THREE ESSAYS ON INFORMATION DISSEMINATION: EVIDENCE USING TEXTUAL ANALYSIS

Daqi Xin

Submitted in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Approved by:
Dr. Bill B. Francis, Co-Chair
Dr. Thomas D. Shohfi, Co-Chair
Dr. Qiang Wu
Dr. Thomas J. Boulton

Lally School of Management
Rensselaer Polytechnic Institute
Troy, New York

[August 2020]
Submitted June 2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

It has been a long journey and finally, I came to this point. This accomplishment would not be possible without the support of many people during my Ph.D. studies at Rensselaer.

First and foremost, I am deeply indebted to my co-chairs, Dr. Bill Francis and Dr. Tom Shohfi. Dr. Francis has been astonishing me with his encyclopedic knowledge of literature. His acute intuition for interesting ideas laid the foundation of this dissertation. The passion and enthusiasm for research he passed on will be my lifelong assets. Dr. Shohfi is not only an advisor but also a life mentor. He taught me how to develop an idea, how to do academic writing, and how to be a professional researcher. He pushed me and supported me with incredible patience and deep caring. With a dedication to top-end research and superior work-life management, Dr. Shohfi sets an admirable example I will strive to follow in my future career. Furthermore, I want to express my gratitude to Dr. Shohfi's family for establishing the Don Shohfi Financial Research Fund which greatly facilitated my data processing by setting up the server. I must also thank the other two committee members, Dr. Qiang Wu and Dr. Tom Boulton, whose insights contributed to this dissertation and my understanding of finance and accounting.

I am fortunate to have the opportunities to learn from and work with so many Rensselaer faculty members—Chanaka Edirisinghe, Shyam Kumar, Pengfei Ye, Yingda Lu, Dorit Nevo, Nishtha Langer, Johan Maharjan, Brian Clark, Chad Stecher, Tim Golden, and Murali Chari. Thank you for sharing your wisdom and I will benefit from what I learned from each of you. Thank the staff at Rensselaer for coordinating everything for me throughout the program.

I would like to thank my fellow students—Wenyao, Haimeng, Majeed, Liang, Amy, Chaoqun, Joe, Abena, Gilna, Josh, Juhi, Lianlian, Qinglin, Billy, Garo, Bok, and Tahseen. I will not forget the time we spent together and the generous help you provided.

I would like to extend my deepest gratitude to my family. I am lucky to be raised in a family that values knowledge and education. To my parents, Rishang Xin and Lichuan Wang: as the most resilient and courageous parents, you are always there when I need you. To my grandpa, Yuqin Wang: you taught me to use a dictionary when I was a kid and I believe you will be proud of Dr. Xin now. To my grandma, Aiyu Wu: you have been giving me unconditional love for thirty years and I will do all I can to repay you.

Finally, my deepest appreciation goes to my wife, Xi Yang, who is a gorgeous woman with a PhD in Management. We knew each other when we were in high school, made the lifetime commitment after we graduated from college, and chose to challenge the most dangerous thing in the world—pursuing PhD at the same time 2000 miles apart. We loved and supported each other through the entire process and I believe nothing will change in the rest of our lives.

I dedicate this dissertation to my two sons, Daniel and Derek, and their sister, 洛岚, who might be born in 2022.

ABSTRACT

This dissertation consists of three essays on information dissemination in financial markets. Specifically, I employ textual analysis to tease out various subtle information that is difficult to gather with traditional approaches and then examine how this information helps us answer important questions in the disciplines in finance and accounting.

In the first chapter, I use the presence of a Wikipedia article for initial public offering (IPO) firms to test theories of information asymmetry and investor awareness. While I find limited support for the former, my results provide strong support for theories of investor awareness. Specifically, IPO firms with a Wikipedia article exhibit significantly higher underpricing than do IPO firms without a Wikipedia article. Investor awareness has positive long-term effects, including greater analyst following and institutional ownership for up to three years after the offering. The effect is robust to firm-specific Google search volume, news coverage, propensity score matching, and an instrumental variable approach.

In Chapter Two, I investigate gender issues in interactions between two high-profile professions—sell-side analysts and public firm executives using a large sample of quarterly earnings conference call transcripts. I find that women are generally less “visible” on conference calls. Specifically, female analysts have fewer conference call participation opportunities. Conditional on participation, female analysts are allowed fewer opportunities to ask follow-up questions and speak less compared with male counterparts. Female analysts speak with more positive tone, less uncertainty, less numerical content, fewer speech hesitations, and fewer back and forth conversations with firm management. Female executives have shorter discourses and receive more rounds of questions from analysts. However, female executives exhibit more certainty and hesitate less, indicating superior abilities in answering analysts’ questions. My

analysis of speech interruptions finds that female analysts are interrupted less by female, but not male, executives. Moreover, female executives receive more interruptions from both male analysts and executives and are more likely to be challenged by male subordinates. The equity market also discounts female analysts' participation. Overall, my results are consistent with gender-based discrimination.

Chapter Three examines gender differences in textual characteristics of analyst reports. I find female analyst reports are more readable but shorter. Consistent with an "ethical standard" explanation, the textual sentiment of female analyst reports is less optimistic. Moreover, female analyst reports contain less financial content and are more long-term oriented. Male analysts' reports induce stronger market reactions when readability is higher, but the opposite is true for female analysts. Female analysts improve report readability more over their career than male analysts do. Out results provide evidence of gender stereotyping in the analyst profession.

CHAPTER ONE

INVESTOR AWARENESS OR INFORMATION ASYMMETRY? WIKIPEDIA AND IPO UNDERPRICING

1.1 Introduction

The way the world consumes information has changed dramatically since researchers began studying initial public offering (IPO) underpricing (Logue, 1973; Ibbotson, 1975). Possibly, no invention since the television in the 1920s has done more to democratize the availability of information than the Internet. In recent years, the Internet has evolved from a medium to consume information passively to a place where users collaborate to create content. There is perhaps no better example of this collaborative effort than Wikipedia, the leading free online encyclopedia where anyone can create and edit content.

As private companies or subsidiaries of public companies, information about IPO firms is often limited. For many potential investors, the issuer's carefully crafted registration statement is the primary source of information used to evaluate the IPO as an investment opportunity. Federal securities laws limit the information that issuers and their representatives can share with the public between the time the registration statement is filed and declared effective by the U.S. Securities and Exchange Commission (SEC, 2017). However, these limitations do not apply to the collaborative efforts of the Wikipedia community, which makes Wikipedia a potentially valuable source of information for IPO investors.

Portions of this chapter previously appeared as: Boulton, T. J., Francis, B. B., Shohfi, T., & Xin, D. (2019). *Investor awareness or information asymmetry? Wikipedia and IPO underpricing.* SSRN Electronic Journal. Retrieved June 25, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2908624

Launched on January 15, 2001, Wikipedia currently ranks among the ten most popular websites worldwide and is the world's leading reference source (Alexa, n.d.). Despite Wikipedia's potential importance as a source of information for investors, its impact on the information environment of IPO firms is an unexplored issue. However, evidence suggests that potential investors reference a company's Wikipedia page around their IPO. To illustrate, I report traffic to LinkedIn's Wikipedia page from January 1 through June 30, 2011 in Figure 1.1. This period includes LinkedIn's S-1 filing date (January 27, 2011) and IPO issue date (May 19, 2011). For most of this period, LinkedIn's Wikipedia page attracts approximately 3,000 page views per day. However, on LinkedIn's IPO issue date, the company's Wikipedia page attracts over 11,000 page views. I observe similar spikes in Wikipedia traffic for other IPO firms during my sample, which motivates us to examine the impact of Wikipedia on the pricing and long-run performance of IPOs.

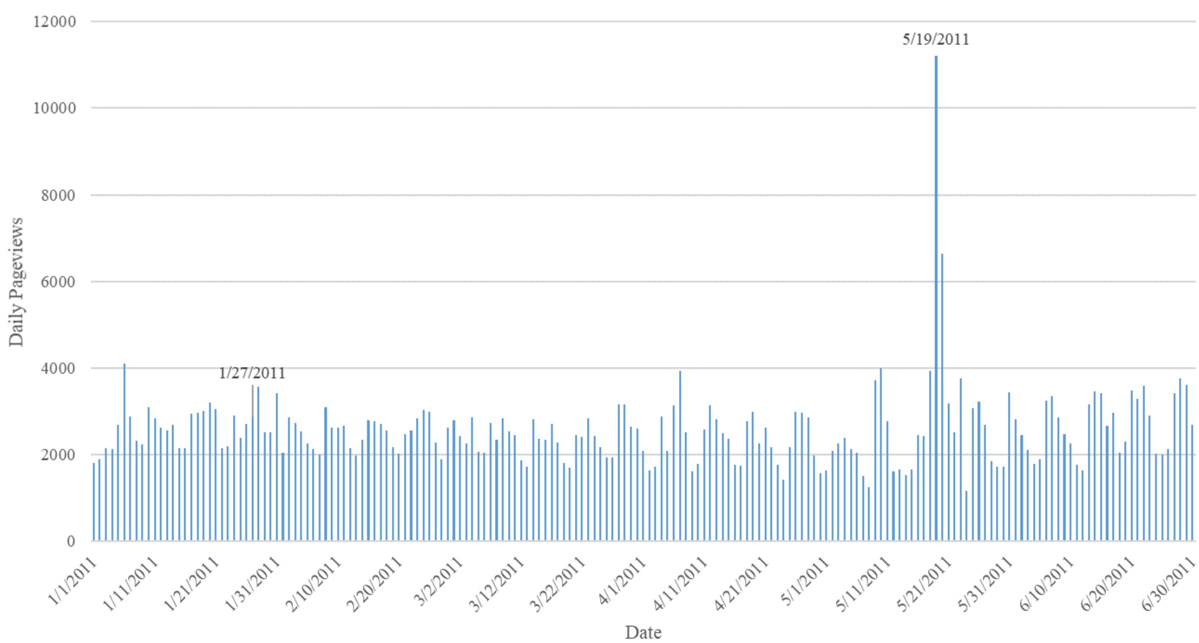


Figure 1.1: Wikipedia article traffic example

I contend that Wikipedia could level the information playing-field between IPO issuers and the investment banks they employ (Baron, 1982), between IPO issuers and potential investors

(Welch, 1989), and between different investor groups (Rock, 1986). Information disparities make it difficult to precisely price a firm's IPO (Bradley, Cooney, Jordan, and Singh, 2004) and are believed to contribute to underpricing that results in the large first-day gains exhibited by many IPOs (Ljungqvist, 2007). The economic consequences of these information effects are significant. For instance, Ritter (1987) and others find that underpricing is the largest single cost of going public for the majority of IPO issuers.

Wikipedia also has the potential to increase investor attention to IPO firms. Prior research finds that stock prices do not fully reflect value-related information until the information grabs investor attention (Hirshleifer, Hou, Teoh, and Zhang, 2004; Hong, Torous, and Valkanov, 2007; Frederickson and Zolotoy, 2016). Recent studies show that the media plays an important role in the information environment of capital markets (Tetlock, 2007; Bhattacharya, Galpin, Ray, and Yu, 2009; Fang and Peress, 2009; Engelberg and Parsons, 2011). Thus, the presence and content of a Wikipedia article could have a significant effect on IPO offer prices and secondary market prices due to Wikipedia's impact on the information environment and investor attention surrounding IPO firms. Note that the investor attention induced by Wikipedia article is different from the attention induced by other media platforms. One concern is that Wikipedia articles do not induce investor attention like newspaper reporting which is "pushed" to potential investors because all Wikipedia traffic is a result of active searching, suggesting that Wikipedia visitors already have attention to IPO firms *before* they visit Wikipedia and Wikipedia traffic is simply a reflection of existing investor attention. I argue that Wikipedia can induce investor like newspapers because of its high visibility. Besides, Wikipedia accredits IPO firms for potential investor. In other words, investors who have "general attention" to IPO firms before they read Wikipedia articles may develop "investing attention" which leads to secondary market investment. This investing attention cannot be fully captured by either news or Google search volume (Da, Engelberg, and Gao, 2011; Liu, Sherman, and Zhang, 2014)

I find that firms that have a Wikipedia article when they go public experience significantly higher underpricing than firms without a Wikipedia article (21.0% vs 12.7%). The association between the existence of a Wikipedia article and IPO underpricing is also present in multivariate regressions that control for firm- and issue-related factors that have been shown to affect underpricing, and to a variety of robustness checks including controlling for abnormal firm-specific Google search volume (Da et al., 2011) and news outlet activity around the IPO. Additionally, I employ instrumental variable regression and propensity score matching methods. My results consistently point to a statistically significant and economically large positive relation between the presence of a pre-IPO Wikipedia article and IPO underpricing. Because this is inconsistent with the notion that a Wikipedia article reduces information asymmetry, I draw on prior research on investor attention to explain the positive link between a Wikipedia article and IPO underpricing.

Da et al. (2011) and Liu et al. (2014) provide evidence that investor attention is positively correlated with IPO underpricing. An important difference between these studies is that they differ on whether transient retail investors or longer-horizon institutional investors drive the impact of investor attention on IPO outcomes. Da et al. (2011) find that high initial returns are followed by long-run underperformance for IPOs that receive high investor attention. This is consistent with Barber and Odean (2008) who find that individual investors tend to buy “attention-grabbing” stocks, which generates temporary price pressure that leads to higher stock prices and lower future returns. However, Liu et al. (2014) provide evidence consistent with Merton’s (1987) investor recognition model that predicts that increased investor attention has positive long-term effects for firms. Consistent with this explanation, I find that the presence of a pre-IPO Wikipedia article is associated with long-term benefits, including greater analyst following and more institutional investors compared to IPO firms without a Wikipedia article.

To understand the nature of the content reported in IPO firms' Wikipedia articles, I also perform textual analysis that compares firms' Wikipedia articles to their registration statements. I find that the tone of Wikipedia articles differs substantially from that of issuers' S-1 registration statements. Specifically, compared to Wikipedia articles I find that registration statements are proportionally more negative than positive and tend to use more words that are uncertain and litigious. It should also be noted that I find that it is general (Harvard GI) rather than financial context specific (Loughran and McDonald, 2011) sentiment that helps to explain the positive relation between Wikipedia articles and IPO underpricing.

To summarize, I provide evidence on the importance of Wikipedia's user-generated content to the information environment for primary capital market participants. The fact that the presence of a Wikipedia article is associated with significantly higher IPO underpricing demonstrates the importance of the collaborative efforts of the Wikipedia community. My results complement recent studies of internet stock message boards (Antweiler and Frank, 2004), financial blogs (Saxton and Anker, 2013), crowd-sourced financial research websites (Chen, De, Hu, and Hwang, 2014), Facebook (Zhou, Lei, Wang, Fan, and Wang, 2015), and Twitter (Blankespoor, Miller, and White, 2014). Similar to Wikipedia, these platforms use decentralized channels to aggregate information. This differs from corporate disclosure and media coverage where information is diffused unidirectionally, which raises concerns about objectivity and validity (Verrecchia, 1983; Gao and Ritter, 2010; Brown and Hillegeist, 2007).

To the best of my knowledge, I provide the first in-depth analysis of the relation between the most prominent source of user-generated content on the World Wide Web and IPO underpricing. My finding that IPO firms with Wikipedia articles have significantly higher underpricing, contrasts with the notion that underpricing is compensation to investors for limited information (Rock, 1986) and is consistent with the notion that Wikipedia articles increase investor attention (Liu et al., 2014). One possible concern is the reliability of information reported on Wikipedia. Greenstein and Zhu (2012) identify three tenets that Wikipedia articles seek to attain: a neutral point of view, verifiability, and the absence of original research. Although there is some evidence of bias and slant in Wikipedia articles on controversial topics involving subjective information (Greenstein and Zhu, 2012, 2018), information provided on Wikipedia tends to be accurate in the areas of history (Holman Rector, 2008), medicine (Devgan, Powe, Blakely, and Makary, 2007), pharmacology (Clauson, Polen, Boulos, and Dzenowagis, 2008), politics (Brown, 2011), and science (Giles, 2005). Thus, given that information contained in articles about companies with a forthcoming IPO is unlikely to be controversial or subjective I contend that my findings are not driven by bias or slant in the Wikipedia articles.

My study contributes to the burgeoning literature on the role of media in financial markets. For example, studies find that local media coverage is associated with local trading activities (Engelberg and Parsons, 2011), that media sentiment predicts stock returns and trading behavior (Tetlock, 2007), and that traditional media coverage predicts lower subsequent stock volatility and turnover (Jiao, Veiga, and Walther, 2018). In the case of IPOs, studies show that more media coverage during the quiet period is associated with more attention-driven retail purchases (Bushee, Cedergren, Michels, 2019), that media sentiment and first-day returns are positively correlated (Bajo and Raimondo, 2017), and that long-run returns are lower for IPOs with more pre-IPO newspaper articles (You, Coakley, Firth, Fuertes, and Shen, 2018). However, the unique information structure of Wikipedia distinguishes it from traditional media and most social media platforms. Specifically, Wikipedia is organized by topic and information accumulates over time due to the contributions of the Wikipedia user community. Comparatively, information on traditional media and social media is more dispersed. Given that individuals have limited information processing ability (Hirshleifer and Teoh, 2003; Hong and Stein, 1999), Wikipedia is likely to significantly reduce information acquisition and processing costs for investors (Gu, Konana, Rajagopalan, and Chen, 2007).

The rest of the paper is organized as follows: Section 2 discusses Wikipedia, IPOs, and develops my hypotheses. Section 3 describes the data and methodology. Section 4 presents my main results. Section 5 addresses endogeneity concerns. In Section 6 I present robustness checks. Section 7 concludes.

1.2 Wikipedia and IPO underpricing

1.2.1 Wikipedia

Launched in 2001, Wikipedia is one of the most popular websites in the world with nearly 750 million unique users each month (Cohen, 2014). As of July 2019, Wikipedia has 36.6 million registered users and 123 thousand active contributors (“Special:Statistics”, n.d.). The English Wikipedia, which is one of 288 international editions, includes more than 5 million content pages and typically receives 3-4 billion page views per month (“Page Views for Wikipedia”, n.d.). Wikipedia is so ubiquitous that Time Magazine recently named it the third most influential website of all time (Fitzpatrick, Eadicicco, and Peckham, 2017).

The basic unit of information on Wikipedia is an article, which distinguishes it from social media platforms such as Facebook and Twitter. A link to a Wikipedia article often appears in an info box following a Google search, thus making Wikipedia a primary source of information on a broad range of topics for many Internet users. An important characteristic of Wikipedia is that articles evolve over time from the collaborative effort of the Wikipedia community. Wikipedia applies several mechanisms to improve the authenticity of content. For example, the “pending changes” system requires an established Wikipedia editor to review edits made by new users (Frewin, 2010). Kumar, West, and Leskovec (2016) report that 90% of hoaxes submitted to Wikipedia are caught in under an hour, suggesting that the editorial process is effective in policing user contributions.

Articles describing private and public firms are a major component of Wikipedia. Such articles typically start with a general description of the company, followed by sections that detail company history, events, products, organizations, strategies, and competitors. Important events are usually reported in a standardized format such as “On [Day Month, Year], [The company] [did something].” According to SimilarWeb, 85.89% of Wikipedia’s traffic is from active searching (SimilarWeb, 2019).[1] Wikipedia articles often take a prominent position in search engine results, with the Wikipedia article of a company generally appearing among the first several results when a company name is used in a search query. A study conducted in 2012 finds that Wikipedia pages are present on the first page of results for 99% of Google searches and as the very first result for 56% of searches.

There is reason to believe that Wikipedia is a primary source of company-related information for many in the investment community. For instance, Xu and Zhang (2013) find that management disclosure and investor reaction are influenced by the presence of a Wikipedia page

[1] Although conducting active search of firms indicates people already pay attention to the company, it does not necessarily mean that they will become investors. My argument is that Wikipedia increases potential investors’ familiarity of IPO firms (Huberman, 2001).

about the firm. A survey of business journalists, analysts, and investors that were asked about preferred sources of information other than firms' corporate sites concluded, "Wikipedia is the most popular social media site for individuals looking for such information, used by more than three quarters of respondents." (Bradshaw, 2008)

1.2.2 Initial public offerings

1.2.2.1 Information asymmetry

IPO firms are private companies or subsidiaries of public companies prior to the offering, which means that information about them is limited. Prior research suggests that limited information about IPO firms contributes to information disparities between issuers and underwriters (Baron, 1982), issuers and IPO investors (Welch, 1989), and different investor groups (Rock, 1986). These information disparities are thought to drive the substantial first-day returns observed for many IPO firms (e.g., Ljungqvist, 2007). Because underpricing reduces the IPO proceeds that an IPO firm raises, it represents a substantial portion of the cost of going public for many firms (Ritter, 1987).

Evidence indicates that IPO firms and their representatives take actions to reduce information asymmetry and improve IPO outcomes. Some firms attempt to improve the information environment by providing more timely and informative disclosures to investors (Jog and McConomy, 2003; Leone, Rock, and Willenborg, 2007). The creation and cultivation of a Wikipedia article could have a similar effect on the information environment of IPO firms. If Wikipedia is associated with better information dissemination and, therefore, less information asymmetry, then I should observe a negative relation between a pre-IPO Wikipedia article and IPO underpricing. Thus, my first hypothesis is as follows:

H1: The presence of a pre-IPO Wikipedia article is negatively correlated with initial returns.

1.2.2.2 Investor attention

Investors have limited information processing ability, which makes attention a valuable resource. Merton (1987) points out that investors who are unfamiliar with a firm are unlikely to include it in their portfolio and, due to "set-up" costs, are less likely to respond to firm-specific announcements. In his model, an increase in investor awareness could have positive *long-run* effects for a firm. For example, a larger investor base is associated with a lower cost of capital and higher market valuation. This suggests that firms have incentives to promote investor awareness.

Barber and Odean (2008) find evidence that individual investors tend to buy "attention-grabbing" stocks, which results in an increase in stock prices. Because investors have many choices of stocks to buy, attention helps to narrow investors' choice set (Odean, 1999). A similar search problem does not exist for stock sales because investors can only sell stocks that are already part of their portfolio. Attention-induced price increases should be short-lived if they result from *temporary* price pressure and not information about firm fundamentals.

IPO firms also need to attract attention to sell their shares to investors and create a liquid secondary market. Da et al. (2011) and Liu et al. (2014) study the effect of investor attention on IPO outcomes. Both studies report a positive correlation between measures of investor attention and IPO underpricing.[2] However, their findings differ when it comes to the long-term impact of investor attention on IPO firms. Da et al. (2011) find that short-term investor attention measured by Google search volume predicts long-run underperformance of IPO stocks, while Liu et al. (2014) find that investor attention, measured by pre-IPO media coverage, has a positive effect on IPO firms' long-term value, liquidity, analyst coverage, and institutional ownership. Liu et al.

[2] Chemmanur, Krishnan, and Yu (2018) report that VC-backing leads to higher first-day returns to IPOs due to an increase in investor attention.

(2014) suggest that the difference may result from measuring a different type of attention from different investors.

The presence of a Wikipedia article could be a proxy for investor attention. A Wikipedia article indicates that a collaborative effort is underway to gather and report information on the company. Wikipedia provides a general picture of the firm when available information is limited. Brunner and Ungeheuer (2020) uses hourly Wikipedia page views as a measure of retail investor attention and shows that stocks ranked as daily winners and losers in the Wall Street Journal and New York Times exhibit attention spikes. Moreover, given the high reliability of Wikipedia, Wikipedia could also act as an accreditor for the legitimacy of IPO firms. For instance, based on a survey conducted in 2014, 64% of the British public trust Wikipedia entries more than they trust news media including the BBC (Ali, 2014). In the context of the IPO market, the sparsity of firm information could lead to an increase in the perceived accuracy of Wikipedia content by potential investors. As such, firms with a Wikipedia article are more likely to grab investor attention than firms without a Wikipedia article. Given that Da et al. (2011) and Liu et al. (2014) find that investor attention is positively correlated with first-day returns, I predict the following:

H2: The presence of a pre-IPO Wikipedia article is positively correlated with initial returns.

1.3 Data and methodology

1.3.1 Sample selection

I begin by collecting completed U.S. IPOs with an offer price of at least $5 per share issued between 2006 and 2016 from Thomson Reuters Securities Data Company (SDC) New Issues database. Although Wikipedia was launched in 2001, I begin my IPO sample in 2006 for several reasons. First, there are no IPO firms with a Wikipedia article at the time of their offering before 2004. Second, I find that the informativeness of Wikipedia articles before 2006 is limited. For example, the average number of words in an IPO firm's Wikipedia article increases from 100 in 2005 to 400 in 2006. Third, during the early years, Wikipedia had low awareness among Internet users. According to alexa.com, Wikipedia traffic ranked in the top 500 websites in October 2004, top 100 in April 2005, and top 30 in January 2006 ("Wikipedia.org Is More Popular Than...", n.d.).

Following prior IPO literature, I exclude foreign issuers, American Depository Receipts, closed-end funds, natural resource limited partnerships, real estate investment trusts, unit offers, small best efforts offerings, financial firms, and stocks not covered by The Center for Research in Security Prices (CRSP) database. After imposing these filters, I am left with a sample of 974 IPOs. I retrieve stock price and return data from CRSP and accounting data from Compustat.

1.3.2 Variables

I use a web crawler to search for an IPO firm's Wikipedia article and manually check its accuracy. I assign a Wikipedia article to an IPO if the article is titled with the name of: (1) the IPO firm; (2) the IPO firm's parent company if it is the IPO firm's parent before the first trade date; (3) the IPO firm's major subsidiary;[3] (4) a company from which the IPO firm separates;[4] (5) the firm's predecessor;[5] or (6) the core product or service and primarily contains information about the firm.[6] Some firms do not have an exclusive article but have brief information on a page with other items that are classified as "ambiguous words" by Wikipedia.[7] For these instances, I do not consider the

3 For example, for Hertz Global Holdings Inc. I use the Wikipedia article of The Hertz Corporation. For NYMEX Holdings Inc., I use the Wikipedia article of New York Mercantile Exchange
4 For example, Reliant Energy separated into CenterPoint Energy, Inc. and Reliant Resources, Inc., which is an IPO firm in my sample.
5 For example, for Bakers Footwear Group Inc. I use the Wikipedia article of Edison Brothers Stores, Inc.
6 For Intersections Inc., I use the Wikipedia article of its service "Identity guard." Neurometrix Inc. uses the page "Quell". Lincoln Educational Services Corp has a page titled "Lincoln Group of Schools"
7 For example, Veridian Corp is on a page titled "Veridian." On this page, Veridian Corp. is the first item and the contents include "an American aerospace and defense company, acquired by General Dynamics in 2003. Veridian Engineering, Inc., a subsidiary of American aerospace and defense company Veridian Corporation which was acquired by General Dynamics in 2003."

firm to have a Wikipedia article because much of my focus is on the information provided by Wikipedia instead of simply the existence of an article.

Because my goal is to examine the effect of Wikipedia on the information environment of IPO firms, I need to identify the presence and content of a firm's Wikipedia article at the time of the public offering. To assist with this identification, I use the "date of page creation" provided for each Wikipedia article. Appendix A contains LinkedIn's "page information" which reports basic information including page length, page ID, number of page watchers, page creator, date of page creation, and total number of edits. At the bottom of the information page, there are "external tools" links to revision history, page view statistics, and other information. For IPO firms with a Wikipedia article prior to the first trading day, I set the indicator variable *Wikipedia* equal to 1, and zero otherwise.[8] I identify 330 firms that have a Wikipedia article at the time of their IPO.

The relevant Wikipedia article for my analysis is the last historical version prior to the first-trading day. For each Wikipedia article, I access historical versions by clicking the "view history" tag. Information on the "revision history" page includes time of modification, editor, IP address, flag of minor edit,[9] and net change page size. In Appendix B, for the purpose of illustration, I provide LinkedIn's Wikipedia article as of its IPO date, May 19, 2011. Wikipedia responds promptly to IPO information. In the calendar week prior to the IPO date, the page has 0.9 revisions per day, on average. During the 3-day window centered on LinkedIn's IPO date, there are 19 revisions. Appendix B shows LinkedIn's Wikipedia article with basic information in the box to

[8] For firms with a Wikipedia article prior to the first-trading day but no content or with less than 30 words in the main body, I set the Wikipedia indicator equal to zero. For example, Allot Communications Ltd went public on Nov 15, 2006 and its page was created on Dec 9, 2005. However, this page was a redirect page of "Allotment". The first page revision after its creation was on Sept 17, 2007 after its IPO. Only one IPO firm has its Wikipedia page created on its IPO date (Zoetis Inc.).

[9] An editor can mark a page modification as "minor edit" if she believes differences between the new version and previous version do not require the review of other editors. Typographical corrections and reformatting are common examples of minor edits.

the right of the article and a brief introduction, including key events, in the first several paragraphs. Consistent with Wikipedia's verifiability tenant (Greenstein and Zhu, 2012), the references used to compose the article are listed at the end of the article. IPO-related information was added to the end of the introduction and more details were added as a separate paragraph in the section on company history.

Typically, information from a company's S-1 filing is rapidly integrated into its Wikipedia page. For example, LinkedIn filed its S-1 with the U.S. Securities and Exchange Commission on January 27, 2011. On the next day, the following was added to the "Company background" section of LinkedIn's Wikipedia page: *"LinkedIn filed for IPO on 27 January 2011. The listing could raise $175 million. According to the prospectus, the company's revenue doubled for the first nine months of 2010.*" On May 19, information about the initial pricing and trading of LinkedIn's IPO was added. Given Wikipedia's detailed editing history, I can check any historical version of a Wikipedia page and compare any two different versions. Appendix C demonstrates LinkedIn's Wikipedia page revision history. Specifically, Appendix C.1 shows a list of historical versions of LinkedIn page and Appendix C.2 shows the comparison of two historical pages. To capture the information aggregation and quality of a Wikipedia article, I construct three variables based on the latest historical version of an IPO firm's Wikipedia page. First, *wiki_revisions* is equal to the total number of Wikipedia article revisions during the book-building period (e.g., from S-1 filing date to the day before the IPO date). Second, *wiki_references* is the number of references in a Wikipedia article. Third, *wiki_words* is the number of words in the main text of a Wikipedia article.

I also conduct textual analysis to investigate the content of a Wikipedia article. I use dictionaries constructed by Loughran and McDonald (2011) to capture the tone and sentiment of each Wikipedia article: positive, negative, uncertainty, and litigious.[10] Because Wikipedia articles include non-financial information, I also calculate positive and negative sentiment proportions using the Harvard General Inquirer (GI) dictionaries used in the psychology and sociology literatures.

Following prior literature, I construct a number of measures related to the IPO event. These variables include offer price revision, venture capital backing, top tier underwriter, share overhang, IPO proceeds, and other IPO firm and event characteristics. Detailed definitions of all variables are provided in the Appendix.

1.4 Main results

1.4.1 Descriptive statistics

In Figure 1.2, I report the total number of IPOs with and without a Wikipedia article for each year during my sample period. The number of IPOs without a Wikipedia article drops from 105 in 2006 to 11 in 2008 during the financial crisis. As IPO activity resumes following the crisis, the number of IPOs without a Wikipedia article reaches 105 in 2014 before falling to 61 in 2015 and 51 in 2016. Comparatively, the number of IPOs with a Wikipedia article exhibits a similar pattern but with lower volatility. After reaching a trough in 2008 with 6 IPOs, the number of IPOs with a Wikipedia article increases gradually to 62 in 2014 before falling to 19 in 2016.

[10] The Loughran and McDonald (2011) dictionaries can be found at Prof. McDonald's website.

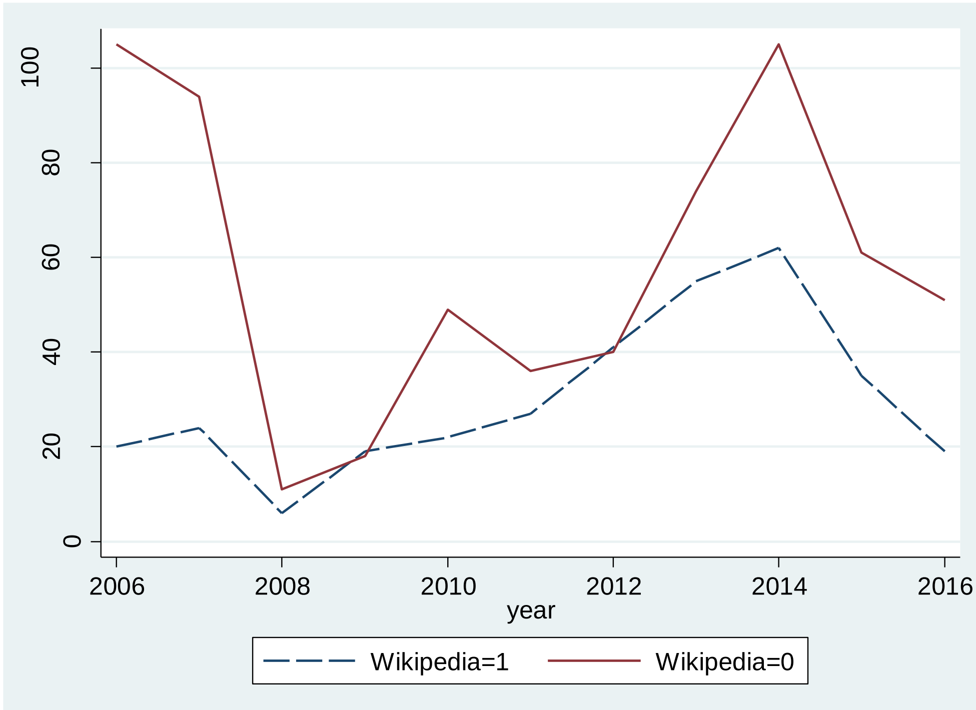


Figure 1.2: Number of IPOs by year with and without a Wikipedia page

I report summary statistics for IPOs with and without a pre-IPO Wikipedia article in Table 1.1 Panel A. I winsorize all continuous variables at the top and bottom 1% to limit the influence of extreme values. The difference in average underpricing for IPOs with and without a Wikipedia article is striking (21.0% and 12.7%, respectively). The average offer price of IPOs with a Wikipedia article is adjusted upward from the midpoint of the initial filing range by 1.76%. This compares to an average downward adjustment of 8.22% for IPOs without a Wikipedia article. Firms with a pre-IPO Wikipedia page are less likely to be backed by a venture capital investor and more likely to have their offer underwritten by a top-tier investment bank. *Overhang* indicates that 4.24 shares are retained for each share sold by firms with a pre-IPO Wikipedia article. This number is greater than the 3.16 shares retained per share sold for firms without a pre-IPO Wikipedia article. IPO firms with a Wikipedia article are more likely to have positive earnings and tend to report greater sales and total assets than IPO firms without a Wikipedia article. Thirteen percent more

IPO firms are classified as high-tech in the Wikipedia sample. IPO firms with a Wikipedia article also have a longer history, more news coverage, and higher levels of institutional ownership compared to those without. In sum, these results indicate that there are significant differences between IPOs with and without a Wikipedia article. Further, the underpricing difference between the two samples provides preliminary evidence in support of the investor attention hypotheses.

In Panel B, I report summary statistics for four Wikipedia-specific variables: *wiki_revisions*, *wiki_references*, *wiki_words*, and *traffic*. These variables suggest that there is substantial heterogeneity with respect to the amount of information aggregated and attention received for IPO firms with Wikipedia pages.

Table 1.1: Descriptive statistics

Panel A. Comparison of Wikipedia and non-Wikipedia IPO characteristics

	Wikipedia IPOs (N=330)		non-Wikipedia IPOs (N=644)			
	Mean	S.D.	Mean	S.D.	Difference	*t*-stat
underpricing	20.99	28.68	12.72	22.23	8.26	4.58***
offer_revision	1.76	19.35	-8.22	20.45	9.97	7.47***
up_revision	8.41	11.91	4.27	8.17	4.14	5.67***
VC	0.45	0.50	0.56	0.49	-0.10	-3.10**
top_tier	0.94	0.23	0.76	0.42	0.18	8.64***
overhang	4.24	2.54	3.16	2.12	1.08	6.61***
pos_EPS	0.48	0.50	0.34	0.47	0.14	4.07***
sales	1210.42	2369.09	274.48	759.11	935.94	7.00***
nasdaq15	0.69	3.30	0.99	3.14	-0.30	-1.36
tech	0.42	0.49	0.29	0.45	0.13	3.97***
age	24.82	27.24	15.73	19.29	9.09	5.40***
news	7.24	11.84	2.86	4.89	4.38	6.44***
assets	3557.96	18316.16	419.79	1613.58	3138.17	3.11**
proceeds	402.41	1293.14	142.97	204.66	259.44	3.62***
instown_pct_post	0.39	0.29	0.34	24.89	0.05	2.70**

Panel B. Wikipedia-specific variables

	n	Mean	S.D.	25th	Median	75th
wiki_revisions	330	18.01	35.28	2.00	7.00	17.00
wiki_references	330	15.87	26.12	3.00	8.00	20.00
wiki_words	330	539.68	621.41	191.00	343.00	626.00
traffic	286	1539.73	9997.81	57.00	222.00	582.00

1.4.2 Determinants of a Wikipedia article

Due to the differences between IPOs with and without a Wikipedia article reported in Table 1.1, I examine the determinants of the existence of a Wikipedia article when firms go public. Table 1.2 reports the results of a probit model with the dependent variable *Wikipedia* that is set equal to 1 for IPOs with a Wikipedia article, and zero otherwise. Independent variables include IPO and firm characteristics defined earlier and the number of news articles about the IPO firm between the S-1 filing and the IPO date (*log_news*).[11] I find that IPO firms with more sales, longer history, top tier underwriters, greater share overhang, and more news coverage are more likely to have a Wikipedia article when they go public. In sum, the evidence indicates that more established firms are more likely to have a Wikipedia article. To the extent that a Wikipedia article is merely a proxy for firm visibility, I would expect to observe lower underpricing for IPOs with a Wikipedia article because more established firms are less risky, on average (Loughran and Ritter, 2004). I address this issue next.

[11] In my sample, 316 out of 330 firms have a Wikipedia article when they file their Form S-1 and some variables in the regression equation are not known at that time. However, companies may choose to disclose financial information voluntarily to bolster the IPO. In addition, variables unknown at the time of S-1 filing are expected to be correlated with predictors that are known when Form S-1 is filed.

Table 1.2: Likelihood of having a Wikipedia page at IPO

VARIABLES	(1) Wikipedia
VC	0.0754
	(0.190)
top_tier	0.472***
	(0.143)
overhang	0.0621**
	(0.0314)
pos_EPS	-0.0255
	(0.161)
log_sales	0.117***
	(0.0444)
tech	0.301
	(0.210)
log_age	0.202***
	(0.0602)
log_news	0.193***
	(0.0595)
Constant	-2.511***
	(0.343)
N	974
Year FE	Yes
Pseudo R^2	0.207

1.4.3 IPO underpricing

In Table 1.1, I report that average IPO underpricing is 12.72% for IPOs without a Wikipedia article and 20.99% for IPOs with a Wikipedia article. The 8.26 percentage point difference is both economically and statistically significant. In Figure 3, I display the mean and median underpricing for IPO firms with and without a Wikipedia article over my sample period. In almost every year, the average and median underpricing is higher for IPO firms with a Wikipedia article than for firms without a Wikipedia article. Although underpricing of Wikipedia IPOs is more volatile, the two samples exhibit similar underpricing patterns. The exception is 2015 and 2016, when underpricing increases substantially for firms with a Wikipedia article and falls for firms without a Wikipedia article.

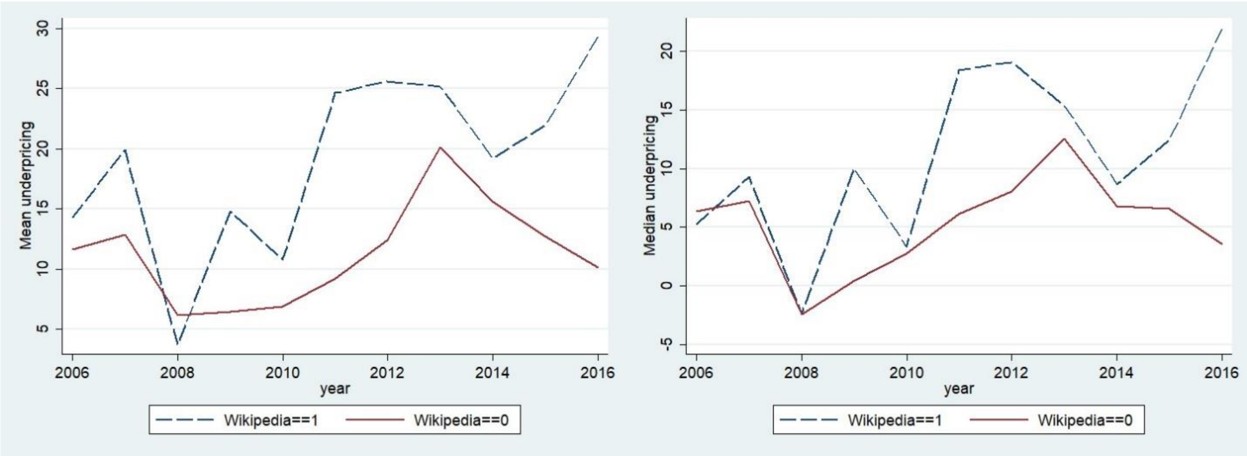


Figure 1.3: Underpricing by year for IPOs with and without a Wikipedia page

In Table 1.3, I report Ordinary Least Squares (OLS) regressions that control for other factors shown in prior studies to be correlated with IPO underpricing. All models also include year fixed effects (unreported) and robust standard errors clustered on IPO year and industry. In Model 1, I report the results of my baseline regression of underpricing on an indicator variable that is set equal to 1 for IPOs with a Wikipedia article, and zero otherwise (*Wikipedia*). Consistent with my univariate results, I find that the presence of a Wikipedia article is associated with underpricing that is 7.7 percentage points higher than for IPOs without a Wikipedia article. In Model 2, I control for the presence of a VC investor and top tier underwriter. I also control for share overhang, firm profitability, sales, recent market performance, presence in a high-tech industry, and the age of the IPO firm.[12] The presence of a Wikipedia article continues to be associated with higher underpricing. Specifically, IPO firms with a Wikipedia article experience underpricing that is 5.9 percentage points higher than IPO firms without a Wikipedia article. To illustrate the economic

[12] Replacing sales or assets with expected IPO market capitalization does not alter my results.

significance of this result, consider the average proceeds raised for my IPO sample ($230.87 million). A firm whose IPO is underpriced by an additional 5.9 percentage points due to the presence of a Wikipedia article would raise $13 million less than a similar firm without a Wikipedia article. In subsequent tests, I consider whether a Wikipedia article also provides long-term benefits to IPO firms that might justify the additional underpricing.

Table 1.3: IPO underpricing

VARIABLES	(1) underpricing	(2) underpricing
Wikipedia	7.686***	5.878**
	(1.788)	(2.162)
VC		8.479***
		(2.517)
top_tier		6.464***
		(1.162)
overhang		1.573***
		(0.295)
pos_EPS		4.414*
		(2.432)
log_sales		-0.0736
		(0.602)
nasdaq15		0.437
		(0.318)
tech		2.348
		(2.966)
log_age		-1.340
		(1.419)
N	974	974
Year FE	Yes	Yes
Adjusted R^2	0.035	0.112

Due to the significant differences between IPOs with and without a Wikipedia article, I also use propensity score matching (Rosenbaum and Rubin, 1983) to construct a matched sample of IPOs with and without a Wikipedia article. Matches are based on IPO firm characteristics and the number of news articles about the IPO firm between the S-1 filing and the IPO date (*log_news*). For each IPO firm with a Wikipedia article, I use the propensity score to identify the nearest match without replacement among IPO firms without a Wikipedia article. This procedure ensures that matched firms have similar characteristics. Table 1.4, Panel A compares the matched samples. I find that the samples are similar except for differences in the following variables: *overhang*, *log_sales,* and *log_news*. I report the results of the underpricing regressions for the matched sample in Table 1.4, Panel B. The results are similar to those reported in Table 1.3 for the full sample. Namely, IPO firms with a Wikipedia article exhibit greater underpricing than their matched counterparts without a Wikipedia article. For example, Model 2 indicates that the presence of a Wikipedia article is associated with first-day returns that are 4.5 percentage points higher.

Table 1.4: IPO underpricing (matched sample)

Panel A.Two-sample mean comparison after propensity score matching

Variable	Wikipedia	non-Wikipedia	Diff t-stat
VC	0.45	0.49	-0.86
top_tier	0.94	0.93	0.48
overhang	4.24	3.61	3.34***
pos_EPS	0.48	0.47	0.31
log_sales	5.41	4.91	2.73***
nasdaq15	0.69	0.90	-0.84
tech	0.42	0.42	0.16
log_age	2.83	2.74	1.46
log_news	1.30	1.03	2.94***

Panel B. OLS regression of underpricing for matched sample

	(1)	(2)
VARIABLES	underpricing	underpricing
Wikipedia	4.964**	4.467*
	(1.661)	(2.209)
VC		9.665***
		(1.978)
top_tier		9.729*
		(5.293)
overhang		1.587**
		(0.683)
pos_EPS		3.373
		(2.562)
log_sales		-1.013
		(0.762)
nasdaq15		0.607
		(0.359)
tech		-0.465
		(3.052)
log_age		-1.456
		(1.839)
log_news		2.646**
		(1.009)
N	660	660
Year FE	Yes	Yes
Adjusted R^2	0.024	0.134

1.4.4 Offer price revision

If an IPO firm's Wikipedia page aggregates information or is related to investor attention, I expect it to be associated with the offer price revision process (Benveniste and Spindt, 1989). In Table 1.5, Panel A, Column 1, I regress o*ffer_revision*, the percentage change from the midpoint of initial filing range to the final offer price, on *Wikipedia* and control variables. I follow Hanley (1993) and Hanley and Hoberg (2010) and include the following control variables: *offer_width* which is the difference between the high and low offer prices quoted in the preliminary prospectus divided by the low offer price; *shares_filed* which is the expected number of shares offered; *per_overallotment* which is the ratio of overallotment option to actual number of shares offered; and *instown_pct_post* which is the percent of institutional ownership at the end of first quarter after IPO.

I find that a pre-IPO Wikipedia article is positively related to the magnitude of the price revision. Specifically, IPO offer prices for firms with a Wikipedia article are revised upward 6.45% more, on average, than IPO offer prices for firms without a Wikipedia article. In Column 2, I examine the effect of Wikipedia-specific variables. The variable *wiki_aggregate* is the union of negative, uncertain, and litigious words based on the Loughran and McDonald (2011) dictionary. More negative sentiment is expected to be negatively related to offer price revision because it reflects higher risk. Due to high correlation between *log_wiki_revisions* and *log_wiki_references*, I only include *log_wiki_revisions.* For non-Wikipedia firms, *wiki_aggregate* and *log_wiki_revisions* are set equal to zero.[13] I find that *log_wiki_revisions* is positively related to offer price revisions. This suggests that, to the extent that Wikipedia revisions reflect higher quality information and more editing effort, IPO firms with higher quality information and more editing

[13] Results are similar if non-Wikipedia IPOs are omitted.

also have higher offer price revisions. It should be noted that this result is consistent with the hypothesis that a Wikipedia article helps to reduce the information asymmetry characterizing IPOs.

Table 1.5: Offer price revision

Panel A. OLS regression

VARIABLES	(1) offer_revision	(2) offer_revision
Wikipedia	6.454***	
	(1.526)	
wiki_aggregate		-0.030
		(0.479)
log_wiki_revisions		2.592**
		(0.908)
VC	0.830	0.707
	(2.859)	(2.940)
top_tier	5.378*	5.783**
	(2.436)	(2.503)
overhang	1.016***	0.954***
	(0.180)	(0.170)
pos_EPS	4.069*	4.070*
	(1.982)	(2.024)
log_sales	0.612	0.635
	(0.419)	(0.425)
nasdaq15	0.303	0.277
	(0.320)	(0.343)
tech	4.064	4.061
	(2.384)	(2.482)
log_age	-3.656***	-3.573***
	(0.673)	(0.738)
offer_width	-0.171	-0.167
	(0.173)	(0.179)
shares_filed	0.034	0.017
	(0.037)	(0.034)
per_overallotment	0.125	0.153
	(0.636)	(0.651)
instown_pct_post	0.028	0.027
	(0.050)	(0.053)
N	974	974
Year FE	Yes	Yes
Adjusted R^2	0.141	0.141

Panel B. Probit regression

VARIABLES	(1) pos_revision	(2) neg_revision	(3) pos_revision	(4) neg_revision
Wikipedia	0.344***	-0.297**		
	(0.129)	(0.122)		
wiki_aggregate			-0.023	0.044
			(0.042)	(0.055)
log_wiki_revisions			0.162**	-0.150**
			(0.066)	(0.064)
VC	0.265***	-0.203*	0.257**	-0.199*
	(0.101)	(0.109)	(0.106)	(0.107)
top_tier	0.616***	-0.276*	0.634***	-0.292*
	(0.148)	(0.159)	(0.157)	(0.157)
overhang	0.061***	-0.043***	0.057***	-0.040***
	(0.010)	(0.011)	(0.012)	(0.012)
pos_EPS	0.272	-0.246	0.272	-0.246
	(0.172)	(0.150)	(0.177)	(0.154)
log_sales	0.030	-0.018	0.031	-0.019
	(0.033)	(0.020)	(0.035)	(0.019)
nasdaq15	0.020	-0.029	0.019	-0.028
	(0.020)	(0.019)	(0.021)	(0.019)
tech	0.147	-0.265*	0.146	-0.268*
	(0.144)	(0.144)	(0.149)	(0.150)
log_age	-0.185***	0.198***	-0.183***	0.196***
	(0.063)	(0.067)	(0.059)	(0.070)
offer_width	-0.010	0.031**	-0.009	0.031**
	(0.014)	(0.013)	(0.013)	(0.013)
shares_filed	0.003**	-0.002	0.002	-0.001
	(0.001)	(0.002)	(0.002)	(0.002)
per_overallotment	-0.027	0.024	-0.024	0.021
	(0.054)	(0.057)	(0.056)	(0.059)
instown_pct_post	0.003*	-0.002*	0.003*	-0.002
	(0.001)	(0.001)	(0.001)	(0.002)
Constant	-1.000*	-0.728	-1.000*	-0.719
	(0.552)	(0.697)	(0.538)	(0.721)
N	974	974	974	974
Year FE	YES	YES	YES	YES
Pseudo R^2	0.120	0.091	0.123	0.093

In Table 1.5, Panel B, I examine the relation between a Wikipedia article and the direction of the offer price revision. The dependent variables are *pos_revision (neg_revision)*, which is an indicator variable set equal to 1 if *offer_revision* is positive (negative) and zero otherwise. In Columns 1 and 2, *Wikipedia* is associated with more upward offer price revisions and fewer downward offer price revisions. Again, *log_wiki_revisions* is positively related to upward offer price revisions and negatively related to downward offer price revisions. Taken together, these results suggest that the presence of a Wikipedia page helps to explain the magnitude and direction of an IPO firm's offer price revision. The results further suggest that offer price revisions are incomplete as evidenced by the positive and significant effect of the presence of a Wikipedia page on the IPO underpricing. These findings are consistent with Bradley and Jordan (2002), who find that IPO prices only partially adjust to public information before the IPO date. The positive relation between a Wikipedia article and IPO underpricing that I observe is consistent with my investor attention hypothesis (H2). Because prior research differs on the long-run impact of investor attention on IPO firms (Da et al., 2011; Liu et al., 2014), the next section explores this issue.

1.4.5 Investor attention

I use *log_news* as a proxy for investor attention to examine whether its inclusion affects the explanatory power of the Wikipedia indicator variable. The results in Table 1.6 Column 1 show that when *log_news* is added to the model, the magnitude and significance of the coefficient on *Wikipedia* decreases but the effect remains statistically significant. This suggests that *log_news* is positively associated with underpricing but that *Wikipedia* captures information beyond what is included in traditional news coverage.

Table 1.6: Wikipedia and investor attention

VARIABLES	(1) underpricing	(2) underpricing	(3) underpricing	(4) log_vol	(5) turnover
Wikipedia	4.950*		3.466**	0.452***	3.859***
	(2.328)		(1.425)	(0.080)	(1.096)
log_news	2.480**	1.889**			
	(0.823)	(0.728)			
log_traffic		1.633***			
		(0.511)			
up_revision			1.331***		
			(0.185)		
Wikipedia×up_revision			-0.164		
			(0.235)		
down_revision			0.339***		
			(0.094)		
Wikipedia×down_revision			0.205*		
			(0.099)		
VC	8.466***	7.140***	4.946**	-0.091	0.864
	(2.493)	(2.032)	(1.768)	(0.100)	(0.986)
top_tier	6.650***	6.879***	1.535*	1.016***	4.839***
	(1.195)	(0.928)	(0.814)	(0.096)	(0.804)
overhang	1.430***	1.172***	0.581	-0.000	-2.275***
	(0.293)	(0.341)	(0.373)	(0.014)	(0.312)
log_sales	-0.201	-0.343	-0.309	0.120***	0.430**
	(0.615)	(0.625)	(0.296)	(0.030)	(0.166)
nasdaq15	0.392	0.462	0.213	0.012	0.342*
	(0.324)	(0.367)	(0.324)	(0.008)	(0.167)
N	974	930	974	974	974
Year FE	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.121	0.133	0.428	0.531	0.218

To the extent that the existence of a Wikipedia article proxies for investor attention, an underlying assumption is that investors refer to Wikipedia to gather information about IPO firms. To provide additional evidence that Wikipedia proxies for investor attention, I extract the number of Wikipedia page views for each IPO firm on the IPO date from the Wikimedia Foundation (the non-profit which runs Wikipedia). Because pageview data are only available after December 2007, I restrict my Wikipedia traffic analysis to IPOs issued after this month (285 IPOs) ("Wikipedia:Pageview statistics", n.d.). The results in Table 1.6, Model 2 indicate that Wikipedia article traffic has considerable incremental explanatory power for underpricing. Specifically, a 10% increase in Wikipedia page views is associated with 0.32 percentage point increase in underpricing.[14] In Model 3 of Table 1.6, I examine information asymmetry by interacting the offer price revision measures with the presence of a Wikipedia article. The lack of (weak) significance on up (down) revisions is less consistent with partial information adjustment models.

If Wikipedia captures retail investor attention, I expect to see a positive association between *Wikipedia* and first-day trading volume. I use *log_vol* (natural logarithm of the first-day trading volume) and *turnover* (the ratio of trading volume to total number of shares outstanding) as proxies for retail investor attention. In Columns 4 and 5 of Table 1.6, I report that IPOs with a Wikipedia article experience significantly higher trading volume during the first day, consistent with the notion that these stocks grab more retail investor attention. The results reported in Table 1.6 are largely consistent with the hypothesis that the existence of a pre-IPO Wikipedia article increases investor attention for IPO firms, particularly among retail investors.

1.4.6 Investor attention and long-run performance

In this subsection I examine the relation between investor attention and long-run IPO performance. Prior research finds that investor attention is positively correlated with IPO underpricing (Da et al., 2011; Liu et al., 2014). However, these studies differ in terms of the long-run effect of investor attention on IPO firms. Specifically, Da et al. (2011) find that high initial returns are followed by long-run underperformance for IPOs that receive high investor attention

[14] Results are similar if non-Wikipedia IPOs (no Wikipedia traffic data) are omitted.

whereas Liu et al. (2014) provide evidence that investor attention has positive long-term effects for IPO firms.

The analysis reported in Table 1.7 follows the approach of Liu et al. (2014), with the primary difference being my proxy for investor attention. They use media coverage as their measure of investor attention, which I control for by using *log_news*, thereby allowing us to isolate the effect of the presence of a Wikipedia article. To the extent that investor attention from a firm's Wikipedia article has positive long-run effects, I would expect to observe a positive relation between *Wikipedia* and post-IPO analyst coverage and institutional ownership, my measures of long-term benefits (Liu et al., 2014). Equally as important, this would provide additional evidence that *Wikipedia* is measuring investor attention beyond what is captured by *log_news*. The remaining variables control for firm characteristics that may influence analyst following and institutional ownership.

Table 1.7: Long-run attention

	Analyst Following			Number of Institutional Investors		
	(1)	(2)	(3)	(4)	(5)	(6)
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
Wikipedia	0.739**	1.479***	1.429***	5.979**	11.329***	20.137**
	(0.272)	(0.279)	(0.388)	(1.975)	(3.336)	(7.776)
log_news	0.367	0.419*	0.697***	8.865***	14.814***	15.832***
	(0.232)	(0.209)	(0.192)	(1.088)	(1.996)	(3.039)
overhang	0.108	0.184	0.475	-0.583	-0.177	0.832
	(0.095)	(0.177)	(0.284)	(0.643)	(1.553)	(1.505)
pos_EPS	0.055	0.092	-0.196	5.730**	8.977**	10.729**
	(0.316)	(0.469)	(0.461)	(2.241)	(3.567)	(4.506)
log_sales	0.082	-0.009	0.081	0.238	1.203	2.115
	(0.059)	(0.067)	(0.115)	(0.694)	(1.260)	(2.000)
nasdaq15	-0.042	-0.074	-0.099	-1.233*	-1.379	-1.517
	(0.050)	(0.078)	(0.088)	(0.592)	(1.204)	(1.926)
tech	0.736*	0.880***	0.811	13.057***	17.703**	17.803**
	(0.340)	(0.263)	(0.583)	(3.986)	(5.496)	(5.696)
log_assets	0.757***	0.898***	0.682***	10.016***	17.676***	21.780***
	(0.176)	(0.104)	(0.159)	(1.287)	(2.322)	(3.183)
top_tier	0.330	0.188	0.833**	4.734*	7.838	7.870
	(0.208)	(0.399)	(0.345)	(2.464)	(4.449)	(4.537)
log_age	-0.686**	-0.851*	-0.873*	-1.951	-4.340	-5.576
	(0.240)	(0.384)	(0.454)	(2.834)	(4.571)	(6.316)
VC	0.415	0.926*	1.151*	7.403	12.022	12.946
	(0.410)	(0.469)	(0.604)	(4.992)	(7.151)	(10.640)
NASDAQ	-0.353	-0.951**	-1.022*	-7.154**	-5.927	-4.938
	(0.316)	(0.303)	(0.497)	(3.084)	(7.194)	(6.101)
AMEX	-0.875	-0.957	-2.157*	-28.219**	-18.343	-23.439
	(0.642)	(1.031)	(0.959)	(10.219)	(22.339)	(24.365)
N	967	860	740	820	732	630
Year and Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.350	0.380	0.330	0.365	0.400	0.399

The results are consistent with the notion that investor attention has positive long-run effects for IPO firms. Specifically, I find that firms with a pre-IPO Wikipedia article have greater analyst following and more institutional investors for at least 3 years after the IPO.[15] The

[15] Untabulated results show marginally significant increases in long-term firm value (Price/EBIT) and liquidity (turnover).

remaining variables are generally consistent with expectations. For example, size, underwriter reputation, and VC backing (firm age, NASDAQ-, and AMEX-listings) are positively (negatively) correlated with post-IPO analyst coverage and institutional ownership.

The evidence to this point indicates that a pre-IPO Wikipedia article is positively correlated with IPO underpricing and long-run performance. Next, I investigate whether there is a price reversal for IPOs that receive investor attention, as reported by Da et al. (2011). In Table 1.8, the dependent variable is the cumulative IPO return from weeks 5 to 52 after the IPO event. The remaining variables follow Da et al. (2011). The primary variables of interest are the Wikipedia indicator and its interaction with IPO underpricing.

Table 1.8: Post-IPO stock performance

VARIABLES	(1) cret	(2) cret	(3) cret	(4) cret	(5) cret	(6) cret
underpricing	0.038	0.083	0.138	0.119	0.016	0.318
	(0.122)	(0.227)	(0.163)	(0.174)	(0.090)	(0.297)
Wikipedia	1.561	0.165	2.149	2.236	0.992	-1.745
	(3.183)	(3.784)	(3.414)	(3.303)	(2.918)	(2.775)
Wikipedia × underpricing		0.082				-0.237
		(0.254)				(0.261)
log_news	4.366	4.294	6.105	4.430	4.485	5.875
	(2.858)	(2.931)	(3.925)	(2.875)	(2.836)	(4.072)
log_news × underpricing			-0.100			-0.090
			(0.197)			(0.202)
offer_revision	-0.409	-0.408	-0.405	-0.318	-0.396	-0.307
	(0.228)	(0.269)	(0.229)	(0.371)	(0.270)	(0.356)
offer_revision × underpricing				-0.007		-0.007
				(0.013)		(0.014)
d_sentiment	-0.027	-0.034	-0.049	-0.052	-0.427	-0.407
	(0.794)	(0.785)	(0.807)	(0.850)	(0.851)	(0.867)
d_sentiment × underpricing					0.023	0.019
					(0.023)	(0.025)
top_tier	20.779	20.853	20.822	20.398	20.190	20.197
	(9.781)	(9.734)	(9.808)	(9.908)	(9.558)	(9.753)
VC	-3.803	-3.679	-3.739	-3.735	-3.418	-3.016
	(6.168)	(6.126)	(6.193)	(6.012)	(6.288)	(6.153)
overhang	-1.422	-1.470	-1.327	-1.384	-1.367	-1.394
	(1.047)	(1.181)	(1.238)	(1.060)	(1.045)	(1.297)
cret_pre_ind	-0.472	-0.473	-0.468	-0.469	-0.466	-0.464
	(0.674)	(0.631)	(0.633)	(0.595)	(0.634)	(0.612)
N	831	831	831	831	831	831
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.027	0.026	0.028	0.028	0.028	0.028

The evidence reported in Table 1.8 indicates that investor attention from applicable Wikipedia content is not correlated with post-IPO returns (i.e., the coefficient on the interaction of the Wikipedia indicator and underpricing is not statistically significant). This differs from Da et al. (2011), who report price reversals for IPOs with both high investor attention and high first-day returns. These results, combined with those reported in Table 1.7, support prior research which posits that pre-IPO investor attention has lasting positive benefits for IPO firms, and as such does not represent overreaction but rather firm fundamentals.

1.4.7 Wikipedia sentiment

I also consider the possibility that my results are driven by the sentiment expressed in firms' Wikipedia articles. I examine differences between the language used in firms' S-1 registration statements and their Wikipedia articles. Given the different nature of Wikipedia articles and S-1 registration statements, I employ both the Loughran and McDonald (2011) (LM) dictionary, which is specifically designed for financial filings, and the Harvard General Inquirer (Harvard GI) dictionary, which is widely used in social science research. Loughran and McDonald (2013) find that IPO firms with a high level of aggregate uncertainty in S-1 filings experience larger first-day returns. I measure sentiment as the percentage of words in a corresponding sentiment dictionary for each firm's Wikipedia article and S-1 filing.

In Table 1.9 Panel A I report the average percentage of words in each sentiment dictionary for both Wikipedia articles and S-1 filings for the 330 sample firms with a Wikipedia article at the time of their IPO. Consistent with issuers' intention to lower litigation risk, S-1 filings are characterized by a negative tone. Specifically, the average percentage of negative words is almost double that of positive words. Note also that Wikipedia articles tend to contain less negative sentiment according to both dictionaries.[16] Overall, these univariate results indicate that Wikipedia and S-1 convey information differently.

---

[16] I report the 20 words that appear most frequently and the corresponding percentage in each sentiment category for both the LM and Harvard GI dictionaries in Appendix D.

Table 1.9: Sentiment analysis

Panel A. S-1 and Wikipedia sentiment comparison

Comparison with LM dictionary				
	WIKIPEDIA	S-1	Difference	t-stat
positive	0.71	0.80	-0.09	-2.61***
negative	0.84	1.54	-0.70	-12.95***
net (pos-neg)	-0.13	-0.74	0.61	9.49***
uncertainty	0.36	1.42	-1.06	-41.27***
litigious	0.29	0.97	-0.68	-22.43***
Comparison with Harvard GI dictionary				
	WIKIPEDIA	S-1	Difference	t-stat
positive_GI	1.81	1.79	0.02	0.23
negative_GI	0.61	0.88	-0.27	-6.26***
net_GI	1.20	0.92	0.29	3.28***

Panel B.** **IPO Underpricing and Wikipedia sentiment

	(1) Positive	(2) Negative	(3) Net	(4) Uncertainty	(5) Litigious	(6) Positive_GI	(7) Negative_GI	(8) Net_GI
sentiment variable	2.35	0.19	1.05	3.20	-0.82	1.71*	-0.86	2.12*
	(2.17)	(0.61)	(1.39)	(2.76)	(1.07)	(0.85)	(0.97)	(0.98)
VC	8.50**	8.50**	8.46**	8.42**	8.47**	8.44***	8.50**	8.44***
	(2.77)	(2.78)	(2.82)	(2.75)	(2.79)	(2.63)	(2.80)	(2.62)
top_tier	6.92***	7.16***	7.13***	7.09***	7.19***	6.79***	7.18***	6.72***
	(1.59)	(1.40)	(1.41)	(1.25)	(1.46)	(1.36)	(1.46)	(1.41)
overhang	1.70***	1.72***	1.74***	1.68***	1.73***	1.64***	1.73***	1.64***
	(0.36)	(0.42)	(0.37)	(0.37)	(0.41)	(0.37)	(0.35)	(0.38)
pos_EPS	4.41	4.42	4.38	4.49	4.37	4.35	4.39	4.28
	(2.53)	(2.59)	(2.61)	(2.51)	(2.62)	(2.48)	(2.64)	(2.51)
log_sales	0.07	0.10	0.12	0.05	0.12	0.02	0.13	0.04
	(0.60)	(0.61)	(0.62)	(0.64)	(0.62)	(0.60)	(0.61)	(0.62)
nasdaq15	0.40	0.39	0.39	0.40	0.39	0.42	0.39	0.42
	(0.30)	(0.31)	(0.30)	(0.34)	(0.35)	(0.29)	(0.30)	(0.29)
tech	2.65	2.85	2.93	2.86	2.89	2.63	2.94	2.71
	(3.17)	(3.15)	(3.20)	(3.19)	(3.22)	(3.06)	(3.19)	(3.01)
log_age	-1.00	-0.99	-0.89	-1.00	-0.93	-1.11	-0.93	-1.05
	(1.46)	(1.42)	(1.43)	(1.47)	(1.41)	(1.43)	(1.42)	(1.42)
N	974	974	974	974	974	974	974	974
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.104	0.102	0.103	0.103	0.102	0.107	0.102	0.110

To provide further insight on this relation I regress underpricing on Wikipedia article sentiment. The results, which are reported in Table 1.9 Panel B, show that positive and net Harvard GI words are positively correlated with first-day returns. Thus, to the extent that Wikipedia articles with general positive tone increase investors' attention to the IPO, these results are consistent with the investor attention hypothesis. The lack of significance of LM dictionaries suggests that investor attention is driven by general content within Wikipedia articles as opposed to financial context-specific content that is more prevalent in S-1 filings.[17]

1.4.8 Underwriter promotion

Given that underwriters assume the risk of selling IPO shares, they are incentivized to act proactively by publicizing the IPO so as to avoid costly price stabilization in the secondary market (Ruud, 1993). If underwriters initiate Wikipedia articles for IPO companies to facilitate the selling of shares, the existence of a pre-IPO Wikipedia page would not be exogenous. According to the editing policy of Wikipedia, insider editing is discouraged although not completely prohibited. Wikipedia requests editors to reveal conflicts of interest that they have with the company and to first discuss edits on the article's "talk page" to get the approval of the Wikipedia community. Despite this request, conflict of interest editing has happened several times in Wikipedia's history. However, the rigorous detection mechanisms discussed previously help to ensure that biased content is removed promptly.[18] In addition, because underwriters are more likely to initiate the IPO firm's Wikipedia page during the bookbuilding period, I examine the time of page creation relative to the S-1 filing date. Of the 330 IPOs in the Wikipedia sample, 316 have a Wikipedia page prior to registration with SEC. Figure 1.4 reports the difference between the S-1 filing date and the date of Wikipedia page creation in days. A positive difference indicates the Wikipedia

[17] Untabulated results show no influence on underpricing by differences in sentiment between S-1 filings and Wikipedia articles.
[18] For example, Wikipedia investigated allegations that Bell Pottinger, the largest UK public relations firm, manipulated its clients' Wikipedia articles.

article is created before the S-1 filing. I find that very few Wikipedia article are created around the S-1 filing date, which alleviates the "self-promotion" concern.[19]

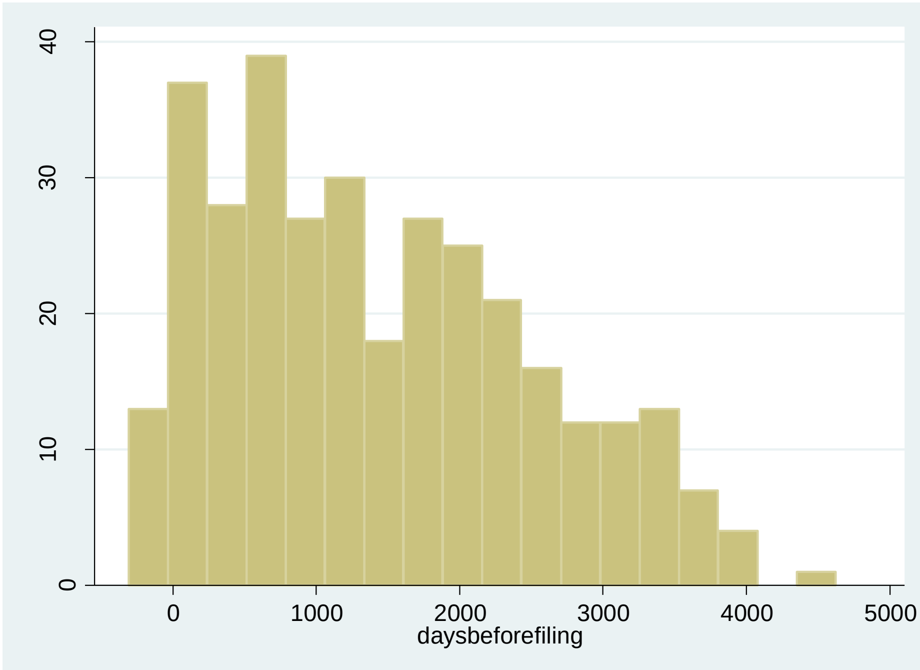


Figure 1.4: Time differences between Wikipedia article creation and S-1 filing (in days)

1.5 Endogeneity concerns

Given the different characteristics of the Wikipedia and non-Wikipedia samples, the existence of a Wikipedia article may not be random. That is, the relation could be endogenous. I employ two identification strategies to account for unobservable variables that could affect both the existence of a Wikipedia article and IPO underpricing: endogenous treatment-regression model and instrumental variables approach.

[19] The results are qualitatively unchanged when I estimate my baseline results using the sample of firms with Wikipedia pages created before the IPO filing date.

1.5.1 Endogenous treatment-regression model

Table 1.10 Panel A contains the results of an endogenous treatment regression. In Model 1, *Wikipedia* is regressed on *VC*, *top_tier*, *log_sales*, *tech*, *log_age*, and *log_news* in a probit model. To control for the endogeneity of the *Wikipedia* indicator variable, the residuals of Model 1 are included in Model 2 where IPO underpricing is the outcome variable and *Wikipedia* is an endogenous treatment variable (Heckman, 1978). Other control variables are identical to the baseline model reported in Table 1.3. The coefficient of the hazard calculated from the treatment model is not significant, indicating that residuals of the selection and outcome models are not significantly correlated, thus, mitigating endogeneity concerns. The results indicate that a Wikipedia page is associated with an 18 percentage point percent increase in underpricing, which is much greater than the estimated coefficient in Table 1.3 Model 2. The underestimation of the coefficient of the Wikipedia indicator variable based on the negative sign of the hazard suggests that unobservable variables that increase the likelihood of having a Wikipedia page also decrease underpricing. This indicates that the significant increase in underpricing due to the existence of a Wikipedia page is not due to unobservable variables.

Table 1.10: Endogeneity

Panel A. Endogenous treatment regression of underpricing

VARIABLES	(1) underpricing	(2) Wikipedia
Wikipedia	18.440**	
	(8.115)	
overhang	1.461***	
	(0.359)	
pos_EPS	4.330**	
	(1.823)	
nasdaq15	0.413*	
	(0.243)	
VC	8.347***	0.113
	(1.910)	(0.117)
top_tier	4.845**	0.508***
	(2.432)	(0.153)
log_sales	-0.496	0.127***
	(0.421)	(0.022)
tech	1.061	0.354***
	(1.983)	(0.105)
log_age	-2.058*	0.173**
	(1.238)	(0.068)
log_news		0.208***
		(0.0447)
hazard	-7.769	
	(4.886)	
Constant	2.323	-2.970***
	(4.846)	(0.291)
N	974	974

Panel B. Instrumental variable regression of underpricing

	Underpricing	
	First stage (1)	Second stage (2)
log_articles_new	-0.288***	
	(0.0496)	
Wikipedia		24.334**
		(10.000)
VC	0.013	8.402***
	(0.034)	(1.966)
top_tier	0.121***	4.002
	(0.039)	(2.594)
Overhang	0.026***	1.151**
	(0.006)	(0.486)
pos_EPS	0.008	3.685*
	(0.033)	(1.920)
log_sales	0.031***	-0.665
	(0.006)	(0.449)
nasdaq15	-0.008*	0.587**
	(0.004)	(0.254)
Tech	0.090***	0.437
	(0.032)	(1.959)
log_age	0.066***	-2.286*
	(0.020)	(1.351)
Constant	1.828***	1.789
	(0.352)	(4.193)
N	974	974
R^2	0.197	0.014

1.5.2 Instrumental variable approach

If IPO firms choose to create and edit their own Wikipedia article and this behavior is correlated with IPO underpricing, I might mistakenly infer a Wikipedia causal effect. Moreover, given the information aggregation function of Wikipedia, if there exists another information channel that affects both the likelihood of an IPO firm having a Wikipedia article and IPO underpricing, then I cannot attribute the effect on underpricing to Wikipedia. To address this endogeneity concern, I use the average number of daily "new English" Wikipedia articles created during each month as an instrumental variable for the *Wikipedia* indicator variable. This instrument is correlated with the *Wikipedia* indicator variable because it captures the contribution intensity of the Wikipedia community. However, it is unlikely to be correlated with IPO underpricing because Wikipedia encompasses 12 large categories of articles and an average of 800 (range of 700 to 2,112 in my sample) new articles are created each day. Company-related articles represent a small component of newly created Wikipedia articles ("Wikipedia:Statistics", n.d.; "Wikipedia:Contents/Categories", n.d.). I expect that fewer company specific pages will be created when special events happen that grab the Wikipedia community's attention. The results in Table 1.10 Panel B confirm this expectation. In the first stage of the two-stage model, *log_articles_new* is negatively correlated with *Wikipedia*. In the second stage, the instrumented *Wikipedia* is strongly associated with underpricing. The results support the hypothesis that Wikipedia captures investor attention of IPO firms, which results in larger first-day returns.

1.6 Robustness checks

1.6.1 Correlation with other investor attention measures

1.6.1.1 Google search volume

Da et al. (2011) find that Google Search Volume Index (SVI) has low correlation with other attention measures including abnormal returns, turnover, and news coverage. In addition, they show that SVI mainly captures retail investor attention. Given that Wikipedia articles rank high in Google search results, one concern is that Wikipedia traffic is simply a derivative of Google search activity. I follow Da et al. (2011) approach to construct abnormal SVI (*ASVI*), which is the natural logarithm of SVI during the IPO week minus the median SVI during the prior eight weeks:

$$ASVI_t = \log(SVI_t) - \log(Med(SVI_{t-1}, \ldots, SVI_{t-8})) \quad (1.1)$$

ASVI captures the attention jump due to the IPO event reflected in Google searches. To determine a search for an IPO company, I start with company names in SDC and then match to search terms based on how investors might search for the company in Google. Notably, to capture the attention of investors rather than consumers, I keep suffixes like “Inc” or “Corp” for companies in retail or service industries. For example, the search term of *Tumi Inc* is assumed to be for the firm *Tumi Inc* because search volume of *Tumi* mainly comes from people who are interested in suitcases or bags (i.e., consumers). In addition, retail and service companies may have strong seasonality that could obscure the change in investors’ search volume around an IPO if I use product name or brand name.

I obtain ASVI for 912 IPO companies.[20] Neither of the correlation between *ASVI* and *Wikipedia* or the correlation between *ASVI* and *log_traffic* is significant. There are several reasons

[20] Missing ASVI is due to two reasons. First, 15 IPOs do not have initial SVI data (959 IPOs left). Second, $\log(SVI_t)$ or $\log(Med(SVI_{t-1}, \ldots, SVI_{t-8}))$ is missing when SVI_t or $Med(SVI_{t-1}, \ldots, SVI_{t-8}$ is 0 (912 IPOs left).

for these weak correlations. For instance, one might interpret *ASVI* as a measure of attention change and *Wikipedia* as a measure of the attention level. Second, many people search for companies in Google for reasons other than investing (e.g., researching corporate history, looking for product information). Third, Wikipedia article traffic could indicate a more serious interest in buying IPO stocks. In other words, people who visit an IPO company's Wikipedia page are more likely to buy the company's stock compared to those searching for the company in Google. Table 1.11 Column 1 reports regression results where I add *ASVI* to the baseline model (Table 3, Column 2). The estimated coefficient and significance of Wikipedia dummy are similar to those in baseline regression. However, *ASVI* does not predict underpricing. An interpretation of the discrepancy between Google and Wikipedia attention is that *ASVI* captures a shock to attention that decays relatively quickly and it is less representative of IPO investing intention compared with IPO firm's Wikipedia page usage. In contrast, Wikipedia provides a stable information outlet that aggregates relevant information over time with regard to the pending IPO, resulting in an increase of a company's investor base.

Table 1.11: Robustness check

VARIABLES	(1) underpricing	(2) integer	(3) underpricing	(4) underpricing
asvi	0.086			
	(0.074)			
pctnum_s1			-5.013***	
			(1.204)	
pctnum_wiki				-0.715
				(0.812)
Wikipedia	5.835**	0.019	6.117**	
	(2.232)	(0.127)	(2.187)	
VC	8.993***	0.419***	6.440**	10.390**
	(2.581)	(0.110)	(2.569)	(4.796)
top_tier	7.024***	0.109	5.637***	-1.529
	(1.306)	(0.190)	(1.081)	(8.100)
overhang	1.605***	-0.004	1.521***	1.824*
	(0.276)	(0.021)	(0.294)	(0.938)
pos_EPS	4.078	-0.128	3.869	4.960
	(2.492)	(0.098)	(2.484)	(3.975)
log_sales	-0.026	0.022	0.170	-4.874***
	(0.614)	(0.047)	(0.597)	(1.589)
nasdaq15	0.532	-0.002	0.390	1.632***
	(0.320)	(0.023)	(0.308)	(0.593)
tech	2.305	-0.397***	2.349	-12.916***
	(2.972)	(0.057)	(2.985)	(4.737)
log_age	-1.094	0.093	-0.987	-1.748
	(1.474)	(0.106)	(1.467)	(2.979)
log_wiki_revisions				0.805
				(1.237)
lambda				-34.880***
				(12.562)
Constant	-2.872	1.119***	19.786***	84.334**
	(3.884)	(0.308)	(3.278)	(37.115)
Observations	912	974	974	974
Year FE	Yes	Yes	Yes	Yes
Adjusted R-squared	0.115		0.132	0.050
Pseudo R-squared		0.0972		

1.6.1.2 Institutional investor attention

Ben-Rephael, Da, and Israelsen (2017) find that Bloomberg activity is an effective measure of institutional investor attention. I examine the relation between the presence of a pre-IPO Wikipedia article and Bloomberg activity for my IPO sample. Specifically, I follow Ben-Rephael et al. (2017) and obtain "News Heat-Daily Max Readership" from Bloomberg for the IPO date. This measure captures attention spikes relative to the prior month. Because the data are only available since 2015, I can only retrieve heat data for 187 (out of 974) of my sample IPOs. The correlation between the presence of a Wikipedia article and the heat variable is insignificant (-0.016). Thus, I contend that the presence of pre-IPO Wikipedia article captures a different component of investor attention than the number of news articles, Google search volume, and Bloomberg activity measures used in prior studies.

1.6.2 IPO offer price precision

Bradley, Cooney, Jordan, and Singh (2004) use non-integer offer price as an indicator of high price precision and therefore as a measure of information asymmetry. To examine if a Wikipedia article helps to mitigate information asymmetry and improve price precision, I construct an indicator variable, *integer*, for IPOs priced on an integer and regress it on *Wikipedia*. Table 1.11 Column 2 shows that *Wikipedia* is not a significant determinant of integer offer prices. This provides additional evidence that the primary effect of a Wikipedia article is not a reduction in information asymmetry during the bookbuilding/price setting process.

1.6.3 Qualitative vs. quantitative information

Managers have incentives to delay the disclosure of bad news (Kothari, Shu, and Wysocki, 2009) and to publicize good news (Solomon, 2012). One technique managers apply to hide bad performance is to use "soft talk", which is equivocal, unverifiable, and biased upward (Dambra, Wasley, and Wu, 2013). Consequently, markets react more strongly to numbers (Hutton, Miller, and Skinner, 2003). To examine the effect of quantitative information in Wikipedia articles and S-1 filings, I count numbers that start with a space and can consist of numeric characters (0-9), plus and minus signs (+ and -), currency symbols ($€£), commas (,) and periods (.). In addition, I count the total number of words identified in the Loughran and McDonald (2011) 10-K dictionary. I exclude numbers in the range of 1900 to 2020 to avoid counting years. The measure of quantitative information is constructed as follows:

$$pctnum = \frac{Total\ count\ of\ numbers}{Total\ count\ of\ numbers + total\ count\ of\ words} \times 100 \qquad (1.2)$$

Results in Table 1.11 Model 3 show that a 1% increase in numbers in a company's S-1 is associated with a 4.91 percentage point decrease of underpricing, consistent with the argument that numbers convey information that can mitigate information asymmetry. The percentage of numbers in Wikipedia articles, however, is not significantly related to underpricing, which suggests that investors are more attentive to general qualitative information within IPO firms' Wikipedia articles. Notice also that the total number of words is not associated with underpricing.

1.6.4 Industry variation

I next examine industry dispersion for IPO firms with a Wikipedia page. Because a Wikipedia article is a collaborative work, I expect that companies with products more closely related to everyday life are more likely to have a Wikipedia page. Figure 1.5 reports the proportions for the Top 10 industries in the Wikipedia sample (blue bar) based on Fama and French (1997) 48-industry classifications and the corresponding proportion for each industry in the full IPO sample (red bar). If an industry's proportion of the Wikipedia sample is greater than the proportion in the full sample, it indicates that the industry is overrepresented in Wikipedia sample. Consistent with my expectations, the results indicate that IPO firms in business services, retail, restaurants and hotels, computers, communication entertainment, and personal services industries are more likely to have a Wikipedia page compared to those in pharmaceutical products and electronic equipment industries.

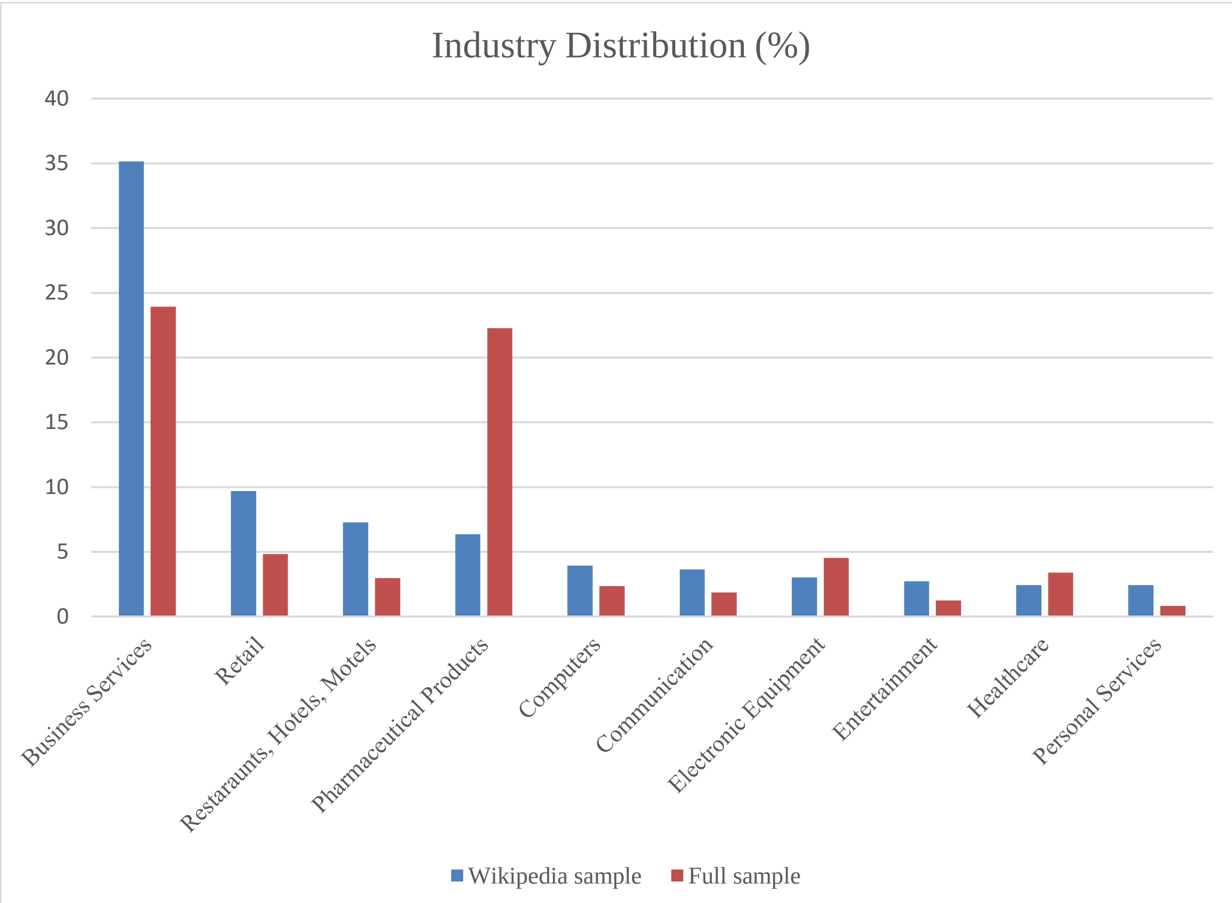


Figure 1.5: Industry distribution of IPO firms

1.6.5 Quiet period expiration

Bradley, Jordan, and Ritter (2003) find that firms with analyst coverage experience a 4.1% cumulative abnormal return after the expiration of the IPO quiet period, compared to 0.1% for those without analyst coverage. Moreover, pre-event run-ups indicate that the market can predict whether analysts will cover a firm. Given the short time period between the quiet period expiration date and the IPO date (39.11 days in my sample), I use the market-adjusted return as the abnormal return measure. Because 64.6% of IPOs in my sample are listed on Nasdaq, I use the Nasdaq Composite Index return as my market return benchmark. Figure 1.6 demonstrates the cumulative abnormal return (*CAR*) for the Wikipedia and non-Wikipedia samples over the [-10,+10] window surrounding the end of the quiet period. Both samples experience a run-up that starts two days prior to the expiration date and partially reverts two days after the quiet period expires.

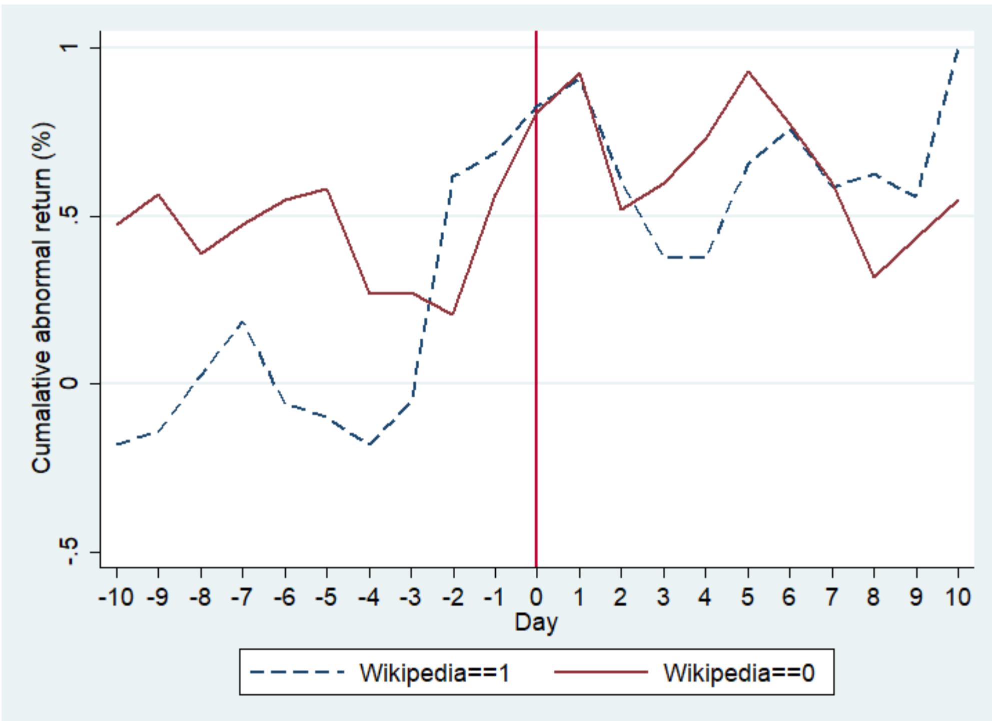


Figure 1.6: Cumulative abnormal returns: Wikipedia and non-Wikipedia samples

Table 1.12 reports the daily *MARs* and *CMARs* (Panel A) over various windows around the expiration of the quiet period (Panel B). The Wikipedia sample has a significant *MAR* on Day -2, while the non-Wikipedia sample has stronger fluctuations but a weak overall magnitude. When I examine multi-day windows, I find that *CMAR* is significant for both samples over the [-2,+1] window; however, the results are different depending on the event window. In sum, I do not observe a strong market reaction around the quiet period expiration date or a difference in reactions between Wikipedia and non-Wikipedia firms.

Table 1.12: Quiet period expiration event study results

Panel A. Market -adjusted returns (MAR)

	Wikipedia (N=330)		non-Wikipedia (N=644)	
Day	Average MAR (%)	t-stat	Average MAR (%)	t-stat
-10	-0.18	-0.83	0.47	2.67***
-9	0.04	0.18	0.09	0.52
-8	0.17	0.98	-0.18	-1.20
-7	0.16	0.94	0.09	0.57
-6	0.25	-1.47	0.07	0.45
-5	-0.04	-0.21	0.03	0.20
-4	-0.08	-0.45	-0.31	-2.25**
-3	0.13	0.67	0.00	0.02
-2	0.67	3.38***	-0.07	-0.46
-1	0.07	0.40	0.36	2.54**
0	0.14	0.71	0.24	1.37
1	0.08	0.34	0.12	0.67
2	-0.30	-1.70*	-0.41	-3.01***
3	-0.22	-1.27	0.08	0.55
4	-0.00	-0.02	0.13	0.60
5	0.28	1.47	0.20	1.19
6	0.10	0.51	-0.16	-0.96
7	-0.17	-0.86	-0.17	-1.19
8	0.04	0.20	-0.28	-1.75*
9	-0.07	-0.40	0.12	0.84
10	0.44	2.32**	0.11	0.78

Panel B. Cumulative market-adjusted returns (CMAR)

	Wikipedia (N=330)		non-Wikipedia (N=644)	
Window	Average CMAR (%)	t-stat	Average CMAR (%)	t-stat
[-2,+2]	0.65	1.43	0.25	0.69
[-2,-1]	0.74	2.91***	0.29	1.42
[-2,+1]	0.96	2.36**	0.65	1.95*
[0,+2]	-0.09	-0.23	-0.04	-0.15
[-1,+1]	0.29	0.79	0.72	2.32**
[-10,+10]	0.99	1.11	0.55	0.74

1.6.6 Long-run performance

I follow Ritter and Welch (2002) to examine long-run performance of IPOs. First, I compare the buy-and-hold abnormal return between the Wikipedia and non-Wikipedia samples.[21] For each IPO, I calculate the buy-and-hold return (*BHR*) from the first closing price to the earlier of the three-year IPO anniversary (756 trading days) or the last available price on CRSP. Buy-and-hold abnormal return (*BHAR*) is the difference between *BHR* of IPOs and the compounded daily return of the CRSP value-weighted index (*BHRM)*. In Table 1.13 Panel A1, I report that IPOs experience a 15.51% average three-year buy-and-hold return, which is 2.55% lower than the CRSP value-weighted return. Panel A2 compares the *BHR* and *BHAR* between Wikipedia and non-Wikipedia samples. The two samples do not show a significant *BHR* (*BHAR*) difference.

[21] The sample period used in Table 1.13, Panel A ends in 2013 to ensure that the long-run performance window is consistent for IPOs through all years.

Table 1.13: IPO long-run performance

Panel A. Three-year buy-and-hold return for IPO from 2006 to 2013

Panel A1. Summary statistics

Variable	n	Mean	S.D.	Min	0.25	Median	0.75	Max
BHR	641	15.51	126.66	-99.89	-62.96	-13.56	55.50	1556.86
BHRM	641	18.06	30.64	-50.02	-13.48	26.11	44.76	78.22
BHAR	641	-2.55	122.45	-155.38	-70.86	-25.53	33.27	1533.65

Panel A2. Average three-year buy-and-hold return

	non-WIKIPEDIA (N=427)	WIKIPEDIA (N=214)	Difference	t-stat
BHR	10.77	24.95	-14.18	-1.50
BHAR	-2.22	-3.20	0.98	0.11

Panel B. Multifactor regression of equally-weighted IPO portfolio returns

Panel B1. Full sample

VARIABLES	(1) ret_rf	(2) ret_rf	(3) ret_rf
intercept	-0.000	-0.013	-0.034
	(0.449)	(0.339)	(0.322)
mktrf	1.351***	1.198***	1.139***
	(0.103)	(0.091)	(0.088)
smb		1.212***	1.221***
		(0.149)	(0.141)
hml		-0.382***	-0.517***
		(0.128)	(0.128)
mom			-0.228***
			(0.068)
N	96	96	96
R^2	0.646	0.805	0.827

Panel B2. Wikipedia vs. non-Wikipedia

	Wikipedia			non-Wikipedia		
VARIABLES	(1) ret_rf	(2) ret_rf	(3) ret_rf	(4) ret_rf	(5) ret_rf	(6) ret_rf
Constant	-0.165	-0.167	-0.184	0.068	0.056	0.034
	(0.409)	(0.305)	(0.292)	(0.520)	(0.413)	(0.398)
mktrf	1.310***	1.158***	1.110***	1.375***	1.208***	1.146***
	(0.094)	(0.082)	(0.080)	(0.120)	(0.111)	(0.109)
smb		1.136***	1.144***		1.312***	1.323***
		(0.134)	(0.128)		(0.182)	(0.175)
hml		-0.329***	-0.438***		-0.406**	-0.548***
		(0.115)	(0.116)		(0.156)	(0.158)
mom			-0.185***			-0.239***
			(0.062)			(0.084)
N	96	96	96	96	96	96
R^2	0.673	0.825	0.841	0.584	0.747	0.768

Due to the overlap of buy-and-hold returns, I follow Ritter and Welch (2002) and conduct a 4-factor time-series regression. The dependent variable is the equally-weighted monthly return in excess of the risk-free rate for the portfolio of IPOs that go public during the prior 36 months.[22] The regression period is 96 months (2009 through 2016). I report the results in Table 1.13, Panel B1. The intercept in Column 3 implies an abnormal return of -3.4 basis points per day (0.4% per year). Although, none of the estimated intercept coefficients are statistically significant, they are nevertheless economically meaningful. Next, I construct portfolios based on the Wikipedia indicator variable and repeat the time-series regressions. I report the results in Table 1.13 Panel B2. Neither the Wikipedia nor the non-Wikipedia subsamples exhibit significant abnormal returns. However, non-Wikipedia IPOs have a relatively higher abnormal return compared with Wikipedia IPOs, consistent with the investor attention model prediction that "more widely-known firms with larger investor bases will have lower alphas" (Merton, 1987).

1.6.7 Alternative information channels

Wikipedia is not the only online platform where an IPO company might receive investor attention. Other platforms include company websites, Facebook, Twitter, etc. Company websites are one of the primary information sources for investors and I find that all sample firms have a company page prior to their IPO. Given the lack of cross-sectional variability, the existence of a company's website is not likely to explain the IPO underpricing difference between the Wikipedia and non-Wikipedia samples.

It is also possible that Wikipedia traffic is correlated with the traffic of other social media platforms. However, unlike a company's website or social media page, a Wikipedia article is not likely to be created or regularly modified by the firm itself. Therefore, Wikipedia articles are less

[22] One issue with calendar-time regression is look-ahead bias. Specifically, an IPO that stops trading prior to its three-year anniversary is excluded from the sample. For example, Traffic.com went public on Jan 25, 2006 and the last available return is on March 6, 2007.

likely to be biased and thus the information provided are more likely to be interpreted as credible by investors.

1.6.8 Emerging growth companies

I note that the JOBS Act was passed during my sample period with the goal of facilitating IPOs of emerging growth companies (EGCs). I examine whether Wikipedia has a different effect on underpricing for EGCs and non-EGCs. Because the JOBS Act was signed on Apr 5, 2012, I only include IPOs issued after that date. I add an EGC indicator and its interaction with the *Wikipedia* indicator to my baseline regressions. In untabulated results, I do not find a difference in the effect of Wikipedia on EGCs and non-EGCs.

1.7 Conclusion

I investigate the impact of Wikipedia on initial public offerings. Because a firm's Wikipedia article is a collaborative effort of the Wikipedia community, it is a potentially valuable source of information beyond the carefully crafted regulatory filings that accompany IPOs. On the one hand, a Wikipedia article may reduce information disparities among IPO participants, which allows for more precise offer prices and also mitigates information effects that contribute to underpricing (Ljungqvist, 2007). On the other hand, Wikipedia has the potential to increase investor attention of IPO firms which prior research has associated with larger first-day returns (Da et al., 2011; Liu et al., 2014).

I find that firms that have a Wikipedia article when they go public experience significantly higher underpricing than firms without a Wikipedia article. This underpricing effect is greater when the firm's Wikipedia page receives more visits and contains more qualitative information. I draw on prior research on investor attention to explain the positive link between a Wikipedia article and IPO underpricing. Da et al. (2011) find that high initial returns are followed by long-run underperformance for IPOs that receive high investor attention, while Liu et al. (2014) provide evidence that investor attention has positive long-term effects for IPO firms. Consistent with the latter study, I find that IPO firms with a Wikipedia article benefit from greater analyst following and attract more institutional investors for up to three years following the offering compared to IPO firms without a Wikipedia article. Importantly, these results are robust to a battery of robustness checks. While my results for the reduction of information asymmetry are marginal, my results are consistent with the Merton (1987) investor attention model which predicts that higher investor attention shifts the demand curve and has positive long-run effects.

Jimmy Wales and his staff at Wikipedia have fundamentally changed the information environment of the world. Future research may examine other effects of Wikipedia articles on capital markets, firm activities, and general economic activity. Wikipedia continues to provide detailed access to page histories, traffic, and other useful data for research applications. As Wikipedia continues to grow, its influence will expand as well.

CHAPTER TWO

GENDER AND EARNINGS CONFERENCE CALLS

"Forget the board room. Women's voices are barely even present on conference calls."

—Marnaz and Greenfield (2018, para. 1)

2.1 Introduction

Despite the substantial progress women have made in the labor market, gender discrimination is not eliminated (Bertrand, Chugh, and Mullainathan, 2005; Goldin, 2014; Bertrand, 2018). Issues such as harassment, slow promotion, and unequal pay are widely reported on Wall Street (Boorstin, 2018). However, unlike entry-level jobs, studies on jobs near the glass ceiling are difficult to conduct in lab and field experiments (Bertrand and Mullainathan, 2004; Bertrand and Duflo, 2017). To provide evidence of gender discrimination for high-earning professionals, researchers have leveraged novel approaches and data sources regarding senior management (Matsa and Miller, 2011), surgeons (Sarsons, 2017), musicians (Goldin and Rouse, 2000), entrepreneurs (Hebert, 2020), mutual fund managers (Niessen-Ruenzi and Ruenzi, 2019), and financial advisers (Egan, Matvos, and Seru, 2018).

In this paper, I examine the gender discrimination issue in prestigious professions of the business world (CEOs and financial analysts) using a unique setting—earnings conference calls. Specifically, I investigate five questions: (1) Are female analysts more or less likely to appear on conference calls? (2) Conditional on participation, are female analysts treated equally compared

Portions of this chapter previously appeared as: Francis, B. B., Shohfi, T., & Xin, D. (2020). *Gender and earnings conference calls.* SSRN Electronic Journal. Retrieved June 25, 2020, from https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3473266 with their male counterparts? (3) Do female analysts and executives exhibit different communication patterns compared to male analysts and executives? (4) Are male participants more likely to discriminate against female participants and vice versa? (5) Does the market interpret female and male conference call participants' information equally?

Conference calls have emerged as a popular and influential disclosure channel for public firms since the passage of Regulation Fair Disclosure (Regulation FD) in 2000. One unique feature of conference calls is that, following management's presentation session, managers will answer questions from the public, typically sell-side analysts, in a question-and-answer (Q&A) session. Prior studies have established that earnings conference calls convey price-related information beyond press releases (Bowen, Davis, and Matsumoto, 2002; Bushee, Matsumoto, Miller, 2004; Kimbrough, 2005) and that the Q&A session is more informative than the presentation portion (Matsumoto, Pronk, and Roelofsen, 2011). Various dimensions of conference calls examined in the literature include analyst participation (Mayew, 2008), linguistic characteristics (Allee and DeAngelis, 2015; Bochkay, Chava, and Hales, 2020; Brochet, Loumioti, and Serafeim, 2015; Bushee, Gow, and Taylor, 2018; Davis, Ge, Matsumoto, and Zhang, 2015; Mayew and Venkatachalam, 2012), and information transfer (Brochet, Kolev, and Lerman, 2018).

To the best of my knowledge, this paper is the first to investigate gender discrimination issues through an earnings conference call lens. Conference calls have several features that provide us with a unique setting to study gender discrimination. First, on conference calls two parties—analysts and executives—participate together, which makes conference calls different from other disclosure venues in which only one party is involved at a time. As such, I can directly observe the interaction between analysts and management with various gender combinations. Second, during the Q&A session, analysts and managers interact in real-time without rehearsal or scripting.

Matsumoto et al. (2011) argue that the spontaneous nature of the Q&A part of a conference call leads to more information disclosure by managers because they prefer to withhold bad news in prepared statements. Therefore, I expect this more stressful part of a conference call to elicit genuine behavioral patterns and gender attitudes of analysts and management. Third, a speaker's voice makes gender more visible when investors listen to conference calls compared with when they read written communication (e.g., regulatory filings by public companies, analyst recommendations, etc.), making gender attitudes more salient.

I analyze a large sample of more than 60,000 conference call transcripts collected from Capital IQ for the period 2008 to 2016, from which, using multiple algorithms based on first names, I determine the gender of participating analysts and executives. My analyses proceed in several stages. First, I examine whether there exist gender differences in the probability of analysts' conference call participation. I follow Mayew (2008) using the Institutional Brokers Estimate System (I/B/E/S) to identify a sell-side analyst population who are interested in participating in conference calls. I find that female analysts are 2% less likely to participate in conference calls, representing a 5% relative reduction from the unconditional mean participation probability of 41%.

Next, I conduct a conditional analysis on conference call participants' behavior. Given the lower participation probability of female analysts in a male-dominated profession, I examine whether they are treated and/or do they behave differently from their male counterparts. By analyzing analyst participation prioritization, I find that although female analysts are equally likely to ask the first question on conference calls, they have fewer opportunities to ask follow-up questions and their statements are shorter. Prior studies find that connections with firm management are valuable capital for sell-side analysts (Mayew, 2008; Green, Jame, Markov, and Subasi, 2014; Fang and Huang, 2017). Thus, if female analysts encounter discrimination, they could behave less aggressively to maintain a favorable relationship with management. Consistent with this notion, I find that female analysts exert less pressure on firm management. Specifically, female analysts have more positive tone and less uncertain tone, discuss less numerical content, hesitate less, and have fewer back-and-forth Q&As with firm management.

I next turn my gender analysis toward executives. In Q&A sessions, analysts initiate questions, which executives answer. Therefore, analysts set the atmosphere of Q&A sessions, which is consistent with the argument that analyst tone, instead of executive tone, moves the market (Chen, Nagar, and Schoenfeld, 2018). This suggests that if analysts discriminate against female executives, they would likely be under stricter scrutiny. Consistent with this idea, I find female executives experience more back-and-forth Q&As. However, female executives exhibit superior ability when handling analysts' questions in that they hesitate less and are less uncertain in tone. Kumar (2010) argues, and provide supporting evidence, that gender discrimination raises the evaluation standard of females in male-dominated professions and only females with superior ability self-select into these professions. My finding is in line with this self-selection hypothesis.

Given that the analysis of conference call participation, tone, speech hesitations, and back-and-forth Q&As provide indirect evidence of gender discrimination, I provide direct evidence by examining interruptions during analyst-executive interactions. Besides "taste-based" and "statistical" discriminations, which are explicit, discrimination can be implicit (Bertrand et al., 2005). This type of unconscious discrimination is more fundamental and difficult to conceal. As such, previous studies have used interruptions to capture discrimination in casual conversation (Zimmerman and West, 1975) and among Supreme Court justices (Jacobi and Schweers, 2017). Using interruptions as a measure of discrimination, I observe an in-group favoritism—female analysts receive fewer interruptions from female executives compared with male executives. Male executives treat male and female analysts equally. However, male analysts and executives are more likely to interrupt female executives. I also find evidence consistent with the “power jockeying” phenomenon—that male executives interrupt their female colleagues, particularly those in superior roles, more than they interrupt female analysts.

Last, I examine the market reaction to female and male analysts’ conference call participation. If market participants misinterpret analysts’ conference call participation due to discrimination or stereotyping, female analysts’ participation may lead to a weaker market reaction. I control for both analyst and executive tone to separate the informational influence of gender differences across roles. I find that there is a negative relationship between the percentage of female analysts participating on the call and the absolute short-term market reaction. Additionally, female analyst tone is associated with a weaker directional market reaction than male analyst tone. This finding is consistent with a gender-stereotyping hypothesis and contrasts with the self-selection hypothesis that the market values female analysts’ superior ability more (Kumar, 2010). It also suggests that gender stereotyping occurs more readily when gender characteristics on conference calls (e.g., voice) are more salient than in other analyst outputs such as forecasts and recommendations, for example.

My paper contributes to the literature in several aspects. First, I add to the gender discrimination literature on high-earning professionals. Extant research predominantly focuses on indirect evidence of gender discrimination. For example, one stream of literature argues that gender discrimination raises the evaluation standard of females in male-dominated professions, thus, females competing successfully in these professions must possess superior abilities (e.g., Kumar, 2010; Hengel, 2020). Other works reveal subtle but direct evidence of gender discrimination in unique settings including comments by economists on Internet forums (Wu, 2018), physicians' referrals to surgeons (Sarsons, 2017), and punishments for financial advisors (Egan et al., 2018). I leverage earnings conference calls, a real-time communication environment, to investigate participants' gender attitudes. My results provide both indirect evidence—female analysts have few participation opportunities and speak less—and direct evidence of discrimination—both female analysts and executives are interrupted more frequently by their male counterparts. Moreover, my conference call setting allows us to study two parties—analysts and executives—at the same time.

Second, I extend earnings conference call literature by introducing gender effects. While prior research on earnings conference calls focuses on incremental information and compares the informational roles of various participants (Matsumoto et al., 2011; Chen et al., 2018) my paper focuses on gender differences in participation, communication, and discrimination. Given that prior studies using private data document gender differences in the upper echelons and other high-profile professions within the financial industry (Kumar, 2010; Huang and Kisgen, 2013; Jeong and Harrison, 2017), it is more surprising to see a gender effect persist in the scrutinized public forum of earnings conference calls.

The rest of the paper proceeds as follows. I review the literature and develop hypotheses in Section 2. Sections 3 describes the data. In Section 4, I present the empirical analysis. Sections 5 concludes.

2.2 Literature review and hypothesis development

2.2.1 Gender discrimination in business

Early studies on gender discrimination provide only indirect evidence by controlling for gender differences in observed characteristics and considering unexplained gender differences, such as the gender pay gap, as discrimination. For example, studies examining gender wage gaps usually control for education, experience, and other variables that are reflective of productivity (Guryan and Charles, 2013). However, this approach will overestimate discrimination if men have higher unobserved productivity or underestimate discrimination if women have higher unobserved productivity (Blau and Kahn, 2017). The unexplained labor market gap can also underestimate discrimination if it in turn affects control variables (Blau and Kahn, 2017). However, even if after controlling for productivity-related characteristics there is no evidence of a gender gap, it does not rule out discrimination through gender segregation and unequal promotion (Bertrand and Hallock, 2001). For management-level positions, human capital, career motivation, and other individual unobservable characteristics are more homogeneous compared with entry-level jobs (Blau and Khan, 2017). Therefore, unexplained gender gaps observed in compensation can be interpreted as evidence of discrimination (Bertrand and Hallock, 2001). However, some vestige of omitted-variable concerns remains.

Recent studies seeking to provide evidence of discrimination have turned to other labor market outcomes using novel approaches, which have led to two streams of literature. One stream of literature examines negative outcomes. For example, Egan et al. (2018) identify a "gender punishment gap". They find that compared to their male counterparts, female financial advisers are more likely to be fired despite engaging in less costly misconduct and have lower likelihood of repeat offenses. In the same vein, Bloomfield et al. (2020) conduct an experiment and find that, in contrast to male analysts, investment professionals evaluate female analysts as less promotable when they lack persistence in pitching a stock, consistent with gender stereotyping. The other stream contends that gender discrimination leads to a phenomenon in which highly qualified women self-select into male-dominated professions (Kumar, 2010; Blau and Khan, 2017). For instance, Kumar (2010) argues that in the male-dominated financial services industry, female analysts, along with having above average abilities (relative to their male counterparts) are not representative of average women who are risk averse. Consistent with the self-selection hypothesis, he finds that female analysts issue bolder and more accurate forecasts, and are more likely to cover large stocks with higher institutional ownership even in early stages of their careers. He further shows that the market reacts, both in the short and long term, more strongly to female analysts' forecast revisions even when they attract less media coverage. In addition, he documents that female analysts are more likely to be promoted to prestigious brokerage firms and less likely to receive a demotion to less prestigious ones.

Discrimination can be explicit or implicit (Bertrand et al., 2005). Implicit discrimination is unconscious and difficult to hide. For example, Sarsons (2017) in investigating physicians' referrals to surgeons finds physicians' evaluation of a surgeon's ability declines more after a patient death for female surgeons compared with male surgeons, controlling for surgeon specialty. Moreover, physicians give evaluation that is more positive to male surgeons after a successful surgery outcome. Wu (2018) examines anonymous discussion about female and male economists on the Economics Job Market Rumors Internet forum and finds pervasive gender discrimination. She documents that users discuss non-academic information more for female economists and academic information more for male economists. In sum, novel datasets and settings are useful tools to identify gender discrimination within high-paying professions.

2.2.2 Earnings conference calls

Earnings conference calls are one of the most important venues for firms to communicate with investors (Matsumoto et al., 2011). The majority of conference calls follow immediately after a quarterly earnings release. A conference call usually starts with a presentation session in which executives discuss current operations and make forward-looking statements. After management's presentation, analysts and investors can communicate with firm management in a Q&A session. Prior studies show that conference calls provide value-relevant information to capital markets (Frankel et al., 1999; Bushee et al., 2004; Kimbrough, 2005). Matsumoto et al. (2011) find that both presentation and Q&A sessions have incremental information over press releases, with the Q&A discussions being particularly informative. They further show that the informativeness of a Q&A session is associated with the number of analysts following the firm and that analysts' active role in conference calls contributes significantly to their informativeness.

From the perspective of analysts, Bowen et al. (2002) show that conference calls increase analysts' forecast accuracy and decrease forecast dispersion. However, analysts' participation is not random, and hosting firms have discretion to determine who ask questions on conference calls (Brown et al., 2019). Mayew (2008) shows that during conference calls, firms discriminate by providing analysts who issue favorable stock recommendations with more opportunities to ask questions. Further, Mayew et al., (2013) find that analysts who participate in conference calls by asking questions issue more accurate and timelier earnings forecasts than non-participating analysts, suggesting participating analysts may possess superior information.

Another stream of literature examines soft information embedded in conference calls. For example, Larcker and Zakolyukina (2012) classify CEO and CFO narratives from conference call transcripts into "deceptive" and "trustful" parts based on psychological and linguistic word lists, and they find that the deception measure can predict subsequent financial restatements. Allee and DeAngelis (2015) document that tone dispersion, which is the degree to which tone is spread evenly in a narrative, is associated with firm performance, managers' financial reporting choices, and managers' incentive to influence the perception of the firm. Mayew and Venkatachalam (2012) show that managers' affective states in conference calls can predict future firm performance and the effect is more prominent in the Q&A session when managers are under great scrutiny by analysts. Davis et al. (2015) show that there exists a manager-specific component in the tone of earnings conference calls that current performance, future performance, or strategic incentives cannot explain. They further add that demographic characteristics including career experience and charitable organization involvement are the driving forces behind the relationship with the manager-specific factor. Additionally, the authors also note that the tone of executives on earnings conference calls is associated with their level of optimism. However, with regard to gender, they document only weak evidence that female executives use less favorable language.

2.2.3 Hypothesis development

Gender discrimination is ubiquitous among male-dominated industries. Equity analysts are a male-dominated occupation. Given extensive historical gender discrimination and an “old boys’ network”, establishing connections with firm management provides fewer rewards, perhaps even punishment, to female analysts (Fang and Huang, 2017) and may therefore decrease their incentives to build connections. Moreover, because managers have discretion over analysts’ conference call participation (Mayew, 2008), connections are a key determinant of their participation. Along the same line, sell-side analysts avoid asking difficult questions on conference calls to maintain a good relationship with management and leave tough questions to private communication instead (Brown et al., 2015). Given gender stereotyping within the analyst industry, I propose:

H1: Female analysts are less likely than male analysts to participate in earnings conference calls.

Firm management has the discretion to determine which analysts they will prioritize on conference calls. Firms are very sensitive with regard to information disclosure on conference calls given that both solid and soft information are disseminated to the public (Suslava, 2017; Zhou, 2018).[23] To avoid disclosing unfavorable information, management regularly chooses not to answer certain analysts' questions (Gow et al., 2019; Hollander et al., 2010) or disproportionately prioritizes optimistic analysts (Cohen et al., 2020; Mayew, 2008). According to firms' Investor Relations Officers (IROs), analysts who have a long coverage history with the firm usually receive priority in the question queue (Brown et al., 2019).

Previous studies consider three dimensions of analyst participation prioritization: asking the first question, asking multiple rounds of questions, and having long conversation with firm management (e.g., Call et al., 2018). Managing conference calls is the primary task of IROs and prioritizing analysts in the Q&A queue selectively is an important component (Brown et al., 2019). Asking the first question in a conference call sends a strong signal of a firm's special attention and connection with the analyst (Call et al., 2018; Cen et al., 2020). Similarly, given the time constraint, asking a second round of questions also implies a friendly relationship between analysts and the management. Note also that long conversations signify that firms are willing to provide analysts with more visibility. Because of the lower benefits of connections to management for female analysts (Fang and Huang, 2017), and potential in-group bias (Jannati et al., 2020), female analysts may have less of an opportunity to build these connections. If analyst gender is a barrier to building these connections, I expect to observe less favorable treatment of female analysts by management

[23] For example, Elon Musk, the CEO of Tesla, Inc., said questions from analysts were asking "boring, bonehead questions" in its 2018 Q1 earnings conference call on May 2nd, 2018. Tesla stock price plunged 5.6% on the following day.

on earnings conference calls in term of conference call participation prioritization. Therefore, I hypothesize:

H2a: Females analysts are less likely to ask the first question on earnings conference calls.

H2b: Females analysts are less likely to have follow-up interactions on earnings conference calls.

H2c: Female analysts' interactions with management on conference calls are shorter than male analysts' interactions with management.

The manner of communication between analysts and firm management plays a crucial role in conference calls. Although analysts are under the pressure of their buy-side to ask acute questions, it should not happen at the expense of the relationship with firm management (Brown et al., 2015). This is the case because the value of a firm's access to analysts benefits from connections with management, both from the perspective of research informativeness (Green et al., 2014) and compensation (Groysberg et al, 2011). Under Regulation FD, although firms must open conference calls to all interested members of the general public (Bushee et al., 2004), the complementing role of public information to private information (i.e., mosaic theory) on earnings conference calls remains essential for analysts (Mayew, 2008). Connections of analysts are also associated with their forecast accuracy and career advancement. Sell-side analysts have strong incentives to curry favor from their buy-side clients (Groysberg et al., 2011). A considerable amount of compensation paid by buy-side clients to sell-side firms is for corporate access (Brown et al., 2019).

To retain connections with management, analysts must not interrogate executives and/or cast them in an unfavorable light. As Soltes (2014) points out: "Assuming you want management to continue speaking with you, you have to avoid making the C-suite lose face on the call…if you have difficult questions and you want management to speak openly, you have to do that off-line." (p. 265). Women value social connections and relationships more in communication compared to their male counterparts (Leaper, 1991). Conversations between women are more fluent and affirmative compared with mixed-gender and male-only pairs (Hirschman, 1994). To the extent that female analysts are at a disadvantage in participating in conference calls, they may choose to initiate a relatively relaxed conversation with management in accordance with the "theater" nature of conference calls (Brown et al., 2019). Consequently, female analysts may discuss less numerical content that is "solid" and give firm management more freedom to provide "soft" statements (Zhou, 2018). Because asking harsh questions can be counterproductive to building a good relationship with management, mild questions may lead to less cognitive dissonance (Festinger, 1957; Chang, Solomon, and Westerfield, 2016), which in turn leads to less uncertain sentiment and less hesitation (Lay and Paivio, 1969). Along the same lines, analysts' pursuit of harmony with firm management may decrease the toughness of their questions and thus lower the possibility of tug-of-war (i.e., fewer back-and-forth comments). Therefore, I hypothesize:

H3a: The tone of female analysts' interaction with management on conference calls is less negative than the tone of male analysts' interaction with management.

H3b: Female analysts discuss less numerical information with firm management.

H3c: Female analysts exhibit less uncertainty in their narratives.

H3d: Female analysts exhibit less frequency of speech hesitation in their interactions with firm management.

H3e: Female analysts have fewer back-and-forth comments with firm management.

In conference call Q&A sessions, firm management generally responds to analysts' questions in a passive manner. Female executives self-select into the pursuit of C-suite positions, which suggests that they possess superior ability than an average C-suite executive (Kumar, 2010). This suggests that female executives are, therefore, more capable at handling analyst inquiries and as such, exhibit less uncertain sentiment and fewer hesitations. Moreover, the possibility of male analysts' discriminatory bias against female executives can also lead to more difficult questions asked and thus more back-and-forth comments (Jannati et al., 2020). Therefore, I hypothesize:

H4a: Female executives exhibit less uncertainty in their narratives.

H4b: Female executives exhibit less frequency of speech hesitations.

H4c: Female executives have more back-and-forth comments with analysts.

Men and women have different views on the purpose of conversation. Women seek social connections and relationships in communication while men prefer to exhibit power (Leaper, 1991). Consequently, women are more expressive and polite in conversation while men are more aggressive (Basow and Rubenfeld, 2003). In line with this, prior studies have shown that men are much more likely to interrupt women than vice versa. Specifically, men generally desire to demonstrate power and control the topics of conversations by interrupting women (Zimmerman and West, 1975). Jacobi and Schweers (2017) examine oral arguments at the U.S. Supreme Court and show that male justices and male advocates disproportionately interrupt female justices. Therefore, I expect women, either female analysts or female executives, to receive more interruptions. This leads to the following hypotheses:

H5a: Female analysts are more likely to be interrupted.

H5b: Female executives are more likely to be interrupted.

Investors respond to a wide range of analyst characteristics including reputation (Gleason and Lee, 2003; Stickel, 1992), connections with firm management (Fang and Huang, 2017), underwriting relationships (Lin and McNichols, 1998), brokerage affiliation (Clement and Tse, 2003), gender (Kumar, 2010), name favorability (Jung et al., 2019), and political preferences (Jiang et al., 2016), among others. Prior studies find that subjective feelings influence investment decisions and that investors seek consistency in how easily perceived characteristics, such as gender, affect their decisions (Alter and Oppenheimer, 2006). Given that men dominate sell-side analysts, gender stereotyping could lead to lower evaluation of female analysts' participation on earnings conference calls. Therefore, I hypothesize:

H6: Market reaction to female analyst participation in conference calls is weaker.

2.3 Data

2.3.1 Sample selection

I collect earnings conference call transcripts of Standard and Poor's 500 (S&P 500) constituent firms from Capital IQ over the 2008 to 2016 time-period. In addition, I collect transcripts of over 2,700 random firms that are not included in S&P 500 index but appear in the Center for Research in Security Prices (CRSP) database. My initial sample includes 81,677 earnings conference call transcripts for 3,346 unique publicly traded companies. I remove firms without data in I/B/E/S or CRSP. For each transcript, I record the call date, time stamp, names of firm executives, names of analysts participating in the question-and-answer (Q&A) session, and analyst affiliation.

To determine analyst gender, I extract the first name from each analyst's full name and apply multiple algorithms sequentially—R package *gender*, Python package *gender-guesser*, and *gender-api.com*. I use these tools, publicly available government databases, and social network data to construct first name-gender pairs. Because a probability is given for each gender guess tool (i.e., Prob(male)+Prob(female)=1), I assign the gender with higher probability to each first name.[24] No gender is assigned to androgynous first names (i.e., Prob(male)=Prob(female)=50%). Appendix H describes the gender determination process. For executives who appear in conference calls, I match names with Execucomp records that have gender and other information. Finally, I complement missing analyst and executive gender data by manually searching a variety of sources including S&P Capital IQ, LinkedIn, Bloomberg, and Seeking Alpha. I successfully identify the gender of 98.5% (99.4%) analyst (executive) conference call participations.[25]

In order to investigate the dynamics of analyst-management interactions on conference calls, I construct a call-analyst level sample. I proceed in several steps. First, I parse all conference call transcripts into question-answer blocks. In conference call transcripts, each narrative starts with the name, title, and affiliation of the speaker in separate lines. Before an analyst asks questions, the conference call operator introduces the analyst. Thus, the appearance of the operator can serve as a delimiter for conversation blocks. Specifically, each conversation block starts with the analyst name and ends with the operator's introduction of the next analyst.[26] In other words, a block is a group of continuous back-and-forth comments between the focal analyst and one or more executives. Hereafter, I designate each block an interaction.

Second, I scan each conference call transcript to identify all interactions. Because analysts may have back-and-forth statements or questions with one or more executives in each block, I separately record each analyst's narrative and narratives of different executives in each interaction block and then collapse multiple observations related to one analyst (or executive) to one

[24] *Gender-guesser* does not provide a probability of gender but rather gives five possible results: male, female, mostly male, mostly female, and androgynous. I assign "male" ("female") to a first name if *gender-guesser* gives "male" or "mostly male" ("female" or "mostly female").
[25] Analysts with unidentifiable gender are recorded in transcripts as "Unidentified Analysts", "Unknown Speaker" or "Unknown Analyst" or with a name abbreviation. Unidentifiable company participants are recorded as "Unidentified Company Representative", "Unknown Executive", "Attendees", "Unknown Speaker", etc.
[26] I remove all names, titles, and affiliations to keep narratives only for my textual analysis applications.

observation. For analysts who ask more than one round of questions (i.e., analysts involved in two or more non-continuous interactions in one conference call), observations are aggregated to generate one observation for each analyst in each conference call. My final conference call sample contains 442,211 call-analyst level observations representing 62,644 conference calls and 2,836 unique firms. Appendix E contains a summary of the sample selection process.

2.3.2 Variables

My key variables are indicator variables *FemaleAna*, which is equal to 1 if the analyst is female and a continuous variable in the range of [0,1], *FemaleExe*, which is the percentage of female executives' narratives related to the corresponding analyst based on number of words spoken.[27] Analyst questions that are answered exclusively by male (female) executives have *FemaleExe* equal to 0 (1).

The extant literature suggests that other analyst's characteristics could vary systematically with gender. To the extent that this is the case, the relationship between analyst gender and earnings conference call or market outcomes, is likely biased. I follow Mayew (2008) and include variables, related to analyst characteristics. *AllStar* is an indicator variable for Institutional Investor All-American analysts in a given year. *BrokerSize* is the number of analysts employed by the brokerage firm of an analyst in the prior calendar year of the conference call. *GenExp* is the number of years between the conference call date and the date on which the analyst issues his or her first forecast on I/B/E/S. *FirmExp* is the number of years between the conference call date and the date on which the analyst issues his or her first forecast for the firm on I/B/E/S. *IndCover* is the number of Fama-French 48 industries covered by an analyst in the prior calendar year of the conference call.

[27] For example, suppose an analyst asks questions and two executives, one man and one woman, answer the questions. If the male executive's narrative consists of 40 words and the female executive's narrative consists of 60 words, *FemaleExe* will be equal to 0.6.

CompCover is the number of unique firms covered by an analyst in the prior calendar year of the conference call. *CCuser* is the number of other conference calls on which the analyst participates in the same calendar quarter as the focal conference call. *Rec* is the analyst's latest stock recommendation of the firm holding the conference call on an integer range from -2 to +2 representing strong sell to strong buy. *RecHorizon* is the number of days from the issue date of the latest stock recommendation to the conference call date. To measure analyst forecast performance, I follow Clement (1999) and construct a forecast accuracy measure, which is equal to the negative value of the absolute forecast error demeaned by the same quarter-firm forecast average:

$$ForeAcc_{ijt} = -\frac{|ForeError_{ijt}| - \overline{|ForeError_{jt}|}}{\overline{|ForeError_{ijt}|}} \tag{2.1}$$

where $|ForeError_{ijt}|$ is the absolute forecast error (the absolute difference between the last earnings per share (EPS) forecast and actual EPS) for analyst *i* of firm *j* in quarter *t*, and $\overline{|ForeError_{jt}|}$ is the mean $|ForeError_{ijt}|$ (average $|ForeError_{ijt}|$ across all analysts covering firm *j* in quarter *t*). A positive (negative) value of *ForeAcc* indicates that an analyst's forecast is more (less) accurate than other analyst forecasts of the same firm in the same quarter. This measure of forecast accuracy is relative to other analysts and eliminates heteroscedasticity across firm-quarters (Ke and Yu, 2006).

2.3.3 Analyst gender distribution

I first examine the gender distribution for analysts appearing on earnings conference calls in my sample. Table 2.1 reports the call-analyst level analyst gender distribution by year (Panel A), Global Industry Classification Standard (GICS) sector (Panel B), and brokerage affiliation (Panel C). Percentage of participation observations represented by female analysts (*%FemalePart*) and percentage of unique female analysts (*%FemaleUnique*) are shown separately. Corresponding percentage of female forecasts (*%FemaleFollowIBES*) and percentage of unique female analysts (*%FemaleUniqueIBES*) in the I/B/E/S sample are also reported. Panel A shows that although there is a slight increase over time in the percentage of unique participating female analysts, there is a steady decline in female analysts participation from 12.15% to 10.20%, indicating that over time female analysts participate less frequently on earnings conference calls than their male counterparts. The percentage of female analyst following in I/B/E/S also exhibits a similar decline. Panel B shows gender distribution across 11 GICS sectors. Female analysts are more concentrated in Consumer Staples and Consumer Discretionary, followed by Health Care. This evidence is consistent with that of Kumar (2010) who shows that female analysts are more heavily represented in these sectors. In Panel C, I follow Green et al. (2009) and rank brokerage firms in the I/B/E/S database based on the number of affiliated analysts in each year separating Top 10 and other brokerages. The proportion of female analysts in large brokerage firms is higher than that in other brokerage firms in both samples. Green et al. (2009) suggest that the relatively high representation of female analysts in large brokerages is because of emphasis on employee diversity and better working conditions, which are attractive to women. The proportion of female participation in Panel C is consistently lower than that of unique female analysts, indicating a lower participation level across both brokerage-ranking groups.

Table 2.1: Gender distribution

Panel A. Conference call gender distribution by year

year	%FemalePart	%FemaleUnique	%FemaleFollowIBES	%FemaleUniqueIBES
2008	12.15%	11.89%	11.53%	13.02%
2009	11.82%	11.19%	10.72%	12.33%
2010	11.63%	11.22%	10.08%	11.71%
2011	11.10%	10.59%	9.79%	11.12%
2012	10.50%	11.22%	9.43%	10.94%
2013	10.23%	11.17%	9.43%	10.85%
2014	10.13%	11.90%	9.50%	11.07%
2015	10.36%	11.79%	9.70%	11.17%
2016	10.20%	12.48%	9.60%	11.53%

Panel B. Conference call gender distribution by sector

Sector	%FemalePart	%FemaleUnique	%FemaleForecastIBES	%FemaleUniqueIBES
Consumer Discretionary	18.47%	16.25%	17.07%	16.72%
Consumer Staples	24.17%	19.20%	23.93%	20.43%
Energy	6.90%	8.68%	7.04%	8.09%
Financials	8.50%	9.02%	7.45%	11.29%
Health Care	12.35%	14.80%	11.22%	15.98%
Industrials	7.96%	8.31%	7.22%	8.61%
Information Technology	6.72%	8.37%	6.45%	8.14%
Materials	7.20%	9.21%	5.09%	8.85%
Real Estate	9.74%	10.86%	4.39%	8.10%
Telecommunication Services	9.65%	7.29%	8.03%	5.17%
Utilities	8.05%	13.35%	11.69%	16.94%

Panel C. Conference call analyst gender distribution by brokerage firms

	%FemalePart		%FemaleUnique		%FemaleForecastIBES		%FemaleUniqueIBES	
	Top 10	Others	Top 10	Others	Top 10	Others	Top 10	Others
2008	16.17%	10.48%	17.96%	10.47%	14.13%	10.23%	16.87%	11.40%
2009	14.30%	10.26%	14.50%	10.50%	12.25%	10.02%	14.85%	11.29%
2010	13.45%	10.41%	14.47%	9.95%	11.77%	9.45%	14.38%	10.77%
2011	13.30%	9.93%	14.11%	9.74%	12.47%	8.89%	14.71%	9.86%
2012	11.29%	9.75%	14.85%	10.43%	12.39%	8.44%	13.79%	9.95%
2013	10.88%	9.35%	14.44%	10.62%	10.99%	8.83%	12.69%	10.16%
2014	10.30%	9.56%	13.19%	11.58%	9.52%	9.50%	11.82%	10.72%
2015	11.46%	9.65%	14.64%	11.51%	9.82%	9.65%	12.68%	10.45%
2016	13.01%	8.89%	16.41%	11.37%	10.88%	9.06%	14.47%	10.39%

2.3.4 Descriptive statistics

Table 2.2 presents descriptive statistics for conference call variables (Panel A), and firm variables (Panel B). Regarding conference call characteristics, the mean number of words spoken in the Q&A session is 3,835 (*WordsQNA*). Panel A shows that on average, 7.6 non-continuous interactions (*FollowupCall*) are made by 7.2 analysts (*AnaCount*) with 3.4 executives (*ExeCount*). The number (percentage) of female analysts per call is 0.76 (9.8%) (*FemaleAnaCount* and *FemaleAnaPct*). The average number of participating female executives is 0.44 (*FemaleExeCount*), representing 12.8% of all executives (*FemaleExePct*). Turning to CEOs and CFOs, I see that 59.6% (58.1%) of conference calls have the firm's CEO (CFO) participating (*CEOPart* and *CFOPart*) and 52.2% have both the CEO and CFO present (*CEOCFOPart*).[28] The average for the number of CEO and CFO participating in a conference call (*CEOCFOCount*) is 1.2, while the number (percentage) of female CEOs or CFOs is just 0.075 (4.1%) (*FemaleCEOCFOCount* and *FemaleCEOCFOPct*). It is important to note that the percentage of female CEOs or CFOs is much lower than the percentage of female executives, which is consistent with the lower participation rate of women in the labor force and lower representation in corporate C-suites. On the other hand, it could also be due to the relatively high proportion of women among investor relations officers (Brown et al., 2019).[29]

Panel B, which contains firm level results, shows that an average firm has market capitalization of approximately $6.9 billion (*MktCap*), a leverage ratio of 2.6 (*Leverage*), market-to-book ratio of 2.9 (*MB*), and return on assets of 0.01 (*ROA*). It also shows that, 21.7% of firms

[28] Because Capital IQ gives up-to-date executive titles but not the title as of the conference call date, I match executive names with Execucomp. Specifically, I follow Jiang, Petroni, and Wang (2010) and use Execucomp variables CEOANN, CFOANN, and TITLEANN to determine CEOs and CFOs. *CEOPart* are *CFOPart* are lower than the actual participation rates because my method does not assign CEO or CFO flags to interim CEOs or CFOs.
[29] Investor relations officers (IROs) are listed as executives at the beginning of conference call transcripts.

are S&P 500 constituents, with institutional ownership accounting for 66.6% of total shares (*InstOwn*), and that on average, 10.7 analysts in I/B/E/S (*AnaCover*) cover each firm,. The average standardized unexpected earnings (actual earnings minus consensus earnings scaled by quarter-end stock price) is approximately 0.035 (*SUE*). Mean consensus stock recommendation (on an integer range from -2 to +2 representing strong sell to strong buy) is 0.7 (*RecCon*). The stock run-up prior to conference call is -0.007 (*Runup*). A mean (median) of 42.7 (14) other conference calls within the same 3-digit SIC code as the focal conference call are held in the same calendar quarter (*CallCluster*). Appendix F contains extended variable definitions.

Table 2.2: Descriptive statistics

Panel A. Conference call variables

	mean	Q1	median	Q3
WordsQNA	3835.324	2534.000	3756.000	4988.000
FollowupCall	7.602	5.000	7.000	10.000
AnaCount	7.167	4.000	7.000	9.000
FemaleAnaCount	0.763	0.000	0.000	1.000
FemaleAnaPct	0.098	0.000	0.000	0.167
ExeCount	3.411	3.000	3.000	4.000
FemaleExeCount	0.439	0.000	0.000	1.000
FemaleExePct	0.128	0.000	0.000	0.250
CEOPart	0.596	0.000	1.000	1.000
CFOPart	0.581	0.000	1.000	1.000
CEOCFOPart	0.522	0.000	1.000	1.000
CEOCFOCount	1.188	0.000	2.000	2.000
FemaleCEOCFOCount	0.075	0.000	0.000	0.000
FemaleCEOCFOPct	0.041	0.000	0.000	0.000

Panel B. Firm variables

	mean	Q1	median	Q3
MktCap	6929.626	469.768	1415.135	4542.676
Leverage	2.601	1.242	1.562	2.357
MB	2.862	1.165	1.925	3.375
ROA	0.010	0.001	0.016	0.043
SP500	0.217	0.000	0.000	0.000
InstOwn	0.666	0.526	0.738	0.878
AnaCover	10.720	5.000	8.000	15.000
SUE	0.035	-0.042	0.042	0.219
RecCon	0.721	0.380	0.730	1.000
Runup	-0.007	-0.078	0.000	0.066
CallCluster	43.110	5.000	15.000	69.000

2.3.5 Univariate analysis

I next compare the mean of a series of analyst-call level variables between male and female analysts. Table 2.3 Panel A contains the results. Consistent with prior work (e.g., Bosquet et al, 2014; Kumar, 2010; Mayew, 2008), I find that female analysts are much more likely to be all-star analysts, are hired by large brokerage firms, have less general experience but similar firm-specific experience, cover fewer industries and companies, are more accurate in earnings forecasts, and issue less favorable stock recommendation with shorter horizons.

Table 2.3 Panel B reports gender comparisons for analysts' participation variables in which I use various variables to capture participation characteristics. Specifically, I use first questioner indicator (*First*), the number of non-continuous interactions between analyst and managers (*InterAna*), the number of words spoken by each analyst (*WordsAna*), and the average number of back-and-forth comments in an interaction (*RallyAna*). I expect the number of back-and-forth statements to reflect the intensity of an interaction with management. I find female analysts are less likely to ask the first question, are less likely to have follow-up interactions with executives, have shorter interaction length, and have fewer rounds of back-and-forth comments in each interaction.

In addition, I introduce two new characteristics of analyst-manager interactions: interruption and hesitation. In a conference call, when a manager (analyst) interrupts an analyst or manager, it indicates that managers (analysts) strongly disagree with an analyst's (manager's) comments and/or want to cut short the conversation. Importantly, it can also reflect how disrespectful the interrupter is toward the interruptee. To proxy for interruptions, I follow the lexical symbols used by Capital IQ. Capital IQ uses an ellipsis (…) at the end of a sentence to indicate that speakers have cut off each other. I construct a variable, *InterruptAna*, which is the total number of times an executive interrupts an analyst.[30] I measure hesitation by the appearance of two consecutive hyphens (--) to represent a self-correction or broken thought. *HesitAna* is the number of hesitations exhibited by the analyst. I provide examples of both interruptions and hesitations in Appendix G using excerpts from a conference call transcript.

[30] I do not find evidence of analysts interrupting each other in my sample.

Panel C contains analysts' textual characteristics comparisons. I measure sentiment with three Loughran and McDonald (2011) (LM) dictionaries: positive, negative, and uncertainty. I calculate analyst tone as:

$$ToneAna = \frac{Tone\ WordsAna\ Count}{WordsAna} \times 100\% \tag{2.2}$$

where *ToneAna* is *positiveAna*, *negativeAna*, or *uncertaintyAna*. Prior research has established that the LM dictionary is an effective measure of financial context sentiment. Given that LM designed their dictionary specifically for financial statements, and conference call transcripts are derived from verbal communication, I also use the Harvard General Inquirer (Harvard GI) dictionary to measure sentiment. To capture general sentiment, I construct a net tone measure, which is the difference between positive and negative tone (*net* and *netGI*). Positive net tone indicates that an interaction exhibits more positive sentiment than negative sentiment. In addition, I follow Zhou (2018) to examine the percentage of numbers or numeric phrases in interactions (*number*). I expect that numbers will contain more specific, value-relevant information than lexical content.

Panel C of Table 3 shows that female analysts are interrupted less by executives and exhibit fewer hesitations. Female analysts use more positive and negative words but do not exhibit a difference in net tone compared with male analysts. Using the Harvard GI dictionary, female analysts exhibit more positive sentiment but less negative sentiment, and therefore a more positive net sentiment. Less numeric content is included in female analysts' comments. This evidence is in line with hypothesis H3 that female analysts desire to establish more harmonious conversation with managers. In sum, the univariate analysis results are consistent with the hypothesis that female analysts' questions are less aggressive on conference calls.

In Table 3 Panel D, I report executive narrative variables for female and male executives. I construct a call-executive level sample including only executives who speak in the Q&A portion of earnings conference calls.[31] This sample contains number of words (*WordsExe*), number of interruptions received (*InteruptExe*), number of hesitations (*HesitExe*), and tone variables. I then make comparisons between female and male executives using these variables. In general, the number of words spoken by male executives is much larger than that of female executives (1037 vs. 593). Female executives receive fewer interruptions and exhibit fewer hesitations. For executive tone, female executives are less positive based on the LM dictionary but are more positive based on the Harvard GI dictionary. Moreover, I find that female executives are more affirmative compared to their male counterparts by exhibiting less uncertain sentiment. Taken together, the univariate comparisons in Table 3 are largely consistent with my hypotheses. To confirm these findings, I turn to multiple regression analyses.

[31] Investor Relations personnel and other firm participants who do not speak in the Q&A portion of the call are not included.

Table 2.3: Analyst gender differences in conference calls

Panel A. Analyst characteristics

	N	Male	Female	Difference	t-stat
AllStar	327757	0.150	0.219	-0.069	-33.827***
BrokerSize	327757	61.729	67.746	-6.017	-23.618***
GenExp	327757	14.372	13.458	0.913	16.861***
FirmExp	327757	5.041	5.266	-0.225	-8.249***
CompCover	327757	16.216	14.653	1.562	33.958***
IndCover	327757	3.056	2.706	0.350	30.079***
ForeAcc	327757	0.101	0.110	-0.009	-2.590**
CCUser	327757	6.562	6.079	0.483	18.332***
Rec	327757	0.504	0.438	0.066	16.661***
RecHorizon	327757	517.082	530.001	-12.919	-4.326***

Panel B. Analyst participation variables

	N	Male	Female	Difference	t-stat
First	442211	0.142	0.128	0.014	8.109***
Followup	442211	1.054	1.042	0.012	11.142***
WordsAna	442211	157.400	136.722	20.678	49.269***
RallyAna	442211	3.387	3.085	0.302	28.809***

Panel C. Analyst narrative variables

	N	Male	Female	Difference	t-stat
InterruptAna	442211	0.022	0.017	0.005	5.949***
HesitAna	442211	0.923	0.706	0.217	29.484***
positiveAna	442211	1.088	1.148	-0.060	-11.503***
negativeAna	442211	1.284	1.343	-0.059	-10.732***
netAna	442211	-0.195	-0.194	-0.001	-0.112
uncertainAna	442211	1.643	1.592	0.051	7.931***
positiveGIAna	442211	3.089	3.108	-0.019	-2.199*
negativeGIAna	442211	0.930	0.915	0.014	2.922**
netGIAna	442211	2.160	2.193	-0.033	-3.366***
numberAna	442211	0.751	0.630	0.121	24.021***

Panel D. Executive narrative variables

	N	Male	Female	Difference	t-stat
WordsExe	169432	1036.805	593.494	443.311	51.610***
InterruptExe	169432	0.058	0.047	0.011	3.888***
HesitExe	169432	5.957	2.941	3.016	35.691***
positiveExe	169432	1.377	1.170	0.207	25.564***
negativeExe	169432	0.828	0.915	-0.086	-12.875***
netExe	169432	0.549	0.263	0.286	26.750***
positiveGIExe	169432	3.234	3.409	-0.175	-12.679***
negativeGIExe	169432	0.941	0.825	0.116	18.818***
netGIExe	169432	2.296	2.596	-0.300	-19.302***
uncertainExe	169432	0.909	0.798	0.111	18.151***

2.4 Empirical findings

2.4.1 Conference call participation

I first examine the determinants of female analysts' earnings conference call participation. I follow Mayew (2008) to use I/B/E/S as the universe of sell-side analysts who are potentially interested in attending conference calls and construct a corresponding I/B/E/S sample. For the initial I/B/E/S sample, I require each firm-quarter-analyst observation to have both an existing earnings forecast and stock recommendation. An analyst is considered as actively following the firm if his/her earnings forecast is issued within one year of a given fiscal quarter end. Only the most recent forecasts prior to an earnings conference call are used.

To determine analyst gender within I/B/E/S, I need to obtain the first name of each analyst. However, I/B/E/S only provides each analyst's last name and first initial (item "ANALYST" in I/B/E/S). I exclude observations with missing brokerage ID (ESTIMID in I/B/E/S) or analyst name. In addition, I eliminate forecasts made by research teams.[32] To ensure the accuracy of analyst gender, I remove analysts for which two or more analysts (indicated by analyst code in I/B/E/S) share the same first initial and last name in the same brokerage (Bradley, Gokkaya, and Liu, 2017). Next, to determine the first name of analysts in I/B/E/S, I match analyst names within earnings call transcripts with analysts in I/B/E/S at the brokerage level. I check unmatched analysts manually with Capital IQ, LinkedIn, Bloomberg, Seeking Alpha, among others. Gender is then determined as described in Appendix D. I successfully identify the full name and gender for 5,687 analysts (99.8% of 5,722 unique sell-side analysts appearing in sample conference calls) in

[32] Analyst names for forecast issued by teams are recorded in I/B/E/S as a combination of two or more last names or a department name (e.g., "GERRY/ADKINS", "RESEARCH DEPT").

I/B/E/S. The final I/B/E/S sample includes over 671,550 analyst-firm-quarter observations for the 62,644 conference calls.[33]

I model conference call participation probability of analyst *i* following firm *j* in quarter *t*. I estimate the following pooled cross-sectional logit regression model:

$$
\begin{aligned}
Participate_{ijt} = \beta_0 &+ \beta_1 FemaleAna_{i,j,t} + \beta_2 Rec_{i,j,t} + \beta_3 AllStar_{i,j,t} \\
&+\beta_4 ForeAcc_{i,j,t} + \beta_5 GenExp_{i,j,t} + \beta_6 FirmExp_{i,j,t} \\
&+\beta_7 IndCov_{i,j,t} + \beta_8 CompCov_{i,j,t} + \beta_9 BrokerSize_{i,j,t} \\
&+\beta_{10} RecHorizon_{i,j,t} + \beta_{11} CCUser_{i,j,t} + \beta_{12} SUE_{i,j,t} \\
&+\beta_{13} Afternoon_{i,j,t} + \beta_{14} AnaCover_{i,j,t} \\
&+\beta_{15} WordsQNALog_{i,j,t} + \epsilon_{i,j,t}
\end{aligned}
\tag{2.3}
$$

The dependent variable, *Participate*, is an indicator variable equal to 1 if an analyst asks a question on an earnings conference call, zero otherwise. Year, industry (3-digit SIC), and brokerage fixed effects are included in all specifications. Standard errors are clustered at the firm level. In Model 1, I include *FemaleAna* and ten control variables capturing analyst characteristics. To examine how analyst gender affects the relationship between analyst reputation and conference call participation, I add an interaction term between *FemaleAna* and *AllStar* in the second specification. In the third model, I further include the firm level variable—*SUE*—and three conference call variables—*Afternoon*, *AnaCover*, and *WordsQNALog*. *Afternoon*, an indicator variable, which is equal to 1 if the conference call is initiated at or after 12 p.m.; this is controlled for because of potential diurnal influence on participation. The number of analysts following the company and

[33] Given the sizes of the I/B/E/S and conference call samples, the average analyst participation rate is about 65.8% (=442,211/671,550) which is higher than the mean of *Participate*, 41%, described below. The difference can be attributed to two reasons. First, some participants (e.g. buy-side analysts, sell-side/independent analysts not qualifying for I/B/E/S inclusion, media, etc.) in conference calls are not in I/B/E/S. For example, only 83% of participating analysts in my sample are sell-side. Second, I/B/E/S does not include all brokerage houses (e.g., Cowen & Co.). The mean of *Participate* in my I/B/E/S sample is close to the 38.1% documented in Mayew (2008).

the number of words spoken reflect how competitive it is for analysts to participate in a conference call. Participation opportunities should be fewer if more analysts follow the firm. *WordsQNALog* captures the time allocated to each Q&A session because analysts are likely to have more opportunities to participate in longer earnings conference calls (Mayew, 2008).

I present the results in Table 2.4. Focusing on my testing variable *FemaleAna*, I see that in all three columns the estimated coefficient is negative and significant at the 1% level. The marginal effect is also meaningful. The predicted probabilities of participation for female and male analysts are 37.0% and 39.0%. The 2.1% difference represents an approximate 5% disadvantage in participation probability for female analysts at the sample mean. The estimated coefficient of all-star analyst is positive and significant at the 1% level. The magnitude of the estimated coefficient of the interaction term is comparable to that of *AllStar*, suggesting that the benefit of being an all-star analyst for female analysts is almost double that of male analysts with regard to the likelihood of participation. In other words, female all-star analysts have a higher participation likelihood than male all-star analysts, thus supporting Hypothesis H1.

Examining other variables in Table 4, I see that across all three columns that the likelihood of conference call participation increases with stock recommendation favorableness (*Rec*), prior forecast accuracy (*ForeAcc*), firm-specific experience (*FirmExp*), frequency of conference call participation (*CCUser*), and length of the Q&A session (*WordsQNALog*). Interestingly, general analyst experience (*GenExp*) has a negative effect on participation likelihood. This finding is consistent with Mayew (2008) who suggests that analysts with more general experience may have lower demand for firm-specific information. I also see that analysts covering more companies (*CompCover*) or industries (*IndCover*) and issuing less timely coverage (*RecHorizon*) have lower participation probability.[34] Consistent with my expectations, there is a positive relationship between earnings surprise and participation, and a negative relationship between high analyst coverage (*AnaCover*) and participation. Notice also that analysts are less likely to participate in conference calls initiated in the afternoon (Column 3). One explanation is that it is because analysts are subject to diurnal influence. This is consistent with the notion that depletion of personal resources and circadian rhythms lead to less participation later in the day (Chen, Demers, Lev, 2018).

[34] Replacing *CompCover* with *SameDayCall*, which is the number of conference calls held by other firms covered by the analyst on the same day, yields similar results.

Table 2.4: Analyst gender and conference call participation

	(1)	(2)	(3)
VARIABLES	Participate	Participate	Participate
FemaleAna	-0.071***	-0.119***	-0.124***
	(0.024)	(0.027)	(0.028)
AllStar	0.286***	0.257***	0.277***
	(0.023)	(0.024)	(0.024)
FemaleAna×AllStar		0.230***	0.253***
		(0.057)	(0.060)
Rec	0.313***	0.310***	0.318***
	(0.008)	(0.008)	(0.008)
ForeAcc	0.166***	0.162***	0.175***
	(0.006)	(0.006)	(0.006)
GenExp	-0.019***	-0.018***	-0.020***
	(0.002)	(0.002)	(0.002)
FirmExp	0.027***	0.027***	0.034***
	(0.002)	(0.002)	(0.002)
IndCover	-0.008	-0.002	-0.012**
	(0.005)	(0.005)	(0.005)
CompCover	-0.019***	-0.022***	-0.021***
	(0.001)	(0.001)	(0.001)
BrokerSize	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)
RecHorizon	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
CCUser	0.015***	0.018***	0.017***
	(0.002)	(0.002)	(0.002)
SUE			0.011**
			(0.005)
Afternoon			-0.055***
			(0.019)
AnaCover			-0.047***
			(0.001)
WordsQNA			0.158***
			(0.004)
Constant	-2.872***	-2.924***	-3.046***
	(0.487)	(0.480)	(0.475)
Observations	668,551	668,551	668,551
Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Brokerage FE	Yes	Yes	Yes
Pseudo R-squared	0.077	0.072	0.093

Collectively, my participation analysis indicates that female analysts generally participate on earnings calls less frequently than male counterparts do, and that Institutional All-star recognition is more important for female analysts than male analysts with regard to conference call participation access. The finding is consistent with the notion that female analysts are in a relatively weaker position with respect to conference call participation.

2.4.2 Conference call prioritization

Next, I examine whether firm management prioritizes female analysts and provides them with more interaction opportunities on conference calls. I use three dependent variables, as my measure of prioritization: *First*, *FollowUp*, and *AbnLength*. Firm-level controls, year-quarter fixed effects, and firm fixed effects are included in all models.[35] Table 2.5 reports the results. Column 1 results where *First* is the dependent variable show that gender is not significant in explaining the likelihood of asking the first question on a conference call and thus hypothesis H2a is not supported. Column 2 reports Poisson model results for the number of interactions. I include initial question position (*Order*) because analysts who ask a question early in the queue are more likely to have a follow-up opportunity.[36] I find that *FemaleAna* is negatively associated with *InterAna*, thus supporting Hypothesis H2b. I further examine the interaction length between analysts and executives by counting the total number of words within each interaction. For analysts who have multiple interactions with executives, I aggregate word counts in all interactions to generate an analyst-level count. I then follow Call et al. (2018) to define abnormal interaction length as:

$$AbnLength = \frac{Words\ in\ all\ interactions\ with\ managers\ for\ the\ analyst}{\left(\frac{Words\ in\ Q\&A\ session}{Number\ of\ participating\ analysts}\right)} - 1 \qquad (2.4)$$

[35] As a robustness check, I replace firm fixed effects with call fixed effects and the results remain similar.

[36] Untabulated results show no gender difference when *Order* is the dependent variable.

AbnLength controls for systematic differences in interaction length due to Q&A session length and the number of participating analysts. A positive value of *AbnLength* indicates that an analyst's interaction length is above the average among all analysts. I regress *AbnLength* on the female analyst indicator and other control variables (Model 3). On average, female analysts' interactions with executives have 492 words and are 2.589% shorter than the within-conference call average compared with 537 words and 0.744% longer than average for male analysts. I add *FemaleExe* and its interaction with *FemaleAna* as additional controls because the presence of female executives could affect female analyst priority (Model 4). In both specifications 3 and 4, I find female analyst interactions are about 4.1% shorter.[37] Analyst interactions with only female managers are 8.1% shorter compared with those with only male management. The insignificant interaction term implies that a more female-dominated environment does not help improve female analysts' priority. In sum, these findings provide strong support for hypothesis H2c.[38]

[37] The average predicted mean *AbnLength* for male analysts are 0.8% and -3.2% for female analysts.

[38] Because CEO gender may affect the general gender attitude, I conduct a subsample analysis based on the CEO gender who is present in a conference all. No significant difference is found between these two samples.

Table 2.5: Participation prioritization of conference calls

	(1)	(2)	(3)	(4)
VARIABLES	First	FollowUp	AbnLength	AbnLength
FemaleAna	-0.005	-0.005***	-4.140***	-4.114***
	(0.031)	(0.002)	(0.458)	(0.480)
FemaleExe				-8.060***
				(0.981)
FemaleAna×FemaleExe				-0.297
				(1.756)
Rec	0.242***	0.006***	1.414***	1.413***
	(0.011)	(0.001)	(0.153)	(0.153)
AnaCountLog	-1.307***	-0.044***	9.275***	9.321***
	(0.009)	(0.003)	(0.262)	(0.262)
WordsQNALog	-0.045***	0.064***	-2.383***	-2.501***
	(0.007)	(0.003)	(0.145)	(0.149)
Order		-0.008***	-2.547***	-2.543***
		(0.000)	(0.061)	(0.061)
Constant	1.059***	-0.334***	12.937***	14.253***
	(0.048)	(0.019)	(1.066)	(1.104)
Observations	442,211	442,211	442,211	442,211
Firm controls	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Pseudo-R2	0.068	0.002		
Adjusted R2			0.028	0.029

2.4.3 Analysts' narrative characteristics

I next test hypothesis H3a by examining the influence of analysts' gender on, the tone of interactions between analyst and executives, uncertainty, quantitative information, the frequency of back-and-forth comments, and hesitations. I report the results in Table 2.6. As shown in the second column, where I use Harvard GI dictionaries to measure sentiment, female analysts convey sentiment that is more positive in their interactions with management. However, using the LM dictionaries in the first column, there is no evidence that gender differences exist in tone with regard to male and female analysts' interactions with management. This difference in results between the two dictionaries could be because LM specifically designed their dictionaries for financial statements. Note that the differences in results across the two dictionaries suggest that female analysts are more positive in nonfinancial context but are similar to their male counterparts in financial topics, thus partially supporting Hypothesis H3a. Columns 3 and 4 report results for uncertainty (Loughran and McDonald, 2011) and numerical content (Zhou, 2018), respectively. Consistent with hypotheses H3b and H3c, they show that female analysts' narratives are more certain and include less numerical content.

Turning to the relationship between speech hesitation and analyst gender, Column 5 shows that *FemaleAna* negatively predicts *HesitAna*. This result is consistent with the notion that female analysts ask fewer aggressive questions that may lead to fewer hesitations, thus providing support for Hypothesis H3e.[39] Finally, in Column 6, I report Poisson regression results for *RallyAna*. Note that *netAna* is included in both Columns 5 and 6 because topics that are more negative could lead to more hesitation and more intense back-and-forth battles between analysts and executives. I also see that female analysts make 0.02 fewer comments in their interaction with management ($p<0.01$), consistent with Hypothesis H3e. In sum, the analysis of analyst narratives suggests that female analysts create a relatively more relaxed environment on conference calls and exert less pressure on executives. This is consistent with the notion that female analysts value connections with firm management more and are conservative when asking questions.

[39] One concern is that hesitations are representative of lack of ability. I re-run the regression with all analyst-level control variables for all I/B/E/S analysts and observe similar results.

Table 2.6: Analyst gender and textual characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	netAna	netGIAna	uncertainAna	numberAna	HesitAna	RallyAna
FemaleAna	-0.014	0.036***	-0.052***	-0.082***	-0.177***	-0.020***
	(0.013)	(0.013)	(0.011)	(0.007)	(0.017)	(0.006)
Rec	0.033***	0.025***	0.028***	-0.007***	0.017***	0.009***
	(0.004)	(0.005)	(0.004)	(0.002)	(0.005)	(0.002)
AnaCountLog	-0.158***	-0.016	-0.040***	-0.016	-0.842***	-0.475***
	(0.016)	(0.020)	(0.013)	(0.010)	(0.019)	(0.008)
WordsQNALog	0.028**	-0.039**	-0.002	0.064***	0.829***	0.466***
	(0.014)	(0.018)	(0.011)	(0.009)	(0.016)	(0.007)
netAna					-0.011***	0.007***
					(0.002)	(0.001)
Constant	-0.071	2.650***	1.668***	0.171***	-4.895***	-1.500***
	(0.100)	(0.132)	(0.077)	(0.062)	(0.114)	(0.046)
Observations	442,211	442,211	442,211	442,211	442,211	442,211
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.033	0.028	0.017	0.043		
Pseudo R2					0.138	0.108

2.4.4 Executives' narrative characteristics

Hypothesis H4 predicts that female executive conference call participation is associated with less uncertain tone, more back-and-forth comments, and fewer speech hesitations. Table 2.7 present regression results testing this hypothesis. Similar to Table 2.6, I add *netAna* as a control variable because analysts lead the direction of discussion with firm management and thus, I can regard executives' narratives as a response to analysts' questions. In Column 1, I find, consistent with Hypothesis H4a, that *FemaleExe* negatively affects the percentage of uncertain sentiment.[40] Similarly, as shown in Column 2, there is a negative relationship between hesitations and female executives, supporting Hypothesis H4b. The effect is also economically important. Specifically,

[40] Untabulated results show no difference in executive tone or numerical content by gender.

given a mean count of 2.1 hesitations for executives, female executives exhibit 0.24, or 12%, fewer hesitations. Consistent with Hypothesis H4C, the results in Column 3 indicate that interactions with only female executives have 0.036 more back-and-forth comments compared to interactions with only male executives. In sum, the results of Table 2.7 support hypotheses H4a through H4c that female executives are under greater pressure from analysts but are still more professional in answering questions compared with male executives.

Table 2.7: Executive gender and narrative variables

	(1)	(2)	(3)
VARIABLES	uncertainExe	HesitExe	RallyAna
FemaleExe	-0.040***	-0.242***	0.036***
	(0.015)	(0.032)	(0.011)
netAna	-0.013***	0.004***	0.007***
	(0.001)	(0.001)	(0.001)
Rec	-0.002	0.026***	0.009***
	(0.001)	(0.003)	(0.002)
AnaCountLog	0.011	-1.093***	-0.475***
	(0.008)	(0.015)	(0.008)
WordsQNALog	-0.007	1.142***	0.467***
	(0.007)	(0.014)	(0.007)
Constant	0.968***	-6.188***	-1.508***
	(0.052)	(0.100)	(0.046)
Observations	442,211	442,211	442,211
Firm controls	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Adjusted R2	0.085		
Pseudo R2		0.270	0.108

2.4.5 Analyst-management interaction interruptions

I next examine Hypothesis H5 regarding whether female participants are interrupted more than their male counterparts are. Table 2.8 contains the results. The dependent variable, *InterruptAnaExe*, is the total number of interruptions made by all executives and received by the focal analyst in a conference call.[41] Because statements that are more negative may incur more interruptions, I control for the net tone of each analyst. Given that it is reasonable to believe that longer discourses are positively associated with interruptions, I also control for longer discourses with log-transformed total number of words the analyst speaks, *WordsAnaLog*.

Panel A reports Poisson regression results.[42] The estimated coefficients of *netAna* and *WordsAnaLog* are as expected. Unconditionally, I find that there is no difference in the interruptions of female analysts compared with their male counterparts. To investigate how female and male executives interrupt analysts' statements differently, I further separate interruptions made by female (*InterruptAnaFemaleExe*) and male (*InterruptAnaMaleExe*) executives and report results in Column 2 and Column 3, respectively. I add *FemaleExeCount* (*MaleExeCount*), the number of female (male) executives, to the corresponding model to eliminate its effect on the number of interruptions. The results indicate that a female analyst is interrupted 34% less when counting female executives' interruptions ($p<0.05$). However, I do not observe more interruptions made by male executives. In summary, the finding is generally unsupportive of hypothesis H5a and is consistent with an in-group favoritism explanation (Jannati et al., 2020) in which female executives treat female analysts more favorably by interrupting them less.

[41] Interruptions made by conference call operators are excluded (0.0014% of 9,965 interruptions).
[42] I create indicator variables, which are equal to 1 if a corresponding participant is interrupted and 0 otherwise, run logit regressions, and yield similar results in terms of gender difference.

Table 2.8: Interruptions

Panel A. Analyst interruptions

VARIABLES	(1) InterruptAnaExe	(2) InterruptAnaFemaleExe	(3) InterruptAnaMaleExe
FemaleAna	-0.062	-0.338**	-0.043
	(0.047)	(0.167)	(0.049)
netAna	-0.041***	-0.044	-0.041***
	(0.008)	(0.039)	(0.008)
WordsAnaLog	1.245***	1.411***	1.235***
	(0.029)	(0.113)	(0.030)
Rec	-0.034**	-0.019	-0.036**
	(0.017)	(0.070)	(0.017)
FemaleExeCount		1.069***	
		(0.188)	
MaleExeCount			0.014
			(0.025)
Constant	-8.733***	-11.416***	-8.775***
	(0.272)	(0.987)	(0.287)
Observations	442,211	442,211	442,211
Firm controls	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Pseudo R^2	0.191	0.191	0.193

Panel B. Executive interruptions

VARIABLES	(1) E	(2) EA	(3) EE	(4) EFA	(5) EMA	(6) EFE	(7) EME
FemaleExeDummy	0.236***	0.207*	0.237***	0.103	0.222**	-2.170***	0.790***
	(0.077)	(0.108)	(0.092)	(0.221)	(0.113)	(0.358)	(0.109)
CEO	0.328***	0.132**	0.500***	0.149	0.129*	0.329	0.536***
	(0.053)	(0.063)	(0.070)	(0.159)	(0.067)	(0.257)	(0.073)
CFO	0.440***	0.474***	0.440***	0.221	0.504***	0.541**	0.472***
	(0.047)	(0.060)	(0.060)	(0.163)	(0.064)	(0.235)	(0.062)
netExe	-0.126***	-0.188***	-0.081***	-0.214***	-0.184***	-0.016	-0.081***
	(0.016)	(0.023)	(0.021)	(0.066)	(0.024)	(0.101)	(0.021)
WordsExeLog	0.657***	0.876***	0.506***	0.907***	0.871***	0.822***	0.504***
	(0.020)	(0.028)	(0.024)	(0.087)	(0.029)	(0.106)	(0.025)
AnaCount		0.066*					
		(0.037)					
ExeCount			0.054*				
			(0.028)				
FemaleAnaCount				0.435***			
				(0.076)			
MaleAnaCount					0.115***		
					(0.025)		
FemaleExeCount						0.888***	
						(0.266)	
MaleExeCount							0.109***
							(0.030)
Constant	-6.315***	-8.419***	-5.973***	-9.921***	-8.464***	-9.192***	-6.270***
	(0.213)	(0.279)	(0.281)	(0.676)	(0.298)	(1.091)	(0.280)
Observations	169,432	169,432	169,432	169,432	169,432	169,432	169,432
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo-R^2	0.250	0.206	0.244	0.201	0.204	0.279	0.241

Panel C. Challenging vs. dominating interruptions by male executives

VARIABLES	(1) Challenge	(2) Dominate
FemaleExeDummy	0.945***	0.503***
	(0.195)	(0.142)
CEO	4.153***	
	(0.243)	
CFO	2.597***	0.933***
	(0.254)	(0.076)
netExe	-0.132***	-0.072***
	(0.043)	(0.024)
WordsExeLog	0.726***	0.266***
	(0.061)	(0.024)
MaleExeCount	0.240***	0.025
	(0.045)	(0.038)
Constant	-11.958***	-4.655***
	(0.631)	(0.338)
Observations	169,432	169,432
Firm controls	Yes	Yes
Year-quarter FE	Yes	Yes
Firm FE	Yes	Yes
Pseudo R2	0.391	0.201

Table 2.8 Panel B examines interruptions received by executives. I aggregate all statements of a participating executive to generate one call-executive observation in the dataset. *FemaleExeDummy* is an indicator variable equal to 1 if the executive is female. The dependent variable, *InterruptExe*, is the number of interruptions the executive receives in a conference call. I include CEO and CFO dummies because interruptions are less likely to occur the higher is the status of the executive. I also add executive statement tone and length as additional controls as I expect that there will be more interruptions when the executive's tone is more negative. Results in Column 1 show that female executives receive 24% more interruptions compared with male analysts, supporting Hypothesis H5b.

Because both analysts and executives could interrupt a speaking executive, I separately count interruptions made by analysts (*InterruptExeAna* or *EA* in Column 2) and executives (*InterruptExeExe* or *EE* in Column 3). Among 10,178 interruptions made to executives, analysts account for 4,745 (47%) and executives account for 5,433 (53%). I add potential interrupter count (e.g. *AnaCount*, *ExeCount*, etc.) as a control in corresponding specifications because more potential interrupters may lead to more interruptions. Interestingly, interruptions made by analysts and executives exhibit similar gender bias as female executives receive 21% and 24% more interruptions made by analysts and other executives, respectively. To evaluate gender differences for interrupters, interruptions made by analysts and executives are each split based on the gender of interrupter (i.e., *InterruptExeFemaleAna* or *EFA* in Column 4, *InterruptExeMaleAna* or *EMA* in Column 5, *InterruptExeFemaleExe* or *EFE* in Column 6, and *InterruptExeMaleExe* or *EME* in Column 7). For example, *InterruptExeFemaleAna* denotes the number of interruptions made by female analysts towards the focal executive. Column 4 shows that there is no gender difference in female analyst interruptions, whereas Column 5 shows that female executives can expect to be interrupted 22% more by male analysts. In addition, when the interrupters are other female executives, female executives only receive 11.4% (=$e^{-2.17}$) of interruptions received by male executives (Column 6), again consistent with in-group favoritism. In contrast, male executives will interrupt their female colleagues 79% more often compared with interrupting their other male colleagues (Column 7). I therefore find support for Hypothesis H5b.

Collectively, my results show that men and women exhibit different patterns of interrupting other conference call participants. Women are more reluctant to interrupt other female participants while male executives are more inclined to interrupt women, especially their female colleagues. These findings echo prior evidence that women have a strong in-group favoritism (Rudman and Goodwin, 2004; Tannen, 1990) and demonstrate a potential "internecine conflict" (or male "power jockeying") and discrimination against women within C-suites.[43]

Since male executives are more likely to have high ranks in C-suite, this internecine conflict can be a result of high-ranked male executives' dominance over low-ranked female executives. To examine whether low-ranked male executives also interrupt high-ranked female executives, I first assign a rank score to each executive based on his/her title: CEO (*Rank*=3), CFO (*Rank*=2), and others (*Rank*=1). Next, I classify all interruptions made by male executives based on the relative rank between interrupters and interruptees. Specifically, interruptions made by male executives with lower rank score are called "challenging" interruptions and interruptions made by male executives with same or higher rank score are called "dominating" interruptions. *Challenge* (*Dominate*) is the number of "challenging" ("dominating") interruptions. I re-run the regressions

[43] One concern about the internecine rivalry in the C-suite is that female participants are more likely to be interrupted due to a hierarchy effect. Specifically, IR officers, who have relatively low status and are more likely to be women, are more likely to be interrupted by CEOs and CFOs, who have relatively high status and are more likely to be men. I believe this is not likely to be the case for two reasons. First, I check a small random sample of Q&A session transcripts and do not find IR officers speaking during interruption events. Second, the estimated coefficient of *CEO* and *CFO* are positive in Models 3, 6, and 7, indicating firm participants with a high status are more likely, rather than less likely, to be interrupted. See Table 8 Panel C for further analysis.

in Table 8 Panel B by replacing *InterruptExeMaleExe* with *Challenge* and *Dominance* respectively. Results in Panel C show that the estimated coefficient of *FemaleExeDummy* in the *Challenge* model is 1.88 times as large as that of the *Dominance* model (0.94 vs. 0.50). Economically, when considering only challenging (dominating) interruptions made by male executives, female executives receive 95% (50%) more interruptions compared with male executives. While both coefficients are positive and significant at 1% level, the results suggest that male executives have a stronger intention to interrupt their female colleagues with higher rank than with lower rank, consistent with both gender discrimination and male "power-jockeying".

2.4.6 Market reaction

In Table 2.9, I provide evidence on Hypothesis 6 where I examine the relationship between female analyst earnings conference call participation and the market reactions associated with the conference call. Specifically, I estimate and compare the market reaction to female and male analysts' tone. Market reaction is measured by the 4-factor adjusted CRSP value-weighted cumulative abnormal stock return over a [-1,+1] window around each conference call date (*CAR*). I use weighted average net tone (positive tone minus negative tone) of all participating female analysts in a call, *netFemaleAnaCall*, to proxy for the opinion of participating analysts. The net tone of male analysts, *netMaleAnaCall*, and the net tone of executives, *netExeCall*, are included along with firm and call controls.[44,45] In addition, I use the proportion of female analysts, *FemaleAnaPct*, to examine how the absolute value of *CAR*, |*CAR*|, is affected. If the market is less sensitive to female analysts' participation, I expect that the estimated coefficient of

[44] *netFemaleAnaCall* and *netMaleAnaCall* have a weak correlation (ρ=0.0095).

[45] Replacing each LM tone variable with its corresponding Harvard GI variable yields qualitatively similar results.

netFemaleAnaCall will be smaller than that of *netMaleAnaCall* and that *FemaleAnaPct* will be negative.

Column 1 contains *CAR* results and shows, consistent with my expectations, that *netFemaleAnaCall*, *netMaleAnaCall*, and *netExeCall*, are all positive and significant. Comparing the coefficients, I see that the market reacts similarly to male analyst and executive tones (p=0.17 for F-test). However, the market reaction to analyst tone is significantly different by analyst gender (p<0.001 for F-test). Specifically, a 1% increase in female (male) analyst tone is associated with 0.5% (1.3%) higher market reaction. In Column 2, I find that all-female-analyst conference calls are associated with 0.58% lower market reaction magnitude, which translates to a 40 million dollar market capitalization difference at the sample mean. In sum, my market reaction analysis indicates that investors discount female analyst participation, thus providing support for Hypothesis H6.

Table 2.9: Market reaction

	(1)	(2)
VARIABLES	CAR	|CAR|
netFemaleAnaCall	0.499***	
	(0.068)	
netMaleAnaCall	1.271***	
	(0.054)	
FemaleAnaPct		-0.578**
		(0.234)
netExeCall	1.147***	
	(0.066)	
AnaCountLog	1.030***	0.108
	(0.153)	(0.104)
WordsQNALog	-0.941***	0.520***
	(0.129)	(0.085)
Constant	6.435***	0.478
	(0.918)	(0.603)
Observations	62,644	62,644
Firm controls	Yes	Yes
Year-quarter FE	Yes	Yes
Firm FE	Yes	Yes
Adjusted R2	0.096	0.224

2.5 Robustness tests

2.5.1 Gender as probability

One concern of my results is that may not be accurate because my initial assignment of gender is a binary variable and I draw inferences based on probabilities. To provide evidence on the robustness of my results, I replace the indicator variable, *FemaleAna*, with a continuous variable, *FemaleProb*, as the probability of being female given by each gender algorithm. Analysts determined manually as female (male) are assigned a probability of 1 (0).[46] Multivariate analysis results remain unchanged when I replace *FemaleAna* with *FemaleProb*.

2.5.2 Firms' gender attitude and analyst conference call participation

My evidence of gender discrimination may result from a firm's general social responsibility characteristics. To capture firm social responsibility, I follow Lins et al. (2017) using the corporate social responsibility (CSR) score based on MSCI ESG Stats Database (formerly known as KLD).[47] To the extent that CSR score captures firm gender attitudes, more socially responsible firms may exhibit less discrimination against females.

I add an interaction term for *FemaleAna* (or *FemaleExe*) and CSR score in all regressions. Untabulated results indicate that the interaction term is not significant in almost all models. The only exception is that the disadvantage of female analysts in abnormal interaction length is weaker for high CSR firms. Specifically, a one-standard-deviation increase of CSR score is associated with a 1.17% increase in abnormal interaction length for female analysts. Apart from this latter result, the results indicate that firm-level gender characteristics and policies have little influence on direct and indirect gender discrimination on earnings conference calls. Collectively, I interpret these results as evidence that gender discrimination is deeply rooted in interpersonal communication as a micro-institution of gender-power relationships in society (Jacobi and Schweers, 2017; Zimmerman and West, 1975). However, the corporate gender equality movement is still relatively young and thus, my findings suggest that firm-level gender attitude is still

[46] Because *Gender-guesser* gives five possible results (male, female, mostly male, mostly female, and androgynous) instead of probability, I assign 0.25 (0.75) to "mostly male" ("mostly female"). Using other probability including 0.2/0.8 and 0.33/0.67 does not qualitatively change the results.
[47] See Krüger (2015) for a detailed discussion regarding MSCI ESG ratings.

drowned out in a larger gender inequality backdrop that is manifested in the earnings conference call environment.

2.6 Conclusion

In this paper I use a large sample of quarterly earnings conference call transcripts to investigate gender discrimination issues within the interactions between two high-profile professions—sell-side analysts and public firm executives. First, I find that women are at a disadvantage in conference call participation. Second, I investigate the linguistic characteristics of analysts and executives by parsing conference call transcripts into conversation blocks. I find conditional on analysts' participation that management of firms treats female analysts with less respect compared to their male counterparts during conference calls. Specifically, female analysts have fewer follow-up opportunities to interact with executives and speak less. Consistent with a relatively weaker status and a desire to be more agreeable in a male-dominated profession, female analysts' narratives have more favorable tone, less numerical content, fewer speech hesitations, and fewer back-and-forth comments. I also find evidence consistent with gender stereotyping in firm management, with female executives displaying less uncertain sentiment and fewer speech hesitations when answering analysts' questions. However, they appear to be under more pressure from analysts with more back-and-forth comments.

I also examine the occurrence of interruptions during analyst-executive interactions in earnings conference calls and find that female analysts receive fewer interruptions from female executives compared with male executives—an in-group favoritism. In addition, I find that male executives treat male and female analysts equally in terms of interruptions. With regard to analysts' interruptions of executives, I find that female analysts interrupt female and male executives to a similar extent, but male analysts interrupt female executives more. Interestingly, while female executives tend to interrupt their female colleagues less, male executives are more likely to interrupt a female colleague than a male one, particularly when the female executive is in a superior role, suggesting an "internecine conflict" or gender-based "power jockeying". I also find that the stock market underreacts to female analysts' participation on conference calls and that my results are. Finally, my results are robust to using my gender variable as a probability and to a firms' general CSR attitude.

In sum, my results indicate that, although prior studies find that women possess superior ability as analysts (i.e., superior forecast accuracy, large brokerage affiliation, and all-star designation) that is valued by firm management (Fang and Huang, 2017; Green et al., 2009; Kumar, 2010), they are in general less "visible" and poorly treated during conference calls relative to male peers. Similarly, female executives are under pressure from both analysts and their male colleagues.

CHAPTER THREE

GENDER AND ANALYST REPORTS

3.1 Introduction

Gender differences in writing abilities have long been recognized despite males and females having similar psychological attributes and cognitive abilities (Hyde, 2005, 2014; Reilly, Neumann, and Andrew, 2019; Reynolds, Scheiber, Hajovsky, Scwartz, and Kaufman, 2015). The superiority of females' writing skills are salient from an early age to adulthood (Reynolds et al., 2015; Scheiber, Reynolds, Hajovsky, and Kaufman, 2015). In addition, gender differences are also observed in writing styles of formal written text (Argamon, Koppel, Fine, and Shimoni, 2003).

Writing communication skills have been consistently listed as one of the most valued skills on Wall Street (Alsop, 2004; Weber and Cutter, 2019). Companies have begun to hire more liberal arts graduates for their communication skills (Waller, 2016). A burgeoning literature of textual analysis examines various types of written communication including news articles, financial filings, earnings conference call transcripts, and social media information (Antweiler and Frank. 2004; Li, 2008; Loughran and McDonald, 2011, 2013, 2016; Matsumoto, Pronk, and Roelofsen, 2011; Tetlock, 2007). However, an examination of how author gender plays a role in written financial communication is still lacking.

In this paper, I examine gender differences in an important written communication form in the business world—sell-side analyst reports (hereinafter analyst reports). Specifically, I investigate the following research questions: (1) Are there gender differences in writing abilities of analyst reports? (2) Do female and male analysts exhibit different attitudes toward covered companies in reports? (3) Do female and male analysts have difference focuses of topics in reports? (4) Does the market react to female analysts' reports differently?

Sell-side analysts are pivotal participants in capital markets (Barber, Lehavy, McNichols, and Trueman, 2001; Brown, Call, Clement, and Sharp, 2015; Stickel, 1991; Womack, 1996). Writing research reports is a primary task of analysts. In a typical report, analysts provide quantitative measures (i.e., earnings forecasts, stock recommendations, and price targets) and written analysis (Huang, Zang, and Zheng, 2014). Although quantitative information is the essence of an analyst report, written text serves as the foundation of the analysis and is informative to the market (Asquith, Mikkhail, and Au, 2005; De Franco, Hope, Vyas, and Zhou; 2015; Huang et al., 2014).

In addition to the general writing skill gender gap, gender differences can exist in analyst reports for other reasons. First, females in the financial sector are faced with gender discrimination. Previous studies provide extensive evidence in various professions including firm executives (Catalyst, 2020), mutual fund managers (Niessen-Ruenzi and Ruenzi, 2019), financial advisers (Egan, Matvos, and Seru, 2018), and financial analysts (Kumar, 2010). Because the financial analyst profession is dominated by men, the ability and opinion of female analysts may be undervalued. Moreover, subject to gender stereotypes, investors might also scrutinize female analysts' reports more carefully and regard them as less credible (Bloomfield, Rennekamp, Steenhoven, and Stewart, 2020). In other words, if investors undervalue reports issued by female analysts, female analysts may adjust their writing quality proactively to adapt to the high standard required by investors. Both low evaluation and high assessment standards for females may cause female analysts to exert more effort and subsequently produce reports with higher quality (Green, Jegadeesh, and Tang, 2009). For example, Hengel (2020) finds that female-authored academic papers are better written and female authors' writing improves over time due to a tougher peer review process.

Second, men and women have different inherent and socially-learned characteristics. In general, women are found to be more conservative (Croson and Gneezy, 2009; Faccio et al., 2016; Johnson and Powell, 1994), less overconfident (Barber and Odean, 2001; Huang and Kisgen, 2013), more ethically-oriented (Dollar, Fisman, and Gatti, 2001; Franke, Crown and Spake, 1997; Reiss and Mitra, 1998), and less competitive (Gneezy, Niederle and Rustichini, 2003; Gneezy and Rustichini, 2004; Niederle and Vesterlund, 2007). Different characteristics of male and female analysts may affect two aspects of report contents: time horizon and conflict of interest. Report compiling is a subjective process in which analysts integrate information collection channels, information topics, and information interpretation in heterogeneous ways (Asquith et al., 2005; Huang, Lehavy, Zang, and Zheng, 2018). For example, one important dimension of analyst reports is time horizon which describes how an individual values the future relative to the present (Chen, Jung, Lim, and Yu, 2020; Graham, Harvey, and Rajgopal, 2005; He and Tian, 2013; Jung, Shane, and Yang, 2012). Because females are more sensitive to risk, female analysts may pay more attention to firms' long-term prospects that merit more uncertainty resolution. In addition, analysts suffer from conflict of interest. Analysts have incentives to curry favor firm management of firms they follow to gain access to management and thus informational advantages (Green, Jame, Markov, and Subasi, 2014; Mayew, 2008). Analysts may issue optimistic stock recommendations to attract underwriting relationships and boost trading activities (Cowen, Groysberg, and Healy, 2006; Michaely and Womack, 1999; Groysberg, Healy, and Maber, 2011). Analysts also "walk down" their earnings forecasts so that firms can beat estimates at earnings announcements (Bradshaw, Lee, and Peterson, 2016; Richardson, Teoh, Wysocki, 2004). Therefore, analysts are reluctant to issue unfavorable quantitative information (Barber et al., 2001; Hong and Kubik, 2003; Malmendier and Shanthikumar, 2007). However, analysts believe issuing unfavorable recommendations is a major approach through which they gain perceived credibility of their clients (Brown et al., 2015). I expect to see that female analysts exhibit a more consistent opinion reflected in both report text and quantitative measures.

I analyze a large sample of analyst reports written from 1990 to 2018 to study gender differences among analysts. First, I test whether female analyst reports have higher readability as a measure of writing ability. Readability is an import measure to capture how easily a reader can understand the opinion in written text (Li, 2008; De Franco et al. 2015). Using five readability scores, I find female analysts write more readable reports, consistent with the notion that females attempt to offset higher evaluation standards with more readable reports (Hengel, 2020). Further, I find that female analysts issue shorter report in terms of both number of words and number of pages. These results suggest that female analysts tend to substitute quality for quantity in their reports.

Next, I examine gender differences in report sentiment. Consistent with the notion that female analysts are less susceptible to conflicts of interest, I find that the tone of female analyst reports is less favorable. Furthermore, because analysts may include various topics in their reports, I compare the percentage of financial and nonfinancial content (Huang et al., 2014). Specifically, I use two measures to capture financial content: the financial dictionary of Matsumoto et al. (2011) and numerical content (Zhou, 2018). I find that female analysts discuss less financial content in reports. In addition, I examine gender differences in short-termism. I find female analyst reports are associated with lower proportion of short-term content but higher proportion of long-term content.

Last, I compare the market reaction to qualitative content of reports. I find no gender differences in terms of market reaction to report sentiment and length. However, more readable reports are related to stronger market reaction for male analysts but weaker market reaction for female analysts. My analyses are robust to the inclusion of firm and analyst characteristics as well as year, firm, and brokerage fixed effects.

This paper is the first attempt to compare gender differences in the qualitative content of analyst reports. My study contributes to the literature in two ways. First, I add to the literature of gender issues in the workplace, especially among high-paying professionals. Given the low representation of women in high-paying jobs, whether there exists gender discrimination or gender differences in ability has become a long-standing issue (Bertrand, Black, Jensen, and Lleras-Muney 2019; Matsa and Miller, 2011; Adams and Funk, 2012). My results suggest that females may respond to gender discrimination proactively by improving writing quality.

Second, I contribute to the literature on financial analysts. Whether financial analysts are informative is open to debate (Altınkılıç and Hansen 2009; Bradley, Clarke, Lee, and Ornthanalai 2014). Prior studies document that gender is a dimension which predicts report timing, forecast boldness, stock recommendation favorableness, career advancement, and market reaction (Bosquet, de Goeij, and Smedts, 2014; Kumar, 2010, Green et al. 2009; Li et al. 2013). My findings suggest that analyst gender also predicts writing styles with regard to textual sentiment, readability, and informativeness.

3.2 Literature review

3.2.1 Gender differences in communication

Hyde (2005) proposes the gender similarity hypothesis, that males and females "are similar on most, but not all, psychological variables. That is, men and women, as well as boys and girls, are more alike than they are different" (p. 581). Regarding verbal performance, she reviews meta-analyses of gender differences in various cognitive attributes and finds that although gender differences in vocabulary and reading comprehension are trivial, moderate gender differences in writing performance exists. Gender differences in writing performance are also documented in other studies. For example, Reynolds et al. (2015) compare the performance of children and adolescents from age 7 to 19 in Kaufman intelligence and achievement tests and find that girls perform better than boys in spelling and written expression with an effect size of 0.46 inconsistent with the gender similarities hypothesis.

If gender differences in writing result from gender stereotyping, gender differences are expected to decline as social expectations for females change (Feingold, 1988). However, Reilly et al. (2019) conduct large sample research on student achievement in writing from the National Assessment of Education Progress (NAEP) from 1988 to 2011 and find that gender differences are consistent over time at a medium level (d=0.55). Moreover, multiple studies document a developmental trend that female advantages in writing performance appear at a young age (i.e., 6 to 10 years old), widen until high school, and stabilize in adolescence (Scheiber et al., 2015; Reilly et al., 2019; Peterson, 2018).

In addition to female advantages in writing abilities, gender differences also exist in writing styles. Argamon et al. (2003) examine a large sample of writing in the British National Corpus of books and articles. They find men use more noun specifiers and women use more pronouns.[48] They further argue that the results are consistent with earlier findings that women pay more attention the relationships than men do (Tannen, 1990).

[48] "Pronouns send the message that the identity of the 'thing' involved is known to the reader, while specifiers provide information about 'things' that the writer assumes the reader does not know." (Argamon et al., 2003, p. 323)

3.2.2 Gender differences in analysts

Gender differences are substantial among analysts. First, women are significantly underrepresented. Prior studies show that women account for less than 15% of analysts in the Institutional Brokers Estimate System (I/B/E/S) database (Fang and Huang, 2017; Green et al., 2009; Kumar, 2010). Second, female analysts exhibit heterogeneity in industry coverage distribution. Particularly, female analysts have a relatively higher concentration in retail, clothing, publishing, and textiles while they are substantially underrepresented in coal, metals, automobiles, and defense (Green et al., 2009; Kumar, 2010). Third, female analysts cover large firms and are hired by larger brokerage houses (Francis, Shohfi, and Xin, 2020; Kumar, 2010). Fourth, female analysts are more likely to be designated as Institutional Investor all-stars (Fang and Huang, 2017; Green et al., 2009; Kumar, 2010). Fifth, female analysts cover a smaller number of firms and rely more on independent research instead of earnings news (Green et al., 2009). Although gender differences in role and industry-selection preferences provide an explanation to female underrepresentation, whether a gender difference exists with regard to analyst forecast ability or market reaction to reports is unclear (Green et al., 2009; Kumar, 2010; Fang and Huang, 2017; Li et al., 2013).

3.2.3 Analyst reports

Analyst reports are a major information outlet through which analysts propagate their insights about covered firms to investors. To achieve career advancement, writing informative reports is a fundamental requirement for analysts (Brown et al., 2015; Mikhail, Walther, and Wills, 1999; Hong and Kubik, 2003). Analyst reports include both quantitative measures—earnings forecasts, stock recommendations, and price targets—and written analysis of the firm (Asquith et al., 2005; Huang et al., 2014; De Franco et al., 2015; Huang et al., 2018; Twedt and Rees, 2012).

Prior studies find that these quantitative outputs are informative to the stock market (Womack, 1996; Brav and Lehavy, 2003; Li, Ramesh, Shen, and Wu, 2015). However, the main body of analyst reports is written analysis of the company which underlies the quantitative measures. “In the end, stock ratings and target prices are just the skin and bones of analysts' research. The meat of such reports is in the analysis, details, and tone. Investors who are willing to spend the time can easily figure out what an analyst really thinks about a stock by reading a research report.” (Tsao, 2002)

If all information in report text is also reflected in quantitative measures, I do not expect to observe a significant market reaction when controlling for relevant quantitative information. However, prior studies show that analyst reports cover a wide range of financial and nonfinancial topics including performance, strategy, risk, management, competitive position, stakeholders, and economic conditions (Asquith et al. 2005; Previts, Bricker, Robinson, and Young, 1994) and textual content in analyst reports is incrementally informative to the market (Asquith et al., 2005; Caylor, Cecchini, and Winchel, 2017; De Franco et al., 2015; Huang et al. 2014). This implies that the text in analyst reports contains subtle information which is valuable to investors.

Further, investors may regard writing as the most valuable information embedded in an analyst report because investors do not simply follow analyst’s conclusions but construct their own investment decisions only partly based on information provided in analyst reports (Huang et al., 2014). According to *Institutional Investor* magazine’s annual survey of institutional investors, writing useful reports is considered more important as an All-Star analyst voting criterion than stock recommendation profitability.

3.3 Hypothesis development

Writing reports is a core task for sell-side analysts. However, report writing entails a large amount of effort. Extant literature finds that women are more conscientious than men. Women conduct more organizational citizenship behavior and more discretionary work (Lovell et al., 1999; Kmec and Gorman, 2010). Moreover, female directors have higher board input (Adams and Ferreira, 2009). Financial analysts have traditionally been regarded as a “boys club” profession (Fang and Huang, 2017). The higher standard of scrutiny exerted by investors may further entail female analysts to invest more time and effort when writing reports (Hengel, 2020).

Information processing is costly (Hirshleifer and Teoh, 2003). Complex text significantly increases information processing cost of readers (Lehavy, Li, and Merkley, 2011). Firms issue annual reports which are often difficult to read, obscuring value-relevant information (Li, 2008; Lo, Ramos, and Rogo, 2017). To attract investor attention and increase influence, analysts are expected to issue more readable reports. Females are faced with a higher evaluation standard than males (Bloomfield et al., 2020; Hengel, 2020; Madera, Hebl, Dial, Martin, Valian, 2019). In addition, females have an advantage of writing skills which I expect to result in better written analyst reports (Peterson, 2018; Reilly et al., 2019). Therefore, I propose:

Hypothesis 1: Female analysts issue more readable reports.

A higher evaluation standard may introduce a quantity-quality tradeoff for women. Specifically, women may reduce the number of outputs but put more effort into each of them to increase quality (Hengel, 2020). Prior studies report evidence consistent with this tradeoff that female analysts are less likely to revise earnings forecasts, issue few stock recommendations, but have higher forecast accuracy (Kumar, 2010; Li et al., 2013). To improve report quality, female analysts may spend more effort in issuing reports to support the quantitative outputs and thus issue longer reports. Based on the above discussion, I have:

Hypothesis 2a: Female analysts issue longer reports.

Hypothesis 2b: Female analysts issue shorter reports.

Investors are the primary consumers of analyst reports. Buy-side clients refer to industry knowledge and forecasts provided in analyst reports to make their own investment decisions (Brown et al., 2015). Because sell-side analysts are particularly vulnerable to conflicts of interest, they develop more credibility with buy-side clients when they issue forecasts or recommendations that are less favorable than consensus (Brown et al., 2015). Women have higher ethical standards than men (Dollar et al., 2001; Franke et al., 1997; Reiss and Mitra, 1998) and thus are less likely to be influenced by conflicts of interest. Specifically, the likelihood of issuing optimistic stock recommendations is significant lower for female analysts and the likelihood of issuing bolder forecasts is significantly higher for female analysts (Bosquet et al., 2014; Kumar, 2010). Therefore, female analysts may exhibit more negative sentiment in their reports.

Hypothesis 3: The tone of female analysts' reports is less positive than that of male analysts.

When writing reports, analysts gather a wide range of information. The information can be broadly classified into financial information and nonfinancial information (Huang et al., 2014). Nonfinancial information is not included in a firm's financial reporting system (Stocken and Verrecchia, 2004). However, nonfinancial information, such as customer satisfaction, is value relevant (Ittner and Larcker, 1998; Cao, Myers, and Omer, 2012; Dhaliwal, Li, Tsang, and Yang, 2011; Park, Eisingerich, Pol, and Park, 2013). Compared with financial information, nonfinancial information is more about relationships with stakeholders.

Cognitive differences between women and men suggest that information acquisition methods between the two can be different. Comparatively, women are characterized by a stronger focus on relationships with others (Tannen, 1990). For example, female writers encode readers into text and use more pronouns while men use more noun specifiers in formal writings (Argamon et al. 2003; Tannen, 1990). Nonfinancial information may require more effort for analysts to collect and analyze because the disclosure is not mandatory (Huang et al., 2014). I therefore have:

Hypothesis 4a: Female analysts discuss less financial content in reports.

Another important dimension of analyst reports is forecast horizon. While the majority of forecast are short-term oriented, long-term forecasts are also informative (Chen, Jung, Lim, and Yu, 2020; Chen, Shane, Yang, and Zhang, 2017). However, forecasting long-term activities such as innovation is difficult. Previous studies show that women are more conservative and risk-averse (Croson and Gneezy, 2009; Faccio et al., 2016; Johnson and Powell, 1994) and are therefore expected to focus more on short-term performance. On the contrary, because managers are in general short-term oriented, female analysts who are less overconfident and are less susceptible to conflicts of interest are expected to focus more on long-term related topics. Thus, I propose the hypothesis with a neutral form:

Hypothesis 4b: Female analysts discuss similar amount of content in terms of forecast horizon in reports.

Prior studies find that report text provides incremental information beyond quantitative summary measures (Huang et al., 2014). Investors may put less weight on reports issued by female analysts due to gender stereotypes especially since gender can be easily inferred based on analyst name(s) on each report. For example, Bloomfield et al. (2020) find that in an experimental setting, female analysts are evaluated by investment professionals as less promotable when they exhibit unexpected behavior. Men, who account for a large proportion of investor community, also exhibit bias against female analysts (Luo and Salterio, 2020).

However, female analysts may adapt to discrimination and gender stereotyping by creating better and/or more efficient reports. The ability and opinion of female analysts may be undervalued because the financial analyst profession is dominated by men. Moreover, subject to gender stereotypes, investors might also scrutinize female analysts' reports more carefully and perceive their credibility as lower. Given potential discrimination against women, female analysts may not be representative of average women because analysts are a competitive profession (Kumar, 2010). Because female analysts compete in an industry dominated by men, they are likely to be more competent than their male counterparts—a "self-selection" phenomenon (Kumar, 2010). In a similar vein, if investors undervalue reports issued by female analysts, I expect female analysts to adapt to the high standard required by investors and improve their analysis and writing skills over time (Hengel, 2020). This improved ability suggests that the market may react more strongly to female analyst reports. Due to the competing arguments, whether the market reacts differently to male and female analysts' report content is an empirical issue. Therefore, I propose

Hypothesis 5a: Market reaction to female analyst text is stronger.

Hypothesis 5b: Market reaction to female analyst text is weaker.

3.4 Sample selection and variable descriptions

3.4.1 Sample selection

I collect a random sample of sell-side analyst reports issued between 1982 and 2018 from Thomson One Investext. I use header information provided by Investext to match analyst reports and other datasets. All analyst reports are downloaded as portable document format (PDF) files.

For text-based PDF documents (i.e., text is searchable), I use *pdftotext* to convert them into text files. For image-based PDFs, I use *Tesseract*, an open-source optical character recognition (OCR) engine, to convert PDFs to plain text. Header information includes the title, report issue date, number of pages, brokerage firm of the analyst, analyst name, a unique number assigned to the report, 6-digit historical Committee on Uniform Security Identification Procedures (CUSIPs)—NCUSIPs, etc. I remove non-English reports. Reports covering multiple stocks are also removed because it is difficult to distinguish firm-specific information (Huang et al., 2014). I remove reports issued by more than one lead analyst because gender-diverse teams may introduce noise to my analysis of gender effects.[49] Reports prior to 1994 and after 2018 are removed due to disproportionally small number of observations.

I match analyst-company pairs in the report sample to I/B/E/S by analyst names and NCUSIPs and verify the matching with broker names. Unmatched reports are deleted in the report sample. I then match analyst reports with I/B/E/S earnings-per-share (EPS) forecast, stock recommendation, and price target datasets.[50] Specifically, for each valid forecast, I/B/E/S only records its announcement date (ANNDATS), the date on which the analyst issues a forecast, and review date (REVDATS), the most recent date on which the analyst confirms the forecast as valid. In other words, multiple reports may share the same record in I/B/E/S. I follow Huang et al. (2014) and use the matching window spanning from two days before the announcement date to two days after the review date.[51] I further define all reports issued with the two-day window as "revision reports" and other reports as "review reports". I only retain reports matched with at least one

[49] Fang and Hope (2020) find that 73% of annual earnings forecasts for U.S. firms from I/B/E/S over the period of 2013 to 2016 are issued by teams. However, the majority of analyst teams are led by one analyst who is in charge. For example, RBC Capital Markets issued a report on Nov. 25th, 2013 and the analyst team consists of a senior analyst, Nik Modi, and three associates. The corresponding record in I/B/E/S only lists Nik Modi as the unique analyst.
[50] I use one-year-ahead EPS forecasts and one-year-ahead price target forecasts.
[51] Price target records in I/B/E/S do not have review dates and hence I only consider a matching window 5 days around the announcement dates.

I/B/E/S earnings forecast, recommendation, or price target. I then take the interaction of the reports and CRSP/COMPUSTAT dataset to obtain stock return and financial data. Reports with less than 100 words are excluded because they are less likely to convey value-relevant information to the market except for templated language. Last, I limit my sample to report observations without missing values for all variables. My final sample consists of 430,356 reports related to 1,696 firms, 3,622 analysts, and 318 brokerages.

3.4.2 Gender determination

To determine analyst gender, I extract first names from full names in report header information and apply gender-API, a gender inference service based on more than 2 million names collected from government records and social networks. Prior studies find that gender-API has superior accuracy compared with other algorithms (Bonham and Stefan, 2017; Santamaría and Mihaljević, 2018). Specifically, gender-API provides an accuracy score ranging from 0 to 100 to exhibit how reliable the gender guess is. All first names with a score less than 80 are manually checked according to Internet search and Capital IQ. An indicator variable, *Female*, is set to 1 (0) for female (male) analysts.

3.4.3 Qualitative information

Readability is generally associated with two indicators—sentence length and number of complicated words. I follow Hengel (2020) and measure analyst report readability with five widely used indices: Gunning Fog (*Fog*), Flesch-Kincaid Grade Level (*FKGL*), Flesch Reading Ease (*FRE*), Dale-Chall (*Dale*), and Simple Measure Gobbledegook (*SMOG*). Because more readable text obtains higher Flesch Reading Ease score but lower scores for other four indices, I multiply the four grade-level scores by negative one. A high score indicates an analyst report is more readable.

Next, I examine two dimensions of textual information of analyst reports: report length and sentiment. Report length is measured with the log number of words in a report (*Word*) and the log number of pages of a report (*Page*). Although previous studies use report length as a measure of readability (Li, 2008; De Franco et al., 2015), it can also represent the effort put forth by analysts, especially when analysts have relatively less intention to obfuscate their readers (i.e. the report is more readable) (Twedt and Rees, 2012). Sentiment of analyst reports is captured by Loughran and McDonald (2011) dictionaries. *Pos* and *Neg* are the ratio of positive or negative words, in percentage. Net sentiment (*Net*) is defined as the difference between *Pos* and *Neg*.

I further investigate report content across three dimensions: financial information, numerical information, and the timeframe information. Financial information is measured with the percentage of financially-oriented words based on the Matsumoto et al. (2011) dictionary (*Fin*). Numerical information is the percentage of numerical information as described by Zhou (2018) (*Number*).[52] Analyst time horizon is measured as the percentage of short-term-oriented words (*Short*) and long-term-oriented words (*Long*) developed by Brochet, Loumioti, and Serafeim (2015).

3.4.4 Descriptive statistics

Table 3.1 shows descriptive statistics for variables by gender (See Appendix I for all variables definitions). Continuous variables are winsorized at 1st and 99th percentiles. Consistent with prior studies on gender issues for analysts, 11.15% of reports in my sample are written by female analysts (Fang and Huang, 2017; Kumar, 2010; Francis et al., 2020). Female analysts issue more readable reports than male analysts but produce reports with fewer words and fewer pages.

[52] *Number* also contains the numerical information embedded in tables, which is not included in other measures.

Textual characteristics also exhibit significant gender differences. The percentage of positive and negative words are both lower for female analysts. Net tone is more positive for female analyst reports. Female analyst reports contain fewer financial words and less numerical content. Female and male analysts are different across various analyst-level characteristics. 22.9% of female analyst reports are written by Institutional Investor All-Star analysts compared with 17.1% of male analyst reports (*Star*). Female analysts are associated with fewer years of forecasting experience (*GenExp*), less firm-specific experience (*FirmExp*), larger brokerage houses (*Broker*), a smaller number of firms covered (*FirmCover*), more reports issued per year (*Frequency*), less accuracy (*Accuracy*).

Covered firm characteristics also exhibit differences by analyst gender. Female reports are associated with lower book-to-market ratio (*BM*), larger logarithm of market capitalization (*Size*), less institutional ownership (*InstOwn*), and more unique industry segments of the firm covered (*Segment*). In sum, the univariate descriptive statistics are consistent with prior studies that female analysts issue bolder and more optimistic forecasts, cover larger firms, and are hired by larger brokerage houses (Kumar, 2010).

Table 3.1: Descriptive statistics

	Male	Female	Difference	t-stat
Readability				
Fog	-15.302	-14.908	-0.393	-19.867***
FKGL	-11.854	-11.485	-0.370	-21.954***
FRE	56.133	57.175	-1.042	-21.380***
Dale	-10.626	-10.586	-0.040	-9.479***
SMOG	-12.934	-12.692	-0.242	-27.445***
Length				
Word	8.032	8.004	0.029	7.732***
Page	1.845	1.803	0.042	13.385***
Textual characteristics				
Pos	0.074	0.071	0.003	8.024***
Neg	0.163	0.158	0.005	8.927***
Net	-0.089	-0.087	-0.002	-3.015**
Topic				
Fin	0.811	0.763	0.048	17.286***
NonFinSentPct	76.556	76.801	-0.245	-4.516***
Number	5.618	5.559	0.059	2.698**
ShortTerm	0.017	0.017	-0.000	-1.259
LongTerm	0.034	0.039	-0.005	-17.452***
Report characteristics				
EFRev	-0.002	-0.002	-0.000	-1.771
RecRev	0.063	0.085	-0.023	-5.245***
PTRev	-0.014	-0.017	0.003	2.848**
EA	0.475	0.441	0.035	14.407***
SameDayReport	4.225	4.322	-0.097	-4.622***
Concentration	5.622	5.771	-0.150	-5.384***
SameDayAnaReport	1.588	1.754	-0.166	-26.968***
CAR	-0.032	-0.114	0.082	2.634**
Runup	0.054	0.055	-0.001	-0.021
Analyst characteristics				
Star	0.171	0.229	-0.058	-31.428***
GenExp	14.263	12.984	1.279	29.593***
FirmExp	4.387	4.100	0.287	12.939***
BrokerSize	63.672	70.832	-7.161	-30.420***
IndCover	3.029	2.800	0.229	23.903***
FirmCover	16.433	14.863	1.570	45.476***
Frequency	56.000	59.913	-3.913	-17.725***
Accuracy	0.014	-0.001	0.015	4.654***
HHI	26.304	29.433	-3.129	-32.445***
Firm characteristics				
BM	0.459	0.406	0.053	30.006***
Size	7.924	8.066	-0.142	-15.715***
Segment	1.500	1.607	-0.106	-25.097***
InstOwn	0.737	0.735	0.002	1.448

3.4.5 Time trend of reports

I plot textual characteristics over time for female and male analysts, respectively. Figure 3.1 plots the number of reports written by female analysts, by male analysts, and the percentage of female reports. Number of reports exhibits a sharp upward trend for both female and male analysts, especially after 2000. However, the proportion of female analyst reports decreases from 10%-15% in the early sample period to around 10% in the last period.[53] The ratio of female reports is consistent with prior studies on gender representation of financial analysts based on the I/B/E/S sample (Fang and Huang, 2017; Kumar, 2010).

[53] The average of number of reports an analyst issues per year exhibits no significant gender difference.

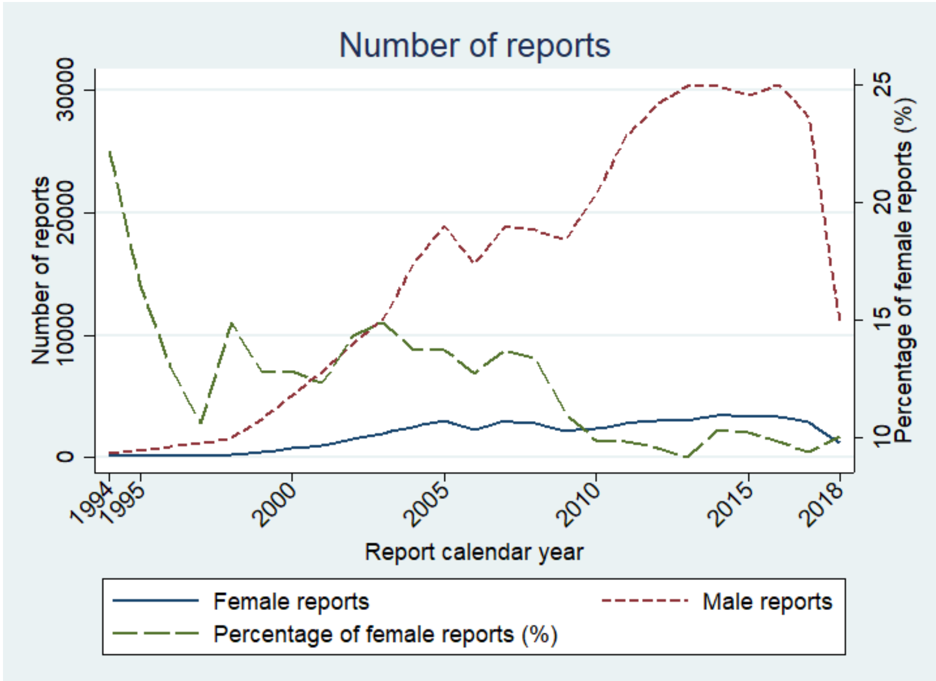


Figure 3.1: Number of analyst reports by gender and year

3.4.6 Report timing

Prior studies find that earnings forecasts and stock recommendations are concentrated around earnings announcement dates (Green et al., 2009; Ivković and Jegadeesh, 2004). Although analysts react to and issue reports for various company events, quarterly earnings announcements are a primary determinant of analyst report issuance. I follow Ivković and Jegadeesh (2004) to calculate the number of trading days relative to quarterly earnings announcement dates (EAD) and report the distribution by gender in Figure 3.2. The results indicate that both female and male analyst reports cluster in the first trading week (i.e., [0,+5] trading days) relative to EAD, followed by the week prior to EAD. The proportion of reports issued in other time periods are lower but do not exhibit salient variance. Compared with male analysts, female analysts are less likely to issue reports immediately around EAD, consistent with the notion that female analysts depend less on earnings news and other information released during EAD window but rely more on independent information gathering (Green et al., 2009).

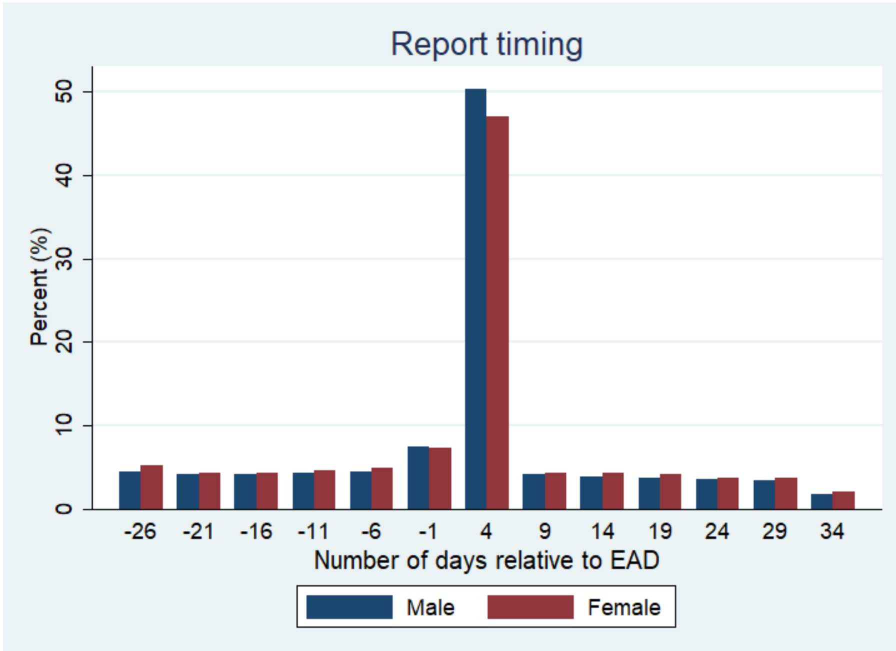


Figure 3.2: Report timing distribution by gender

3.5 Empirical analysis

3.5.1 Analyst report readability and length

To test H1 and H2, I examine how analyst gender is related to report readability and length. Readability is measured by five readability scores. Year, firm, and brokerage fixed effects are also included in all specifications. Standard errors are clustered at the firm level for all OLS models. Results are reported in Table 3.2. I find female analyst reports are more readable in terms of all five readability measures. Column 6 and Column 7 report that female analyst reports are 4% shorter than male analyst reports. H1 and H2b are supported. My results indicate female analysts choose to issue shorter but more readable reports. I interpret the pattern as a quality versus quantity tradeoff for female analysts.

Table 3.2: Analyst report readability and length

VARIABLES	(1) Fog	(2) FKGL	(3) FRE	(4) Dale	(5) SMOG	(6) Word	(7) Page
Female	0.14**	0.13**	0.49**	0.03**	0.08***	-0.04***	-0.05***
	(0.06)	(0.06)	(0.20)	(0.01)	(0.03)	(0.01)	(0.01)
Star	0.23***	0.20***	0.60***	-0.01	0.11***	0.02	0.03***
	(0.04)	(0.03)	(0.10)	(0.01)	(0.02)	(0.01)	(0.01)
GenExp	0.01***	0.01***	0.04***	0.00***	0.00***	0.00***	0.00**
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
FirmExp	0.00	0.00	-0.00	0.00**	-0.00	-0.00***	-0.00***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
BrokerSize	0.00***	0.00***	0.01***	0.00***	0.00***	-0.00***	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
IndCover	-0.05***	-0.05***	-0.15***	0.00	-0.02***	-0.00	-0.00
	(0.01)	(0.01)	(0.03)	(0.00)	(0.01)	(0.00)	(0.00)
FirmCover	-0.01*	-0.01**	-0.00	-0.00	-0.00	0.00***	0.00***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Frequency	0.00*	0.00**	0.00	0.00**	0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BM	0.06	0.04	0.14	0.02	0.02	-0.03***	-0.03***
	(0.07)	(0.06)	(0.16)	(0.01)	(0.03)	(0.01)	(0.01)
Size	-0.02	-0.02	-0.12	0.00	-0.02	0.01	0.00
	(0.03)	(0.03)	(0.08)	(0.01)	(0.01)	(0.01)	(0.00)
Segment	-0.00	-0.00	-0.00	-0.01	-0.01	0.00	0.00
	(0.05)	(0.04)	(0.10)	(0.01)	(0.02)	(0.01)	(0.01)
InstOwn	-0.11	-0.13	0.30	0.03	0.02	-0.04**	-0.03**
	(0.10)	(0.08)	(0.23)	(0.02)	(0.04)	(0.02)	(0.01)
Constant	-15.27***	-11.81***	55.94***	-10.83***	-12.91***	8.15***	1.91***
	(0.27)	(0.23)	(0.67)	(0.07)	(0.12)	(0.05)	(0.04)
Observations	430,356	430,356	430,356	430,356	430,356	430,356	430,356
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Broker FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.216	0.268	0.246	0.231	0.292	0.351	0.297

3.5.2 Analyst report sentiment

H3 argues that the tone of female analyst reports is less positive. I regress net tone on *Female*, analyst variables, and firm variables. Quantitative information is controlled for with the total number of upward revisions minus the total number of downward revisions among earnings forecasts, stock recommendations and price target revisions (*RevFavor*). Results are reported in Table 3.3. Consistent with H3, the female indicator variable negatively predicts report sentiment. However, economic significance of the gender difference is weak. No gender difference is found for negative sentiment (untabulated results). The finding is consistent with the argument that female analysts are less likely to curry favor with firm management with overly optimistic language.

Table 3.3: Analyst report sentiment and content

VARIABLES	(1) Net	(2) Fin	(3) Number	(4) ShortTerm	(5) LongTerm
Female	-0.001**	-0.014***	-0.127***	-0.001***	0.002***
	(0.001)	(0.003)	(0.021)	(0.000)	(0.000)
RevFavor	0.002***	-0.007***	-0.019***	-0.000	-0.000***
	(0.000)	(0.001)	(0.005)	(0.000)	(0.000)
Star	-0.009***	-0.004	-0.072***	0.001***	-0.000*
	(0.001)	(0.003)	(0.020)	(0.000)	(0.000)
GenExp	0.000***	0.001***	0.002**	-0.000*	0.000***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
FirmExp	-0.000***	-0.002***	-0.024***	0.000***	0.000***
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)
BrokerSize	0.000***	0.000***	0.002***	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
IndCover	0.000	0.007***	-0.013**	0.000*	-0.000***
	(0.000)	(0.001)	(0.005)	(0.000)	(0.000)
FirmCover	-0.000	0.001***	0.020***	-0.000*	0.000***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Frequency	0.000***	-0.001***	-0.004***	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Accuracy	-0.001***	0.001	0.006	0.000	-0.000***
	(0.000)	(0.001)	(0.009)	(0.000)	(0.000)
BM	-0.013***	-0.004	-0.128***	0.001***	0.001***
	(0.001)	(0.004)	(0.031)	(0.000)	(0.000)
Size	0.001*	-0.010***	-0.072***	0.000*	-0.000
	(0.000)	(0.002)	(0.014)	(0.000)	(0.000)
Segment	0.001	-0.005**	0.010	-0.000**	-0.001***
	(0.000)	(0.002)	(0.016)	(0.000)	(0.000)
InstOwn	-0.005***	0.007	0.362***	0.000	-0.001
	(0.001)	(0.006)	(0.047)	(0.000)	(0.001)
Constant	-0.094***	0.862***	5.836***	0.014***	0.037***
	(0.003)	(0.016)	(0.122)	(0.001)	(0.002)
Observations	430,356	430,356	430,356	430,356	430,356
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Broker FE	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.277	0.227	0.202	0.168	0.227

3.5.3 Analyst report content

Analysts may have their own unique forecasting approaches. One taxonomy of information is that which is recognized by financial reporting systems (Huang et al., 2014). Because nonfinancial content encompasses a broad scope of topics and is thus difficult to capture, I employ an indirect approach by examining the percentage of financial content with two measures: the percentage of numbers and the percentage of financial words (Matsumoto et al., 2011). In Table 3.3, I report OLS regression results of *Fin* and *Number* on *Female*. Female analyst reports include less financial content and numerical content. The results provide indirect evidence that female analysts value nonfinancial topics more in their reports than male analysts. H4a is supported.

One important dimension of nonfinancial content is forecast horizon. Forecast horizon can be affected by the performance measures stipulated in an accounting system which is essentially short-term oriented (Kaplan, 1984; Marginson and McAulay, 2008). In other words, financial content is more related to short-termism and nonfinancial content is more related to long-termism. If females care less about financial performance, they may be more long-term oriented. I examine the percentage of short-term and long-term oriented words developed by Brochet, Loumioti, and Serafeim (2015). Column 4 and 5 reports the results. Female analysts use fewer short-term oriented words and more long-term oriented words. It suggests that female analysts are more concerned with long-run risk when writing their reports. In sum, although I find statistical significance for various topic measures, the gender difference is not economically significance. One reason is that the word count for each topic is excessively small.

3.5.4 Market reaction

Prior studies find that female analyst forecast revisions elicit stronger market reactions (Kumar, 2010). In Table 3.4, I examine how the market reacts differently to qualitative information within female and male analyst reports. I regress the absolute value of *CAR*, the Fama-French 3 factor and momentum factor adjusted cumulative abnormal return over the [0,+1] window, on analyst gender indicator and three variables related to qualitative information—*Word*, absolute value of *Net* (*NetAbs*), and readability variables. Interaction terms are included to examine how markets react differently to female and male analysts' report content. Because the market may react more strongly when more reports are issued, I also control for the number of reports for the target firm over a [0,+1] window relative to the report date (*Concentration*).[54] Table 3.4 reports the OLS regression results. Readability variables are positive and significant in all specifications. The negative interaction term between female dummy and readability scores suggests that more readable reports written by female analysts induce weaker market reactions.[55] The result is consistent with the notion that female analyst reports are undervalued. In sum, H5b is supported.

[54] I also conduct a firm-day level regression analysis by aggregating reports for the same firm on the same day. Results are similar.

[55] In untabulated analysis, I exclude the reports in earnings announcement windows and re-run models in Table 3.4. My inferences are unaltered.

Table 3.4: Market reaction

VARIABLES	(1) CARabs	(2) CARabs	(3) CARabs	(4) CARabs	(5) CARabs
Female	-0.610	-0.564	0.147	-1.153	-0.883
	(0.570)	(0.565)	(0.570)	(0.765)	(0.615)
Netabs	1.301***	1.302***	1.331***	1.310***	1.321***
	(0.166)	(0.166)	(0.167)	(0.166)	(0.167)
Female×Netabs	-0.246	-0.250	-0.320	-0.241	-0.288
	(0.422)	(0.422)	(0.419)	(0.425)	(0.420)
Word	-0.016	-0.015	-0.021	-0.032	-0.019
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Female×Word	0.043	0.040	0.042	0.062	0.042
	(0.066)	(0.066)	(0.066)	(0.067)	(0.066)
Fog	0.007**				
	(0.003)				
Female×Fog	-0.014*				
	(0.008)				
FKGL		0.007*			
		(0.004)			
Female×FKGL		-0.017*			
		(0.010)			
FRE			0.004***		
			(0.001)		
Female×FRE			-0.009***		
			(0.003)		
Dale				0.053***	
				(0.017)	
Female×Dale				-0.057	
				(0.041)	
SMOG					0.017**
					(0.008)
Female×SMOG					-0.039**
					(0.019)
Concentration	0.254***	0.254***	0.254***	0.254***	0.254***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Constant	12.343***	12.317***	12.039***	12.939***	12.480***
	(0.356)	(0.356)	(0.364)	(0.433)	(0.368)
Observations	430,356	430,356	430,356	430,356	430,356
Controls	Yes	Yes	Yes	Yes	Yes
FE	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.121	0.121	0.121	0.121	0.121

3.6 Additional analyses

3.6.1 Time trend of readability

Given the higher standard females must meet when being evaluated, women exhibit a superior learning curve to adjust writing style both proactively and gradually (Hengel, 2020). To explore whether female analysts improve their writing ability over time, I include interaction terms between *Female* and the number of years since the analyst's first report date in my sample (*Career*). Results are reported in Table 3.6 Panel A. The positive interaction suggests that female analysts start their career with similar writing ability compared with male analysts but their writing quality improves over time. However, female analyst report length does not significantly change with experience.

Table 3.5: Additional tests

Panel A. Time trend of analyst report readability and effort

VARIABLES	(1) Fog	(2) FKGL	(3) FRE	(4) Dale	(5) SMOG	(6) Word	(7) Page
Female	-0.050	-0.062	0.117	0.001	0.036	-0.035*	-0.038**
	(0.095)	(0.086)	(0.277)	(0.017)	(0.046)	(0.020)	(0.020)
Career	0.008*	0.009*	0.035***	0.004***	0.006***	-0.000	-0.000
	(0.005)	(0.004)	(0.012)	(0.001)	(0.002)	(0.001)	(0.001)
Female×Career	0.036***	0.036***	0.070**	0.005*	0.009*	-0.001	-0.001
	(0.013)	(0.012)	(0.034)	(0.003)	(0.005)	(0.003)	(0.002)

Panel B. Readability and earnings announcement

VARIABLES	(1) Fog	(2) FKGL	(3) FRE	(4) Dale	(5) SMOG
Female	0.196**	0.176**	0.749***	0.035***	0.131***
	(0.077)	(0.071)	(0.255)	(0.013)	(0.042)
EA	0.187***	0.156***	1.117***	0.044***	0.138***
	(0.023)	(0.020)	(0.051)	(0.005)	(0.008)
Female×EA	-0.123*	-0.102*	-0.551***	-0.016	-0.102***
	(0.065)	(0.059)	(0.196)	(0.010)	(0.032)

Panel C. Report length distribution across firm

VARIABLES	(1) HHI
Female	-0.080
	(0.757)
Constant	67.095***
	(1.100)
Observations	19,641
Analyst controls	Yes
Year FE	Yes
Industry FE	Yes
Broker FE	Yes
Adjusted R^2	0.456

3.6.2 Analyst report around earnings announcement

Previous studies argue that forecasts not issued around earnings announcement dates are more likely to based on independent research instead of earnings news (Green et al., 2009). To examine this in the context of gender, I include an indicator which is equal to 1 if a report is issued within two days of an earnings announcement (*EA*). If female analysts have better writing skill, the difference should be larger for more independent research reports. In Table 3.5 Panel B, I find that reports issued around earnings announcement dates are more readable. Moreover, the advantage of female report readability is weaker for earnings-driven reports (in all but the *Dale* readability measure), consistent with my prediction.

3.6.3 Length dispersion

Female analysts may choose to distribute their effort evenly across firms which is likely observed in firm-specific variations in report length. To examine this possibility, I create a Herfindahl–Hirschman Index (HHI) for analyst report length. Specifically, I first calculate the average number of words for all firms an analyst covers in each year and then calculate the HHI at analyst-year level based on the average number of words for each firm. Table 3.5 Panel C shows OLS regression results at analyst-year level. No gender difference in report length dispersion is observed.

3.6.4 Text-recommendation consistency

Previous studies find that analysts may issue inconsistent stock recommendations and earnings forecasts to balance the interests of the firm and the investors (Malmendier and Shanthikumar, 2014). If female analysts are less influenced by the conflict of interest, I expect to observe that female analysts issue more consistent reports. I construct a dummy variable—

Consistency—which is equal to 1 if both recommendation and report tone are above or below the mean of a report. I find no gender difference in this text-recommendation measure.

3.6.5 Propensity score matching

Because analyst characteristics are significantly different between female and male analysts, I conduct a one-to-one propensity score matching with replacement. I re-run all analyses of readability, tone, topics, and market reaction for matched sample and results largely hold.

3.7 Concluding remarks

Motivated by existing evidence of gender differences in analyst quantitative outputs, I compare the textual characteristics of analyst reports between female and male analysts. Controlling for quantitative measures including earnings forecasts, stock recommendations, and price targets, I find female analysts issue more readable reports and improve report readability over time relative to male counterparts. However, female analyst reports are shorter, consistent with a "quality over quantity" approach. The textual sentiment of female analyst reports is also less optimistic, suggesting that they are more resistant to conflicts of interest common in sell-side analysis. Moreover, female analyst reports contain less financial content and are more long-term oriented. Female analysts do not benefit from high report readability in terms of market influence.

Overall, my findings contribute to a better understanding of how gender differences in writing abilities and gender stereotyping collectively affect gender differences in analyst report text characteristics and market reaction. Future research may attempt to explore how female and male analysts consider various topics when compiling their reports.

REFERENCES

Adams, R. B., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, *94*(2), 291–309. https://doi.org/10.1016/j.jfineco.2008.10.007

Adams, R. B., & Funk, P. (2012). Beyond the glass ceiling: Does gender matter? *Management Science*, *58*(2), 219–235. https://doi.org/10.1287/mnsc.1110.1452

Alexa. (n.d.). *Wikipedia.org competitive analysis, marketing mix and traffic*. Retrieved June 25, 2020, from http://www.alexa.com/siteinfo/wikipedia.org

Ali, H. (2014, August 11). *Brits trust Wikipedia more than the news: Survey*. CNBC. Retrieved June 25, 2020, from https://www.cnbc.com/2014/08/11/brits-trust-wikipedia-more-than-the-news-survey.html

Allee, K. D., & Deangelis, M. D. (2015). The structure of voluntary disclosure narratives: Evidence from tone dispersion. *Journal of Accounting Research*, *53*(2), 241–274. https://doi.org/10.1111/1475-679X.12072

Alsop, R. (2004, September 22). *How to get hired*. Wall Street Journal. Retrieved June 25, 2020, from https://www.wsj.com/articles/SB109577501492723498?ns=prod/accounts-wsj

Alter, A. L., & Oppenheimer, D. M. (2006). Predicting short-term stock fluctuations by using processing fluency. *Proceedings of the National Academy of Sciences*, *103*(24), 9369–9372. https://doi.org/10.1073/pnas.0601071103

Altınkılıç, O., & Hansen, R. S. (2009). On the information role of stock recommendation revisions. *Journal of Accounting and Economics*, *48*(1), 17–36. https://doi.org/10.1016/j.jacceco.2009.04.005

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

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text & Talk*, *23*(3), 321–346. https://doi.org/10.1515/text.2003.014

Asquith, P., Mikhail, M. B., & Au, A. S. (2005). Information content of equity analyst reports. *Journal of Financial Economics*, *75*(2), 245–282. https://doi.org/10.1016/j.jfineco.2004.01.002

Bajo, E., & Raimondo, C. (2017). Media sentiment and IPO underpricing. *Journal of Corporate Finance*, *46*, 139–153. https://doi.org/10.1016/j.jcorpfin.2017.06.003

Barber, B., Lehavy, R., McNichols, M., & Trueman, B. (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns. *The Journal of Finance*, *56*(2), 531–563.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, *116*(1), 261–292.

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, *21*(2), 785–818. https://doi.org/10.1093/rfs/hhm079

Baron, D. P. (1982). A model of the demand for investment banking advising and distribution services for new issues. *The Journal of Finance*, *37*(4), 955–976. https://doi.org/10.1111/j.1540-6261.1982.tb03591.x

Basow, S. A., & Rubenfeld, K. (2003). “troubles talk”: Effects of gender and gender-typing. *Sex Roles*, *48*(3–4), 183–187.

Ben-Rephael, A., Da, Z., & Israelsen, R. D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *The Review of Financial Studies*, *30*(9), 3009–3047. https://doi.org/10.1093/rfs/hhx031

Benveniste, L. M., & Spindt, P. A. (1989). How investment bankers determine the offer price and allocation of new issues. *Journal of Financial Economics*, *24*(2), 343–361.

Bertrand, M. (2018). The glass ceiling. *Economica*, *85*(338), 205–231. https://doi.org/10.1111/ecca.12264

Bertrand, M., Black, S. E., Jensen, S., & Lleras-Muney, A. (2019). Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway. *The Review of Economic Studies*, *86*(1), 191–239. https://doi.org/10.1093/restud/rdy032

Bertrand, M., Chugh, D., & Mullainathan, S. (2005). Implicit discrimination. *American Economic Review Papers and Proceedings*, *95*(2), 94–98. https://doi.org/10.1257/000282805774670365

Bertrand, M., & Duflo, E. (2017). Field experiments on discrimination. In *Handbook of field experiments* (Vol. 1, pp. 309–393). North-Holland. https://doi.org/10.3386/w22014

Bertrand, M., & Hallock, K. F. (2001). The gender gap in top corporate jobs. *ILR Review*, *55*(1), 3–21.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, *94*(4), 991–1013.

Bhattacharya, U., Galpin, N., Ray, R., & Yu, X. (2009). The role of the media in the internet IPO bubble. *The Journal of Financial and Quantitative Analysis*, *44*(3), 657–682.

Blankespoor, E., Miller, G. S., & White, H. D. (2014). The role of dissemination in market liquidity: Evidence from firms’ use of Twitter™. *The Accounting Review*, *89*(1), 79–112. https://doi.org/10.2308/accr-50576

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, *55*(3), 789–865. https://doi.org/10.1257/jel.20160995

Bloomfield, R., Rennekamp, K., Steenhoven, B., & Stewart, S. (in press). Penalties for unexpected behavior: Double standards for women in finance. *The Accounting Review*. https://doi-org.libproxy.rpi.edu/10.2308/tar-2018-0715

Bochkay, K., Hales, J., & Chava, S. (2020). Hyperbole or reality? Investor response to extreme language in earnings conference calls. *The Accounting Review*, *95*(2), 31–60. https://doi.org/10.2308/accr-52507

Boehmer, E., & Fishe, R. P. H. (2000). *Do underwriters encourage stock flipping? A new explanation for the underpricing of IPOs*. SSRN Electronic Journal. Retrieved June 25, 2020, from http://www.ssrn.com/abstract=228434

Bonham, K. S., & Stefan, M. I. (2017). Women are underrepresented in computational biology: An analysis of the scholarly literature in biology, computer science and computational biology. *PLOS Computational Biology*, *13*(10), 1–12. https://doi.org/10.1371/journal.pcbi.1005134

Boorstin, J. (2018, June 26). *After Time's Up and #MeToo, women in entertainment still see gender issues that men don't*. CNBC. Retrieved June 25, 2020, from https://www.cnbc.com/2018/06/25/surveyon-wall-street-workplace-biases-persist---but-men-dont-see-t.html

Bosquet, K., de Goeij, P., & Smedts, K. (2014). Gender heterogeneity in the sell-side analyst recommendation issuing process. *Finance Research Letters*, *11*(2), 104–111. https://doi.org/10.1016/j.frl.2013.11.004

Bowen, R. M., Davis, A. K., & Matsumoto, D. A. (2002). Do conference calls affect analysts' forecasts? *The Accounting Review*, *77*(2), 285–316.

Bradley, D., Clarke, J., Lee, S., & Ornthanalai, C. (2014). Are analysts' recommendations informative? Intraday evidence on the impact of time stamp delays. *The Journal of Finance*, *69*(2), 645–673. https://doi.org/10.1111/jofi.12107

Bradley, D., Gokkaya, S., & Liu, X. (2017). Before an analyst becomes an analyst: Does industry experience matter? *The Journal of Finance*, *72*(2), 751–792. https://doi.org/10.1111/jofi.12466

Bradley, D. J., & Jordan, B. D. (2002). Partial adjustment to public information and IPO underpricing. *The Journal of Financial and Quantitative Analysis*, *37*(4), 595–616. https://doi.org/10.2307/3595013

Bradley, D. J., Jordan, B. D., & Ritter, J. R. (2003). The quiet period goes out with a bang. *The Journal of Finance*, *58*(1), 1–36. https://doi.org/10.1111/1540-6261.00517

Bradley, D. J., Jr., J. W. C., Jordan, B. D., & Singh, A. K. (2004). Negotiation and the IPO offer price: A comparison of integer vs. Non-integer IPOs. *The Journal of Financial and Quantitative Analysis*, *39*(3), 517–540.

Bradshaw, M. T., Lee, L. F., & Peterson, K. (2016). The interactive role of difficulty and incentives in explaining the annual earnings forecast walkdown. *The Accounting Review*, *91*(4), 995–1021. https://doi.org/10.2308/accr-51398

Bradshaw, T. (2008). *Companies woo investors via social websites*. Financial Times. Retrieved June 25, 2020, from https://www.ft.com/content/c9d0271a-bf49-11dd-ae63-0000779fd18c

Brav, A., & Lehavy, R. (2003). An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics. *The Journal of Finance*, *58*(5), 1933–1967. https://doi.org/10.1111/1540-6261.00593

Brochet, F., Kolev, K., & Lerman, A. (2018). Information transfer and conference calls. *Review of Accounting Studies*, *23*(3), 907–957. https://doi.org/10.1007/s11142-018-9444-4

Brochet, F., Loumioti, M., & Serafeim, G. (2015). Speaking of the short-term: Disclosure horizon and managerial myopia. *Review of Accounting Studies*, *20*(3), 1122–1163. https://doi.org/10.1007/s11142-015-9329-8

Brown, A. R. (2011). Wikipedia as a data source for political scientists: Accuracy and completeness of coverage. *PS: Political Science & Politics*, *44*(02), 339–343. https://doi.org/10.1017/S1049096511000199

Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2015). Inside the "black box" of sell-side financial analysts. *Journal of Accounting Research*, *53*(1), 1–47. https://doi.org/10.1111/1475-679X.12067

Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2019). Managing the narrative: Investor relations officers and corporate disclosure. *Journal of Accounting and Economics*, *67*(1), 58–79. https://doi.org/10.1016/j.jacceco.2018.08.014

Brown, S., & Hillegeist, S. A. (2007). How disclosure quality affects the level of information asymmetry. *Review of Accounting Studies*, *12*(2–3), 443–477. https://doi.org/10.1007/s11142-007-9032-5

Brunner, F., & Ungeheuer, M. (2020). *Information, trade, and salient returns*. SSRN Electronic Journal. Retrieved June 25, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931547

Bushee, B., Cedergren, M., & Michels, J. (2020). Does the media help or hurt retail investors during the IPO quiet period? *Journal of Accounting and Economics*, *69*, 101261. https://doi.org/10.1016/j.jacceco.2019.101261

Bushee, B. J., Gow, I. D., & Taylor, D. J. (2018). Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research*, *56*(1), 85–121. https://doi.org/10.1111/1475-679X.12179

Bushee, B. J., Matsumoto, D. A., & Miller, G. S. (2004). Managerial and investor responses to disclosure regulation: The case of Reg FD and conference calls. *The Accounting Review*, *79*(3), 617–643.

Call, A. C., Sharp, N. Y., & Shohfi, T. D. (2018). *Which buy-side institutions participate in public earnings conference calls? Implications for capital markets and sell-side coverage*. SSRN Electronic Journal. Retrieved June 25, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2731930

Cao, Y., Myers, L. A., & Omer, T. C. (2012). Does company reputation matter for financial reporting quality? Evidence from restatements. *Contemporary Accounting Research*, *29*(3), 956–990. https://doi.org/10.1111/j.1911-3846.2011.01137.x

Caylor, M., Cecchini, M., & Winchel, J. (2017). Analysts' qualitative statements and the profitability of favorable investment recommendations. *Accounting, Organizations and Society*, *57*, 33–51. https://doi.org/10.1016/j.aos.2017.03.005

Cen, L., Chen, J., Dasgupta, S., & Ragunathan, V. (in press). Do analysts and their employers value access to management? Evidence from earnings conference call participation. *Journal of Financial and Quantitative Analysis*. https://doi.org/10.1017/S0022109020000198

Chang, T. Y., Solomon, D. H., & Westerfield, M. M. (2016). Looking for someone to blame: Delegation, cognitive dissonance, and the disposition effect. *The Journal of Finance*, *71*(1), 267–302. https://doi.org/10.1111/jofi.12311

Chemmanur, T. J., Krishnan, K., & Yu, Q. (2018). *Venture capital backing, investor attention, and initial public offerings*. SSRN Electronic Journal. Retrieved June 25, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2851196

Chen, H., De, P., Hu, Y. (Jeffrey), & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*(5), 1367–1403. https://doi.org/10.1093/rfs/hhu001

Chen, J., Demers, E., & Lev, B. (2018). Oh what a beautiful morning! Diurnal influences on executives and analysts: Evidence from conference calls. *Management Science*, *64*(12), 5899–5924. https://doi.org/10.1287/mnsc.2017.2888

Chen, J. V., Nagar, V., & Schoenfeld, J. (2018). Manager-analyst conversations in earnings conference calls. *Review of Accounting Studies*, *23*(4), 1315–1354. https://doi.org/10.1007/s11142-018-9453-3

Chen, J. Z., Shane, P. B., Yang, L. L., & Zhang, J. H. (2017). *Financial analysts' long-term growth forecasts and market and analyst efficiency with respect to firms' innovative efficiency*. SSRN Electronic Journal. Retrieved June 25, 2020, from https://sci-hub.st/https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2847334

Chen, S., Jung, J. H., Lim, S. S., & Yu, Y. (2020). *Analysts' cultural attitudes to time orientation*. SSRN Electronic Journal. Retrieved June 25, 2020, from https://www.ssrn.com/abstract=3551566

Chen, X., Cheng, Q., & Lo, K. (2010). On the relationship between analyst reports and corporate disclosures: Exploring the roles of information discovery and interpretation. *Journal of*

Accounting and Economics, *49*(3), 206–226. https://doi.org/10.1016/j.jacceco.2009.12.004

Clauson, K. A., Polen, H. H., Boulos, M. N. K., & Dzenowagis, J. H. (2008). Scope, completeness, and accuracy of drug information in Wikipedia. *Annals of Pharmacotherapy*, *42*(12), 1814–1821. https://doi.org/10.1345/aph.1L474

Clement, M. B. (1999). Analyst forecast accuracy: Do ability, resources and portfolio complexity matter? *Journal of Accounting and Economics*, *27*(3), 285–303.

Clement, M. B., & Tse, S. Y. (2003). Do investors respond to analysts' forecast revisions as if forecast accuracy is all that matters? *The Accounting Review*, *78*(1), 227–249. https://doi.org/10.2308/accr.2003.78.1.227

Cohen, L., Lou, D., & Malloy, C. J. (in press). Casting conference calls. *Management Science*. https://doi.org/10.1287/mnsc.2019.3423

Cohen, N. (2014, February 10). *Wikipedia vs. The small screen*. The New York Times. Retrieved June 25, 2020, from https://www.nytimes.com/2014/02/10/technology/wikipedia-vs-the-small-screen.html

Cowen, A., Groysberg, B., & Healy, P. (2006). Which types of analyst firms are more optimistic? *Journal of Accounting and Economics*, *41*(1–2), 119–146. https://doi.org/10.1016/j.jacceco.2005.09.001

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448–474.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Dambra, M., Wasley, C. E., & Wu, J. S. (2013). Soft-talk management cash flow forecasts: Bias, quality, and stock price effects. *Contemporary Accounting Research*, *30*(2), 607–644. https://doi.org/10.1111/j.1911-3846.2012.01167.x

Davis, A. K., Ge, W., Matsumoto, D., & Zhang, J. L. (2015). The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies*, *20*(2), 639–673. https://doi.org/10.1007/s11142-014-9309-4

De Franco, G., Hope, O.-K., Vyas, D., & Zhou, Y. (2015). Analyst report readability. *Contemporary Accounting Research*, *32*(1), 76–104. https://doi.org/10.1111/1911-3846.12062

Devgan, L., Powe, N., Blakey, B., & Makary, M. (2007). Wiki-surgery? Internal validity of Wikipedia as a medical and surgical reference. *Journal of the American College of Surgeons*, *205*(3), S76–S77. https://doi.org/10.1016/j.jamcollsurg.2007.06.190

Dhaliwal, D. S., Li, O. Z., Tsang, A., & Yang, Y. G. (2011). Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *The Accounting Review*, *86*(1), 59–100. https://doi.org/10.2308/accr.00000005

Dollar, D., Fisman, R., & Gatti, R. (2001). Are women really the “fairer” sex? Corruption and women in government. *Journal of Economic Behavior & Organization*, *46*(4), 423–429. https://doi.org/10.1016/S0167-2681(01)00169-X

Egan, M. L., Matvos, G., & Seru, A. (2018). *When Harry fired Sally: The double standard in punishing misconduct*. NBER. Retrieved June 25, 2020, from https://www.nber.org/papers/w23242

Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, *66*(1), 67–97. https://doi.org/10.1111/j.1540-6261.2010.01626.x

Faccio, M., Marchica, M.-T., & Mura, R. (2016). CEO gender, corporate risk-taking, and the efficiency of capital allocation. *Journal of Corporate Finance*, *39*, 193–209. https://doi.org/10.1016/j.jcorpfin.2016.02.008

Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, *43*(2), 153–193.

Fang, B., & Hope, O.-K. (2020). *Analyst teams*. SSRN Electronic Journal. Retrieved June 25, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3187324

Fang, L. H., & Huang, S. (2017). Gender and connections among Wall Street analysts. *The Review of Financial Studies*, *30*(9), 3305–3335. https://doi.org/10.1093/rfs/hhx040

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, *64*(5), 2023–2052. https://doi.org/10.1111/j.1540-6261.2009.01493.x

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, *43*(2), 95–103.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Fitzpatrick, A., Eadicicco, L., & Peckham, M. (2017, October 20). *The 15 most influential websites of all time*. Time. Retrieved June 25, 2020, from http://time.com/4960202/most-influential-websites/

Francis, B., Hasan, I., Park, J. C., & Wu, Q. (2015). Gender differences in financial reporting decision making: Evidence from accounting conservatism. *Contemporary Accounting Research*, *32*(3), 1285–1318. https://doi.org/10.1111/1911-3846.12098

Francis, B., Shohfi, T., & Xin, D. (2020). *Gender and earnings conference calls*. SSRN Electronic Journal. Retrieved June 25, 2020, from https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3473266

Franke, G. R., Crown, D. F., & Spake, D. F. (1997). Gender differences in ethical perceptions of business practices: A social role theory perspective. *Journal of Applied Psychology*, *82*(6), 920–934.

Frankel, R., Johnson, M., & Skinner, D. J. (1999). An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research*, *37*(1), 133–150.

Frederickson, J. R., & Zolotoy, L. (2016). Competing earnings announcements: Which announcement do investors process first? *The Accounting Review*, *91*(2), 441–462.

Frewin, J. (2010, June 15). *Wikipedia unlocks divisive pages for editing*. BBC. Retrieved June 25, 2020, from https://www.bbc.com/news/10312095

Gao, X., & Ritter, J. R. (2010). The marketing of seasoned equity offerings. *Journal of Financial Economics*, *97*(1), 33–52.

Gelb, D. S., & Strawser, J. A. (2001). Corporate social responsibility and financial disclosures: An alternative explanation for increased disclosure. *Journal of Business Ethics*, *33*(1), 1–13.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*, 900–901.

Gleason, C. A., & Lee, C. M. C. (2003). Analyst forecast revisions and market price discovery. *The Accounting Review*, *78*(1), 193–225. https://doi.org/10.2308/accr.2003.78.1.193

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, *118*(3), 1049–1074. https://doi.org/10.1162/00335530360698496

Gneezy, Uri, & Rustichini, A. (2004). Gender and competition at a young age. *The American Economic Review*, *94*(2), 377–381.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*(4), 1091–1119. https://doi.org/10.1257/aer.104.4.1091

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of “blind” auditions on female musicians. *American Economic Review*, *90*(4), 715–741.

Goodwin, D. (2012, February 13). *Wikipedia appears on Page 1 of Google for 99% of searches [study]*. Search Engine Watch. Retrieved June 25, 2020, from https://www.searchenginewatch.com/2012/02/13/wikipedia-appears-on-page-1-of-google-for-99-of-searches-study/

Gow, I. D., Larcker, D. F., & Zakolyukina, A. A. (2019). *Non-answers during conference calls*. SSRN Electronic Journal. Retrieved June 25, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3310360

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, *40*(1–3), 3–73. https://doi.org/10.1016/j.jacceco.2005.01.002

Green, C., Jegadeesh, N., & Tang, Y. (2009). Gender and job performance: Evidence from Wall Street. *Financial Analysts Journal*, *65*(6), 65–78.

Green, T. C., Jame, R., Markov, S., & Subasi, M. (2014). Access to management and the informativeness of analyst research. *Journal of Financial Economics*, *114*(2), 239–255. https://doi.org/10.1016/j.jfineco.2014.07.003

Greenstein, S., & Zhu, F. (2012). Is Wikipedia biased? *American Economic Review*, *102*(3), 343–348. https://doi.org/10.1257/aer.102.3.343

Greenstein, S., & Zhu, F. (2018). Do experts or crowd-based models produce more bias? Evidence from Encyclopædia Britannica and Wikipedia. *MIS Quarterly*, *42*(3), 945–959.

Groysberg, B., Healy, P. M., & Maber, D. A. (2011). What drives sell-side analyst compensation at high-status investment banks? *Journal of Accounting Research*, *49*(4), 969–1000. https://doi.org/10.1111/j.1475-679X.2011.00417.x

Gu, B., Konana, P., Rajagopalan, B., & Chen, H.-W. M. (2007). Competition among virtual communities and user valuation: The case of investing-related communities. *Information Systems Research*, *18*(1), 68–85. https://doi.org/10.1287/isre.1070.0114

Guryan, J., & Charles, K. K. (2013). Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *The Economic Journal*, *123*(572), F417–F432.

Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review*, *9*(2), 193–206.

Hanley, K. W. (1993). The underpricing of initial public offerings and the partial adjustment phenomenon. *Journal of Financial Economics*, *34*, 231–250.

Hanley, K. W., & Hoberg, G. (2010). The information content of IPO prospectuses. *Review of Financial Studies*, *23*(7), 2821–2864. https://doi.org/10.1093/rfs/hhq024

He, J. (Jack), & Tian, X. (2013). The dark side of analyst coverage: The case of innovation. *Journal of Financial Economics*, *109*(3), 856–878. https://doi.org/10.1016/j.jfineco.2013.04.001

Hebert, C. (2020). *Gender stereotypes and entrepreneur financing*. SSRN Electronic Journal. Retrieved June 25, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3318245

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, *46*(4), 931–959. https://doi.org/10.2307/1909757

Hengel, E. (2020). *Publishing while female. Are women held to higher standards? Evidence from peer review* [Working Paper]. Retrieved June 25, 2020, from http://www.erinhengel.com/research/publishing_female.pdf

Hirschman, L. (1994). Female-male differences in conversational interaction. *Language in Society*, *23*(3), 427–442. https://doi.org/10.1017/S0047404500018054

Hirshleifer, D., Kewei Hou, Teoh, S. H., & Yinglei Zhang. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, *38*, 297–331. https://doi.org/10.1016/j.jacceco.2004.10.002

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*(1–3), 337–386. https://doi.org/10.1016/j.jacceco.2003.10.002

Hollander, S., Pronk, M., & Roelofsen, E. (2010). Does silence speak? An empirical analysis of disclosure choices during conference calls. *Journal of Accounting Research*, *48*(3), 531–563. https://doi.org/10.1111/j.1475-679X.2010.00365.x

Holman Rector, L. (2008). Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review*, *36*(1), 7–22. https://doi.org/10.1108/00907320810851998

Hong, H., & Kubik, J. D. (2003). Analyzing the analysts: Career concerns and biased earnings forecasts. *The Journal of Finance*, *58*(1), 313–351. https://doi.org/10.1111/1540-6261.00526

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, *54*(6), 2143–2184.

Hong, H., Torous, W., & Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics*, *83*(2), 367–396.

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*, *64*(6), 2833–2855. https://doi.org/10.1287/mnsc.2017.2751

Huang, A. H., Zang, A. Y., & Zheng, R. (2014). Evidence on the information content of text in analyst reports. *The Accounting Review*, *89*(6), 2151–2180. https://doi.org/10.2308/accr-50833

Huang, J., & Kisgen, D. J. (2013). Gender and corporate finance: Are male executives overconfident relative to female executives? *Journal of Financial Economics*, *108*(3), 822–839. https://doi.org/10.1016/j.jfineco.2012.12.005

Huberman, G. (2001). Familiarity breeds investment. *The Review of Financial Studies*, *14*(3), 659–680.

Hutton, A. P., Miller, G. S., & Skinner, D. J. (2003). The role of supplementary statements with management earnings forecasts. *Journal of Accounting Research*, *41*(5), 867–890. https://doi.org/10.1046/j.1475-679X.2003.00126.x

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592. https://doi.org/10.1037/0003-066X.60.6.581

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, *65*(1), 373–398. https://doi.org/10.1146/annurev-psych-010213-115057

Ibbotson, R. G. (1975). Price performance of common stock new issues. *Journal of Financial Economics*, *2*(3), 235–272. https://doi.org/10.1016/0304-405X(75)90015-X

Ittner, C. D., & Larcker, D. F. (1998). Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research*, *36*, 1–35. https://doi.org/10.2307/2491304

Ittonen, K., Vähämaa, E., & Vähämaa, S. (2013). Female auditors and accruals quality. *Accounting Horizons*, *27*(2), 205–228. https://doi.org/10.2308/acch-50400

Ivković, Z., & Jegadeesh, N. (2004). The timing and value of forecast and recommendation revisions. *Journal of Financial Economics*, *73*(3), 433–463. https://doi.org/10.1016/j.jfineco.2004.03.002

Jacobi, T., & Schweers, D. (2017). Justice, interrupted: The effect of gender, ideology, and seniority at supreme court oral arguments. *Virginia Law Review*, *103*(7), 1379–1485.

Jannati, S., Kumar, A., Niessen-Ruenzi, A., & Wolfers, J. (2020). *In-group bias in financial markets*. SSRN Electronic Journal. Retrieved June 25, 2020, from http://www.ssrn.com/abstract=2884218

Jeong, S.-H., & Harrison, D. A. (2017). Glass breaking, strategy making, and value creating: Meta-analytic outcomes of women as CEOs and TMT members. *Academy of Management Journal*, *60*(4), 1219–1252. https://doi.org/10.5465/amj.2014.0716

Jiang, D., Kumar, A., & Law, K. K. F. (2016). Political contributions and analyst behavior. *Review of Accounting Studies*, *21*(1), 37–88. https://doi.org/10.1007/s11142-015-9344-9

Jiao, P., Veiga, A., & Walther, A. (2020). Social media, news media and the stock market. *Journal of Economic Behavior & Organization*, *176*, 63–90. https://doi.org/10.1016/j.jebo.2020.03.002

Jog, V., & McConomy, B. J. (2003). Voluntary disclosure of management earnings forecasts in IPO prospectuses. *Journal of Business Finance & Accounting*, *30*(1–2), 125–168. https://doi.org/10.1111/1468-5957.00486

Johnson, J. E., & Powell, P. L. (1994). Decision making, risk and gender: Are managers different? *British Journal of Management*, *5*(2), 123–138.

Jung, B., Shane, P. B., & Yang, Y. S. (2012). Do financial analysts' long-term growth forecasts matter? Evidence from stock recommendations and career outcomes. *Journal of Accounting and Economics*, *53*(1–2), 55–76. https://doi.org/10.1016/j.jacceco.2011.11.002

Jung, J. H., Kumar, A., Lim, S. S., & Yoo, C.-Y. (2019). An analyst by any other surname: Surname favorability and market reaction to analyst forecasts. *Journal of Accounting and Economics*, *67*(2–3), 306–335. https://doi.org/10.1016/j.jacceco.2019.02.002

Kaplan, R. S. (1984). The evolution of management accounting. *The Accounting Review*, *59*(3), 586–621. https://doi.org/10.1007/978-1-4899-7138-8_27

Ke, B., & Yu, Y. (2006). The effect of issuing biased earnings forecasts on analysts' access to management and survival. *Journal of Accounting Research*, *44*(5), 965–999. https://doi.org/10.1111/j.1475-679X.2006.00221.x

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

Kmec, J. A., & Gorman, E. H. (2010). Gender and discretionary work effort: Evidence from the United States and Britain. *Work and Occupations*, *37*(1), 3–36. https://doi.org/10.1177/0730888409352064

Kothari, S. P., Shu, S., & Wysocki, P. D. (2009). Do managers withhold bad news? *Journal of Accounting Research*, *47*(1), 241–276. https://doi.org/10.1111/j.1475-679X.2008.00318.x

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of Financial Economics*, *115*(2), 304–329. https://doi.org/10.1016/j.jfineco.2014.09.008

Kumar, A. (2010). Self-selection and the forecasting abilities of female equity analysts. *Journal of Accounting Research*, *48*(2), 393–435. https://doi.org/10.1111/j.1475-679X.2009.00362.x

Kumar, S., West, R., & Leskovec, J. (2016). Disinformation on the web: Impact, characteristics, and detection of Wikipedia hoaxes. *Proceedings of the 25th International Conference on World Wide Web*, 591–602. https://doi.org/10.1145/2872427.2883085

Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls: Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, *50*(2), 495–540. https://doi.org/10.1111/j.1475-679X.2012.00450.x

Lay, C. H., & Paivio, A. (1969). The effects of task difficulty and anxiety on hesitations in speech. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *1*(1), 25–37. https://doi.org/10.1037/h0082683

Leaper, C. (1991). Influence and involvement in children's discourse: Age, gender, and partner effects. *Child Development*, *62*(4), 797–811.

Lehavy, R., Li, F., & Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review*, *86*(3), 1087–1115. https://doi.org/10.2308/accr.00000043

Leone, A. J., Rock, S., & Willenborg, M. (2007). Disclosure of intended use of proceeds and underpricing in initial public offerings. *Journal of Accounting Research*, *45*(1), 111–153. https://doi.org/10.1111/j.1475-679X.2007.00229.x

Li, E. X., Ramesh, K., Shen, M., & Wu, J. S. (2015). Do analyst stock recommendations piggyback on recent corporate news? An analysis of regular-hour and after-hours revisions. *Journal of Accounting Research*, *53*(4), 821–861. https://doi.org/10.1111/1475-679X.12083

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, *45*(2–3), 221–247.

Li, X., Sullivan, R. N., Xu, D., & Gao, G. (2013). Sell-side analysts and gender: A comparison of performance, behavior, and career outcomes. *Financial Analysts Journal*, *69*(2), 83–94.

Lin, H., & McNichols, M. F. (1998). Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics*, *25*(1), 101–127. https://doi.org/10.1016/S0165-4101(98)00016-0

Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *The Journal of Finance*, *72*(4), 1785–1824. https://doi.org/10.1111/jofi.12505

Liu, L. X., Sherman, A. E., & Zhang, Y. (2014). The long-run role of the media: Evidence from initial public offerings. *Management Science*, *60*(8), 1945–1964. https://doi.org/10.1287/mnsc.2013.1851

Ljungqvist, A. (2007). IPO underpricing. In B. E. Eckbo (Ed.), *Handbooks in Finance: Empirical Corporate Finance* (pp. 375–422). Elsevier.

Lo, K., Ramos, F., & Rogo, R. (2017). Earnings management and annual report readability. *Journal of Accounting and Economics*, *63*(1), 1–25. https://doi.org/10.1016/j.jacceco.2016.09.002

Logue, D. E. (1973). On the pricing of unseasoned equity issues: 1965-1969. *The Journal of Financial and Quantitative Analysis*, *8*(1), 91. https://doi.org/10.2307/2329751

Loughran, T., & Mcdonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, *109*(2), 307–326. https://doi.org/10.1016/j.jfineco.2013.02.017

Loughran, T., & Mcdonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187–1230. https://doi.org/10.1111/1475-679X.12123

Loughran, T., & Ritter, J. (2004). Why has IPO underpricing changed over time? *Financial Management*, *33*(3), 5–37.

Lovell, S. E., Kahn, A. S., Anton, J., Amanda, D., Dowling, E., Post, D., & Mason, C. (1999). Does gender affect the link between organizational citizenship behavior and performance evaluation? *Sex Roles*, *41*(5–6), 469–478.

Luo, Y., & Salterio, S. E. (2020). *The effect of gender on investors' judgements and decision making*. SSRN Electronic Journal. Retrieved June 25, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3036218

Madera, J. M., Hebl, M. R., Dial, H., Martin, R., & Valian, V. (2019). Raising doubt in letters of recommendation for academia: Gender differences and their impact. *Journal of Business and Psychology*, *34*(3), 287–303. https://doi.org/10.1007/s10869-018-9541-1

Malmendier, U., & Shanthikumar, D. (2007). Are small investors naive about incentives? *Journal of Financial Economics*, *85*(2), 457–489. https://doi.org/10.1016/j.jfineco.2007.02.001

Malmendier, U., & Shanthikumar, D. (2014). Do security analysts speak in two tongues? *Review of Financial Studies*, *27*(5), 1287–1322. https://doi.org/10.1093/rfs/hhu009

Maranz, F., & Greenfield, R. (2018, September 13). *Men get the first, last and every other word on earnings calls*. Bloomberg. Retrieved June 25, 2020, from https://www.bloomberg.com/news/articles/2018-09-13/men-get-the-first-last-and-every-other-word-on-earnings-calls

Marginson, D., & McAulay, L. (2008). Exploring the debate on short-termism: A theoretical and empirical analysis. *Strategic Management Journal*, *29*(3), 273–292. https://doi.org/10.1002/smj.657

Matsa, D. A., & Miller, A. R. (2011). Chipping away at the glass ceiling: Gender spillovers in corporate leadership. *American Economic Review*, *101*(3), 635–639.

Matsumoto, D., Pronk, M., & Roelofsen, E. (2011). What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *The Accounting Review*, *86*(4), 1383–1414. https://doi.org/10.2308/accr-10034

Mayew, W. J. (2008). Evidence of management discrimination among analysts during earnings conference calls. *Journal of Accounting Research*, *46*(3), 627–659. https://doi.org/10.1111/j.1475-679X.2008.00285.x

Mayew, W. J., Sharp, N. Y., & Venkatachalam, M. (2013). Using earnings conference calls to identify analysts with superior private information. *Review of Accounting Studies*, *18*(2), 386–413. https://doi.org/10.1007/s11142-012-9210-y

Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *The Journal of Finance*, *67*(1), 1–43. https://doi.org/10.1111/j.1540-6261.2011.01705.x

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, *42*(3), 483–510. https://doi.org/10.1111/j.1540-6261.1987.tb04565.x

Michaely, R., & Womack, K. L. (1999). Conflict of interest and the credibility of underwriter analyst recommendations. *The Review of Financial Studies*, *12*(4), 653–686.

Mikhail, M. B., Walther, B. R., & Willis, R. H. (1999). Does forecast accuracy matter to security analysts? *The Accounting Review*, *74*(2), 185–200.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, *122*(3), 1067–1101.

Niessen-Ruenzi, A., & Ruenzi, S. (2019). Sex matters: Gender bias in the mutual fund industry. *Management Science*, *65*(7), 3001–3025. https://doi.org/10.1287/mnsc.2017.2939

Odean, T. (1999). Do investors trade too much? *American Economic Review*, *89*(5), 1279–1298.

Page views for Wikipedia. (n.d.). In *Wikipedia*. Retrieved June 25, 2020, from https://stats.wikimedia.org/EN/TablesPageViewsMonthly.htm

Park, C. W., Eisingerich, A. B., Pol, G., & Park, J. W. (2013). The role of brand logos in firm performance. *Journal of Business Research*, *66*(2), 180–187. https://doi.org/10.1016/j.jbusres.2012.07.011

Petersen, J. (2018). Gender difference in verbal performance: A meta-analysis of United States state performance assessments. *Educational Psychology Review*, *30*(4), 1269–1281. https://doi.org/10.1007/s10648-018-9450-x

Petroni, K. R., Wang, I. Y., & Jiang, J. (2010). CFOs and CEOs: Who have the most influence on earnings management? *Journal of Financial Economics*, *96*(3), 513–526.

Previts, G. J., Brioker, R. J., Robinson, T. R., & Young, S. J. (1994). A content analysis of sell-side financial analyst company reports. *Accounting Horizons*, *8*(2), 55–70.

Pyramid: Women in S&P 500 companies. (2020). Catalyst. Retrieved June 25, 2020, from https://www.catalyst.org/research/women-in-sp-500-companies/

Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the national assessment of educational progress (NAEP). *American Psychologist*, *74*(4), 445–458. https://doi.org/10.1037/amp0000356

Reiss, M. C., & Mitra, K. (1998). The effects of individual difference factors on the acceptability of ethical and unethical workplace behaviors. *Journal of Business Ethics*, *17*(14), 1581–1593.

Reynolds, M. R., Scheiber, C., Hajovsky, D. B., Schwartz, B., & Kaufman, A. S. (2015). Gender differences in academic achievement: Is writing an exception to the gender similarities hypothesis? *The Journal of Genetic Psychology*, *176*(4), 211–234. https://doi.org/10.1080/00221325.2015.1036833

Richardson, S., Teoh, S. H., & Wysocki, P. D. (2004). The walk-down to beatable analyst forecasts: The role of equity issuance and insider trading incentives. *Contemporary Accounting Research*, *21*(4), 885–924. https://doi.org/10.1506/KHNW-PJYL-ADUB-0RP6

Ritter, J. R. (1987). The costs of going public. *Journal of Financial Economics*, *19*(2), 269–281.

Ritter, J. R., & Welch, I. (2002). A review of IPO activity, pricing, and allocations. *The Journal of Finance*, *57*(4), 1795–1828. https://doi.org/10.1111/1540-6261.00478

Rock, K. (1986). Why new issues are underpriced. *Journal of Financial Economics*, *15*(1–2), 187–212. https://doi.org/10.1016/0304-405X(86)90054-1

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology*, *87*(4), 494–509. https://doi.org/10.1037/0022-3514.87.4.494

Ruud, J. S. (1993). Underwriter price support and the IPO underpricing puzzle. *Journal of Financial Economics*, *34*(2), 135–151.

Santamaría, L., & Mihaljević, H. (2018). Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science*, *4*, e156.

Sarsons, H. (2017). *Interpreting signals in the labor market: Evidence from medical referrals* [Working Paper]. Retrieved June 25, 2020, from https://scholar.harvard.edu/sarsons/publications/interpreting-signals-evidence-medical-referrals

Saxton, G. D., & Anker, A. E. (2013). The aggregate effects of decentralized knowledge production: Financial bloggers and information asymmetries in the stock market. *Journal of Communication*, *63*(6), 1054–1069. https://doi.org/10.1111/jcom.12060

Scheiber, C., Reynolds, M. R., Hajovsky, D. B., & Kaufman, A. S. (2015). Gender differences in achievement in a large, nationally representative sample of children and adolescents: Gender and achievement. *Psychology in the Schools*, *52*(4), 335–348. https://doi.org/10.1002/pits.21827

SEC. (2017, May 18). *Quiet period*. Retrieved June 25, 2020, from https://www.sec.gov/fast-answers/answersquiethtm.html

SimilarWeb. (n.d.). *Wikipedia.org*. Retrieved July 2, 2019, from https://www.similarweb.com/website/wikipedia.org

Solomon, D. H. (2012). Selective publicity and stock prices. *The Journal of Finance*, *67*(2), 599–637.

Soltes, E. (2014). Private interaction between firm management and sell-side analysts. *Journal of Accounting Research*, *52*(1), 245–272. https://doi.org/10.1111/1475-679X.12037

Special:Statistics. (n.d.). in *Wikipedia*. Retrieved June 25, 2020, from https://en.wikipedia.org/wiki/Special:Statistics

Stickel, S. E. (1991). Common stock returns surrounding earnings forecast revisions: More puzzling evidence. *The Accounting Review*, *66*(2), 402–416.

Stickel, S. E. (1992). Reputation and performance among security analysts. *Journal of Finance*, *47*(5), 1811–1836.

Stocken, P. C., & Verrecchia, R. E. (2004). Financial reporting system choice and disclosure management. *The Accounting Review*, *79*(4), 1181–1203. https://doi.org/10.2308/accr.2004.79.4.1181

Suslava, K. (2017). *Stiff business headwinds and unchartered economic waters': The use of euphemisms in earnings conference calls*. SSRN Electronic Journal. Retrieved June 25, 2020, from http://www.ssrn.com/abstract=2876819

Tannen, D. (1990). *You just don't understand: Women and men in conversation*. William Morrow

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168.

Tsao, A. (2002). *When a stock's rating and target collide*. Bloomberg. Retrieved June 25, 2020, from https://www.bloomberg.com/news/articles/2002-04-24/when-a-stocks-rating-and-target-collide

Twedt, B., & Rees, L. (2012). Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports. *Journal of Accounting and Public Policy*, *31*(1), 1–21. https://doi.org/10.1016/j.jaccpubpol.2011.10.010

Verrecchia, R. E. (1983). Discretionary disclosure. *Journal of Accounting and Economics*, *5*, 179–194. https://doi.org/10.1016/0165-4101(83)90011-3

Waller, N. (2016, October 25). *Hunting for soft skills, companies scoop up English majors*. Wall Street Journal. Retrieved June 25, 2020, from https://www.wsj.com/articles/hunting-for-soft-skills-companies-scoop-up-english-majors-1477404061

Weber, L., & Cutter, C. (2019, May 10). *A wake-up call for grads: Entry-level jobs aren't so entry level any more*. Wall Street Journal. Retrieved June 25, 2020, from https://www.wsj.com/articles/a-wake-up-call-for-grads-entry-level-jobs-arent-so-entry-level-any-more-11557480602

Welch, I. (1989). Seasoned offerings, imitation costs, and the underpricing of initial public offerings. *The Journal of Finance*, *44*(2), 421–449. https://doi.org/10.1111/j.1540-6261.1989.tb05064.x

Wikipedia:Contents/Categories. (n.d.). in *Wikipedia*. Retrieved June 25, 2020, from https://en.wikipedia.org/wiki/Portal:Contents/Categories

Wikipedia:Pageview statistics. (n.d.). in *Wikipedia*. Retrieved June 25, 2020, from https://en.wikipedia.org/wiki/Wikipedia:Pageview_statistics

Wikipedia:Statistics. (n.d.). in *Wikipedia*. Retrieved June 25, 2020, from https://en.wikipedia.org/wiki/Wikipedia:Statistics

Wikipedia.org is more popular than... (n.d.). in *Wikipedia*. Retrieved June 25, 2020, from https://meta.wikimedia.org/wiki/Wikipedia.org_is_more_popular_than...

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value? *The Journal of Finance*, *51*(1), 137–167. https://doi.org/10.2307/2329305

Wu, A. H. (2018). Gendered language on the economics job market rumors forum. *American Economic Association Papers and Proceedings*, *108*, 175–179. https://doi.org/10.1257/pandp.20181101

Xu, S. X., & Zhang, X. (Michael). (2013). Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction. *MIS Quarterly*, *37*(4), 1043–1068. https://doi.org/10.25300/MISQ/2013/37.4.03

You, J., Coakley, J., Firth, M., Fuertes, A.-M., & Shen, Z. (2018). *Driving the presence of investor sentiment: The role of media tone in IPOs*. SSRN Electronic Journal. https://www.ssrn.com/abstract=3221073

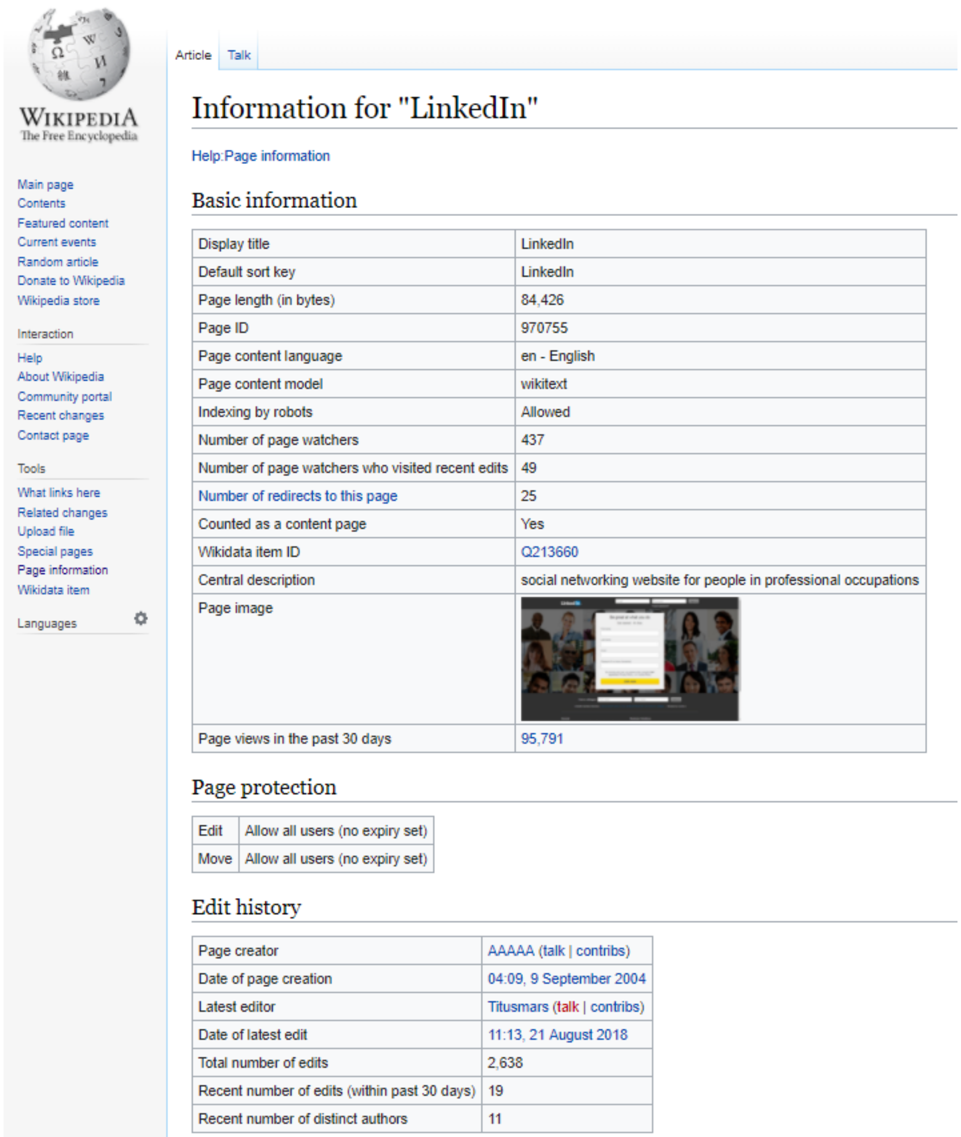
Zhou, D. (2018). *Do numbers speak louder than words?* SSRN Electronic Journal. Retrieved June 25, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2898595

Zhou, M. (Jamie), Lei, L. (Gillian), Wang, J., Fan, W. (Patrick), & Wang, A. G. (2015). Social media adoption and corporate disclosure. *Journal of Information Systems, 29*(2), 23–50. https://doi.org/10.2308/isys-50961

Zimmerman, D. H., & West, C. (1975). Sex roles, interruptions and silences in conversation. In B. Thorned & N. Henley (Eds.), *Language and Sex: Difference and Dominance* (pp. 105–129). Newbury House.

APPENDICES

Appendix A: Wikipedia "Page Information" Example



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Page information
Wikidata item

Languages

Article Talk

Information for "LinkedIn"

Help:Page information

Basic information

Display title	LinkedIn
Default sort key	LinkedIn
Page length (in bytes)	84,426
Page ID	970755
Page content language	en - English
Page content model	wikitext
Indexing by robots	Allowed
Number of page watchers	437
Number of page watchers who visited recent edits	49
Number of redirects to this page	25
Counted as a content page	Yes
Wikidata item ID	Q213660
Central description	social networking website for people in professional occupations
Page image	
Page views in the past 30 days	95,791

Page protection

Edit	Allow all users (no expiry set)
Move	Allow all users (no expiry set)

Edit history

Page creator	AAAAA (talk | contribs)
Date of page creation	04:09, 9 September 2004
Latest editor	Titusmars (talk | contribs)
Date of latest edit	11:13, 21 August 2018
Total number of edits	2,638
Recent number of edits (within past 30 days)	19
Recent number of distinct authors	11

Figure A: Wikipedia "page information" page of LinkedIn

Appendix B: Wikipedia Page Example

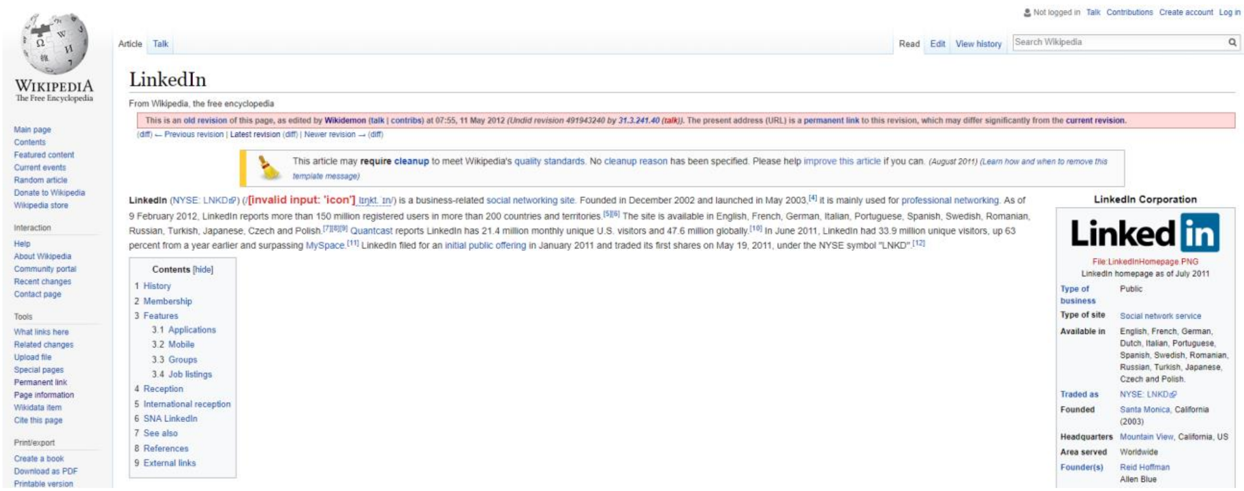


Figure B: The historical Wikipedia page of LinkedIn on May 19th, 2011

Appendix C: Example of Wikipedia Page Revision History

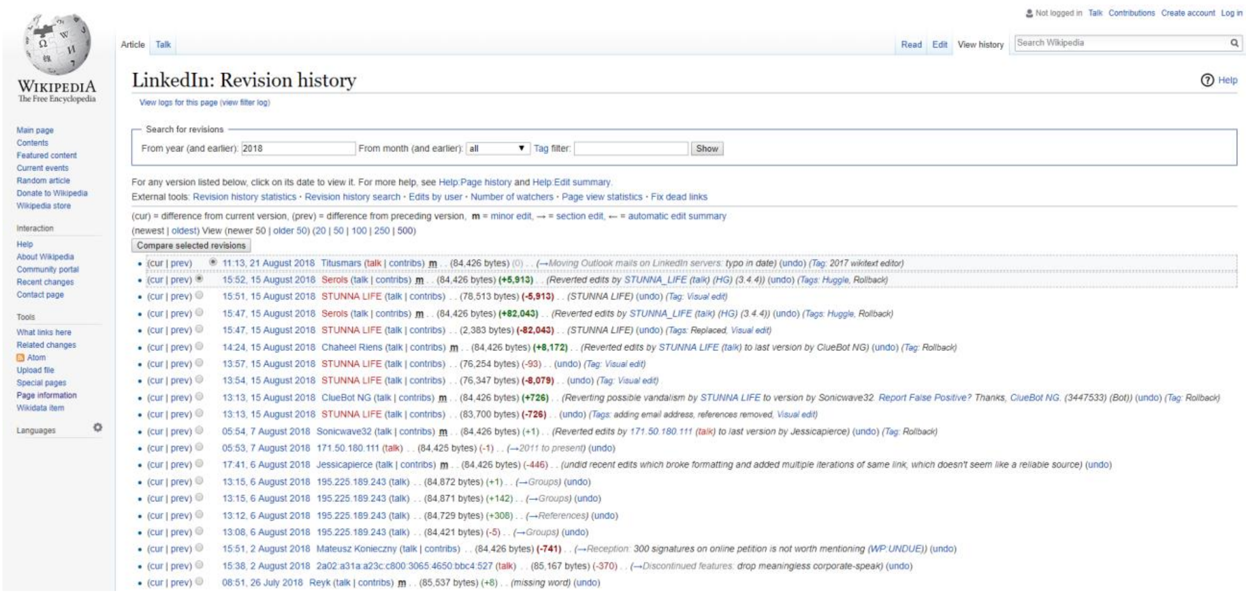


Figure C.1: Revision history page of LinkedIn

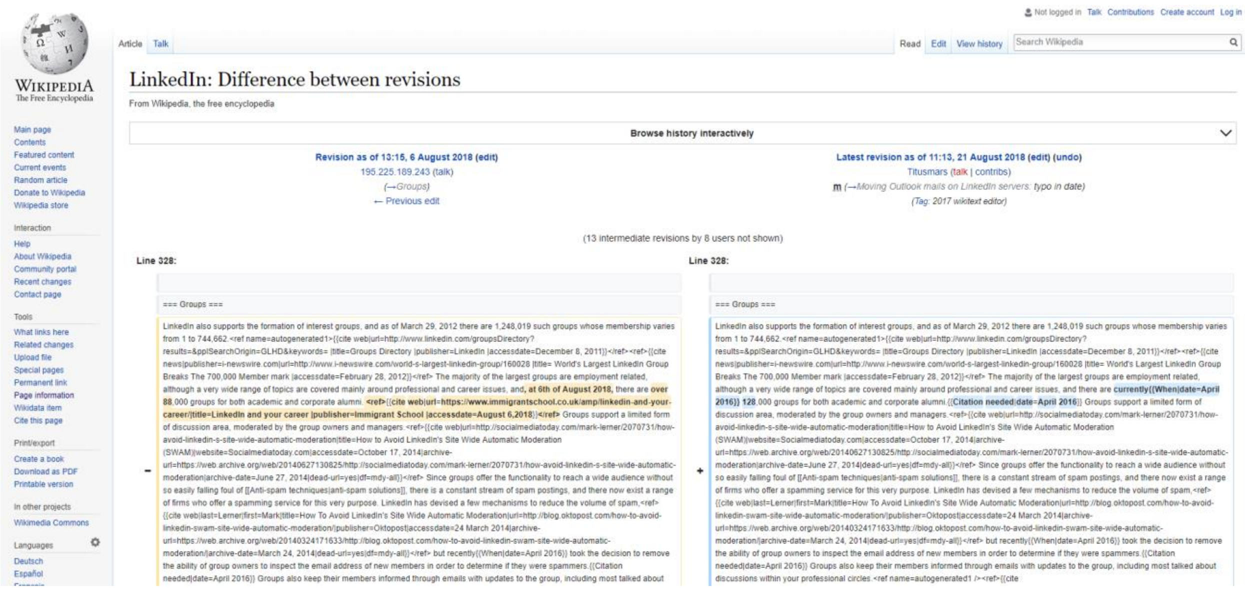


Figure C.2: Comparison between historical Wikipedia pages of LinkedIn

Appendix D: Top Words in Wikipedia and S-1

Table D.1: Percentage of top words in Wikipedia

Positive %		*Negative %*		*Uncertainty %*		*Litigious %*		*positive_GI %*		*negative_GI %*	
BEST	10.4	ERRORS	2.8	APPROXIMATELY	10.0	CONTRACT	8.9	HEALTH	4.9	DEAD	5.5
GREAT	3.3	AGAINST	2.6	MAY	8.8	CLAIMS	6.6	HOME	4.7	DIVISION	4.8
LEADING	3.2	LATE	2.1	RISK	8.8	LAWSUIT	6.5	BEST	4.4	AGAINST	3.6
BETTER	2.9	CLOSED	2.0	COULD	8.6	COURT	5.6	ENTERTAINMENT	2.3	YELP	2.9
POPULAR	2.7	CLAIMS	1.8	UNKNOWN	4.9	LEGAL	5.2	PROTECTION	1.8	BANKRUPTCY	2.5
WINNER	2.6	BANKRUPTCY	1.8	POSSIBLE	4.9	LAW	5.0	SUPER	1.8	CANCER	1.9
LEADERSHIP	2.4	SPAM	1.7	NEARLY	4.4	CONTRACTS	4.5	COMMUNITY	1.7	INVALID	1.9
WINNERS	2.3	LACKING	1.6	SUGGESTED	3.3	SETTLEMENT	3.5	EDUCATION	1.6	LOSS	1.8
INNOVATION	2.2	FORCE	1.3	ALMOST	2.9	DOCKET	3.3	GRAND	1.6	CONTROVERSY	1.6
ABLE	2.1	INVALID	1.3	MIGHT	1.7	CLAIM	2.3	PARTNERSHIP	1.4	HUNGRY	1.6
GOOD	2.1	LOSS	1.3	DEPENDING	1.7	REGULATORY	2.3	FRESH	1.4	THEFT	1.4
SUCCESSFUL	2.0	IGNORED	1.2	ROUGHLY	1.6	SUED	2.2	PRIVACY	1.4	ERROR	1.4
ALLIANCE	1.9	CONTROVERSY	1.2	SOMETIMES	1.3	BREACH	1.9	CREATE	1.3	EMERGENCY	1.4
SUCCESS	1.8	ERROR	1.0	VARIABLE	1.2	LAWS	1.5	INTELLIGENCE	1.3	COMPETITOR	1.4
ENABLES	1.4	CUT	0.9	VARY	1.1	LAWSUITS	1.4	PARTNER	1.2	COMPETITION	1.4
INNOVATIVE	1.3	LOST	0.8	BELIEVED	1.0	AMENDMENT	1.4	SOLUTION	1.2	SAP	1.1
EXCLUSIVE	1.3	FRAUD	0.8	APPEARS	0.9	JUSTICE	1.3	POPULAR	1.1	WAR	1.1
GREATER	1.3	PROBLEMS	0.7	SPECULATION	0.9	ALLEGED	1.3	HUMAN	1.0	DEATH	1.1
POSITIVE	1.2	CRITICISM	0.7	AMBIGUOUS	0.9	REGULATORS	1.3	ABILITY	0.9	COMPETE	1.1
EASY	1.1	CONCERNS	0.7	APPEARED	0.9	LITIGATION	1.0	FITNESS	0.9	FRAUD	1.1

Table D.2: Percentage of top words in S-1

Positive %		Negative %		Uncertainty %		Litigious %		positive_GI %		negative_GI %	
EFFECTIVE	10.2	LOSS	5.3	MAY	33.1	AMENDED	6.1	SIGNIFICANT	5.6	LOSS	8.9
BENEFIT	5.2	AGAINST	2.9	COULD	12.6	REGULATORY	5.9	EFFECTIVE	4.7	LIABILITY	6.2
ABLE	5.1	CLAIMS	2.8	APPROXIMATELY	7.5	LAWS	5.8	ABILITY	4.4	AGAINST	4.8
GREATER	3.4	ADVERSELY	2.7	BELIEVE	6.3	REGULATIONS	4.9	PRO	3.1	ADVERSE	3.8
GAIN	2.6	RESTATED	2.6	RISK	4.6	CLAIMS	4.5	ABLE	2.3	COMPETITIVE	2.9
BENEFICIAL	2.4	LOSSES	2.6	RISKS	2.8	LAW	4.3	OBTAIN	2.3	FAILURE	2.6
SUCCESSFUL	2.2	ADVERSE	2.3	ASSUMPTIONS	1.9	CONTRACT	3.7	HEALTH	2.1	EXCESS	2.5
SUCCESS	2.1	TERMINATION	2.2	ASSUMED	1.7	CONTRACTS	3.5	RELEVANT	1.6	DEPRECIATION	2.1
OPPORTUNITIES	2.1	CLOSING	2.0	INTANGIBLE	1.4	LEGAL	3.4	PROPRIETARY	1.4	DIFFICULT	1.8
ACHIEVE	2.0	IMPAIRMENT	1.8	ASSUMING	1.2	SHALL	3.0	BONUS	1.3	COMPETITION	1.8
SUCCESSFULLY	2.0	FAILURE	1.6	MIGHT	1.1	INDEMNIFICATION	2.7	PARTNER	1.2	LIQUIDATION	1.8
GOOD	1.8	UNABLE	1.6	ANTICIPATE	1.1	AMENDMENT	2.3	REASONABLE	1.1	COMPETE	1.7
BENEFICIALLY	1.7	LITIGATION	1.4	POSSIBLE	1.1	LITIGATION	2.2	SUFFICIENT	1.1	CANCER	1.4
OPPORTUNITY	1.7	LIMITATIONS	1.3	DEPEND	1.0	CONSENT	1.8	PARTNERSHIP	1.1	DISEASE	1.4
PROFITABILITY	1.7	TERMINATE	1.3	ANTICIPATED	0.9	CONTRACTUAL	1.7	OFFSET	1.1	VOLATILITY	1.3
LEADING	1.7	DECLINE	1.2	VOLATILITY	0.9	REGULATION	1.6	BENEFICIAL	1.1	BREACH	1.3
BEST	1.7	DIFFICULT	1.1	PENDING	0.9	SETTLEMENT	1.5	COMPREHENSIVE	1.1	DIFFER	1.3
ENABLE	1.5	DELAY	1.1	DIFFER	0.8	CLAIM	1.5	AUTHORITY	1.0	DEFICIT	1.1
FAVORABLE	1.5	TERMINATED	1.1	FLUCTUATIONS	0.8	COURT	1.4	PROTECTION	1.0	NEGATIVE	1.1
IMPROVE	1.4	LIQUIDATION	1.1	VARIABLE	0.7	BREACH	1.2	SUCCESSFUL	1.0	LIMITATION	1.0

Appendix E: Sample Construction

	Conference calls
Initial sample	81,677
Merge with I/B/E/S to obtain quarterly forecasts / recommendations related to earnings conference calls	70,224
Remove observations without at least one corresponding quarterly earnings forecast issued within 365 days prior to the earnings conference call. Remove estimates without analyst name, brokerage ID (ESTIMID). Remove estimates made by team (i.e., analyst name is "RESEARCH DEPARTMENT" or two last names separated by "/")	70,023
Drop observations for which two or more analysts have the same first initial and last name at the same brokerage	69,995
Remove observations for which the firm is covered by only one analyst for a fiscal quarter end	66,813
Remove observations with no Compustat/CRSP data	65,888
Keep the last quarterly forecast prior to conference call date	63,720
Remove observation with missing values	**62,644**

Appendix F: Variable Definitions for Chapter Two

Variable	Definition
Conference call level variables	
MktCap	Market value of equity, in million dollars
Leverage	Book value of debt and equity divided by the market value of equity.
MB	Ratio of market value of equity to book value of equity.
ROA	Net income in the most recent quarter divided by total assets
SP500	Indicator variable equal to 1 if a firm is a component of Standard and Poor's 500 index and 0 otherwise.
InstOwn	Percentage of aggregate institutional ownership in shares outstanding of firm in the Thomson Reuters 13-F filing immediately prior to conference call date.
AnaCover	Number of analysts issuing one-quarter-ahead or two two-quarter-ahead forecast and having an outstanding stock recommendation for the current fiscal quarter
SUE	Actual quarterly EPS minus consensus EPS forecast, scaled by the stock price at the quarter end
RecCon	Mean stock recommendation scaled into [-2,+2] discrete interval as of the conference call date. -2 indicates strong sell and +2 indicates strong buy.
Runup	Fama-French 4-factor adjusted cumulative return during the [-42,-2] window relative to the conference call date
CallCluster	Number of other conference calls with the same 3-digit SIC code as the focal conference call held in the same calendar quarter
WordsQNA	Log-transformed number of words spoken in question-and-answer portion of conference call, in thousands
FollowupCall	Number of non-continuous interactions between analysts and executives in a call
AnaCount	Number of analysts in the conference call
IBESCount	Number of IBES analysts in the conference call
IBESPart	Indicator variable equal to 1 if at least one IBES analyst participates
ExeCount	Number of executives in the conference call
CEOPart	Indicator equal to 1 if CEO attends the conference call
CFOPart	Indicator equal to 1 if CFO attends the conference call
CEOCFOPart	Indicator equal to 1 if both CEO and CFO attend the conference call
FemaleAnaPct	Proportion of female analysts, in decimal
CAR	Fama-French 4-factor adjusted cumulative return during the [-1,+1] event window relative to the conference call date
netAnaCall	Weighted average net tone (positive tone minus negative tone) of all participating analysts in a call
netExeCall	Weighted average net tone (positive tone minus negative tone) of all participating executives in a call
CsrDiv	Net CSR score based on *Diversity* category in MSCI ESG Stats Database

Csr5	Net CSR score based on *Community, Diversity, Employee Relations, Environment, Human Rights* categories in MSCI ESG Stats Database
Csr7	Net CSR score based on seven major categories in MSCI ESG Stats Database
CsrAll	Net CSR score based on seven major categories and six Controversial Business Issues categories in MSCI ESG Stats Database

Analyst-call level variables

FemaleAna	Dummy variables equal to 1 if the analyst is female
FemaleExe	Proportion of executive narratives accounted by female executives
Participate	Indicator variable equal to 1 if an analyst asks a question in firm's quarterly earnings conference call and 0 otherwise.
First	Indicator equal to 1 if this is the analyst is the first questioner in the call
Order	Order of analyst interaction with management in the call
Words	Number of words spoken by the analyst (with suffix *Ana*) or executives (with suffix *Exe*)
AbnLength	Abnormal interaction length for each participant, measured as the standardized difference between the participant's actual length of interactions and the average interaction length for the call
RallyAna	Number of back-and-forth comments between the analyst and executive for the analyst
Interrupt	Number of times analyst (with suffix *Ana*) or executives (with suffix *Exe*) is interrupted by another conference call participant for the analyst. See Table 9 for detailed definitions
Hesit	Number of times analyst (with suffix *Ana*) or executives (with suffix *Exe*) self-corrects or has a broken thought in this conversation
Words	Number of words spoken by analyst (with suffix *Ana*) or executives (with suffix *Exe*)
numberAna	Percentage of numbers the analyst speaks in this conversation/interaction
numberExe	Percentage of numbers the executive speaks in this conversation/interaction
Tone	Percentage of sentiment words in the analyst's (with suffix *Ana*) or executives' (with suffix *Exe*) narrative based on Loughran and McDonald (2011) dictionary. *Tone* can be positive, negative, or net sentiment
ToneGI	Percentage of sentiment words in the analyst's (with suffix *Ana*) or executives' (with suffix *Exe*) narrative based on Harvard GI dictionary. *Tone* can be positive, negative, or net sentiment
net	Net sentiment of combined analyst's and executives' narratives
Rec	I/B/E/S stock recommendation score prior to the conference call in [-2, +2] interval. 2 indicates strong buy, 1 indicates buy, 0 indicates hold, -1 indicates sell, and -2 indicates strong sell.

AllStar	Indicator variable equal to 1 if an analyst is voted as Institutional Investor All-American research team in the prior calendar year of the conference call.
ForeAcc	Negative value of the absolute forecast error demeaned by same quarter-firm average forecast for previous quarter
BrokerSize	Number of analysts hired by affiliated brokerage firm of an analyst in the prior calendar year of the conference call.
GenExp	Number of years between the analyst's first forecast date for the firm and the conference call date.
FirmExp	Number of years between the first forecast date of an analysts and the conference call date.
CompCover	Number of firms covered by an analyst in the prior calendar year of the conference call.
IndCover	Number of Fama-French 48 industries covered by an analyst in the prior calendar year of the conference call.
RecHorizon	Number of days between most recent recommendation announcement date and conference call date
CCUser	Number of other conference calls on which the analyst participates in the same calendar quarter as the focal conference call
Executive-call level variables	
CEO	Indicator equal to 1 if the executive is the CEO in most recent fiscal year
CFO	Indicator equal to 1 if the executive is the CFO in most recent fiscal year
FemaleExeDummy	Indicator equal to 1 if the executive is female

Appendix G: Interruption and Back-And-Forth Comments in Conference Calls

Appendix G show an excerpt for the interaction between BMO Capital Markets analyst, Richard C. Anderson, and two company participants, Timothy M. Schoen and James F. Flaherty, on the quarterly earnings conference call for HPC, Inc. on May 1st, 2012. Richard and James are interrupted by each other twice (identified by "…" and coded as *InterruptAnaExe*=2 and *InterruptExeAna=2*). Richard exhibits two hesitations (identified by "--"and coded as *HesitAna*=2) and exhibits James exhibits six hesitations (*HesitExe*=6). Seth makes seven statements resulting in a value of 7 for *RallyAna*.

Timothy M. Schoen
Chief Financial Officer and Executive Vice President
The insurance recovery and the Google payment was in our guidance.

Richard C. Anderson
BMO Capital Markets U.S.
Okay. That's what I thought. And, Jay, just maybe to refine the acquisition question a little bit for you, what -- of the 5x5 matrix that you talk about, what property type within that, do you think fits best in an environment that you're describing, with a lot of uncertainty, that you would say, this is **the -- maybe** the least risky or the best fit in the environment that you're in right now?

James F. Flaherty
Former Director
Well, if you want to call the ballgame based on lowest risk, that would probably **be...**

Richard C. Anderson
BMO Capital Markets U.S.
I think risk is part of the conversation, **but...**

James F. Flaherty
Former Director
I'm just thinking your question. You defined the question in terms of risk. If you want to know what the lowest risk piece of our economic business model is, it's probably on-campus medical office buildings where the hospital is the #1 or #2 market share hospital system in a growing area. But that's just, that **was -- from** our standpoint, we wouldn't stop there. (continued)

Richard C. Anderson
BMO Capital Markets U.S.
Okay. But what would be some of those other elements where you'll pull the trigger in this environment?

James F. Flaherty
Former Director

Valuation, condition, i.e. fiscal obsolescence, CapEx obsolescence of the portfolio that we're acquiring, quality of the counter party, both from the standpoint **of -- you've** heard me talk forever about, we want to have counter parties that have 3 criteria: quality outcomes, efficient operations and critical mass. So those are the **whole -- it's** kind of **a -- it's** a large algorithm that comes into play.

Richard C. Anderson
BMO Capital Markets U.S.
Do you think life sciences is well placed right now in this environment?

James F. Flaherty
Former Director
I think life science located in one of the 4 or 5 concentrations that are the recipients of the NIH grants is how I'd start that discussion. However, then you have to get and look at the characterization of the tenants. Are they more VC-backed private companies that are working 1 or 2 drugs through a Phase I, Phase II, Phase III ultimately FDA approval process? That would have a lot of risk associated with it. Or are they very substantial companies like Amgen, like Genentech, like Takeda, like Pfizer, like Google, like LinkedIn, sorry. That would have a different element to it. And then I think you really need to think **about -- this** isn't going to impact anything in terms of 2012, 2013. (continued)

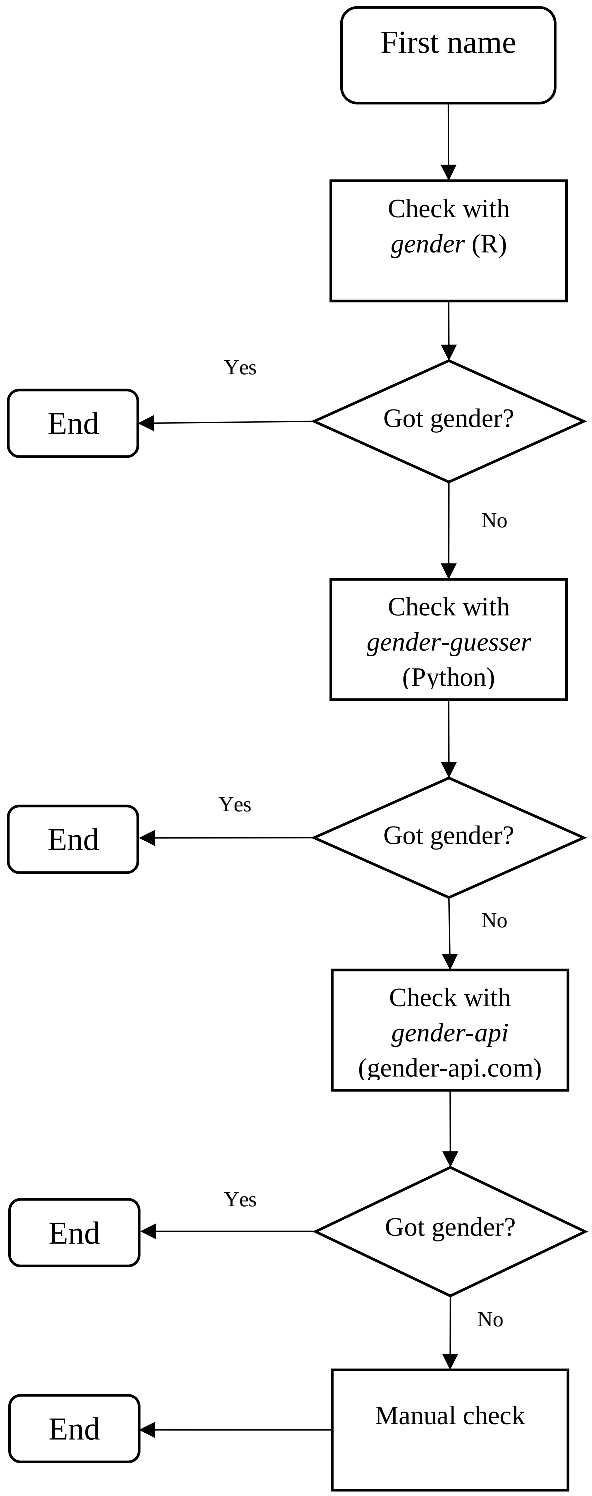
Richard C. Anderson
BMO Capital Markets U.S.
And then lastly, just, I think more of a comment. (continued)

James F. Flaherty
Former Director
Well, we'll certainly take it under advisement. They're just not -- as you know, there's not a lot moving around. But we like to give guidance on our company's results. We think it's a little inappropriate for us to be giving guidance on another company's results, particularly when they have their own strategic plan that **they're...**

Richard C. Anderson
BMO Capital Markets U.S.
You can take their results out of it, and just say what happens to your results in terms of coverage. **Anyway...**

James F. Flaherty
Former Director
We're certainly willing to take a peek at that.

Appendix H: Gender Determination Procedure



Appendix I: Variable Definitions for Chapter Three

Variable	Definition
CAR	The cumulative abnormal return over the [0,+1] window relative to the report date based on Fama-French 4-factor model
Runnp	The cumulative abnormal return over the [-10,-1] window relative the report date based on Fama-French 4-factor model
EA	Indicator variable equal to 1 if the firm issues an earnings announcement over the [-2,+2] window centered on report date
Forecast variables	
EFRep	Indicator variables if the analyst report contains a one-year-ahead annual earnings forecast
RecRep	Indicator variables if the analyst report contains a stock recommendation
PTRep	Indicator variables if the analyst report contains price target forecast with a 12-month horizon
EFRep	Indicator variable if an analyst issues a one-year-ahead annual earnings forecast in the report
RecRep	Indicator variable if an analyst issues a stock recommendation in the report
PTRep	Indicator variable if an analyst issues a 12-month-ahead price target in the report
EF	One-year-ahead annual earnings forecast scaled by the stock price 50 days before the report date
EFRev	Earnings forecast revision calculated as *EF* minus last *EF* for the same fiscal year end
Rec	Stock recommendation from I/B/E/S. 1-Sell, 2-Underperform, 3-Hold, 4-Buy, 5-Strong Buy
RecRev	Stock recommendation revision calculated as *Rec* minus last *Rec*
PT	Price target over a 12-month horizon scaled by the stock price 50 days before the report date
PTRev	Price target revision calculated as *PT* minus last *PT*
Analyst variables	
Female	Indicator variable equal to 1 if the author of the report is female and 0 otherwise
Star	Indicator variable equal to 1 if the analyst is ranked as an *Institutional Investor* All-Star in the current year
BrokerSize	The number of analysts issuing earnings forecasts from the report analyst's brokerage house in the report year
GenExp	The number of years between the analyst's first forecast date on I/B/E/S and the report date
FirmExp	The number of years between the analyst's first forecast date for the firm and the report date

FirmCover	The number of firms covered by an analyst in the prior calendar year of the conference call.
IndCover	The number of Fama-French 48 industries covered by an analyst in the prior calendar year of the report
RepFreq	The number of earnings forecast revisions issued by the analysts the report year
NumReports	The number of report an analysts issues in a year
Textual variables	
Word	Logarithm of the number of words in the report
Page	Number of pages of the report
Fog	The Gunning-Fog index
NonFin	Ratio of sentences without "%" or "$", in percentage (%)
Number	Ratio of numerical content, in percentage (%)
Pos	The percentage of positive words in the report based on Loughran and McDonald (2011) dictionary
Neg	The percentage of negative words in the report based on Loughran and McDonald (2011) dictionary
Net	The difference between *Pos* and *Neg*
Firm variables	
Leverage	The book value of debt to market value of equity at the end of last fiscal year end
BM	The book value of equity to market value of equity at the end of last fiscal year end
MktCap	Logarithm of the market value of equity, in millions, at the end of last fiscal year end
ROA	Return on assets at the end of last fiscal year end
InstOwn	The percentage of institutional ownership at the end of last fiscal year end
NumAna	Number of analysts issuing at least one earnings forecast in the last fiscal year
SegInd	Number of unique 4-digit SIC industry segments