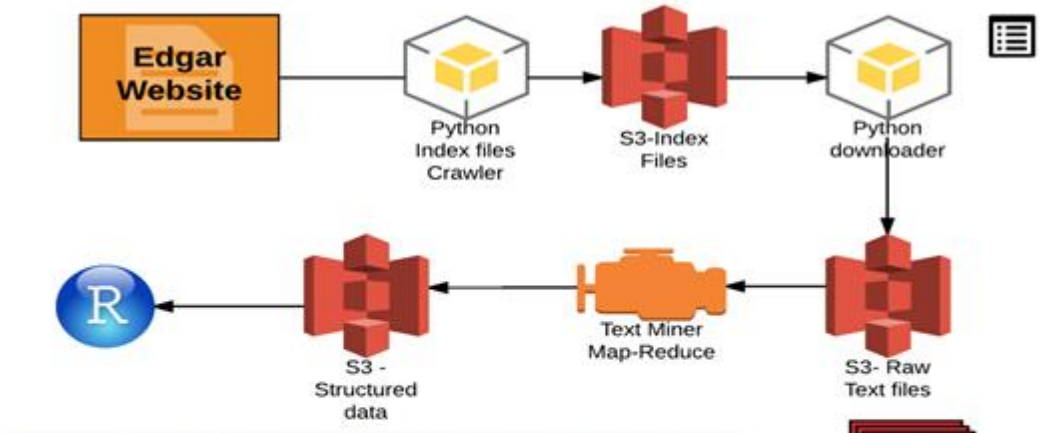
April 20, 2017

TEXT MINING UNSTRUCTURED CORPORATE FILING DATA

Systematically Profiling the EDGAR Database

- **Web Scraping Unstructured Textual Information.** With the rapid development and innovation in computing power and machine learning algorithms, processing unstructured textual information to generate useful numerical signals becomes increasingly important. In this research we take advantage of distributed cloud computing and advanced Natural Language Processing (NLP) algorithms to systematically analyze corporate filings from the EDGAR database – the official corporate filing database maintained by the SEC.
- **NLP Stock Selection Signals.** Equipped with our own NLP algorithms, we study a wide range of models based on corporate filing data: measuring the document tone or sentiment with finance oriented lexicons; investigating the changes in the language structure; computing the proportion of numeric versus textual information, and estimating the word complexity in corporate filings; and lastly, using machine learning algorithms to quantify the informative contents.
- **Systematic Profiling EDGAR Composite (SPEC) Model.** We create a composite stock selection model – SPEC, by combining the best text mining signals. Our SPEC model generates strong and consistent performance, with slow signal decay and low portfolio turnover. It is uncorrelated to traditional factors and highly complementary to traditional quantitative models.
- **A Useful Tool for Fundamental Managers.** Rather than manually reading and sifting through 5,000 lengthy corporate filings (there are 150,000 words in an average 10-K document) every day, our NLP models help discretionary portfolio managers make more timely and accurate investment decisions.



CIK	Fiscal Date	Report Date	Document	Section	Word	Count
0000000020	19951231	19960328	10-K	1	INCLUDED	1
0000000020	19951231	19960328	10-K	15	ACCOUNTED	2
0000000020	19951231	19960328	10-K	15	AFFILIATES	3
0000000020	19951231	19960328	10-K	15	DEPRECIATED	1
0000000020	19951231	19960328	10-K	15	DISPOSED	1
0000000020	19951231	19960328	10-K	15	FINANCING	4
0000000020	19951231	19960328	10-K	15	OVERALL	2

Gaurav Rohal, CFA
GRohal@wolferesearch.com

Yin Luo, CFA, CPA
YLuo@wolferesearch.com

Javed Jussa
JJussa@wolferesearch.com

Sheng Wang
SWang@wolferesearch.com

QES Desk Phone: 1.646.582.9230
Luo.QES@wolferesearch.com

Source: Wolfe Research Luo's QES

Table of Contents

A LETTER TO OUR READERS

Today's investing world is characterized by information overload. Traditionally, investors focus on numeric data, using either fundamental analysis or quantitative models. However, the vast majority of the available information is in unstructured formats, e.g., text, audio, video, and image. For example, there are almost 5,000 documents filed by public companies in the US every day. On average, there are over 150,000 words in a typical 10-K – the standard annual filing of a company's performance required by the SEC (Securities and Exchange Commission). As argued in our previous research (see Luo, et al [2017a, 2017b, and 2017c], the future of active investing rests on how to best incorporate unique data with sophisticated modeling techniques.

In this paper, we showcase how web scraping, distributed cloud computing, natural language processing (NLP), and machine learning techniques can be applied to systematically analyze corporate filings from the EDGAR (Electronic Data Gathering, Analysis, and Retrieval) database. The EDGAR system is used by the SEC to disseminate business relevant information. In addition to the detailed accounting data presented in the financial statements, firms are also required to provide thorough descriptive information in their filings.

We parse, store and sift through these filings for relevant qualitative information. The focus is how to best quantify descriptive textual documents into investment intelligence. The overall process is intricate, as companies often tread far from the standard format specified by the SEC, which forces us to make custom adjustments.

We explore a range of innovative NLP algorithms to quantify the textual information embedded in the annual and quarterly corporate filings. We use financial lexicons to categorize the text contained in these sections with sentiments – positive or negative (bullish or bearish). We look at the overall proportion of textual content to the numeric data. Lastly, when firms make an active decision to significantly change the wording and language embedded in their regular filings, these conscious adjustments have considerable implications for future firm behavior.

We find that not all sections of firm filings are created equally. Beyond the numeric financial data, MD&A (Management's Discussion and Analysis) and FSS (Financial Statements and Supplemental) sections contain useful information. However, the surprising part is that the one section with the most predictive textual information is the "Risk Factors", which is arguably among the least read categories by investors[1].

We explore various ways to combine the disparate signals we create from the textual content of the 10-K and 10-Q filings. In the base case, we define two equally weighted composite signals (EQ10K and EQ10Q), using the most prominent factors. Alternatively, we explore nonlinear machine learning algorithms, for automatic feature selection. We also evaluate the readability or complexity of the filing statements, to improve the explanatory power of our models. In the end, we introduce our SPEC (Systematic Profiling EDGAR Composite) model – a stock selection model based on sophisticated web scraping, cloud computing, and NLP algorithms. The SPEC model has strong and consistent performance, with low turnover, slow decay, and is uncorrelated to traditional factors.

[1] To protect themselves, companies typically list all possible potential risk factors to their operations and businesses in the "Risk Factors" section. Because the contents are extremely broad and generic, few investors even bother to read this section.

Last but not least, for fundamental discretionary portfolio managers, rather than manually reading thousands of lengthy corporate filing documents, our NLP models help them make more timely and accurate investment decisions.

We have a regular data feed and a web portal based on the SPEC model and its underlying components. Interested readers should contact your sales representative or Luo.QES@wolferesearch.com for more details.

Regards,

Yin, Gaurav, Javed, Sheng, and Luo's QES Team

INTRODUCTION

The vast majority of information related to the world of investing is arguably unstructured in the form of text, image, audio, or video. How to transform unstructured information into numeric data in real time requires a suite of integrated systems, from web scraping, data collection, distributed parallel computing, advanced Natural Language Processing (NLP), to machine learning techniques. As discussed in Luo, et al [2017a, 2017b, and 2017c], the underwhelming performance of traditional factors and the technological advancements have unleashed an arms race to acquire and process alternative data sources. In this research, we demonstrate how NLP and machine learning can be used in processing corporate filing data from the EDGAR database to generate highly effective, low turnover, and orthogonal stock selection signals.

LITERATURE REVIEW

Most empirical studies in corporate finance and asset pricing focus on quantitative information from corporate accounting statements and press releases. Only recently some attention has been given to the qualitative or textual information embedded in these documents. One of the early works in this area of textual and sentiment analysis was done by Antweiler and Frank [2004]. The authors use linguistic methods to study the effect of messages posted on *Yahoo! Finance* and *Raging Bull*, for the companies in the Dow Jones Industrial Average and the Dow Jones Internet Index.

Another notable pioneer paper was Tetlock [2007], conducting textual analysis on the content of a popular *Wall Street Journal* column with the Harvard psychosocial lexicon. Tetlock [2007] finds that high media pessimism predicts downward movement of stock market, followed by a reversion to fundamentals, and unusually high or low pessimism predicts high market trading volume. In another study (see Tetlock, et al [2008]), researchers find that higher percentages of negative words in firm news predict lower future quarterly earnings. The general-purpose Harvard dictionary used by Tetlock [2007] and Tetlock, et al [2008] was further refined by Loughran and McDonald [2011]. They suggest that the Harvard psychological dictionary may not be suitable for finance and accounting applications. Since the meaning of positive and negative words may have very different connotations in a financial context. Loughran and McDonald [2011] re-classify positive and negative words from a finance perspective. They were also among one of the first to apply NLP on the SEC Form 10-K filings and link the signals to stock returns, trading volume, return volatility, fraud, material weakness, and unexpected earnings.

Recently, there is a tremendous growth in academic literature on textual analytic studies, with one group focusing on news, financial columns and social media. Chen, et al [2013] investigates the extent to which investor opinions transmitted through the social media on the predictability of future stock returns and earnings surprises. Garcia [2013] finds that the predictability of stock returns using news content is concentrated in recessions. Boudoukh, et al [2013] find that news that can be identified and classified in certain categories has a higher impact on stock markets than unidentified news. Chouliaras and Grammatikos [2013] argue that higher news pessimism in times of crisis is associated with lower stock returns.

The other branch of study focuses on corporate financial disclosure through earnings reports, press release and conference calls. Li [2010] studies the effect of forward-looking statements in corporate flings on future earnings and liquidity. Feldman, et al [2010] find that a positive tone in the MD&A

section is associated with higher contemporaneous and future returns, and that an increasingly negative tone is associated with lower contemporaneous returns. Davis and Tama-Sweet [2011] suggests a negative relationship between the level of pessimistic language in the MD&A section of SEC filing and future firm performance, controlling for pessimistic language in the corresponding earnings press release. Mayew, et al [2013] examine the pitch of CEO's voices and their labor market success.

Furthermore, textual analysis has been used for the study of initial public offerings (IPOs). Loughran and McDonald [2013] find that IPOs with high levels of uncertain text have higher first-day returns, absolute offer price revisions, and subsequent volatility. Finally, Ahern and Sosyura [2014] show evidence of firms manipulating media coverage to achieve better stock prices during mergers and acquisitions negotiations.

Beyond words, a phrase frequency count can be done on documents, but it increases computational complexity greatly. In order to control complexity, topic modeling on select phrases can be performed. Israelsen [2014] extracts a comprehensive set of disclosed risk factors into which all firms' risk disclosures can be classified. He finds disclosed risks about credit markets and constraints explain much of the variation in the Fama-French factors, suggesting small and value firms may be more risky in downturns. Ball, et al [2014] also use a topic model on the MD&A section of firms' 10-K filings. Huang, et al [2014] fit a topic model to transcripts from analyst conference calls and analyst reports; to examine sell-side equity analysts' information interpretation and discovery roles. Hoberg and Maksimovic [2014] implement proximity search (another form of topic modeling) on the 10-Ks to identify financially constrained firms. They focus their word search on the Liquidity and Capitalization Resource Subsection in the MD&A section. The authors create a measure of delayed investment where they search for words such as *delay, abandon, eliminate,* or *postpone* within 12 words of investment-type words like *construction* or *expansion*.

In this paper, we focus on corporate disclosures through SEC filings and at the same time, try to keep our readers abreast with the overall process of text mining and sentiment analysis.

THE SEC EDGAR CORPORATE FILING DATABASE

We use the EDGAR (the Electronic Data Gathering, Analysis, and Retrieval) system, as our source of corporate filing information. The EDGAR is used by the US SEC (Securities and Exchange Commission) to disseminate business relevant information. The database is freely available to the public via the Web or FTP.

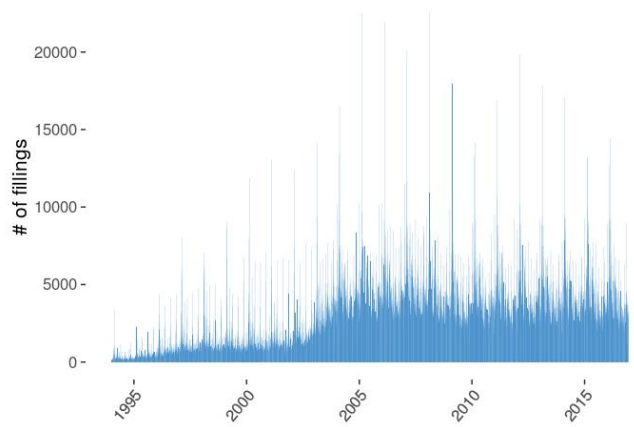
The EDGAR system for electronic filing was developed in early 1990s. Almost all companies with public securities in the US are required to file various documents with the SEC via the EDGAR system. The vast majority of documents are now filed electronically, with around 5,000 filings per day (see Figure 1 A). Filers in the EDGAR system are identified using the Central Index Key (CIK). Only a fraction of these fillers are publicly traded firms and many filers are private firms, mutual funds and individuals. With more than 100 types of forms submitted to EDGAR, Form 4 is the most common type (see Figure 1 B).

Form 4 is related to insider transactions. Directors, officers or owners of more than 10% of any class of equity securities registered under Section 12, are required to report purchases or sales of securities using Form 4. The second most common EDGAR filing is Form 8-K. Public companies use

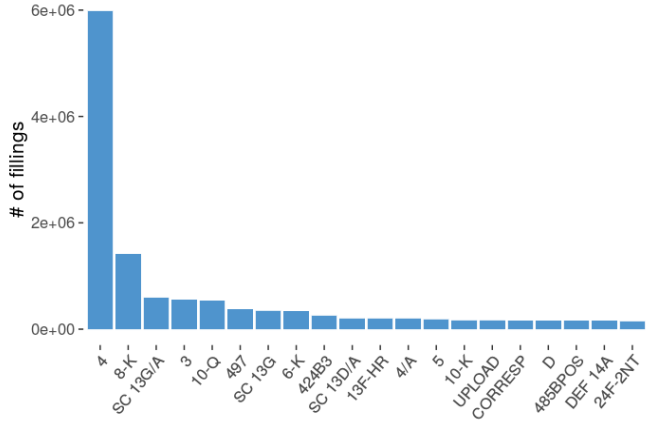
this form to notify investors on events that are of material importance to the firm. The 8-K statements include information on management changes, new contracts, M&A, lawsuits, impairments, and other significant events. In a recent research (see Jussa, et al [2017]), we conduct a corporate governance and fraud detection study using the information derived from the Form 8-K. Other common form types include Form 3 (used to report preliminary insider transaction) and Form Schedule 13G (filed by anyone who acquires ownership in a public company of more than 5% of the outstanding stock).

Figure 1 Number of filings in the EDGAR database

A) Number of filings (daily)



B) Number of filings (form type)

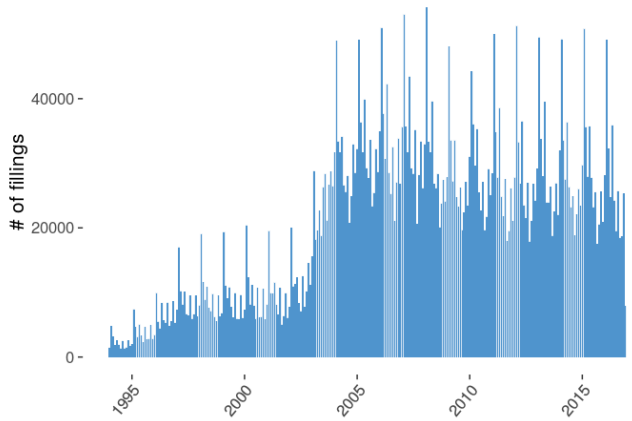


Sources: EDGAR, Wolfe Research Luo's QES

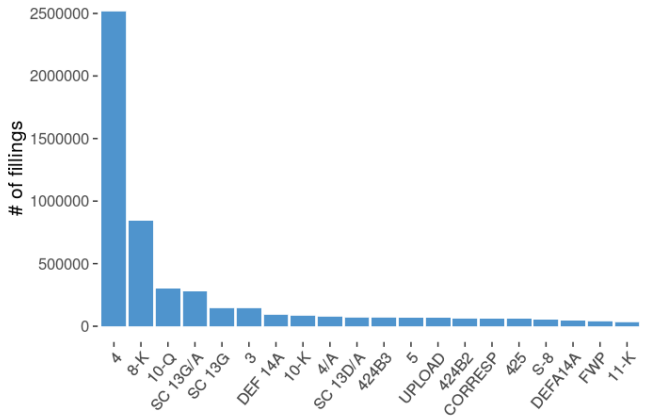
Form 10-K (annual company financial statements) and Form 10-Q (quarterly company financial statements) also appear prominently in the EDGAR filing system, especially by public companies. Figure 2 shows the EDGAR filing frequencies for public companies in the Russell 3000 index. There has been a sharp rise in number of filings since early 2000s following the Sarbanes–Oxley Act of 2002, also known as the "Public Company Accounting Reform and Investor Protection Act". This regulation imposes new and expanded requirements for all U.S. public company boards, management and public accounting firms.

Figure 2 Number of filings by the Russell 3000 index constituents in the EDGAR database

A) Number of filings (monthly)



B) Number of filings (form type)

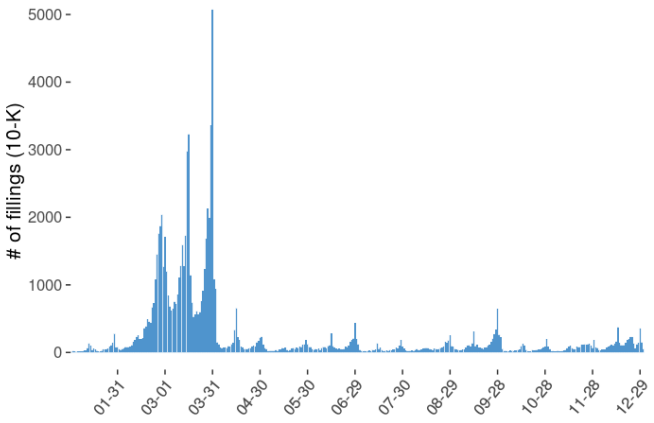


Sources: FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

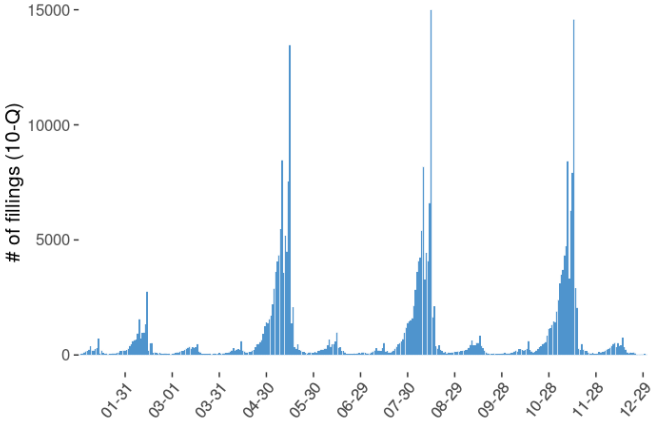
Figure 3, shows the filing frequencies for the 10-K and 10-Q over time. As expected, nearly all the 10-K filings occur at the start of the year in the month of February and March, since most companies have a December 31 fiscal year end. The 10-K filings peak around end of March, which coincides with the 90-day SEC filing deadline for annual reporting. While quarterly filings peak in the second month post each quarter end, peaking around the SEC's 45-day deadline for interim reporting.

Figure 3 Seasonality of the 10-K and 10-Q EDGAR filings by the companies in the Russell 3000 index

A) Number of 10-K filings around the year



B) Number of 10-Q filings around the year



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

WEB SCRAPING EDGAR

We certainly do not want to manually download thousands of filings from the EDGAR database every day. Therefore, the first task that we need to do, before we can conduct our NLP and machine learning exercise is to web scrape the EDGAR system automatically, based on a similar framework as the L-Scholar (see Luo, et al [2016]).

EDGAR DATA STRUCTURE

EDGAR provides daily/quarterly master index files to effectively download company filings from its website. The daily index file contains information about all the filing done on that specific date. Figure 4 shows a sample master index file for Apple (CIK 320193).

Figure 4 Sample master index files

CIK	COMPANY_NAME	FORM_TYPE	DATE_FILED	EDGAR_LINK
320193	APPLE INC	8-K	4/26/2016	edgar/data/320193/0001193125-16-556520.txt
320193	APPLE INC	8-K	6/22/2016	edgar/data/320193/0001193125-16-628957.txt
320193	APPLE INC	FWP	6/7/2016	edgar/data/320193/0001193125-16-615362.txt
320193	APPLE INC	S-3ASR	4/28/2016	edgar/data/320193/0001193125-16-564177.txt
320193	APPLE INC	10-Q	7/27/2016	edgar/data/320193/0001628280-16-017809.txt
320193	APPLE INC	424B2	7/28/2016	edgar/data/320193/0001193125-16-661201.txt

Sources: EDGAR, Wolfe Research Luo's QES

For example, Apple filled a 10-Q report on July 27, 2016. The complete submission text file associated with this filing can be accessed using the URL https://www.sec.gov/Archives/ followed by EDGAR link. The same file can also be accessed through an FTP site.

In the next section, we highlight the 10-K and the 10-Q form format along with our parsing process.

Form 10-K structure

Form 10-K is reported in the following formats with a list of required Items. It has sections from one to 15 as specified below. All 15 sections are required except in certain special situations.

PART I

- Item 1. Business
- Item 1A. Risk Factors
- Item 1B. Unresolved Staff Comments
- Item 2. Properties
- Item 3. Legal Proceedings
- Item 4. Mine Safety Disclosures

PART II

- Item 5. Market for Registrant's Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities
- Item 6. Selected Financial Data
- Item 7. Management's Discussion and Analysis of Financial Condition and Results of Operations
- Item 7A. Quantitative and Qualitative Disclosures About Market Risk
- Item 8. Financial Statements and Supplementary Data
- Item 9. Changes in and Disagreements With Accountants on Accounting and Financial Disclosure
- Item 9A. Controls and Procedures
- Item 9B. Other Information

PART III

- Item 10. Directors, Executive Officers and Corporate Governance
- Item 11. Executive Compensation
- Item 12. Security Ownership of Certain Beneficial Owners and Management and Related Stockholder Matters
- Item 13. Certain Relationships and Related Transactions, and Director Independence
- Item 14. Principal Accounting Fees and Services

PART IV

- Item 15. Exhibits, Financial Statement Schedules

Form 10-Q structure

Form 10-Q is reported in the following formats with a list of required Items. It has two parts with sections from one to four and one to six as specified below. Similar to the 10-K, all sections are required, except in certain special situations.

PART I

- Item 1. Financial Statements
- Item 2. Management's Discussion and Analysis of Financial Condition and Results of Operations
- Item 3. Quantitative and Qualitative Disclosures about Market Risk
- Item 4. Controls and Procedures

PART II

- Item 1. Legal Proceedings
- Item 1A. Risk Factors
- Item 2. Unregistered Sales of Equity Securities and Use of Proceeds
- Item 3. Defaults upon Senior Securities
- Item 4. Mine Safety Disclosures
- Item 5. Other Information
- Item 6. Exhibits

MAP-REDUCE FRAMEWORK FOR TEXT MINING THE SEC 10-K/10-Q FILINGS

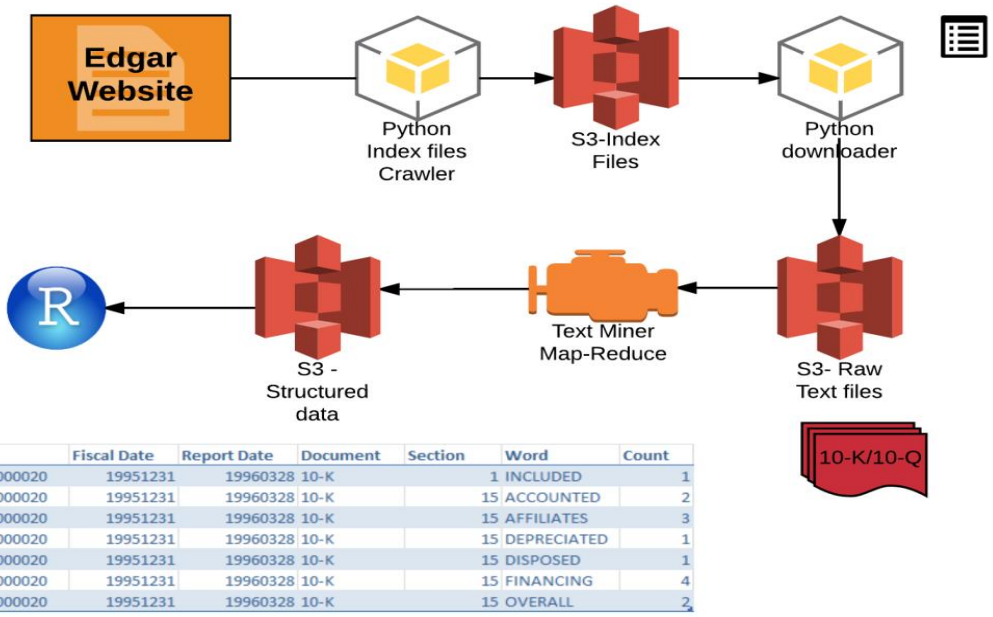
The readily available resources provided by the cloud computing providers such as Amazon, Microsoft, and Google have democratized the access to distributed computing. Once exclusively employed by a few selected elites, now the power of large scale data mining and analytics is within the reach of many firms. This has opened up a great number of opportunities in the investment world that were once deemed intractable or impractical. The problem of extracting relevant information from textual information is notably "embarrassingly parallel" and hence should greatly benefit from distributed computing.

Methodology

We utilize Hadoop framework for performing large scale data mining, as shown in Figure 5. Amazon Web Services (AWS) offers a simple off-the-shelf Hadoop implementation called Elastic Map Reduce (EMR). EDGAR website provides daily/quarterly master index files that are used to identify relevant filing documents and location.

- EDGAR Document Index files are used to identify the list of relevant documents to download. We write a simple Python script to download all index files and merge them into one single document.
- Based on the list, we develop a Python based web crawler to download filing documents from the website. The documents are then stored into a distributed file system.
- We create a suite of customized mapper functions for Map-Reduce framework. The mapper functions classify and extract relevant keywords, pairs and numbers from the previously downloaded text files. Note that since the text files are stored in a distributed file system, we are able to leverage parallelized computing to speed up the computations involved in mining process by an order of 100x or more.
- We then build a customized reduction function that aggregates words based on company, type of document, date of filing and reporting period and section. Structured data is then pulled into R for further processing.

Figure 5 Map-Reduce framework schema



CIK	Fiscal Date	Report Date	Document	Section	Word	Count
0000000020	19951231	19960328	10-K	1	INCLUDED	1
0000000020	19951231	19960328	10-K	15	ACCOUNTED	2
0000000020	19951231	19960328	10-K	15	AFFILIATES	3
0000000020	19951231	19960328	10-K	15	DEPRECIATED	1
0000000020	19951231	19960328	10-K	15	DISPOSED	1
0000000020	19951231	19960328	10-K	15	FINANCING	4
0000000020	19951231	19960328	10-K	15	OVERALL	2

Sources: EDGAR, Wolfe Research Luo's QES

EDGAR DATA PARSING

Once we transform unstructured textual information into structured data, we can then extract the key words and phrases, and more importantly, use NLP to develop numerical signals.

Extracting components from the "complete submission" text files

- The SEC provides a full text version of filings called complete submission file as well as a browser friendly version. For example here is an example of complete submission text file for Apple and here is the browser friendly version of the same document.
- The browser friendly version lists the core 10-K filing document in a HTML format, including eight exhibits, two graphics files, and six XBRL files. All of these files are also contained in a complete submission text file with the embedded HTML, XBRL, exhibits, and the ASCII-encoded graphics. Due to this embedding feature, the actual text contained in the complete submission text file is much smaller (less than 10% of the browser friend version). The bulk of the complete submission text file is composed of HTML codes, XBRL tags and ASCII-encoded graphics.
- XBRL (eXtensible Business Reporting Language) is a markup language that provides standardized specifications to describe financial information for public and private companies. It is a variant of XML and therefore follows the XML syntax and technologies. It provides semantic context for data reported within a financial statement such as the 10-Ks.

- For the scope of this research paper, we focus only on the textual content of the document, and therefore we filter out all the unwanted content from the complete submission text file. We can use a HTML parser to extract all the textual information from the report. Alternatively, we can remove all the markup tags (HTML, XBRL, XML), ASCII-encoded graphics from the file and use the remaining sections for textual analysis.

- We also need to decide whether to exclude all the tables from the text file, since tables mostly contains numeric data. However, the tables still contains some textual information and provides highlights of key company performance parameters, which we shall exploit in our research. Therefore, we include the information contained in the tables at this stage.

- In some instances, the sections begin with a table of data, where the Item name demarcation appears within the table. If we exclude all the tables at this stage, we need to ensure that those tables containing Item tags are *not* deleted; otherwise there can be a significant loss of data for some sections of the document.

- The goal is to capture each section of the report, from Item 1 to 15 in the case of the 10-K filing. Hering [2016] and McDonald [2012] give detailed overview of the process in their respective research papers. McDonald [2012] provides extensive information on extracting textual information from EDGAR filings on his website.

As outlined above, the parsing process is intricate, as companies often tread far from the standard format specified by the SEC. This non-standardization complicates the whole process further and we are forced to choose between quality and quantity. Keeping the parsing methodology too tight will result in poor coverage, while loosening it too much is prone to errors. Since the EDGAR filings are done by thousands of independent fillers, the natural language text and financial tables are weakly structured and filled with formatting errors and inconsistencies (see Stümpert [2008] and O´Riain [2012]). This is especially true for early years when the use of markup languages was less common (see Loughran and McDonald [2014]). A limited number of tagged items, formatting errors and other inconsistencies lead to difficulties in accurately identifying and parsing common textual subjects across multiple filings over a long history.

Natural language processing

Once we have parsed the text, it is the time for Natural Language Processing (NLP) to go on the stage. NLP mostly starts with a frequency count of words appearing in a textual document. Statistical programming languages such as R and Python have many libraries for text computing. Our NLP algorithm does the following steps:

- Create a Corpus which is a collection of documents containing natural language text.

- Next we remove punctuation and numbers, and further convert all the text to upper cases.

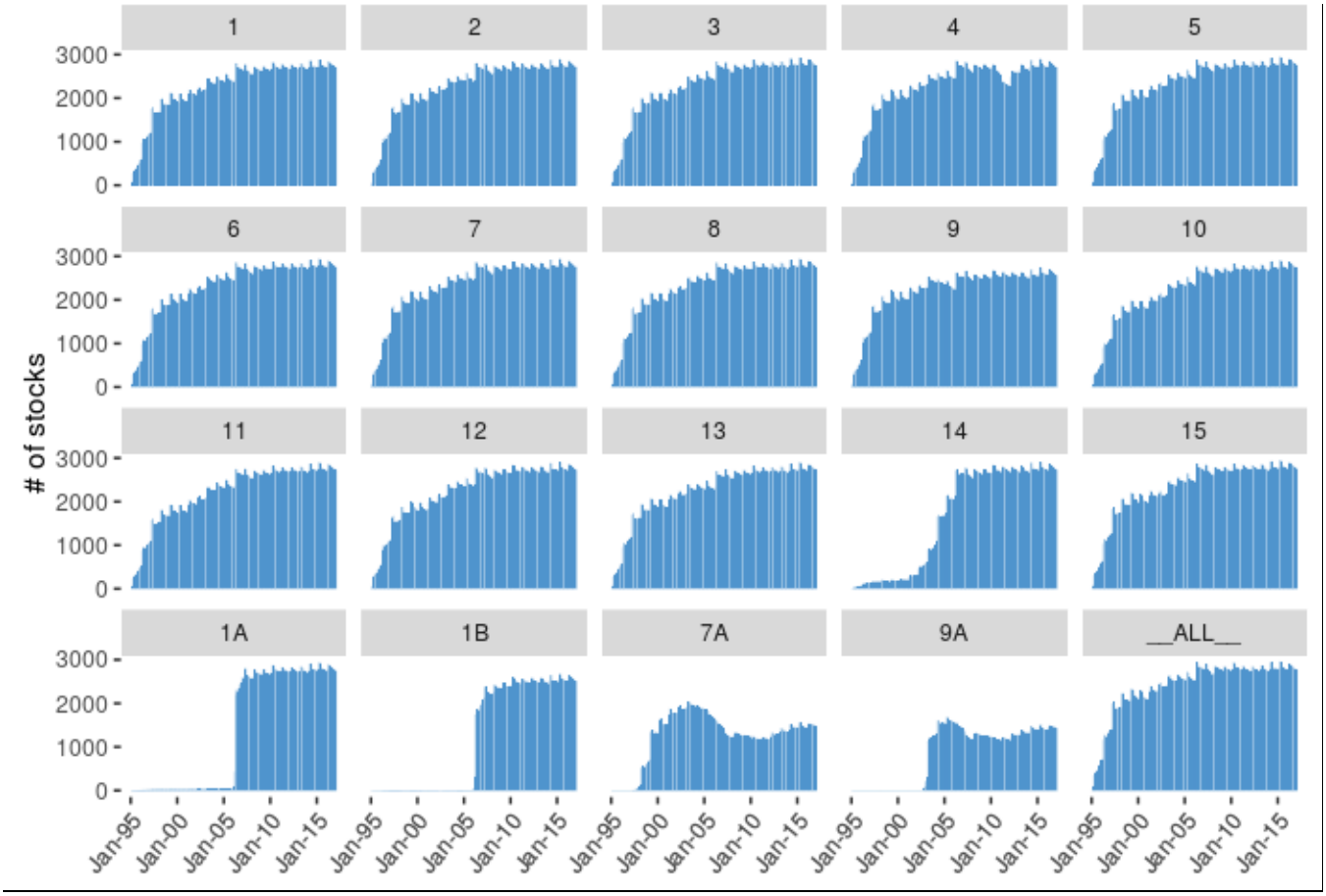
- Furthermore, we remove all the English stop words from the text document. We also remove words which are not relevant for our needs, such as generic words, names, geographies. We again refer our readers to Bill McDonald's website for a list of relevant stop words.

- The next step is to stem the document. Stemming is the process of reducing words to their base or roots, so that to group together words with similar meaning. For this paper, we decide not to stem the document, as we wish to keep the words with their original connotation.
- At the end of this exercise, we should have a list of words and their frequency count for each text section.

OVERVIEW OF THE FORM 10-K SECTIONS

Figure 6 show the coverage of the 10-K filings successfully parsed by our program for stocks in the Russell 3000 index. Coverage has gradually improved from close to zero in 1995 to almost complete in the past 10 years for almost all individual sections. Some sections start very late. Section 7A starts in late 1990s, while Sections 1A, 1B and 9A begin around 2005. Beginning 2006, the SEC requires firms to include in their annual 10-K filing a discussion of the significant factors that make the offering speculative or risky. This "Risk Factors" section (Item 1A) is now mandatory for all but the smallest filers.

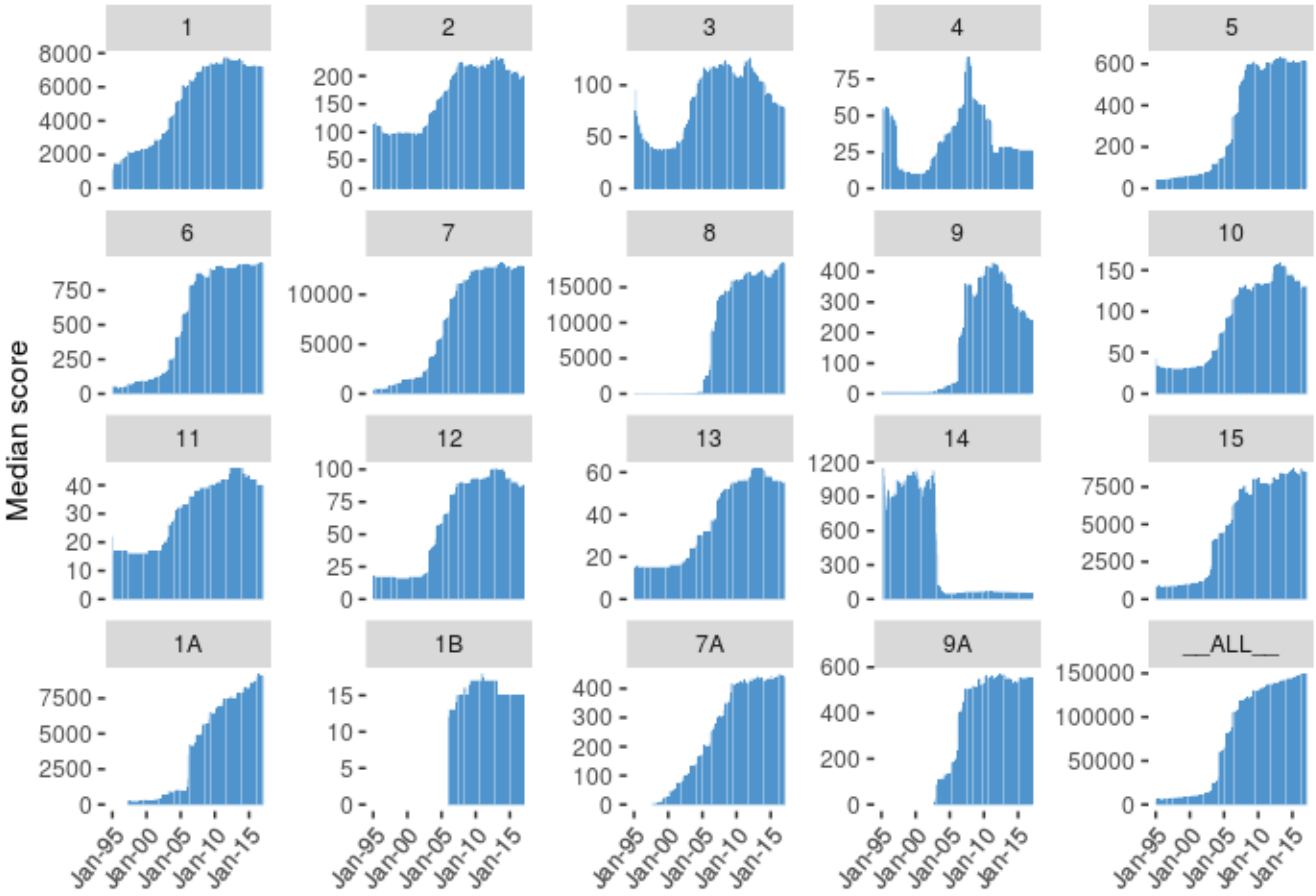
Figure 6 Coverage by each section of the 10-K filings, Russell 3000 firms



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 7 shows the number of words in each section of the 10-K Annual report. Clearly the complexity of the 10-K is growing over time as the number of words in each section has spiked multiple folds. Hence a systematic analysis of textual content becomes ever more relevant. Not all sections are given equal attention by the filers. There are few words in Items 1B, 9B, 4, and 11 to 14. Some sections only pertain to certain industries, e.g., Item 4 (Mine Safety Disclosures); and therefore are irrelevant for most firms.

Figure 7 Average number of words in each section of the 10-K filings, Russell 3000 firms



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Word cloud of textual documents

Another way to quickly glance the common features of text documents is a word cloud. The word cloud is simply a term frequency plot across all the documents. Figure 8 shows the word clouds for the "Risk Factors" (Item 1A) and "Market Risk" (Item 7A) sections of the 10-K filings. Obviously, the Risk factor section in 10 K filings is dominated by terms such as customer, services, products, future ability and regulatory concerns. On the other hand, the "Market Risk" section is dominated by terms such as interest rate, market risk, debt, currency and foreign exposures.

Figure 8 World clouds of the "Risk Factors" and "Market Risk" sections in 10-K filings

A) "Risk Factors" section

B) "Market Risk" section

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

OVERVIEW OF THE FORM 10-Q SECTIONS

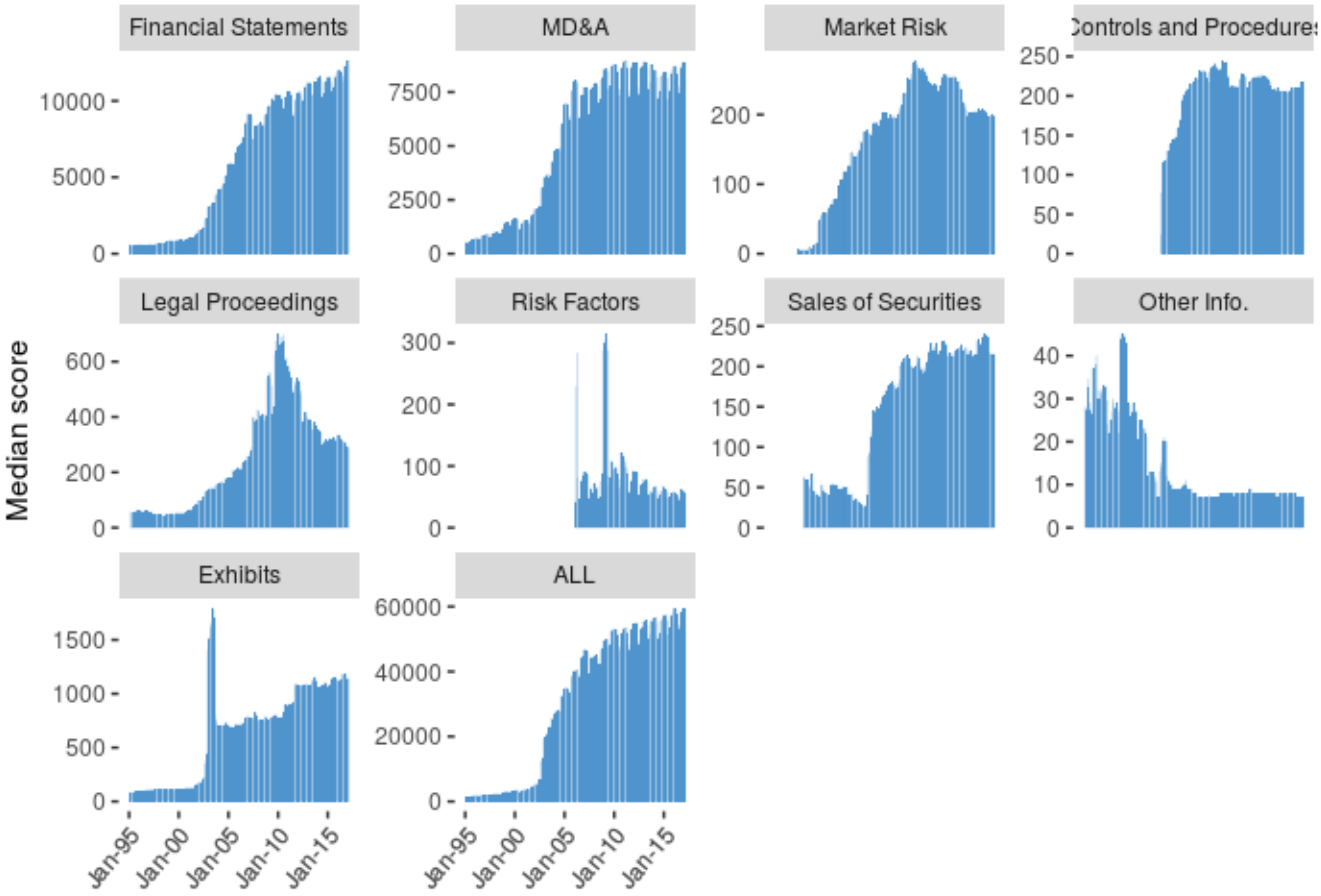
Figure 9 shows the coverage of the 10-Q filings parsed based on the firms in the Russell 3000 index. Similarly to the 10-K's, the coverage has gradually improved due to more standardized reporting by companies. We are able to better capture certain sections such as the MD&A and "Market Risk" than other sections such as the Risk factors. In addition, reporting for certain sections such as the "Risk Factors" and "Control and Procedures" again start only in recent years.

Figure 9 Coverage by each section of 10-Q, Russell 3000 firms

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 10 shows the number of words in each section of the 10-Q quarterly filing. The patterns are similar to what we observe in the 10-K, e.g., rising complexity over time.

Figure 10 Average number of words in each section of 10-Q, Russell 3000 firms



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Overall, based on the availability and relevance, the prominent sections for the 10-K filings seem to be Item 1 (Business), Item 1A (Risk Factors), Item 7 (MD&A), Item 7A (Quantitative and Qualitative Disclosures about Market Risk), Item 8 (Financial Statements and Supplementary Data) and Item 9A (Controls and Procedures).

Similarly, for the 10-Q filings, the salient sections are FSS, "Risk Factors", MD&A, "Quantitative and Qualitative Disclosures about Market Risk", "Legal proceedings", and "Controls and Procedures".

TEXT MINING AND NLP SIGNALS

Often before a company files its form 10-K/10-Q with the SEC, it already discloses nearly all of the key numerical information to the public. Sales, net income, EPS, dividends, and other key measures of financial performance are often disclosed in earnings announcements and conference calls weeks or even months before the filing of 10-K/10-Q. As a result, few investors actually read the voluminous 10-K/10-Q filings and they tend to underestimate the importance these documents, leading to market under-reaction (see You and Zhang [2007]). The length and complexity of the filings make them an ideal place for text mining via NLP.

SENTIMENT AND TONE ANALYSIS

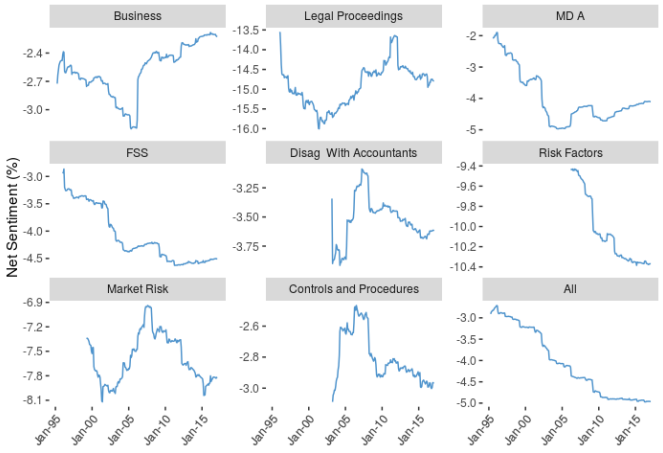
One of the most popular methods to extract information out of natural language is sentiment or tone analysis. It aims to objectively characterize the message conveyed by descriptive information. Then the attention is shifted to examine how the stock market reacts to such quantified qualitative information. Individual words are simply classified based on the sentiment they deliver, positive or negative. The market reaction is then a function of the relative number of positive and negative words in the document, normalized by the total number of words. The foremost work in this direction has been done by Loughran and McDonald [2011], and most recently in our research (see Luo, et al [2017a, 2017b, and 2017c]). Following Loughran and McDonald [2011], we conduct our sentiment analysis on 10-K/10-Q filings. In addition, we also compute more interesting characteristics such as uncertainty, litigious, constraining, superfluous, and interesting.

We combine the words characterized as Positive or Interesting, to measure the overall positive tone of the document. Similarly we also blend words classified as Negative, Uncertainty, Litigious, Constraining or Superfluous to measure overall negative tone. Positive sentiment is the ratio of total number of words classified as Positive divided by the total number of words in a particular section. Similarly we define the Negative sentiment as the ratio of words with Negative tone divided by total number of words.

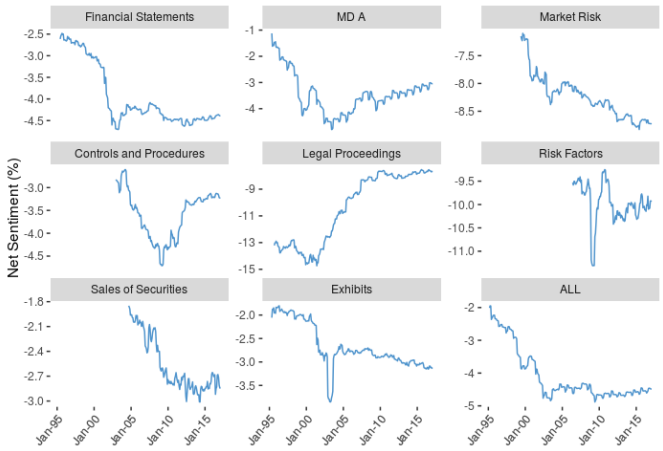
Figure 11 shows the Net tone of the various sections in the 10-K and 10-Q filings, by aggregating all companies in the Russell 3000 index. Net tone is computed as the difference between the positive and negative tone. Since the nature of the filings is more for disclosure and downside protection, there are more negative words in general and that's why the Net tone almost always stays below zero. In addition, the Net tone is trending down over time for most sections and at the aggregate level. The proportion of negative words is less common in the "Business" and "MD&A" sections than in the "Risk Factors" and "Legal Proceedings" parts.

Figure 11 Historical net sentiment for the various sections in the 10-K and 10-Q filings

A) Net sentiment in 10-K filings



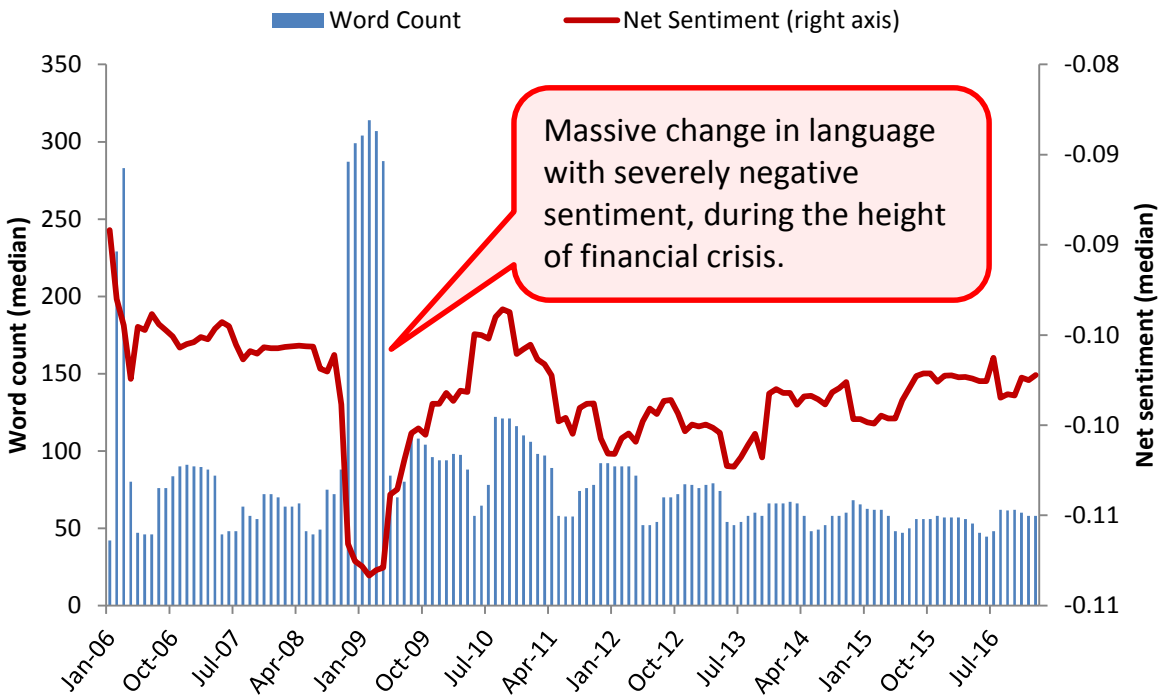
B) Net sentiment in 10-Q filings



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Interestingly, we see a plunge in the Net tone of the "Risk Factors" section during the height of the 2008 Financial Crisis (see Figure 12). This coincides with the introduction of massive number of new textual descriptions in this particular section from October 2008 to April 2009. These new words have predominantly negative tones, expressing the concerns about the recession and its impact on the company financials.

Figure 12 Word count and net sentiment for the "Risk Factors" section of the 10-Q filings



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 13 gives an example of the "Risk Factors" section of 10-Q filings of Avery Dennison Corp, before, during, and after the 2008 Financial Crisis. Interestingly, Avery Dennison Corp is not even from the financial services sector, which had the largest change in language during the crisis. We can see the sharp change in the language for the Quarter ending September 2008, from a year ago. The description quickly reverts back to business as usual in the subsequent year, albeit with more details added.

Figure 13 Risk factor section of the 10-Q filing for Avery Dennison Corp during height of the financial crisis

Quarter ending: 2007-09

Our ability to attain our goals and objectives is materially dependent on numerous factors and risks, including but not limited to matters described in Part I, Item 1A, of the Company's Form 10-Q for the period ended June 30, 2007 and the Form 10-K for the fiscal year ended December 30, 2006.

Quarter ending: 2008-09

Our ability to attain our goals and objectives is materially dependent on numerous factors and risks, including but not limited to matters described in Part I, Item 1A, of the Company's Form 10-K for the fiscal year ended December 29, 2007. Set forth below is an update to such risk factors.

Adverse conditions in the global economy and disruption of financial markets could negatively impact our customers, suppliers, and our business.

Financial markets in the United States, Europe and Asia have experienced extreme disruption in recent months, including, among other things, extreme volatility in security prices, severely diminished liquidity and credit availability, rating downgrades, declines in asset valuations, inflation, reduced consumer spending, and fluctuations in foreign currency exchange rates. While currently these conditions have not impaired our ability to access credit markets and finance our operations, there can be no assurance that there will not be a further deterioration in financial markets in major economies. These economic developments affect our customers and our suppliers and businesses such as ours. In addition, they could have a variety of negative effects such as reduction in revenues, increased costs, lower gross margin percentages, increased allowances for doubtful accounts and/or write-offs of accounts receivable, require recognition of impairments of capitalized assets, including goodwill and other intangibles, and could otherwise have material adverse effects on our business, results of operations, financial condition and cash flows.

We are not able to predict the duration and severity of the current disruption in financial markets and adverse economic conditions in the U.S. and other countries.

Quarter ending: 2009-09

Our ability to attain our goals and objectives is materially dependent on numerous factors and risks, including but not limited to matters described in Part I, Item 1A, of the Company's Form 10-K for the fiscal year ended December 27, 2008. Set forth below is an update to such risk factors.

Proposed changes in U.S. tax legislation could materially impact our results.

Currently, the majority of our revenues is generated from customers located outside of the U.S., and a substantial portion of our assets, including employees, are located outside of the U.S. We have not accrued income taxes and foreign withholding taxes on undistributed earnings for most non-U.S. subsidiaries, because such earnings are intended to be indefinitely reinvested in the operations of those subsidiaries. Certain recently announced proposals could substantially increase our tax expense, which would result in a negative impact on our financial position and results of operations.

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Historical performance of the sentiment factors using 10-K filings

Once we map all the descriptive contents of the filings into a quantitative sentiment score, we can measure the predictive power of our sentiment factor, in a number of ways:

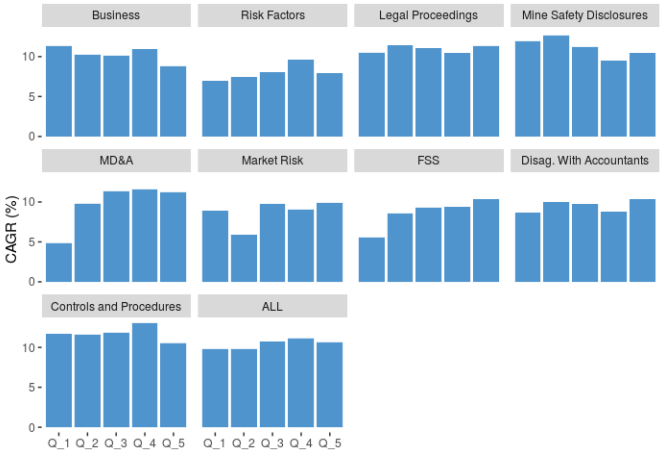
- Positive sentiment: the proportion of positive/interesting words
- Negative sentiment: the proportion of negative words

Figure 14 shows the historical performance of the positive sentiment factor computed on the various sections of the 10-K filings. The positive sentiment factor based on the MD&A and FSS (Financial Statements and Supplemental data) sections is particularly strong at predicting future stock returns.

In addition, contradictory to the results in most sections, a more positive tone in the Business section relates to negative future performance, especially on a risk-adjusted basis. Given the generic and descriptive nature of the Business section, it might reflect a behavioral finance problem where an overly bullish tone without substance is indicative of negative performance.

Figure 14 Historical performance of the positive sentiment indicator, 10-K filings

A) Quintile returns of the sentiment factor



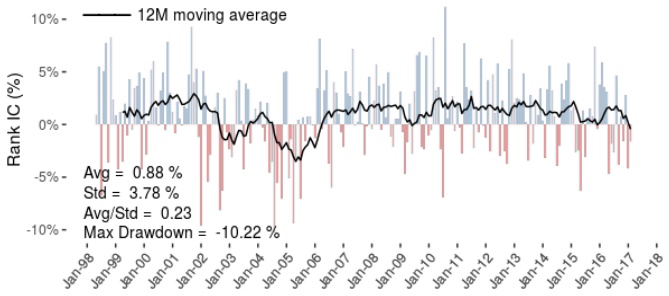
B) Sharpe ratio of the sentiment factor

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

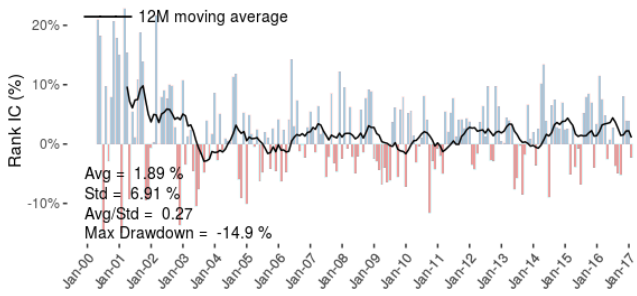
Figure 15 (A) and (B) show the performance (as measured by Rank IC) of our positive sentiment factor based on the MD&A and FSS sections. It is worth noting that the performance has been fairly consistent in recent years, while most conventional factors have shown significant deterioration (see Luo et al [2017b]). Furthermore, given the low frequency nature of the 10-K filings, the positive sentiment factors also demonstrate high signal autocorrelation (i.e., low turnover).

Figure 15 Historical performance of the positive sentiment factor based on the MD&A and FSS sections

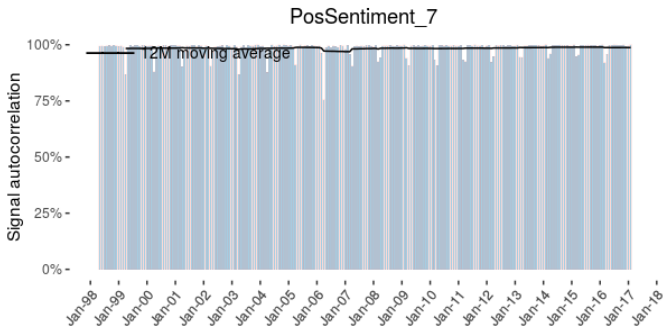
A) The positive sentiment factor (MD&A), rank IC



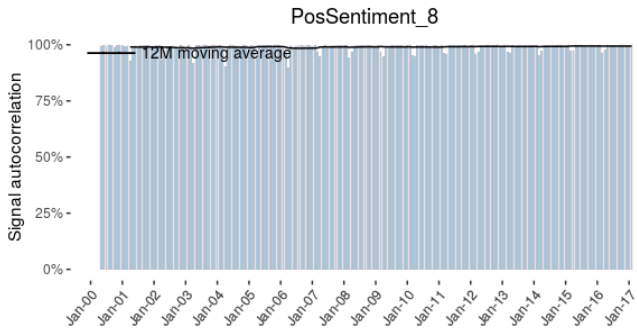
B) The positive sentiment factor (FSS), rank IC



C) The positive sentiment factor (MD&A), signal autocorrelation



D) The positive sentiment factor (FSS), signal autocorrelation



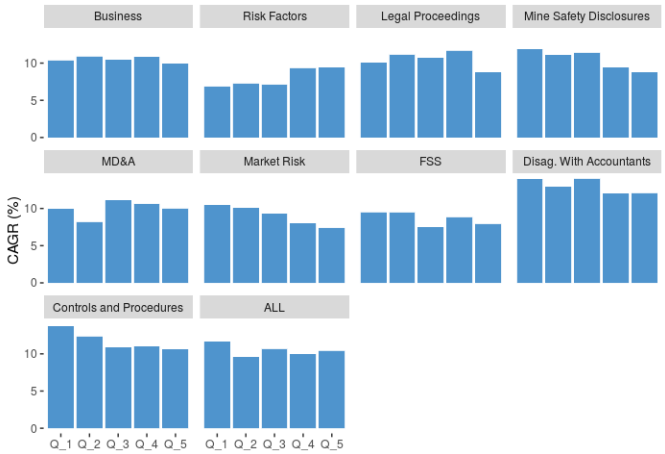
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 16 shows the historical performance of the negative sentiment factor based on various sections of the 10-K filings. Not every section is created equally, but broadly the direction is intuitive, i.e., more negative sentiment leads to lower subsequent stock returns. The most predictive signals are computed on the "Market Risk", "Mine Safety Disclosures" and "Control and Procedures".

Interestingly, for the "Risk Factors" section, a more negative tone is actually positive for future stock returns (see Figure 16). More conservative disclosures in the "Risk Factors" section is perhaps an indicator of management prudence.

Figure 16 Historical performance of the negative sentiment indicator, 10-K filings

A) Quintile returns of negative sentiment factor



B) Quintile Sharpe ratio of negative sentiment factor

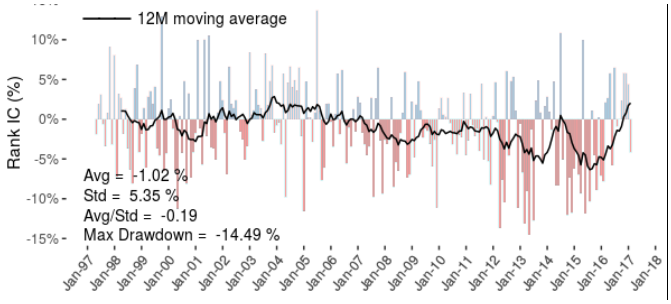
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 17 (A) and (B) show the performance (rank IC) of the negative sentiment factor, computed on the "Mine Safety" and "Controls and Procedure" sections. Again, the performance is modest, but consistent over time, with low turnover (see Figure 17 C and D).

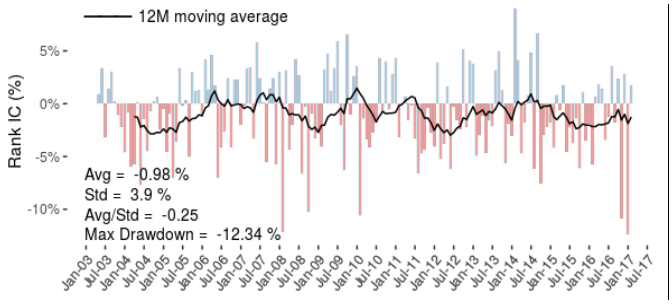
The huge outperformance of the negative sentiment factor (Mine Safety Disclosure) in recent years seems to coincide with the down cycle in the commodities market and the subsequent underperformance of the metals and mining sector.

Figure 17 Historical performance of negative sentiment indicators in 10-K filings

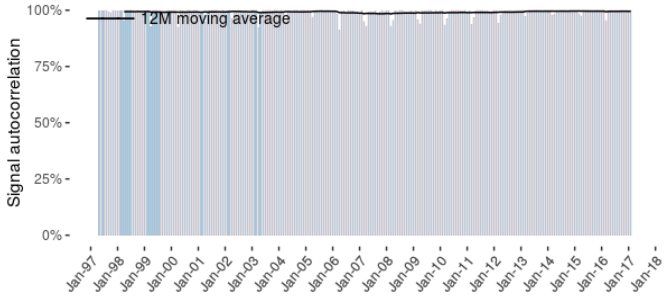
A) The negative sentiment factor (Mine Safety Disclosure), rank IC



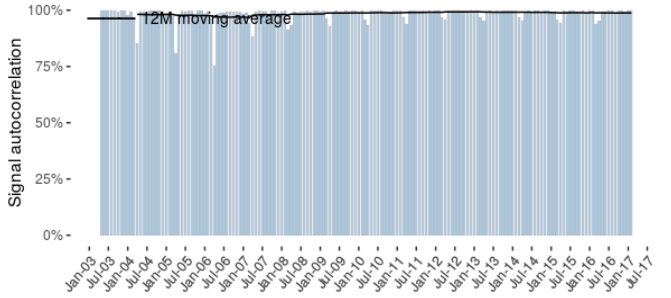
B) The negative sentiment factor (Controls and Procedure), rank IC



C) The negative sentiment factor (Mine Safety Disclosure), signal autocorrelation



D) The negative sentiment factor (Controls and Procedure), signal autocorrelation



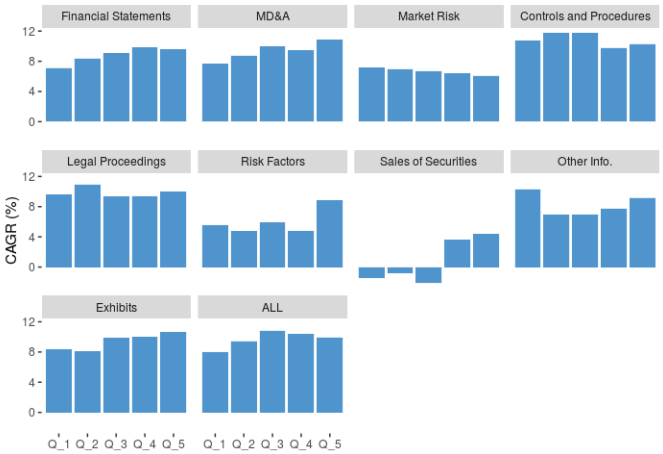
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Historical performance of the sentiment factors using the 10-Q filings

Similar to the 10-K filings, we define the same set of positive and negative sentiment factors using the 10-Q quarterly filings. The results are generally comparable to what we observe in the previous section using 10-K data (see Figure 18).

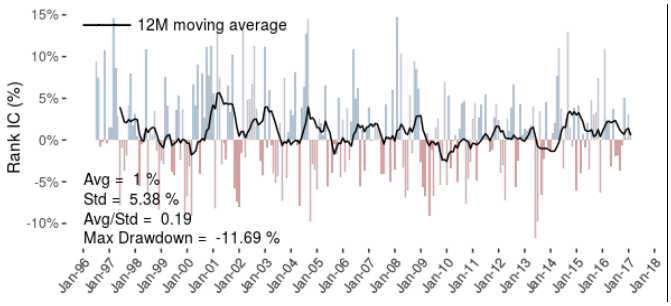
Figure 18 Historical performance of the positive sentiment factor, 10-Q filings

A) Quintile returns of the positive sentiment factor

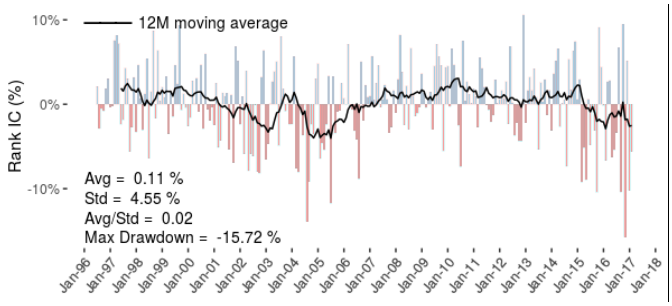


B) Sharpe ratio of the positive sentiment factor

C) The positive sentiment factor (FSS), rank IC



D) The positive sentiment factor (MD&A), rank IC



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 19 shows the quintile portfolio performance of the negative sentiment measures using 10-Q filings. The negative sentiment measures have a more consistent performance profile across sections than the positive sentiment factors.

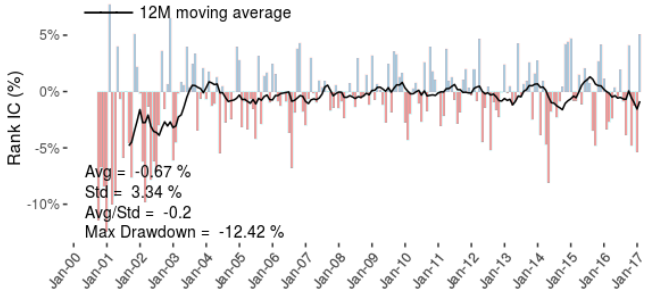
The FSS, MD&A as well as "Market Risk" sections look particularly interesting. It is important to note that the strong performance is dominated by the 2000 technology bubble period (see Figure 19 C and D).

Figure 19 Historical performance of the negative sentiment indicator, 10-Q filings

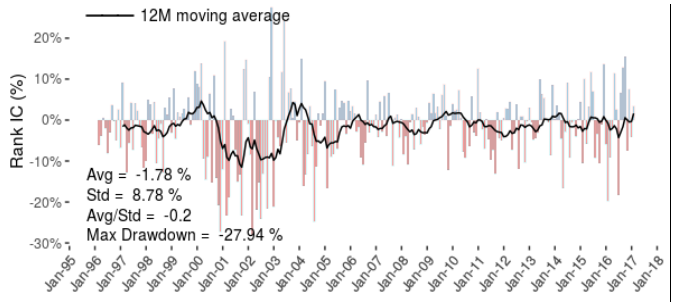
A) Quintile returns of negative sentiment factor

B) Quintile Sharpe ratio of negative sentiment factor

C) The negative sentiment factor (Market Risk), rank IC



D) The negative sentiment factor (Consolidated), rank IC



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

CHANGE IN SENTIMENT

Given the nature that some companies are almost always more conservative than the others, and some sections of the 10-K/10-Q filings have more negative tones than the other parts, maybe the change in sentiment is more relevant than the absolute level. For example, Chouliaras [2015] argues that, although the effect of the 10-K textual sentiment change is insignificant beyond the submission month, the interaction of past momentum with change in sentiment churns out more interesting patterns.

In our analysis, the change in sentiment turns out to be rather weak for the majority of the sections, with the exception of the "Risk Factors". As shown in Figure 20, either an increase or a decrease in sentiment in the "Risk Factors" leads to remarkably negative returns. The payoff pattern is a classic inversed U-shaped curve, the middle quintiles with the least notable change in sentiment (either positive or negative) deliver the highest returns.

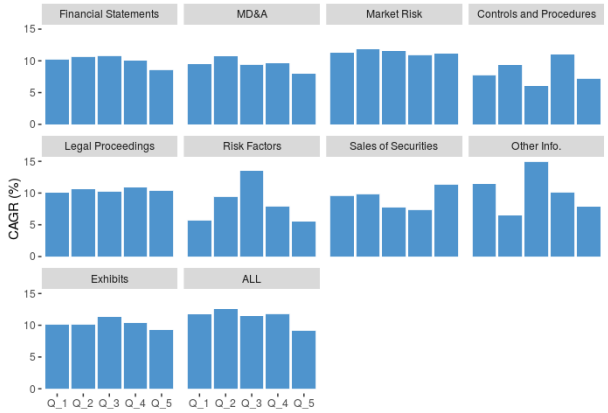
In our view, the change in sentiment factor overlaps with a more elementary factor that simply measures the magnitude of change in each section, regardless of the direction of change. Such a measure might sound overly simplistic at a first glance. However, as we have shown in Figure 20 and

will further elaborate in the next section, this primitive measure is in fact among the most fascinating factor in this paper.

Figure 20 Historical performance of the change in tone measures, 10-Q filings

A) Quintile returns of the change in positive sentiment

B) Quintile returns of the change in negative sentiment



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

DISTANCE MEASURES (YOY CHANGE IN CORPORATE FILING LANGUAGE)

While the bulk of academic and industry studies focus primarily on the sentiment and tone of the corporate filings, Brown and Tucker [2011] is one of the first to examine the year-on-year changes (as opposed to levels) in the text (instead of sentiment) of filings. Brown and Tucker [2011] find that changes in the MD&A section are related to future operating changes in the business (e.g., accounting-based performance, liquidity measures). The authors further indicate that contemporaneous returns around the 10-K filing dates are correlated to the changes in the MD&A section.

Most financial performance numbers are not very informative without a side by side comparison with previous reported data. For instance, an annual revenue figure of $2.5 billion does not tell us much without looking at the prior years' sales. Therefore, traditional financial statement based factors are mostly presented as ratios or percentage changes. Corporate financial statements are also presented with a few periods' data for the ease of interpretation. In contrast, investors do not appear to systematically compare this year's textual description to previous years' filings. The simple comparison may contain rich information about the future.

As argued in Cohen, et al [2010], human beings tend to be "lazy". In a sense that for anything of recurring nature, few people would start from scratch each time. Rather, most would take what they did in the previous years and modify. As a result, the numbers in the document might be quite different, but the structure and language used are similar. For example, firms in general use a well-defined format for their annual and quarterly filings. Almost identical language is simply repeated year after year, until someone actively intervenes and makes changes. When firms do break away from their tradition of textual descriptions in their filings, they typically foresee significant changes in their business, risk, or corporate strategy.

Consistent with the experimental evidence on the importance of active choices, when firms do make an active decision to significantly change the wording and language embedded in their quarterly and annual reports, these active changes have large but subtle implications for future firm behavior (see Cohen, et al [2000]).

To measure these changes in language, we use two elementary document similarity measures, namely Cosine distance and Jaccard distance:

Cosine distance

$$D_1^{TermFreq} = [freqD_1^{t1}, freqD_1^{t2}, \dots\dots, freqD_1^{tn}]$$

$$D_2^{TermFreq} = [freqD_2^{t1}, freqD_2^{t2}, \dots\dots, freqD_2^{tn}]$$

Where D_1 and D_2 are two documents and $t1$ to tn are unions of terms occurring in the two documents. $D_1^{TermFreq}$ and $D_2^{TermFreq}$ are the set of frequencies of terms in the two documents.

$$Cosine\ distance = \frac{D_1^{TermFreq} * D_2^{TermFreq}}{||D_1^{TermFreq}|| * ||D_2^{TermFreq}||}$$

Cosine distance is the scalar product of the two document term frequency vectors divided by the product of their Euclidean norm.

Jaccard distance

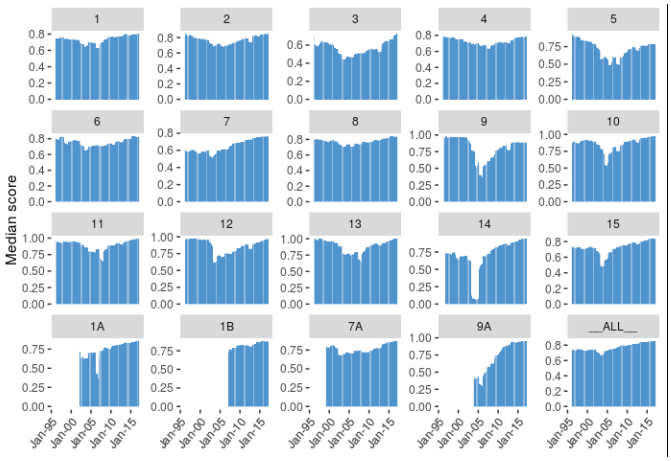
$$Jaccard\ distance = \frac{D_1^{TermFreq} \cap D_2^{TermFreq}}{D_1^{TermFreq} \cup D_2^{TermFreq}}$$

Jaccard distance is simply the intersection of the two document term frequency vectors divided by their unions.

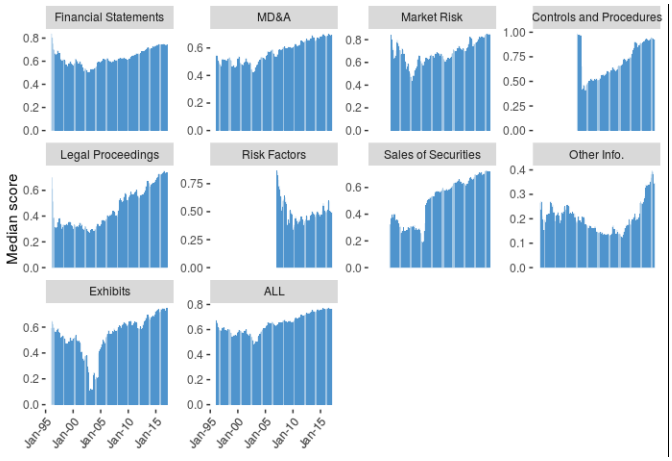
The two measures by definition are highly correlated, with higher values indicating higher similarity between the documents. As shown in Figure 21, the average year-over-year distance measure ranges from 70% to 80% for most sections of 10-K and 10-Q filings. The regulatory filings for the same firm tend to be relatively similar with few changes from one year to another, signifying that the active language change does not occur often. The large dip in similarity measures for most sections around the year 2005 coincides with the significant reporting guidelines changes to the 10-K and 10-Q by the regulator.

Figure 21 Jaccard distance for the companies in the Russell 3000 index

A) 10-K sections



B) 10-Q sections



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

A case study with substantial changes to the 10-Q filings

Using American Express as an example, we can clearly see the impact of substantial modifications to the "Risk Factors" section of the 10-Q filings on the firm's future business, operating, and stock performance. Similar to other companies, Amex provides extensive details on regulatory, business and other significant risks relevant to the company in its Annual 10-K filings. Although the languages used in this section do not change much most of the time, the occasional adjustments reveal important (adverse) information about the company that is unknown to the market. Such changes are often first reported in the quarterly 10-Q filings.

Figure 22 highlights two such occasions in 2014 and then 2015 for the company, along with its stock price performance. The "Risk Factors" section in the 10-Q filings for the period ending June 2014 added new paragraphs discussing the tough market condition for the airline industry, which represents a significant portion of the firm's business. American Express also highlighted the risk due to substantial and increasingly intense competition for partner relationships that could have a material adverse impact on its business. Furthermore, in the 10-Q filing for the quarter ending June 2015, the company discussed the ongoing legal proceedings regarding provisions in the merchant contracts, which could subject the firm to substantial monetary and reputational damages. The market appears to have overlooked the inherited risk in both cases. The sell-off only occurred months after the risk disclosures in the 10-Q documents. This is a classic example of how NLP and machine learning can be more timely and accurate in analyzing textual information.

Figure 22 Changes to Risk factor section of the American Express 10-Q filings and stock performance in recent years

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Historical performance of the distance factor using the 10-K filings

On average, the firms that make the most significant active changes to the language in their filings are associated with the worst future returns, based on most sections of the 10-K filings (see Figure 23 and Figure 24). The performance of the distance based factors is stronger than the sentiment signals discussed in the previous sections. We also find that the changes in the wordings of the Business and MD&A sections are most prominent. Lastly, the performance of both distance measures is largely the same across all sections.

Figure 23 Historical performance of the Jaccard distance factors, 10-K filings

A) Quintile returns of the Jaccard distance factor

B) Sharpe ratio of the Jaccard distance factor

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 24 Historical performance of the Cosine distance factors, 10-K filings

A) Quintile returns of the Cosine distance factor

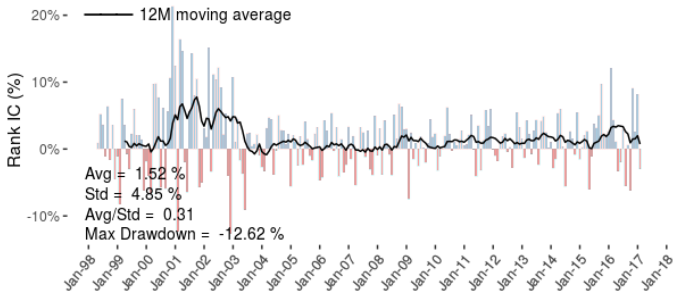
B) Sharpe ratio of the Cosine distance factor

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

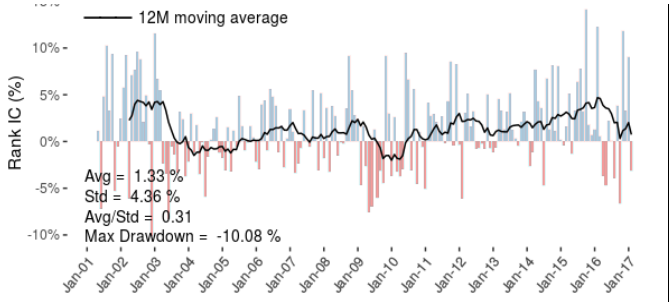
Consistent to what we find from the sentiment factors, the historical performance of the distance measures is strongest around 2000 technology bubble. The performance has also been very strong in the recent few years (see Figure 25).

Figure 25 The Jaccard distance factor performance, rank IC

A) Business section distance factor



B) MD&A section distance factor

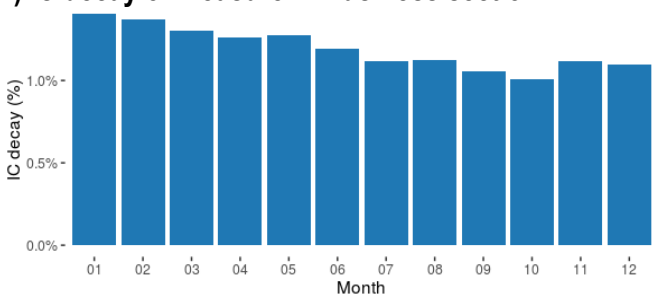


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

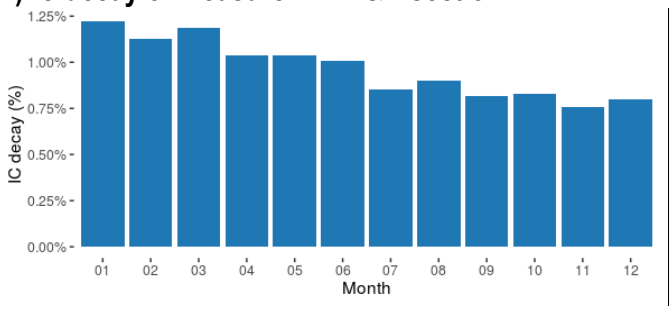
Given the factors are based on the 10-K annual filings, they have a very slow decay (see Figure 26), lasting beyond a year. This implies that the active changes in the languages in the regulatory filings convey very salient information about the company, and these changes get incorporated into asset prices gradually.

Figure 26 IC decay of the Jaccard distance factor, 10-K filings

A) IC decay of measure in Business section



B) IC decay of measure in MD&A section

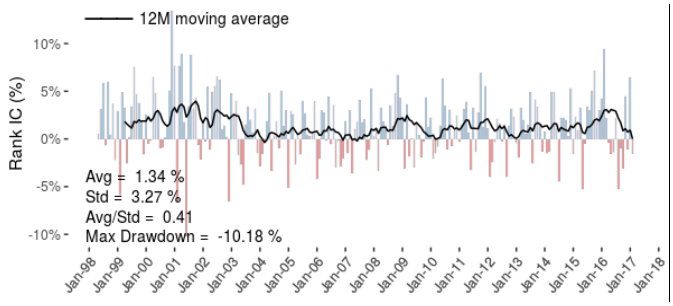


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

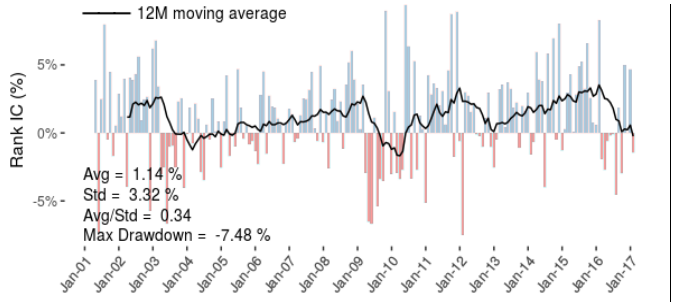
To mitigate the influence of the 2000 technology bubble, we backtest our factors using a sector neutral approach. The performance remains very significant and the risk-adjusted performance is even stronger (see Figure 27).

Figure 27 Sector neutral Jaccard distance factor, rank IC

A) MD&A section distance factor



B) FSS section distance factor



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

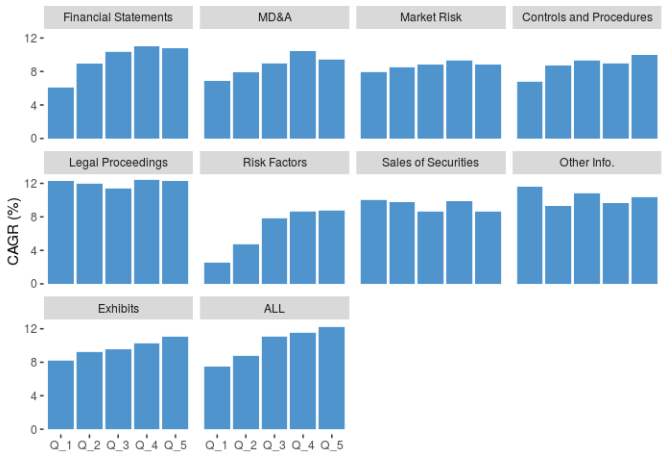
Historical performance of the distance factor using the 10-Q filings

Using a more timely 10-Q filing, we also compute the year-over-year distance measures by aligning the current quarter with the corresponding quarter last year. The performance is particularly strong based on the FSS, MD&A, Controls and Procedures, Risk Factors and the Exhibits sections (see Figure 28 and Figure 29). Computing the distance factor using text from all sectors in the 10-Q also has a strong monotonic performance profile.

The distance factors using the "Risk Factors" section has the strongest performance. Companies with the least changes in the "Risk Factor" section outperform the ones with the most adjustments by 6% per year.

Figure 28 Historical performance of the Jaccard distance factor, 10-Q filings

A) Quintile returns of the Jaccard distance factor

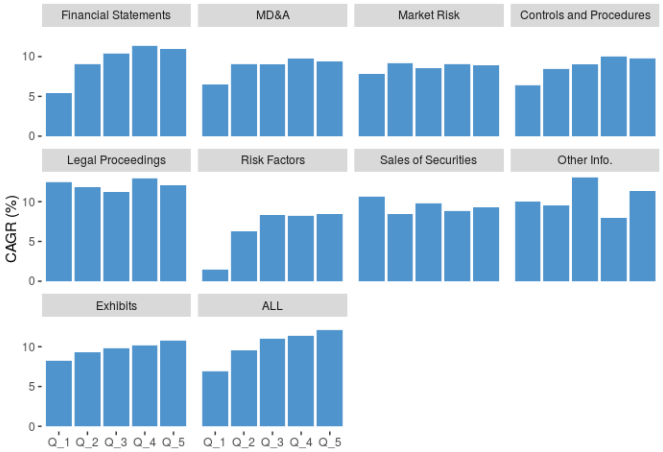


B) Sharpe ratio of the Jaccard distance factor

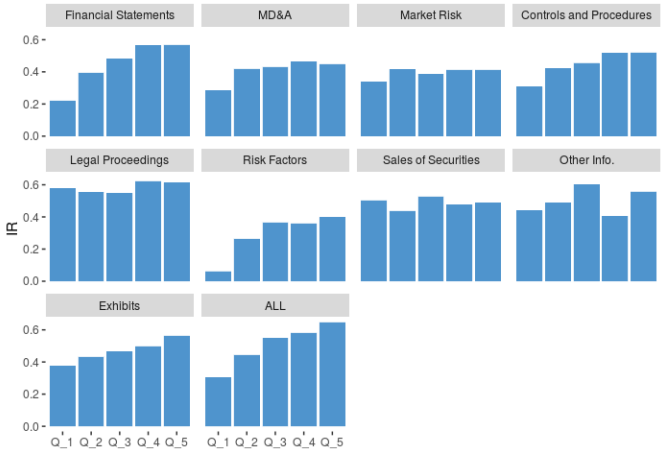
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 29 Historical performance of the Cosine distance factor, 10-Q filings

A) Quintile returns of the Cosine distance factor



B) Sharpe ratio of the Cosiine distance factor



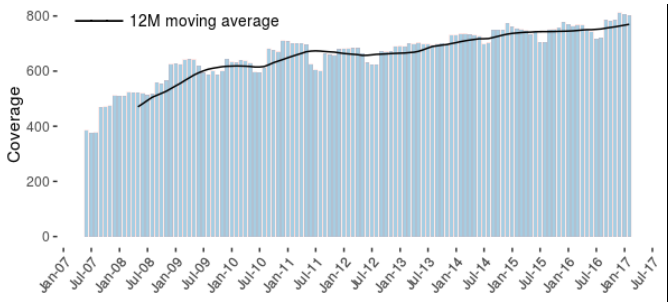
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Although the "Risk Factors" section has the strongest performance, it suffers from low coverage (see Figure 30 A). Reasonable coverage only starts from 2007 and even as of now, we have just around 700 stocks. On the other hand, the measure based on the FSS section has data from 1997 with more than double coverage (see Figure 30 B).

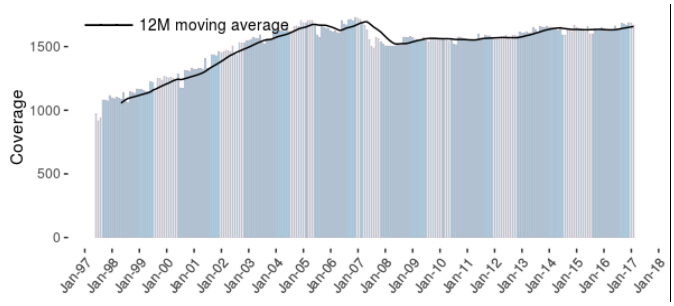
More importantly, unlike most traditional factors, the performance of our distance factors has been particularly strong in recent years (see Figure 30 C and D).

Figure 30 Coverage and performance (rank IC), 10-Q filings

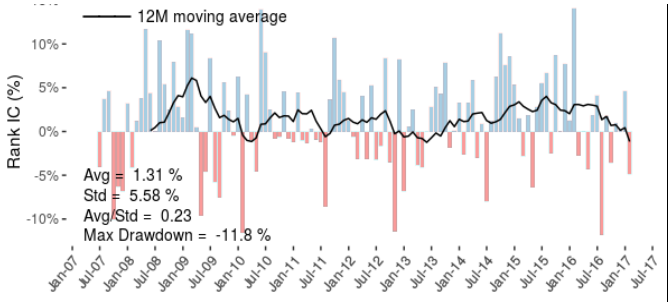
A) Coverage in "Risk Factors" (Jaccard distance)



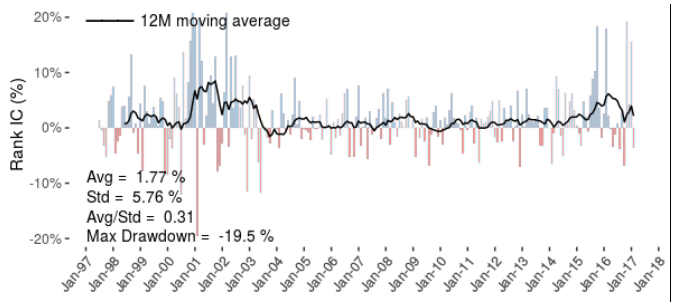
B) Coverage in FSS (Cosine distance)



C) Jaccard distance using "Risk Factors", rank IC



D) Cosine distance using FSS, rank IC

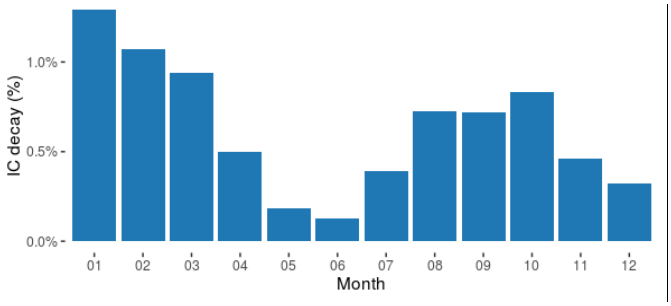


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

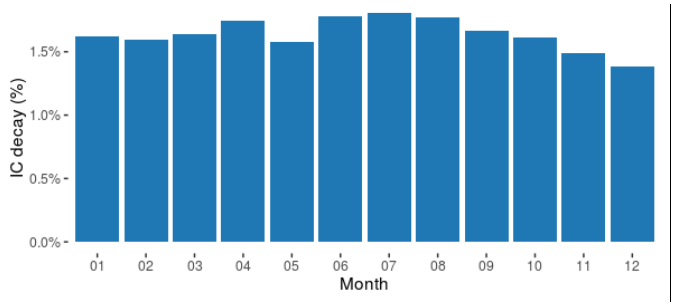
The signal decay is not always uniform across sections in the quarterly filings, Figure 31. As in the case below the signal decays over the quarter for the Risk Factor section while for the financial statement section the decay in signal is very slow even after one year.

Figure 31 IC decay of the distance measures, 10-Q filings

A) Jaccard distance factor "Risk Factors", IC decay



B) Cosine distance factor FSS, IC decay



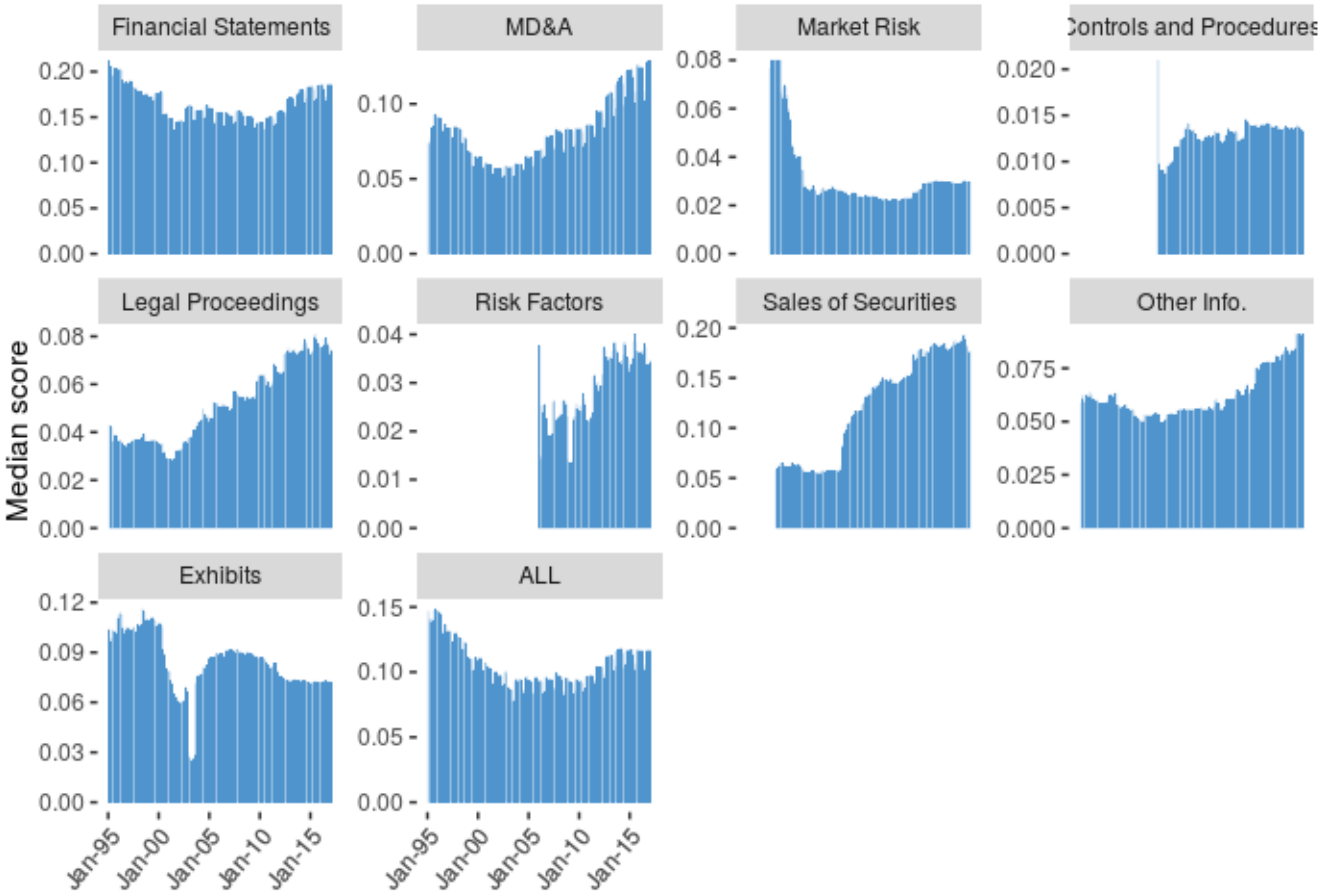
Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

NUMERIC PERCENTAGE MEASURES

In regulatory filings, companies can present their performance and business using either textual description or numerical data. Arguably, when the numbers are weak, firms might be tempted to distract investors' attention with future prospect in writing.

In order to capture this phenomenon, we compute the percentage of numeric data embedded in each section of the filings. We use the ratio of numeric digits to alphabets in a section as our factor (called numeric percentage factor). Figure 32 shows the historical average percentage of numeric data for each section of the 10-Q filing.

Figure 32 Numeric percentage in each section of the 10-Q for the companies in the Russell 3000 index

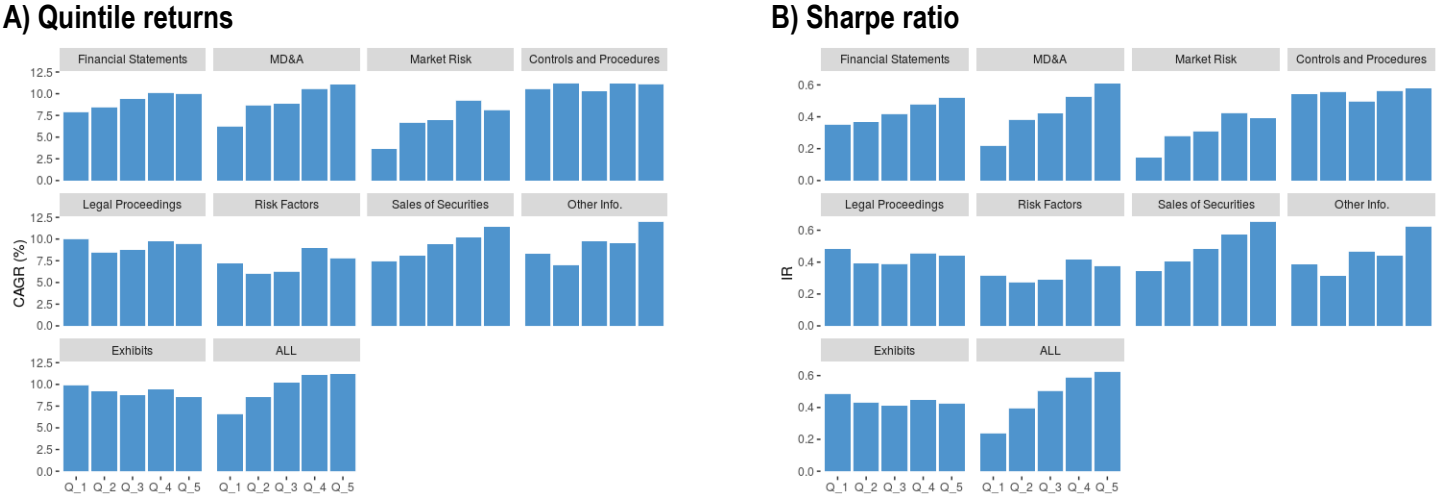


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Historical performance of the numeric percentage measure in 10-Q filings

For most sections, firms with higher percentages of numeric contents have higher subsequent returns, which confirms our hypothesis. This is particularly true for the MD&A, Sales of Securities and Market Risk sections (see Figure 33).

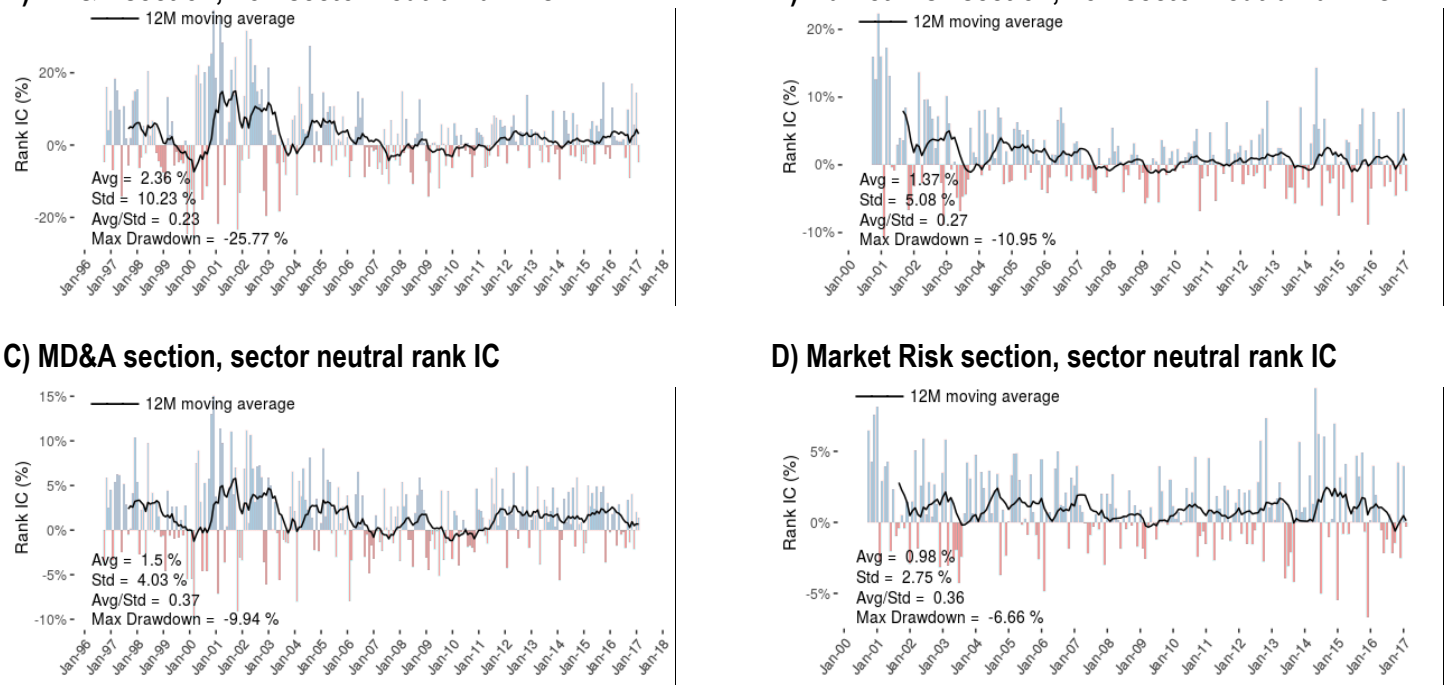
Figure 33 Historical performance of the numeric percentage factor, 10-Q filings



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

As shown in Figure 34, once again we observe a strong performance around the 2000 tech bubble, but the strong performance also persists in recent years.

Figure 34 Performance of the numeric percentage factor, 10-Q filings, rank IC



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

EDGAR PROFILING COMPOSITE

In this section, we discuss various ways to combining the NLP based stock selection factors presented in the previous sections together. We also elaborate a few potential approaches to further improve the performance of our signals using machine learning and language complexity measures.

SYSTEMATIC WEIGHTING OF TEXTUAL CONTENT

Nearly all of the textual factors discussed in this paper and in previous studies are computed as ratios. For example, for the sentiment or tone measure, we use the ratio of negative (or positive) words to the total number of words in a document. Such a method implicitly assumes that all words within a category are equally informative, which is highly unrealistic.

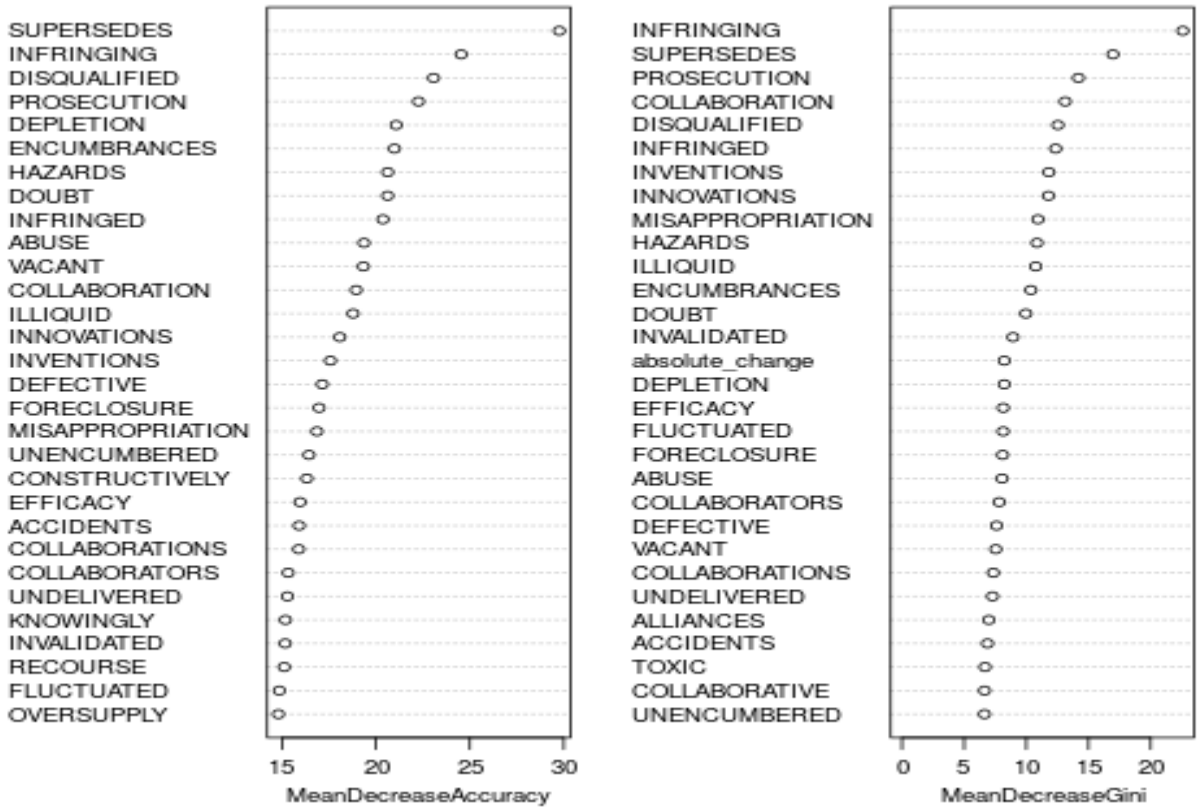
Alternatively, words can be manually categorized and weighted. However, it is not only extremely time consuming, but also suffers from subjective biases.

We can also use models to quantify each word. For example, Jegadeesh and Wu [2013] examine the market reaction to the 10-K's containing specific words, and use this to develop a weighting scheme.

Lastly, nonlinear machine learning algorithms can be applied for automatic feature selection. A properly designed machine learning algorithm, in general tends to be more robust to dimensionality (i.e., large number of factors) and multicollinearity (see Luo, et al [2017c] and Jussa, et al [2017]). We develop a random forest algorithm to assign weights to the words appearing in the 10-K filings. We use the post 10-K filing stock return as the dependent variable, and the proportion of each individual word in the filing as the independent variable. The 10-K filings contain more than 30,000 unique words in total and therefore we need to find a way to reduce the dimensionality. We only consider those words that convey some kind of a sentiment based on the Lexicon provided by Loughran and McDonald [2011]. We also remove the most frequently used words (as these may be trivial) as well as the rarely used words in the filings.

Figure 35 shows a list of words with the most informative contents based on our algorithm. As we can see, it is dominated by words with severe negative tones, e.g., infringing, prosecution, hazards, disqualified, depletion, foreclosure, abuse, accidents, defective, etc. There are few words with distinctively positive tone such as collaboration, inventions and innovations. While the list of words looks promising, the performance is not much better than the un-weighted version.

Figure 35 The most informative terms in the 10-K filings for the year ending 2016

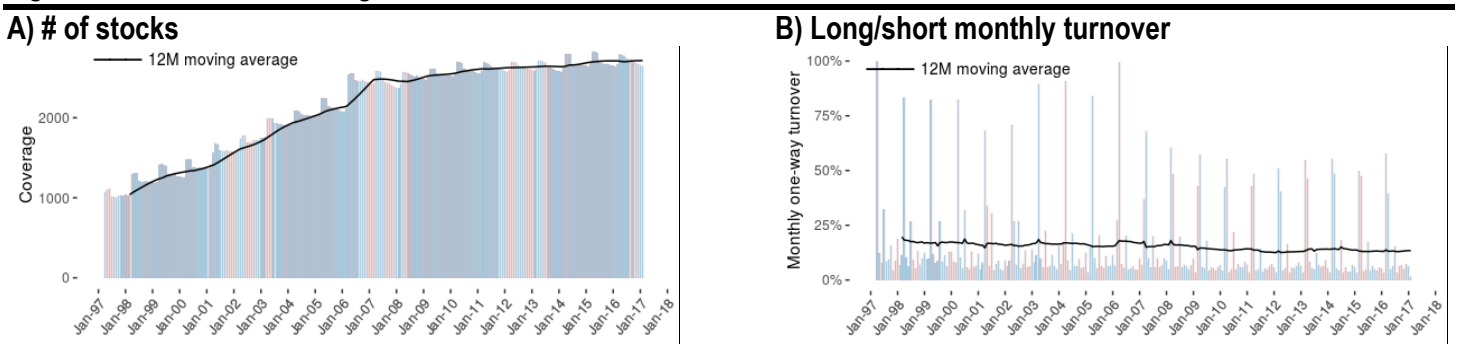


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

EQUAL WEIGHTED COMPOSITE FACTOR BASED ON THE 10-K FILINGS (EQ10K)

Next we explore various ways to combine the factors discussed in the paper so far. The base case combines all the prominent factors equally. Figure 36 to Figure 39 shows the performance of the base case composite factor using 10-K filings – EQ10K. Coverage of the factor is decent, with around 2,700 stocks out of the Russell 3000 in recent years (see Figure 36 A). The average monthly long/short portfolio turnover is less than 20%. Given the annual frequency of the data, the portfolio needs to be rebalanced at least once a year around the annual filing months of February and March (see Figure 36 B).

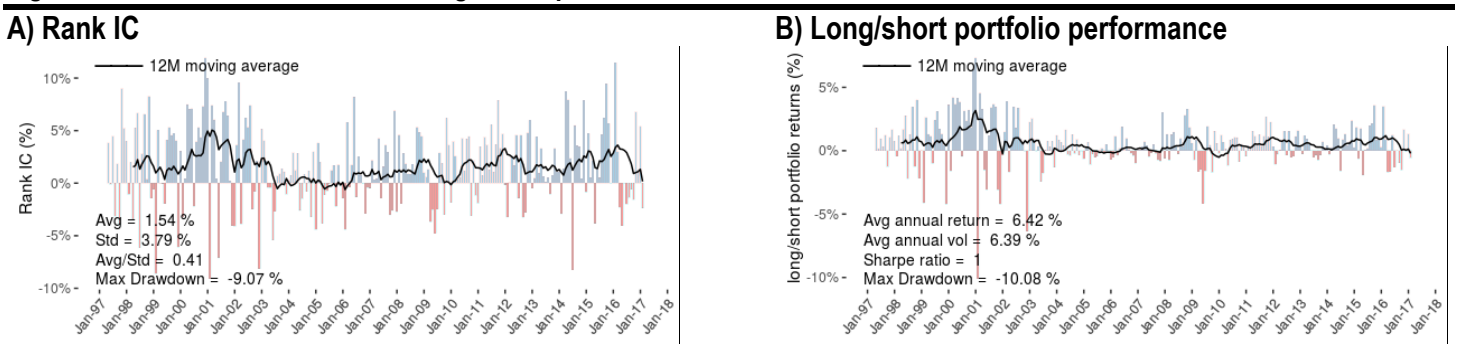
Figure 36 EQ10K factor coverage and turnover



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

A long/short sector neutral quintile portfolio based on the EQ10K factor generates an annual alpha of 6.5% (see Figure 37 B). The performance has been quite consistent in recent years.

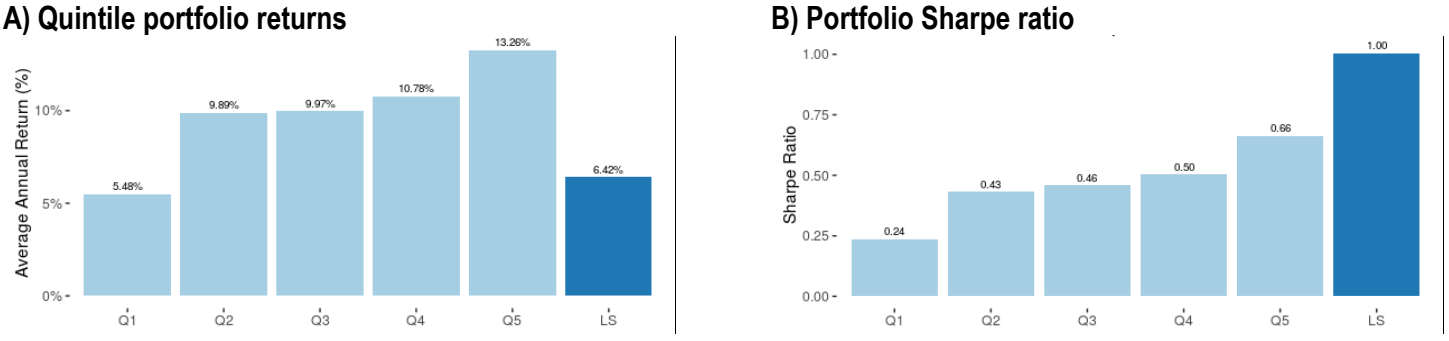
Figure 37 EQ10K factor rank IC and long/short spread



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

The portfolio performance is very monotonic from Quintiles One to Five. A long/short quintile portfolio generates a Sharpe ratio of 1.0x (see Figure 38).

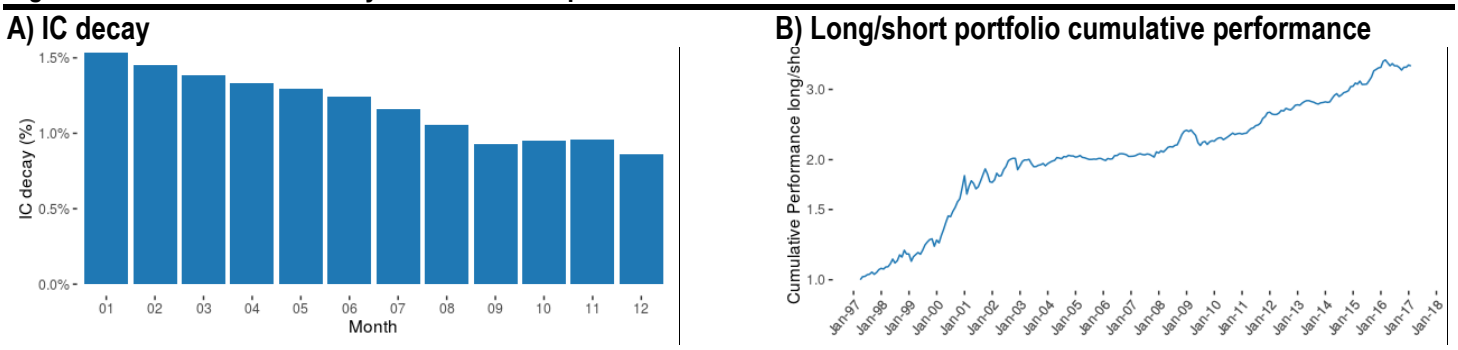
Figure 38 EQ10K factor quintile return and Sharpe ratio



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Lastly, the equally weighted composite signal also has slow decay. Performance lasts beyond a year (see Figure 39).

Figure 39 EQ10K factor IC decay and cumulative performance

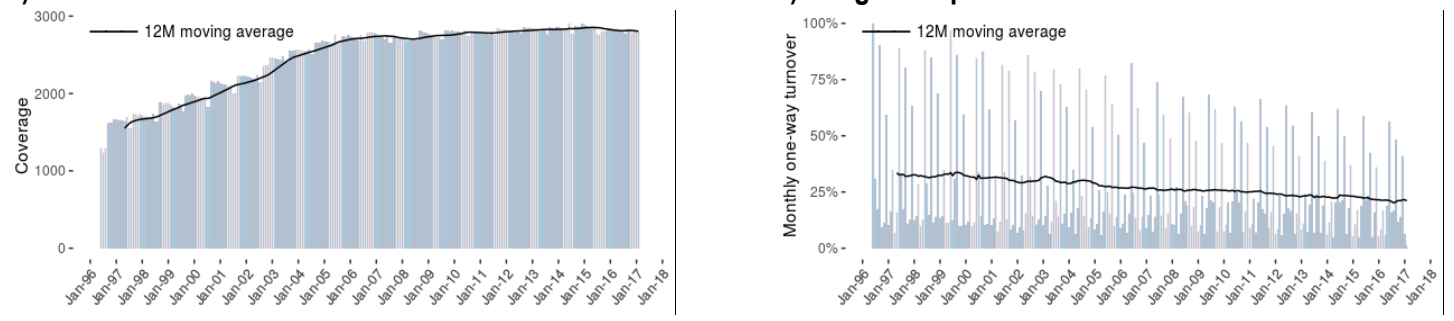


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

EQUAL WEIGHTED COMPOSITE FACTOR BASED ON THE 10-Q FILINGS (EQ10Q)

Similar to EQ10K, we create another equal weighted composite of the prominent signals from 10-Q filings – EQ10Q. Figure 40 to Figure 43 show the performance of the EQ10Q factor on a sector neutral basis. Coverage is around 90% in recent years in the Russell 3000 universe (see Figure 40 A). As expected, the turnover of the EQ10Q factor is higher than EQ10K, but still well below most traditional factors. It aligns with the quarterly reporting frequencies (see Figure 40 B). In the past 20 years, the turnover has been trending down slightly, as company filings are becoming more and more standardized over time.

Figure 40 EQ10Q factor coverage and turnover

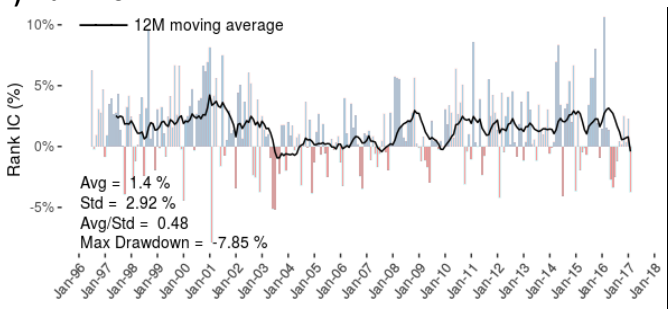


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

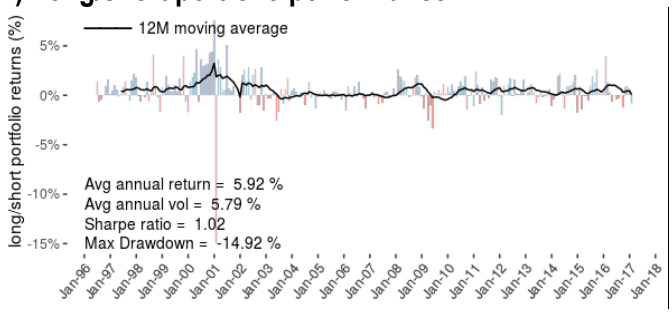
The long/short return of the EQ10Q factor is fairly strong (see Figure 41 A and B), especially post 2007. One of the most prominent factors in EQ10Q is the distance measure based on the "Risk Factors" section, where data only becomes widely available from 2007.

Figure 41 EQ10Q factor rank IC and long/short spread

A) Rank IC



B) Long/short portfolio performance

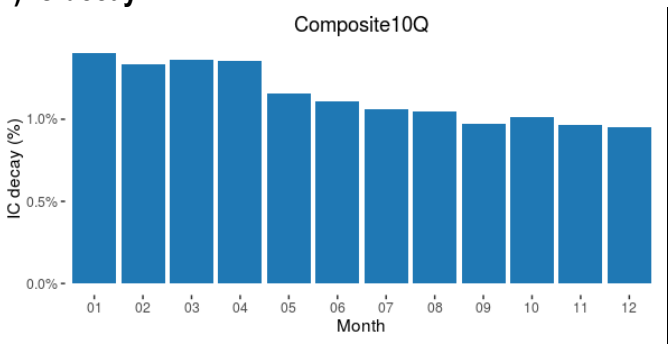


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

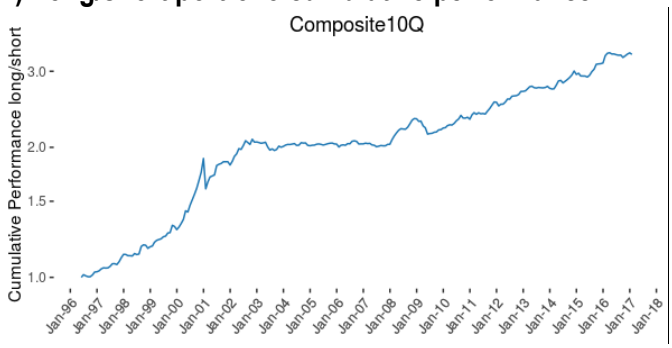
Lastly, the EQ10Q signal also has a very slow decay profile, with forecasting power lasting beyond a year (see Figure 42 A).

Figure 42 EQ10Q factor IC decay and cumulative performance

A) IC decay



B) Long/short portfolio cumulative performance

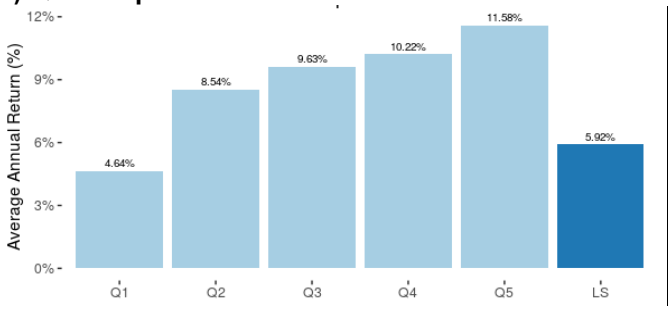


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

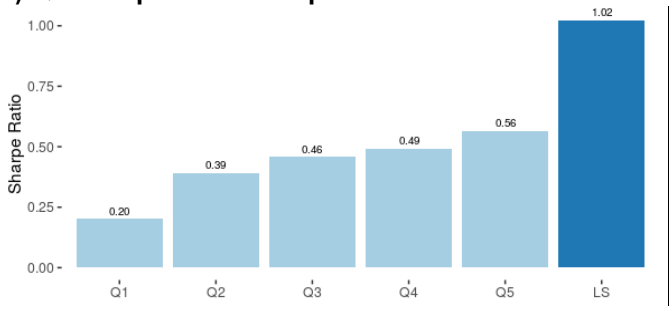
The EQ10Q signal delivers an annual alpha of almost 6% (see Figure 43 A) and a Sharpe ratio of 1.0x (see Figure 43 B).

Figure 43 EQ10Q factor quintile portfolio return and Sharpe ratio

A) Quintile portfolio returns



B) Quintile portfolio Sharpe ratio



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

COMPLEXITY MEASURE

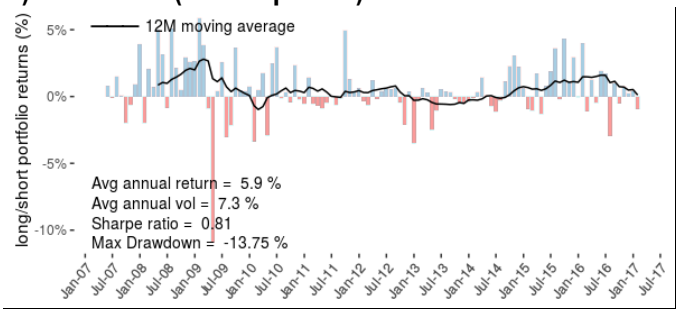
Readability or complexity of the filings can hinder manual information processing. Hence systematic processing of textual contents from these more complex filings can be more predictive for future returns, due to the "limits to arbitrage" behavioral finance argument.

Investors underreact to the information in the 10-K filings, when that information is more complex, measured using simple word counts (see You and Zhang [2009]). Retail investors' trading behavior is affected by the clarity of financial disclosures in the 10-K (see Lawrence [2013]). The size of the filing statements is a good enough measure of readability (see Loughran and McDonald [2014]).

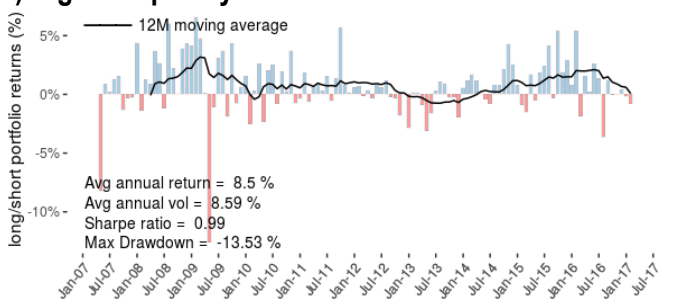
Based on the textual contents, we divide our universe in to two groups – High complexity and Low complexity. We simply remove the bottom quintile of stocks based on a complexity measured as the number of words in each section of the filing from our investment universe. The remaining 80% stocks are defined as High complexity universe. Figure 44 shows the improvement in performance of the distance factor using the "Risk Factors" section of 10-Q filings. The Long/short portfolio return jumps from 5.9% for the base case to 8.5% for the high complexity measure – an increase of 44%. Overall, focusing on the companies with more complex filings does boosts the performance for the distance measures based on the 10-Q filing, but not much for the 10-K filings.

Figure 44 The Cosine distance factor based on the "Risk Factor" section, 10-Q filings

A) Base case (all companies)



B) High complexity universe



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

SYSTEMATIC PROFILING EDGAR COMPOSITE (SPEC) MODEL

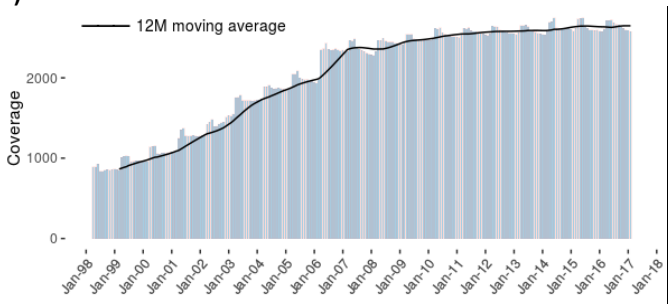
Next we create a systematic profiling EDGAR composite (SPEC) model by combining the two composites based on the 10-Q and 10-K filings, i.e., EQ10Q and EQ10K.

We could have combined the EQ10Q and EQ10K based on the timing of annual/quarterly fillings, e.g., using the EQ10K model for first quarter and EQ10Q model for the other three quarters. However, given the slow decay nature of both EQ10Q and EQ10K, we take a simpler approach of equal weighting the two models.

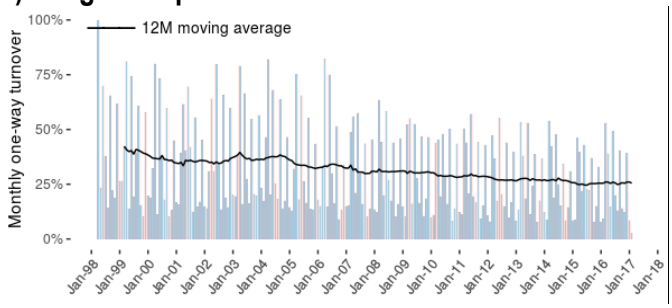
The coverage of the SPEC model is around 2,600 stocks in recent years for the Russell 3000 index (see Figure 45 A), with monthly portfolio turnover of 25% (see Figure 45 B).

Figure 45 Systematic profiling EDGAR composite model (sector neutral)

A) Number of stocks



B) Long/short portfolio turnover

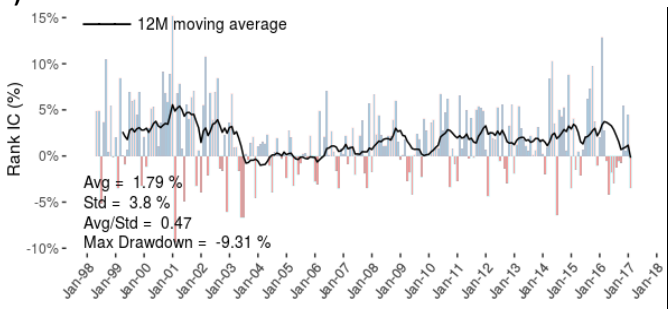


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

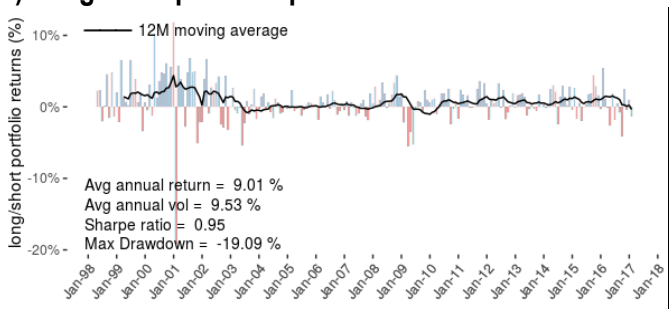
The performance of the SPEC model has been fairly consistent over the past 20 years, especially post 2007 (see Figure 46 A and B).

Figure 46 Systematic profiling EDGAR composite model (sector neutral)

A) Rank IC



B) Long/short portfolio performance

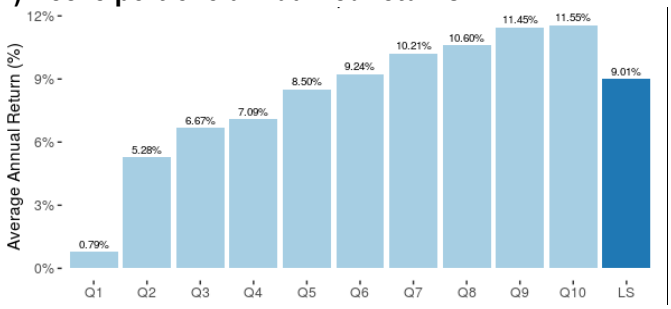


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

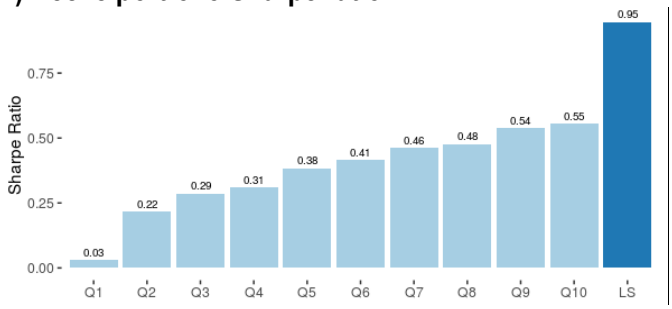
The long/short decile portfolio delivers an annualized returns of 9% (see Figure 47 A), with a Sharpe ratio of 1.0x (see Figure 47 B).

Figure 47 Systematic profiling EDGAR composite model (sector neutral)

A) Decile portfolio annualized returns



B) Decile portfolio Sharpe ratio

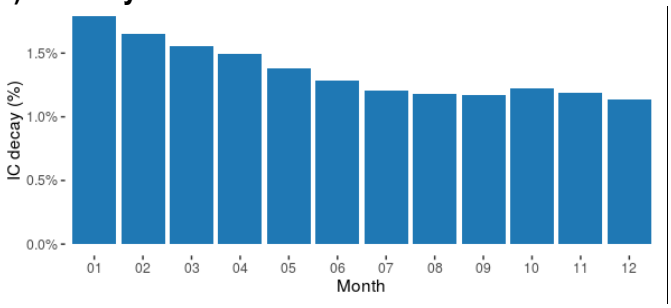


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

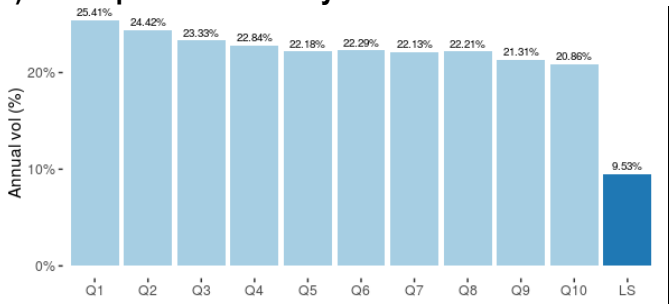
The SPEC model has a slow IC decay profile, with forecasting power lasting beyond a year (see Figure 48 A).

Figure 48 Systematic profiling EDGAR composite model (sector neutral)

A) IC decay



B) Decile portfolio volatility

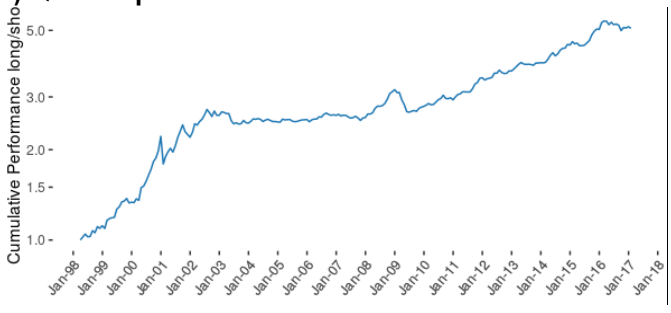


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

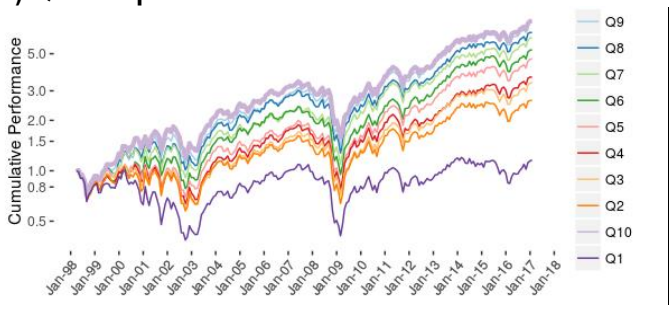
Lastly, as shown in Figure 49, the performance has been reasonably consistently over the past 20 years.

Figure 49 SPEC model with complexity constraint (sector neutral)

A) Quintile portfolio annualized returns



B) Quintile portfolio wealth curve

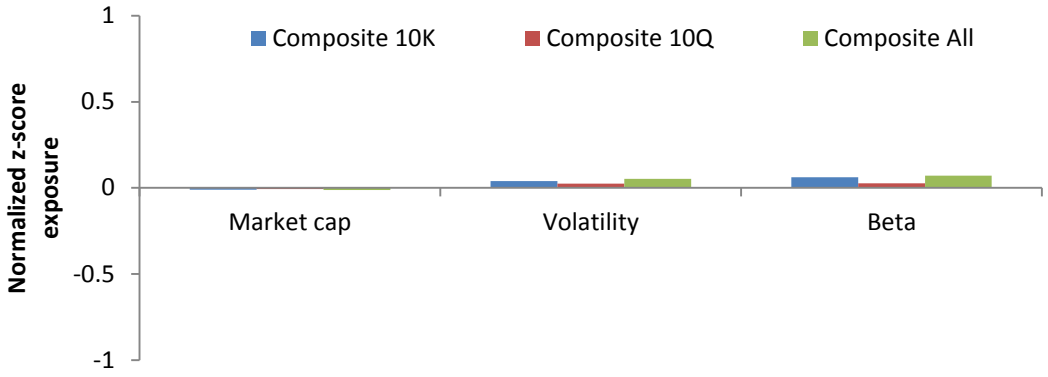


Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

INTERACTION WITH TRADITIONAL FACTORS

In the last section we look at the interaction of our text based model with conventional factors. The universe for our composite models starts from 1997 and the coverage is almost the complete Russell 3000 index in recent years. It is interesting to note that our factor universe has almost no exposure to classic risk factors such as size, beta or volatility (see Figure 50).

Figure 50 Factor exposure of composite universe relative to the Russell 3000 index



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

FACTOR CORRELATIONS

One of the most important reasons of relying on alternative Big Data sources rests on the diversification benefit, in that they are expected to have minimal correlation with traditional factors. A low correlation with traditional factors means a unique source of alpha that is not simply a linear combination of traditional factors.

As expected, our signals based on EDGAR text mining are almost uncorrelated to any of our traditional factors (see Figure 51), with only slight positive tilts towards value, quality, and low volatility factors.

Figure 51 Factor correlations with EDGAR composite models

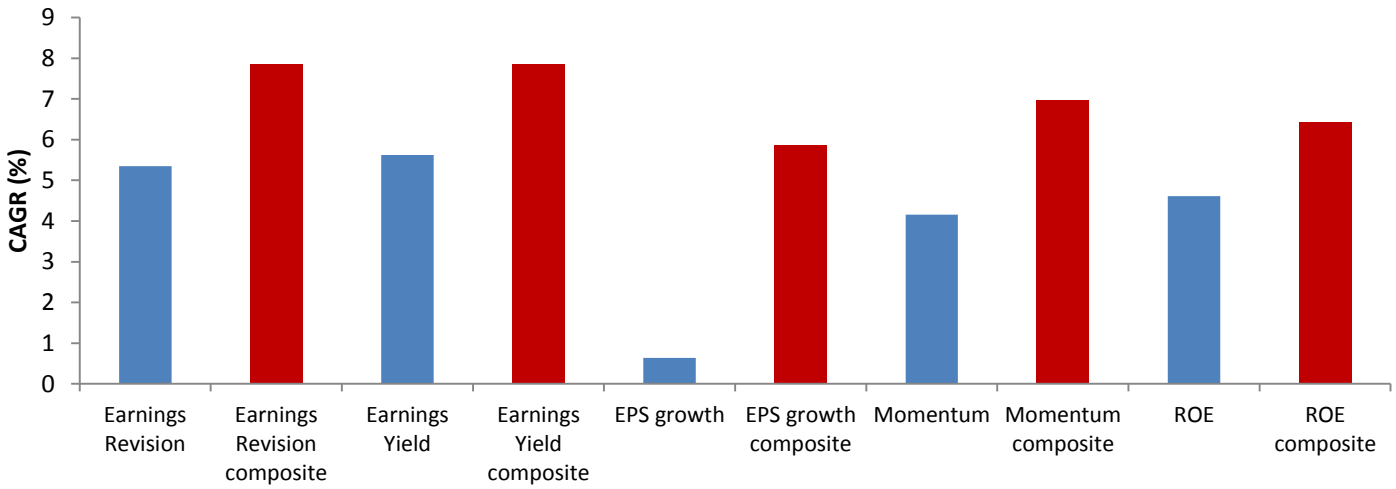
	Earnings Yield	Earnings Revision	Momentum	1M Reversal	ROE	EPS growth	Market cap	Volatility
Composite 10K	11%	6%	4%	1%	10%	-3%	11%	-14%
Composite 10Q	8%	3%	3%	1%	7%	-2%	8%	-10%
Composite All	11%	5%	5%	2%	11%	-2%	11%	-13%

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

COMPLEMENTING TRADITIONAL ALPHA FACTORS

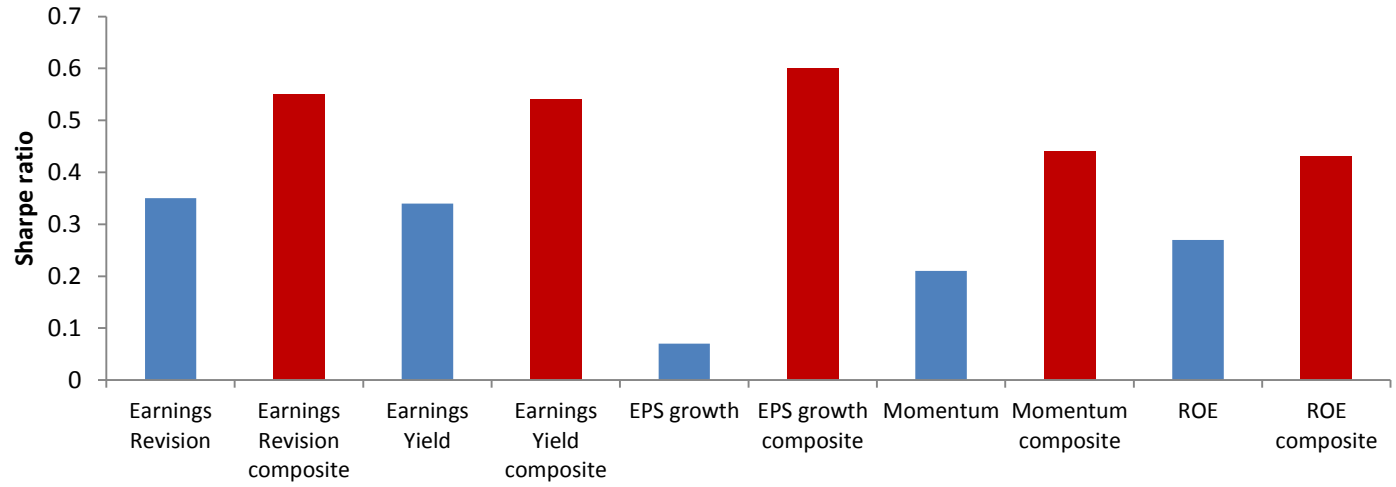
Finally, as a unique alpha source, the SPEC model should complement and add value to traditional factors. As sown in Figure 52, the performance of all traditional factors improves remarkably, when the SPEC is added. Growth factors witness the largest improvement, with a jump of annual alpha from 1% to 6%. Improvement in risk-adjusted performance is even more significant. The Sharpe ratio improves by 50% for the majority of the conventional factors (see Figure 53).

Figure 52 Annualized returns of Style composites vs base factors



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 53 Sharpe ratio of Style composites vs base factors



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

BIBLIOGRAPHY

Ahern, K., and Sosyura, D. [2014]. "Who Writes the News? Corporate Press Releases during Merger Negotiations", Journal of Finance, http://onlinelibrary.wiley.com/doi/10.1111/jofi.12109/abstract

Amel-Zadeh, A., and Faasse, J. [2016]. "The Information Content of 10-K Narratives: Comparing MD&A and Footnote Disclosures", SSRN, https://ssrn.com/abstract=2807546

Antweiler, W., and Frank, M. [2004]. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards", The Journal of Finance, https://www.jstor.org/stable/3694736?seq=1#page_scan_tab_contents

Ball, C., Hoberg, G., and Maksimovic, V., [2014]. "Disclosure, business change and earnings quality", SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2260371

Bodnaruk, A., Loughran, T., and McDonald, B. [2014]. "Using 10-K Text to Gauge Financial Constraints", SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2331544

Boudoukh, J., Feldman, R., Kogan, S., and Bodnaruk, M. [2013]. "Which News Moves Stock Prices? A Textual Analysis", The National Bureau of Economic Research, http://www.nber.org/papers/w18725

Brown, S., and Tucker, J. [2011]. "Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications", Journal of Accounting Research, http://onlinelibrary.wiley.com/doi/10.1111/j.1475-679X.2010.00396.x/abstract

Chen, H., De, P., Hu, Y., and Hwang, B. [2013]. "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media", SSRN, https://academic.oup.com/rfs/article-abstract/27/5/1367/1581938/Wisdom-of-Crowds-The-Value-of-Stock-Opinions

Chouliaras, A., and Grammatikos, T. [2015]. "News Flow, Web Attention and Extreme Returns in the European Financial Crisis", SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2348189

Chouliaras, A. [2015]. "The Pessimism Factor: SEC EDGAR Form 10-K Textual Analysis and Stock Returns", SSRN, http://ssrn.com/abstract=2627037

Cohen, L., Malloy, C., and Nguyen, Q. [2010]. "Lazy Prices", SSRN, http://ssrn.com/abstract=1658471

Davis, A., and Tama-Sweet, I. [2011]. "Managers' Use of Language Across Alternative Disclosure Outlets: Earnings Press Releases versus MD&A", Contemporary accounting research, http://onlinelibrary.wiley.com/doi/10.1111/j.1911-3846.2011.01125.x/abstract

Feldman, R., Govindaraj, S., Livnat, J., and Segel, B. [2010]. "Management's tone change, post earnings announcement drift and accruals", Review of Accounting Studies, 15(4), 915-953

Garcia, D., and Norli, O. [2013]. "Crawling EDGAR", The Spanish Review of Financial Economics (SRFE), http://leeds-faculty.colorado.edu/garcia/paper_edgar_v07.pdf

Hering, J. [2016]. "The Annual Report Algorithm: Retrieval of Financial Statements and Extraction of Textual Information", Academy and Industry research collaboration center, http://airccj.org/CSCP/vol7/csit76615.pdf

Hoberg, G., and Maksimovic, V. [2014]. "Redefining Financial Constraints: A Text-Based Analysis", Review of Financial studies, https://academic.oup.com/rfs/article-abstract/28/5/1312/1867105/Redefining-Financial-Constraints-A-Text-Based

Huang, A., Lehavy, R., Zang, A. Y., and Zheng, R., [2014]. "A thematic analysis of analyst information discovery and information interpretation roles", SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2409482&rec=1&srcabs=2665128&alg=1&pos=10

Israelsen, R. [2014]. "Tell It like It Is: Disclosed Risks and Factor Portfolios", SSRN, http://ssrn.com/abstract=2504522

Jegadeesh, N., and Wu, D. [2013]. "Word Power: A New Approach for Content Analysis", SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1787273

Jussa, J., Luo, Y., and Wang, S. [2017]. "QUANT CSI: Company Profiling Using Alternative Data", Wolfe Research Luo's QES, March 13, 2017

Lawrence, A. [2013]. "Individual investors and financial disclosure", Journal of Accounting and Economics, http://www.sciencedirect.com/science/article/pii/S0165410113000359

Li, F. [2010]. "The Information Content of Forward-Looking Statements in Corporate Filings—A Naïve Bayesian Machine Learning Approach, http://onlinelibrary.wiley.com/doi/10.1111/j.1475-679X.2010.00382.x/abstract

Loughran, T., and McDonald, B. [2011]. "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks", SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1331573

Loughran, T., and McDonald, B. [2013]. "IPO First-Day Returns, Offer Price Revisions, Volatility, and Form S-1 Language", SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2128766

Luo, Y., Jussa, J., and Wang, S. [2016]. The Journal of Quantitative, Economics, and Strategy, Wolfe Research Luo's QES, December 12, 2016

Luo, Y., Jussa, J., and Wang, S. [2017a]. "The Big and The Small Sides of Big Data", Wolfe Research Luo's QES, February 8, 2017

Luo, Y., Jussa, J., and Wang, S. [2017b]. "Signal Research and Multifactor Models", Wolfe Research Luo's QES, February 16, 2017

Luo, Y., Jussa, J., and Wang, S. [2017c]. "Style Rotation, Machine Learning, and The Quantum LEAP" Wolfe Research Luo's QES, February 24, 2017

Mayew, W., Parsons, C., and Venkatachalam, M. [2013]. "Voice pitch and the labor market success of male chief executive officers", Journal of the Human behavior and Evolution Society, http://www.sciencedirect.com/science/article/pii/S1090513813000238

Purda, L., and Skillicorn, D. [2012]. "Accounting Variables, Deception, And A Bag of Words: Assessing the tools of Fraud Detection", SSRN, http://ssrn.com/abstract=1670832

Tetlock, P. [2007]. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market", The Journal of Finance, http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2007.01232.x/abstract

Tetlock, P., Saar-Tsechansky, M and Macskassy, S. [2008]. "More Than Words: Quantifying Language to Measure Firms' Fundamentals", The Journal of Finance, https://www0.gsb.columbia.edu/faculty/ptetlock/papers/Tetlock_et_al_JF_08_More_Than_Words.pdf

You, H., and Zhang, X. [2007]. "Financial reporting complexity and investor under reaction to 10-K information", SSRN, http://ssrn.com/abstract=985365

APPENDIX

Figure 54 World clouds of Business and MD&A sections in 10-K filings

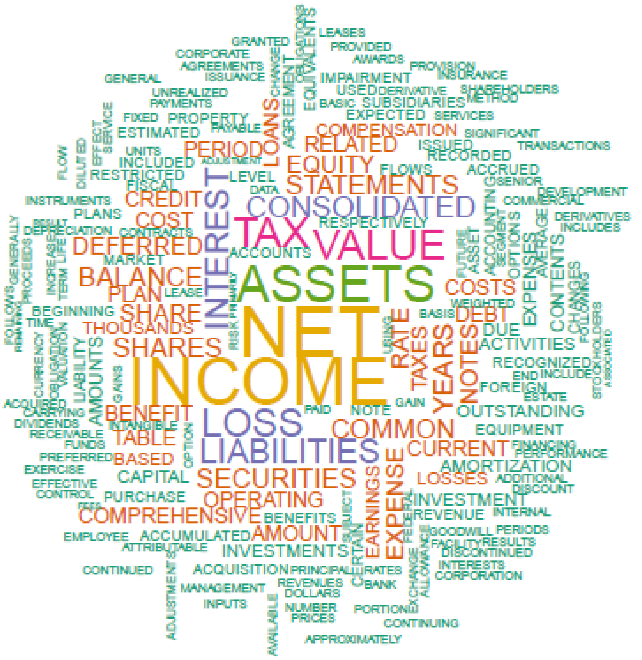
A) Business section

B) MD&A

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 55 World clouds of Financial Statement and Control and Procedures sections in 10-K filings

A) Financial Statement



B) Control and Procedures

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 56 World clouds of MD&A and Legal Proceedings sections in 10-Q filings

A) MD&A

B) Legal Proceedings

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 57 World cloud of Financial Statement and Control and Procedures sections in 10-Q filings

A) Financial Statement

B) Control and Procedures

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

Figure 58 World clouds of Market Risk and Risk Factors sections in 10-Q filings

A) Market Risk

B) Risk Factors

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, EDGAR, Wolfe Research Luo's QES

DISCLOSURE SECTION

Analyst Certification:
The analyst of Wolfe Research primarily responsible for this research report whose name appears first on the front page of this research report hereby certifies that (i) the recommendations and opinions expressed in this research report accurately reflect the research analysts' personal views about the subject securities or issuers and (ii) no part of the research analysts' compensation was, is or will be directly or indirectly related to the specific recommendations or views contained in this report.

Other Disclosures:
Wolfe Research, LLC does not assign ratings of Buy, Hold or Sell to the stocks it covers. Outperform, Peer Perform and Underperform are not the respective equivalents of Buy, Hold and Sell but represent relative weightings as defined above. To satisfy regulatory requirements, Outperform has been designated to correspond with Buy, Peer Perform has been designated to correspond with Hold and Underperform has been designated to correspond with Sell.

Wolfe Research Securities and Wolfe Research, LLC have adopted the use of Wolfe Research as brand names. Wolfe Research Securities, a member of FINRA (www.finra.org) is the broker-dealer affiliate of Wolfe Research, LLC and is responsible for the contents of this material. Any analysts publishing these reports are dually employed by Wolfe Research, LLC and Wolfe Research Securities.

The content of this report is to be used solely for informational purposes and should not be regarded as an offer, or a solicitation of an offer, to buy or sell a security, financial instrument or service discussed herein. Opinions in this communication constitute the current judgment of the author as of the date and time of this report and are subject to change without notice. Information herein is believed to be reliable but Wolfe Research and its affiliates, including but not limited to Wolfe Research Securities, makes no representation that it is complete or accurate. The information provided in this communication is not designed to replace a recipient's own decision-making processes for assessing a proposed transaction or investment involving a financial instrument discussed herein. Recipients are encouraged to seek financial advice from their financial advisor regarding the appropriateness of investing in a security or financial instrument referred to in this report and should understand that statements regarding the future performance of the financial instruments or the securities referenced herein may not be realized. Past performance is not indicative of future results. This report is not intended for distribution to, or use by, any person or entity in any location where such distribution or use would be contrary to applicable law, or which would subject Wolfe Research, LLC or any affiliate to any registration requirement within such location. For additional important disclosures, please see www.wolferesearch.com/disclosures.

The views expressed in Wolfe Research, LLC research reports with regards to sectors and/or specific companies may from time to time be inconsistent with the views implied by inclusion of those sectors and companies in other Wolfe Research, LLC analysts' research reports and modeling screens. Wolfe Research communicates with clients across a variety of mediums of the clients' choosing including emails, voice blasts and electronic publication to our proprietary website.