrn.2340754

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383. doi:10.2307/2325486

Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review, 10*(1), 1. doi:10.2307/2525569

Fiol, C. M. (1995). Corporate communications: Comparing executives' private and public statements. *Academy of Management Journal, 38*(2), 522-536. doi:10.5465/256691

Frazier, K. B., Ingram, R. W., & Tennyson, B. M. (1984). A methodology for the analysis of narrative accounting disclosures. *Journal of Accounting Research, 22*(1), 318. doi:10.2307/2490713

Gooding, S., & Briscoe, T. (2019). Active learning for financial investment reports. Retrieved March 12, 2021, from https://www.aclweb.org/anthology/W19-6404.pdf

Granger, C. W. (1969). Investigating causal relations By econometric models And cross-spectral methods. *Econometrica,37*(3), 424. doi:10.2307/1912791

Hanley, K. W. (2016). Dynamic interpretation of emerging systemic risks. *SSRN Electronic Journal*. doi:10.2139/ssrn.2792943

Hoberg, G., & Maksimovic, V. (2014). Redefining financial constraints: A text-based analysis. *The Review of Financial Studies, 28*(5), 1312-1352. doi:10.1093/rfs/hhu089

Huang, A., Lehavy, R., Zang, A. Y., & Zheng, R. (2014). A thematic analysis of analyst information discovery and information interpretation roles. *SSRN Electronic Journal*. doi:10.2139/ssrn.2409482

Hüfner, B. (2007). The SEC's MD&A: Does it meet the INFORMATIONAL demands of investors? *Schmalenbach Business Review, 59*(1), 58-84. doi:10.1007/bf03396742

Jin, F., Self, N., Saraf, P., Butler, P., Wang, W., & Ramakrishnan, N. (2013). Forex-foreteller. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. doi:10.1145/2487575.2487710

Ke, Z. T., Kelly, B., & Xiu, D. (2019). Predicting returns with text data. doi:10.3386/w26186

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems, 104*, 38-48. doi:10.1016/j.dss.2017.10.001

Loughran, T., & McDonald, B. (2011). Measuring readability in financial disclosures. *SSRN Electronic Journal*. doi:10.2139/ssrn.1920411

Maia, M., Freitas, A., & Handschuh, S. (2018). FinSSLx: A sentiment analysis model for the financial domain using TEXT SIMPLIFICATION. *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*. doi:10.1109/icsc.2018.00065

Malo, P., Sinha, A., Korhonen, P., Wallenius, J., & Takala, P. (2013). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for* Information Science and Technology, 65(4), 782-796. doi:10.1002/asi.23062

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge (Massachusetts): MIT Press.

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin, 82*(2), 213-225. doi:10.1037/h0076486

Mintzberg, H. (1990). The design school: Reconsidering the basic premises of strategic management. *Strategic Management Journal, 11*(3), 171-195. doi:10.1002/smj.4250110302

Neuhierl, A., Scherbina, A. D., & Schlusche, B. (2010). Market reaction to corporate press releases. *SSRN Electronic Journal*. doi:10.2139/ssrn.1556532

Oliveira, N., Cortez, P., & Areal, N. (2016, March 05). Stock market sentiment lexicon acquisition using microblogging data and statistical measures. Retrieved March 12, 2021, from https://www.sciencedirect.com/science/article/pii/S0167923616300240

Peteraf, M. A. (1993). The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal, 14*(3), 179-191. doi:10.1002/smj.4250140303

Porter, M. E., (1979). How Competitive Forces Shape Strategy, Harvard Business Review, vol. 57, issue 2, pp. 137-145.

Porter, M. E., (1985). Competitive Advantage: Creating and Sustaining Superior Performance, Free Press, New York.

Porter, M. E., (1996). What is Strategy?, Harvard Business Review, vol. 74, issue 6, pp. 61-78

Routledge. R. B., Sacchetto, S & Smith, A, N. (2013). Predicting Merger Targets from Acquirers from Text,

Salancik, G. R., & Meindl, J. R. (1984). Corporate attributions as strategic illusions of management control. *Administrative Science Quarterly, 29*(2), 238. doi:10.2307/2393176

Sinha, A., Kedas, S., Kumar, R., & Malo, P. (2019). Buy, Sell or Hold: Entity-Aware Classification of Business News. Technical report, Working Paper. Indian Institute of Management Ahmedabad, Research and Publication Department.

Staw, B. M., McKechnie, P. I., & Puffer, S. M. (1983). The justification of organizational performance. *Administrative Science Quarterly, 28*(4), 582. doi:10.2307/2393010

Tetlock, P. C. (2005). Giving content to investor sentiment: The role of media in the stock market. *SSRN Electronic Journal*. doi:10.2139/ssrn.685145

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2007). More than words: Quantifying language to measure firms' fundamentals. *SSRN Electronic Journal*. doi:10.2139/ssrn.923911