ESSAYS ON EMPIRICAL ASSET PRICING

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Prasad Hegde

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Sugato Chakravarty, Chair

Department of Consumer Science

Dr. Richard Feinberg

Department of Consumer Science

Dr. Jonathan Bauchet

Department of Consumer Science

Dr. Jiong Sun

Department of Consumer Science

APPROVED BY

Dr. Jonathan Day

To my parents, who have always been there for me. Thank you for all of the love, support, encouragement and dedication.

To my wife, who has been a constant source of support and encouragement. Thank you for all of your love, help, trust and dedication.

This is a tribute to the three of you.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my dissertation committee members, Dr. Sugato Chakravarty, Dr. Richard Feinberg, Dr. Jonathan Bauchet and Dr. Jiong Sun for their support of this work. I am deeply indebted to my advisor, Dr.Sugato Chakravarty, for his unwavering guidance throughout my PhD. This work would not have been possible without his unrelenting support and nurturing. I wish to thank Yang Xie, for his help in the development of this work. I had a great pleasure of working with Dr.Brian Hunter, Saikat Ghosh Roy and Ryan Umbeck during my PhD. I am grateful to my friends and family for their support and encouragement. I would also like to thank the entire faculty, staff and graduate students at the division of Consumer Science at Purdue University for their support.

TABLE OF CONTENTS

Page

LIST OF TABLES vii

ABSTRACT viii

INTRODUCTION x

1 THE POWER OF THE WRITTEN WORD: DOCUMENT TONALITY AS A PRICED RISK FACTOR 1

1.1 Introduction 1

1.2 Background literature on textual analysis 6

1.3 Constructing the *TONE* factor 10

1.4 Results 13

1.4.1 *Summary Statistics* 13

1.4.2 *Portfolio regressions based on one- and two-level sorts* 14

1.4.3 *Factor spanning regressions on six factors* 17

1.4.4 *Individual stock-level regressions* 21

1.4.5 *Directly confronting the factor zoo problem* 22

1.5 Weighted *TONE* factor 23

1.6 Conclusion 25

2 PRICE DISCOVERY IN CORPORATE BOND MARKETS 35

2.1 Introduction 35

2.2 Background literature and hypotheses 38

2.3 Data description and sample construction 43

2.3.1 *Price discovery through information share approach* 44

2.4 Results 47

2.4.1 *Estimating information share* 47

2.4.2 *Estimating the lead-lag return relationship* 49

Page

2.5 Robustness – Component Share approach 52
2.6 Conclusion 54
REFERENCES 58

LIST OF TABLES

Table — Page

1.1 Descriptive statistics . . . 27

1.2 Regressions for decile portfolios formed on *pessimism* . . . 28

1.3 Regressions for decile portfolios formed on *pessimism* . . . 29

1.4 Regression results for double sorts on *pessimism*, *size*,*bm* and *pr* . . . 30

1.5 Regression results for triple sorts on *size*, *bm* and *pr* . . . 32

1.6 Stock-level regressions of *TONE* on *EX_RET* . . . 33

1.7 Alternative definition of *TONE* . . . 34

2.1 Descriptive statistics of information share . . . 55

2.2 Lead-lag relationship between stock and bond returns . . . 56

2.3 Robustness : Component Share . . . 57

ABSTRACT

Hegde, Prasad Ph.D., Purdue University, December 2019. ESSAYS ON EMPIRICAL ASSET PRICING. Major Professor: Sugato Chakravarty.

This dissertation involves two topics in investments. In Chapter 1 we establish that a quantitative measure capturing the tonality (*TONE*) of a publicly traded company's annual 10-K documents is indeed a priced risk factor employing a large panel and multiple econometric approaches. Using both the conventional portfolio level, as well as the more recent individual stock level, regressions, we show that *TONE* captures risk associated with a company that is unique from those captured by the other established risk factors like: *HML*, $R_m - R_f$, *SMB*, *MOM*, *RMW* and *CMA*. We also establish that *TONE* is distinct from the recently identified plethora of factors (the "factor zoo" problem). We find that the coefficient associated with *TONE* is significantly negatively (positively) correlated with excess stock returns for the most (least) pessimistic stock portfolios. Further, *TONE* is negatively correlated with returns for portfolios with the smallest stocks and those with the smallest book-to-market ratios.

In Chapter 2, we investigate the contribution of bond markets to price discovery in equity markets, using Hasbrouck (1995) information share approach. Based on corporate bond and equity transaction for S&P 500 stocks over a sample period between 2002 and 2014 we find that the monthly average information share (IS) of corporate bond markets is about 19%. We also find that the information share of high-yield bonds is significantly greater than investment grade bonds. We also examine the relative informational efficiency in corporate bonds through the lead-lag relationship between their daily returns. Our results demonstrate empirically the

theoretical arguments that informed investors trade in both stock and corporate bond markets, suggesting an important informational role for corporate bonds.

INTRODUCTION

The dissertation involves two topics in asset markets. The first topic examines the role of a the qualitative information in a textual document in explaining the cross-section of stock returns. To do so, in Chapter 1 we quantify the textual tonality of 10-K filings and show that such tonality has significant explanatory power over and above the other well-known risk factors.

The second topic, investigates two important questions related to the informational advantage of US corporate bonds. We investigate the contribution of corporate bonds in price discovery. We also investigate a related question whether the corporate bonds have informational advantage.

Since the publication of the influential Fama and French (1993) paper documenting three specific factors that arguably explain a significant cross section of excess stocks returns, a large body of research has developed around the usage of these factors along with three other more recently documented factors (six in total) as control variables for various asset pricing questions.[1] However, these factors are all related to quantifiable information captured from firms balance sheets and/or income statements. The qualitative aspects of a firm like potential information conveyed through firms written documents have been largely ignored until recently. Tetlock, Saar-Tsechansky and Macskassy (2008), for example, report that the fraction of negative words in firm-specific news stories forecasts low firm earnings and that firms'

[1] For a small sampling, see Asness, Frazzini and Pedersen (2014); Lewellen (2011); Tetlock, Saar-Tsechansky and Macskassy (2008); George, Hwang, and Li (2018); Ma, Wang, and Zhang (2017); Asness, Frazzini, Israel, Moskowitz, and Pedersen (2018); Ranganathan, and Singh (2018); Barillas, and Shanken, (2017, 2018); Stambaugh, and Yuan (2016) and Feng; Giglio and Xiu (2019); Asness and Frazzini (2013); and Novy-Marx (2013).

stock return predictability from negative words is largest for stories that focus on firm fundamentals. They go on to argue that linguistic media content captures otherwise hard to quantify aspects of firms' fundamentals which then map into stock returns.[2] Loughran and McDonald (LM, 2011) build on previous research by underscoring the fact that word lists developed for other disciplines misclassify tonality (especially negative tonality) when applied to financial texts. Accordingly, they develop an alternative negative word list, along with five other word lists, that better reflect tone in financial texts.

We demonstrate that an appropriately constructed tonality factor significantly explains the cross section of excess (and abnormal) stock returns after controlling for the original Fama and French (1992) factors given by: the *HML*, $R_m - R_f$ and *SMB*; as well as: *MOM* (the momentum factor, Carhart, 1997); and two additional factors introduced by Fama and French (2015): *RMW* and *CMA*.

We construct the factor model using the afore-mentioned six risk factors along with *TONE*. To do so, we follow a similar approach as used in Carhart (1997) in forming decile stock portfolios based on the previous year's fraction of negative words. Thus, for example, Decile 1 (Decile 10) would comprise of stocks with the most (least) pessimism. We then estimate a (current year) time-series regression for each of these 10 deciles, where the dependent variable is the monthly portfolio excess return while the independent variables are the six and *TONE* based on current year values. We find that the coefficient associated with *TONE* is significantly negatively (positively) correlated with excess stock returns for the most (least) pessimistic stock portfolio. We also use the more recent methodology of employing individual stock-level regressions and confirm that *TONE* indeed plays a positive and significant role in explaining excess returns. Overall, we are able to show that qualitative information improves the model's explanatory power over and above the quantifiable factors that are commonly

[2]See, for example, Admati and Pfleiderer (2001), Antweiler and Frank (2004, 2005), Das and Chen (2007), Tetlock (2007), Tetlock, Saar-Tsechansky and Macskassy (2008), Li (2008, 2010).

used in explain stock returns.

We also create portfolios based on the fundamental measures of a firm such as size, book-to-market ratio, past performance as well as the fraction of negative words, a firm-specific measure used by Tetlock, Saar-Tsechansky and Macskassy (2008) and Loughran and McDonald (2016). Through the various sorts, we show that TONE has a positive (negative) and significant correlation with returns for portfolio consisting of small (large) stocks. Therefore it appears that management in large firms are more likely to be cautious in their discussions (resulting in relatively less optimism) because larger firms with an eye to reducing potential lawsuits are likely to present a more pessimistic picture in their 10-Ks. Also, *TONE* has a negative and significant correlation with returns for portfolios of high book-to-market stocks. A reasonable interpretation is that value stocks that provide an optimistic view in their 10-K filings and are more likely to portray a rosy outlook relative to those associated with growth stocks. Finally, we show that *TONE* has negative and significant correlation with returns for most pessimistic as well as low past performance portfolios. This could mean stocks that witness a higher past performance coupled with relatively less pessimism in the 10-Ks tend to perform well in the future.

In Chapter 2 we investigate the contribution of corporate bond market's price discovery using the Hasbrouck's (1995) information share approach. Our research question is motivated by the strand of literature that provides evidence of significant informed trading in US corporate bond markets prior to earnings announcements (Wei and Zhou, 2012), prior to acquisition announcement (Kedia and Zhou, 2014). Bodnaruk and Rossi (2013) document that institution holding both equity and debt securities of a given firm benefit significantly from the price appreciation during the M&A events.

Another objective of this study is to investigate the lead-lag correlation between stocks and bonds of a given firm. Since, stocks and bonds account for a dominant share in all traded financial assets, understanding the correlation between equity and corporate debt securities play an important role in investors' diversification, risk management and asset allocation decisions. Although, conventional wisdom and academic studies [Fleming, Kirby and Ostdiek (2003); Gulko (2002); Li (2002) and Hartmann, Straetmans and Devries (2001)] indicate a negative correlation between stock returns and long-term treasuries, there is substantial variation in the relationship between stock and bond returns over the short term. Not surprisingly, given the macro nature of these studies we cannot extrapolate the firm-specific correlations between stocks and bonds.

We extend prior work by examining whether the variation in stock-bond return relation is prevalent in a specific type of firm. Our motivation follows from literature on cross-market hedging by Fleming, Kirby and Ostdiek (1998) and Chordia, Sarkar and Subrahmanyam (2001), and stock market uncertainty (Connolly, Stivers and Sun, 2005, Vironesi, 1999).

Although, the evidence provided by existing studies on aggregate levels of correlations provide insight into the determination of the stock-bond correlations. As mentioned earlier the previous studies do not analyze at the micro level (i.e. return correlation between stock and bonds across same firms). This approach may very well explain the broad market behavior, but understanding the specific type of firm(s) susceptible to such variations in correlations can provide insights for investors.

By employing a large dataset of corporate bonds traded during the sample period 2002 through 2014, we find that, US corporate bonds have an average information share of 19%. We also find that the information share of high-yield bonds is significantly greater than investment grade bonds. It appears that the corporate bond

markets have a significant role in price discovery. Additionally, through our daily vector auto-regression (VAR) estimates we find that corporate bonds have a significant informational advantage for investment grade bonds where we find that bond return leads stock returns. Finally, through a unique dataset that covers institutional trading in bonds and stocks of the same firm we find that corporate bonds have informational advantage for high-yield bonds.

1. THE POWER OF THE WRITTEN WORD: DOCUMENT TONALITY AS A PRICED RISK FACTOR

1.1 Introduction

Since the publication of the influential Fama and French (1993) paper documenting three specific factors that arguably explain a significant cross section of excess stocks returns, a large body of research has mushroomed on the usage of these factors along with three other more recently documented factors (six in total) as control variables for various asset pricing questions.[1] An important caveat is that these factors are all related to quantifiable information captured mostly from firms' balance sheets and/or income statements. The qualitative aspects of a firm, like potential information conveyed through firms' written documents have, until recently, been largely ignored. Tetlock, Saar-Tsechansky and Macskassy (2008) report that the fraction of negative words in firm-specific news stories forecasts low firm earnings and that firms' stock return predictability from negative words is largest for stories that focus on firm fundamentals. They go on to argue that linguistic media content captures otherwise hard to quantify aspects of firms' fundamentals which then map into stock returns.[2] Loughran and McDonald (LM, 2011) build on previous research by underscoring the fact that word lists developed for other disciplines misclassify tonality (especially negative tonality) when applied to financial texts. Accordingly, they develop an alternative negative word list, along with five other word lists, that better

[1]For a small sampling, see Asness, Frazzini and Pedersen (2014); Lewellen (2011); Tetlock, Saar-Tsechansky and Macskassy (2008); George, Hwang, and Li (2018); Ma, Wang, and Zhang (2017); Asness, Frazzini, Israel, Moskowitz, and Pedersen (2018); Ranganathan, and Singh (2018); Barillas, and Shanken, (2017, 2018); Stambaugh, and Yuan (2016) and Feng; Giglio and Xiu (2019); Asness and Frazzini (2013); and Novy-Marx (2013).

[2]See, for example, Admati and Pfleiderer (2001), Antweiler and Frank (2004, 2005), Das and Chen (2007), Tetlock (2007), Tetlock, Saar-Tsechansky and Macskassy (2008), Li (2008, 2010).

reflect tone in financial texts.

The purpose of this paper is to demonstrate that an appropriately constructed tonality factor significantly explains the cross section of excess (and abnormal) stock returns after controlling for the original Fama and French (1993) factors given by: the *HML*, $R_m - R_f$ and *SMB*; as well as: *MOM* (the momentum factor, Carhart, 1997); and two additional factors introduced by Fama and French (2015): *RMW* and *CMA* because these risk factors are known to significantly explain the cross-section of stock returns.[3] Our sample period is from 1994 to 2016 and we include all stocks listed on the NYSE, NASDAQ and AMEX, which represents over 4,000 distinct firms. For any given firm-year, we use the LM dictionary to obtain the fraction of negative words (pessimism) in the 10-K document from the first quarter of that year and classify each document into the top 30 percent (most pessimistic) and the bottom 30 percent (least pessimistic). We then compute the monthly value-weighted portfolio returns for each of these two extreme portfolios. The monthly return difference of the top 30 percent (least) from the bottom 30 percent (most) forms our monthly factor, *TONE*. Next, we form the decile stock portfolios based on pessimism (i.e. the previous year's fraction of negative words). Thus, for example, Decile 1 (Decile 10) would comprise of stocks with the most (least) pessimism. We estimate portfolio based time-series regressions to determine the relationship between *TONE* and portfolio returns. In each of the regressions, the dependent variable is the monthly portfolio excess return while the independent variables are the six afore-mentioned factors along with *TONE*.[4]

We find that the coefficient associated with *TONE* is significantly negatively (positively) correlated with excess stock returns for the most (least) pessimistic stock

[3]Barillas and Shanken (2018) employing a Bayesian approach document that the Fama and French (2015) factors (i.e. *HML*, $R_m - R_f$, *SMB*, *CMA* and *RMW*), *MOM* by Carhart (1997) and investment factor (*CMA*) proposed by Hou, Xue and Zhang (2015) dominate several other factors used in asset pricing tests. The authors show that a model with these six-factors has the highest explanatory power.

[4]In other words, the factors and the decile portfolios are created on previous year's measures while the monthly performance is estimated based on the following year values.

portfolio. To allay any potential concerns that *TONE* might be capturing some of the risks captured by the other established factors we perform spanning regressions to establish that our factor captures information that is orthogonal to the risks captured by factors such as *SMB*, *HML*, $R_m - R_f$, *CMA*, *RMW* and *MOM*.[5]

We further perform a double-sort along pessimism and successively with size, book-to-market, past performance, operating profitability and asset growth.[6] In the pessimism/size sort, we show that *TONE* has a positive (negative) and significant coefficient for small (large) stocks associated with the least pessimism; *TONE* also has a positive coefficient for large stocks associated with the most pessimism. In the pessimism/book-to-market sort we show that *TONE* has a positive and significant correlation with returns for the most pessimistic and growth stocks. In contrast, *TONE* has a negative and significant correlation with returns for the least pessimistic and value stocks. In the pessimism/past performance sorts we show that *TONE* has negative and significant correlation with returns for most pessimistic as well as low past performance portfolios. Therefore, while past performance does not appear to matter for the *TONE*-excess return relationship, the pessimism/profitability and the pessimism/ asset growth sorts show that the least pessimistic firms appear to signal their value through positive statements and effective communication of risk-related

[5]To put these adjusted R^2- square numbers in perspective, Carhart (1997) reports an average adjusted R^2 of 0.92 in his time-series regressions of average mutual fund excess returns on *SMB*, *HML*, $R_m - R_f$ and *MOM* for the decile sorted portfolios. Also, Hou et al. (2015) report an average R^2 of 0.94 for their monthly regressions for their four-factor model. Similarly, Lewellen (1999) reports an average R^2 of 0.90 for the three-factor model estimated on the industry average excess returns for the industry-sorted portfolios.

[6]We simply follow a time-honored tradition of testing for the explanatory power of the risk factors is through double sorting the stocks into portfolios based on the firm characteristics. See for example, Li (2007) who sorts the stocks based on firm's financial constraints and R&D to examine the returns within each portfolios. Nielsen (2007) employs sorting on payout policy and leverage, Lee and Swaminathan (2000) constructs portfolios using double sort on momentum and trading volume. Fama and French (1993) and Daniel, Titman, Grinblatt and Wermers (1997) were the early papers that popularized the double sorting approach. For a detailed discussion on double portfolio sorts and tests of cross-sectional returns, see Patton and Timmerman (2008).

information.[7]

It also appears that the management in large firms are more likely to be cautious in their public pronouncements (resulting in relatively greater pessimism) since larger firms are more susceptible to class-action lawsuits (Dey, 2008; Watts and Zimmerman, 1990).[8] Finally, consistent with the notion of growth firms are abnormally risky, the managers of such firms appear to have an incentive to reasonably lower such risk perceptions by clearly disclosing risk related information in their respective annual filings.[9]

We ensure that *TONE* is not simply one of the hundreds of factors that have been discovered in the past decade and have been collectively labeled as a "factor zoo" by Cochran (2011) in his AFA presidential address. A notable feature of these factors within the zoo is that they have been discovered primarily through data mining exercises and are mostly devoid of any meaningful economic content. Specifically, Harvey, Liu and Zhu (2016) and Hou, Xue and Zhang (2018) have documented over three hundred such factors. The collective conclusion of these studies is a majority of these factors fail to replicate, falling well short of acceptable standards of empirical finance. In particular, Harvey et al. (2016) establish the standard for concluding

[7] Recently investors have witnessed a surge in apple prices and several analysts have raised their third quarter 2019, earnings estimates for Apple's stock based on the firm's optimistic outlook in their 10-K despite the decline in iPhone sales.
(See https://markets.businessinsider.com/news/stocks/apple-stock-price-1-year-high-rising-iphone-11-demand-2019-10-1028591463).

[8] Dey (2008) and Watts and Zimmerman (1990) document that agency conflicts are more likely to arise in bigger firms and the managers of such firms have a greater incentive to reduce litigation costs. Such an increased potential for litigation may encourage such firms to be more cautious in discussing future events in the Management Discussion and Analysis (MD&A) section of the 10-Ks (Li, 2010)

[9] For example, Facebook in their 2013, 10-K stated that "We anticipate that our active user growth rate will continue to decline over time as the size of our active user base increases, and as we achieve higher market penetration rates. If people do not perceive our products to be useful, reliable, and trustworthy, we may not be able to attract or retain users or otherwise maintain or increase the frequency and duration of their engagement. A number of other social networking companies that achieved early popularity have since seen their active user bases or levels of engagement decline, in some cases precipitously. There is no guarantee that we will not experience a similar erosion of our active user base or engagement levels."

if a given factor is "truly significant" then the t-statistic associated with the average return should be above 3.0. In fact, they go even further, and propose that the statistical significance of the returns associated with a given factor (being distinct from zero) be set at a level higher than 3.0 because a falsely discovered factor could potentially clear the 3.0 cut-off. Therefore, to alleviate concerns of a factor, created through mostly mechanical means, from clearing such a hurdle, they employ multiple testing methods and derive a cutoff value even higher at a t-statistic of 3.54, 3.2 and 2.67 corresponding to Bonferroni's, Holm's and Benjamini, Hochberg and Yekutieli's (BHY) adjustments, respectively.[10] Comparing against such standards, the risk premium associated with *TONE* displays a t-statistic = 3.5 which clears the threshold t-statistic of at least 3.0 and implies that *TONE* is a true discovery outside the purview of the factor zoo. Therefore, it is the relative lack of significance and importance of these factors that lead us to omit them in our model.

Finally, as a robustness, we re-construct *TONE* using the alternative approach, where we employ the Jegadeesh and Wu (2013) weighting methodology and define the most (least) pessimistic portfolios based on the largest (smallest) weighted score. Then we re-estimate our regressions and confirm that our results are robust to alternative definition of tonality. Additionally, we also use the more recent methodology of employing individual stock-level regressions and confirm that *TONE* indeed plays a positive and significant role in explaining excess returns. Overall, we are able to show that qualitative information improves the model's explanatory power over and above the quantifiable factors that are commonly used to explain stock returns.

The remainder of this paper is organized in four sections. Section 1.3 discusses the details of constructing the tonality factor and provides the data sources. Section

[10] Hou et al. (2018) report that 65 per cent of their list of over 400 factors cannot clear the single test hurdle of t-statistic = 1.96. This percentage increases significantly if the authors eliminate "microcap" stocks from their study. Additionally, the authors show that, regardless of microcaps, most of their factors fail to replicate if they set a higher hurdle of a cutoff t-statistic = 2.78 through multiple testing as identified in their Table 3 (p. 17).

1.4 provides the results and Section **??** provides the results of our robustness tests. Section **??** presents our concluding remarks.

1.2 Background literature on textual analysis

Textual analysis has been in vogue in some form or the other across many disciplines such as natural language processing, content analysis or information retrieval analysis (Loughran and McDonald, 2015). Specifically, content analysis objectively characterizes the tone of a specific text description.[11] With the advent of the data mining techniques and increase in computational power, textual analysis has emerged as a research area in accounting and finance. Following the earlier papers by Frazier, Ingram and Tennyson (1984), Antweiler and Frank (2004), Das and Chen (2007), Tetlock (2007) and Li (2008), Loughran and McDonald (2011) several accounting and finance studies have explored the impact of qualitative information on stock returns, earnings and predict fraudulent activities. As far as finance literature is concerned, Tetlock (2007) in his pioneering study uses content analysis to measure the interaction between the media content and stock market activity. Specifically the author constructs a 'pessimism' measure from the content of the Wall Street Journal (WSJ) column and argue that high media pessimism can predict downward price pressure on NYSE stocks, in addition to unusually high market trading volume. Similarly, Mayew and Venkatachalam (2009), analyze conference call audio files to form opinion on positive and negative cues revealed during the conference calls.

Furthermore, other researchers have studied the linkages of tonal content of newspaper articles or company press releases including SEC form 10-K and IPO prospectus. For example, Kothari, Li and Short (2008) construct firm specific financial disclosure measures based on exhaustive set of sources such as management press releases, news reports, and analyst reports etc. The authors document that nega-

[11] Since textual analysis is a part of content analysis, in this study we use the term textual analysis and content analysis interchangeably.

tive news disclosure is strongly weighted by the market and positive disclosure is discounted. Similarly, Demers and Vega (2008) examine the impact of the tonality of the text contained in the managements' quarterly earnings press releases on abnormal stock price and volatility. The authors study whether the soft information derived from the textual analysis has incremental information relative to the hard information provided by the management. Further, Henry (2008) analyses the tone of earnings press releases and document that firm's abnormal market return increases as the tone of the press release becomes more positive. In a similar vein, Li (2008, 2009) examine the relationship between annual report readability and firm performance by measuring the readability through a statistic that combines the number of words per sentence and the number of syllables per word. On similar lines, Hanley and Hoberg (2010), uses content analysis to measure the informative and standard components of the initial public offering (IPO) prospectus and show that greater informative content of the IPO prospectus results in accurate offer prices and less underpricing and vice versa for standard content.

In general, the literature on finance and accounting have accessed various sources of text to make predictions about future stock performance and to better understand the management's outlook and disclosure policies. Lu, Chen, Chen, Hung and Li (2010) has broadly divided the finance related textual analysis studies into three classifications: (a)content generated by the firm's management,(b) internet forums, twitter etc., (c) news articles, and analyst research reports. The focus of the current study falls under the second classification (i.e. content generated by firm's management), which has superior reliability (see Das, 2014 for a comprehensive survey on textual analysis) relative to other classifications (i.e. internet sources and news articles).

Much of textual analysis in finance have used message posted on internet forums and news articles. For example earlier works by Tumarkin and Whitelaw (2001), Antweiler and Frank (2004, 2005), Das and Chen (2007) and Das, Martinez-Jerez and Tufano (2005) have focused on extracting information from messages posted on internet forums and message boards. Other works such as Leinweber and Sisk (2011) uses Thompson-Reuters NewsScope Sentiment Engine (RNSE) to analyze the impact of tonality on stock returns. Similarly, Brown (2012) looks at the relationship between tweets and stock prices through several measures. In a similar vein, Bollen, Mao and Zeng (2010) shows that the direction of stock prices can be predicted using tweets with an accuracy of 87.6%.

A majority of previous studies focus on predicting future stock price movements and an overwhelming majority of such studies have relied on the word classification in the Harvard Psycho sociological Dictionary to categorize the words as either positive or negative. Recently, Loghran and McDonald [2011, hereafter LM (2011)] show that the Harvard dictionary list may not be suitable to accurately classify the positive and negative tone of the words in the financial context.[12] Therefore LM (2011) create a comprehensive wordlist of positive, negative, modal, litigious, uncertainty and constraining based on the SEC 10-K reports. [13] Specifically LM (2011) use the inverse document frequency (idf) as a weighting approach for word to modify the term frequency counts in the 10-K documents. The *idf* for word j would be

$$w_j^{idf} = ln(\frac{N}{df_i})$$

Where N is the total number of documents, and df_i is the number of documents containing word j. The idf weighting scheme was proposed by Manning and and Schütze (1999). Through the *idf* weighting scheme LM (2011) arrive at the following weight on word j

$$w_{ij} = max(0, 1 + ln(f_{ij})w_j^{idf})$$

[12]LM (2011) show that almost 75% of the negative word counts in 10-K reports based on the Harvard wordlist are not negative in financial context. For example, depreciation, tax, liability and board are considered negative under Harvard wordlist.

[13]Available at https://sraf.nd.edu/textual-analysis/resources/

Where f_{ij} is the frequency of the word j in document i. Following the above weight LM (2011) compute the document score as follows:

$$S^{LM} = \frac{1}{1 + ln(a_i)} \sum_{j=1}^{J} w_{ij}$$

Where S^{LM} stands for LM (2011) score, a_i is the total number of words in the document i, and J is the total number of words in the lexicon. Overall, LM (2011) in their study show that negative wordlist has improved explanatory power relative to the Harvard list. On similar lines, Jegadeesh and Wu (2013, hereafter JW) argue that the previous studies on textual analysis typically measure the tone of the document through the ratio of negative or positive words to total number of words within a given document. This approach implicitly assumes equal weighting scheme for all words within a given category. However, the authors propose a new approach, which builds on the positive and negative wordlist by LM (2011) and objectively determines the strength of various words in the lexicon to measure the toe of 10-Ks. Specifically JW (2013) compute a document score as specified below:

$$S^{JW} = \frac{1}{a_i} \sum_{j=1}^{J} w_{ij} f_{ij}$$

Where, S^{JW} stands for JW (2013) score, w_i is the weight for word j. Furthermore, JW (2013) find several new results in their study, first they find that their document score has very low correlation with the score computed by LM (2011) using an identical wordlist. Second, the authors document that stock market does not 'react' during the 10-K filing period. In addition, JW (2013) extend their lexicon to compute the document score of IPO prospectuses and find negative relationship between tone and IPO underpricing.

As highlighted in LM (2011) and earlier studies textual analysis can help us to understand the information content in company released documents such as 10-K, 8-

K reports and earnings press releases. In the current study, we build on the previous work of LM (2011) and JW (2013) to examine the relationship between tonality of 10-K reports and quarterly fund returns. To examine whether textual tone has significant correlation with fund returns we draw the intuition from earlier works of Fama and French (1993) and Carhart (1997). Given the importance of size, book value and risk factors highlighted by Fama and French (1993) and momentum anomaly by Jegadeesh and Titman (1993) in explaining stock returns, we argue that the qualitative aspect of textual tone could have incremental explanatory power. Since firms use their annual filings (i.e. 10-K reports) to disclose information to current and future investors, understanding the tonality of such annual filings can provide better insights into investor's portfolio returns in the future.

1.3 Constructing the *TONE* factor

Our empirical tests use data from three sources. First, we obtain the Loughran and McDonald's (hereafter LM wordlist) positive/negative wordlist and from the authors' website. Second, we obtain the monthly Fama and French (1993 and 2015) factors (i.e. SMB, HML, $R_m - R_f$, CMA, and RMW) and momentum factor (MOM) from Kenneth French's website for the sample period January 1994 through December 2016. Third, we obtain the monthly stock returns, monthly index returns, month end market value from the Center for Research in Security Prices (CRSP) as well as accounting information such as annual book value and shares outstanding from the Compustat Annual Fundamental files. Following prior studies, we restrict the sample to NYSE, AMEX and NASDAQ stocks only with share codes of 10 and 11 both on CRSP and Compustat. In addition, we use the following criteria to construct the sample for the analyses.

- We consider only the initial 10-K filing in a given year by a given firm. That is, we exclude subsequent amendments to the initial 10-K that are filed by many firms over the course of the year.
- Since SEC, EDGAR identifies the firms through Central Index Key (CIK). We use the WRDS CRSP-COMPUSTAT database to match the CIKs in EDGAR and WRDS. We exclude all firms for which we not able to match the CIKs between the two databases.
- The analyses use various accounting and financial variables such as market capitalization and stock returns. We exclude all firms for which we do not have these data.
- We require that the firms should have a closing price of at least $3 on the filing date to be included in the sample. Eliminating low-priced firms is standard in textual analyses in finance and empirical asset pricing literature to reduce the role of bid-ask bounce.

Specifically, we start with a sample of 182,267 10-Ks (excluding duplicates) during the sample period April 1994 through December 2016. However, the above listed filters along with the requirement of matching the firms with CRSP/COMPUSTAT reduces the original sample to 94,239 stock-years.

The fundamental variable we use to identify the tonality of the 10-K document is the net negative words (pessimism), defined as the number of negative words minus positive words divided by the total number of effective words in a given 10-K (i.e. each year for a given firm).[14] Since the 10-Ks that are published in a given year (t) contains accounting and other information corresponding to the previous year (t-1) along with the forward looking statements in the Management Discussion and Analysis (MD&A), we create our year t factors based on year t-1 information similar to

[14]The fraction of negative words has also been used by Tetlock, Saar-Tsechansky and Macskassy (2008) and Loughran and McDonald (2016).

the creation of the momentum factor by Carhart (1997). Our goal therefore is to determine whether the information contained in year t in a given 10-K, is able to explain stock returns.

At the end of every June, we sort the stocks into the top (and bottom) 30 percent portfolios based on pessimism corresponding to the 10-K of previous year.[15] Next, we compute the monthly value weighted (by market size) portfolio returns for the top 30 percent (most pessimistic) and the bottom 30 percent (least pessimistic). The monthly factor *TONE* is then constructed as the difference in value weighted monthly returns between the top and bottom 30 percent portfolios. We employ two models of performance measurement: the first model consists of the five-factors described in Fama and French (2015) plus the Carhart (1997) factor as a benchmark model and in the second model, we add *TONE* to the first model.[16] We use the firm-specific characteristics in creating double-sort portfolios. Firm size (*size*) is defined as market capitalization of equity at the end of June; book to market (*bm*) is defined at the annual book value of equity divided by market capitalization of equity; operating profitability (*op*) is defined as is annual revenues minus cost of goods sold, interest expense, and selling, general, and administrative expenses divided by book value of equity for the last fiscal year end in t-1; investments (*inv*) is defined as change in total assets from the fiscal year ending in year t-2 to the fiscal year ending in t-1, divided by t-2 total assets at the end of each June. Finally, prior year returns () is defined as

[15] Specifically, we sort each 10-K based on a net measure of the number of negative words minus the number of negative words divided by the total number of effective words in the document. This net measure can be more or less net negative. Therefore, the top 30 percent portfolios have stocks with 10-Ks having the most net negative words and the bottom 30 percent of the portfolios have stocks with 10-Ks having the least net negative words.

[16] Thus, *SMB* is the difference between the returns on the three 'small' portfolios minus the average return on the three 'large' portfolios. *HML* is the average return on the two value portfolios minus the average return on the two growth portfolios. $R_m - R_f$ is the excess market return. *MOM* is computed using prior year cumulative returns for the months t-2 through t-12. *RMW* is the average return on the two 'high' operating profitability portfolios minus the average return on the two 'low' operating profitability portfolios. *CMA* is the average return on the two 'low' investment portfolios minus the average return on the two 'high' investment portfolios.

t-1 to t-11 month stock returns. These definitions are standard in the asset pricing literature and are consistent with Fama and French (2015) and Carhart (1993).

1.4 Results

1.4.1 *Summary Statistics*

Panel A of Table 1.1 presents the average time-series returns associated with each of the factors used in the current paper. We see that the factor (*TONE*) earns a statistically significant return of 4.08% (t = 3.5) over the sample period of 1994 through 2016. In comparison, *SMB* earns an average of 9.02% (t= 2.36), *HML* 1.93% (t= 2.93), *MOM* earns 4.32% (t =5.02) and CMW and RMA earns 3% (t = 4.26) and 4.52% (t =3.23), respectively. Note that all of these factors clear the stringent Bonferroni adjusted t-statistic of 3.43 set by Harvey et al. (2016) which implies that *TONE* is a significant factor.[17]

Panel B of Table 1.1 reports the time series averages of the pairwise Spearman rank correlations between all of the factors that we use. This is to ensure that the underlying information in *TONE* is not subsumed in the other factors (correlation ranges between 0.07 and 0.21 with the other factors). While the correlation between *CMA* and *HML* at 0.62 might appear, high but is consistent with 0.69 that has been reported by Hou et al. (2015).

[17] We also demonstrate that the average control factor returns from our data and those reported by other studies on the same factors appear to be comparable. For example, Lewellen (2011) reports an average *SMB* of 1.9% (t = 0.94), *HML* of 4.6% (t = 6.35) and 9.7% (t = 7.14) *MOM* returns for his sample: 1980 to 2007. Similarly, Hou et al. (2015) report that the profitability (*RMW*) and investment (*CMA*) factors earn an average of 7% (t = 4.95) and 5.5% (t = 4.81).

1.4.2 *Portfolio regressions based on one- and two-level sorts*

Portfolio sorting approach has been a popular tool in empirical finance to examine the exposures of a given portfolio returns to the explanatory risk factors. In practice, a common approach is to sort the stocks into portfolios on a single firm characteristic and estimate the regression coefficients of portfolio average returns on risk factors (see Cochrane 2011, p. 1061). Carhart (1997) employs time-series regressions in decile-sorted portfolios to show that the factor risk premia of the 4-factor model indicate the average returns attributable to the risk factors. Similarly, Lakonishok, Schleifer and Vishny (1994) examine the returns of the decile-sorted portfolios based on bm, earnings to price (E/P) and cash flow to price (C/P).

However, Asparouhova, Bessembinder and Kalcheva (2010) show that noisy prices lead to both biased intercepts and biased coefficient estimates in OLS regressions with returns as the dependent variable. They, therefore, suggest estimating such regressions using weighted least squares (WLS) since WLS assigns appropriate weights for small and large firms, thereby correcting for any biases created by noisy stock prices. They consider three weighting methods: value weighting (weights based on prior-month market value), return weighting (weights based on prior-month returns) and annual value weighting (weights based on prior-December market value) methods in the WLS regressions. The authors conclude that value-weighting and return-weighting methods have the least bias relative to the equally weighted approach of OLS. We too use the WLS regression approach using value-weighting.

At the beginning of every June, we form ten value-weighted (decile) portfolios of stocks using pessimism. The portfolios are rebalanced every year given the annual nature of the 10-K. Stocks in Decile 1 (Decile 10) are further divided into 1A, 1B and 1C (10A, 10B and 10C) corresponding to the top 30%, middle 40% and bottom 30%. These sub-portfolios are formed simply to examine the relative correlation between the risk factors and excess stock returns. Each decile portfolio (and sub-portfolio) consists of a monthly time-series of 273 returns from 1994 to 2016. Next, we estimate the portfolio performance by employing the following two models for each of the portfolios:

$$\begin{aligned} EX\_RET = & a_t + b_t(R_m - R_f) + s_t(SMB) + h_t(HML) + m_t(MOM) + c_t(CMA) \\ & + r_t(RMW) \end{aligned} \tag{1.1}$$

$$\begin{aligned} EX\_RET = & a_t + b_t(R_m - R_f) + s_t(SMB) + h_t(HML) + m_t(MOM) + c_t(CMA) \\ & + r_t(RMW) + e_t(TONE) \end{aligned} \tag{1.2}$$

where *EX_RET* is the value-weighted average return on a portfolio in excess of the one-month T-bill return; $(R_m - R_f)$ is the excess return on the value-weighted market proxy and *SMB*,*HML*,*MOM*, *CMA*, *RMW* and *TONE* are factors capturing size, value, momentum, operating profitability, investment and pessimism, respectively.

In Table 1.2, we report the time-series regressions for each of the 10 portfolios and the sub-portfolios, where Decile 1 consists of stocks with the smallest *pessimism* (i.e. least pessimistic portfolio) and Decile 10 consists of stocks displaying the largest *pessimism*(i.e. most pessimistic portfolio). In Eq. (1.1), the spread in the alpha (the intercept) between the Decile 1 and Decile 10 is 27 basis points per month (3.2 percent per year). In contrast, the spread in Eq.(1.2) is 5 basis points per month (0.6 percent per year), indicating that *TONE* captures more of the spread in the seven-factor model. Further, within Decile 1 (and Decile 10) we construct three sub-portfolios namely 1A, 1B and 1C (10A, 10B and 10C). For example, the sub-portfolios 1A (1C) correspond to least (most) pessimism. Similarly, within Decile 10, the sub-portfolios 10A (10C) correspond to the least (most) pessimism. We compare the spreads in intercept between extreme portfolios that is the portfolio capturing the least pessimism (1A) and most pessimism (10C). We find that the six-factor model has a larger intercept relative to the seven-factor model (i.e. portfolio 1A has the monthly intercept of 1.12 basis points and portfolio 10C has the intercept of 0.66 basis points), suggesting that the model with *TONE* explains most of the spread. In other words, the lower spread in intercept shows that *TONE*, when added to the six-factor model, is able to explain more of the cross-sectional return variation of the corresponding portfolio. In addition, the factor sensitivities in the seven-factor model explain a significant proportion of the return variation suggested by a higher adjusted R^2 of 0.92 compared to 0.87 in the six-factor model. The pattern in the extreme portfolios is still evident in Deciles 1 to 3 (Deciles 6 – 10) for strong positive (negative) correlation between *TONE* and *EX_RET*.

The positive relationship between *TONE* and portfolio excess returns for the least pessimistic group can be explained by the findings in LM (2011), who show that firms reporting fewer negative words display a more positive reaction over the filing date window. Similarly, Heston and Sinha (2017), using the Harvard dictionary, show that the positive net sentiment measure (positive words minus negative words) results in higher stock returns over a window of one to four trading days. The negative relationship can be explained in light of the findings in Tetlock et al. (2008), who use the Wall Street Journal (WSJ) and Dow Jones News Service (DJNS) news stories about the S&P 500 firms between 1980 and 2004 and show that a higher percentage of negative words results in a lower standardized unexpected earnings. Similarly, LM (2011) show that 10-K documents that use more negative words have lower filing date returns. Thus, a non-positive relationship in five out of the 10 decile portfolios suggests that investors view relative pessimism as a signal of reduced future cash flows (see also, Huang, Teoh and Zhang, 2014). We note that the control risk factors are consistent in sign and significance to the various existing studies cited in the paper. Upon re-estimating the model by using *AB_RET* and find (unreported) similar results as our regressions with *EX_RET*.

1.4.3 *Factor spanning regressions on six factors*

A skeptic might suggest that the risk elements captured in *TONE* are not unique and are instead elements from the other established risk factors that are being channeled through an alternative qualitative route. In other words, the tonality of the documents is picking up the same risk factors through a different channel. To investigate this, we estimate a monthly time-series regression with *TONE* as a dependent variable and the remaining six factors as the RHS variables. The intuition of this regression estimation is that if these six factors can explain *TONE*, then these factors can also absorb the portion *TONE* explains excess returns.[18] Thus if *TONE* captures incremental information about excess stock return Fama and French (2015) argue that the five-factor model does not improve the model's performance compared to the model where *HML* is dropped (i.e. four-factor model with the factors *SMB*, $R_m - R_f$, *RMW* and *CMA*). The authors show that *HML* is a redundant factor through estimate regressions of each of the five factors on the other four. urns, over and beyond the other factors, we should expect a statistically significant intercept from the regression of the six factors on *TONE*.

Table 1.3 shows the regressions of the six factors on *TONE*. Specifically, we use *TONE* as the dependent variable with the remaining six factors as the RHS variables. In the monthly time-series regression we find that the intercept is 0.21 (t = 2.14). The significant intercept suggests that the explanatory power of *TONE* is not absorbed by the other six factors. Further, in the *TONE* regression, we find that the coefficients for *HML* and *CMA* are positive and significant at 1%, while the coefficient for *SMB* is significant at 10%. However, we find that *RMW* has a negative and significant relationship with *TONE*. The regression in Table 6 shows that *SMB*, *HML*, *CMA* and *RMW* explain some of the variation in *TONE*. The result from this test provides

[18]Fama and French (2015) argue that the five-factor model does not improve the model's performance compared to the model where *HML* is dropped (i.e. four-factor model with the factors *SMB*, $R_m - R_f$, *RMW* and *CMA*). The authors show that *HML* is a redundant factor through estimate regressions of each of the five factors on the other four.

evidence that the information content in *TONE* is not subsumed by other factors and strengthens our argument that quantifiable information when coupled with qualitative information provides a better description of the excess stock returns.

The extant empirical asset pricing literature establishes that firm specific characteristics like *size*, *bm*, *op*, *inv*and *pr*(these are the primitives behind the factors, *SMB*, *HML*, *RMW*, *CMA* and *MOM*, respectively) explain most of the variation in stock returns. Additionally, the textual analysis literature has shown that tonality explains variations in stock returns (see, Das, 2015, for a review). Therefore, we too control for both the pessimism as well as each of the five primitives, one at a time (i.e., double sorts with pessimism as the first sort in each instance), to investigate the explanatory power of *TONE* in the double sorted portfolios in the extreme tertiles.

Panel A - E in Table 1.4 present the results for each of the double sorted portfolios on the lines described above. Specifically, from Panel A we find that the coefficient estimate of *TONE* in the large stock – most pessimism portfolio is 0.627 (t - stat = 18.36); for the large stock – least pessimism portfolio is -0.373 (t-stat = -11.65); for the small stock - least pessimism portfolio is 0.194 (t - stat = 1.9). Thus, for large stocks *TONE* has a significant relationship with returns, whereas for the small stocks *TONE* has significant relationship with only the least pessimistic portfolio. In Panel B, we find that the coefficient of *TONE* for large *bm*- least pessimism is -0.378 (t - stat = -4.15); for small *bm*– least pessimism is -0.295 (t-stat = -4.75); for small *bm*– most pessimism is 0.62 (t - stat = 8.13). Thus, for least (most) pessimism stocks *TONE* appears to have a significantly negative (positive) relationship with excess returns. Likewise, in Panel C, we find that the coefficient of *TONE* in the small *pr*– most pessimism portfolio is -1.252 (t-stat = -4.85); for the large *pr*– most pessimism portfolio is -0.813 (t-stat = -4.52). Thus, independent of past returns, *TONE* appears to have a negative relationship with returns for the most pessimistic stocks. In Panel D, the coefficient estimate of *TONE* for the large op – least pessimism portfolio is

-0.368 (t-stat = -7.45); for the small op – most pessimism portfolio it is 0.711 (t-stat = 4.99); for the large op – most pessimism portfolio it is 0.62 (t-stat = 9.99). Thus, the most pessimistic portfolio with profitable stocks appear to have a negative relationship between *TONE* and excess returns. Finally, in Panel E, we find that the coefficient estimate of *TONE* in large *inv*– least pessimism is -0.278 (t-stat = -3.87); for the small *inv*– least pessimism portfolio it is -0.354 (t-stat = -3.22); for the large *inv*– most pessimism portfolio it is 0.619 (t-stat = 6.89); and for the small *inv*– most pessimism portfolio it is 0.771 (t-stat = 5.72). For the most pessimistic stocks with large investments, *TONE* has a positive relationship with excess returns.

The results in Panel A can be explained by previous literature, which has documented that firm size and book-to-market are known to be the main determinants of the document tonality (Das, 2015; Tetlock et al. 2008, JW, 2013). Firm size proxy for the risks and book-to-market proxy for growth. Among these firm-specific characteristics, size is the fundamental determinant of disclosure information in 10-K filings. The earlier works of Singhvi and Desai (1971) and Moore and Buzby (1972) show that size is a primary determinant of corporate disclosure. Specifically, these authors document that large firms have higher disclosure quality relative to small firms and large firms have stronger motivation to disclose more information. Therefore, we see that the large stocks have significant relationship between *TONE* and returns. In addition, large firms tend to be more complex and have more varied operations. This characteristic implies higher risk levels, which translate into higher information asymmetry among investors (Deumes and Knechel, 2008). According to agency theory, risk reporting may reduce agency costs and information asymmetry between managers and shareholders (Watts and Zimmerman 1983). Management of large firms are more likely to be cautious in their discussions (resulting in relatively less optimism) because larger firms are sensitive to lawsuits (Buskirk and Zechman, 2011).[19] Therefore, to mitigate litigation risk and agency costs large firms increases

[19]The authors use 165 lawsuits filed between 2003 and 2008 and show that firms with optimistic disclosure tone have a higher likelihood of shareholder litigation than other firms.

its risk-related disclosure (Oliveira, Rodrigues and Craig, 2011). Likewise, the results in Panel B imply that the smaller *bm*stocks are valued for their growth opportunities and are more likely to explain the risks in their businesses. For example, Khurana, Pereira and Martin (2006) show that risk disclosure enables growth firms to obtain external financing easily.Additionally, the small *bm* stocks are typically growth firms and are risky so the managers in such firms have an incentive to lower the uncertainty. Therefore, more pessimism in risky stocks may reduce uncertainty regarding cash flows and business outlook. Additionally, theoretical and empirical research argue that firms with high levels of risk disclosures reduces information asymmetry between informed and uninformed investors and improves stock liquidity (Diamond and Verrecchia, 1991; Kim and Verrecchia, 1994; Healy and Pelpu, 2001).[20] Thus, we see that the portfolio with smaller *bm* stocks have a significantly positive relationship between *TONE* and returns.

In Panel C, among the most pessimistic group, there is a negative relationship between current excess returns and *TONE*, that is independent of past performance (captured by Momentum), or positive excess returns is correlated with lower *TONE*. Thus, past performance does not appear to matter in the relationship between *TONE* and excess returns. One possible explanation of this finding is that *TONE* dominates momentum.

The results in Panels D might be explained in light of managers in most profitable and least pessimistic firms have greater incentive to signal the quality of their performance and convey good news (relative to bad news) in order to avoid undervaluation of their shares (Giner, 1997). Elshandidy, Fraser, and Hussainey (2013) argue that risk disclosure may reduce uncertainty regarding future cash flows. In

[20]A study ("User perspective on financial instrument risk disclosures under international financial reporting standards") by CFA institute highlight that market participants such as investors and analysts prefer to have the risk information explained qualitatively and summarized in the 10-k documents.

addition, managers are inclined to communicate risk-related information to improve the corporation's image (Iatridis, 2008). In contrast, the result for most profitable firms disclosing bad news results in a lower return associated with *TONE* consistent with the signaling theory. Therefore, firms that explain the risks through usage of negative words have a negative relationship with performance.

The result in Panel E is implies that firms with the least pessimism are profitable in comparison to firms with most pessimism.

1.4.4 *Individual stock-level regressions*

In a stark deviation from the portfolio based approach of investigating for the significance of factors in explaining excess returns, Ang, Liu and Schwarz (2019) take the position that creating portfolios would result in loss of information and estimation efficiency. They further show, both theoretically and through simulations of empirical data, that the estimates for $R_m - R_f$ have a positive and significant relationship with excess returns when individual stocks are used while these estimates are statistically insignificant when used in portfolio level regressions. The authors show that that individual stock regressions display smaller standard errors of the factor risk premia while the act of forming portfolios washes away firm-specific information. Accordingly, they propose and estimate two-pass individual stock level regressions where, in the first step, they estimate individual stock time-series of excess returns on the factors. This yields the estimated coefficients (i.e. βs) for each stock. In the second step, they estimate a cross-sectional regression of the excess stock returns on these βs.

Thus, to ensure that our portfolio level regressions are not similarly affected, we follow and amplify the Ang et al. (2019) regression method by using both WLS and OLS estimations. Table 5 reports the results for individual stock regression estimations. In Col 2, of Table 5 we present the present the Ang et al. (2013) OLS estimation and in Col 4 we present the estimates of the WLS regression. We find that the coefficient estimate of *TONE* for Ang et al. method is 0.01, (t - stat = 2.5) and the coefficient estimate for WLS method is 1.565, (t - stat = 1.98), confirming that our estimations are not biased. In sum, the individual stock-level regressions provide further evidence that *TONE* can significantly explain the excess stock returns when other risk factors are appropriately accounted for.

1.4.5 *Directly confronting the factor zoo problem*

We have so far estimated the marginal contribution of our *TONE* factor against the six established factors in the asset pricing literature which includes the five Fama and French (2015) factor and Carhart's (1997) momentum factor. Therefore, for example, the *size* factor premium is the difference in returns between the largest stocks and the smallest stocks in the CRSP universe. Similarly, *TONE* factor premium is the difference in returns between the stocks associated with the most pessimism in their 10-Ks and the stocks associated with the least pessimism in the same annual filings. However, given the extensive documentation of the factor zoo issue whereby hundreds of factors have been discovered over the past decade, mostly through data mining, or the zoo of factors, it is reasonable to ask how *TONE* compares relative to the zoo.

Harvey, Liu and Zhu (2016) present a multiple testing framework to derive a threshold t – stat cutoff to classify the truly significant factors. The authors evaluate 316 significant published factors and highlight that the usual cutoff level of statistical significance is not appropriate, and any newly constructed factor should clear the threshold t - stat of at least 3.0. Therefore, they provide a guidance as to the appropriate significance level a given factor should be compared against. To do so, they follow the statistics literature in employing three p-value adjustment methods: Bonferroni's adjustment, Holm's adjustment and Benjamini, Hochberg and Yekutieli's adjustment.[21] Furthermore, based on the p-value adjustments Harvey et al. obtain three benchmark p-values, with a corresponding t - stat of 3.54, 3.20 and 2.67 for Bonferroni, Holm and Benjamini, Hochberg and Yekutieli's adjustments, respectively. In general Bonferroni's and Holm's adjustments result in higher rejection rates than the third method (see also Sethuraman et al., 2019).[22]

We also provide another way of establishing that *TONE* is a true factor. Hou et al. (2018) replicate the 452 distinct factors that have been documented in the recent finance literature. They use a multiple testing framework to derive a benchmark t-stat of at least 2.78 for the return difference between any given factor's top and bottom decile portfolios. In addition, they also suggest employing value-weighted returns in portfolio sorts rather than equal-weighted returns and estimate cross-sectional WLS regressions to clear the acceptable standard. By that standard, we find that the value-weighted returns associated with *TONE* displays a t- stat of 3.5 and the t - stat is 13.4 for multivariate returns associated with *TONE* between the top and bottom decile portfolios indicating that *TONE* clears the minimum threshold of 2.78. In addition, we also construct our portfolios using value-weighted returns to account for large and small firms in section 1.4.2 and we estimate the cross-sectional regressions using WLS method in section 4. Therefore, based on all of these tests the level of significance of *TONE* implies that *TONE* might be deserving of being labeled as a true factor.

1.5 Weighted *TONE* factor

In this section, we use the approach taken by Jegadeesh and Wu (2013) who objectively determine term weights associated with each of the positive and negative

[21]Specifically, the Bonferroni's adjustment controls the family-wise error rate (FWER), whereas the Holm's and Benjamini, Hochberg and Yekutieli's adjustment control for false discovery rate (FDR). In general, the FWER is a more stringent adjustment than FDR. Also, the Bonferroni's adjustment is known to be the most stringent test among the three methods (Sethuraman, Gonzalez, Grenier, Kansagra, Mey, Nunez-Zavala and Wulf, 2019).

[22]Harvey et al. (2016) suggest that all is not lost if a factor fails to clear the p-value adjustments as long as the factor clears the t-stat of 3.0 and is motivated by a theory rather than by a purely empirical exercise.

words found in the LM wordlist and apply it to 10-Ks. The authors argue that weighting the positive and negative words appropriately, can provide reliable estimation. Furthermore, they show that their approach reliably quantifies the tone of IPO prospectuses as well, and we find that the document score is negatively related to IPO underpricing. Accordingly, we first obtain the term weightings for each negative word for the sample period of 1995 through 2010.[23] Further, we extract the negative words for each firm during the sample period and assign the weights. Finally, we compute a weighted document Score for the negative words for each of the 10-Ks.[24] To construct the weighted measure of *TONE*, we follow similar steps as detailed in Section 1.3. Next, we estimate Eq. (1.1) and (1.2) on the decile sorted portfolios. Decile 1 (Decile 10) corresponds to stocks associated with the least (most) pessimism.

From Table 1.7 we find that the correlation between *TONE* and *EX_RET* in the decile portfolios is similar to that in Table 1.2. For example, we find that the average excess returns for Deciles 2 through 4 is strongly positively correlated with *TONE* with estimates of (0.195, 0.491, and 0.721). While, the average excess portfolio returns for Deciles 7 through 10 is strongly negatively correlated with *TONE* at the 1% level (estimate = -0.23, -0.479, -0.424, -0.389). Other factors display similar signs and significance as in Table 1.2. Further, in unreported tests, we repeat the regressions for the double sort approach as in Table 1.4 and find similar results. In addition to the portfolio regressions, we also perform the individual stock level estimation procedures for the reasons mentioned in section 1.4.4. In unreported results, we find that the coefficient estimates for *TONE* is positive and significant under both the estimation

[23]The term weightings for the positive and negative wordlists are available on Di Wu's website.

[24]Specifically, we compute the weighted Score as follows:

$$Score_i^{(tf,idf)} = \frac{1}{1 + loga_i} \sum_{j=1}^{J} w_j F_{i,j}$$

where a_i is the total number of words (positive or negative) in document j and I is the total number of positive or negative words in the lexicon, w_j is the weight for each word (JW list) and $F_{(i,j)}$ is the number of occurrences of word j in i^{th} document.

methods. For example, the coefficient estimate for Ang et al. method is 0.08, (t-stat = 4.73); the coefficient for the WLS method is 1.592, (t - stat = 2.72) and the). More importantly, we find that the t-statistics of *TONE* is higher relative to those in Table 1.6. Hence, our results appear robust to alternative approaches to the construction of *TONE*.

1.6 Conclusion

In this paper, we consider potentially the most significant document released by a publicly traded entity: its annual 10-K filing, and construct a document tonality factor *TONE*, which we then use to demonstrate that *TONE* significantly explains the cross section of stock excess (and abnormal) returns even after controlling for previously known factors such as the *HML*, $R_m - R_f$ and *SMB* (the original FF factors); the momentum factor and the two additional factors introduced by Fama and French: *RMW* and *CMA*. We further show that *TONE* is positively (negatively) related with returns for stocks with least (most) pessimistic 10-Ks. Through sorts based on specific firm characteristics where *TONE* is allowed to float freely, we show that the portfolio of small, and low book-to-market, stocks present a cautious outlook of their risk factors by employing a relatively pessimistic view in their 10-Ks. We also show that portfolios of stocks with weak past performance and high pessimism are both associated with lower returns. Finally, we show that *TONE* is a truly discovered factor based on tests suggested by Harvey, Liu and Zhu (2016) which are also consistent with cutoffs provided in Hou, Xue and Zhang (2018).

Our study is the first to our knowledge that rigorously establishes the descriptive aspects of a publicly traded entity's 10-K document as a priced risk factor that stands apart from the other quantitative risk factors. Our research opens the door to future work establishing how qualitative information associated with companies can be used to provide investors, academics and policy makers with a relatively complete picture of a publicly traded company's operations in an increasingly complex and global environment.

Table 1.1.: Descriptive statistics

This table presents descriptive statistics of the factors used in our analyses. Panel A reports the average annual factor returns over the sample period 1994 to 2016. *TONE* is constructed as the difference in value weighted stock returns between the firms that are in top 30% of the *pessimism* every month and the firms in bottom 30% of *pessimism*. *SMB* is the difference between the returns on the three 'small' portfolios minus the average return on the three 'big' portfolios. *HML* is the average return on the two value portfolios minus the average return on the two growth portfolios. $R_m - R_f$ is the excess market return. *MOM* is computed using prior year cumulative returns for the months $t-2$ through $t-12$. *RMW* is the average return on the two 'high' operating profitability portfolios minus the average return on the two 'low' operating profitability portfolios. *CMA* is the average return on the two 'low' investment portfolios minus the average return on the two 'high' investment portfolios. Panel B, reports time-series averages of Spearman rank correlation between factors.

Panel A: Average annual factor returns		
Factors	Average returns (%)	T-stats ($H_0 = 0$)
TONE	4.08	3.50
SMB	9.02	2.36
HML	1.93	2.93
MOM	4.32	4.32
CMA	3.00	4.26
RMW	4.52	3.23

Panel B: Spearman rank correlation between factors						
SMB	*HML*	*MOM*	$R_m - R_f$	*CMA*	*RMW*	*TONE*
SMB	1					
HML	0.449	1				
MOM	-0.461	-0.499	1			
$R_m - R_f$	0.074	-0.031	-0.350	1		
CMA	0.302	0.620	0.006	-0.286	1	
RMW	-0.241	0.390	0.141	-0.628	0.323	1
TONE	0.216	-0.198	0.072	-0.145	0.134	-0.173

Table 1.2.: Regressions for decile portfolios formed on *pessimism*

This table reports the time-series regressions of *TONE* on *EX_RET* for decile-based portfolios sorted on *pessimism* each year in April. *pessimism* is defined as the difference between total number of negative words and positive words divided by total effective words in a given 10-K document each year. *EX_RET* is the value weighted average monthly stock returns in excess of one-month U.S. Treasury bill rate. Stocks with lowest *pessimism* comprise Decile 1 and stocks with highest *pessimism* comprise Decile 10. *TONE* is constructed as the difference in value weighted stock returns between the firms that are in top 30% of the *pessimism* every year and the firms in bottom 30% of *pessimism*. *SMB* is the difference between the returns on the three small portfolios minus the average return on the three large portfolios. *HML* is the average return on the two value portfolios minus the average return on the two growth portfolios. $R_m - R_f$ is the excess market return. *MOM* is computed using prior year cumulative returns for the months $t-2$ through $t-12$. *RMW* is the average return on the two high operating profitability portfolios minus the average return on the two low operating profitability portfolios. *CMA* is the average return on the two low investment portfolios minus the average return on the two high investment portfolios. Alpha is the intercept of the model. The standard errors are robust heteroscedasticity-adjusted.

Portfolio	Parameter r	*Alpha*	*TONE*	*SMB*	*HML*	*RMW*	*CMA*	*R_m-R_f*	*MOM*	*Adjusted* R^2
Decile 1	Estimate	0.663	0.300	0.113	-0.057	0.158	0.133	0.891	-0.058	0.870
	t Value	6.650	3.520	2.960	-1.220	2.710	2.070	33.970	-2.000	
Decile 1A	Estimate	1.152	0.220	0.294	-0.834	-0.013	0.941	0.059	1.152	
	t Value	4.770	2.160	2.520	-4.850	-0.110	13.050	0.820	4.770	
Decile 1B	Estimate	1.213	0.074	-0.222	0.072	0.365	1.018	-0.169	1.213	
	t Value	7.300	1.240	-2.420	0.630	2.710	20.090	-4.300	7.300	
Decile 1C	Estimate	0.663	-0.300	0.113	-0.057	0.158	0.133	0.891	-0.058	
	t Value	6.640	-3.520	2.970	-1.210	2.720	2.070	33.970	-2.010	
Decile 2	Estimate	0.839	0.365	-0.098	-0.074	-0.006	0.163	0.985	0.035	0.879
	t Value	8.670	4.230	-2.110	-1.500	-0.090	2.450	32.390	1.020	
Decile 3	Estimate	0.872	0.473	0.076	0.050	-0.129	-0.167	1.037	-0.013	0.866
	t Value	7.310	4.500	1.380	0.780	-1.310	-1.890	29.900	-0.350	
Decile 4	Estimate	0.671	0.085	0.059	0.008	0.055	0.075	0.986	-0.002	0.891
	t Value	6.950	1.270	1.540	0.160	0.810	0.990	38.470	-0.080	
Decile 5	Estimate	0.730	-0.025	0.022	-0.163	0.042	0.288	0.986	-0.009	0.875
	t Value	6.970	-0.270	0.490	-2.510	0.560	3.830	31.390	-0.330	
Decile 6	Estimate	0.644	-0.224	0.070	-0.017	0.048	-0.033	0.952	0.008	0.894
	t Value	6.600	-2.870	1.820	-0.300	0.820	-0.500	30.240	0.300	
Decile 7	Estimate	0.884	-0.308	0.147	0.040	-0.063	-0.046	0.948	-0.067	0.879
	t Value	7.050	-3.100	2.600	0.690	-0.710	-0.570	23.450	-1.700	
Decile 8	Estimate	0.863	-0.561	0.017	-0.015	-0.010	-0.093	0.960	-0.009	0.901
	t Value	8.320	-6.430	0.370	-0.250	-0.140	-1.310	29.410	-0.270	
Decile 9	Estimate	0.845	-0.558	0.096	-0.077	-0.060	0.038	1.020	-0.014	0.899
	t Value	7.610	-5.210	1.650	-1.180	-0.640	0.440	28.320	-0.280	
Decile 10	Estimate	0.703	-0.693	-0.027	-0.082	0.097	0.234	0.951	-0.016	0.867
	t Value	6.580	-7.880	-0.520	-1.370	1.190	3.240	32.210	-0.490	
Decile 10A	Estimate	0.714	-0.429	0.299	-0.092	-0.013	0.009	0.909	-0.099	
	t Value	3.600	-2.400	3.420	-1.010	-0.110	0.070	16.660	-1.420	
Decile 10B	Estimate	0.607	-0.328	0.120	0.019	0.332	0.267	0.946	-0.015	
	t Value	4.030	-2.880	2.100	0.250	3.790	2.550	21.860	-0.430	
Decile 10C	Estimate	0.663	-0.300	0.113	-0.057	0.158	0.133	0.891	-0.058	
	t Value	6.640	-3.520	2.970	-1.210	2.720	2.070	33.970	-2.010	

Table 1.3.: Regressions for decile portfolios formed on *pessimism*

This table presents the spanning regressions for the factors, where the five factors explain the returns on the sixth. *TONE* is constructed as the difference in value weighted stock returns between the firms that are in top 30% of the pessimism every month and the firms in bottom 30% of pessimism. *SMB* is the difference between the returns on the three small portfolios minus the average return on the three large portfolios. *HML* is the average return on the two value portfolios minus the average return on the two growth portfolios. $R_m - R_f$ is the excess market return. *MOM* is computed using prior year cumulative returns for the months t-2 through t-12. *RMW* is the average return on the two high operating profitability portfolios minus the average return on the two low operating profitability portfolios. *CMA* is the average return on the two low investment portfolios minus the average return on the two high investment portfolios. The standard errors are robust and heteroscedasticity-adjusted. *,**,** represent statistical significance at 10%, 5% and 1% respectively.

Dependent variable	*Intercept*	*TONE*	*SMB*	*HML*	*RMW*	*CMA*	*Rm-Rf*	*Mom*	Adjusted R2
TONE	0.219**		0.062	0.181	-0.372	0.173	0.045	-0.024	0.372
T-stat	2.140		1.770	3.860	-7.690	2.730	1.640	-1.180	
SMB	0.291	18.545		0.069	-0.544	0.048	-0.012	0.069	0.249
T-stat	1.630	1.280		0.760	-4.500	0.390	-0.210	1.190	
HML	-0.246*	0.037	29.062		0.511	0.773	0.108	-0.113	0.577
T-stat	-1.680	0.760	3.060		7.140	8.840	2.390	-4.100	
RMW	0.518***	-0.236	0.413	-48.401		-0.055	-0.176	0.030	0.594
T-stat	4.680	-4.940	6.770	-5.420		-0.570	-5.560	0.730	
CMA	0.226**	0.014	0.430	-0.038	15.515		-0.144	0.044	0.504
T-stat	2.120	0.390	11.330	-0.570	2.560		-5.020	1.750	
Rm-Rf	1.022***	-0.019	0.326	-0.656	21.952	-0.779		-0.139	0.363
T-stat	4.760	-0.220	2.450	-5.850	1.720	-4.750		-2.690	
Mom	0.537	0.203	-0.617	0.201	-21.216	-0.252	0.434		0.153
T-stat	1.610	1.330	-3.420	0.750	-0.570	-2.660	1.450		

Table 1.4.: Regression results for double sorts on *pessimism*, *size*,*bm* and *pr*

This table reports monthly time-series regressions (276 months) of *TONE* on *EX_RET* for the 9 *pessimism-size*, *pessimism-bm* and *pessimism-pr* groups. Every year stocks are allocated to three *pessimism* groups and three Size groups (Small to Big) or three *bm* or *pr* groups. *pessimism* is defined as the difference between total number of negative words and positive words divided by total effective words in a given $10-K$ document each year. *size* is calculated every year as stock price times shares outstanding and the groups are formed based on NYSE break points, where small stocks are in the 30^{th} percentile and big stocks are in the 70^{th} percentile range. *bm* is calculated as the ratio of annual book-equity and market-equity, again the groups are formed based on NYSE break points, where Low-*bm* stocks are in the 30^{th} percentile and Hi-*bm* stocks are in the 70^{th} percentile range. *pr* is one-year prior rolling returns of each stock every month and the three groups are formed based on $30^{t}h$ and 70^{th} percentile range. The intersections of the two sorts produce 9 *pessimism-size*, *pessimism-bm* and *pessimism–pr* portfolios. *EX_RET* is defined as the monthly stock returns in excess of one-year U.S. Treasury bill rate. *TONE* is constructed as the difference in value weighted (by *size*) stock returns between the firms that are in top 30% of the *pessimism* every month and the firms in bottom 30% of *pessimism*. *SMB* is the difference between the returns on the three small portfolios minus the average return on the three big portfolios. *HML* is the average return on the two value portfolios minus the average return on the two growth portfolios. $R_m - R_f$ is the excess market return. *MOM* is computed using prior year cumulative returns for the months $t-2$ through $t-12$. *RMW* is the average return on the two 'high' operating profitability portfolios minus the average return on the two 'low' operating profitability portfolios. *CMA* is the average return on the two 'low' investment portfolios minus the average return on the two 'high' investment portfolios. Panel A reports the regressions for portfolios formed on *pessimism-size*, Panel B reports the regressions for portfolios formed on *pessimism-bm* and Panel C reports the regressions for portfolios formed on *pessimism-pr*. The standard errors are clustered robust heteroscedasticity-adjusted. The coefficients and t-statistics are reported.

Panel A: 3x3 sorts on *pessimism- size*						
	Intercept			T-stats		
pessimism	Small	Medium	Big	Small	Medium	Big
Positive	0.416	0.779	0.782	2.38	9.25	14.48
	0.378	0.824	0.785	2.35	9.27	12.32
Negative	0.431	1.051	0.834	2.1	8.67	14.49
	TONE			T-stats		
Positive	0.194	-0.043	-0.373	1.87	-0.87	-11.65
	-0.022	0.057	0.106	-0.22	1.09	2.81
Negative	-0.025	-0.144	0.627	-0.21	-2.01	18.36
	SMB			T-stats		
Positive	0.784	0.961	-0.029	13.08	33.27	-1.58
	0.845	0.965	0.016	15.28	31.7	0.74
Negative	0.844	0.994	-0.045	11.97	23.93	-2.3
	HML			T-stats		
Positive	0.011	0.05	-0.02	0.14	1.28	-0.8
	0.114	0.05	-0.044	1.52	1.19	-1.49
Negative	0.123	0.072	-0.065	1.27	1.28	-2.43

Panel B: 3x3 sorts on *pessimism- bm*						
	Intercept			T-stats		
pessimism	Low-*bm*	Medium *bm*	Hi-*bm*	Low-*bm*	Medium *bm*	Hi-*bm*
Positive	0.984	0.074	-0.71	9.38	0.54	-4.63
	1.109	0.43	-0.897	3.85	3.24	-5.27
Negative	1.214	0.378	-0.36	9.46	2.6	-1.64
	TONE			T-stats		
Positive	-0.295	-0.439	-0.378	-4.75	-5.38	-4.15
	0.028	-0.158	-0.117	0.16	-2	-1.16
Negative	0.619	0.098	0.192	8.13	1.13	1.47
	SMB			T-stats		
Positive	-0.05	0.121	0.278	-1.38	2.57	5.3
	0.045	0.107	0.242	0.46	2.36	4.15
Negative	-0.032	0.075	0.212	-0.74	1.5	2.82
	HML			T-stats		
Positive	-0.103	0.17	0.464	-2.11	2.65	6.47
	-0.12	0.255	0.547	-0.89	4.1	6.88
Negative	-0.356	0.15	0.689	-5.92	2.2	6.72

Panel C: 3x3 sorts on *pessimism- pr*						
	Intercept			T-stats		
pessimism	Low *pr*	Med *pr*	High *pr*	Low *pr*	Med *pr*	High *pr*
Positive	-2.907	0.395	5.053	-17.61	3.76	23.65
	-2.782	0.461	4.619	-17.41	4.38	22.56
Negative	-3.158	0.668	5.087	-17.75	5.94	24.11
	TONE			T-stats		
Positive	-0.309	0.084	0.355	-1.21	0.72	1.65
	-0.566	-0.115	0.134	-2.26	-0.91	0.67
Negative	-1.252	-0.543	-0.813	-4.85	-4.35	-4.52
	SMB			T-stats		
Positive	0.649	0.539	0.872	9.3	13.51	10.26
	0.657	0.492	0.756	9.35	13.07	8.84
Negative	0.711	0.542	0.716	9.59	13.51	9.45
	HML			T-stats		
Positive	-0.038	0.003	-0.257	-0.49	0.06	-2.64
	-0.035	0.034	-0.176	-0.49	0.58	-1.73
Negative	-0.098	0.076	-0.244	-1.11	1.22	-2

Table 1.4 :Continued

	$R_m - R_f$			T-stats			$R_m - R_f$			T-stats			$R_m - R_f$			T-stats		
Positive	0.798	1.004	0.963	16.94	44.25	66.16	0.929	0.947	0.959	32.91	25.58	23.23	0.02	0.316	-0.246	0.16	4.42	-1.75
	0.838	1.028	0.958	19.28	42.95	55.84	0.844	0.881	1.009	10.8	24.59	22.02	-0.01	0.27	-0.317	-0.09	4.44	-2.1
Negative	0.834	1.024	0.981	15.04	31.37	63.22	0.902	1.012	0.895	26.09	25.82	15.14	0.026	0.192	-0.213	0.2	3.27	-1.79
	CMA			T-stats			*CMA*			T-stats			CMA			T-stats		
Positive	-0.092	-0.029	0.045	-0.84	-0.55	1.32	-0.06	0.304	-0.199	-0.91	3.52	-2.07	0.861	0.79	0.874	16.1	28.61	14.72
	-0.024	-0.051	0.055	-0.24	-0.91	1.38	-0.189	-0.061	-0.025	-1.04	-0.73	-0.23	0.871	0.797	0.907	16.6	23.77	16.1
Negative	-0.045	-0.014	0.084	-0.35	-0.19	2.33	0.141	0.063	-0.378	1.75	0.69	-2.75	0.9	0.8	0.882	15.62	25.71	15.13
	RMW			T-stats			*RMW*			T-stats			*RMW*			T-stats		
Positive	-0.093	0.058	-0.005	-1.03	1.32	-0.18	0.118	0.151	0.023	2.15	2.11	0.29	-0.479	-0.169	0.027	-8.69	-7.12	0.33
	-0.173	0.013	0.007	-2.06	0.28	0.2	0.012	0.164	0.101	0.08	2.37	1.14	-0.527	-0.144	0.167	-9.01	-4.88	2.97
Negative	-0.389	-0.228	0.002	-3.63	-3.62	0.06	0.155	0.152	0.028	2.33	2.01	0.25	-0.541	-0.17	0.153	-10.1	-6.17	2.66
	MOM			T-stats			*MOM*			T-stats			*MOM*			T-stats		
Positive	-0.258	-0.136	0.009	-7.37	-8.05	0.8	0.018	-0.113	-0.178	0.84	-4.1	-5.79	-0.097	0.08	0.248	-0.88	1.14	1.73
	-0.254	-0.161	-0.001	-7.86	-9.01	-0.1	0.013	-0.135	-0.262	0.23	-5.06	-7.69	-0.105	0.135	0.361	-0.99	1.95	2.6
Negative	-0.367	-0.306	-0.001	-8.9	-12.6	-0.05	0.037	-0.227	-0.382	1.45	-7.77	-8.69	-0.201	0.064	0.408	-1.72	0.81	2.92

Table 1.5.: Regression results for triple sorts on *size*, *bm* and *pr*

This table reports monthly time-series regressions (276 months) of *TONE* on *EX_RET* for the 6 *size-bm* and *size-MOM*. Stocks are allocated to two *size* groups (Small and Big) using the NYSE median as the market cap break-point. Small and big stocks are allocated independently to three *bm* or *pr* groups (Low *bm* to High *bm* or Low *pr* to High *pr*). The intersections of the two sorts produce 6 *size-bm* or *size-MOM* portfolios. The LHS variables in the 6 regressions are the excess returns on the portfolios. Size is calculated as the monthly closing stock price times shares outstanding and the groups are formed based on NYSE break points, where Small stocks are in the 30^{th} percentile and Big stocks are in the 70^{th} percentile range. *bm* is calculated as the ratio of annual book-equity and market-equity, again the groups are formed based on NYSE break points, where Low-*bm* stocks are in the 30^th percentile and Hi-*bm* stocks are in the 70^{th} percentile range. *pr* is a rolling 12-month; prior returns of each stock in a given month and the three groups are formed annually based on 30^{th} and 70^{th} percentile range. *EX_RET* is defined as the annual (calendar year) stock returns in excess of one-year U.S. Treasury bill rate. *TONE* is constructed as the difference in value weighted stock returns between the firms that are in top 30% of the *TONE* (defined as the difference between positive words and negative words for a given stock every year) every year and the firms in bottom 30% of *TONE*. *SMB* is the difference between the returns on the three small portfolios minus the average return on the three large portfolios. *HML* is the average return on the two value portfolios minus the average return on the two growth portfolios. $R_m - R_f$ is the excess market return. *MOM* is computed using prior year cumulative returns for the months $t-2$ through $t-12$. *RMW* is the average return on the two high operating profitability portfolios minus the average return on the two low operating profitability portfolios. *CMA* is the average return on the two low investment portfolios minus the average return on the two high investment portfolios. Panel A reports the regressions for portfolios formed on *size-bm*, Panel B reports the regressions for portfolios formed on *size-pr*. The standard errors are clustered robust heteroscedasticity-adjusted. The coefficients and t-statistics are reported.

Panel A: *size-bm* portfolios

	Small-Low *bm*		Small- Medium *bm*		Small- High *bm*		Large- Low *bm*		Large- Medium *bm*		Large- High *bm*	
Parameter	Estimate	t-Value	Estimate	t-Value	Estimate	t-Value	Estimate	t-Value	Estimate	t-Value	Estimate	t-Value
Alpha	0.091	0.54	0.297	2.85	0.57	4.07	0.162	2.16	-0.027	-0.27	-0.043	-0.48
TONE	-0.715	-2.14	-0.276	-1.87	-0.438	-2.6	-0.016	-0.17	0.196	1.97	-0.091	-0.97
SMB	0.926	12.73	0.831	21.41	0.746	15.47	0.25	6.54	0.231	5.88	0.262	6.56
HML	-0.417	-5.01	0.074	1.07	0.312	3.82	-0.229	-5.78	0.238	3.71	0.623	11.12
RMW	-0.331	-3	0.032	0.53	-0.113	-1.33	0.055	1.19	0.319	5.35	0.138	2.65
CMA	-0.212	-1.59	0.019	0.23	0.05	0.45	-0.179	-2.8	0.206	2.98	0.048	0.64
$R_m - R_f$	0.88	12.54	0.86	23.74	0.769	16.83	1.051	38.24	1.094	32.18	1.083	34.24
MOM	-0.271	-4.77	-0.188	-5.6	-0.143	-3.92	-0.177	-7.43	-0.18	-6.34	-0.178	-5.46

Panel B: *size-pr*portfolios

	Small- Weak *pr*		Small- Average *pr*		Small- Strong *pr*		Big- Weak *pr*		Big- Average *pr*		Big- Strong *pr*	
Parameter	Estimate	t-Value	Estimate	t-Value	Estimate	t-Value	Estimate	t-Value	Estimate	t-Value	Estimate	t-Value
Alpha	0.367	1.7	0.327	3.69	0.471	4.67	-0.036	-0.37	0.02	0.27	-0.018	-0.25
TONE	-0.797	-2.39	-0.155	-1.56	-0.258	-2.07	0.025	0.23	0.213	3.09	0.201	3.08
SMB	0.861	11.11	0.698	22.1	0.854	16.23	0.267	6.19	0.203	6.46	0.38	12.1
HML	-0.114	-0.9	0.22	3.81	0.057	1.12	-0.002	-0.05	0.189	4.67	0.036	0.86
RMW	-0.258	-2.04	0.097	1.64	-0.147	-1.75	0.123	2.1	0.286	7.47	-0.033	-0.77
CMA	-0.078	-0.46	0.042	0.59	-0.001	-0.02	0.029	0.4	0.132	2.3	0.004	0.08
$R_m - R_f$	0.813	11.08	0.792	27.1	0.929	22.9	1.144	37.3	1.003	41.91	1.08	47.86
MOM	-0.621	-9.65	-0.094	-3.59	0.23	8.77	-0.689	-26.72	-0.161	-7.8	0.32	11.52

Table 1.6.: Stock-level regressions of *TONE* on *EX_RET*

This table reports the cross-sectional regressions of *TONE* on *EX_RET* for all the stocks during the sample period 1994 -2016. We stack all the stocks' excess returns into one panel and run a first-pass time series regression with average value weighted monthly *EX_RET* as the dependent variable for each stock across the 273 months during our sample period. We use the first-pass OLS estimates of betas and estimate the coefficients in a second-pass cross-sectional regression. *EX_RET* is the monthly stock returns in excess of one-month U.S. Treasury bill rate. *TONE* is constructed as the difference in value-weighted (by *size*) stock returns between the firms that are in top 30% of the *pessimism* every month and the firms in bottom 30% of *pessimism*. In the second column we present the estimates from their second pass stock-level regression model using the Newey- West adjusted version of the Fama and McBeth (1972) regressions. In the fourth column we also present the Fama and McBeth (1972), weighted least squares (WLS) regressions using $t-1$ stock returns as the weighting variable as suggested by Asparouhova et al. (2010). In the sixth column we present the two-pass estimation method as in Ang et al. (2019) on a stock by stock basis. *SMB* is the difference between the returns on the three small portfolios minus the average return on the three large portfolios. *HML* is the average return on the two value portfolios minus the average return on the two growth portfolios. $R_m - R_f$ is the excess market return. *MOM* is computed using prior year cumulative returns for the months $t-2$ through $t-12$. *RMW* is the average return on the two high operating profitability portfolios minus the average return on the two low operating profitability portfolios. *CMA* is the average return on the two low investment portfolios minus the average return on the two high investment portfolios. Alpha is the intercept of the model. The standard errors are robust heteroscedasticity-adjusted.

	Fama-MacBeth (OLS)		Fama-MacBeth (WLS)		Ang-Liu-Schwarz (OLS)	
Variable	Estimate	t-stats	Estimate	t-stats	Estimate	t-stats
Intercept	-0.082	-1.65	-0.03	-1.598	-0.21	-519.28
TONE	1.513	1.84	1.565	1.892	0	2.5
SMB	0.002	0.03	0.054	0.082	0.022	1.79
HML	-0.016	-0.45	0.036	0.398	-0.032	-2.61
RMW	0.002	0.15	0.054	0.202	-0.032	-2.45
CMA	-0.008	-0.81	0.044	-0.758	-0.014	-1.75
$R_m - R_f$	0.001	0.43	0.053	0.482	0.037	4.81
MOM	-0.003	-1.68	0.049	1.628	-0.042	-3.36

Table 1.7.: Alternative definition of *TONE*

This table reports the monthly time-series regressions (276 months) of *TONE* on *EX_RET* for decile-based portfolios sorted on the negative SCORE each year in April. SCORE is constructed following the definition of JW (2013). *EX_RET* is the value weighted average monthly stock returns in excess of one-month U.S. Treasury bill rate. Stocks with lowest SCORE comprise Decile 1 and stocks with highest SCORE comprise Decile 10. *TONE* is constructed as the difference in value weighted stock returns between the firms that are in 70^{th} percentile of the SCORE every month and the firms that are in 30^{th} percentile of the SCORE. *SMB* is the difference between the returns on the three small portfolios minus the average return on the three large portfolios. *HML* is the average return on the two value portfolios minus the average return on the two growth portfolios. $R_m - R_f$ is the excess market return. *MOM* is computed using prior year cumulative returns for the months $t-2$ through $t-12$. *RMW* is the average return on the two high operating profitability portfolios minus the average return on the two low operating profitability portfolios. *CMA* is the average return on the two low investment portfolios minus the average return on the two high investment portfolios. Alpha is the intercept of the model. The standard errors are robust heteroscedasticity-adjusted. The t-statistics are in parentheses.

Dependent variable: *EX_RET*

Portfolio	Parameter	*Alpha*	*TONE*	*SMB*	*HML*	*RMW*	*CMA*	$R_m - R_f$	*MOM*	*Adjusted* R^2
Decile 1	Estimate	1.037	0.195	0.431	-0.005	0.131	0.081	-0.04	0.062	0.33
	t Value	-5.77	-1.19	-6.07	(-0.080)	-0.85	-0.87	(-0.720)	-1.7	
Decile 2	Estimate	0.582	0.491	0.698	0.122	0.047	-0.01	0.939	-0.163	0.915
	t Value	-4.5	-5.49	-18.73	-2.2	-0.66	(-0.130)	-23.2	(-6.540)	
Decile 3	Estimate	0.571	0.721	0.719	0.14	-0.081	-0.011	0.99	-0.222	0.927
	t Value	-4.41	-7.45	-17.88	-2.47	(-1.070)	(-0.140)	-28.3	(-6.850)	
Decile 4	Estimate	0.712	0.367	0.7	0.085	-0.07	0.141	0.995	-0.17	0.922
	t Value	-5.5	-3.94	-15.76	-1.56	(-0.950)	-1.64	-25.15	(-4.900)	
Decile 5	Estimate	0.581	0.111	0.748	0.125	-0.086	0.021	0.979	-0.144	0.938
	t Value	-5.2	-1.19	-16	-2.29	(-1.410)	-0.31	-25.77	(-5.710)	
Decile 6	Estimate	0.804	0.033	0.803	0.027	-0.239	0.11	0.976	-0.15	0.941
	t Value	-6.07	-0.33	-17.36	-0.44	(-3.700)	-1.23	-22.59	(-4.300)	
Decile 7	Estimate	0.766	-0.23	0.72	0.022	-0.117	0.026	0.941	-0.199	0.935
	t Value	-5.02	(-2.070)	-14.91	-0.32	(-1.550)	-0.25	-21.33	(-4.720)	
Decile 8	Estimate	0.875	-0.479	0.714	0.087	-0.069	0.033	0.902	-0.144	0.938
	t Value	-5.94	(-4.300)	-17.3	-1.39	(-0.900)	-0.32	-20.45	(-6.020)	
Decile 9	Estimate	0.514	-0.424	0.712	0.123	0.01	0.034	0.936	-0.193	0.945
	t Value	-3.95	(-4.820)	-15.98	-2.3	-0.13	-0.42	-23.78	(-6.400)	
Decile 10	Estimate	0.565	-0.389	0.746	0.116	-0.126	-0.112	0.977	-0.234	0.95
	t Value	-3.82	(-3.470)	-16.33	-1.83	(-1.750)	(-1.250)	-23.15	(-4.880)	

2. PRICE DISCOVERY IN CORPORATE BOND MARKETS

2.1 Introduction

Efficient market hypothesis suggests that the information that is relevant to one market should also affect the other market. Theoretical and anecdotal evidence of informed trading suggests that an informed investor who holds both the debt and the corresponding equity securities of an issuer can maximize his profits (see for example Back, 1993, Kraus and Smith, 1996, Grossman, 1988, Chang and Yu, 2010). Informed traders who have access to private information can choose to trade in debt or equity markets because an informed trader might reveal his information by aggressively trading in stocks or options. Additionally, corporate bond markets are subject to less scrutiny for insider trading. Therefore, we would expect some of the new information about the stock price to be reflected in corporate bond markets first. An important approach in examining how new information is incorporated into security prices is through determining the price discovery across markets. Hasbrouck (1995) pioneered the early price discovery literature, contemporaneously Gonzalo and Granger (1995) proposed an approach to measure the contribution of each market's price impact.

The current paper examines the level of price discovery in stock and corporate bonds in an effort to understand where the informed traders trade and investigate the role of corporate bond market's price discovery using the Hasbrouck's (1995) information shares approach. Our research question is motivated by the strand of literature that provides evidence of significant informed trading in US corporate bond markets prior to important corporate events such as earnings announcements (Wei and

Zhou, 2016), prior to acquisition announcement (Kedia and Zhou, 2014). Similarly, Bodnaruk and Rossi (2013) document that institution holding both equity and debt securities of a given firm benefit significantly from the price appreciation during the MA events.

This paper contributes to the research in various ways. This is the first paper that measures directly the percentage of price discovery of corporate bonds and provide evidence of price discovery in bond markets. Another objective of this study is to investigate the lead-lag correlation between stocks and bonds of a given firm. Since, stocks and bonds account for a dominant share in all traded financial assets, understanding the correlation between equity and corporate debt securities play an important role in investors' diversification, risk management and asset allocation decisions. Although, conventional wisdom and academic studies (Fleming, Kirby and Ostdiek, 2003; Gulko, 2002; Li, 2002 and Hartmann, Straetmans and Devries, 2001) indicate a negative correlation between stock returns and long-term treasuries, there is substantial variation in the relationship between stock and bond returns over the short term. Not surprisingly, given the macro nature of these studies we cannot extrapolate the firm-specific correlations between stocks and bonds. We extend prior work by examining whether the variation in stock-bond return relation is prevalent in a specific type of firm. Our motivation follows from literature on cross-market hedging by Fleming, Kirby and Ostdiek (1998) and Chordia, Sarkar and Subrahmanyam (2001), and stock market uncertainty (Connolly, Stivers and Sun, 2005; Vironesi, 1999).

Most existing studies on joint correlation between stock and bond have taken a traditional, fundamental approach and have examined monthly or annual return data. For example, Campbell and Ammer (1993) discuss several offsetting effects behind correlation between stock and bond returns. In their monthly, return sample over 1952 to 1987 the authors find that the overall correlation changes as macro-economic conditions change. The authors show that changes in real interest rates promotes positive correlation, change in expected inflation promotes negative correlation and expected returns promotes positive correlation. Similarly, previous studies explore the overall linkages between stock and bond markets (Shiller and Beltratti, 1992, Fama and French, 1989, Barsky, 1989 and Keim and Stambaugh, 1986). Further, Scruggs and Glabadanidis (2003) find that the conditional correlation between US stock market and bond market returns has varied considerably over the post-war period. While the aggregate stock-bond correlation in important in asset allocation the relevance of firm specific return relationship is not clear from the existing studies. In the current study, we propose an empirical investigation at a micro level the return correlation between corporate bond securities and equities.

Although, the above-mentioned studies have focused on capturing the informativeness of corporate bond market using the lead-lag relationship, the results are at best inconclusive regarding the lead-lag relationship between the bonds and equity of a given firm. For example, Ronen and Hotchkiss (2002) do not find any lead-lag correlation between stocks and bonds for high-yield firms, whereas Downing, Underwood and Xing (2009) find that high-yield stocks lead bonds. Additionally, as Hasbrouck (1995), Chakravarty et al (2004) and several subsequent studies have pointed out that the information share and the lead-lag relationships capture different aspects of informativeness. For example, the vector error correction model (VECM) generally followed by empirical studies to measure the price discovery focuses on the permanent component of price changes. In contrast, the lead-lag relationship tend to combine the permanent and temporary price changes. Also, Hasbrouck (1995) argues that when cointegration is present in the price series information share provides a precise estimate of price discovery.

Previous studies that have examined the informational role of corporate bond markets are restricted to relatively smaller sample and to high yield firms (for example Ronen and Hotchkiss, 2002; Alexander, Edwards and Ferri, 2000). The recent study by Downing et al (2009) use the over the counter (OTC) bond market data by TRACE during July-2002 and December-2005.[1] In our study by employing a large dataset of corporate bonds traded during the sample period 2002 through 2014, we find that, US corporate bonds have an average information share of 18.68%. We also find that the information share of high-yield bonds is significantly greater than investment grade bonds. It appears that the corporate bond markets have a significant role in price discovery. Additionally, through our daily vector auto-regression (VAR) estimates we find that corporate bonds have a significant informational advantage for investment grade bonds where we find that bond return leads stock returns for high-yield bonds.

The rest of this paper is organized into four sections. Section 2.2 provides the background literature on price discovery and lead-lag studies. Section 2.3 provides details of the data sources and our sample construction. Section 2.4 discuss the summary statistics and the empirical results. Section **??** provides the robustness tests and section **??** provides concluding remarks.

2.2 Background literature and hypotheses

Our paper is related to price discovery in corporate bonds and stocks and informational advantage in bonds. There is limited prior evidence regarding the contribution of corporate bonds in price discovery. The seminal work of Merton (1974) posits that both equity and debt securities are derivatives of the underlying assets of the firm. They imply that both of these securities are sensitive to information about the assets of the firm, thus providing an opportunity for the investor to trade and benefit in these two securities. Several studies have focused on examining whether informed

[1] Asquith, Covert and Pathak (2019) document that FINRA disseminated trade related information through the TRACE data in three phases and the final phase was completed in August 2004. Therefore, most studies that have used TRACE data before April 2004 will not likely have the complete picture of various types of corporate bonds.

traders simultaneously trade in multiple markets. For example, Back (1993), Kraus and Smith (1996), Grossman (1988), Chakravarty et al. (2004), Chang and Yu (2010). Specifically, Back (1993) theoretically show optimal trading strategy in stock and option markets for an informed trader. Chakravarty et al. (2004) using transaction data for stocks listed in NYSE and having options trading in Chicago Board Options Exchange (CBOE) show that informed traders trade in both stocks and options and find that options market has a price discovery of 10% to 20%. Chang and Yu (2010) investigate how firm's capital structure policy is related to informational efficiency of prices. The authors argue that informed traders optimally choose to trade in a firm's stocks or bonds based on the information conveyed by the security prices.[2]

The relatively sparse research on the informational role of corporate bond markets focuses on the lead-lag relationship between debt and equity markets. For example, Alexander, Edwards and Ferri (2000) using daily data of 51 high yield bonds for the sample period 1994 to 1997 study the correlation between the returns on the debt and the returns on stocks of the issuing firms. The authors find a positive correlation between the excess returns of individual firm's bonds with the stock returns. Kapadia and Pu (2012) show that bonds and stocks have a negative correlation in the short run. There are also studies that examine the informational role of corporate bonds using other approach than the lead-lag correlation. For example, Blanco, Brennan and March (2005) test the theoretical arbitrage relationship between credit default swap (CDS) and credit spreads. The authors find that CDS market leads bond market in the determination of credit risk and the CDS market contributes on average around 80% of the price discovery.

[2]Other studies that focus on multi-security trading include Boot and Thakor (1993), Admati (1985), Goldstein, Li and Yang (2013). In a related stream of literature, several studies have focused on the lead-lag relationships between two markets in determining price discovery. For example Manaster and Rendleman (1982), show that call option prices contain fundamental information about stock prices, However, Stephan and Whaley (1990) and Chan, Chung, and Johnson (1993), have analyzed the lead-lag relationship between options and stocks, and do not find any evidence that option markets lead price changes in stock markets.

Evidence of significant informed trading in corporate bonds is provided by previous studies such as Wei and Zhou (2012), Kedia and Zhou (2014), Bodnaruk and Rossi (2013), Acharya and Johnson (2010) and Han and Zhou (2014). Specifically, Wei and Zhou (2012) examine the trading activity in corporate bonds prior to earnings announcements and find a strong evidence of informed trading in bonds. The authors also show that pre-announcement trading activity can predict the earnings surprises. Similarly, Kedia and Zhou (2014) show significant informed trading occur in corporate bonds prior to acquisition announcements. Using mergers and acquisition (M&A) transactions during 1994 and 2006 the authors find that target bonds experience an abnormal trading volume prior to the public announcement. Bodnaruk and Rossi (2013) using a sample of institutional equity holdings during 1999 and 2009 document that often institutions hold both equity and debt securities of a given firm and the institutions benefit significantly from the price appreciation during the MA events. However, none of the above papers focuses on the contribution of corporate bond markets in price discovery.

Investigating the contribution of corporate bonds in price discovery has several practical implications. First, corporate bond markets provide an additional venue to exploit the information and offers profitable trading opportunities. As a result, an informed trader can trade in corporate bonds in addition to stocks and options. Second, the literature on informed trading suggest optimal trading strategies including optimal order size. Extant theoretical research argues that an informed trader tends to trade gradually in order to profit (Kyle, 1985, Admati and Pfleiderer, 1988, Foster and Viswannathan, 1993).[3] So an informed trader trading aggressively (large order size) in stocks or options can reveal the information through his trades (Chakravarty,

[3]Several other papers also examined the impact of trades on prices. See for example, Easley and O'Hara (1987), Easley and O'Hara (1992a), Easley and O'Hara (1992b), Burdette and O'Hara (1987), Holthausen, Leftwich, Mayers (1987), Ball and Finn (1989), Seppi (1990), Hasbrouck (1991a), Hasbrouck (1991b), Grossman (1992), Madhavan and Smidt (1991), Madhavan and Smidt (1993), Chan and Lakonishok (1993), Chan and Lakonishok (1995), Huang and Stoll (1994) and Keim and Madhavan (1996).

2001). Additionally, an informed investor can substitute a certain amount of trading in stocks and options with corporate bonds to maximize the profits. Third debt securities are subject to less scrutiny for insider trading relative to stock or option markets. Hence, we posit that an informed trader chooses to take advantage of the informational role of corporate bonds. In addition, Caballe and Krishnan (1994) highlight this point in a generalized setting, where they imply that it is optimal for an informed investor to trade in all of the correlated security to potentially benefit from the information. Thus, combining the theoretical frameworks of Caballe, and Krishnan (1994) and Merton (1974), we suggest that corporate bonds should have a significant price discovery thus leading lead to our first hypothesis

Hypothesis 1: Corporate bond's information share is significantly greater than zero. Given the large size of corporate bond market and the participation of institutional investors, there is anecdotal evidence of informed traders' preference to high-yield corporate bonds.[4], [5] In addition, it has been documented in previous studies that high-yield bonds have higher sensitivity to firm-specific information relative to other bonds. Therefore, we expect to see a significantly higher information share for high-yield corporate bonds relative to others. In other words, we should expect the informed trader to trade more in the high-yield bonds resulting in such bonds to have relatively higher contribution in price discovery. This argument leads to our next hypothesis:

Hypothesis 2: The information share of high-yield bonds is significantly greater than non-high-yield bonds. The seminal works by Markowitz (1952) and Sharpe (1964) highlight the importance of stock-bond return in determining the diversification benefits and to hedge the common exposures across two asset classes. Since imperfect correlation of asset returns is the key assumption in portfolio theory, there has been a wide interest across researchers and practitioners to investigate the stock-bond return correlation. The finance literature has extensively studied the stock-bond return cor-

[4]A bond rated BB or lower because of its high default risk. Also known as a high-yield bond, or speculative bond.

[5]See https://www.thegentlemansjournal.com/article/story-michael-milken-junk-bond-king/

relation at the aggregate level for example, Shiller and Beltratti (1992) study annual stock-bond returns in USA and UK and argue that according to their rational expectations present value model stock prices drop (rise) when long-term interest rates rise (drop). The authors however find modest correlation between stock prices and changes in inflation. Further, the authors also find that the excess returns in stock market correlate too much with excess bond returns. Campbell and Ammer (1993), using a vector auto-regressive model decompose the excess stock and bond returns on long-term assets. The authors find that the variance of excess stock returns is attributable to changing expectations of future returns. More recent evidence by Connolly, Stivers and Sun (2005,2007) and Gulko (2002) show that future stock-bond correlations decreases with increasing stock market uncertainty in US. The authors argue that flight-to-quality could partially explain this phenomenon. Similarly, Kim, Moshirian and Wu (2006) show that stock market uncertainty drives the international stock-bond correlation. Along similar lines, Guidolin and Timmermann (2005) document that the monthly correlations between UK stock and bond returns are positive and significant in the normal and bull market conditions, whereas the correlations are negative during bear market conditions. Overall, the emerging explanation from previous studies on stock-bond correlation provide mixed findings. As noted by previous literature the stock-bond correlation may be challenging to estimate. Although several factors influence the stock-bond relationship earlier studies [Shiller and Beltratti (1993); Leibowitz, Sorensen, Arnott and Hanson (1989); Li (2002)] focus broadly on macro-economic factors such as real-interest rates, unemployment, inflation and GDP.

However, evidence on firm level correlation between the two asset classes is scant. Exceptions include Kwan (1996), Hotchkiss and Ronen (2002) and Schaefer and Strebulaev (2008). For example, Kwan (1996) in their 1986-1990 sample study investigate the common element of firm-specific information that drives stock and bond prices. The authors study the cross-sectional and contemporaneous correlations between individual stocks and bonds issued by same firm. They find that stock market leads the bond market in transmitting firm-specific information. In contrast, Hotchkiss and Ronen (2002) using a high-frequency data for 55 high-yield bonds over the sample period January through October 1995 document that stocks do not lead bonds. Further, the authors also find that firm-specific information around earnings releases is quickly incorporated in stocks and bonds contemporaneously. Schaefer and Strebulaev (2008) find that the magnitude of the stock-bond correlation is consistent with the Merton (1974) model. Further Bao and Hou (2014) extend the Merton (1974) model to show that corporate bond maturity and credit risk have stronger correlation.

In the current study, we estimate the correlations between stock returns and bond returns of a given firm traded in US markets. Further, we also explore whether institutional traders make use of such correlations in their trading portfolios. Our study identifies the correlations at the micro-level by directly isolating the equity and bond securities of a given issuer. By closely analyzing the stock and bond correlations, we identify the investors' expectations on the portfolio return and understand the diversification effects. Furthermore, we note that the correlation between bond returns and stock returns is critical for institutional investors (pension funds, mutual funds etc.) in their asset allocation decisions. If the bond and stock markets are informationally efficient then we should expect to see a positive correlation between the firm's stock and bond prices. The above discussion leads to our third hypothesis:

Hypothesis 3: There exists a significantly positive relationship between stocks and bonds.

2.3 Data description and sample construction

We utilize five main datasets in this study. The first dataset is the stock market transaction level data for S&P 500 stocks, obtained from Trades and Quote (TAQ). The second dataset is the corporate bond transaction data from Trade Reporting and Compliance Engine (TRACE) through Wharton Research Data Services (WRDS) for the S&P 500 firms. The TRACE data provides over the counter (OTC) corporate bond market real-time prices.[6] To examine the price discovery of bonds in equity prices we use a sample period of over 1,000 trading days from January 2004 through December 2008.

Our third data source is the institutional level transaction data from ANcerno, which provides transactional level trade data for corporate bonds and stocks for the first quarter of 2006 through the third quarter of 2010. Several studies have used equity transaction dataset to examine the ANcerno institutional trading behavior. See for example Puckett and Yan (2011), Bethel, Hu and Wang (2009), Chemmanur, He and Hu (2009), Goldstein, Irvine, Kandel and Wiener (2009). Additionally, Hu, Jo, Wang and Xie (2018) provide a comprehensive review of ANcerno dataset. The fourth source of data comes from Mergent Fixed Income Security Database (FISD), which provides details of bond characteristics and credit ratings from standard and poor's (S&P) and Moody's. Finally, we obtain the daily stock returns data from center for security prices (CRSP) database and match it with the daily bond returns to examine the lead-lag relationships.

2.3.1 *Price discovery through information share approach*

Hasbrouck (1995) estimates the information share to examine stocks listed on multiple exchanges and he determines that New York Stock Exchange accounts for substantial amount of price discovery. Several empirical works have implemented the information share approach in multi-security setting. For example, Chakravarty, Gulen and Mayhew (2004) investigate the contribution of equity option market's price discovery. These authors show that option markets have a significant (17% on an average) price discovery. Hasbrouck (2003) examine the price discovery across index futures and exchange traded funds and find a significant price discovery in futures

[6]We follow the data filtering process suggested by Dick-Nielsen (2014) to clean the enhanced TRACE data.

market. However, none of these studies estimates the contribution of corporate bond markets. Given the large size of US corporate bond markets and previous evidence of informed trading in corporate bonds we propose to fill the gap in the literature by directly examining the role of corporate bonds in price discovery. [7]

Price discovery is known as a process that efficiently incorporates the new information into asset prices. The information share (IS) approach is based on a common implicit efficient price that is contained in the observed price of a security and can be estimated using a vector error correction model (VECM) framework. Hasbrouck (1995) provides an econometric method to estimate the contribution of securities traded in multiple markets. He has documented in the context of stocks trading in multiple exchanges. The application of the estimation method through SAS program is provided in his website. Subsequently, several studies have used the Hasbrouck's information share approach to determine the role of a given market in a multi-security setting.[8] IS measures the variance of the efficient price, and identifies the proportion of the efficient price variance that can be attributed to different markets. For example, in the current study the IS measures the variance of the stock and bond prices of a given firm. Specifically, IS of a given market is the percentage contribution of the variance of the given asset.

Formally, Hasbrouck (1995) defines IS as:

[7] Based on 2019 article ("U.S. Corporate Debt Market: The State of Play In 2019") by S&P Global, the outstanding US corporate debt market is $9.2 trillion. In comparison, the US equity market is approximately $32 trillion. See for example Back and Crotty (2015), Bittlingmayer and Moser (2014), Zhou (2005) for evidence of informed trading in corporate bonds.

[8] See Chakravarty et al. (2004), Eun and Sabherwal (2003), Kryzanowski and Lazrak (2011), Chen and Choi (2012)

Let $p_{i,t} = (p_{1,t}, p_{2,t})'$ denote a vector of (log) prices of stocks and bond prices of an issuer. In case of multiple bonds for a given issuer at a given time, we consider the volume weighted average price across all the bonds for the issuer.

$$p_{i,t} = (p_{1,t}, p_{2,t}) \sim I(1)$$

Since the prices represent the same assets. The two prices are cointegrated and the linear relationship is expressed as:

$$p_{i,t} = \beta p_{2,t} + \mu_t$$

Where $\mu_t \sim$ I(0) and β is the cointegrating vector, $\beta = (1, -\beta)'$. The vector error correction model (VECM) representation for 'k'' lags is

$$\Delta p_t = \alpha\beta' p_{(t-1)} + \Gamma_1 \Delta p_{(t-1)} + \cdots + \Gamma_k \Delta p_{(t-k)} + \epsilon_t$$

where α represents the coefficient associated with the error correction term. ϵ is a 2 x 1 vector of the residuals with $\epsilon_t \sim N(0, \Omega)$. The price changes are covariance stationary, the vector moving average (VMA) is

$$\Delta p_t = \Psi(L)\epsilon_t$$

Hasbrouck (1995) shows that since both the price series represent the same asset, the long run impact of ϵ_t on each of the price series should be the same. Denote $\Psi(1) = (\Psi 1, \Psi 2)$ as the common row vector of $\Psi(1)$. $\Psi\epsilon_t$ is the incremental change in price that is impounded into the security prices. The author proposes the use of the structure of the variance of this component to derive the price discovery. The variance of $\Psi\epsilon_t$.

$$VAR = (\Psi\epsilon_t) = (\Psi\Omega\Psi')$$

If Ω is diagonal, then $\Psi\Omega\Psi'$ will consist of 'n' terms, each of which would represent contribution to the efficient price innovation from each market. The proportion of the variance of the efficient price that can be attributed from an innovation from market 'j', is formally defined as

$$IS_j = \frac{\Psi_j^2 \Omega_{jj}}{\Psi\Omega\Psi'}$$

Where Ψ_j is the j^{th} element of the Ψ vector and Ω_{jj} is the variance of the j^{th} market. We follow the standard estimation method and compute the information share bounds are every month for each firm using intra-day transactions data. Monthly estimates are then averaged across time for each stock.

2.4 Results

2.4.1 *Estimating information share*

We estimate the two-market (stock and bond) IS at the firm-month level. The Hasbrouck's (IS) methodology computes the upper and lower bounds when the information across the markets are correlated. Panel A of Table 2.1 reports the distributional statistics of the corporate bonds used in our analyses. From Panel A we find that there are 42,000 distinct bonds in our sample over 3,170 firms, with an average of 12 bonds per issuer. We also find that that our sample has about 55% firms consisting of investment grade firms (those with a credit rating for A- or above). Further, we find that the average yield of investment grade (high yield) bonds is 7% (18%), while the overall average is 9%. To put this in perspective, Holden, Mao and Nam (2018) using NYSE bond data report that 67% of their sample consists of investment grade bonds. Bittlingmayer and Moser (2014) report an average yield of 5.74% in their 2002-2008 sample. Similarly, Tolikas (2017) report an average yield of 6.45% for investment grade bonds and 10% for high yield bonds. Overall, we find that the firms in our dataset have similar characteristics to those reported by above-mentioned studies.

In Panel B, we report the average IS for the full sample and over time. We find that the mean IS is 18.68% and is statistically significant different from zero at 1% level of significance. The median IS in Panel B is about 1.55%, which is also statistically significant from zero, based on Wilcoxon signed rank test. Thus, in Panel B, we find that the mean (and median) information share for the corporate bonds are significant at 1% level consistent with our first hypothesis of significant price discovery in US corporate bonds. In Panel C, we report the IS for a sub-sample of high yield firms. We classify the high yield bonds based on the issuer's credit rating at a given time. The high yield bonds are those with a credit rating of A- or above. We find that the mean IS for high yield bonds are significantly higher relative to full sample with an average information share of 33% compared to 18.68%. In Panel B we also find that the median IS for high-yield bonds is 9% and statistically significant at 1% level. The results in Panels A and B suggests that the US corporate bond markets provide significant price discovery and the bonds with lower credit rating have higher informed trading than the investment grade bonds. We find a significantly higher IS compared to previous studies that have examined the price discovery between multi-market of a given issuer. For example, Chakravarty et al. (2004) report an average information share of 17% in option markets. Kryzanowski, Perrakis and Zhong (2017) report a price discovery of 50% in CDS markets. Our results support the view that corporate bonds have significant informed trading.

Overall, the results in Table 2.1 suggest that US corporate bonds have significant contribution in price discovery. Additionally, our evidence suggests that high-yield bonds have higher informed traders, resulting in a significantly higher (relative to investment grade bonds) price discovery. Therefore, the results from Table 2.1 confirms to our first and second hypothesis discussed in section 2.2.

2.4.2 *Estimating the lead-lag return relationship*

We now turn to examining the lead-lag relationship between corporate bonds and stock returns for a given firm. To do so, we employ the bivariate vector autoregressions for stock and bond returns over a sample period of 2002 through 2014. Following previous literature such as Hotchkiss and Ronen (2002), Downing, Underwood and Xing (2009) and several others we estimate the following vector autoregressive system:

$$z_{j,t} = c_j + \sum_{i=1}^{L} b_{i,j} R_{b,t-i,j} + \sum_{i=1}^{L} s_{i,j} R_{s,t-i,j} + \epsilon_{j,t} \tag{2.1}$$

Where $z_{j,t}$ is the vector $[R_{(B,t,j)}, R_{(S,t,j)}]'$, $R_{(B,t,j)}$ is the return on bond j at time t, and $R_{(S,t,j)}$ is the return on stock 'j' at time t. [9] The slope coefficients are b_i and s_i associated with the lagged bond and stock returns. The lag-length is 5 for the daily returns and 1 for the weekly returns. The choice of lag-lengths is based on Akaike Information Criterion (AIC) and suggestions by previous studies such as Back and Crotty (2015). In addition, for robustness, we also estimate alternative lag-lengths and we find that our results are not driven by changes in lag-lengths.

In order to calculate the daily bond returns we follow the standard approach of using the last bond trade in a given day. Following studies mentioned above we assume a zero return for trading intervals where no trades occur. Similarly, for our institutional trading data we calculate the weekly bond returns using last trade of the week approach. In addition, we also require a bond to trade at least twice in a given week. We obtain the daily and weekly stock returns from CRSP. Our null hypothesis is that the bond and stock markets are equally efficient and react to similar information. In Table 2.2, we start by estimating the VAR across all the firms during our sample. The results in Table 2.2 exhibit a substantial relationship between equity and bond returns for the high yield firms. For example, in Panel A the coefficient

[9] For firms with multiple outstanding bonds we calculate the volume weighted returns across all the bonds for a given firm consistent with Kalimipalli and Nayak (2014) and Bessembinder, Kahle, Maxwell and Xu (2008).

estimate on the first three lags of stock returns is positive and significant at 1% level. The estimates are 0.029 (t-stat = 3.65), 0.044 (t-stat = 5.56) and 0.019 (t-stat = 2.37) respectively. In Panel A we also find that the coefficient estimates of bond returns for lags 1, 4 and 5 are positive and statistically significant with estimates of 0.001 (t-stat = 2.56), 0.001 (t-stat = 2.84) and 0.002 (t-stat = 4.62) respectively. However, the coefficient estimates in Panel B for the bond returns in the investment grade firms are insignificant. The results in Table 2 suggests that for high yield bonds equity returns lead bond returns and it appears that the high-yield corporate bonds act like equity and have a positive and significant correlation with equity securities. In contrast, the investment grade bonds behave like treasury securities and do not show any significant correlation with equity.

From our two tests on the informational role of corporate bond market, we find evidence of significant price discovery in the corporate bond markets, however we do not find that corporate bond returns lead stock returns but for a group of high-yield bonds in our sample. That is high yield bonds in our sample appear to have a significant price discovery as well as incorporate new information before the stocks. However for the investment grade bonds, although we determine a significant price discovery there is little evidence to suggest that the new information is incorporated in the bond prices before the stock prices. To reconcile the results we offer some of the plausible explanations for these findings.

First, our results in this section hinges on the evidence that there are several empirical and anecdotal evidence regarding the informed trading in high yield firms relative to investment grade firms. For example, Han and Zhou (2014) argue that institutional investors predominantly trade high yield bonds; therefore, the possibility of informed trading is higher in high yield bonds relative to investment grade bonds. Similarly, Datta and Iskandar-Datta (1996) argue that corporate bond markets provides an opportunity for insider trading in the absence of any regulatory reporting requirements of bond transactions. Consistent with this argument, former Securities and Exchange Commission (SEC) chairperson Arthur Levitt stated that SEC has found evidence of "possible misuse of inside information in the high yield debt market". Anecdotal evidence such as, the recent issue of Six Flag's Inc.'s bankruptcy case and the Delphi Corporation case suggests that there is significant informed trading in the low-rated (high yield) bonds.[10] Additionally several empirical studies that focuses on corporate announcements have documented that significant informed trading is mainly prevalent in high-yield bonds during pre- earnings announcement and prior to take over announcements (Wei and Zhou, 2016; Kedia and Zhou, 2014).

Hendershott, Kozhan and Raman (2019) argue that the 2007 financial crisis post Lehman Brothers collapse increased risk associated with the high-yield bonds and pushed such bonds closer to default. [11] Therefore, a very large potential profitability in high-yield bonds provided the short sellers a greater opportunity to take advantage of the information due to higher informational sensitivity and reduced competition. Further, the theoretical argument by Merton (1974) that corporate debt is valued as a risky free debt and a short position in put option on the firm's equity. Therefore any information associated with the firm should affect its bonds values, thus the relevance of information for bondholders depends on the probability of default of the firm. In other words, bondholders make use of the potential informational advantage when the firm is closer to default. Therefore, the degree of informed trading in bonds are inversely correlated with the bond's credit risk.

Therefore, despite the evidence of significant informed trading in corporate bond market, there is little evidence to support that the new information is incorporated

[10]In the Six Flag's Inc.'s case a hedge fund sold low-rated bonds of Six Flags after obtaining information regarding the firm's reorganization plan. Similarly, in 2008 Delphi Corporation accused institutional investors of insider trading and alleged that some of the institutional investors shorted the bonds after obtaining confidential information of the firm's bankruptcy financing.

[11]The Lehman bankruptcy increased funding costs and led to greater price distortions (see for example, Brunnermeier, 2009; Mitchell, Pedersen and Pulvino, 2007; Mitchell and Pulvino, 2012)

first in the bond returns before stock returns in a lead-lag analysis. However, for the reasons mentioned above investors in high yield bond markets potentially benefit from the informational advantage. Previous evidence on the lead-lag relationship is mixed, for example, Ronen and Zhou (2013), Hotchkiss and Ronen (2002) and Kedia and Zhou (2014) support that bond markets are as informationally efficient as equity markets. In contrast, Kwan, Alexander, Edwards and Ferri (2000) and Downing, Underwood and Xing (2009) conclude that stock markets lead bond markets. However, Ronen and Zhou (2013) and Wei and Zhou (2016) have expressed concerns that VAR models do not capture the lead-lag relationships efficiently. Thus our results for he high yield bond markets is consistent with the view of Ronen and Zhou (2013), Hotchkiss and Ronen (2002) and Kedia and Zhou (2014).

2.5 Robustness – Component Share approach

Hasbrouck's information share approach has been a widely used measure of price discovery. However, an alternative approach known as component share (CS) is also in vogue to measure the price discovery of multi-security of a given issuer. Several studies such as Booth, So, and Tse (1999), Chu, Hsieh, and Tse (1999), Harris, McInish, and Wood (2002) adopt the decomposition technique proposed by Gonzalo and Granger (1995), which focuses on the composition of the price impact of one market's contribution to the other using component weight of a given market. Although the IS and CS differ in their approach both methods use cointegration technique to restrict multiple market price information to share an efficient price. In addition, both of these approaches use the commonly applied vector error correction model (VECM) for estimating the price discovery. Under the CS approach, the price (p_t) of a given issuer in equity and bond markets takes the following form:

$$p_t = V_1 f_t + V_2 z_2$$

where, $f = \gamma' p_t \; I(1)$is the permanent component and the $z_t \; I(0)$ is the transitory component, whereas V_1 and V_2 are the loading matrices. Gonzalo and Granger (1995) defined γ as $\gamma = (\alpha'\beta)^{-1}\alpha$ such that $\alpha'\alpha = 0$ and $\beta'\beta = 0$. Gonzalo-Granger and several other studies that are mentioned above formally define CS of a market 'j' as follows:

$$CS_j = \frac{\alpha_j}{\alpha_j + \beta_j}, j = 1, 2$$

Similar to section 2.4.1 we estimate the CS measure for firms in sample every month between the equity and bond prices. In Panel A of Table 2.3 we find that the average CS for the full sample is 13% and is statistically significant different from zero at 1% level of significance. The median CS in Panel A is about 2.20%, which is also statistically significant from zero, based on Wilcoxon signed rank test. Thus, in Panel A, we find that the mean (and median) information share for the corporate bonds are significant at 1% level consistent with the Hasbrouck's IS measure.

In Panel B, we report the CS for a sub-sample of high yield firms. We classify the high yield bonds based on the issuer's credit rating at a given time. The high yield bonds are those with a credit rating of A- or above. As in IS measure in Table 2.1, here too we find that the mean CS for high yield bonds are significantly higher relative to full sample with an average information share of 27% compared to 13%. In Panel B we also find that the median CS for high-yield bonds is 16.10% and statistically significant at 1% level. The results in Panels A and B suggests that the US corporate bond markets provide significant price discovery and the bonds with lower credit rating have higher informed trading than the investment grade bonds. In addition, the results from the alternative measure of price discovery is consistent with those reported in Table 2.1. Overall, the results in Table 2.3 suggest that US corporate bonds have significant contribution in price discovery and the consistent results from the alternative measure of component share approach provides confidence on our main results.

2.6 Conclusion

In the current study, we investigate the daily stock returns and corporate bond returns for securities traded in US markets. As noted by previous studies that stock and bond return correlations play an important role in diversification and asset allocation decisions. Various studies have focused extensively on aggregate levels of stock and bond return relation [for example Campbell, 1987; Shiller and Beltratti, 1992 and Campbell and Ammer, 1993 among others. However, there is very scant research that has focused on the contemporaneous correlations between stock and bonds of a same issuer. This cross sectional and time series correlation provides insights into firm-specific information flows and investor preference for one asset over another in the light of asymmetric information and risk uncertainty. We aim to examine the cross-market correlation between daily stock returns and bond returns using a broader sample of firms that issues stocks and bonds over our sample period between 2002 and 2012.

A secondary purpose of this study is to opine on where the portfolio diversification and other strategic trading occurs across the two markets. Not surprisingly, given the aggregate nature of previous studies there do not appear to exist a consensus on what the correlations are between stocks and bonds of a same issuer. We propose a micro approach to the question of trading across publicly traded companies' stocks and bonds both by all classes of traders and by institutions only to see how frequent such trading is and what kind of stocks are particularly susceptible to such behavior. Doing so allows us to make sharper observations about the kind of investor behavior alluded to by past studies.

Table 2.1.: Descriptive statistics of information share

This table presents the corporate bond market characteristics and the average information share for the firms used for our analyses. Panel A presents the bond characteristics used in our sample. Panels B and C present the mean, median, min, max and standard deviation (SD) of Hasbrouck's information share for the full-sample and high-yield firms respectively. *,** and *** mean significant at 10%, 5% and 1% respectively, based on the t-test for the means and Wilcoxon signed rank test for the medians

Panel A : Bond Characteristics	Overall	Investment grade	High-Yield
Number of bonds	42,000	23,100	18,900
Number of firms	3,170	1,744	1,426
Average trade size ($ millions)	647	720	550
Average yield (%)	9	7	18

Panel B : Information share : Full sample	Overall	2004	2005	2006	2007
Mean	18.68%***	35.06%***	7.85%***	18.01%***	15.23%***
Median	1.55%***	1.69%***	0.53%***	2.54%***	1.35%***
Min	0.00%	0.00%	0.00%	0.00%	0.00%
Max	99.00%	99.00%	99.00%	99.00%	99.00%
SD	32.06%	44.84%	20.86%	28.64%	27.82%

Panel C : Information share : High Yield	Overall	2004	2005	2006	2007
Mean	33.00%***	46.00%***	32.00%***	37.00%***	22.00%***
Median	9.00%***	10.00%***	7.00%***	12.00%***	9.00%***
Min	0.00%	0.00%	0.00%	0.00%	0.00%
Max	99.00%	99.00%	99.00%	99.00%	99.00%
SD	17.00%	21.00%	8.00%	19.00%	21.00%

Table 2.2.: Lead-lag relationship between stock and bond returns

This table presents the present the coefficient estimate and the corresponding t-stat for the multivariate vector autoregressions (VAR) for the daily stock and bond returns. Panel A consists of high-yield firms and Panel B presents the investment grade firms. We report only the estimates for lagged bond returns (b1 through b5) and lagged stock returns (s1 through s5). *,** and *** mean significant at 10%, 5% and 1% respectively, based on the t-test for the means and Wilcoxon signed rank test for the medians

	Lagged bond returns					Lagged stock returns				
	b1	b2	b3	b4	b5	s1	s2	s3	s4	s5
Panel A : High Yield firms										
Estimate - Stock	0.001	0.000	0.000	0.001	0.002	-0.053	-0.039	-0.004	-0.019	-0.025
T-stat	2.560	0.050	0.900	2.840	4.620	-29.020	-21.710	-2.410	-10.220	-13.580
Estimate - Bond	-0.141	0.129	0.186	0.180	0.152	0.029	0.044	0.019	0.000	0.018
T-stat	-78.540	72.120	105.090	101.070	84.500	3.650	5.560	2.370	0.030	2.230
Panel B :Investment grade firms										
Estimate - Stock	0.254	0.124	-0.256	0.110	0.052	-0.042	-0.102	-0.030	0.042	0.005
T-stat	0.980	0.412	-0.780	0.386	0.254	-0.742	-1.760	-0.523	0.723	-0.081
Estimate - Bond	-0.116	0.010	0.040	0.030	0.027	0.000	-0.009	0.013	0.004	0.022
T-stat	-1.970	0.176	0.672	0.462	0.482	0.024	-0.724	1.130	1.134	0.188

Table 2.3.: Robustness : Component Share

This table presents the mean, median, min, max and standard deviation (SD) of component share (CS) for the full-sample and the high-yield firms respectively. *,** and *** mean significant at 10%, 5% and 1% respectively, based on the t-test for the means and Wilcoxon signed rank test for the medians

Full - Sample	High-Yield
13.0%***	27.2%***
2.2%***	16.1%***
0.0%	3.0%
100.0%	100.0%
22.5%	26.7%

REFERENCES

REFERENCES

[1] A. R. Admati and P. C. Pfleiderer, "Disclosing information on the internet: is it noise or is it news," *Technical Report, Graduate School of Business, Stanford University*, 2001.

[2] R. B. Barsky, "Why don't the prices of stocks and bonds move together," *The American Economic Review*, vol. 79, no. 5, p. 1132, 1989.

[3] H. Bessembinder, K. M. Kahle, W. F. Maxwell, and D. Xu, "Measuring abnormal bond performance," *The Review of Financial Studies*, vol. 22, no. 10, pp. 4219–4258, 2008.

[4] J. Y. Campbell and J. Ammer, "What moves the stock and bond markets? a variance decomposition for long-term asset returns," *The Journal of Finance*, vol. 48, no. 1, pp. 3–37, 1993.

[5] M. M. Buehlmaier and T. M. Whited, "Are financial constraints priced? evidence from textual analysis," *The Review of Financial Studies*, vol. 31, no. 7, pp. 2693–2728, 2018.

[6] K. Back, "Asymmetric information and options," *The Review of Financial Studies*, vol. 6, no. 3, pp. 435–472, 1993.

[7] A. Kraus and M. Smith, "Heterogeneous beliefs and the effect of replicatable options on asset prices," *The review of financial studies*, vol. 9, no. 3, pp. 723–756, 1996.

[8] E. C. Chang, J. W. Cheng, and Y. Yu, "Short-sales constraints and price discovery: Evidence from the hong kong market," *The Journal of Finance*, vol. 62, no. 5, pp. 2097–2121, 2007.

[9] J. Gonzalo and C. Granger, "Estimation of common long-memory components in cointegrated systems," *Journal of Business & Economic Statistics*, vol. 13, no. 1, pp. 27–35, 1995.

[10] S. Kedia and X. Zhou, "Informed trading around acquisitions: Evidence from corporate bonds," *Journal of Financial Markets*, vol. 18, pp. 182–205, 2014.

[11] A. Bodnaruk and M. Rossi, "Dual ownership, returns, and voting in mergers," *Journal of Financial Economics*, vol. 120, no. 1, pp. 58–80, 2016.

[12] J. Fleming, C. Kirby, and B. Ostdiek, "The economic value of volatility timing using "realized" volatility," *Journal of Financial Economics*, vol. 67, no. 3, pp. 473–509, 2003.

[13] L. Gulko, "Decoupling," *The Journal of Portfolio Management*, vol. 28, no. 3, pp. 59–66, 2002.

[14] P. Hartmann, S. Straetmans, and C. d. Vries, "Asset market linkages in crisis periods," *Review of Economics and Statistics*, vol. 86, no. 1, pp. 313–326, 2004.

[15] T. Chordia, A. Sarkar, and A. Subrahmanyam, "Common determinants of bond and stock market liquidity: the impact of financial crises, monetary policy, and mutual fund flows," Staff Report, Tech. Rep., 2001.

[16] R. Connolly, C. Stivers, and L. Sun, "Stock market uncertainty and the stock-bond return relation," *Journal of Financial and Quantitative Analysis*, vol. 40, no. 1, pp. 161–194, 2005.

[17] P. Veronesi, "Stock market overreactions to bad news in good times: a rational expectations equilibrium model," *The Review of Financial Studies*, vol. 12, no. 5, pp. 975–1007, 1999.

[18] R. J. Shiller and A. E. Beltratti, "Stock prices and bond yields: Can their comovements be explained in terms of present value models?" *Journal of Monetary Economics*, vol. 30, no. 1, pp. 25–46, 1992.

[19] E. F. Fama and K. R. French, "Business conditions and expected returns on stocks and bonds," *Journal of financial economics*, vol. 25, no. 1, pp. 23–49, 1989.

[20] D. B. Keim and R. F. Stambaugh, "Predicting returns in the stock and bond markets," *Journal of financial Economics*, vol. 17, no. 2, pp. 357–390, 1986.

[21] J. T. Scruggs and P. Glabadanidis, "Risk premia and the dynamic covariance between stock and bond returns," *Journal of Financial and Quantitative Analysis*, vol. 38, no. 2, pp. 295–316, 2003.

[22] E. S. Hotchkiss and T. Ronen, "The informational efficiency of the corporate bond market: An intraday analysis," *The Review of Financial Studies*, vol. 15, no. 5, pp. 1325–1354, 2002.

[23] C. Downing, S. Underwood, and Y. Xing, "The relative informational efficiency of stocks and bonds: An intraday analysis," *Journal of Financial and Quantitative Analysis*, vol. 44, no. 5, pp. 1081–1102, 2009.

[24] S. Chakravarty, H. Gulen, and S. Mayhew, "Informed trading in stock and option markets," *The Journal of Finance*, vol. 59, no. 3, pp. 1235–1257, 2004.

[25] G. J. Alexander, A. K. Edwards, and M. G. Ferri, "The determinants of trading volume of high-yield corporate bonds," *Journal of Financial Markets*, vol. 3, no. 2, pp. 177–204, 2000.

[26] P. Asquith, T. Covert, and P. Pathak, "The effects of mandatory transparency in financial market design: Evidence from the corporate bond market," National Bureau of Economic Research, Tech. Rep., 2013.

[27] J. Wei and X. Zhou, "Informed trading in corporate bonds prior to earnings announcements," *Financial Management*, vol. 45, no. 3, pp. 641–674, 2016.

[28] V. V. Acharya and T. C. Johnson, "More insiders, more insider trading: Evidence from private-equity buyouts," *Journal of Financial Economics*, vol. 98, no. 3, pp. 500–523, 2010.

[29] S. Han and X. Zhou, "Informed bond trading, corporate yield spreads, and corporate default prediction," *Management Science*, vol. 60, no. 3, pp. 675–694, 2013.

[30] T. Chordia, A. Sarkar, and A. Subrahmanyam, "An empirical analysis of stock and bond market liquidity," *The Review of Financial Studies*, vol. 18, no. 1, pp. 85–129, 2004.

[31] S. Das, M. Kalimipalli, and S. Nayak, "Did cds trading improve the market for corporate bonds?" *Journal of Financial Economics*, vol. 111, no. 2, pp. 495–525, 2014.

[32] M. Guidolin and A. Timmermann, "Economic implications of bull and bear regimes in uk stock and bond returns," *The Economic Journal*, vol. 115, no. 500, pp. 111–143, 2004.

[33] J. Fleming, C. Kirby, and B. Ostdiek, "Information and volatility linkages in the stock, bond, and money markets," *Journal of financial economics*, vol. 49, no. 1, pp. 111–137, 1998.

[34] S.-J. Kim, F. Moshirian, and E. Wu, "Evolution of international stock and bond market integration: Influence of the european monetary union," *Journal of Banking & Finance*, vol. 30, no. 5, pp. 1507–1534, 2006.

[35] M. Guidolin and A. Timmermann, "Term structure of risk under alternative econometric specifications," *Journal of Econometrics*, vol. 131, no. 1-2, pp. 285–308, 2006.

[36] M. L. Leibowitz, E. H. Sorensen, R. D. Arnott, and H. N. Hanson, "A total differential approach to equity duration," *Financial Analysts Journal*, vol. 45, no. 5, pp. 30–37, 1989.

[37] L. Li, "Macroeconomic factors and the correlation of stock and bond returns," 2002.

[38] S. H. Kwan, "Firm-specific information and the correlation between individual stocks and bonds," *Journal of financial economics*, vol. 40, no. 1, pp. 63–80, 1996.

[39] S. M. Schaefer and I. A. Strebulaev, "Structural models of credit risk are useful: Evidence from hedge ratios on corporate bonds," *Journal of Financial Economics*, vol. 90, no. 1, pp. 1–19, 2008.

[40] J. Bao and K. Hou, "Comovement of corporate bonds and equities," *Charles A. Dice Center Working Paper*, no. 2013-11, 2013.

[41] J. Dick-Nielsen, "How to clean enhanced trace data," *Available at SSRN 2337908*, 2014.

[42] A. Puckett and X. Yan, "The interim trading skills of institutional investors," *The Journal of Finance*, vol. 66, no. 2, pp. 601–633, 2011.

[43] J. E. Bethel, G. Hu, and Q. Wang, "The market for shareholder voting rights around mergers and acquisitions: Evidence from institutional daily trading and voting," *Journal of Corporate Finance*, vol. 15, no. 1, pp. 129–145, 2009.

[44] T. J. Chemmanur, S. He, and G. Hu, "The role of institutional investors in seasoned equity offerings," *Journal of Financial Economics*, vol. 94, no. 3, pp. 384–411, 2009.

[45] M. A. Goldstein, P. Irvine, E. Kandel, and Z. Wiener, "Brokerage commissions and institutional trading patterns," *The Review of Financial Studies*, vol. 22, no. 12, pp. 5175–5212, 2009.

[46] G. Hu, K. M. Jo, Y. A. Wang, and J. Xie, "Institutional trading and abel noser data," *Journal of Corporate Finance*, vol. 52, pp. 143–167, 2018.

[47] C. S. Eun and S. Sabherwal, "Cross-border listings and price discovery: Evidence from us-listed canadian stocks," *The Journal of Finance*, vol. 58, no. 2, pp. 549–575, 2003.

[48] L. Kryzanowski and S. Lazrak, "Informed traders of cross-listed shares trade more in the domestic market around earnings releases," *Review of Quantitative Finance and Accounting*, vol. 36, no. 1, pp. 1–31, 2011.

[49] H. Chen, P. M. S. Choi, and Y. Hong, "How smooth is price discovery? evidence from cross-listed stock trading," *Journal of International Money and Finance*, vol. 32, pp. 668–699, 2013.

[50] L. Kryzanowski, S. Perrakis, and R. Zhong, "Price discovery in equity and cds markets," *Journal of Financial Markets*, vol. 35, pp. 21–46, 2017.

[51] S. Mayhew, A. Sarin, and K. Shastri, "The allocation of informed trading across related markets: An analysis of the impact of changes in equity-option margin requirements," *The Journal of Finance*, vol. 50, no. 5, pp. 1635–1653, 1995.

[52] C. Chang and X. Yu, "Informational efficiency and liquidity premium as the determinants of capital structure," *Journal of Financial and Quantitative Analysis*, vol. 45, no. 2, pp. 401–440, 2010.

[53] N. Kapadia and X. Pu, "Limited arbitrage between equity and credit markets," *Journal of Financial Economics*, vol. 105, no. 3, pp. 542–564, 2012.

[54] A. W. Boot and A. V. Thakor, "Security design," *The Journal of Finance*, vol. 48, no. 4, pp. 1349–1378, 1993.

[55] A. R. Admati, "A noisy rational expectations equilibrium for multi-asset securities markets," *Econometrica: Journal of the Econometric Society*, pp. 629–657, 1985.

[56] I. Goldstein, Y. Li, and L. Yang, "Speculation and hedging in segmented markets," *The Review of Financial Studies*, vol. 27, no. 3, pp. 881–922, 2013.

[57] S. Manaster and R. J. Rendleman Jr, "Option prices as predictors of equilibrium stock prices," *The Journal of Finance*, vol. 37, no. 4, pp. 1043–1057, 1982.

[58] J. A. Stephan and R. E. Whaley, "Intraday price change and trading volume relations in the stock and stock option markets," *The Journal of Finance*, vol. 45, no. 1, pp. 191–220, 1990.

[59] K. Chan, Y. P. Chung, and H. Johnson, "The intraday behavior of bid-ask spreads for nyse stocks and cboe options," *Journal of Financial and Quantitative Analysis*, vol. 30, no. 3, pp. 329–346, 1995.

[60] A. S. Kyle, "Continuous auctions and insider trading," *Econometrica: Journal of the Econometric Society*, pp. 1315–1335, 1985.

[61] A. R. Admati and P. Pfleiderer, "Selling and trading on information in financial markets," *The American Economic Review*, vol. 78, no. 2, pp. 96–103, 1988.

[62] F. D. Foster and S. Viswanathan, "Variations in trading volume, return volatility, and trading costs: Evidence on recent price formation models," *The Journal of Finance*, vol. 48, no. 1, pp. 187–211, 1993.

[63] S. Chakravarty, "Stealth-trading: Which traders' trades move stock prices?" *Journal of Financial Economics*, vol. 61, no. 2, pp. 289–307, 2001.

[64] J. Caballe and M. Krishnan, "Imperfect competition in a multi-security market with risk neutrality," *Econometrica (1986-1998)*, vol. 62, no. 3, p. 695, 1994.

[65] R. C. Merton, "On the pricing of corporate debt: The risk structure of interest rates," *The Journal of finance*, vol. 29, no. 2, pp. 449–470, 1974.

[66] D. Easley and M. O'hara, "Price, trade size, and information in securities markets," *Journal of Financial economics*, vol. 19, no. 1, pp. 69–90, 1987.

[67] ——, "Time and the process of security price adjustment," *The Journal of finance*, vol. 47, no. 2, pp. 577–605, 1992.

[68] D. Easley and M. O'Hara, "Adverse selection and large trade volume: The implications for market efficiency," *Journal of Financial and Quantitative Analysis*, vol. 27, no. 2, pp. 185–208, 1992.

[69] R. W. Holthausen, R. W. Leftwich, and D. Mayers, "The effect of large block transactions on security prices: A cross-sectional analysis," *Journal of Financial Economics*, vol. 19, no. 2, pp. 237–267, 1987.

[70] R. Ball and F. J. Finn, "The effect of block transactions on share prices: Australian evidence," *Journal of Banking & Finance*, vol. 13, no. 3, pp. 397–419, 1989.

[71] D. J. Seppi, "Equilibrium block trading and asymmetric information," *the Journal of Finance*, vol. 45, no. 1, pp. 73–94, 1990.

[72] J. Hasbrouck, "The summary informativeness of stock trades: An econometric analysis," *The Review of Financial Studies*, vol. 4, no. 3, pp. 571–595, 1991.

[73] S. J. Grossman, "The informational role of upstairs and downstairs trading," *Journal of Business*, pp. 509–528, 1992.

[74] A. Madhavan and S. Smidt, "A bayesian model of intraday specialist pricing," *Journal of Financial Economics*, vol. 30, no. 1, pp. 99–134, 1991.

[75] ——, "An analysis of changes in specialist inventories and quotations," *The Journal of Finance*, vol. 48, no. 5, pp. 1595–1628, 1993.

[76] L. K. Chan and J. Lakonishok, "Institutional trades and intraday stock price behavior," *Journal of Financial Economics*, vol. 33, no. 2, pp. 173–199, 1993.

[77] ——, "The behavior of stock prices around institutional trades," *The Journal of Finance*, vol. 50, no. 4, pp. 1147–1174, 1995.

[78] R. D. Huang and H. R. Stoll, "Market microstructure and stock return predictions," *The Review of Financial Studies*, vol. 7, no. 1, pp. 179–213, 1994.

[79] D. B. Keim and A. Madhavan, "The upstairs market for large-block transactions: Analysis and measurement of price effects," *The Review of Financial Studies*, vol. 9, no. 1, pp. 1–36, 1996.

[80] H. Markowitz, "Portfolio selection," *The journal of finance*, vol. 7, no. 1, pp. 77–91, 1952.

[81] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *The journal of finance*, vol. 19, no. 3, pp. 425–442, 1964.

[82] J. Yang, Y. Zhou, and Z. Wang, "The stock–bond correlation and macroeconomic conditions: One and a half centuries of evidence," *Journal of Banking & Finance*, vol. 33, no. 4, pp. 670–680, 2009.

[83] J. Hasbrouck, "One security, many markets: Determining the contributions to price discovery," *The journal of Finance*, vol. 50, no. 4, pp. 1175–1199, 1995.

[84] ——, "Intraday price formation in us equity index markets," *The Journal of Finance*, vol. 58, no. 6, pp. 2375–2400, 2003.

[85] G. Bittlingmayer and S. M. Moser, "What does the corporate bond market know?" *Financial Review*, vol. 49, no. 1, pp. 1–19, 2014.

[86] K. Back and K. Crotty, "The informational role of stock and bond volume," *The Review of Financial Studies*, vol. 28, no. 5, pp. 1381–1427, 2014.

[87] X. Zhou, "Information, liquidity and corporate yield spreads," *Rutgers University*, 2005.

[88] A. J. Patton and A. Timmermann, "The resolution of macroeconomic uncertainty: Evidence from survey forecast," *CREATES Research Paper*, no. 2008-54, 2008.

[89] J. Boudoukh, R. Michaely, M. Richardson, and M. R. Roberts, "On the importance of measuring payout yield: Implications for empirical asset pricing," *The Journal of Finance*, vol. 62, no. 2, pp. 877–915, 2007.

[90] N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *The Journal of finance*, vol. 48, no. 1, pp. 65–91, 1993.

[91] E. Asparouhova, H. Bessembinder, and I. Kalcheva, "Liquidity biases in asset pricing tests," *Journal of Financial Economics*, vol. 96, no. 2, pp. 215–237, 2010.

[92] C. M. Lee and B. Swaminathan, "Price momentum and trading volume," *the Journal of Finance*, vol. 55, no. 5, pp. 2017–2069, 2000.

[93] J. L. Rogers, A. Van Buskirk, and S. L. Zechman, "Disclosure tone and shareholder litigation," *The Accounting Review*, vol. 86, no. 6, pp. 2155–2183, 2011.

[94] C. S. Asness, A. Frazzini, and L. H. Pedersen, "Low-risk investing without industry bets," *Financial Analysts Journal*, vol. 70, no. 4, pp. 24–41, 2014.

[95] W. Antweiler and M. Frank, "The market impact of corporate news stories," working paper, University of British Columbia, Tech. Rep., 2005.

[96] J. H. Cochrane, "Presidential address: Discount rates," *The Journal of finance*, vol. 66, no. 4, pp. 1047–1108, 2011.

[97] A. Ang, J. Liu, and K. Schwarz, "Using stocks or portfolios in tests of factor models," *Journal of Financial and Quantitative Analysis, Forthcoming*, 2019.

[98] W. Antweiler and M. Z. Frank, "The market impact of corporate news stories," 2004.

[99] ——, "Is all that talk just noise? the information content of internet stock message boards," *Journal of Finance*, vol. 59, no. 3, pp. 1259–1294, 2004.

[100] C. Asness, A. Frazzini, R. Israel, T. J. Moskowitz, and L. H. Pedersen, "Size matters, if you control your junk," *Journal of Financial Economics*, vol. 129, no. 3, pp. 479–509, 2018.

[101] C. Asness and A. Frazzini, "The devil in hml's details," *Journal of Portfolio Management*, vol. 39, no. 4, pp. 49–68, 2013.

[102] C. S. Asness, A. Frazzini, and L. H. Pedersen, "Quality minus junk," *Review of Accounting Studies*, vol. 24, no. 1, pp. 34–112, 2019.

[103] R. W. Banz, "The relationship between return and market value of common stocks," *Journal of Financial Economics*, vol. 9, no. 1, pp. 3–18, 1981.

[104] F. Barillas and J. Shanken, "Which alpha?" *Review of Financial Studies*, vol. 30, no. 4, pp. 1316–1338, 2017.

[105] L. Pástor and R. F. Stambaugh, "Comparing asset pricing models: an investment perspective," *Journal of Financial Economics*, vol. 56, no. 3, pp. 335–381, 2000.

[106] M. E. Blume, "Portfolio theory: a step toward its practical application," *Journal of Business*, vol. 43, no. 2, pp. 152–173, 1970.

[107] A. Bodnaruk, T. Loughran, and B. McDonald, "Using 10-k text to gauge financial constraints," *Journal of Financial and Quantitative Analysis*, vol. 50, no. 4, pp. 623–646, 2015.

[108] M. J. Brennan, T. Chordia, and A. Subrahmanyam, "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns," *Journal of Financial Economics*, vol. 49, no. 3, pp. 345–373, 1998.

[109] M. M. Carhart, "On persistence in mutual fund performance," *Journal of Finance*, vol. 52, no. 1, pp. 57–82, 1997.

[110] L. K. Chan, Y. Hamao, and J. Lakonishok, "Fundamentals and stock returns in japan," *Journal of Finance*, vol. 46, no. 5, pp. 1739–1764, 1991.

[111] J. Chen, H. Hong, and J. C. Stein, "Breadth of ownership and stock returns," *Journal of Financial Economics*, vol. 66, no. 2-3, pp. 171–205, 2002.

[112] J. L. Coles, N. D. Daniel, and L. Naveen, "Boards: Does one size fit all?" *Journal of Financial Economics*, vol. 87, no. 2, pp. 329–356, 2008.

[113] K. Daniel and S. Titman, "Evidence on the characteristics of cross sectional variation in stock returns," *Journal of Finance*, vol. 52, no. 1, pp. 1–33, 1997.

[114] K. Daniel, S. Titman *et al.*, "Testing factor-model explanations of market anomalies," *Critical Finance Review*, vol. 1, no. 1, pp. 103–139, 2012.

[115] K. Daniel, M. Grinblatt, S. Titman, and R. Wermers, "Measuring mutual fund performance with characteristic-based benchmarks," *Journal of Finance*, vol. 52, no. 3, pp. 1035–1058, 1997.

[116] S. R. Das *et al.*, "Text and context: Language analytics in finance," *Foundations and Trends® in Finance*, vol. 8, no. 3, pp. 145–261, 2014.

[117] S. R. Das and M. Y. Chen, "Yahoo! for amazon: Sentiment extraction from small talk on the web," *Management Science*, vol. 53, no. 9, pp. 1375–1388, 2007.

[118] A. Dey, "Corporate governance and agency conflicts," *Journal of Accounting Research*, vol. 46, no. 5, pp. 1143–1181, 2008.

[119] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, vol. 33, no. 1, pp. 3–56, 1993.

[120] ——, "Size and book-to-market factors in earnings and returns," *Journal of Finance*, vol. 50, no. 1, pp. 131–155, 1995.

[121] ——, "Size, value, and momentum in international stock returns," *Journal of financial economics*, vol. 105, no. 3, pp. 457–472, 2012.

[122] ——, "A five-factor asset pricing model," *Journal of Financial Economics*, vol. 116, no. 1, pp. 1–22, 2015.

[123] ——, "Dissecting anomalies with a five-factor model," *Review of Financial Studies*, vol. 29, no. 1, pp. 69–103, 2016.

[124] ——, "Choosing factors," *Journal of Financial Economics*, vol. 128, no. 2, pp. 234–252, 2018.

[125] E. F. Fama and J. D. MacBeth, "Risk, return, and equilibrium: Empirical tests," *Journal of Political Economy*, vol. 81, no. 3, pp. 607–636, 1973.

[126] G. Feng, S. Giglio, and D. Xiu, "Taming the factor zoo: A test of new factors," National Bureau of Economic Research, Tech. Rep., 2019.

[127] T. J. George, C.-Y. Hwang, and Y. Li, "The 52-week high, q-theory, and the cross section of stock returns," *Journal of Financial Economics*, vol. 128, no. 1, pp. 148–163, 2018.

[128] M. R. Gibbons, S. A. Ross, and J. Shanken, "A test of the efficiency of a given portfolio," *Econometrica: Journal of the Econometric Society*, pp. 1121–1152, 1989.

[129] C. R. Harvey, Y. Liu, and H. Zhu, "... and the cross-section of expected returns," *Review of Financial Studies*, vol. 29, no. 1, pp. 5–68, 2016.

[130] S. L. Heston and N. R. Sinha, "News vs. sentiment: Predicting stock returns from news stories," *Financial Analysts Journal*, vol. 73, no. 3, pp. 67–83, 2017.

[131] K. Hou, C. Xue, and L. Zhang, "Digesting anomalies: An investment approach," *Review of Financial Studies*, vol. 28, no. 3, pp. 650–705, 2015.

[132] J. Lakonishok, A. Shleifer, and R. W. Vishny, "Contrarian investment, extrapolation, and risk," *Journal of Finance*, vol. 49, no. 5, pp. 1541–1578, 1994.

[133] A. Klein, "Firm performance and board committee structure," *Journal of Law and Economics*, vol. 41, no. 1, pp. 275–304, 1998.

[134] N. Jegadeesh and D. Wu, "Word power: A new approach for content analysis," *Journal of Financial Economics*, vol. 110, no. 3, pp. 712–729, 2013.

[135] X. Huang, S. H. Teoh, and Y. Zhang, "Tone management," *The Accounting Review*, vol. 89, no. 3, pp. 1083–1113, 2013.

[136] J. Lewellen, "The time-series relations among expected return, risk, and book-to-market," *Journal of Financial Economics*, vol. 54, no. 1, pp. 5–43, 1999.

[137] ——, "Institutional investors and the limits of arbitrage," *Journal of Financial Economics*, vol. 102, no. 1, pp. 62–80, 2011.

[138] F. Li, "Annual report readability, current earnings, and earnings persistence," *Journal of Accounting and Economics*, vol. 45, no. 2-3, pp. 221–247, 2008.

[139] ——, "The information content of forward-looking statements in corporate filings—a naïve bayesian machine learning approach," *Journal of Accounting Research*, vol. 48, no. 5, pp. 1049–1102, 2010.

[140] A. W. Lo and A. C. MacKinlay, "When are contrarian profits due to stock market overreaction?" *Review of Financial Studies*, vol. 3, no. 2, pp. 175–205, 1990.

[141] T. Loughran and B. McDonald, "When is a liability not a liability? textual analysis, dictionaries, and 10-ks," *Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.

[142] ——, "The use of word lists in textual analysis," *Journal of Behavioral Finance*, vol. 16, no. 1, pp. 1–11, 2015.

[143] ——, "Textual analysis in accounting and finance: A survey," *Journal of Accounting Research*, vol. 54, no. 4, pp. 1187–1230, 2016.

[144] R. Novy-Marx, "The other side of value: The gross profitability premium," *Journal of Financial Economics*, vol. 108, no. 1, pp. 1–28, 2013.

[145] K. Ranganathan and P. Singh, "Anchoring in mergers and acquisitions: Does the regulatory environment matter?" *Journal of Accounting, Auditing & Finance*, p. 0148558X18798998, 2018.

[146] B. Rosenberg, K. Reid, and R. Lanstein, "Persuasive evidence of market inefficiency," *Journal of Portfolio Management*, vol. 11, no. 3, pp. 9–16, 1985.

[147] R. F. Stambaugh and Y. Yuan, "Mispricing factors," *Review of Financial Studies*, vol. 30, no. 4, pp. 1270–1315, 2016.

[148] P. C. Tetlock, "Giving content to investor sentiment: The role of media in the stock market," *Journal of Finance*, vol. 62, no. 3, pp. 1139–1168, 2007.

[149] P. C. Tetlock, M. Saar-Tsechansky, and S. Macskassy, "More than words: Quantifying language to measure firms' fundamentals," *Journal of Finance*, vol. 63, no. 3, pp. 1437–1467, 2008.

[150] S. Titman, K. J. Wei, and F. Xie, "Capital investments and stock returns," *Journal of Financial and Quantitative Analysis*, vol. 39, no. 4, pp. 677–700, 2004.

[151] B. Twedt and L. Rees, "Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports," *Journal of Accounting and Public Policy*, vol. 31, no. 1, pp. 1–21, 2012.

[152] M. Vassalou and Y. Xing, "Default risk in equity returns," *Journal of Finance*, vol. 59, no. 2, pp. 831–868, 2004.

[153] R. L. Watts and J. L. Zimmerman, "Positive accounting theory: a ten year perspective," *The Accounting Review*, pp. 131–156, 1990.

[154] Q. Ma, H. Wang, and W. Zhang, "Trading against anchoring," *Review of Behavioral Finance*, vol. 9, no. 3, pp. 242–261, 2017.

[155] F. Barillas and J. Shanken, "Comparing asset pricing models," *Journal of Finance*, vol. 73, no. 2, pp. 715–754, 2018.

[156] N. Jegadeesh and S. Titman, "Momentum," *Annual Review of Financial Economics*, vol. 3, no. 1, pp. 493–509, 2011.