Essays on Information Transmission & Machine Learning in Finance

Inaugural-Dissertation
zur Erlangung des Doktorgrades
des Fachbereichs Wirtschaftswissenschaften
der Johann Wolfgang Goethe Universität
Frankfurt am Main

vorgelegt von
Sasan Mansouri
aus dem
Iran

Frankfurt am Main, 2021

Erstgutachter:
Prof. Dr. Mark Wahrenburg

Zweitgutachter:
Prof. Dr. Alexander Hillert

Tag der Promotion:

Acknowledgment

The creation of this dissertation is made possible, to a large extent, by the support and commitment of a large number of people and I would like to take this opportunity to express my sincere gratitude to all of them.

First and foremost, I would like to thank my doctoral supervisor, Prof. Dr. Mark Wahrenburg, for giving me the opportunity to write this dissertation at his chair. Through constant inspiration, exciting discussions, and considerable academic freedom, you have contributed significantly to the success of this work. I would also like to thank Prof. Dr. Alexander Hillert, who has supported me for many years and contributed significantly to my interest in textual analysis topics. My heartfelt thanks also go to Prof. Dr. Alexander F. Wagner and Prof. Dr. Paul P. Momtaz, for their time and support.

Furthermore, I would like to thank Dr. Andreas Barth, who accompanied me throughout my academic education as a mentor and friend. My interest and progress have certainly been strongly influenced by you. An extraordinary thank you goes to my colleagues Dr. Fabian Wöbbeking, Valerie Laturnus, Arne Reichel, Janika Bockmeyer, Dr. Severin Zörgiebel, Dr. Florian Balke, Milad Goodarzi, Begüm Sapancilar, Kai Rieck, Farrokh Emami, and Sara Fadavi. Without your support and friendship, this promotion would not have been so enjoyable. Thanks also to Kristine Rach-Emma and Valérie Bignon for your support and cooperation at the chair.

Last but not least, I would like to address the people in my immediate personal environment, without whose support and motivation my doctorate would have been impossible. My special thanks also go to Khashaiar & Kourosh Mansouri, I am infinitely proud and happy to be your brother. Finally, my deepest thanks go to the people who made this promotion and so much more possible for me: my parents Azam Bahrabadi and Abolghasem Mansouri. Their unrelenting support and backing paved the way that led me to my doctorate. This dissertation is dedicated to them.

Frankfurt am Main, July. 2021

Sasan Mansouri

Contents

Introduction and Summary

This dissertation consists of five chapters on information transmission and machine learning applications in finance. The first three chapters are concerned with information transmission in quarterly earnings conference calls. The *first chapter* analyzes whether context-specific language or jargon is used in verbal firm disclosures to obfuscate or to efficiently transfer information. The *second chapter* is a follow-up paper where the jargon measure developed in the first chapter is used as an input to a machine learning approach to classify non-answers. This paper provides a glossary of 1,364 trigrams that are found to be frequently used to refrain from factually answering a question. The *third chapter* is concerned with the professional and non-professional business media attention to firms, and its association with the richness in the firm's information environment. The final two chapters are concerned with information transmission in the context of startups' token offerings. The *fourth chapter* investigates the role of freelancing analysts as information intermediaries. Finally, the *fifth chapter* analyzes the economic attractiveness of the ESG related issues for the startups and their investors.

The paper of the first chapter is titled 'How to talk down your stock performance', and co-authored with Andreas Barth, Fabian Woebbeking, and Severin Zoergiebel. For this paper, a revision has been requested by the *Journal of Banking & Finance*. The paper of the second chapter, ' "Let me get back to you" - A machine learning approach to measuring non-answers', co-authored with Andreas Barth and Fabian Woebbeking, is currently under revision for resubmission to *Management Science*. The paper of the third chapter, 'Does firm's silence drive media's attention away?' shall be submitted in the near future. The paper of the fourth chapter, 'ICO analysts', co-authored with Andreas Barth, Valerie Laturnus, and Alexander F. Wagner, is submitted and is currently under review at the *Journal of Accounting Research*. Finally, the fifth paper, "Financing sustainable entrepreneurship: ESG measurement, valuation, and performance in token offerings", co-authored with Paul P. Momtaz is currently under review at the *Strategic Entrepreneurship Journal*.

Chapter I: How to talk down your stock performance

The first chapter is an empirical analysis co-authored with Andreas Barth, Fabian Woebbeking, and Severin Zoergiebel. In this paper, we use natural language processing to study the use of financial jargon in verbal firm disclosures and their impact on financial markets. Firm disclosures, in general, are a fundamental tool to transfer information to investors. The overarching question of how financial jargon is used to obfuscate or efficiently transfer information is therefore of great importance to the formation of efficient financial markets.

The paper analyzes two competing hypotheses of how markets perceive (a lack of) factual language when senior management verbally conveys information in earnings calls. On one hand, avoiding complex terminology in favor of a broader language could improve the acces-

sibility of financial disclosures. On the other hand, markets could interpret the absence of precise financial terminology as blathering, i.e. obfuscating information by 'beating around the bush.'

The paper utilizes the unique settings of financial firms' quarterly earnings calls to study the market's understanding of financial jargon. These calls have a semantically homogeneous context, where any discussion on either financial health, future strategy, or new products always involves financial terminology, and there is little interest in topics outside the contextual domain of finance. Thus, the most efficient way to transfer information is the usage of financial jargon. For each earnings call in our sample, we calculate the *JargonRatio*, defined as the ratio of finance-related words to total words in the responses of the management team to the calls' participants. Finance-related words are identified with the Hypertextual Finance Glossary by Campbell Harvey, the largest available finance glossary with more than 8,500 entries.

We observe lower cumulative abnormal returns and higher implied volatility following earnings calls where managers use less financial jargon, supporting the argument that excessive use of non-factual language is perceived as blathering that retards the reduction of information asymmetries. 'Beating around the bush' is particularly pronounced when earnings management is more likely, when analysts' questions are tougher, and when last quarters' return on equity was poor.

This paper adds – with management blathering – an important dimension to the literature on disclosure choices and information transmission in financial markets. Beyond the scope of this paper, the metric for factual language that we develop here already proved to be a useful input for further research. For example, in the follow-up paper presented in the next chapter, we are using the jargon measure as an input to a machine learning approach to classify non-answers.

The paper in this chapter is available online via SSRN and has been presented at international conferences, such as the 21st Annual Conference of the Swiss Society for Financial Market Research, the 2018 Conference of the International Finance and Banking Society, the 7th Paris Financial Management Conference (PFMC-2019), and the 37th International Conference of the French Finance Association. This paper is currently invited for resubmission to the *Journal of Banking & Finance*.

Chapter II: "Let me get back to you" - A machine learning approach to measuring non-answers

The second chapter is co-authored with Andreas Barth and Fabian Woebbeking. In this paper, we build on the foundation of research in linguistics to better understand how humans request and process information through questions and answers. We use a supervised

machine learning framework called Multinomial Inverse Regression (MNIR) to develop a measure of "non-answers", which identifies the absence of requested information in the management's response to analysts' questions during earnings conference calls.

We train our MNIR model using two symptoms of the non-answer; first, "rejection phrases" such as direct refusals, and second, "blathering phrases" with a low signal-to-noise ratio about the underlying context. The training set includes all Q&As of earnings conference calls for financial firms in the S&P 500 from 2002 to 2019. The final result of this procedure is a list of 1,364 trigrams such as 'back to you', 'do not know', 'hard to predict', etc., which are found to be frequently used to refrain from answering a question concisely and factually.

We conduct a variety of tests to evaluate the plausibility of our glossary. First, we investigate to which questions managers try to avoid a response during an earnings conference call. Second, we examine cumulative abnormal stock returns and implied volatilities following earnings conference calls for the validation set of non-financial S&P 500 firms. Results show that, first, within an earnings call, non-answers are observed more prevalently for tougher and more critical questions. More specifically, non-answers are more frequent for managements' responses to follow-up questions by the same analyst and managements' responses to more negative questions. In addition, we observe more non-answers for questions with forward-looking sentences, i.e., questions that refer to (potentially unknown) future outcomes. Second, we observe a strong negative impact on stock returns for the measure derived from our glossary, i.e., not answering analysts' questions leads on average to negative abnormal stock returns after an earnings call. Furthermore, linking our measure to option implied volatilities after earnings conference calls shows that investor uncertainty increases if the requested information in the call has not been provided by the management. Financial analysts, too, perceive non-answers as a negative signal. In particular, we find that analysts are less likely to modify their EPS forecasts upwardly following a call with many non-answers.

Our metric is designed to be of general applicability for Q&A situations, and hence, is capable of identifying non-answers outside the contextual domain of financial earnings conference calls. To corroborate this claim, we briefly explore textual data from presidential interviews, thereby motivating additional research into alternative settings such as central banks press conferences and U.S. Senate hearings. We provide a machine-readable version of the glossary at econlinguistics.org as well as an online application for measuring the non-answer score of the user-input texts at econlinguistics.shinyapps.io/econlinguistics.

This paper has been awarded the 2021 Lazaridis Institute Prize for best paper on accounting issues relevant to technology firms at the 2021 annual meeting of the Canadian Academic Accounting Association. Moreover, the paper has been awarded the best conference paper award of the 82nd annual meeting of the German Academic Association of Business Research (VHB).

The paper in this chapter is available online via SSRN and has been accepted for presentation at international conferences, such as the 18th Paris December Finance Meeting, the European Accounting Association's (EAA) first Virtual Annual Congress, the 2021 CAAA (Canadian Academic Accounting Association) Annual Conference, the 2021 annual AFAANZ (Accounting & Finance Association of Australia and New Zealand) Conference, and the 27th Annual Meeting of the German Finance Association (DGF). This paper is currently invited for resubmission to *Management Science*.

Chapter III: Does firm's silence drive media's attention away?

The third chapter is a single-authored paper. This paper sheds light on the business media coverage of firms and verifies whether the robustness of the information environment surrounding a firm skews the coverage a firm receives.

Media platforms enjoy the broadest audience compared to other information intermediaries, and there is abundant literature showing the effects of media attention on firms' security prices, corporate governance, and investors' attention. However, we know little about the driving forces behind media coverage. In this paper, I examine the media coverage of big corporations using two competing hypotheses to verify whether the business media/press coverage of firms is demand- or supply-driven. Specifically, media attention is "demand-driven" if business media respond to the stakeholders' demand for more information and analyses about firms with less robust information environment. For firms with a less robust information environment, it is more challenging for media sources to gather enough publishable materials. In other words, a firm's media exposure is "supply-driven", if the media sources, as suppliers of information, reduce their coverage of firms that are more difficult to cover.

To measure the quality of the information environment surrounding a firm, I use the non-answer score developed in the previous chapter. Results of the empirical study in chapter two indicate that a higher non-answer score for a firm's earnings call reflects higher demand for the information among firms' stakeholders. Furthermore, I collect the media coverage information before and after each earnings call using the Ravenpack dataset.

The empirical analysis shows that first, while the demand for firm-specific information increases due to non-answers, the media also deliver less content due to supply-side difficulties of information acquisition. Second, differentiating between different types of coverage based on their production costs for the media sources, i.e. full articles versus other types, confirms the supply-driven coverage hypothesis for full articles only. Third, I differentiate between the professional business media like Dow Jones Newswire and the non-professional business media including blog posts written by a non-professional analyst (e.g. Seeking-Alpha), and

show that firms with higher non-answer earnings calls, witness a lower share of professional coverage in the next quarter.

The paper in this chapter is available online via SSRN and has been accepted for presentation at international conferences, such as the 2021 Annual Meeting of the Swiss Society for Financial Market Research (SGF), the 37th International Conference of the French Finance Association (affi), the 2021 Annual Conference of the German Economic Association (VfS), the 2021 CAAA (Canadian Academic Accounting Association) Annual Conference, 2021 Australasian Meeting of the Econometric Society, 14th International Risk Management Conference (IRMC), and the 2021 annual AFAANZ (Accounting & Finance Association of Australia and New Zealand) Conference.

Chapter IV: ICO analysts

The fourth chapter is co-authored with Andreas Barth, Valerie Laturnus, and Alexander F. Wagner. In this paper, we add novel insights to the literature on how analysts contribute to the functioning of capital markets. We use the setting of Initial Coin Offerings (ICO) to investigate determinants and consequences of the quantitative and qualitative aspects of investment ratings issued by human experts. The ICO setting has the particularly valuable feature that investors can observe the track record of analysts and their potentially conflicting activities, allowing us to analyze whether this information is valuable for market participants.

ICOs are token sale events on an own or existing blockchain to facilitate the financing of an entrepreneurial venture. The lack of regulation, easy accessibility to funds, and the widespread excitement about the rise of Bitcoin and other cryptocurrencies has spurred the growth of the ICO market since the mid-2010s.

As in many FinTech markets, the problem of asymmetric information looms particularly large. This paper focuses on the role of freelancing human experts (ICO analysts) as information intermediaries. We collect the data on ICOs, ICO ratings, and ICO analysts from the platform ICObench.com. Our data covers 5,384 ICOs.

While investors on average follow analysts' ratings, there are numerous cases where ICOs were not successful despite positive (machine-generated and human expert) ratings, or where ICOs were successful despite negative ratings. Therefore, we investigate which characteristics of analysts or of the ICO itself lead to the discrepancy between analysts' advice and the market, i.e., what determines the probability that a human analyst gives a sell (buy) recommendation but the ICO becomes successful (fails).

As a novel contribution, this paper addresses potential conflicts of interest in ICO analyst ratings, and documents that ratings vary in quality and exhibit biases due to reciprocal interactions of ICO analysts with ICO team members. We show that investors assign more weight to ratings provided by high-quality analysts while discounting reciprocal ratings,

so the market discipline works only to some extent and hence, intermediaries can play an important role in mitigating information asymmetries.

The paper in this chapter has been published as CEPR Discussion Paper 16200 and has been accepted for presentation at international conferences, such as the Jahrestagung UFSP Finanzmarktregulierung, the 2021 annual conference of the German Economic Association (VfS), and the American Finance Association(AFA) 2022 annual meeting. The paper is currently under review at the *Journal of Accounting Research*.

Chapter V: Financing sustainable entrepreneurship

The fifth chapter is co-authored with Paul P. Momtaz. In this paper, we investigate the economic attractiveness of environmental, social and governance (ESG) related issues for entrepreneurs and investors.

Sustainable Entrepreneurship (SE) is characterized by profit-seeking entrepreneurial activity that embraces the broader (non-financial) environment, society, and governance goals of our time. The entrepreneurship literature's tenet is that market failure to solve those ESG challenges creates entrepreneurial opportunities. An important research gap is whether ESG-driven opportunities are economically attractive for entrepreneurs and investors in the first place. This is ambiguous because (i) ESG goals impose binding restrictions upon entrepreneurs that limit the scope of viable routes to (economic) success, and (ii) entrepreneurs largely fail to internalize ESG rents because they come as positive externalities. Therefore, we directly pose the research question: How (economically) attractive is SE for entrepreneurs and investor?

A key contribution of our paper is that we have developed a machine-learning algorithm to estimate ESG ratings for startups from text data, such as a startup's mission statement, pitch deck, whitepaper, and so forth. This is important given that there are no commercially or otherwise available ESG ratings for startups as there are for larger corporations. We trained a computer to identify ESG-specific language from the universe of all Financial Times articles with the tags "Moral Money" and "ESG Investing." We find that our machine-learning approach is extremely reliable and robust. We make the Python source code available via our GitHub page, and also developed an easy-to-use web app so that future research can apply our method by simply copying&pasting text on our website `www.SustainableEntrepreneurship.org`.

Empirically, we examine a large sample of 1,043 token offerings (i.e., blockchain-based crowdfunding events) over the 2016-2020 period. We find that startups with salient ESG properties benefit from substantially higher valuations. A one-standard-deviation increase in the ESG metric is associated with a 28% increase in the funding amount, which corresponds to around $4.2 million (relative to the mean funding amount of $15.2 million in our sample).

Interestingly, startups with pronounced ESG properties underperform post-funding. The price of a token (i.e., one unit of asset sold in the crowdfunding event) decreases by 16% in the year after the crowdfunding campaign.

Another main finding is intriguing: We analyze how pre-existing binding constraints related to the startups technology, network, and governance impact the valuation and performance of SE. Such preexisting binding constraints hurt both valuation and performance, which is predicted in the SE literature but marks a stark contrast to conventional entrepreneurship. For example, venture capital backing is often associated with better valuation and performance, but in sustainability-oriented startups it hurts.

In addition to our novel machine-learning approach to quantify startups' ESG properties, we also ensure that our results are not biased by observed and unobserved heterogeneity. Our results are robust in many additional sensitivity checks (which are partly documented in the appendix).

The paper in this chapter is available online via SSRN and has been accepted for presentation at international conferences, such as the Entrepreneurial Finance Association 2021 Annual Meeting and the 27th Annual Meeting of the German Finance Association (DGF). The paper is currently under review at the *Strategic Entrepreneurship Journal*.

Chapter I

How to talk down your stock performance

How to talk down your stock performance*

A. Barth†, S. Mansouri†, F. Woebbeking†, S. Zoergiebel†

January 18, 2021

Abstract

We process the natural language of verbal firm disclosures in order to study the use of context specific language or jargon and its impact on financial performance. We observe that, within the Q&A of earnings conference calls, managers use less jargon in responses to tougher questions, and after a quarter of bad economic success. Moreover, markets interpret the lack of precise information as a bad signal: we find lower cumulative abnormal returns and a higher implied volatility following earnings calls where managers use less jargon. These results support the argument that context specific language or jargon helps to efficiently and precisely transfer information.

Keywords: financial jargon, natural language processing, textual analysis, stock returns, implied volatility, information exchange, corporate disclosure

JEL-Classification: D82, G12, G14, G30

*We would like to thank Renee Adams, Daniel Andrei, Doron Avramov, Turan Bali, Justin Birru, Wayne Guay, Alexander Hillert, Kari Heimonen, Michael Kirchler, Christian Leuz, Hugues Langlois, Philippe Mueller, Christian Schlag, Oliver Spalt, Alexander Wagner, Mark Wahrenburg, Zexi Wang, and Xiao-Jun Zhang for their valuable comments and suggestions.

†Goethe University Frankfurt, House of Finance, Theodor-W.-Adorno-Platz 3, 60629 Frankfurt, Germany. Send correspondence to woebbeking@finance.uni-frankfurt.de, Tel: +49 (0)69 798-33731.

1 Introduction

"[...] The language of money is a powerful tool, and it is also a tool of power. Incomprehension is a form of consent. If we allow ourselves not to understand this language, we are signing off on the way the world works today [...]"

— The New Yorker, *"Money Talks – Learning the language of finance", published in the print edition of the August 4, 2014, issue.*

The efficient and ultimately successful transfer of information is the foundation for any kind of (economic) cooperation and often hinges on the common understanding of a specific language or jargon. For instance, a patient might have a hard time understanding a doctor who uses a high degree of medical jargon. However, being versed in the contextual vocabulary allows the recipient to access information at a much higher rate and precision. In finance, too, financial jargon could retard the flow of information by adding complexity, conversely, the absence of financial jargon could indicate the unwillingness or inability to disclose precise information. Both too complex (Li, 2008; Bloomfield, 2008; Loughran and McDonald, 2014; Bonsall, Leone, Miller and Rennekamp, 2017) or too little information (Hollander, Pronk and Roelofsen, 2010) could be understood as obfuscation, and thus, the perception of jargon highly depends on the financial literacy of market participants. We know, however, only little about the understanding of financial jargon, and there is hardly guidance on tools that help managers to efficiently convey information.

This paper analyzes whether the usage of financial jargon in verbal management disclosures makes management communication more efficient, or, alternatively, increases the complexity of conveyed information. We find evidence that market participants understand financial jargon, and a lack of jargon is perceived as bad signal. Managers seem to be aware of that and avoid jargon in disadvantageous situations, that is, situations where vague language might be used to conceal disadvantageous facts.

Earnings calls of financial firms offer a unique setting to study the market's understanding of financial jargon. More specifically, these calls have a semantically homo-

geneous context, where any discussion on either financial health, future strategy or new products always involves financial terminology, and there is little interest in topics outside the contextual domain of finance. Thus, the most efficient way to transfer information is the usage of financial jargon.[1]

We collect transcripts of earnings conference calls for *financial firms* in the S&P 500 for the period 2002 to 2019 and focus within these calls on the Q&A session, i.e. the spontaneous answers of the management to analysts' questions. For all answers within a call, we calculate the $JargonRatio$, defined as the ratio of finance-related words to total words. Finance-related words are identified with the Hypertextual Finance Glossary by Harvey (2016), the largest available finance glossary with more than 8,500 entries.

Consider the response of Ian Lowitt, the CFO of Lehman Brothers to a question by Deutsche Bank analyst Mike Mayo in an earnings call in June 2008 as one example of an answer with a low $JargonRatio$ of 3.2%. To the question:

> *"It would be helpful to have some carrying values for all these charts. It's good detail but without knowing the carrying values we're going to have to make a lot of assumptions when we estimate write-downs where they might – should be,"*

he responded with the words

> *"I think, Mike, we recognize that we've taken the step on the path around providing additional detail and we don't think we've reached the end of that. And it is something that we're going to look to find ways in which we can make these things more apparent to* ***investors****. I would say, even though we will get back to you with the exact* ***composition*** *of each of these different items, whatever that turns out to be, we have seen a huge amount of flow through all of those different categories and that's what's provided the information that*

[1] As a consequence, the interpretation of the inverse of financial jargon as unwillingness to disclose information holds only true for financial firms. For other industries, precise answers will certainly contain non-financial terminology, for example a description of new products or technologies. This idiosyncratic jargon, however, is less constant over time and might change very rapidly.

> *we needed in order to mark them to where the* ***market*** *is currently* ***trading*** *and the fact that there's so much flow going through gives you a sense of the* ***markets*** *are in fact trading and there is great price transparency around those and we are marked to what that set of data is telling us."*

This is an example for an elaborate answer with little information content. For now, we leave it to the reader to decide if this constitutes a satisfactory answer. In the remainder of the paper, we shall provide evidence on how markets in general perceive answers with a low (high) amount of jargon.

We begin by investigating managers use of financial jargon and find that managers avoid a precise and factual language in disadvantageous situations: *Within* an earnings call, answers to more critical questions contain less jargon. In particular, managers use less jargon when responding to later questions,[2] and to more negative toned questions. Likewise, we observe less jargon in earnings calls following a quarter with poor performance.

Next, we analyze market reactions to the usage of jargon. We document that equity markets appreciate the use of factual language and punish the lack of jargon. We also find a higher implied volatility after earnings calls with less jargon, i.e. investors are willing to pay a higher premium in order to insure themselves against stock price changes after a conference call where the management uses less factual language.

All results are robust to controlling for a number of confounding factors, such as tone and uncertainty in the language used by managers, the ratio of numeric to textual content (Zhou, 2018), or the linguistic complexity of the language, as measured by the Gunning (1952) *Fog* Index. When analyzing market reactions, we also control for the return in the previous quarter as well as for analysts' expectations, and use different definitions of abnormal returns. In all analyses, we further absorb a number of observable and

[2] As shown by Mayew (2008) and Cohen, Lou and Malloy (2013), managers have an incentive to 'cast' analysts during a call so that preferred analysts can ask questions first. Thus, the more critical questions appear rather at the end of the call. Supporting this presumption, we observe a more negative tone in later questions within a call.

unobservable factors by including several fixed effects: we remove common time trends, control for any time-constant firm-specific factors, and include management team fixed effects to absorb any management team specific factors such as the general style to respond to questions in conference calls or the taste of the management team for using finance terminology. As an additional robustness test, we remove all answers to non-finance-related questions and focus only on answers where the question contains at least one finance-related word, which ensures that the manager was not 'invited' by the questioner to use non-factual terminology.

All our findings indicate that market participants are well versed in financial jargon and hence, that the usage of financial jargon by managers positively affects the information environment. That is, financial jargon is an efficient tool of transferring information to market participants. Managers might obfuscate information by avoiding a precise language, i.e. by blathering. However, markets understand the lack of precise jargon as unwillingness to disclose the requested information, leading to higher uncertainty and a lower stock returns. This conclusion is further supported by evidence for a heterogeneous market reaction to the usage of jargon in response to contextual versus context-free questions. We find that financial jargon in response to finance-related questions is appreciated by the market, however, a higher $JargonRatio$ in response to finance-unrelated questions has no effect on stock prices. We conclude that market participants are well versed and, hence, appreciate the use of financial jargon, especially in response to topical questions.

The remainder of this paper is organized as follows: we review the relevant literature in Section 2. We describe the data used in Section 3, and lay the competing hypothesis to be analyzed out in Section 4. Section 5 presents the empirical model and our empirical results, respectively. Section 6 concludes.

2 Background and literature

Earnings conference calls are voluntary disclosures that aim to reduce information asymmetries among investors (Brown, Hillegeist and Lo, 2004).[3] The calls are typically attended by institutional investors who hold a large stake in the company and therefore wish to improve their 'understanding' of the company" (Barker, Hendry, Roberts and Sanderson, 2012), and by financial analysts, who base their earnings forecasts – in part – on information extracted from the call, thereby decreasing the dispersion among their forecasts (Bushee, Matsumoto and Miller, 2004). Earnings conference calls usually consist of two parts: they start with a short (prepared) presentation by the management, followed by a Q&A session. The Q&A, where the management answers questions from investors and analysts, is relatively more informative than the presentation, as it relies on spontaneous responses rather than a prepared text by the investor relation division (Matsumoto, Pronk and Roelofsen, 2011).

Textual analysis helps to derive information from earnings calls in addition to the reported numbers. Research in this area primarily focuses on the tone and vagueness of the language used by the management. For example, several paper show that the tone of managers used in the Q&A session of earnings calls has predictive power for cumulative abnormal returns (Price, Doran, Peterson and Bliss, 2012; Blau, DeLisle and Price, 2015; Druz, Petzev, Wagner and Zeckhauser, 2020), or that vague communication, as measured by a more frequent usage of words like 'vague' and 'uncertainty,' leads to a lower firm valuation (Loughran and McDonald, 2011; Dzieliński, Wagner and Zeckhauser, 2017). In addition, when classifying financial statements as 'deceptive' or 'truthful', it has been shown that linguistic methods are at least as predictive and often outperform accounting based measures (Larcker and Zakolyukina, 2012). This linguistic research often relies on glossaries that identify a specific topic or sentiment. The Harvard-IV and Lasswell

[3]There has been some earlier research analyzing a firm's decision to voluntarily disclose information in general (Skinner, 1994) or through hosting conference calls in particular (Frankel, Johnson and Skinner, 1999). Nowadays, however, it is standard to disclose information through earnings calls.

dictionaries, as part of the Harvard General Inquirer Word Lists, for example, help to identify general linguistic sentiments such as positivity/negativity. More finance-specific dictionaries have been developed by Henry (2008), Loughran and McDonald (2011) and Harvey (2016), where the latter captures factual finance terminology (financial jargon).

We further add to the literature focusing on the precise and efficient transfer of information, as well as the literature on linguistic complexity. It has been established that mangers steer information in an opportunistic manner. For example, Mayew (2008) show that managers choose more favorable analysts to ask questions on the call. Likewise, managers obscure inconvenient information or a poor performance by using a more complex language in written disclosures (Li, 2008; Bloomfield, 2008; Loughran and McDonald, 2014; Bonsall et al., 2017) or by avoiding factual, quantitative statements (Zhou, 2018).[4] The very extreme form of reducing the informativeness of a response is to openly refuse an answers (Hollander et al., 2010; Gow, Larcker and Zakolyukina, 2019). Hollander et al. (2010), for example, manually classify answers where managers withhold information, and show that markets interpret this silence negatively.

We add to this literature by exploring how financial markets interpret the usage of financial jargon. While jargon is in its nature complex, it remains an open question if its use is perceived as *obfuscation* or *information*. Bushee, Gow and Taylor (2018) also decompose the source of complexity into these two latent factors, arguing that complex responses to complex questions should be understood as information, whereas complex responses to simpler questions should be understood as obfuscation. While their paper investigates *when* grammatically complex phraseology is used to obfuscate, we ask *whether* financial jargon is a source of complexity that can be used to obfuscate.

[4]There is an ongoing discussion on the measure of linguistic complexity: while Li (2008) and Bloomfield (2008) use the popular Gunning (1952) Fog Index, this measure has been challenged by Loughran and McDonald (2014) as being "poorly specified in financial applications."

3 Data

3.1 Jargon Ratio

We collect transcripts of all earnings calls held by financial companies listed in the S&P 500 available from Thomson Reuters' 'StreetEvents' for the period 2002 to 2019.[5] These calls are released quarterly and usually take place on the same day as the corresponding earnings release.[6] As we focus on managements' spontaneous answers to investors' questions rather than the prepared presentation, we exclude all earnings calls without a Q&A session. We further drop calls with less than 500 words to mitigate any bias in our text measure, and remove answers with less than 20 words. This leaves us in our final dataset with 39,845 answered questions in 2,121 earnings calls for 63 financial firms.

We use the Hypertextual Finance Glossary by Campbell R. Harvey (2016) with more than 8,500 entries to identify financial jargon,[7] but remove common English stop-words that could bias our metric.[8] We then quantify our measure $JargonRatio$ for the management response to a question q in the earnings call by company i in quarter t by counting the number of finance-related words as a share of total words:

$$JargonRatio_{qit} = \frac{\text{Finance glossary words}_{qit}}{\text{Total words}_{qit}},$$

where $Finance\ Glossary\ Words$ is the frequency of words that belong to the Harvey finance glossary, and $Total\ Words$ is the number of all words in the answer to a specific question in the respective call.[9] For most of our analyses, we aggregate the jargon measure

[5] We classify financial firms based on the Fama-French industry classification (industry *44* and *47*). A detailed industry definition and corresponding SIC codes are available at: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_48_ind_port.html`.

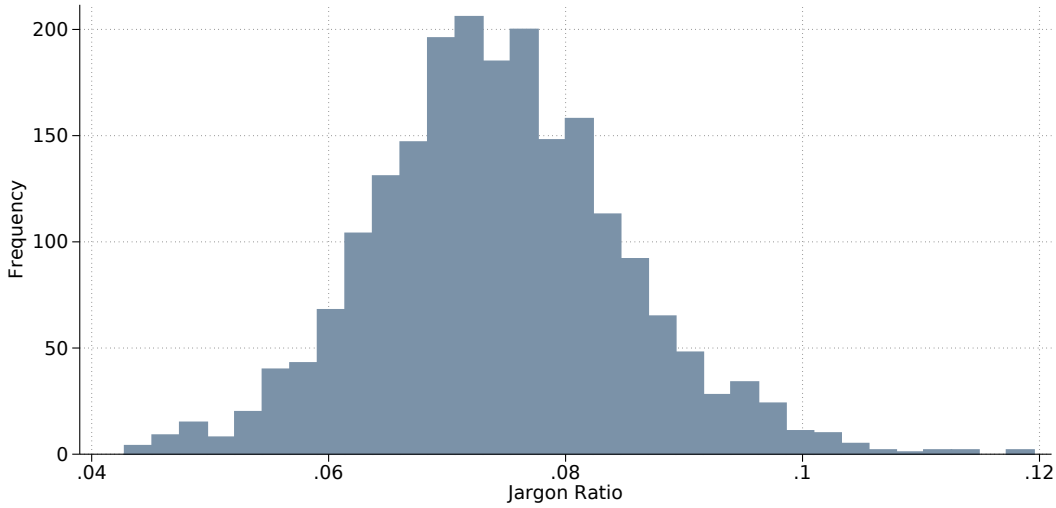
[6] 92.5% of all calls in our sample take place on the same day as the earnings release, while 7.3% take place one day after the earnings announcements. In only six cases is the call scheduled more than one day after the earnings announcement.

[7] The glossary is available at `http://people.duke.edu/~charvey/`

[8] For example, the words 'MY' or 'ARE' are defined in Harvey's finance glossary as "the two-character ISO 3166 country code for MALAYSIA" or "the three-character ISO 3166 country code for UNITED ARAB EMIRATES", respectively. See Table 1 for a list of all 45 English stop-words in Harvey's finance glossary.

[9] One could likewise define a $BlatheringScore$ as absence of jargon, $Blathering_{qit} = 1 -$

Figure 1: Histogram of the $JargonRatio$.



over all answers in a call, $JargonRatio_{it}$, defined as the ratio of finance-related words over total words in all management responses in the call by company i at time t.

$JargonRatio_{it}$ is distributed around a mean and median of 7.3% with a quite sizable heterogeneity of jargon across earnings calls, ranging from 4.2% to 13.1% (Figure 1).

Note again that we restrict our sample to financial firms. For financial firms, words from the contextual domain of finance targets both the business model and company financials, so that all factual information can be provided using financial jargon. Non-financial firms, however, might provide factual and relevant information on their business model, products and strategy using idiosyncratic terminology (e.g. iPad, Starlink, etc.) in addition to financial jargon.

3.2 Alternative speech characteristics

Investors recognize tone sentiment and vagueness of the language used in earnings calls, as well as the share of numbers relative to text.[10] We compute standard metrics from the literature to capture these alternative language characteristics.

$JargonRatio_{qit}$ and interpret the results in terms of non-answers.

[10]See, inter alia, Price et al. (2012), Blau et al. (2015), Brockman, Li and Price (2015) or Davis, Ge, Matsumoto and Zhang (2015) for evidence on tone sentiment, Dzieliński et al. (2017) for evidence on uncertainty, and Zhou (2018) for the link between numeric information in earnings calls and stock returns.

For tone, we count the number of words that appear on the positive and negative word list by Loughran and McDonald (2011), and define $Tone$ of the earnings call by company i in quarter t as the difference between positive and negative words relative to total words,

$$Tone_{it} = \frac{\text{Positive words}_{it} - \text{Negative words}_{it}}{\text{Total words}_{it}}.$$

For some analyses, we measure the tone of the question and quantify $Tone$ for the question q in the earnings call by company i at time t.

The uncertainty metric is also based on the Loughran and McDonald (2011) word list.[11] Uncertainty in the earnings call of company i at time t is defined as the ratio of uncertain words to total words,

$$Uncertainty_{it} = \frac{\text{Uncertain words}_{it}}{\text{Total words}_{it}}.$$

The word lists by Loughran and McDonald (2011) are widely recognized dictionaries that are used for sentiment analysis in a finance context, while the Hypertextual Finance Glossary by Harvey (2016) is used less frequently. One would think that there is a significant overlap in the dictionaries by Loughran and McDonald (2011) and Harvey (2016). The Loughran and McDonald (2011) word lists, however, contain only few financial words. In fact, just around 3% of the words on the Loughran and McDonald (2011) word lists are also in the Harvey (2016) finance glossary. These are 59 negative, 11 positive, and 14 uncertain words. Table 2 in the appendix summarizes the overlapping words.

We follow the approach in Zhou (2018) to generate a variable $Numbers$ that accounts for the usage of numbers in managements' answers relative to textual words. Specifically, we use a regular expression to capture numerical values and calculate $Numbers_{it}$ for the

[11] Note that this word list contains also the Loughran and McDonald (2011) weak modal word list.

earnings call of company i at time t as follows:

$$Numbers_{it} = \frac{\text{Number count}_{it}}{\text{Total words}_{it} + \text{Numbers count}_{it}}.$$

We further calculate the Gunning (1952) Fog index, which indicates how easy or difficult a text is to understand in a linguistic sense.[12] The Fog index is a function of the number of words per sentence (length of a sentence) and the share of complex words (words with more than two syllables) relative to total words, and has been commonly used to proxy for the linguistic complexity of a text. The Fog index for the earnings call of company i at time t is calculated as follows:

$$Fog_{it} = 0.4 \times \left(\frac{\text{Total words}_{it}}{\text{Total sentences}_{it}} + \frac{\text{Complex words}_{it}}{\text{Total words}_{it}} \right).$$

3.3 Earnings surprise and firm characteristics

We collect analyst data from IBES to calculate quarterly earnings surprises as the difference between the actual and consensus forecast earnings, divided by the share price five trading days prior to the announcement. Thus, any positive (negative) number indicates a better (worse) performance than expected. As in Dzieliński et al. (2017), we rank all firms' earnings surprises into deciles and categorize earnings surprises from 1 (most negative) to 5 (least negative) and from 6 (least positive) to 10 (most positive).

In addition, we collect quarterly balance sheet statistics (total assets, book equity, and return on equity) as well as banks' market capitalizations from Compustat and calculate firm characteristics as for example the book-to-market ratio and the natural logarithm of total assets.

[12] Note that the Fog index proxies linguistic complexity, while the jargon measure might be understood as factual complexity.

3.4 Cumulative abnormal returns

We obtain daily adjusted stock returns from CRSP and define daily abnormal returns in several ways. First, we calculate the abnormal return for stock i at time t as the adjusted return on the stock minus the index return on the S&P 500, $r_{i,t}^{abnormal} = r_{i,t} - r_{S\&P,t}$. Second, we replace the S&P 500 return with the Fama-French five-factor (2015) model return.[13] Third, as additional robustness tests, we (i) replace the S&P 500 counterfactual with the value weighted CRSP return (VWRETD) and (ii) use the Fama-French three-factor model (1993).

We aggregate abnormal returns from the day of the earnings call to the day thereafter to derive cumulative abnormal returns $CAR_{i,t}^{0;1}$:

$$CAR_{i,t}^{0;1} = \prod_{d=0}^{1} \left(1 + r_{i,t+d}^{abnormal}\right) - 1.$$

3.5 Option implied volatility

The implied volatility reflects the premium that investors are willing to pay for insuring against price movements in the underlying and thus proxies for investor uncertainty.[14] We collect the daily implied volatility $\sigma_{i,t}$ derived from liquid at-the-money options in tenures of 91 days from OptionMetrics. We calculate two measures to capture the instantaneous update of investors believes on future volatility after a conference call. The first approach follows Rogers et al. (2009), and compares the implied volatility of company i on the day just after the call to the implied volatility on the day just before the call,

$$IV_{i,t}^{-1;1} = \ln\left(\frac{\sigma_{i,t+1}}{\sigma_{i,t-1}}\right).$$

[13]The model is calibrated to 60 trading days preceding an earnings call, with data from the Fama-French data library at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html` (2019-01-08).

[14]See Rogers, Skinner and Buskirk (2009) for a discussion of why to prefer implied volatility over other possible measures such as realized volatility or the dispersion in analyst forecasts to measure investor uncertainty.

Second, we compare the change in $\sigma_{i,t}$ with a counterfactual change in the implied volatility, which we calculate as the average change in implied volatility for the 60 trading days proceeding the earnings call,

$$\Delta IV_{i,t} = \frac{\sigma_{i,t+1} - \sigma_{i,t-1}}{2} - \frac{\sigma_{i,t-60} - \sigma_{i,t-1}}{59}.$$

3.6 Descriptive statistics

Table 1 presents descriptive statistics for the variables in our analysis. On average, 7.4% of all words in management responses to investors' questions are classified as financial jargon. Of course, this low percentage of jargon is not surprising, as many non financial words are necessary to formulate a meaningful sentence. In our sample, measures for *Tone* and *Uncertainty* show an average of -1.1% and 1.7%, which are comparable to values found in the literature (Price et al., 2012; Dzieliński et al., 2017). We find in our sample of earnings calls a somewhat smaller share of numbers to text compared to the metric in Zhou (2018).[15] Finally, we note that all measures for cumulative abnormal returns share very similar distributional characteristics.

In terms of linear correlations, we observe a positive but weak correlation between our *JargonRatio* and the *Fog* index, which indicates that terminological and linguistic complexity are two separate aspects that need to be distinguished. It is also worth mentioning that our *JargonRatio* is only weakly correlated with the two speech characteristic measures, uncertainty and tone, and with the Zhou (2018) number measure[16].

[15] We get similar estimates as in Zhou (2018) if we count a sequence of digits that is separated by a non-numeric character as a new number, e.g. counting 100,000 as two numbers rather than one.

[16] The Pearson correlation between *JargonRatio* and *Fog* is -0.23, the correlation between *JargonRatio* and *Uncertainty*, *Tone*, or *Numbers* is -0.02, 0.14, and -0.13, respectively.

Table 1: Descriptive Statistics

Notes: This table presents descriptive statistics. *Jargon_All* is defined as the ratio between finance-related words to total words and *Jargon_Fin* is the same ratio but considering only answers to questions with at least one finance-related phrase. *Tone* represents the difference between positive to negative words as a ratio to total words. *Uncertainty* measures the share of words classified as vague to total words. *EarnSurp* represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). *Numbers* is the share of numbers to the sum of total words and numbers. *Fog* is the Gunning-Fog Index as a measure of linguistic complexity. $CAR^{0;1}$ ($CAR^{2;60}$) is the cumulative abnormal returns in the [0;1] ([2;60]) interval around the earnings call, with abnormal returns measured as stock return over the S&P 500 index return. $VW-CAR$ denotes the abnormal return over the CRSP value weighted return. $FF3-CAR$ and $FF5-CAR$ use Fama-French three (1993) and five (2015) factor model returns respectively. $IV^{-1;1}$ and $\Delta\ IV$ are the change in implied volatility as defined in subsection 3.5. *BTM* defined as total Common/Ordinary Equity divided by the market value of equity. *ln(Assets)* is the natural logarithm of total assets. *RoE* denotes return on equity. All return variables are winsorized at the 1/99% percentiles. Panel B shows the *JargonRatio* and *Tone* variable at the Q&A level, where $Tone_Q$ is the tone measure for analysts' questions.

Variable	Obs.	Mean	Std. Dev.	Min	P10	P50	P90	Max
Panel A: Firm-quarter data								
$Jargon_{All}$	2,121	.074	.011	.043	.062	.074	.088	.12
$Jargon_{Fin}$	2,121	.075	.011	.043	.062	.075	.089	.13
$Jargon_{NonFin}$	1,982	.061	.03	-.00029	.022	.061	.093	.25
$Tone$	2,121	-.011	.0098	-.063	-.024	-.01	.00063	.023
$Uncertainty$	2,121	.017	.0054	.003	.01	.016	.024	.04
$EarnSurp$	2,121	5.8	2.9	1	2	6	10	10
$Numbers$	2,121	.011	.0048	.00099	.005	.01	.017	.032
FOG	2,121	16	2.7	10	13	15	19	31
$CAR^{0;1}$	2,121	.000011	.065	-.49	-.053	-.00046	.055	1.2
$VW-CAR^{0;1}$	2,121	-.00045	.051	-.18	-.053	-.00055	.054	.17
$FF3-CAR^{0;1}$	2,121	-.001	.043	-.16	-.047	-.00097	.047	.14
$FF5-CAR^{0;1}$	2,121	-.00094	.041	-.14	-.045	-.0009	.043	.13
$CAR^{2;60}_{t-1}$	2,121	.0036	.099	-.33	-.098	.003	.099	.37
$IV^{-1;1}$	1,931	.96	.063	.81	.89	.96	1	1.2
ΔIV	1,931	-.0054	.014	-.066	-.017	-.0049	.0061	.054
BTM	2,121	.74	.53	-.18	.25	.64	1.3	5
ROE	2,121	.013	.029	-.17	.0023	.018	.028	.07
$Ln(Assets)$	2,121	11	1.8	6.2	9	12	14	15
Panel B: Q&A-level data								
$Jargon_{All}$	39,845	.072	.037	0	.029	.069	.12	.36
$Jargon_{Fin}$	38,101	.073	.036	0	.03	.07	.12	.36
$Tone_Q$	39,845	-.015	.039	-.52	-.063	-.013	.027	.24

4 Hypothesis

Our empirical analysis aims to explore whether the usage of financial jargon eases or retards the information flow from managers to investors, and whether it affects investors' uncertainty.[17] By definition, jargon helps to efficiently and precisely transfer information, as long as all participants are sufficiently versed in the specific terminology. Likewise, jargon-free language makes communication less factual, as it reduces the 'news part' of a response without adding content to the answer, thereby increasing information processing costs.

Jargon Efficiency Hypothesis Jargon is the most efficient way to transfer information as long as the vocabulary is well understood. If market participants understood financial jargon, avoiding factual language would make a message less precise. In the context of earnings conference calls, more background noise in the information increases uncertainty about the true message that the management intends to convey. Likewise, if financial jargon is an efficient tool to transfer information, investors might interpret a lack of jargon as a negative signal for the managers' private information. More specifically, a manager's perceived attempt to obfuscate unfavourable information by *avoiding* a precise language, leads to uncertainty, which in turn results in a higher risk premium demanded by investors and to a higher premium that investors are willing to pay for insurance against adverse price movements in the stock. This line of reasoning leads us to the *"Jargon Efficiency Hypothesis:"*

Hypothesis ('Jargon Efficiency') *Investors are versed in financial terminology, so that financial jargon increases the efficiency of information flow in conference calls and thereby helps to reduce information asymmetries.*

[17] We use the implied volatility as a direct measure of investor uncertainty as well as stock returns as an indirect measure. See Andrei and Hasler (2014) for a theoretical asset pricing model showing the positive impact of uncertainty on risk premium and Bali, Brown and Tang (2017) for empirical evidence for this prediction. For a discussion on the impact of policy uncertainty or for the effect of disagreement as some special cases of uncertainty on asset prices, see Pástor and Veronesi (2013), Liu, Shu and Wei (2017), or Carlin, Longstaff and Matoba (2014).

Jargon Complexity Hypothesis The *Jargon Complexity Hypothesis* offers a conflicting view to the *Jargon Efficiency Hypothesis.* Jargon is designed by introducing a specific and potentially unique set of vocabulary. Therefore, jargon would increase the complexity of a conversation, if it was not equally understood by market participants – think about that medical diagnosis conveyed to a patient in Latin. In economics, there exists evidence that linguistic complexity is employed to obfuscate unfavorable information (Li, 2008; Bloomfield, 2008; Loughran and McDonald, 2014; Bonsall et al., 2017). While their research relies on measures for grammatical complexity such as the Gunning (1952) Fog Index, investors would perceive jargon itself as complex if they were not able to understand financial terminology. If that was the case, managers would obfuscate unfavorable information by *using* jargon. Likewise, investors want to be compensated for the uncertainty arising from a too complex language, and more jargon should lead to a higher premium that investors are willing to pay for insuring against price movements in the underlying. This line of reasoning leads us to the *"Jargon Complexity Hypothesis:"*

Hypothesis ('Jargon Complexity') *Financial terminology is difficult to process for investors, so that financial jargon increases the complexity of a message and thereby retards the reduction of information asymmetries.*

5 Empirical analysis

The two competing hypotheses outlined in Section 4 provide several predictions that are analyzed in this section. We first investigate when managers use financial jargon in their responses in Section 5.1, followed by an analysis of market reactions to the usage of jargon in Section 5.2 (stock returns) and Section 5.3 (implied volatility). Finally, we analyze the heterogeneous impact of jargon on the transfer of information for situations when jargon is requested or not in Section 5.4.

5.1 Usage of jargon

It is well documented in the literature that managers strategically manage the flow of information. Jargon as a relevant information transmission tool could either improve (*Jargon Efficiency Hypothesis*) or retard (*Jargon Complexity Hypothesis*) the flow of information. We therefore analyze in which situations jargon is used throughout an earnings call and what motivates managers to refrain from factual language.

First, favorable analysts are typically allowed to ask questions earlier in a call (Cohen et al., 2013), and managers might want to be most precise to these questions. The *Jargon Efficiency Hypothesis* would imply that managers use more jargon responding to these questions, and avoid jargon in response to disadvantageous and tougher questions, which they want to evade answering.[18] Contrary to that, the *Jargon Complexity Hypothesis* would predict less jargon in answers to early questions in order to avoid complexity, and more jargon-complex responses to the critical questions at the end of the call.

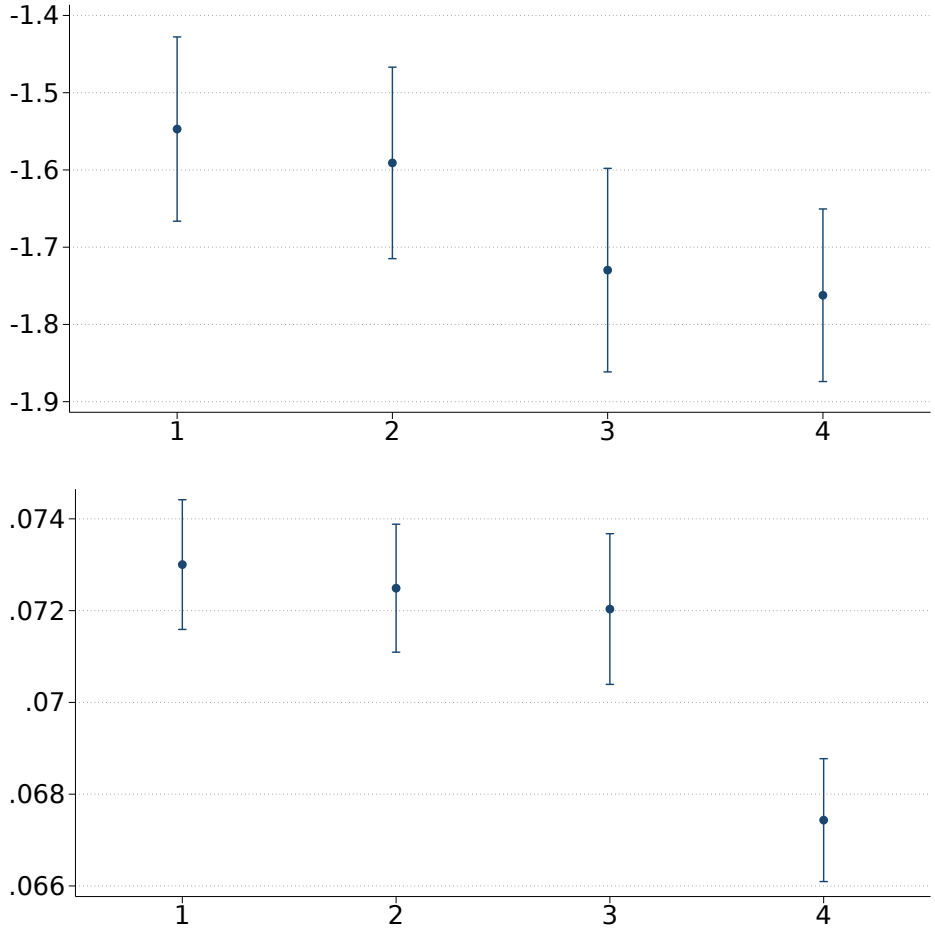
Figure 2 displays the variation of jargon *within* earnings calls. We divide the number of questions in each conference call into quartiles according to the order of their appearance and show the average negative tone of analysts' questions (Figure 2, upper panel) as well as the average $JargonRatio$ in managements' responses (Figure 2, lower panel). We observe that questions become more negative the later they are asked in the call. At the same time, we observe less financial jargon in response to later questions.

We investigate the statistical significance of this observation by analyzing calls at the most granular level, i.e. single questions and their respective answers. In particular, we correlate the usage of jargon in a response to question q within the earnings call by company i at time t to the tone of the respective question,

$$Jargon_{qit} = \alpha + \beta_1 \cdot Tone_{qit} + \mu_{it} + \epsilon_{qit}. \tag{1}$$

[18]The easiest way not to answer a question would be to refuse the answer directly. However, as Hollander et al. (2010) show that silence is interpreted negatively by investors, managers might prefer to refuse a response in a less obvious manner.

Figure 2: Tone of questions (upper panel) and *JargonRatio* of managements' answers (lower panel) for chronological order of questions in conference calls.



Note that the unit of observation is an answer to a question, which allows us to use variation within earnings calls, i.e. we control for many observable and unobservable firm, management, or earnings call-specific factors using call fixed effects, μ_{it}. To account for autocorrelations in the errors, we employ two-way clustering (Cameron, Gelbach and Miller, 2011) and cluster standard errors at the firm and time dimensions.

Table 2, column (1) summarizes the result of this exercise. In line with the descriptive analysis in Figure 2, we find a significant positive coefficient for the tone of questions, implying that the management uses less jargon in response to negative questions. In Table 2, column (2), we redo the exercise but focus only on answers where the question contains at least one finance-related word to ensure that the manager was not 'invited' by the questioner to use non-finance terminology.

Next, we analyze the usage of jargon after a quarter of bad success and poor performance. To this extent, we regress our $JargonRatio$ of company i at time t on the cumulative abnormal stock return in the quarter prior to the earnings call,

$$Jargon_{it} = \alpha + \beta_1 \cdot CAR_{it-1} + X_{it} + \mu_i + \mu_t + \mu_m + \epsilon_{it}. \tag{2}$$

We control for firm-quarter characteristics X_{it} ($EarningSurpise$, BTM, RoE, and $ln(Assets)$) and absorb firm-specific, time-specific, and management-specific variation using fixed effects. Standard errors are again clustered at the firm and time dimensions.

Results are presented in Table 2 for the $JargonRatio$ in all responses [column (3)] and for the $JargonRatio$ in only those responses where the question was related to a finance content [column (4)]. We find a significantly larger amount of jargon after quarters with high returns. Note that this result again holds true if we exploit only the variation within a management team, i.e. if we compare the same management's usage of jargon after a good and after a bad quarter.

To sum up, we find that managers avoid jargon in response to critical questions and in quarters following a poor performance. This first set of results point toward the *Jargon Efficiency Hypothesis:* managers deliberately obfuscate information by avoiding a factual language and precise terminology.

5.2 Jargon and stock returns

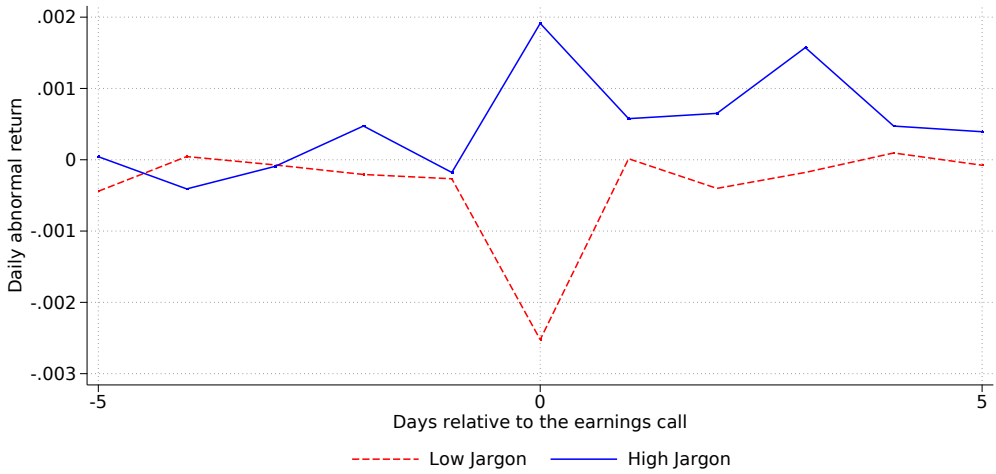
This section investigates how stock markets perceive the usage of financial jargon (or lack thereof). First descriptive evidence is presented in Figure 3, where we show daily abnormal returns around the days of earnings calls for firms on the high and low side of the $JargonRatio$ distribution. The returns of these two groups show hardly any difference in the days prior to and after the call. On the event day, however, stock returns of high and low jargon firms diverge, and returns of high jargon firms outperform returns of low jargon firms.

Table 2: Incentives to use jargon

Notes: Column 1 shows the regression of the *JargonRatio* in answers by the management on the *Tone* of the question. Column 2 repeats the same regression but only for answers to those questions that have at least one financial-related word. In both columns, we include earnings call fixed effect. In columns 3 and 4, the dependent variables is the *JargonRatio* in all answers and in answers to finance-related questions within a call. $CAR_{t-1}^{2;60}$ is the cumulative abnormal return over S&P500 from the 2nd day until the 60th day after the previous earnings conference call and it captures the stock performance in the previous quarter. Firm control variables include BTM, RoE, $ln(Assets)$, and $EarnSurp$. BTM is defined as total Common/Ordinary Equity divided by the market value of equity. RoE denotes return on equity. $ln(Assets)$ is the natural logarithm of total assets. $EarnSurp$ represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). t-statistics are given in parentheses; Standard errors are clustered on the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Observation Unit	Answer		Earnings Call	
	Jargon		*Jargon*	
	_All	*_Fin*	*_All*	*_Fin*
	(1)	(2)	(3)	(4)
$Tone_Q$	0.0097**	0.0075*		
	(2.31)	(1.85)		
$CAR_{t-1}^{2;60}$			0.0065***	0.0053**
			(3.19)	(2.32)
N	39845	38101	1910	1910
R^2	0.094	0.096	0.492	0.484
Firm_Controls	Implied	Implied	Yes	Yes
Firm_FE	Implied	Implied	Yes	Yes
Management_FE	Implied	Implied	Yes	Yes
Quarter_FE	Implied	Implied	Yes	Yes
EarningsCall_FE	Yes	Yes		

Figure 3: Average daily abnormal returns in a 5-day window before and after an earnings call for firms with above and below median within firm usage of jargon.



In order to rule out that this descriptive observation is driven by potential confounding factors, we analyze the stock market reaction to the usage of jargon more formally in a regression setting and model cumulative abnormal returns of firm i in quarter t:

$$CAR_{it} = \alpha + \beta_1 \cdot Jargon_{it} + \delta \cdot X_{it} + \theta \cdot Z_{it} + \mu_i + \mu_m + \mu_t + \epsilon_{it}. \tag{3}$$

We control for a number of important variables to ensure that the effect of jargon is not driven by outside factors. First, we control with X_{it} for several time-varying firm characteristics that might correlate with investor's demand for information, as for example bank size and the book-to-market ratio. It also includes a metric that captures investors' expectations of future earnings, and the return on equity in the past quarter. Both analysts' expectations and the RoE absorb concerns of a spurious relation between the usage of jargon and abnormal returns due to poor performance in the previous quarter.

Furthermore, Z_{it} controls for earnings call specific variables that have been shown to impact returns during and after a call. This vector consists of $Tone$, $Uncertainty$, $Numbers$, and the Gunning (1952) Fog Index.

Finally, we remove all observable and unobservable firm-specific time-constant variation by including firm fixed effects, μ_i, as well as common time trends by including time

(quarter-year) fixed effects, μ_t. We further include management team fixed effects, μ_m, in order to absorb all time-constant management-specific variation that can neither be explained by the current and future performance of the company nor by any strategic incentives (Davis et al., 2015). This enables us to separate the usage of jargon from personal specific unobservable time-constant characteristics. To account for autocorrelations in the errors, we employ two-way clustering (Cameron et al., 2011) and cluster standard errors at the firm and time dimensions.

Table 3 displays the results of the regression using the S&P 500 index return (columns (1)-(3)) and the Fama-French five-factor model returns (columns (4)-(6)) as a predictor of normal returns. We find a positive and highly significant coefficient for the $JargonRatio$, highlighting the positive effect that jargon has on short-term cumulative abnormal returns. In the most saturated regression in columns (3) and (6), a one standard deviation increase in $JargonRatio$ increases the short term cumulative abnormal returns on average by 0.38(0.22) percentage points, ceteris paribus. This result again supports the *Jargon Efficiency Hypothesis,* suggesting that investors understand and value financial jargon.

As discussed in Section 3.4, we ensure that the results are not biased by our definition of abnormal returns and use with a value-weighted cumulative abnormal return as well as the Fama-French three factor model various alternative counterfactuals for the 'normal' return. All results remain virtually unchanged and are shown in the appendix, Table 1.

5.3 Jargon and option implied volatility

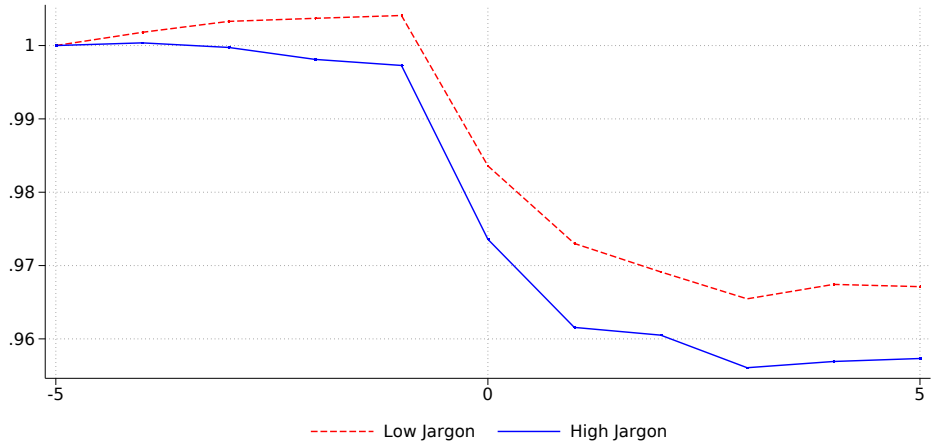
The empirical asset pricing literature suggests that the effect on stock returns is driven by investor uncertainty. We therefore analyze in this section whether the usage of jargon affects implied volatility as a direct measure for investor uncertainty.

Figure 4 shows the implied volatility for earnings calls with a high versus a low $JargonRatio$. Not surprising, implied volatility is highest just before the conference call and drops on the day of the call, when private information about the firm is revealed. The decrease of investor uncertainty, however, is much stronger for earnings call with a

Table 3: Jargon and the short-term returns ($CAR^{0;1}$)
Notes: OLS regressions for Equation (3). The dependent variable is the abnormal return over S&P500 and the Fama-French five factor model (2015) return cumulated from the day of the earnings call to the day thereafter, $CAR^{0;1}$ and $FF5 - CAR^{0;1}$, respectively. *Jargon_All* is the ratio of finance-related words in managements' answers. *Tone* is the difference between the ratio of positive words minus the ratio of negative words in managements' answers. *Uncertainty* is the ratio of uncertain words to total words. List of positive, negative, and uncertain words are from the Loughran and McDonald (2011) word-lists. *Numbers* is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). *Fog* is the Gunning-Fog Index as a measure of linguistic complexity. Firm control variables include *EarnSurp*, *BTM*, *RoE*, and *ln(Assets)*. *EarnSurp* represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). *BTM* is defined as total Common/Ordinary Equity divided by the market value of equity. *RoE* denotes return on equity. *ln(Assets)* is the natural logarithm of total assets. t-statistics are given in parentheses; Standard errors are clustered on the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	$CAR^{0;1}$			$FF5 - CAR^{0;1}$		
	(1)	(2)	(3)	(4)	(5)	(6)
Jargon_All	0.2261**	0.2278**	0.3405**	0.1716*	0.1741**	0.2007**
	(2.05)	(2.31)	(2.51)	(1.88)	(2.07)	(2.20)
Tone		0.0695	0.1797		0.2221**	0.3063**
		(0.59)	(1.44)		(2.50)	(2.53)
Uncertainty		-0.1034	0.1921		-0.0333	0.0532
		(-0.29)	(0.46)		(-0.14)	(0.19)
Numbers		0.1907	-0.1287		0.2248	0.0672
		(0.63)	(-0.25)		(1.26)	(0.27)
Fog		-0.0005	-0.0006		-0.0005	-0.0006
		(-0.69)	(-0.68)		(-1.21)	(-0.99)
N	2121	2121	1910	2121	2121	1910
R^2	0.105	0.106	0.182	0.117	0.120	0.160
Firm_Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm_FE	Yes	Yes	Yes	Yes	Yes	Yes
Management_FE	No	No	Yes	No	No	Yes
Quarter_FE	Yes	Yes	Yes	Yes	Yes	Yes

Figure 4: Daily implied volatility of at the money options in a 5-day window before and after an earnings call for firms with above and below median within firm usage of jargon.



high *JargonRatio*.

We test this visual result formally in Table 4, where we analyze the short term change as well as the abnormal change in implied volatility, similar to the analysis of abnormal returns described in Equation 3. We find that a high *JargonRatio* leads to a significantly lower implied volatility, which again is in line with the *Jargon Efficiency Hypothesis:* Investors understand financial terminology, so that the use of jargon improves the flow of information and reduces information asymmetries between the management and investors.

5.4 Heterogeneous impact of jargon

We finally ask whether there is a heterogeneous impact of jargon on financial markets for situations where financial jargon is demanded by investors. In the spirit of Bushee et al. (2018), who show that complex responses to complex questions should be understood as information, whereas complex responses to simpler questions should be understood as obfuscation, we differentiate between finance-related questions and non-financial questions. If market participants understood financial jargon, the *Jargon Efficiency Hypothesis* would predict that the usage of jargon never increases uncertainty. Moreover, financial jargon would be the most precise and efficient way of transferring information

Table 4: Jargon and option implied volatility ($\Delta\ IV$ and $IV^{-1;1}$)
Notes: OLS regressions with the dependent variable $IV^{-1;1}$ in the columns (1) to (3), and ΔIV in the columns (4) to (6), indicating the change in implied volatility as defined in subsection 3.5. *Jargon_All* is defined as the ratio between finance-related words to total words in managements' answers. *Tone* is the difference between the ratio of positive words minus the ratio of negative words in managements' answers. *Uncertainty* is the ratio of uncertain words to total words. List of positive, negative, and uncertain words are from the Loughran and McDonald (2011) word-lists. *Numbers* is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). *Fog* is the Gunning-Fog Index as a measure of linguistic complexity. Firm control variables include *EarnSurp*, *BTM*, *RoE*, and *ln(Assets)*. *EarnSurp* represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). *BTM* is defined as total Common/Ordinary Equity divided by the market value of equity. *RoE* denotes return on equity. *ln(Assets)* is the natural logarithm of total assets. *t*-statistics are given in parentheses; Standard errors are clustered on the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	$\Delta\ IV$			$IV^{-1;1}$		
	(1)	(2)	(3)	(4)	(5)	(6)
Jargon_All	-0.0689*	-0.0664**	-0.0674*	-0.2878**	-0.2914**	-0.3362**
	(-1.91)	(-2.02)	(-1.77)	(-2.02)	(-2.20)	(-2.02)
Tone		0.0267	0.0316		-0.0320	-0.0797
		(0.85)	(0.85)		(-0.19)	(-0.41)
Uncertainty		0.0009	-0.0199		-0.0417	0.0309
		(0.01)	(-0.16)		(-0.11)	(0.07)
Numbers		0.0388	0.1096		0.2669	0.3657
		(0.61)	(1.36)		(0.89)	(0.91)
Fog		-0.0002	-0.0001		-0.0005	-0.0003
		(-0.78)	(-0.58)		(-0.83)	(-0.48)
N	1931	1931	1725	1931	1931	1725
R^2	0.160	0.161	0.208	0.179	0.180	0.230
Firm_Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm_FE	Yes	Yes	Yes	Yes	Yes	Yes
Management_FE	No	No	Yes	No	No	Yes
Quarter_FE	Yes	Yes	Yes	Yes	Yes	Yes

in response to topical questions.

We measure for each earnings call the $JargonRatio$ of responses to finance-related questions and non-finance questions, i.e. questions that do not include a finance-specific word, and differentiate the effect of the $JargonRatio$ on abnormal returns and implied volatility. The results for abnormal returns are shown in Table 5 (CAR using market return and Fama-French five-factor model),[19] and results for implied volatility are shown in Table 6.

We find that a strong usage of jargon increases returns and decreases implied volatility only in responses to finance-related questions. The usage of jargon to finance-unrelated questions, however, has no statistically significant effect on returns or implied volatility.

[19]See Table 2 in the appendix for results of CAR using value-weighted market return and Fama-French three-factor model.

Table 5: Jargon and the short-term returns ($CAR^{0;1}$)

Notes: OLS regressions for Equation (3). The dependent variable is the abnormal return over S&P500 and the Fama-French five factor model (2015) return cumulated from the day of the earnings call to the day thereafter, $CAR^{0;1}$ and $FF5 - CAR^{0;1}$, respectively. $Jargon\_Fin$ is the ratio of finance-related words in managements' answers to the questions that have at least a word categorised as financial-related. $Jargon\_NonFin$ is the same ratio for the management's answers to the questions with no finance-related word. Financial jargon is identified using Harvey (2016) finance glossary. $Tone$ is the difference between the ratio of positive words minus the ratio of negative words in managements' answers. $Uncertainty$ is the ratio of uncertain words to total words. List of positive, negative, and uncertain words are from the Loughran and McDonald (2011) word-lists. $Numbers$ is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). Fog is the Gunning-Fog Index as a measure of linguistic complexity. Firm control variables include $EarnSurp$, BTM, RoE, and $ln(Assets)$. $EarnSurp$ represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). BTM is defined as total Common/Ordinary Equity divided by the market value of equity. RoE denotes return on equity. $ln(Assets)$ is the natural logarithm of total assets. t-statistics are given in parentheses; Standard errors are clustered on the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	$CAR^{0;1}$			$FF5 - CAR^{0;1}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$Jargon\_Fin$	0.2446**	0.2563***	0.4210***	0.1705*	0.1811**	0.2092**
	(2.15)	(2.71)	(2.73)	(1.77)	(2.02)	(2.01)
$Jargon\_NonFin$	-0.0017	-0.0035	-0.0195	-0.0201	-0.0225	-0.0108
	(-0.03)	(-0.05)	(-0.26)	(-0.49)	(-0.56)	(-0.25)
$Tone$		0.0987	0.2250		0.2635**	0.3620***
		(0.77)	(1.59)		(2.60)	(2.67)
$Uncertainty$		-0.0624	0.2556		-0.0140	0.0729
		(-0.17)	(0.59)		(-0.06)	(0.26)
$Numbers$		0.1082	-0.1806		0.1940	0.0670
		(0.36)	(-0.34)		(1.11)	(0.28)
Fog		-0.0006	-0.0007		-0.0006	-0.0007
		(-0.51)	(-0.83)		(-1.48)	(-1.14)
N	1981	1981	1776	1981	1981	1776
R^2	0.110	0.110	0.187	0.124	0.128	0.167
Firm_Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm_FE	Yes	Yes	Yes	Yes	Yes	Yes
Management_FE	No	No	Yes	No	No	Yes
Quarter_FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 6: Jargon and option implied volatility ($\Delta\ IV$ and $IV^{-1;1}$)

Notes: OLS regressions with the dependent variable $IV^{-1;1}$ in the columns (1) to (3), and ΔIV in the columns (4) to (6), indicating the change in implied volatility as defined in subsection 3.5. $Jargon\_Fin$ is the ratio of finance-related words in managements' answers to the questions that have at least a word categorized as financial-related. $Jargon\_NonFin$ is the same ratio for the management's answers to the questions with no finance-related word. Financial jargon is identified using Harvey (2016) finance glossary. $Tone$ is the difference between the ratio of positive words minus the ratio of negative words in managements' answers. $Uncertainty$ is the ratio of uncertain words to total words. List of positive, negative, and uncertain words are from the Loughran and McDonald (2011) word-lists. $Numbers$ is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). Fog is the Gunning-Fog Index as a measure of linguistic complexity. Firm control variables include $EarnSurp$, BTM, RoE, and $ln(Assets)$. $EarnSurp$ represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). BTM is defined as total Common/Ordinary Equity divided by the market value of equity. RoE denotes return on equity. $ln(Assets)$ is the natural logarithm of total assets. t-statistics are given in parentheses; Standard errors are clustered on the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	$\Delta\ IV$			$IV^{-1;1}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$Jargon\_Fin$	-0.0558*	-0.0547**	-0.0604**	-0.1963	-0.2050*	-0.2500**
	(-1.99)	(-2.08)	(-2.21)	(-1.58)	(-1.88)	(-2.21)
$Jargon\_NonFin$	-0.0085	-0.0090	-0.0121	-0.0471	-0.0491	-0.0644
	(-0.70)	(-0.71)	(-0.86)	(-1.10)	(-1.13)	(-1.20)
$Tone$		0.0318	0.0272		-0.0110	-0.1127
		(0.95)	(0.67)		(-0.06)	(-0.56)
$Uncertainty$		0.0119	-0.0028		0.0605	0.1913
		(0.11)	(-0.02)		(0.16)	(0.44)
$Numbers$		0.0710	0.1232		0.3995	0.4101
		(1.24)	(1.35)		(1.31)	(0.90)
Fog		-0.0002	-0.0001		-0.0006	-0.0004
		(-0.94)	(-0.50)		(-0.91)	(-0.70)
N	1794	1794	1593	1794	1794	1593
R^2	0.163	0.164	0.211	0.180	0.181	0.234
Firm_Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm_FE	Yes	Yes	Yes	Yes	Yes	Yes
Management_FE	No	No	Yes	No	No	Yes
Quarter_FE	Yes	Yes	Yes	Yes	Yes	Yes

6 Conclusion

Asymmetric information urges companies to regularly disclose information to their investors (Healy and Palepu, 2001), commonly via quarterly earnings calls. In this study, we analyze whether market participants are well versed in financial jargon and thus, whether the usage of jargon increases efficiency in communication or retards the flow of information.

Earnings conference calls of financial firms provide an ideal setting for the question at hand. In contrast to other industries, calls of financial firms have a semantically homogeneous context, where any discussion on either financial health, future strategy or new products always involves finance terminology. We are able to identify this jargon with the use of the Hypertextual Finance Glossary by Harvey (2016), without the risk of falsely classifying the industry specific or idiosyncratic jargon of non-financial firms as uninformative.

Our results show that managers avoid the usage of financial jargon, i.e. a clear and factual statement, in disadvantageous situations. In particular, they use less financial jargon in response to more negative questions and after a quarter of poor performance. Moreover, we document that the usage of jargon decreases the implied volatility, and results in a higher abnormal stock return after the conference call. It is, however, only the usage of jargon in response to contentual questions that affects stock prices, but not the usage of jargon to finance-unrelated questions.

These findings strongly support the *"Jargon Efficiency Hypothesis:"* market participants understand financial jargon and expect its usage as a tool for efficient information transfer. The absence of jargon, i.e. a less factual language, is punished by market participants. Similar to the open refusal to respond in Hollander et al. (2010), this 'hidden' withholding of information is interpreted as obfuscation.

The identification strategy in this paper relies on the specific linguistic context of financial firms, but our results invite future research to extend the scope and usage of

the *JargonRatio* in different ways. For example, Barth, Mansouri and Woebbeking (2020) use several linguistic metrics, including the *JargonRatio*, to train a machine learning model for the identification of non-answers. We also leave it to future research to investigate, whether financial markets are versed to interpret financial jargon, but lack the understanding of jargon of other industries.

References

Andrei, D., Hasler, M., 2014. Investor Attention and Stock Market Volatility. Review of Financial Studies 28, 33–72.

Bali, T.G., Brown, S.J., Tang, Y., 2017. Is economic uncertainty priced in the cross-section of stock returns? Journal of Financial Economics 126, 471 – 489.

Barker, R., Hendry, J., Roberts, J., Sanderson, P., 2012. Can company-fund manager meetings convey informational benefits? exploring the rationalisation of equity investment decision making by uk fund managers. Accounting, Organizations and Society 37, 207 – 222.

Barth, A., Mansouri, S., Woebbeking, F., 2020. "let me get back to you - a machine learning approach to measuring non-answers. Proceedings of Paris December 2020 Finance Meeting EUROFIDAI - ESSEC, Available at SSRN: 3567724 .

Blau, B.M., DeLisle, J.R., Price, S.M.K., 2015. Do sophisticated investors interpret earnings conference call tone differently than investors at large? Evidence from short sales. Journal of Corporate Finance 31, 203–219. URL: `http://dx.doi.org/10.1016/j.jcorpfin.2015.02.003`, doi:`10.1016/j.jcorpfin.2015.02.003`.

Bloomfield, R., 2008. Discussion of "annual report readability, current earnings, and earnings persistence". Journal of Accounting and Economics 45, 248–252.

Bonsall, S.B., Leone, A.J., Miller, B.P., Rennekamp, K., 2017. A plain english measure of financial reporting readability. Journal of Accounting and Economics 63, 329–357.

Brockman, P., Li, X., Price, S.M., 2015. Differences in conference call tones: Managers vs. analysts. Financial Analysts Journal 71, 24–42.

Brown, S., Hillegeist, S.A., Lo, K., 2004. Conference calls and information asymmetry. Journal of Accounting and Economics 37, 343–366.

Bushee, B.J., Gow, I.D., Taylor, D.J., 2018. Linguistic complexity in firm disclosures: Obfuscation or information? Journal of Accounting Research 56, 85–121.

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, 617–643.

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2011. Robust inference with multiway clustering. Journal of Business & Economic Statistics 29, 238–249.

Carlin, B.I., Longstaff, F.A., Matoba, K., 2014. Disagreement and asset prices. Journal of Financial Economics 114, 226 – 238.

Cohen, L., Lou, D., Malloy, C., 2013. Playing favorites: How firms prevent the revelation of bad news. Working Paper. National Bureau of Economic Research.

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, 639–673. URL: http://dx.doi.org/10.1007/s11142-014-9309-4, doi:10.1007/s11142-014-9309-4.

Druz, M., Petzev, I., Wagner, A.F., Zeckhauser, R.J., 2020. When managers change their tone, analysts and investors change their tune. Financial Analysts Journal 76, 47–69. doi:10.1080/0015198X.2019.1707592.

Dzieliński, M., Wagner, A.F., Zeckhauser, R.J., 2017. Straight talkers and vague talkers: The effects of managerial style in earnings conference calls. Working Paper. National Bureau of Economic Research.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

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

Gow, I.D., Larcker, D.F., Zakolyukina, A.A., 2019. Non-answers during conference calls. Working Paper. Chicago Booth.

Gunning, R., 1952. The technique of clear writing .

Harvey, C.R., 2016. Harvey's hypertextual finance glossary. URL: https://people. duke. edu/˜ charvey/Classes/wpg/bfglosm. htm [Version 8 November 2016.] .

Healy, P.M., Palepu, K.G., 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. Journal of Accounting and Economics 31, 405–440.

Henry, E., 2008. Are Investors Influenced By How Earnings Press Releases Are Written? Journal of Business Communication 45, 363–407. URL: `http://job.sagepub.com/content/45/4/363.abstract{%}5Cnhttp://job.sagepub.com/cgi/doi/10.1177/0021943608319388`, doi:`10.1177/0021943608319388`.

Hollander, S., Pronk, M., Roelofsen, E., 2010. Does silence speak? an empirical analysis of disclosure choices during conference calls. Journal of Accounting Research 48, 531–563.

Larcker, D.F., Zakolyukina, A.A., 2012. Detecting deceptive discussions in conference calls. Journal of Accounting Research 50, 495–540.

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

Liu, L.X., Shu, H., Wei, K.J., 2017. The impacts of political uncertainty on asset prices: Evidence from the bo scandal in china. Journal of Financial Economics 125, 286 – 310.

Loughran, T., McDonald, B., 2011. When is a Liability not a Liability? Journal of Finance 66, 35 – 65.

Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures. Journal of Finance 69, 1643–1671.

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, 1383–1414.

Mayew, W.J., 2008. Evidence of management discrimination among analysts during earnings conference calls. Journal of Accounting Research 46, 627–659.

Pástor, L., Veronesi, P., 2013. Political uncertainty and risk premia. Journal of Financial Economics 110, 520 – 545.

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

Rogers, J.L., Skinner, D.J., Buskirk, A.V., 2009. Earnings guidance and market uncertainty. Journal of Accounting and Economics 48, 90 – 109. URL: http://www.sciencedirect.com/science/article/pii/S0165410109000342, doi:https://doi.org/10.1016/j.jacceco.2009.07.001.

Skinner, D.J., 1994. Why firms voluntarily disclose bad news. Journal of Accounting Research 32, 38–60.

Zhou, D., 2018. Do numbers speak louder than words? Available at SSRN 2898595 .

A Sample of Financial Firms

Table 1: List of firms in the sample

AFFILIATED MANAGERS GROUP INC	KEYCORP NEW
ALTRIA GROUP INC	LEGG MASON INC
AMBAC FINANCIAL GROUP INC	LEHMAN BROTHERS HOLDINGS INC
AMERICAN EXPRESS CO	M & T BANK CORP
AMERIPRISE FINANCIAL INC	M S C I INC
AMSOUTH BANCORPORATION	MARSHALL & ILSLEY CORP
B B & T CORP	MASCO CORP
BANK NEW YORK INC	MELLON FINANCIAL CORP
BANK OF AMERICA CORP	MERRILL LYNCH & CO INC
BLACKROCK INC	MORGAN STANLEY DEAN WITTER & CO
BLOCK H & R INC	N Y S E EURONEXT
C B O E GLOBAL MARKETS INC	NASDAQ O M X GROUP INC
C B RICHARD ELLIS GROUP INC	NORTHERN TRUST CORP
C I T GROUP INC NEW	P N C FINANCIAL SERVICES GRP INC
CAPITAL ONE FINANCIAL CORP	PEOPLES UNITED FINANCIAL INC
CHARTER ONE FINANCIAL INC	PRUDENTIAL FINANCIAL INC
CHICAGO MERCANTILE EXCH HLDG INC	RAYMOND JAMES FINANCIAL INC
CITIGROUP INC	REGIONS FINANCIAL CORP
CITIZENS FINANCIAL GROUP INC	S & P GLOBAL INC
COMPASS BANCSHARES INC	S V B FINANCIAL GROUP
DISCOVER FINANCIAL SERVICES	STATE STREET CORP
E TRADE FINANCIAL CORP	STILWELL FINANCIAL INC
FEDERAL HOME LOAN MORTGAGE CORP	SUNTRUST BANKS INC
FEDERAL NATIONAL MORTGAGE ASSN	SYNOVUS FINANCIAL CORP
FEDERATED INVESTORS INC PA	U S BANCORP DEL
FIFTH THIRD BANCORP	VISA INC
FIRST HORIZON NATIONAL CORP	WACHOVIA CORP 2ND NEW
FRANKLIN RESOURCES INC	WASHINGTON MUTUAL INC
GOLDMAN SACHS GROUP INC	WELLS FARGO & CO NEW
HUNTINGTON BANCSHARES INC	WESTERN UNION CO
INTERCONTINENTALEXCHANGE INC	ZIONS BANCORPORATION
J P MORGAN CHASE & CO	

B Word lists

Table 1: List of stop-words in the Harvey finance glossary

a	Fifth letter of a Nasdaq stock symbol specifying Class A shares.
ago	The three-character ISO 3166 country code for ANGOLA.
all	The ISO 4217 currency code for Albanian Lek.
am	The two-character ISO 3166 country code for ARMENIA.
an	The two-character ISO 3166 country code for NETHERLANDS ANTILLES.
and	The three-character ISO 3166 country code for ANDORRA.
are	The three-character ISO 3166 country code for UNITED ARAB EMIRATES.
as	The two-character ISO 3166 country code for AMERICAN SAMOA.
at	The two-character ISO 3166 country code for AUSTRIA.
be	The two-character ISO 3166 country code for BELGIUM.
by	The two-character ISO 3166 country code for BELARUS.
can	The three-character ISO 3166 country code for CANADA.
clear	To settle a trade by the seller delivering securities and the buyer delivering funds in the proper form.
close	The close is the period at the end of the trading session. Sometimes used to refer to closing price.
d	Fifth letter of a NASDAQ stock symbol specifying that it is a new issue, such as the result of a reverse split.
do	The two-character ISO 3166 country code for DOMINICAN REPUBLIC.
i	Fifth letter of a Nasdaq stock symbol specifying that it is the third preferred bond of the company.
in	The two-character ISO 3166 country code for INDIA.
is	The two-character ISO 3166 country code for ICELAND.

it	The two-character ISO 3166 country code for ITALY.
its	Intermarket Trading System (ITS)
m	Fifth letter of a NASDAQ stock symbol specifying that the issue is the company's fourth class of preferred shares.
ma	The two-character ISO 3166 country code for MOROCCO.
me	The two-character ISO 3166 country code for MONTENEGRO.
mean	The expected value of a random variable. Arithmetic average of a sample.
my	The two-character ISO 3166 country code for MALAYSIA.
no	The two-character ISO 3166 country code for NORWAY.
nor	The three-character ISO 3166 country code for NORWAY.
now	Negotiable Order of Withdrawal
o	Fifth letter of a Nasdaq stock symbol specifying that it is the company's second class of preferred shares.
on	Used in the context of general equities. Conjunction that denotes trade execution /indication, usually during a pre-opening look. "Looks 6 on 6000 shares at opening."
or	Operations research: A method of decision-making that uses analytical tactics such as mathematical models and statistical data to reduce risk and assist in answering complex business problems.
out	Used in the context of general equities. (1) No longer obligated to an order, as it has already been canceled: (2) advertised on Autex.
re	The two-character ISO 3166 country code for REUNION.
right	Privilege granted shareholders of a corporation to subscribe to shares of a new issue of common stock before it is offered to the public. Such a right, which normally has a life of two to four weeks, is freely transferable and entitles the holder to buy the new common stock below the public offering price.

run	A run consists of a series of bid and offer quotes for different securities or maturities. Dealers give and ask for runs from each other.
s	Fifth letter of a Nasdaq stock symbol specifying a beneficial interest.
so	The two-character ISO 3166 country code for SOMALIA.
t	Fifth letter of a Nasdaq stock symbol indicating that the stock has warrants or rights.
take	(1) To agree to buy. A dealer or customer who agrees to buy at another dealer's offered price is said to take the offer. (2) Euro bankers speak of taking deposits rather than buying money.
top	The ISO 4217 currency code for the Tonga Pa'anga.
us	The two-character ISO 3166 country code for UNITED STATES.
up	Market indication; willingness to go both ways (buy or sell) at the mentioned volume and market. Print; up on the ticker tape, confirming that the trade has been executed.
ve	The two-character ISO 3166 country code for VENEZUELA.
y	Fifth letter of a Nasdaq stock symbol specifying that it is an ADR

Table 2: List of words which are common in Loughran and McDonald (2011) word lists and Harvey finance glossary

Negative Words	abandonment, arrearage, backdating, bailout, bankruptcy, break, broken, cancel, clawback, correction, default, deficiency, deficit, delinquency, delisting, discrepancy, dishonor, divestiture, downgrade, downsizing, downturn, drawback, encumbered, erosion, expropriation, fail, foreclosure, forfeiting, forfeiture, illiquid, impairment, inquiry, insolvent, kickback, lag, lagging, lie, liquidation, liquidator, loss, manipulation, markdown, monopoly, overage, overrun, overvalued, protest, recession, rejection, restructuring, shrinkage, slippage, solvency, stagnation, stopped, undercapitalized, underperform, usury, volatility
Positive Words	advancement, better, efficiency, enhancement, exclusive, gain, leading, outperform, stability, stabilization, upturn
Uncertain words and weak modals	anticipation, risk, dependent, volatility, speculation, assumption, contingent, variance, contingency, probability, refinancing, fluctuation, speculative, variable

C Robustness

Table 1: Jargon and the short-term returns ($CAR^{0;1}$)

Notes: OLS regressions for Equation (3). The dependent variable is the abnormal returns over the CRSP value weighted returns and the Fama-French three factor (1993) model returns cumulated from the day of the earnings call to the day thereafter, $VW\_CAR^{0;1}$ and $FF3-CAR^{0;1}$, respectively. $Jargon\_All$ is the ratio of finance-related words in managements' answers. $Tone$ is the difference between the ratio of positive words minus the ratio of negative words in managements' answers. $Uncertainty$ is the ratio of uncertain words to total words. List of positive, negative, and uncertain words are from the Loughran and McDonald (2011) word-lists. $Numbers$ is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). Fog is the Gunning-Fog Index as a measure of linguistic complexity. Firm control variables include $EarnSurp$, BTM, RoE, and $ln(Assets)$. $EarnSurp$ represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). BTM is defined as total Common/Ordinary Equity divided by the market value of equity. RoE denotes return on equity. $ln(Assets)$ is the natural logarithm of total assets. t-statistics are given in parentheses; Standard errors are clustered on the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	$VW\_CAR^{0;1}$			$FF3-CAR^{0;1}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$Jargon\_All$	0.2370**	0.2337**	0.2627**	0.1674*	0.1736**	0.1937*
	(2.23)	(2.39)	(2.17)	(1.88)	(2.11)	(1.99)
$Tone$		0.0968	0.2037		0.2010**	0.2836**
		(1.07)	(1.62)		(2.49)	(2.34)
$Uncertainty$		-0.0295	0.1787		-0.0744	0.0505
		(-0.10)	(0.51)		(-0.29)	(0.16)
$Numbers$		0.2006	0.0976		0.1921	0.0443
		(0.82)	(0.23)		(1.02)	(0.16)
Fog		-0.0003	-0.0004		-0.0006	-0.0006
		(-0.55)	(-0.56)		(-1.27)	(-1.03)
N	2121	2121	1910	2121	2121	1910
R^2	0.119	0.120	0.164	0.121	0.123	0.163
Firm_Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm_FE	Yes	Yes	Yes	Yes	Yes	Yes
Management_FE	No	No	Yes	No	No	Yes
Quarter_FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 2: Jargon and the short-term returns ($CAR^{0;1}$)

Notes: OLS regressions for Equation (3). The dependent variable is the abnormal return over the CRSP value weighted return and the Fama-French three factor (1993) model return cumulated from the day of the earnings call to the day thereafter, $VW\_CAR^{0;1}$ and $FF3-CAR^{0;1}$, respectively. $Jargon\_Fin$ is the ratio of finance-related words in managements' answers to the questions that have at least a word categorised as financial-related. $Jargon\_NonFin$ is the same ratio for the management's answers to the questions with no finance-related word. Financial jargon is identified using Harvey (2016) finance glossary. $Tone$ is the difference between the ratio of positive words minus the ratio of negative words in managements' answers. $Uncertainty$ is the ratio of uncertain words to total words. List of positive, negative, and uncertain words are from the Loughran and McDonald (2011) word-lists. $Numbers$ is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). Fog is the Gunning-Fog Index as a measure of linguistic complexity. Firm control variables include $EarnSurp$, BTM, RoE, and $ln(Assets)$. $EarnSurp$ represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). BTM is defined as total Common/Ordinary Equity divided by the market value of equity. RoE denotes return on equity. $ln(Assets)$ is the natural logarithm of total assets. t-statistics are given in parentheses; Standard errors are clustered on the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	$VW\_CAR^{0;1}$			$FF3-CAR^{0;1}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$Jargon\_Fin$	0.2521**	0.2586***	0.3061**	0.1657*	0.1802*	0.2066*
	(2.40)	(2.68)	(2.53)	(1.71)	(2.00)	(1.82)
$Jargon\_NonFin$	-0.0128	-0.0142	-0.0097	-0.0125	-0.0150	-0.0039
	(-0.24)	(-0.27)	(-0.17)	(-0.27)	(-0.33)	(-0.08)
$Tone$		0.1258	0.2459*		0.2472**	0.3460**
		(1.21)	(1.72)		(2.60)	(2.50)
$Uncertainty$		0.0145	0.2428		-0.0586	0.0678
		(0.05)	(0.68)		(-0.23)	(0.21)
$Numbers$		0.1178	0.0588		0.1591	0.0394
		(0.48)	(0.14)		(0.84)	(0.15)
Fog		-0.0004	-0.0005		-0.0007	-0.0008
		(-0.72)	(-0.69)		(-1.57)	(-1.23)
N	1981	1981	1776	1981	1981	1776
R^2	0.127	0.128	0.172	0.129	0.133	0.171
Firm_Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm_FE	Yes	Yes	Yes	Yes	Yes	Yes
Management_FE	No	No	Yes	No	No	Yes
Quarter_FE	Yes	Yes	Yes	Yes	Yes	Yes

Chapter II

“Let me get back to you” - A machine learning approach to measuring non-answers

"Let me get back to you" – A machine learning approach to measuring non-answers*

Andreas Barth† Sasan Mansouri‡ Fabian Woebbeking§

June 30, 2021

Abstract

Using a supervised machine learning framework on a large training set of questions and answers, we identify 1,364 trigrams that signal non-answers. We show that this glossary has economic relevance by applying it to contemporaneous stock market reactions after earnings conference calls. Our findings suggest that obstructing the flow of information leads to significantly lower cumulative abnormal stock returns and higher implied volatility. Our metric is designed to be of general applicability for Q&A situations, and hence, is capable of identifying non-answers outside the contextual domain of financial earnings conference calls.

Keywords: Econlinguistics, textual analysis, natural language processing, multinomial inverse regression, non-answers
JEL-Classification: D80, D82, G10, G14, G30.

*We thank Daniel Andrei, Doron Avramov, Turan Bali, Justin Birru, Alexander Hillert, Gerard Hoberg, Michael Kirchler, Christian Leuz, Tim Loughran, Bill McDonald, Philippe Mueller, Michael Norton, Todd Rogers, Astrid Salzmann, Christian Schlag, Matt Taddy, Alexander Wagner, Mark Wahrenburg, and Xiao-Jun Zhang for valuable comments and suggestions. We also thank Markus Bader and the department of *Psycholinguistics* at Goethe University Frankfurt for providing valuable background information. The paper was previously circulated under the working title 'Econlinguistics'.

†Goethe University Frankfurt, andreas.barth@finance.uni-frankfurt.de

‡Goethe University Frankfurt, mansouri@finance.uni-frankfurt.de

§Goethe University Frankfurt, woebbeking@finance.uni-frankfurt.de

1 Introduction

"Senator, my — I can certainly have my team get back to you on any specifics there that I don't know, sitting here today." — Mark Zuckerberg, *US Senate Hearing, April 2018*

The asymmetric distribution of information is considered a key friction in economics. Different mechanisms aim to transfer information from the better-informed to the less informed agent, where a question and answer (Q&A) setting is the most targeted form of information exchange. Faced with a question, the addressee can respond in two ways. First, they can supply the requested information by faithfully answering the question, which requires having the specific knowledge within the context of the question.[1] Second, they can refuse to supply the requested information, which does not require a context-specific answer. While it is *relatively* easy for humans to detect whether a question has been answered or not, we teach this skill to a machine. We use a supervised machine learning framework on a large textual training set of 64,173 classified responses to questions to identify 1,364 trigrams that signal non-answers.

The Gricean norms in communication describe cooperative principles of how people achieve effective conversational communication (Grice, 1989). These principles state that effective communication (i) contains the appropriate quantity of information, (ii) is truthful, (iii) is delivered in an appropriate manner and (iv) is relevant to the topic at hand. Violating any of these principles results in 'deceptive' communication.

We use violations of Gricean norms to derive a metric that identifies the absence of requested information in an answer, i.e. non-answers. At its core, our glossary based metric employs a trained set of trigrams, which are markers for non-answers. The glossary is derived from financial markets, which are heavily characterized by and sensitive to asymmetric information. More precisely, we derive the glossary from a training set of earnings conference calls, where investors and analysts can directly question senior executives' during Q&A sessions. [2] We conduct several tests to investigate the economic relevance of the metric using a large validation set of earnings conference calls.

We document that markets react to non-answers and observe negative stock returns after calls on which management distinctly avoids answers. Moreover, we find larger implied volatilities after these calls, indicating higher investor uncertainty. Financial analysts, too, perceive non-answers as a negative signal. In particular, we find that analysts are less likely to modify their EPS forecasts upwardly following a call with many non-

[1] Alternatively, the respondent could answer the question with a lie, which also requires knowing the specific context of the question. As lies typically receive heavy sanctions, this paper focuses on refusals to answer and not detecting lies.

[2] Despite being derived from a financial market context, the glossary is applicable to other Q&A contexts.

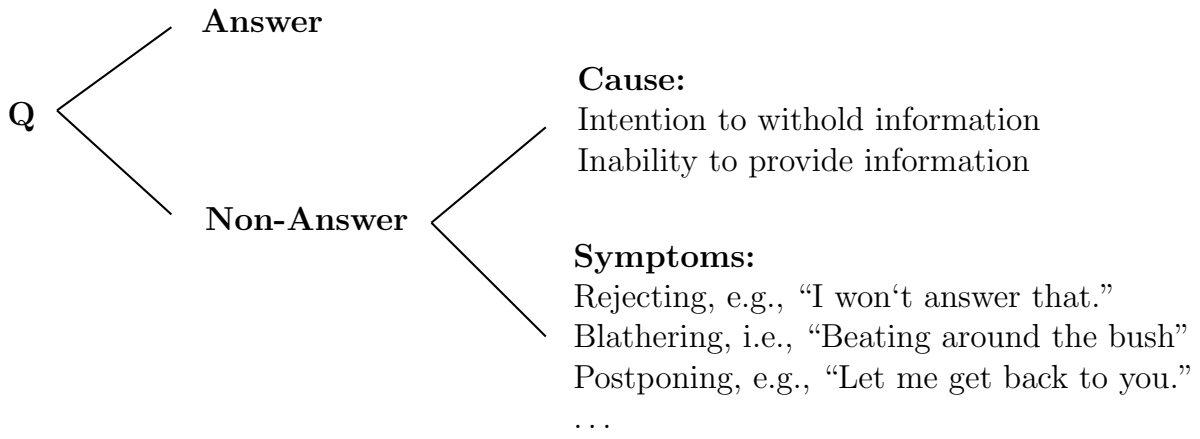


Figure 1: Anatomy of an answer. The response to a question consists of either an attempt to provide the requested information with an appropriate answer or the lag thereof, i.e. a non-answer. Later, we will use the symptoms blathering and rejecting to classify non-answers in our training set and show that this is sufficient to capture additional symptoms, such as postponing. Non-answers have in common that they can be free of context, e.g. one may reject answering without even knowing the questions, which is an important distinction that allows for generalizing our glossary.

answers. We finally investigate to which questions managers provide non-answers and find that, within the same earnings call, managers avoid responding to arguably tougher questions, as indicated by negative-toned questions and follow-up questions by the same analyst. Similarly, non-answers appear more frequent when questions are 'forward looking', hence, when managers' may be less able to provide the requested information.

Conceptually, a question can be understood as an illocutionary act that attempts to extract information from its addressee. The addressee can respond in two ways, as outlined in Figure 1. First, she can supply the requested information by faithfully answering the question, which requires having the specific knowledge within the context of the question. Second, she can refuse to supply the requested information, which, in contrast to effective communication, does not depend on the context of the question. This paper refers to this potentially deceptive communication in answering as a non-answer.

Non-answers are characterized by different symptoms. The most obvious symptom is openly refusing to provide the requested information, as for example, Elon Musk, the CEO of Tesla Inc, did during an earnings call in May 2018.[3] A second symptom of refusing context-specific information is the more indirect and deceptive behavior of dodging a question or "blathering", i.e. 'beating around the bush'.[4]

[3]On the question of Sanford Bernstein's analyst Toni Sacconaghi: "And so where specifically will you be in terms of capital requirements?", Musk replied: "Excuse me. Next. Boring, bonehead questions are not cool. Next?"

[4]Although there may be other symptoms, we later show that a limited number of symptoms is sufficient to train the model.

We base our measure on the two symptoms *'rejecting'* and *'blathering'*, and show that we can construct a metric that quantifies non-answers by focusing on just a few symptoms by using a Multinomial Inverse Regression (MNIR) (Taddy, 2013b). The input for the MNIR are all Q&As of earnings conference calls for financial firms in the S&P 500 for the period 2002 - 2019 (the training set). For each of management's answers during these calls, we quantify the two symptoms rejecting, as described in Gow et al. (2021), and blathering, as outlined in Barth et al. (2021). MNIR, as a supervised generative model, then maps the high-dimensional choice set of available trigrams into the two observable attributes in the classified training set.[5]

This procedure results in a glossary of 1,364 trigrams that deduce a scoring metric for non-answers. The trigrams in the glossary are neither industry- nor context-specific. Thus, it can be applied to any economic sector or other Q&A setting, such as senate hearings, interviews with politicians, or press conferences by central banks.

We conduct various tests to document the plausibility of the glossary. To do so, we collect the earnings conference calls of non-financial companies in the S&P 500, i.e. firms that are not part of the training set, for the period 2002 - 2019 as a validation set. Earnings conference calls provide an ideal setting to test our glossary: the listener is more likely to detect whether a question has been answered when their attention is diverted from social goals (Rogers and Norton, 2011), and thus, we expect an immediate market reaction.

We first document how firms subtly control information flow (Cohen et al., 2020): there are more non-answers in responses to analysts that are more pessimistic in their earnings forecasts. Comparing different assessments by the same analyst helps us to rule out a non-random matching of analysts and firms. Moreover, during an earnings call, managers try to avoid answering tougher and more critical questions. The non-answer score is higher for managements' responses to follow-up questions by the same analyst, i.e. when the analyst asks a (typically more drilling) clarification question, as well as for managements' responses to more negative questions. Furthermore, questions with forward looking sentences, i.e., questions that refer to (potentially unknown) future outcomes, are more likely to receive a non-answer.

Second, in regressions with multiple control variables, we show that not answering analysts' questions leads on average to significantly negative abnormal stock returns following the conference call. We also link our metric to option implied volatilities after earnings conference calls and show that investor uncertainty is greater if the requested

[5]While the usage of multinomial text regressions is novel in finance literature, it has been applied in political science studies to derive the subject-specific document sentiment in political posts (Taddy, 2013b) or to measure time trends in partisanship in congressional speeches (Gentzkow et al., 2019).

information is not provided, i.e. investors are willing to pay more for insurance against adverse stock price movements. Both the stock price reaction as well as higher implied volatility suggest that the non-answer score measures an obstruction of information flow, which retards the reduction of information asymmetries between the management and investors. We validate that the glossary is not a product of sheer randomness in a Monte Carlo simulation, where we draw 1,000 randomly selected dictionaries with 1,364 trigrams from all words that appear at least once in the training set of our earnings calls. We repeat the textual analysis for each of these random draws, derive a corresponding placebo non-answer score and test for an effect of this score on cumulative abnormal returns. It turns out that it would be extremely unlikely to produce economically significant results by randomly drawing a glossary from the universe of trigrams.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature. In Section 3, we describe in detail how we generate our novel glossary. We describe several tests for the validity of the word list in Section 4. Section 5 concludes.

2 Background and literature

Textual analysis is a versatile tool in finance and accounting that transforms qualitative information into quantitative measures. One common approach for quantifying language is word categorization (bag of word / dictionary). For example, the Harvard-IV and Lasswell dictionaries, which are part of Harvard General Inquirer Word Lists, consist of word lists about many psychological and sociological topics.[6]

To overcome issues with noise from general dictionaries, researchers have introduced finance-specific dictionaries to measure the tone of financial reports. Henry (2008), for example, published one such list for the telecommunications and computer services industries. Loughran and McDonald (2011) produced other widely recognized word lists that were extracted from 10-K reports to measure inter alia positive and negative tone, and most recently a word list to measure firm complexity (Loughran and McDonald, 2019). Harvey (2016) also created a glossary of factual finance terminology, which Loughran and McDonald (2014) use, for example, to develop a measure of financial readability of 10-K reports. We extend this literature by providing a novel glossary that quantifies the informational content of a response to a question, and thus, a new dimension of quantifiable language that can be used in several contexts.

Second, we add to the interdisciplinary literature on the precise and efficient transfer of information, or lack thereof. In linguistics, for example, many studies describe how to

[6]For more details on the different available dictionaries and the list of words, see: `http://www.wjh.harvard.edu/~inquirer/homecat.htm`

achieve effective conversational communication, with the cooperative principles by Grice (1989) being some of the most influential content. Building on this work, several studies show how listeners perceive a violation of the cooperative principles, or in which situations a listener is more likely to detect deceptive communication.[7]

The effects of precise information sharing are also of particular interest in the field of economics, where obscuring information is associated with, e.g., lower stock returns, lower earnings and higher risks. The related studies use various proxies for the characterization of imprecise information, such as vague communication measured by the frequency of words such as "vague" and "uncertainty" in management statements (Loughran and McDonald, 2011; Dzielinski et al., 2016), the ratio of numeric to textual content in earnings conference calls (Zhou, 2018), readability of 10-K reports measured by the popular Gunning (1952) "Fog-Index" of linguistic complexity (Li, 2008; Bloomfield, 2008),[8] calling on bullish analysts in conference calls (Cohen et al., 2020), or managers reading from prepared scripts when responding to questions during earnings conference calls (Lee, 2015).

Most closely related to our work in the interdisciplinary literature is Clayman (1993), which shows that evading a question is frequently characterized by the response practice to reformulate the question. In the field of economics, our work is closest to Hollander et al. (2010) and Gow et al. (2021), who measure withholding information in the most direct sense by manually reviewing call transcripts to deduct regular expressions that identify answers such as '*No, we do not want to provide that information.*'

Our glossary is a significant step towards the general identification of non-answers. Our novel approach starts by classifying a training set with different symptoms for non-

[7]See, for example, Buller and Burgoon (1996), who state in their Interpersonal Deception Theory that the process and outcome of interpersonal deception is grounded within a conversational context and an interpersonal relationship. McCornack et al. (1992) show that perceived message deceptiveness and perceived message competence are significantly influenced by the manipulations of amount, veracity, relevance, and clarity of information. The paper by Rogers and Norton (2011) shows that deception is more likely to be detected when listeners' attention was to determine the relevance of the speakers' answers, i.e. if it was diverted from social goals. It is also shown that speakers were more negatively rated once their deception was detected.

[8]Based on this work, several studies offer different measures of linguistic complexity (Loughran and McDonald, 2014; Bonsall et al., 2017) or provide a rationale for using complex language. For example, Bushee et al. (2018) analyze the linguistic complexity of Q&As in earnings calls and argue that the source of complexity can be composed into its latent components *obfuscation* and *information*. Specifically, they argue that complex responses to complex questions should be understood as information, whereas complex responses to simpler questions should be understood as obfuscation. In line with that, Guay et al. (2016) show that managers employ voluntary disclosures as a tool to mitigate the negative impact of their complex financial statements.

answers, including symptoms laid out by the preceding literature.[9] The glossary is then determined by a machine learning algorithm, thereby reducing the subjectivity of a human interpreter. The resulting glossary reflects not only one very specific symptom, e.g. 'rejecting', but captures a non-answer across several symptoms (see Figure 1) and in a much broader sense. We show that this identification is significantly more powerful than previous attempts in the literature. Furthermore, our non-answer metric is only very weakly correlated with other linguistic sentiment measures and, hence, serves as a new dimension in textual analysis.

3 The glossary

A machine readable version of the glossary is available at econlinguistics.org

---

The glossary contains trigrams that are markers for non-answers.[10] This section lays out how these trigrams were identified from a training set of questions and answers.

As outlined in Figure 1, when faced with a question, the addressee can respond using effective communication and supply the requested information, or they can be deceptive by violating any of the four Gricean cooperation principles (Grice, 1989). When developing the glossary, we focus on the first and the fourth Gricean maxims, i.e., whether the respondent provides any information at all and whether their answer provides factual content that is relevant for the topic at hand.

To understand general factuality in a linguistic sense, one would need to understand the context of the question and the expected information gain. The supervised machine learning approach, however, can derive a glossary that is context-independent. Thus, we do not approach factuality from a context and audience specific perspective but rather focus on vocabulary that indicates the intention not to respond to a question in a broader, more general sense.

[9] When applying the regular expressions for rejecting an answer of Gow et al. (2021) to the analysis of economic relevance in Section 4, we find only very weak evidence of abnormal stock returns following an earnings call and obtain hardly any variation in the distribution for any Q&A setting outside the financial domain.

[10] A trigram is a continuous sequence of three elements from a text. The literature on natural language processing shows a significant improvement in modeling language with trigrams compared to unigrams (Dave et al., 2003; Bekkerman and Allan, 2004). While higher-order n-grams better capture the sentiments of expressions, they come at the cost of lower coverage in the data (Pak and Paroubek, 2010). For many years, trigrams have been a favorite model choice, as they can simultaneously reflect the syntax and the pragmatics of the text domain (Jelinek, 1991).

3.1 Multinomial Inverse Regression

Multinomial Inverse Regression is a supervised generative model developed by Taddy (2013b, 2015) that allows for mapping a high-dimensional choice set of words within a text into an observable attribute. A text is defined as a combination of several tokens, where a token is a single word or a combination of n words (n-gram). For a given tokenization, a document i in the universe of all available documents $\mathcal{I}$ is represented by a sparse vector of token counts $\mathbf{x_i} = [x_{i1}, ..., x_{ip}]'$ and frequencies $\mathbf{f_i} = \mathbf{x_i}/m_i$, where $m_i = \sum_{j=1}^{p} x_{ij}$, for all available tokens p in $\mathcal{I}$.

A naive approach would be to fit a linear regression model of the attribute measure on the token counts,

$$\mathbf{y} = \beta \mathbf{x}^{\intercal} + \epsilon,$$

where the factor loading β represents each token's contribution to the attribute measure. However, as the choice set of tokens within a text and thus, the dimension of $\mathbf{x}$, is usually quite large, a normal regression cannot provide an appropriate estimate of the conditional distribution of $\mathbf{y}$.

To shrink dimensionality in pursuit of a parsimonious model, we turn to a least absolute shrinkage and selection operator (Lasso) regression type of model (also known as L1-regularisation), see e.g. Hastie et al. (2009). Building on the pioneering work of Cook et al. (2007), Taddy (2013b, 2015) developed a multinomial inverse regression (MNIR) methodology. MNIR, as a 'Gamma-Lasso' scheme, applies inverse regressions, "wherein the inverse conditional distribution for text given sentiment is used to obtain low-dimensional document scores that preserve information relevant to y." This methodology, which was specifically designed for textual analyses, produces significant computational improvements over the classical Lasso approach.

As described in Taddy (2013b), a basic MNIR model is given by

$$\begin{aligned}
\mathbf{x_y} &\sim MN(\mathbf{q_y}, m_y), \quad \text{with} \\
\mathbf{x_y} &= \sum_{i:y_i=y} \mathbf{x_i}, \\
m_y &= \sum_{i:y_i=y} m_i, \\
q_{yj} &= \frac{exp[\alpha_j + y\phi_j]}{\sum_{k=1}^{p} exp[\alpha_k + y\phi_k]} \\
\text{for } \ j = 1, ..., p, y \in \mathcal{Y} \ \text{ and } \ m_i &= \sum_{j=1}^{p} x_{ij}.
\end{aligned}$$

Each MN is a p-dimension multinomial distribution of size m and probabilities $\mathbf{q}$ that are a linear function of y through a logistic link with token loadings ϕ.

The parameters ϕ of the model, i.e. , each token's contribution to the attribute measure, can be fitted via maximum a posteriori (MAP) estimation. While a classical Lasso estimation can be interpreted as a MAP estimation with independent and identical Laplace priors for the regression parameters, Taddy (2013b) use independent Gamma-Laplace priors to fit the MNIR model.[11]

In this paper, we focus on the refined estimation procedure proposed in Taddy (2015), which is available through the *textir* package in R. In contrast to the procedure in Taddy (2013b), which requires explicit shape and rate hyperpriors for the Gamma distribution, the modified approach in Taddy (2015) only has one relevant parameter, the *"Gamma-Lasso weight"* γ. We follow the author and set $\gamma = 1$; however, we find that alternative specifications (including $\gamma = 0$ as in classical Lasso) do not change our overall results. Ultimately, our results do not hinge on the regularization model, we simply follow Taddy (2015) because it is computationally very efficient, allows more flexibility for the sentiment input (higher number of categories), and requires the fewest assumptions.

The response factor does not have to be a single attribute measure y_i, but MNIR can be generalized to support K-dimensional response factors $\mathbf{v_i}$, in which case the multinomial model collapses to the levels of $\mathbf{x_v}$. In our case, using several response factors has teh benefit that we can employ measures for different violations of the Gricean norms of effective conversational communication (Grice, 1989).

We use measures for the violation of the first and fourth Gricean maxim as response factors.[12] A violation of the first Gricean norm, the lack of appropriate quantity of information, is derived in the spirit of Gow et al. (2021), using regular expressions that identify answers with a direct rejection to provide the requested information. The violation of the fourth Gricean maxim, the relevance of the response to the topic at hand, is proxied with the metric for blathering introduced in Barth et al. (2021).

By using these response factors, the resulting list of trigrams and their corresponding weights indicate the degree to which a given trigram predicts a violation of the effective conversational communication principles of Grice (1989) in any response to a question.

[11]See Taddy (2013a) for a comparison of their Gamma-Lasso framework with standard Lasso and alternatives with concave penalization.

[12]As described below, we derive the glossary based on Q&As in financial markets. As lies typically receive heavy sanctions, and answers by management were usually delivered in an appropriate manner, we abstract from measures of violations of the second and third Gricean norms.

3.2 Training set

The glossary is extracted from textual data that originates from earnings conference calls. These calls offer a relatively standardized Q&A format in a controlled contextual environment, have an economically relevant impact and are available in regular intervals and large numbers. Earnings calls do not have an identical structure, yet they often follow a similar pattern: first, the management, typically the CEO or CFO, presents the latest financial results and earnings outlooks in a speech that is usually prepared by the investor relations department. A question and answer session between the management and financial analysts then follows this presentation.

We collected every transcript of earnings calls held by companies listed in the S&P 500 index available from Thomson Reuters' StreetEvents for the period 2002 - 2019. These calls are released quarterly and usually take place on the same day as the corresponding earnings release.[13] As we focus on a question and answer setup as a specific form of communication rather than the prepared presentation, we exclude all earnings calls without a Q&A session. We also restrict our sample to managements' responses that contain at least five words to mitigate any bias in our attribute measures.[14]

The full sample of earnings calls is divided into a training set $\mathcal{I}$, where we have a clean measure for both response factors, and a validation set that *validates* our glossary by showing its economic relevance. We split the sample across industries and use all financial firms in the training set.[15] More precisely, we use each single answer given by a financial firm's management in response to an analyst's question as an observation in the training set $\mathcal{I}$. We derive our two response factors for these answers.

As a first attribute measure, we use regular expressions to identify the rejections according to Gow et al. (2021). Rejections can take several forms, such as the refusal to provide the requested information ("we do not provide this disclosure") or the inability to provide the requested information ("I do not know"). We flag these answers with a dummy $y_{ijt}^{\text{Rejecting}}$ that equals 1 if the response j in an earnings call of company i at time

[13] 92.8% of all calls in our sample take place on the same day as the earnings release and 6.9% take place one day after the earnings announcements. In only six cases is the call scheduled for more than one day after the earnings announcement.

[14] Shorter answers need to be removed because of the high signal-to-noise ratio for the attribute measure blathering. For example, sentences like "This is correct" result in a blathering score of 1 and would thus appear with a high probability in the non-answer glossary. These sentences, however, are obviously very precise answers.

[15] We classify all firms with industry code 44 or 47 in Fama-French's 48 industry portfolios as financial firms.

t contains any rejection phrase, i.e.

$$y_{ijt}^{\text{Rejecting}} = \begin{cases} 1 & \text{, if rejection phrase} \in \text{response } j \\ 0 & \text{, else.} \end{cases} \tag{1}$$

As a second attribute measure for non-answers, we calculate blathering as introduced by Barth et al. (2021). Blathering – from the Oxford English Dictionary: *"[To] talk in a long-winded way without making very much sense."* – is capturing 'information' that is volunteered but either does not meet or purposefully avoids a precise answer.

The degree of blathering in the response j in an earnings call of company i at time t is defined as

$$y_{it}^{\text{Blathering}} = 1 - \frac{\text{Finance glossary words}_{it}}{\text{Total words}_{it}}, \tag{2}$$

where words are classified as financial words based on the Hypertextual Finance Glossary by Campbell R. Harvey, consisting of more than 8,500 entries.[16] This metric assumes that for *financial firms*, the factual content in managements' responses to analysts' questions is mirrored by the usage of *finance-related words.*

Fitting the model requires us to turn $y_{it}^{\text{Blathering}}$ into a categorical variable. We therefore min-max normalize $y_{it}^{\text{Blathering}}$ and truncate to one decimal place. This leaves us with 10 categories that reflect different blathering intensities.[17]

Non-financial firms might provide factual and relevant information on, e.g., their long-term strategic vision or future product releases, which the blathering metric would falsely classify as non-factual. Thus, the set of non-financial text is only an intersection of non-answers for these firms.

To emphasize the usage of financial firms as a training set, consider the signal-to-noise ratio of the blathering measure for financial versus non-financial firms. The measure would be equally applicable to any industry if earnings calls were solely concerned with a company's financial situation. However, those calls typically include additional topics, such as a discussion of the firm's products and other revenue drivers. The linguistic context of financial firm earnings calls is therefore unique, as their products and other topics are largely finance related. Thus, non-financial words are a strict subset of non-answers, that is, they consist of noise that is not required for answering the initial question, as well as of words that are required to formulate a sentence in plain English. Think in contrast

[16] See Campbell Harvey's webpage at Duke University, `http://people.duke.edu/~charvey/`.

[17] The results are also robust to changes in the number of categories, e.g., when truncating to two digits. We decide in favor of a low number of categories, which, on one hand, ensures sufficient observations per category and, on the other hand, is computationally more efficient.

of a technology firm like Apple, which tends to discuss non-finance product and brand specifics during earnings calls. This discussion provides factual and relevant information, but the blathering metric would falsely classify these as non-factual, resulting in a low signal-to-noise ratio of the measure.

To reduce noise in this attribute measure for non-answers further, we restrict our training set to answers that respond to questions with at least one finance-related word. This ensures that we do not involuntarily assign a high score for blathering to a response to a question that was unrelated to a finance context.

Our training set comprises 64,173 management answers from 2,124 earnings calls for 42 financial firms listed in the S&P 500, which accounts for roughly 10% of the textual data from earnings calls of all S&P 500 firms. As the remaining textual data of earnings calls of non-financial firms in the S&P 500 is used for validation, we end up with a very large validation set, which prevents typical problems of machine learning algorithms, such as over-fitting and data mining.

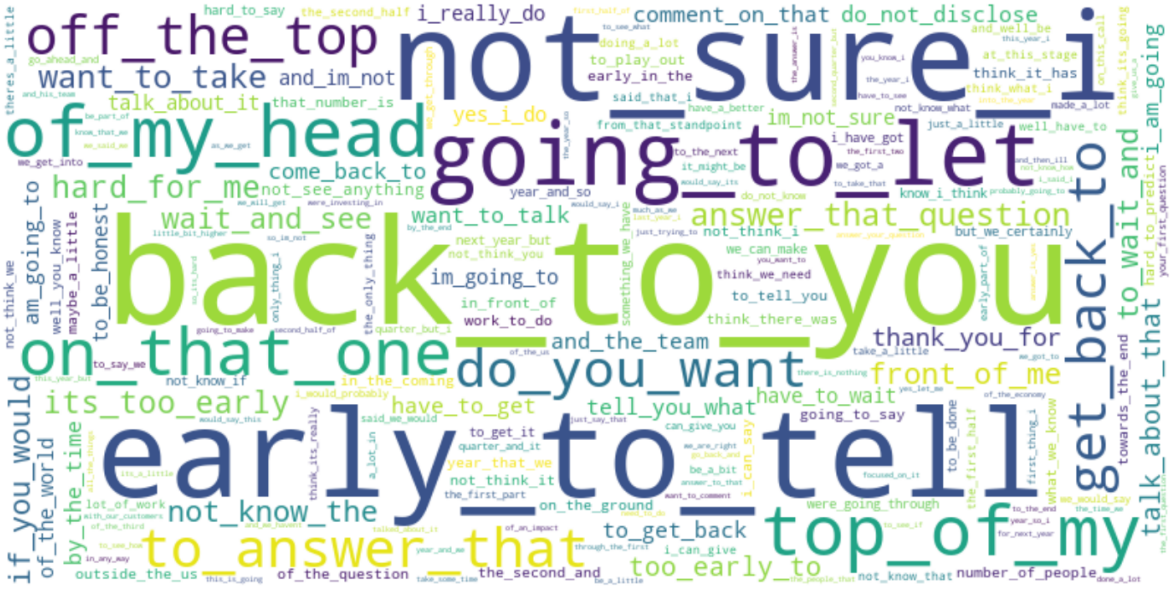
3.3 Fitted glossary

For each answer in the training set, we form an answer-term-matrix of trigrams. The response factors are metrics for rejection and blathering as described above. We employ two cleaning procedures to make sure that the answers contain meaningful words and that the resulting glossary is of general use. First, we aim to avoid company specific trigrams to achieve the most general language in the glossary. Thus, we focus on the most common trigrams of all responses that appear in at least 100 of the answers. Second, we want to remove common trigrams consisting mostly of (meaningless) stop words. In order to provide a directly applicable glossary to spoken English sentences, we do not filter for stop words before forming trigrams, but would like to remove those trigrams from the glossary that appear in more than 50% of the answers.[18] This cleaning procedure leaves us with around 3,400 trigrams.[19]

The model returns 568 (1290) trigrams with a positive (negative) loading for the rejection response factor and 1099 (970) trigrams with a positive (negative) loading for the blathering response factor. Unlike a non-answer, an answer is always specific to the context of the question. Hence, trigrams with a finance meaning show a strong negative loading by construction. However, these trigrams are only meaningful in a context-specific

[18]Note that this restriction is not binding, i.e., we do not delete any trigram by applying this filter.

[19]Loosening the first filtering criteria will result in a less generic and lengthier word-list that contains more trigrams specific to the financial industry, similar to our training set. It does not, however, affect the results shown later in the validation analysis.


back_to_you
early_to_tell
not_sure_i
going_to_let
off_the_top
of_my_head
on_that_one
to_answer_that
top_of_my
do_you_want
get_back_to
talk_about_that
hard_for_me
wait_and_see
answer_that_question
want_to_take
talk_about_it
come_back_to
im_going_to
thank_you_for
front_of_me
its_too_early
have_to_get
tell_you_what
have_to_wait
not_know_the
to_get_back
too_early_to
i_really_do
comment_on_that
do_not_disclose
and_im_not
yes_i_do
im_not_sure
i_have_got
know_i_think
not_know_what
not_know_that
number_of_people
outside_the_us
of_the_question
the_second_and
if_you_would
of_the_world
by_the_time
to_be_honest
am_going_to
i_am_going
to_wait_and
hard_to_predict
hard_to_say
the_second_half
theres_a_little
maybe_a_little
well_you_know
not_think_we
not_think_it
not_think_you
not_know_if
in_the_coming
and_the_team
work_to_do
in_front_of
to_tell_you
think_there_was
think_we_need
we_can_make
but_we_certainly
not_think_i
early_in_the
to_play_out
doing_a_lot
said_that_i
from_that_standpoint
think_it_has
think_what_i
at_this_stage
made_a_lot
well_have_to
not_see_anything
year_and_so
that_number_is
going_to_say
were_going_through
on_the_ground
i_can_give
i_can_say
be_a_bit
year_that_we
to_get_it
quarter_and_it
lot_of_work
with_our_customers
of_the_third
talked_about_it
of_an_impact
through_the_first
the_first_part
the_first_half
what_we_know
towards_the_end
the_time_we
done_a_lot
take_some_time
be_a_little
this_is_going
in_any_way
to_see_how
to_see_what
go_ahead_and
we_get_into
we_said_we
know_that_we
be_part_of
we_got_through
we_got_a
to_take_that
probably_going_to
end_then_ill
not_know_how
i_said_i
the_economy
yes_let_me
there_is_nothing
we_got_to
we_would_say
for_next_year
year_so_i
the_year_i
have_to_see
not_know_what
the_year_so
the_year_is
the_year_i
the_answer_is
second_quarter_but
we_get_through
the_year_so
to_see_what
first_half_of
doing_a_lot
to_the_next
it_might_be
would_say_its
do_not_know
we_got_a
know_i_think
we_have
just_trying_to
not_think_i
answer_your_question
we_can_make
you_want_to
think_we_need
to_tell_you
think_there_was
work_to_do
in_front_of
early_part_of
answer_is_we
your_first_question
the_only_thing
only_thing_i
quarter_but_i
second_half_of
going_to_make
we_will_get
were_investing_in
we_got_into
as_we_get
be_part_of
know_that_we
we_said_we
the_year_so
the_year_is
the_year_i
have_to_see
not_know_what
intro_the_year
made_a_lot
well_have_to
end_then_ill
not_know_how
probably_going_to
i_said_i
the_economy
yes_let_me
there_is_nothing
we_got_to
we_would_say
for_next_year
year_so_i
the_time_we
done_a_lot
take_some_time
be_a_little
this_is_going
in_any_way
to_see_how
to_see_what
go_ahead_and
we_get_into
we_said_we
know_that_we
be_part_of
we_got_through
we_got_a
to_take_that
probably_going_to
end_then_ill
not_know_how

Q&A data from interviews of US presidents.[24] The presidential data shows substantial variation in $NonAnswer$, a necessary condition for the measure to be informative. Particularly high examples for non-answer scores are found in interviews with President Clinton around the time when sexual assault allegations surfaced that later became the basis for an impeachment charge of perjury.[25]

4 Economic relevance

As financial economists, we naturally focus on applying $NonAnswer$ on textual data related to our discipline. Thus, we measure non-answers for management responses in earnings calls and conduct a variety of tests to evaluate the plausibility of our glossary. First, we analyze non-answers in earnings calls by investigating which analysts are more likely to receive non-answers and which questions managers try to avoid. Second, we examine how markets react to non-answers by studying cumulative abnormal stock returns and implied volatilities in the wake of earnings conference calls.

Financial markets are perfectly suited to assess the economic relevance of our measure for two reasons. First, economic theory gives a prior expectation of the effect that we would expect for avoiding answering analysts' questions. Second, in a financial markets' context, 'artful dodgers', as described in Rogers and Norton (2011), should be detected, as social evaluation does not play a role and the listeners' attention is directed towards the goal of identifying whether a person is answering a question. Thus, in the context of finance, we have a clear prior of a negative perception of avoiding an answer.

Investors participate in the Q&A sessions of earnings conference calls in order to reduce uncertainty about a firm's expected future performance. The theoretical asset pricing literature suggests that higher uncertainty translates to larger risk premia (Andrei and Hasler, 2014).[26]

An investor's uncertainty is reduced less by a non-answer compared to a precise,

[24]These interviews were collected by UCSB's American Presidency Project, see www.presidency.ucsb.edu.

[25]For example, during a telephone interview with Morton Kondracke, Bill Clinton responded to the question: *"Okay. Let me just ask you one more question about this. You said in a statement today that you had no improper relationship with this intern. What exactly was the nature of your relationship with her?"* with the words *"Well, let me say, the relationship's not improper, and I think that's important enough to say. But because the investigation is going on and because I don't know what is out—what's going to be asked of me, I think I need to cooperate, answer the questions, but I think it's important for me to make it clear what is not. And then, at the appropriate time, I'll try to answer what is. But let me answer, it is not an improper relationship, and I know what the word means."*.

[26]For a discussion on the impact of policy uncertainty on risk premia, see also Pástor and Veronesi (2013) or Liu et al. (2017). For the effect of disagreement as some special cases of uncertainty on asset pricesCarlin et al. (2014).

context specific response. For a given prior expectation, a context-free response might even increase uncertainty. Thus, we expect market participants to react more negatively in response to non-answers. In fact, empirical literature in line with this expectation shows that not conveying information leads to a negative stock market reaction. For example, Zhou (2018) argues that obscuring information by increasing textual rather than numeric content is associated with lower cumulative abnormal returns around the earnings call date. Similarly, Hollander et al. (2010) shows that stock returns in a 90 or 120-minute window after an earnings conference call react significantly more negatively if the management refused to answer a question in the call. Taking these results at face value, a necessary condition for the validity of our glossary is to observe a negative correlation between $NonAnswer$ and stock returns after an earnings conference call.

4.1 Data

We collect earnings calls of all S&P 500 non-financial companies, i.e., all calls of those S&P 500 firms that were not used to train the model. These 23,815 earnings calls of the 'validation set' is significantly larger than the training set (2,124 earnings calls) and, hence, minimizes the risk of over-fitting.

Non-answer score We apply our glossary to derive a metric that captures non-answers. For the earnings call of company i in quarter t, we count the occurrence of trigrams from the glossary in all responses of the Q&A session and divide by the total number of words, hence,

$$NonAnswer_{it} = \frac{\text{Non-answer glossary tokens}_{it}}{\text{Total words}_{it}}. \tag{3}$$

In addition, to incorporate the information on the loadings, we measure a $NonAnswer^{\phi}$ by weighting each trigram in the glossary with its respective factor loading,

$$NonAnswer^{\phi}_{it} = \frac{\sum_{k=1}^{K} \phi_k \times \text{Non-answer glossary token}_{itk}}{\text{Total words}_{it}}, \tag{4}$$

where ϕ_k is the loading associated with trigram $k \in \{1, 2, ..., K\}$.

The distribution of $NonAnswer$ for the earnings calls in our validation set is shown in Figure A1 and the sample average of $NonAnswer$ is shown in Figure A2. It is interesting to note that we observe a peak of non-answers during and in the immediate aftermath of the financial crisis.

Cumulative abnormal returns We obtain daily adjusted stock returns from CRSP and calculate daily abnormal return for the stock of company i at time t with the Fama-French three-factor (1993) and five-factor (2015) model returns,[27]

$$r_{i,t}^{abnormal} = r_{i,t} - r_{i,t}^{FF}.$$

We investigate the short-term effect using cumulative abnormal returns from the day of the earnings call to the day after, $CAR_{i,t}^{0;1}$. As a robustness test, we enlarge the event window and consider cumulative abnormal returns including the day before the earnings call, $CAR_{i,t}^{-1;1}$.

Option implied volatility The implied volatility derived from prices of exchange-traded equity options reflects the premium that investors are willing to pay for insuring against price movements in the underlying and thus proxies for investor uncertainty.[28] We collect the daily implied volatility $\sigma_{i,t}$ derived from liquid at-the-money options with 91-day maturity from OptionMetrics, LLC. We calculate two metrics to capture the instantaneous update of investors beliefs on future volatility after a conference call. The first approach follows Rogers et al. (2009), and compares the implied volatility of company i on the day just after the call to that of the day just before the call,

$$IV_{i,t}^{-1;1} = \ln\left(\frac{\sigma_{i,t+1}}{\sigma_{i,t-1}}\right).$$

Second, we compare the change in $\sigma_{i,t}$ with a counterfactual change in the implied volatility, which we calculate as the average change in implied volatility for the 60 trading days preceding the earnings call,

$$\Delta IV_{i,t} = \frac{\sigma_{i,t+1} - \sigma_{i,t-1}}{2} - \frac{\sigma_{i,t-1} - \sigma_{i,t-60}}{59}.$$

Alternative speech characteristics The literature provides evidence that investors recognize tone sentiment and the uncertainty of the language used in earnings calls.[29] As we want to test whether our measure of non-answers is not purely capturing management

[27]The model is calibrated to 40 trading days preceding an earnings call, with data from the Fama-French data library at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

[28]See Rogers et al. (2009) for a discussion of the advantages of using implied volatility to measure investor uncertainty compared to other possible measures such as realized volatility or the dispersion in analyst forecasts.

[29]See, e.g., Price et al. (2012), Blau et al. (2015), Brockman et al. (2015) or Davis et al. (2015) for evidence on tone sentiment and Dzielinski et al. (2016) for evidence on uncertainty.

tone and uncertainty, we compute standard metrics from the literature to control for these language characteristics.

For tone, we count the number of negative words in earnings calls that appear on the negative word list by Loughran and McDonald (2011). Then, we define $Negativity$ of company i's earnings call in quarter t as the ratio of negative words relative to total words:

$$Negativity_{it} = \frac{\text{Negative words}_{it}}{\text{Total Words}_{it}},$$

To measure uncertainty we use the word list from Loughran and McDonald (2011).[30] Similar to the tone measure, we quantify the uncertainty of statements by counting the number of words in the earnings call that appear on this word list. $Uncertainty$ for the earnings call of company i at time t is then defined as the ratio of uncertain words to total words:

$$Uncertainty_{it} = \frac{\text{Uncertainty words}_{it}}{\text{Total words}_{it}},$$

We follow the approach in Zhou (2018) to generate a variable $Numbers$ that accounts for managements' usage of numbers relative to textual words in their answers. Specifically, we use a regular expression to capture all numbers preceded by a space or a dollar sign and calculate $Numbers_{it}$ for the earnings call of company i at time t :

$$Numbers_{it} = \frac{\text{Number count}_{it}}{\text{Total words}_{it} + \text{Number count}_{it}}.$$

We further calculate for the responses of the management the complexity score proposed by Loughran and McDonald (2019).[31] Using the list of 255 words that proxy for complexity, we build the measure for the earnings call of company i at time t as follows:

$$Complexity_{it} = \frac{\text{Complex words}_{it}}{\text{Total words}_{it}},$$

Finally, we flag follow-up questions by the same analyst during an earnings call and derive a measure of forward-looking words within analysts' questions. For a single question q during the earnings call of company i at time t, we define the share of forward-looking

[30]Note that this word list contains also the word list of weak modals from Loughran and McDonald (2011).

[31]As an alternative metric of complexity, we also calculate the Gunning (1952) Fog index, which is a function of the number of words per sentence (length of a sentence) and the share of complex words (words with more than two syllables) relative to total words. Using this measure does not change any of our results.

words according to the word-lists provided by Bozanic et al. (2018) and Matsumoto et al. (2011):

$$ForwardSentiment_{qit} = \frac{\text{Forward-looking words}_{qit}}{\text{Total words}_{qit}}.$$

Earnings surprise, analyst forecast revisions, and firm characteristics We collect analyst data from IBES and calculate earnings surprises as the difference between the actual and consensus forecast earnings, divided by the share price at five trading days before the announcement in every quarter. Thus, any positive (negative) number indicates better (worse) performance than expected. As in Dzielinski et al. (2016), we rank all firms' earnings surprises into deciles and categorize earnings surprises from 1 (most negative) to 5 (least negative) and from 6 (least positive) to 10 (most positive). Moreover, we collect analyst EPS forecast data and match the latest EPS forecast prior to the call with the first EPS forecast post call. This allows us to flag whether or not the EPS forecast was revised upwards and to calculate for each call the percentage of analysts with a positive revision.

We further collect quarterly balance sheet statistics as well as firms' market capitalizations from Compustat to calculate the book-to-market ratio, the natural logarithm of total assets and Tobin's Q as additional firm characteristics.

Descriptive statistics Table 1 presents descriptive statistics for the variables in this analysis. Our metric $NonAnswer$ and $NonAnswer^{\phi}$ are similar in magnitude with the other sentiments from the dictionary approach, i.e., $Negativity$ and $Uncertainty$. In our sample, $Negativity$ and $Uncertainty$ metrics show an average of 2.8% and 1.6%, which are comparable to estimates found in the literature (see, e.g., (Price et al., 2012) and (Dzielinski et al., 2016)). One might expect a strong correlation between the $NonAnswer$ metric and other sentiment metrics, in particular $Uncertainty$. Yet, as Table A1 shows, $NonAnswer$ only correlates with other textual measures very weakly, highlighting that the measure captures a new dimension of precise information sharing.[32] All measures for cumulative abnormal returns share very similar distributional characteristics.

4.2 How do non-answers affect stock returns?

We attempt to see whether markets react to non-answers and explain cumulative abnormal stock returns for the day of an earnings conference call. For this purpose, we

[32]We present some examples for answers with high (low) $Uncertainty$ and $NonAnswer$ in Appendix A.3.

Table 1: Descriptive statistics

Variable	Obs.	Mean	Std. Dev.	Min	P10	P50	P90	Max
Panel A: Firm-quarter data								
$NonAnswer$	21,129	.074	.015	.029	.056	.073	.093	.14
$NonAnswer^{\phi}$	21,129	.17	.037	.048	.12	.16	.21	.38
$Negativity$	21,129	.028	.0071	.0075	.02	.028	.037	.074
$Uncertainty$	21,129	.016	.0056	0	.0089	.015	.023	.056
$EarnSurp$	21,129	5.7	2.8	1	2	6	10	10
$FF3 - CAR_{0;1}$	21,129	.00039	.052	-.16	-.06	.00032	.062	.15
$FF5 - CAR_{0;1}$	21,129	.00035	.049	-.15	-.057	.00039	.059	.14
$IV_{-1;1}$	19,373	.95	.065	.79	.88	.95	1	1.2
ΔIV	19,372	-.007	.011	-.044	-.021	-.006	.0045	.026
BTM	21,129	.41	.34	-3.2	.1	.34	.78	12
$Ln(Assets)$	21,129	9.5	1.2	5.8	8	9.4	11	14
Q	21,129	2.2	1.4	.63	1.1	1.7	3.8	36
$Numbers$	21,129	.012	.0058	0	.0052	.011	.019	.046
$Complexity$	21,129	.007	.0038	0	.0027	.0064	.012	.035
$\%Positive\,Revisions$	21,129	.45	.16	0	.26	.44	.66	1
Panel B: Q&A-level data								
$NonAnswer$	621,696	.057	.051	0	0	.05	.12	.67
$NonAnswer^{\phi}$	621,696	.15	.2	0	0	.1	.31	6.4
$IsFollowUp_q$	621,696	.66	.47	0	0	1	1	1
$Negativity_q$	621,696	.028	.036	0	0	.018	.07	1
$ForwardSentiment_q$	621,696	.11	.065	0	.012	.11	.19	1

Notes: **Panel A** shows descriptive statistics for our firm-quarter level data with language measures aggregated over all Q&As within an earnings call. **Panel B** provides summary statistics for the language measures on the dimension of individual Q&As. $NonAnswer$ ($NonAnswer^{\phi}$) is the ratio of trigrams in our non-answer glossary (weighted by loadings) to the total words. $Negativity$ and $Uncertainty$ are the ratio of negative and uncertain words to the total words. The list of negative and uncertain words are from Loughran and McDonald (2011) word-lists. $EarnSurp$ represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). $CAR_{0;1}$ is the cumulative abnormal returns in the [0;1] interval around the earnings call. $FF3\_CAR_{0;1}$ and $FF5\_CAR_{0;1}$ use Fama-French three (1993) and five (2015) factor model returns respectively. $IV_{-1;1}$ and ΔIV are the change in option's implied volatility around the earnings call as defined in Section 4.1. BTM defined as total Common/Ordinary Equity divided by the market value of equity. $ln(Assets)$ is the natural logarithm of total assets. Q is the Tobin's Q. $Numbers$ is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). $Complexity$ is the complexity measure derived from the Loughran and McDonald (2019) complexity word list. $\%PositiveRevisions$ is the share of analysts that revise their EPS forecast upwards. $IsFollowUp_q$ is a dummy variable that equals one if a question is a follow up question by the same analyst. $Tone_q$ is the positivity minus negativity sentiment of the question calculated by word count of the corresponding word-lists provided by Loughran and McDonald (2011). $ForwardSentiment_q$ measures the ratio of forward-looking words in a question according to the word-lists provided by Bozanic et al. (2018) and Matsumoto et al. (2011). All return variables are truncated at the 1/99% percentiles.

model cumulative abnormal returns of firm i with management m around the days of the earnings call in quarter t as follows:

$$\begin{aligned} CAR_{imt} = \alpha &+ \beta_1 \cdot NonAnswer_{imt} + \beta_2 \cdot EarnSurp_{imt} \\ &+ \beta_3 \cdot Negativity_{imt} + \beta_4 \cdot Uncertainty_{imt} \\ &+ \theta \cdot X_{imt} + \mu_i + \nu_m + \gamma_t + \epsilon_{imt}. \end{aligned} \tag{5}$$

CAR_{imt} represents the cumulative abnormal return for the initial, short-term reaction ($CAR_{i,t}^{0;1}$) for firm i with management m in quarter t. $NonAnswer$ is the main variable of interest generated from our glossary, which measures management's degree of non-answers in the call in quarter t. If our glossary generates a valid measure for not conveying information, we should observe lower abnormal stock returns for earnings calls with a high $NonAnswer$ and thus expect a negative coefficient for β_1.

We control for three important variables to ensure that outside factors do not affect $NonAnswer$. First, we include a metric that captures investors' expectations of future earnings. As is standard in the literature, we measure the difference between analysts' expectations about earnings and realized earnings as earnings surprise and, for a given point in time, cluster all firms into 10 different groups, $EarnSurp$, with a larger number indicating a more positive earnings surprise.

Second, we control for two variables that have been shown to impact returns after an earnings conference call. One of these measures is $Negativity$, defined as the ratio of negative words to total words used in management's answers. In line with Price et al. (2012), we expect a negative coefficient for β_3. We also control for the vagueness of managements' language, $Uncertainty$. As Dzielinski et al. (2016) show that uncertainty in managements' answers to investors' questions leads to lower stock returns, we expect a negative coefficient for β_4.

Finally, we remove all observable and unobservable firm-specific time-constant variation by including firm fixed effects, μ_i, as well as time (quarter-year) fixed effects, γ_t, to control for firm-constant factors and common trends of abnormal returns in a given quarter, respectively. We further include CEO fixed effects, ν_m, in order to absorb a manager specific component, which neither the current and future performance of the company nor strategic incentives can explain (Davis et al., 2015). This enables us to separate the effect $NonAnswer$ from personal specific unobservable time-constant characteristics. To account for autocorrelations of the errors, we employ two-way clustering (Cameron et al., 2011) and cluster standard errors at the firm and time dimensions.

Table 2 and Table 3 display the results of the regression model outlined in Equation 5 for $CAR_{i,t}^{0;1}$ using $NonAnswer$ and the loading-weighted $NonAnswer^{\phi}$. In both

tables, we observe a negative and highly significant coefficient, highlighting the negative effect that not answering to analysts' questions has on short-term cumulative abnormal returns. These results are in line with our expectation, provided our glossary measures non-answers.[33] Note that this result also holds if we control for earnings surprises, other characteristics of management language, as well as industry, firm and CEO fixed effects and common time trends by quarter-year fixed effects. Moreover, the coefficients of all control variables are in line with our expectations and with the existing literature: we find a positive and highly significant coefficient for the earnings surprise group, i.e. a greater difference in the actual earnings and earnings expected by analysts leads to more positive abnormal returns. We further obtain a negative point estimate for the tone measure, and, in line with Dzielinski et al. (2016), a negative coefficient for the uncertainty measure.

4.3 Which non-answer symptom drives stock returns?

We derive our glossary from two different symptoms of non-answers, rejecting an answer and blathering. To shed light on which of these violations of the "Gricean norms" is more informative in the earnings call context, we derive a separate glossary for each of the two symptoms and repeat the baseline empirical analysis. The results are shown in Table 4. We find that both symptoms on their own lead to negative abnormal returns. However, the blathering-based glossary appears stronger on its own and dominates the rejecting-derived glossary when combining the two variables in one regression.

Note that an immediate use of the MNIR attribute measures does not provide a clean measure of non-answers in general. This is not surprising for the blathering measure in Equation (2), as it only captures non-answers for financial firms. The approach to measuring the refusal of an answer in Gow et al. (2021) suffers from insufficient power, especially when aggregating questions , c.f. Table 4.

4.4 Can random glossaries produce similar results?

In order to show that negative abnormal returns are indeed due to wording that indicates non-answers, we run a Monte Carlo simulation by randomly drawing 1,364 trigrams from the training set 1000 times. For each of these placebo dictionaries, we run regressions as in column 2 of Table 2. The distribution of t-statistics for the $NonAnswer$ coefficient is roughly normal and centered around zero, as shown in Figure 3.[34] The t-statistic for $NonAnswer$ for our original glossary is -4.27 (see Table 2, column 2). This clearly

[33] As a robustness test, we present cross-sectional Fama-MacBeth regressions in Appendix A.4, which confirm our main result.

[34] We consider the distribution of the t-statistic from the 1,000 regression coefficients as we are not only after the effect of the glossary, but also the precision of the point estimate for each draw (glossary).

Table 2: Management *NonAnswer* and abnormal returns ($CAR_{0;1}$)

	$FF3 - CAR_{0;1}$			$FF5 - CAR_{0;1}$		
	(1)	(2)	(3)	(4)	(5)	(6)
NonAnswer	-0.089***	-0.114***	-0.125***	-0.080***	-0.104***	-0.112***
	(-3.52)	(-4.27)	(-2.92)	(-3.30)	(-3.42)	(-2.77)
Negativity	-0.415***	-0.421***	-0.615***	-0.361***	-0.482***	-0.567***
	(-6.53)	(-6.53)	(-7.54)	(-5.98)	(-7.29)	(-7.21)
Uncertainty	-0.089	-0.038	-0.098	-0.071	-0.087	-0.103
	(-1.26)	(-0.52)	(-1.08)	(-1.08)	(-1.11)	(-1.17)
Numbers		-0.185**	-0.273***	-0.166**	-0.213**	-0.262***
		(-2.59)	(-2.86)	(-2.52)	(-2.56)	(-2.99)
Complexity		0.168	0.406***	0.183*	0.226*	0.375***
		(1.64)	(2.72)	(1.97)	(1.97)	(2.69)
Observations	21191	21035	20004	21191	21191	20004
R^2	0.045	0.048	0.132	0.044	0.080	0.130
FirmControls	Yes	Yes	Yes	Yes	Yes	Yes
QuarterYear FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Implied	No	Implied	Implied
Firm FE	No	No	Yes	No	Yes	Yes
CEO FE	No	No	Yes	No	No	Yes

Notes: OLS regressions for Equation (5). The dependent variable is the abnormal returns over the Fama-French three (1993) and five (2015) factor model returns cumulated from the day of the earnings call to the day after it, $FF3 - CAR_{0;1}$ ($FF5 - CAR_{0;1}$). *NonAnswer* is the ratio of trigrams in our non-answer glossary to the total words. *Negativity* and *Uncertainty* are the ratio of negative and uncertain words to the total words. List of negative and uncertain words are from Loughran and McDonald (2011) word-lists. *EarnSurp* represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). *BTM* defined as total Common/Ordinary Equity divided by the market value of equity. $ln(Assets)$ is the natural logarithm of total assets. Q is the Tobin's Q. *Numbers* is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). *Complexity* is the complexity measure derived from the Loughran and McDonald (2019) complexity word list. t-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Table 3: Management $NonAnswer$ and abnormal returns ($CAR_{0;1}$)

	$FF3 - CAR_{0;1}$			$FF5 - CAR_{0;1}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$NonAnswer^{\phi}$	-0.075***	-0.080***	-0.073***	-0.066***	-0.061***	-0.064***
	(-4.41)	(-4.68)	(-3.44)	(-4.16)	(-3.48)	(-3.29)
$Negativity$	-0.447***	-0.441***	-0.636***	-0.397***	-0.517***	-0.594***
	(-7.16)	(-6.98)	(-7.64)	(-6.80)	(-8.38)	(-7.56)
$Uncertainty$	-0.047	-0.036	-0.086	-0.036	-0.074	-0.096
	(-0.63)	(-0.49)	(-0.84)	(-0.52)	(-0.95)	(-0.97)
$Numbers$		-0.150*	-0.124	-0.136*	-0.085	-0.115
		(-1.84)	(-1.08)	(-1.83)	(-0.98)	(-1.09)
$Complexity$		0.102	0.348**	0.129	0.152	0.336**
		(0.91)	(2.28)	(1.25)	(1.21)	(2.38)
Observations	23815	23689	22557	23815	23815	22557
R^2	0.045	0.047	0.125	0.044	0.077	0.124
FirmControls	Yes	Yes	Yes	Yes	Yes	Yes
QuarterYear FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Implied	Yes	Implied	Implied
Firm FE	No	No	Yes	No	Yes	Yes
CEO FE	No	No	Yes	No	No	Yes

Notes: OLS regressions for Equation (5). The dependent variable is the abnormal returns over the Fama-French three (1993) and five (2015) factor model returns cumulated from the day of the earnings call to the day after it, $FF3 - CAR_{0;1}$ ($FF5 - CAR_{0;1}$). $NonAnswer^{\phi}$ is the ratio of trigrams in our non-answer glossary weighted by loadings to the total words. $Negativity$ and $Uncertainty$ are the ratio of negative and uncertain words to the total words. List of negative and uncertain words are from Loughran and McDonald (2011) word-lists. $EarnSurp$ represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). BTM defined as total Common/Ordinary Equity divided by the market value of equity. $ln(Assets)$ is the natural logarithm of total assets. Q is the Tobin's Q. $Numbers$ is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). $Complexity$ is the complexity measure derived from the Loughran and McDonald (2019) complexity word list. t-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Table 4: Different symptoms of management $NonAnswer$

	$FF5 - CAR_{0;1}$				
	(1)	(2)	(3)	(4)	(5)
$NonAnswer$	-0.090***				
	(-3.67)				
$NonAnswer^{Blathering}$		-0.098***		-0.112***	
		(-3.65)		(-3.09)	
$NonAnswer^{Rejection}$			-0.083*	0.048	
			(-1.82)	(0.76)	
$Rejection$					0.000
					(0.32)
$Negativity$	-0.388***	-0.390***	-0.371***	-0.390***	-0.366***
	(-6.53)	(-6.58)	(-6.24)	(-6.58)	(-6.11)
$Uncertainty$	-0.060	-0.064	-0.061	-0.072	-0.079
	(-0.86)	(-0.91)	(-0.91)	(-1.07)	(-1.13)
Observations	21035	21035	21035	21035	21035
R^2	0.046	0.046	0.046	0.046	0.045
FirmControls	Yes	Yes	Yes	Yes	Yes
QuarterYear FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions for Equation (5). The dependent variable is the abnormal returns over the Fama-French three (1993) factor model returns cumulated from the day of the earnings call to the day after it, $FF3 - CAR_{0;1}$. $NonAnswer^{Blathering}$ ($NonAnswer^{Rejecting}$) is the ratio of trigrams in our non-answer glossary derived from the symptom blathering (rejecting) to the total words. $Negativity$ and $Uncertainty$ are the ratio of negative and uncertain words to the total words. List of negative and uncertain words are from Loughran and McDonald (2011) word-lists. $EarnSurp$ represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement). BTM defined as total Common/Ordinary Equity divided by the market value of equity. $ln(Assets)$ is the natural logarithm of total assets. Q is the Tobin's Q. $Numbers$ is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). $Complexity$ is the complexity measure derived from the Loughran and McDonald (2019) complexity word list. t-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

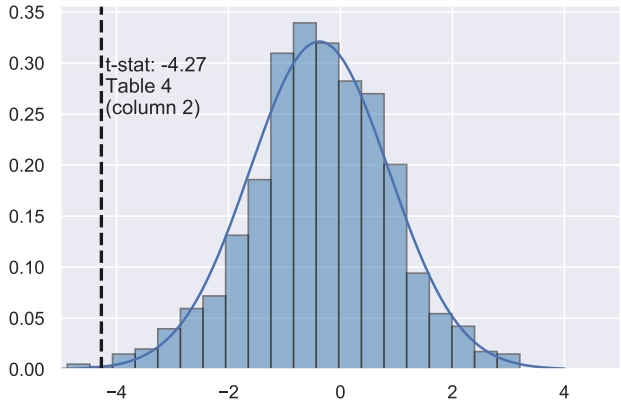


Figure 3: Histogram of t-statistics for the *NonAnswer* coefficient from Equation 5 with 1000 word-lists of 1,364 trigrams randomly selected from the universe of trigrams. The dashed line shows the t-statistic of a regression with our non-answer glossary (column 2 of Table 2).

shows that it would be extremely unlikely to produce economically significant results by randomly drawing a glossary from the universe of trigrams.

4.5 How do non-answers affect expected volatility?

Dodging a question retards the information flow and hinders the reduction of informational asymmetries between the management and investors. We therefore would expect to observe that investor uncertainty is reduced more after earnings calls with low *NonAnswer*. To this extent, we analyze the short-term change as well as the abnormal change in implied volatility, similar to the analysis of abnormal returns above. Table 5 shows the results.

We observe a higher post-earnings-call implied volatility for earnings calls with high *NonAnswer*, i.e., investors are willing to pay a higher premium in order to insure themselves against stock price changes after conference calls in which managers more frequently avoid answering questions.

4.6 How analysts respond to non-answers?

Next, we investigate analysts' responses to non-answers to see whether or not non-answers affect analysts' expectations. In particular, we track analysts' EPS forecasts before and after an earnings call and calculate the percentage of analysts that increase their EPS estimate after the call.

Results are shown in Table 6 for the unweighted (columns 1-2) and weighted non-

Table 5: Management $NonAnswer$ and option implied volatility ($\Delta\ IV$ and $IV_{-1;1}$)

	ΔIV				$IV_{-1;1}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$NonAnswer$	0.024***	0.027***			0.101**	0.116**		
	(2.86)	(3.17)			(2.27)	(2.60)		
$NonAnswer^{\phi}$			0.017***	0.018***			0.067**	0.070***
			(3.51)	(3.67)			(2.57)	(2.66)
$Negativity$		0.018		0.017		0.255**		0.246**
		(0.95)		(0.86)		(2.65)		(2.57)
$Uncertainty$		0.037*		0.035*		0.222**		0.218**
		(1.93)		(1.87)		(2.40)		(2.37)
$Numbers$		0.039		0.037		0.195		0.183
		(1.57)		(1.47)		(1.61)		(1.51)
$Complexity$		0.036		0.037		-0.030		-0.028
		(1.16)		(1.18)		(-0.19)		(-0.18)
Observations	20108	20108	20108	20108	20113	20113	20113	20113
R^2	0.138	0.139	0.138	0.139	0.162	0.164	0.162	0.164
FirmControls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
QuarterYear FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions with the dependent variable $IV_{-1;1}(\Delta IV)$ in the columns 1-4 (5-8) indicating the change in an option's implied volatility around the earnings call as defined in Section 4.1. $NonAnswer$ ($NonAnswer^{\phi}$) is the ratio of trigrams in our non-answer glossary (weighted by loadings) to the total words. $Negativity$ and $Uncertainty$ are the ratio of negative and uncertain words to the total words. List of negative and uncertain words are from Loughran and McDonald (2011) word-lists. $Numbers$ is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). $Complexity$ is the complexity measure derived from the Loughran and McDonald (2019) complexity word list. Firm control variables include $EarnSurp$, BTM, RoE, $ln(Assets)$, and Q. $EarnSurp$ represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement) to 5 negative and 5 (zero and) positive groups. BTM defined as total Common/Ordinary Equity divided by the market value of equity. RoE denotes return on equity. $ln(Assets)$ is the natural logarithm of total assets. Q is the Tobin's Q. t-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

answer score (column 3-4). Calls with more non-answers are significantly more often followed by a non-positive revision of the EPS estimate.

Table 6: Analysts EPS estimate around the call

	%Positive Revisions			
	(1)	(2)	(3)	(4)
NonAnswer	-0.240*	-0.167		
	(-1.71)	(-1.40)		
$NonAnswer^{\phi}$			-0.128**	-0.086*
			(-2.44)	(-1.94)
Negativity	-0.997***	-1.155***	-0.989***	-1.145***
	(-3.48)	(-4.47)	(-3.48)	(-4.50)
Uncertainty	0.098	-0.455	0.128	-0.435
	(0.31)	(-1.63)	(0.40)	(-1.56)
Observations	21129	20973	21129	20973
R^2	0.075	0.112	0.075	0.112
FirmControls	Yes	Yes	Yes	Yes
QuarterYear FE	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes

Notes: This table shows the results for OLS regression with the dependent variable *%PositiveRevisions*, measuring the share of analysts with an upward revision of the EPS forecast after an earnings call. *NonAnswer* ($NonAnswer^{\phi}$) is the ratio of trigrams in our non-answer glossary (weighted by loadings) to the total words. *Negativity* and *Uncertainty* are the ratio of negative and uncertain words to the total words. List of negative and uncertain words are from Loughran and McDonald (2011) word-lists. Firm controls include *EarnSurp*, *BTM*, *ln*(*Assets*) and Tobin's *Q*. *EarnSurp* represents the grouping of all firms in deciles of earnings surprise (defined as the difference between the actual and the consensus forecast earnings as a ratio to the share price 5 trading days before the announcement) to 5 negative and 5 (zero and) positive groups. *BTM* defined as total Common/Ordinary Equity divided by the market value of equity. *RoE* denotes return on equity. *ln*(*Assets*) is the natural logarithm of total assets. *Q* is the Tobin's Q. *t*-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

4.7 Which questions receive non-answers?

Second, we analyze which questions managers avoid precisely responding to during earnings calls. It is reasonable to assume that the management is more likely not to answer

disadvantageous and tougher questions, as they want to evade critical questions.[35] We proxy for critical questions in two ways. First, we calculate the negativity sentiment for each question, assuming that a more negative tone reflects a more critical question. Second, we flag whether a question is a follow-up question by the same analyst.[36]

It is also likely that the management may not be able to answer questions about future outcomes. Thus, we add the ratio of forward-looking phrases in a question as an additional dimension that may result in a non-answer and then regress $NonAnswer$ on these metrics. Note that by analyzing questions and answers within the same earnings conference call, we can control for many observable and unobservable factors, such as management or firm characteristics.

Table 7: $NonAnswer$ in response to follow-up questions.

	$NonAnswer$			$NonAnswer^{\phi}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$IsFollowUp_q$	0.0016***	0.0017***	0.0021***	0.0085***	0.0088***	0.0094***
	(10.66)	(11.63)	(14.13)	(15.95)	(16.54)	(17.49)
$Negativity_q$		0.0207***	0.0194***		0.0450***	0.0432***
		(10.17)	(9.57)		(5.54)	(5.32)
$ForwardSentiment_q$			0.0304***			0.0449***
			(25.72)			(9.86)
Observations	621696	621696	621696	621696	621696	621696
R^2	0.065	0.065	0.066	0.049	0.049	0.049
EarningsCall FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the results for OLS regression with the dependent variable $NonAnswer$ ($NonAnswer^{\phi}$) in columns 1 to 3 (4 to 6). $IsFollowUp_q$ is a dummy variable that equals one if a question is a follow up question by the same analyst. $Negativity_q$ is the negativity sentiment of the question calculated by word count of the corresponding word-lists provided by Loughran and McDonald (2011). $ForwardSentiment_q$ measures the ratio of forward-looking words in a question according to the word-lists provided by Bozanic et al. (2018) and Matsumoto et al. (2011). All the specifications control for earnings call fixed effects. t-statistics are given in parentheses. Standard errors are clustered at the earnings call level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Table 7 shows the regression results for $NonAnswer$ (columns 1 - 3) and $NonAnswer^{\phi}$ (columns 4 - 6). We find that $NonAnswer$ is greater in responses to follow up questions, in line with our expectation. In addition, we observe that questions with a more negative tone are associated with a higher value for $NonAnswer$, indicating that managers more

[35]See, e.g., Mayew (2008) or Cohen et al. (2020) who show that managers try to avoid unfavorable questions by not allowing unfavorable analysts to ask a question.

[36]The second proxy is based on Clayman (1993), who argues that analysts have the capacity to recognize and counter evasive answers by asking a follow-up question, so that follow-up questions by the same analyst are usually more critical.

often dodge critical questions. Finally, we find higher values for *NonAnswer* in response to forward looking questions.

5 Conclusion

The asymmetric distribution of information is considered a key friction in economics. While question and answer (Q&A) settings are designed to remove information asymmetries, the addressee of a question does not necessarily provide the requested information. Building on a large textual dataset of questions and answers and employing a supervised machine learning framework, we generate a generalizable glossary that can identify non-answers.

Using a Multinomial Inverse Regression (Taddy, 2013b), we identify a glossary of 1,364 trigrams such as 'back to you', 'do not know', 'hard to predict', etc., which are frequently used to refrain from answering a question in a concise and factual manner. The glossary is derived from earnings conference calls, where investors and analysts can directly question senior managers' during Q&A sessions. However, as non-answers do not contain any context- or industry-specific vocabulary, the glossary is applicable to a broad Q&A context, such as sports or political interviews and senate hearings.

We provide evidence for the plausibility and economic relevance of the glossary. In particular, we apply the glossary to market reactions after earnings conference calls for a large sample of firms over a span of 16 years. In regressions with multiple control variables, we show a strong negative impact for the measure derived from our glossary, i.e., not answering analysts' questions leads on average to negative abnormal stock returns after an earnings call. We also link our measure to option implied volatilities after earnings conference calls and show that investor uncertainty increases if management fails to provide the information requested in the call. Both results are in line with the theoretical asset pricing literature, which suggests that higher uncertainty translates to larger risk premia (Andrei and Hasler, 2014).

We further observe that financial analysts provide less often a positive update of their EPS estimates after calls with high non-answers and document that non-answers are observed more prevalently for tougher and more critical questions (Mayew, 2008; Cohen et al., 2020). Within the same conference call, *NonAnswer* is higher for managements' responses to follow-up questions by the same analyst and for managements' responses to more negative questions. We also observe higher *NonAnswer* for questions with forward-looking sentences, i.e., questions that refer to (potentially unknown) future outcomes.

A non-answer does not require a specific context. As both our method and glossary are free of financial context, we believe that the measure is applicable to other fields with

a Q&A setup.

References

Andrei, D. and Hasler, M. (2014). Investor Attention and Stock Market Volatility. *Review of Financial Studies*, 28(1):33–72.

Barth, A., Mansouri, S., Woebbeking, F., and Zörgiebel, S. (2021). How to talk down your stock performance. Working paper.

Bekkerman, R. and Allan, J. (2004). Using bigrams in text categorization. Technical report, Technical Report IR-408, Center of Intelligent Information Retrieval, UMass Amherst.

Blau, B. M., DeLisle, J. R., and Price, S. M. K. (2015). Do sophisticated investors interpret earnings conference call tone differently than investors at large? Evidence from short sales. *Journal of Corporate Finance*, 31:203–219.

Bloomfield, R. (2008). Discussion of "annual report readability, current earnings, and earnings persistence". *Journal of Accounting and Economics*, 45(2-3):248–252.

Bonsall, S. B., Leone, A. J., Miller, B. P., and Rennekamp, K. (2017). A plain english measure of financial reporting readability. *Journal of Accounting and Economics*, 63(2-3):329–357.

Bozanic, Z., Roulstone, D. T., and Van Buskirk, A. (2018). Management earnings forecasts and other forward-looking statements. *Journal of Accounting and Economics*, 65(1):1–20.

Brockman, P., Li, X., and Price, S. M. (2015). Differences in conference call tones: Managers vs. analysts. *Financial Analysts Journal*, 71(4):24–42.

Buller, D. B. and Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, 6(3):203–242.

Bushee, B. J., Gow, I. D., and Taylor, D. J. (2018). Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research*, 56(1):85–121.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

Carlin, B. I., Longstaff, F. A., and Matoba, K. (2014). Disagreement and asset prices. *Journal of Financial Economics*, 114(2):226 – 238.

Clayman, S. (1993). Reformulating the question: A device for answering/not answering questions in news interviews and press conferences. *Text - Interdisciplinary Journal for the Study of Discourse*, 13.

Cohen, L., Lou, D., and Malloy, C. J. (2020). Casting conference calls. *Management Science*, 66(11):5015–5039.

Cook, R. D. et al. (2007). Fisher lecture: Dimension reduction in regression. *Statistical Science*, 22(1):1–26.

Dave, K., Lawrence, S., and Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, pages 519–528.

Davis, A. K., Ge, W., Matsumoto, D., and 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.

Dzielinski, M., Wagner, A. F., and Zeckhauser, R. J. (2016). In no (un)certain terms: Managerial style in communicating earnings news. Working Paper.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, 87(4):1307–1340.

Gow, I. D., Larcker, D. F., and Zakolyukina, A. A. (2021). Non-answers during conference calls. *Journal of Accounting Research*, forthcoming.

Grice, P. H. (1989). *Studies in the Way of Words*. Harvard University Press., Cambridge, MA.

Guay, W., Samuels, D., and Taylor, D. (2016). Guiding through the fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics*, 62(2-3):234–269.

Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill, New York.

Harvey, C. R. (2016). Harvey's hypertextual finance glossary. *URL: https://people. duke. edu/˜ charvey/Classes/wpg/bfglosm. htm [Version 8 November 2016.]*.

Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer, 2nd edition.

Henry, E. (2008). Are Investors Influenced By How Earnings Press Releases Are Written? *Journal of Business Communication*, 45(4):363–407.

Hollander, S., Pronk, M., and Roelofsen, E. (2010). Does silence speak? an empirical analysis of disclosure choices during conference calls. *Journal of Accounting Research*, 48(3):531–563.

Jelinek, F. (1991). Up from trigrams!-the struggle for improved language models. In *Second European Conference on Speech Communication and Technology*.

Lee, J. (2015). Can investors detect managers' lack of spontaneity? adherence to predetermined scripts during earnings conference calls. *The Accounting Review*, 91(1):229–250.

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

Liu, L. X., Shu, H., and Wei, K. J. (2017). The impacts of political uncertainty on asset prices: Evidence from the bo scandal in china. *Journal of Financial Economics*, 125(2):286 – 310.

Loughran, T. and McDonald, B. (2011). When is a Liability not a Liability? *Journal of Finance*, 66(1):35 – 65.

Loughran, T. and McDonald, B. (2014). Measuring readability in financial disclosures. *Journal of Finance*, 69(4):1643–1671.

Loughran, T. and McDonald, B. (2019). Measuring firm complexity. Working paper.

Matsumoto, D., Pronk, M., and 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.

Mayew, W. J. (2008). Evidence of management discrimination among analysts during earnings conference calls. *Journal of Accounting Research*, 46(3):627–659.

McCornack, S. A., Levine, T. R., Solowczuk, K. A., Torres, H. I., and Campbell, D. M. (1992). When the alteration of information is viewed as deception: An empirical test of information manipulation theory. *Communication Monographs*, 59(1):17–29.

Pak, A. and Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10, pages 1320–1326.

Pástor, L. and Veronesi, P. (2013). Political uncertainty and risk premia. *Journal of Financial Economics*, 110(3):520 – 545.

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

Rogers, J. L., Skinner, D. J., and Buskirk, A. V. (2009). Earnings guidance and market uncertainty. *Journal of Accounting and Economics*, 48(1):90 – 109.

Rogers, T. and Norton, M. I. (2011). The artful dodger: Answering the wrong question the right way. *Journal of Experimental Psychology: Applied*, 17(2):139 – 147.

Taddy, M. (2013a). The gamma lasso. *arXiv preprint arXiv:1308.5623*.

Taddy, M. (2013b). Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, 108(503):755–770.

Taddy, M. (2015). Distributed multinomial regression. *The Annals of Applied Statistics*, 9(3):1394–1414.

Zhou, D. (2018). Do numbers speak louder than words? Technical report.

Zuckerberg, M. Facebook, social media privacy, and the use and abuse of data. Joint hearing before the COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION UNITED STATES SENATE and the COMMITTEE ON THE JUDICIARY UNITED STATES SENATE, One Hundred Fifteenth Congress, Second Session.

A Appendix

A.1 The Non-Answer Glossary

A lists of all trigrams in the glossary with their corresponding loading ϕ. A machine readable version of the glossary is available at econlinguistics.org

Token	ϕ	Token	ϕ	Token	ϕ	Token	ϕ	Token	ϕ
back.to.you	20.8	i.can.say	6.4	made.a.lot	5.3	of.the.third	4.5	and.then.well	4.1
not.sure.i	13.9	not.think.it	6.4	your.first.que		take.some.time	4.5	get.a.lot	4.1
early.to.tell	13.8	not.think.i	6.4	stion	5.2	we.will.get	4.5	couple.of.weeks	4.1
going.to.let	13.3	well.you.know	6.4	only.thing.i	5.2	to.see.how	4.5	for.us.its	4.1
of.my.head	13.0	work.to.do	6.4	we.get.into	5.2	last.year.i	4.5	i.would.like	4.0
top.of.my	12.3	think.it.has	6.3	a.lot.in	5.1	second.quarter		at.least.for	4.0
on.that.one	12.1	maybe.a.little	6.3	do.not.know	5.1	.but	4.5	im.not.going	4.0
to.answer.that	12.1	early.in.the	6.3	go.ahead.and	5.1	we.got.to	4.5	do.not.think	4.0
off.the.top	12.0	in.the.coming	6.3	quarter.but.i	5.1	through.the.fi		the.top.of	4.0
get.back.to	11.9	were.going.thr		we.get.through	5.1	rst	4.5	think.is.going	4.0
do.you.want	11.6	ough	6.3	just.a.little	5.1	talked.about.it	4.5	the.next.two	4.0
answer.that.qu		not.know.if	6.3	answer.your.qu		the.year.i	4.4	think.you.would	4.0
estion	11.3	know.i.think	6.2	estion	5.1	so.its.hard	4.4	i.tried.to	4.0
want.to.take	11.3	year.and.so	6.2	all.the.things	5.1	for.next.year	4.4	better.than.we	4.0
hard.for.me	10.7	we.can.make	6.1	the.time.we	5.1	the.answer.is	4.4	were.pleased.w	
its.too.early	10.4	i.have.got	6.1	the.first.two	5.0	by.the.end	4.4	ith	4.0
to.wait.and	9.6	next.year.but	6.1	so.im.not	5.0	need.to.do	4.4	to.make.it	4.0
front.of.me	9.3	but.we.certain		we.are.right	5.0	give.us.a	4.4	to.talk.about	4.0
not.know.the	9.0	ly	6.1	second.half.of	5.0	answer.to.that	4.4	to.go.through	3.9
wait.and.see	8.8	to.be.done	6.1	there.is.nothi		probably.going		an.area.that	3.9
if.you.would	8.8	not.think.we	6.0	ng	5.0	.to	4.4	it.takes.a	3.9
too.early.to	8.7	to.get.it	6.0	much.as.we	5.0	the.first.ques		my.guess.is	3.9
talk.about.that	8.6	hard.to.say	6.0	focused.on.it	5.0	tion	4.4	to.come.out	3.9
thank.you.for	8.6	something.we.h		i.said.i	5.0	know.that.we	4.4	not.think.its	3.9
comment.on.that	8.5	ave	6.0	we.said.we	5.0	to.the.end	4.3	quarter.and.i	3.9
want.to.talk	8.2	from.that.stan		just.say.that	4.9	answer.is.yes	4.3	at.the.beginni	
come.back.to	7.9	dpoint	6.0	to.take.that	4.9	year.and.we	4.3	ng	3.9
have.to.get	7.9	doing.a.lot	6.0	done.a.lot	4.9	the.people.that	4.3	but.having.said	3.9
tell.you.what	7.9	think.we.need	6.0	have.to.see	4.9	its.a.little	4.3	to.get.there	3.9
do.not.disclose	7.8	not.know.that	6.0	to.see.what	4.9	as.we.get	4.3	to.comment.on	3.9
to.get.back	7.8	i.can.give	5.9	this.is.going	4.9	to.see.if	4.3	we.havent.seen	3.9
and.the.team	7.8	at.this.stage	5.9	you.want.to	4.9	want.to.go	4.3	the.bank.of	3.9
by.the.time	7.7	the.second.half	5.9	in.any.way	4.9	the.year.and	4.3	say.we.have	3.9
am.going.to	7.7	said.that.i	5.9	go.back.and	4.9	first.quarter.		in.a.little	3.9
i.am.going	7.6	think.what.i	5.9	want.to.comment	4.8	is	4.3	of.our.custome	
im.going.to	7.6	and.well.be	5.9	this.year.i	4.8	to.go.back	4.3	rs	3.9
have.to.wait	7.5	think.its.going	5.9	and.his.team	4.8	know.that.i	4.3	in.i.think	3.9
of.the.world	7.5	not.think.you	5.9	and.then.ill	4.8	bit.more.than	4.3	want.to.add	3.8
i.really.do	7.5	lot.of.work	5.9	would.say.i	4.8	one.way.or	4.2	the.way.they	3.8
im.not.sure	7.4	the.first.half	5.8	of.the.us	4.8	of.the.people	4.2	as.we.work	3.8
talk.about.it	7.3	theres.a.little	5.8	yes.let.me	4.8	you.know.its	4.2	youre.referrin	
yes.i.do	7.2	well.have.to	5.8	would.say.this	4.8	try.to.get	4.2	g.to	3.8
to.be.honest	7.2	the.only.thing	5.7	were.investing		first.part.of	4.2	to.take.some	3.8
and.im.not	7.0	i.would.probab		.in	4.8	the.next.quart		where.we.want	3.8
not.see.anythi		ly	5.7	year.so.i	4.7	er	4.2	you.know.what	3.8
ng	7.0	be.a.bit	5.7	you.know.i	4.7	i.would.be	4.2	so.i.wouldnt	3.8
of.the.question	6.9	have.a.better	5.6	with.our.custo		i.will.let	4.2	think.that.was	3.8
number.of.peop		to.the.next	5.6	mers	4.7	as.good.as	4.2	will.take.a	3.8
le	6.9	think.its.real		of.an.impact	4.7	we.hope.to	4.2	on.the.call	3.8
think.there.was	6.9	ly	5.6	going.to.make	4.7	i.have.seen	4.1	of.these.things	3.8
to.play.out	6.8	quarter.and.it	5.5	into.the.year	4.7	as.we.speak	4.1	got.to.be	3.7
year.that.we	6.8	the.first.part	5.5	and.we.havent	4.7	customers.that		to.the.fourth	3.7
to.tell.you	6.7	we.would.say	5.5	would.say.its	4.7	.we	4.1	the.second.part	3.7
on.the.ground	6.7	to.say.we	5.5	not.know.how	4.7	i.cannot.give	4.1	feel.very.comf	
outside.the.us	6.7	on.this.call	5.5	first.half.of	4.6	not.think.there	4.1	ortable	3.7
in.front.of	6.6	take.a.little	5.5	little.bit.hig		the.year.that	4.1	way.i.would	3.7
that.number.is	6.6	not.know.what	5.4	her	4.6	i.think.well	4.1	first.quarter.	
the.second.and	6.6	early.part.of	5.4	be.a.little	4.6	a.little.higher	4.1	but	3.7
towards.the.end	6.6	first.thing.i	5.4	just.trying.to	4.6	is.the.first	4.1	want.to.give	3.7
hard.to.predict	6.5	we.got.a	5.3	be.part.of	4.6	cannot.give.you	4.1	do.not.give	3.7
going.to.say	6.5	can.give.you	5.3	of.the.economy	4.6	the.number.that	4.1	continue.to.in	
what.we.know	6.5	it.might.be	5.3	this.year.but	4.6	not.sure.that	4.1	vest	3.7
		said.we.would	5.3	the.year.so	4.5	the.answer.to	4.1	year.and.then	3.7

Token	ϕ
back.half.of	3.7
say.that.i	3.7
things.that.i	3.7
the.back.half	3.7
to.work.through	3.7
think.what.were	3.7
think.that.its	3.7
later.in.the	3.7
i.know.you	3.7
i.gave.you	3.7
a.really.good	3.7
think.at.the	3.7
we.might.have	3.6
know.what.the	3.6
but.i.wouldnt	3.6
half.of.the	3.6
get.into.the	3.6
over.a.period	3.6
pleased.with.t	
he	3.6
on.the.second	3.6
when.i.look	3.6
time.and.i	3.6
think.you.can	3.6
we.know.we	3.6
think.all.of	3.6
fair.to.say	3.6
into.the.second	3.6
last.year.that	3.6
just.add.to	3.6
to.happen.in	3.5
what.were.going	3.5
with.the.regul	
ators	3.5
but.i.cannot	3.5
the.thing.that	3.5
second.part.of	3.5
we.have.spent	3.5
it.is.today	3.5
will.say.that	3.5
think.were.goi	
ng	3.5
like.to.see	3.5
have.to.make	3.5
year.and.that	3.5
i.think.whats	3.5
we.can.get	3.5
our.customers.	
and	3.5
give.you.a	3.5
have.a.great	3.5
the.competitiv	
e.environment	3.5
still.going.to	3.5
into.the.fourth	3.5
what.we.said	3.5
have.got.to	3.5
think.there.are	3.5
are.seeing.the	3.5
a.little.better	3.5
its.one.of	3.5
but.you.know	3.4
and.the.first	3.4

Token	ϕ
the.right.way	3.4
we.have.worked	3.4
the.business.so	3.4
was.going.to	3.4
that.we.thought	3.4
do.not.break	3.4
for.the.year	3.4
it.looks.like	3.4
i.should.say	3.4
think.as.i	3.4
quarter.of.last	3.4
cannot.tell.you	3.4
it.will.take	3.4
it.is.hard	3.4
thats.somethin	
g.that	3.4
no.i.think	3.3
want.to.say	3.3
this.year.and	3.3
used.to.be	3.3
not.think.ther	
es	3.3
going.to.work	3.3
of.next.year	3.3
give.you.more	3.3
well.let.me	3.3
going.to.need	3.3
it.could.be	3.3
about.the.fact	3.3
the.third.and	3.3
i.will.say	3.3
as.soon.as	3.3
going.to.happen	3.3
would.just.say	3.3
trying.to.be	3.3
the.next.year	3.3
were.not.seeing	3.3
our.clients.are	3.2
quarter.that.we	3.2
be.a.lot	3.2
to.help.us	3.2
turn.it.over	3.2
i.would.have	3.2
thought.we.wou	
ld	3.2
think.that.were	3.2
think.this.is	3.2
well.be.able	3.2
to.answer.your	3.2
an.opportunity	
.for	3.2
question.i.thi	
nk	3.2
is.a.business	3.2
going.to.do	3.2
were.excited.a	
bout	3.2
the.first.and	3.2
fourth.quarter	
.and	3.2
we.can.see	3.2
whats.going.to	3.2
our.businesses	

Token	ϕ
.and	3.2
around.the.glo	
be	3.2
there.might.be	3.2
think.that.wou	
ld	3.2
think.you.shou	
ld	3.2
right.way.to	3.2
and.kind.of	3.2
think.they.are	3.2
been.focused.on	3.2
were.seeing.so	
me	3.2
and.were.going	3.2
what.we.thought	3.2
of.the.year	3.2
fourth.quarter	
.as	3.2
to.take.a	3.2
way.i.think	3.1
not.think.the	3.1
i.think.i	3.1
the.us.and	3.1
on.that.front	3.1
until.we.get	3.1
said.we.have	3.1
of.the.first	3.1
happen.in.the	3.1
i.have.said	3.1
not.think.were	3.1
the.other.thin	
gs	3.1
on.the.last	3.1
just.looking.at	3.1
i.cannot.tell	3.1
the.commercial	
.side	3.1
that.it.would	3.1
i.wouldnt.expe	
ct	3.1
for.many.years	3.1
a.good.job	3.1
to.be.much	3.1
think.that.the	
res	3.1
of.the.changes	3.1
have.to.go	3.1
be.much.more	3.1
that.right.now	3.1
is.not.somethi	
ng	3.1
for.the.second	3.1
good.about.it	3.1
that.is.going	3.1
we.really.have	
nt	3.1
things.that.are	3.0
need.to.get	3.0
to.come.up	3.0
i.think.thats	3.0
do.not.feel	3.0
other.things.t	

Token	ϕ
hat	3.0
know.if.you	3.0
a.great.questi	
on	3.0
going.to.get	3.0
we.get.to	3.0
i.guess.what	3.0
we.are.going	3.0
its.hard.to	3.0
that.its.going	3.0
an.awful.lot	3.0
the.i.think	3.0
you.a.little	3.0
continue.to.dr	
ive	3.0
have.made.a	3.0
is.a.lot	3.0
in.the.next	3.0
in.new.york	2.9
i.think.were	2.9
but.we.havent	2.9
talked.about.a	2.9
a.couple.years	2.9
of.years.ago	2.9
happy.with.the	2.9
third.and.four	
th	2.9
have.been.pret	
ty	2.9
quarter.i.would	2.9
lot.of.time	2.9
think.it.will	2.9
they.have.got	2.9
it.is.something	2.9
in.the.process	2.9
products.and.s	
ervices	2.9
think.it.is	2.9
next.few.quart	
ers	2.9
parts.of.our	2.9
guess.i.would	2.9
about.a.year	2.9
got.a.very	2.9
we.will.take	2.9
get.a.little	2.9
the.things.we	2.9
in.the.growth	2.9
going.to.start	2.9
thats.really.w	
hat	2.9
say.that.were	2.9
other.parts.of	2.9
think.thats.wh	
at	2.9
in.that.catego	
ry	2.9
but.right.now	2.9
not.a.big	2.9
quarter.last.y	
ear	2.9
think.thats.the	2.9
this.quarter.i	2.8

Token	ϕ
i.think.they	2.8
but.i.think	2.8
number.that.we	2.8
expect.to.be	2.8
we.will.see	2.8
thinking.about	
.it	2.8
of.your.questi	
on	2.8
were.still.in	2.8
do.not.want	2.8
that.business.	
and	2.8
of.people.that	2.8
would.say.it	2.8
really.want.to	2.8
tell.you.is	2.8
and.fourth.qua	
rter	2.8
i.think.it	2.8
that.could.be	2.8
think.i.would	2.8
actions.that.we	2.8
think.it.was	2.8
the.way.i	2.8
there.is.going	2.8
of.things.that	2.8
for.some.time	2.8
the.month.of	2.8
both.sides.of	2.8
right.now.that	2.8
would.add.to	2.8
each.one.of	2.8
to.do.something	2.7
first.and.seco	
nd	2.7
spent.a.lot	2.7
us.i.think	2.7
think.you.know	2.7
the.next.few	2.7
the.united.sta	
tes	2.7
been.a.little	2.7
be.happy.to	2.7
would.have.exp	
ected	2.7
couple.of.mont	
hs	2.7
fourth.quarter	
.so	2.7
and.theres.a	2.7
the.consumer.s	
ide	2.7
continue.to.fo	
cus	2.7
might.have.been	2.7
still.in.the	2.7
other.part.of	2.7
not.want.to	2.7
this.quarter.it	2.7
well.we.are	2.7
would.say.in	2.7
it.i.think	2.7

Token	ϕ	Token	ϕ	Token	ϕ	Token	ϕ	Token	ϕ
way.we.think	2.7	i.think.there	2.5	i.will.tell	2.3	that.i.think	2.2	mean.i.think	2.1
its.not.going	2.7	in.the.numbers	2.5	to.make.sure	2.3	have.a.good	2.2	the.year.we	2.1
of.the.second	2.7	in.the.second	2.5	a.year.or	2.3	like.that.but	2.2	continue.to.do	2.1
its.going.to	2.7	and.so.well	2.5	are.not.seeing	2.3	the.last.quart		were.a.little	2.1
part.of.your	2.7	just.going.to	2.5	i.would.not	2.3	er	2.2	to.drive.that	2.1
with.a.lot	2.7	think.its.fair	2.5	really.focused		right.now.i	2.2	we.need.to	2.1
it.is.still	2.7	but.i.would	2.5	.on	2.3	last.year.and	2.2	to.be.careful	2.1
little.bit.from	2.7	as.strong.as	2.5	little.bit.abo		i.mean.we	2.2	so.i.guess	2.1
for.the.next	2.7	think.that.is	2.5	ut	2.3	to.see.us	2.2	year.and.i	2.0
able.to.do	2.7	into.the.first	2.5	would.probably		on.the.corpora		i.think.about	2.0
spend.a.lot	2.7	for.us.but	2.5	.be	2.3	te	2.2	to.make.a	2.0
i.think.this	2.6	beginning.of.t		going.to.have	2.3	most.important		things.we.have	2.0
very.pleased.w		he	2.5	business.and.t		.thing	2.2	late.in.the	2.0
ith	2.6	wanted.to.make	2.5	hat	2.3	in.the.economy	2.2	make.sure.that	2.0
a.whole.lot	2.6	will.probably.		what.we.expect	2.3	it.is.going	2.2	focused.on.that	2.0
said.i.think	2.6	be	2.5	have.said.that	2.3	in.q.and	2.2	i.said.before	2.0
bit.on.the	2.6	the.economy.is	2.5	and.we.certain		thats.a.good	2.2	second.and.thi	
with.us.and	2.6	confident.that		ly	2.3	not.going.to	2.2	rd	2.0
to.work.on	2.6	.we	2.5	first.quarter.		quarter.i.think	2.2	this.year.so	2.0
well.and.we	2.6	to.sort.of	2.4	and	2.3	make.sure.we	2.2	the.economy.and	2.0
you.know.a	2.6	end.of.this	2.4	us.to.do	2.3	the.end.of	2.2	that.well.be	2.0
that.a.lot	2.6	where.we.need	2.4	say.that.we	2.3	will.see.that	2.2	i.wouldnt.say	2.0
think.youre.go		i.mean.i	2.4	the.right.thing	2.3	was.a.good	2.2	go.back.to	2.0
ing	2.6	i.think.from	2.4	i.think.just	2.3	for.a.second	2.2	have.gone.thro	
we.made.a	2.6	the.business.in	2.4	to.go.into	2.3	last.year.but	2.2	ugh	2.0
they.want.to	2.6	i.guess.i	2.4	a.business.that	2.3	just.in.terms	2.2	we.thought.we	2.0
think.that.you	2.6	its.a.good	2.4	were.very.plea		that.is.probab		right.i.think	2.0
will.tell.you	2.6	i.agree.with	2.4	sed	2.3	ly	2.2	in.the.united	2.0
a.sense.of	2.6	think.that.if	2.4	is.an.area	2.3	want.to.see	2.2	through.the.ye	
well.i.would	2.6	part.of.what	2.4	third.quarter.		little.bit.bet		ar	2.0
to.deal.with	2.6	again.i.think	2.4	so	2.3	ter	2.2	think.we.can	2.0
it.that.way	2.6	that.we.made	2.4	a.little.more	2.3	are.things.that	2.2	right.now.so	2.0
not.see.it	2.6	think.its.a	2.4	a.question.of	2.3	those.things.a		of.our.busines	
on.the.consumer	2.6	this.year.that	2.4	you.know.if	2.3	re	2.2	ses	2.0
and.the.second	2.6	were.thinking.		right.now.but	2.3	last.two.years	2.2	the.beginning.	
yes.i.would	2.6	about	2.4	think.we.were	2.3	and.well.see	2.2	of	2.0
continue.to.ma		the.last.years	2.4	were.trying.to	2.3	i.just.think	2.1	half.of.this	2.0
ke	2.6	try.to.do	2.4	in.the.fourth	2.3	a.good.question	2.1	awful.lot.of	2.0
need.to.be	2.6	the.pace.of	2.4	see.a.little	2.3	going.to.take	2.1	business.and.i	2.0
terms.of.how	2.6	think.the.way	2.4	the.world.and	2.2	is.a.great	2.1	a.bit.more	2.0
quarter.it.was	2.6	quarter.but.we	2.4	the.benefits.of	2.2	would.love.to	2.1	would.like.to	2.0
of.our.growth	2.6	year.but.i	2.4	had.a.little	2.2	thats.a.little	2.1	impact.in.the	2.0
very.difficult		in.europe.and	2.4	two.or.three	2.2	us.we.have	2.1	during.the.cou	
.to	2.6	theres.still.a	2.4	in.the.early	2.2	outside.of.the	2.1	rse	2.0
in.fact.i	2.6	we.expect.it	2.4	you.think.of	2.2	i.think.again	2.1	the.way.you	2.0
that.i.mean	2.6	fourth.quarter		see.how.that	2.2	going.to.keep	2.1	year.in.the	2.0
from.our.persp		.of	2.4	business.that.		lot.of.people	2.1	trying.to.do	2.0
ective	2.6	bit.of.an	2.4	we	2.2	as.you.said	2.1	i.think.what	2.0
one.thing.i	2.6	things.that.we	2.4	the.question.is	2.2	sure.that.were	2.1	every.one.of	1.9
year.or.so	2.6	think.we.do	2.4	in.the.right	2.2	i.think.we	2.1	the.last.three	1.9
that.we.might	2.6	would.say.is	2.4	thats.why.i	2.2	i.can.tell	2.1	fourth.quarter	
have.done.that	2.5	think.we.had	2.4	not.something.		want.to.make	2.1	.is	1.9
have.seen.it	2.5	our.clients.and	2.4	that	2.2	we.will.look	2.1	so.we.expect	1.9
deal.with.the	2.5	i.want.to	2.4	to.see.it	2.2	year.ago.and	2.1	the.high.end	1.9
at.this.time	2.5	the.latter.part	2.4	lot.of.things	2.2	think.were.in	2.1	add.to.that	1.9
and.i.know	2.5	i.said.earlier	2.4	a.bit.in	2.2	i.would.just	2.1	latter.part.of	1.9
talk.a.little	2.5	of.last.year	2.4	the.fourth.qua		time.but.we	2.1	over.a.year	1.9
said.we.are	2.5	think.we.have	2.4	rter	2.2	give.you.the	2.1	little.bit.low	
what.we.need	2.5	the.year.to	2.4	into.next.year	2.2	the.us.we	2.1	er	1.9
the.last.two	2.5	pick.up.in	2.4	would.say.a	2.2	around.the.wor		the.growth.rate	1.9
year.or.two	2.5	something.that		the.process.of	2.2	ld	2.1	see.some.of	1.9
of.this.year	2.5	.we	2.3	i.think.its	2.2	last.year.so	2.1	can.tell.you	1.9
us.so.we	2.5	so.i.cannot	2.3	to.give.you	2.2	like.that.we	2.1	of.the.fourth	1.9

Token	ϕ
think.it.would	1.9
have.a.number	1.9
just.want.to	1.9
sense.for.us	1.9
it.will.contin	
ue	1.9
quarter.so.that	1.9
us.and.i	1.9
and.we.expect	1.9
but.we.feel	1.9
types.of.things	1.9
when.we.get	1.9
talked.about.we	1.9
think.we.are	1.9
what.i.said	1.9
the.good.news	1.9
we.would.certa	
inly	1.9
to.get.that	1.9
will.look.at	1.9
of.the.best	1.9
as.we.looked	1.9
thing.that.i	1.9
look.i.think	1.9
a.very.competi	
tive	1.9
based.on.what	1.9
first.quarter.i	1.9
to.respond.to	1.9
businesses.tha	
t.we	1.8
year.so.we	1.8
yes.so.i	1.8
have.been.sayi	
ng	1.8
its.kind.of	1.8
thats.what.were	1.8
think.about.it	1.8
it.makes.sense	1.8
i.think.for	1.8
next.couple.of	1.8
in.the.sense	1.8
to.invest.in	1.8
were.going.to	1.8
got.a.lot	1.8
have.been.inve	
sting	1.8
i.think.people	1.8
sure.that.we	1.8
business.for.us	1.8
were.focused.on	1.8
i.would.look	1.8
give.you.an	1.8
of.the.growth	1.8
i.do.think	1.8
the.third.quar	
ter	1.8
and.i.think	1.8
going.to.give	1.8
of.our.business	1.8
at.a.time	1.8
last.year.in	1.8
number.of.thin	

Token	ϕ
gs	1.8
we.go.into	1.8
i.think.that	1.8
going.to.come	1.8
to.think.that	1.7
and.i.would	1.7
we.saw.that	1.7
of.the.next	1.7
really.do.not	1.7
number.of.years	1.7
expect.it.to	1.7
i.just.want	1.7
probably.a.lit	
tle	1.7
that.were.look	
ing	1.7
well.i.do	1.7
want.to.do	1.7
so.well.have	1.7
would.say.that	1.7
for.us.and	1.7
our.customer.b	
ase	1.7
we.have.tried	1.7
say.is.that	1.7
i.would.say	1.7
little.bit.more	1.7
i.would.descri	
be	1.7
kinds.of.things	1.7
to.make.that	1.7
the.business.we	1.7
are.a.lot	1.7
come.up.with	1.7
of.the.quarter	1.7
the.first.quar	
ter	1.7
in.the.uk	1.7
have.tried.to	1.7
you.go.back	1.7
the.growth.that	1.7
fourth.quarter	
.but	1.7
course.of.the	1.7
second.quarter	
.and	1.6
the.last.year	1.6
area.where.we	1.6
that.continues	
.to	1.6
we.have.mentio	
ned	1.6
do.not.get	1.6
year.i.think	1.6
initiatives.th	
at.we	1.6
think.you.will	1.6
think.that.we	1.6
we.saw.some	1.6
think.there.is	1.6
us.and.we	1.6
to.our.custome	
rs	1.6

Token	ϕ
want.to.be	1.6
could.be.a	1.6
well.we.have	1.6
because.i.think	1.6
we.are.still	1.6
i.said.we	1.6
that.i.would	1.6
not.a.lot	1.6
the.prior.year	1.6
want.to.get	1.6
everything.tha	
t.we	1.6
we.didnt.have	1.5
the.quarter.but	1.5
to.get.into	1.5
yes.i.think	1.5
going.on.there	1.5
the.course.of	1.5
on.the.first	1.5
was.a.little	1.5
we.tried.to	1.5
time.we.have	1.5
going.to.conti	
nue	1.5
the.other.part	1.5
but.we.think	1.5
in.the.last	1.5
right.now.and	1.5
have.done.a	1.5
are.a.little	1.5
so.a.lot	1.5
a.little.less	1.5
think.we.will	1.5
have.been.talk	
ing	1.5
that.would.be	1.5
thing.i.would	1.5
not.think.that	1.5
i.think.youll	1.5
there.as.well	1.5
quarter.we.have	1.5
in.that.regard	1.5
i.do.believe	1.5
last.year.we	1.5
what.i.would	1.5
to.say.that	1.5
in.the.back	1.5
the.things.that	1.4
quarter.and.th	
en	1.4
i.think.is	1.4
yes.i.mean	1.4
making.sure.th	
at	1.4
it.tends.to	1.4
in.the.year	1.4
of.the.business	1.4
of.the.products	1.4
for.this.year	1.4
i.think.theres	1.4
trying.to.make	1.4
be.the.case	1.4
i.think.youre	1.4

Token	ϕ
and.the.growth	1.4
all.the.time	1.4
said.that.the	1.4
in.the.third	1.4
the.second.qua	
rter	1.4
expect.us.to	1.4
and.we.want	1.3
we.have.got	1.3
i.think.with	1.3
fourth.quarter	
.i	1.3
look.at.this	1.3
business.in.the	1.3
things.i.think	1.3
we.feel.we	1.3
that.we.saw	1.3
let.me.just	1.3
we.saw.in	1.3
i.think.all	1.3
need.to.make	1.3
a.lot.to	1.3
know.we.do	1.3
comment.on.the	1.3
we.are.continu	
ing	1.3
i.think.as	1.3
we.needed.to	1.3
have.i.think	1.3
consistent.wit	
h.what	1.3
the.part.of	1.3
the.next.couple	1.3
in.the.first	1.3
it.would.be	1.3
so.were.going	1.3
we.are.actually	1.3
the.one.thing	1.3
i.think.on	1.3
be.able.to	1.3
i.mean.its	1.3
been.talking.a	
bout	1.2
things.that.we	
re	1.2
would.tell.you	1.2
terms.of.where	1.2
the.institutio	
nal.side	1.2
are.continuing	
.to	1.2
going.through.	
the	1.2
as.i.said	1.2
first.quarter.	
of	1.2
think.what.we	1.2
i.would.tell	1.2
would.think.ab	
out	1.2
coming.out.of	1.2
do.not.really	1.2
have.kind.of	1.2

Token	ϕ
that.is.really	1.2
had.a.lot	1.2
to.start.to	1.2
over.the.next	1.2
really.hard.to	1.2
not.really.have	1.1
to.see.the	1.1
side.i.think	1.1
is.a.little	1.1
end.of.the	1.1
to.come.back	1.1
what.we.saw	1.1
four.or.five	1.1
look.at.it	1.1
for.our.clients	1.1
going.to.be	1.1
the.ones.that	1.1
i.think.you	1.1
were.not.going	1.1
well.i.think	1.1
up.a.little	1.1
think.you.have	1.1
a.little.bit	1.1
we.have.done	1.1
to.think.about	1.1
way.to.think	1.1
of.the.things	1.0
think.as.we	1.0
we.have.already	1.0
the.future.but	1.0
things.like.th	
at	1.0
and.you.know	1.0
is.hard.to	1.0
do.not.like	1.0
to.see.some	1.0
trying.to.get	1.0
over.the.course	1.0
i.would.think	1.0
if.you.want	1.0
we.were.going	1.0
have.a.lot	1.0
we.would.like	1.0
are.going.to	0.9
on.the.institu	
tional	0.9
in.the.us	0.9
very.focused.on	0.9
very.hard.to	0.9
think.thats.a	0.9
we.feel.very	0.9
a.year.ago	0.9
but.let.me	0.9
quarter.so.i	0.9
that.were.going	0.9
go.out.and	0.9
we.have.never	0.8
for.us.to	0.8
at.the.end	0.8
have.got.a	0.8
or.may.not	0.8
a.position.to	0.7
we.really.do	0.7

Token	ϕ
of.those.things	0.7
think.if.you	0.7
couple.of.years	0.7
i.know.that	0.7
may.or.may	0.7
just.give.you	0.7
so.i.think	0.7
going.to.go	0.6
the.short.answ	
er	0.6
we.talked.about	0.6
we.sit.here	0.6
so.we.havent	0.6
this.point.i	0.6
part.of.the	0.6
stuff.like.that	0.6
they.are.going	0.6
we.havent.real	
ly	0.6
rates.are.going	0.6
as.we.sit	0.6
of.factors.that	0.6
figure.out.what	0.6
sit.here.today	0.6
business.so.i	0.6
figure.out.how	0.5
you.a.sense	0.5
depends.on.what	0.5
what.that.means	0.5
that.level.of	0.5
that.we.think	0.5
on.an.annual	0.5
are.not.going	0.5
an.annual.basis	0.5
for.us.i	0.5
going.to.try	0.5
is.going.to	0.5
its.not.someth	
ing	0.5
well.see.how	0.5
and.how.much	0.5
with.the.fed	0.5
we.can.provide	0.5
whether.or.not	0.5
how.much.of	0.5
long.as.we	0.5
have.come.down	0.5
you.can.get	0.5
for.the.future	0.5
in.interest.ra	
tes	0.5
a.few.things	0.5
can.look.at	0.5
the.net.intere	
st	0.5
i.think.the	0.5
what.is.going	0.4
talk.about.the	0.4
in.response.to	0.4
have.given.you	0.4
end.up.with	0.4
like.that.so	0.4
that.you.could	0.4

Token	ϕ
in.a.position	0.4
in.the.pipeline	0.4
the.next.months	0.4
that.will.come	0.4
do.not.believe	0.4
theyre.going.to	0.4
over.time.so	0.4
a.quarterly.ba	
sis	0.4
we.put.out	0.4
its.fair.to	0.4
may.not.be	0.4
a.great.deal	0.4
money.market.f	
unds	0.4
to.figure.out	0.4
step.back.and	0.4
in.a.good	0.4
to.put.a	0.4
but.we.would	0.4
thats.not.a	0.4
we.know.that	0.4
given.that.we	0.4
to.continue.to	0.4
give.you.some	0.4
of.the.industry	0.4
take.into.acco	
unt	0.4
the.federal.re	
serve	0.4
i.would.guess	0.4
at.the.moment	0.4
thats.a.very	0.4
any.sort.of	0.4
at.the.last	0.4
something.in.t	
he	0.4
tell.you.that	0.4
the.mortgage.b	
usiness	0.4
a.level.of	0.4
some.kind.of	0.4
much.of.that	0.4
those.two.thin	
gs	0.3
and.i.believe	0.3
seem.to.be	0.3
going.forward.	
but	0.3
on.a.quarterly	0.3
to.go.in	0.3
having.said.th	
at	0.3
come.in.and	0.3
to.the.market	0.3
it.this.way	0.3
with.regard.to	0.3
net.interest.m	
argin	0.3
at.that.level	0.3
that.people.are	0.3
the.extent.we	0.3
this.point.but	0.3

Token	ϕ
you.get.into	0.3
in.the.mortgage	0.3
that.theres.a	0.3
not.believe.th	
at	0.3
the.timing.of	0.3
that.we.know	0.3
be.i.think	0.3
the.ccar.proce	
ss	0.3
at.some.point	0.3
a.lot.of	0.3
in.the.investm	
ent	0.3
i.would.put	0.3
likely.to.be	0.3
at.this.point	0.3
you.can.look	0.3
so.thats.one	0.3
you.know.that	0.3
would.think.th	
at	0.3
through.the.end	0.3
mentioned.that	
.we	0.3
out.i.think	0.3
and.i.guess	0.3
that.go.into	0.3
the.equity.side	0.3
in.our.numbers	0.3
for.a.number	0.3
going.to.look	0.3
year.but.we	0.3
one.is.the	0.3
would.be.very	0.3
you.know.it	0.3
point.in.time	0.3
on.the.revenue	0.3
to.see.in	0.3
think.about.th	
is	0.3
in.this.case	0.3
this.point.in	0.2
ahead.of.the	0.2
also.have.a	0.2
top.of.that	0.2
lot.of.differe	
nt	0.2
the.longer.term	0.2
in.that.range	0.2
i.think.over	0.2
get.to.a	0.2
but.in.terms	0.2
good.question.i	0.2
on.top.of	0.2
do.not.need	0.2
go.into.the	0.2
when.you.see	0.2
of.the.range	0.2
expectation.is	
.that	0.2
each.and.every	0.2
and.over.time	0.2

Token	ϕ
what.happens.w	
ith	0.2
the.same.kind	0.2
a.bunch.of	0.2
we.believe.the	0.2
one.of.those	0.2
the.percentage	
.of	0.2
i.mentioned.in	0.2
it.depends.on	0.2
thats.what.i	0.2
have.to.take	0.2
trying.to.figu	
re	0.2
we.have.given	0.2
we.said.in	0.2
of.the.way	0.2
the.change.in	0.2
the.things.i	0.2
not.trying.to	0.2
back.into.the	0.2
see.that.we	0.2
like.that.and	0.2
our.view.is	0.2
outlook.for.the	0.2
in.my.remarks	0.2
saw.this.quart	
er	0.2
going.to.change	0.2
same.kind.of	0.2
to.come.down	0.2
is.not.going	0.2
we.would.do	0.2
have.looked.at	0.2
quarter.of.the	0.2
the.industry.a	
nd	0.2
of.that.busine	
ss	0.2
the.size.of	0.2
mentioned.in.my	0.2
it.relates.to	0.2
would.have.to	0.2
that.we.still	0.2
this.point.we	0.2
any.kind.of	0.2
thing.that.we	0.2
i.think.at	0.2
the.rate.of	0.2
this.is.one	0.2
do.not.necessa	
rily	0.2
we.tend.to	0.2
the.revenue.si	
de	0.1
is.to.get	0.1
i.believe.that	0.1
seems.to.be	0.1
are.talking.ab	
out	0.1
years.we.have	0.1
an.environment	
.where	0.1

Token	ϕ
in.the.company	0.1
as.you.know	0.1
terms.of.what	0.1
in.the.future	0.1
and.so.forth	0.1
have.a.pretty	0.1
see.what.the	0.1
is.one.of	0.1
we.believe.that	0.1
that.kind.of	0.1
i.look.at	0.1
i.mentioned.th	
at	0.1
to.add.to	0.1
in.my.prepared	0.1
thats.going.to	0.1
know.we.have	0.1
by.the.way	0.1
that.we.want	0.1
when.it.comes	0.1
change.in.the	0.1
parts.of.the	0.1
you.know.we	0.1
and.the.reason	0.1
i.guess.the	0.1
as.it.relates	0.1
talked.about.t	
hat	0.1
regard.to.the	0.1
other.thing.th	
at	0.1
is.going.on	0.1
how.we.think	0.1
to.try.to	0.1
to.try.and	0.1
at.the.time	0.1
as.much.as	0.1
and.it.really	0.1

A.2 Correlation between sentiment measures

Table A1: Correlation between sentiment measures

	1.	**2.**	**3.**	**4.**	**5.**	**6.**
1. $NonAnswer$						
2. $NonAnswer^{\phi}$	0.918					
3. $Negativity$	-0.131	-0.071				
4. $Uncertainty$	0.040	0.071	0.200			
5. $Numbers$	-0.095	-0.055	0.200	0.104		
6. $Complexity$	-0.088	-0.086	0.133	-0.050	0.021	

Notes: This table presents Pearson correlations between the sentiment variables used in our analysis. $NonAnswer$ ($NonAnswer^{\phi}$) is the ratio of trigrams in our non-answer glossary (weighted by loadings) to the total words. $Negativity$ and $Uncertainty$ are the ratio of negative and uncertain words to the total words. List of negative and uncertain words are from Loughran and McDonald (2011) word-lists. $Numbers$ is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). $Complexity$ is the complexity measure derived from the Loughran and McDonald (2019) complexity word list.

A.3 Uncertain language versus non-answers: examples

A.3.1 Uncertainty without non-answers

Q: "Before i hit on the second question, again the variable piece on that, is that a 4% to 5% number? or how should we think about that?"

A: "That's probably within the range. maybe 3% to 5%, just depending on the quarter."

Q: "Okay. And is that the plan still to sort of bring the quarterly capacity up to about 345 million with the expansion in hillview and cammis?"

A: "Yes, depending how the demand goes, yes, we could do that or more or less depending what we needed."

Q: "First off, what percentage of business is through (inaudible) for oem (ph)?"

A: "Our u.s. business, maybe about 60% through distribution; our international, probably maybe half, distribution."

A.3.2 Non-answers without uncertainty

Q: "Then in terms of your promotional spending in the marketplace, did it tweak up just slightly sequentially from the second quarter level?"

A: "I actually don't know offhand. i don't think so. i don't know that, we can get back to you on that."

Q: "What's your sense on timing? how do you expect it to play out what are the negative – (multiple speakers)."

A: "Too early to tell. i think it's too early to tell. i wouldn't take a guess at it at this point."

Q: "Regionally, do you see one area being more rational than the other on those, rick?"

A: "Yes, but i really don't want to comment on this call about that."

A.3.3 Mixture of non-answers and uncertainty

Q: "John, on the 9 billion units that you mentioned, how much of that is actually cans?"

A: "Good question. the majority of it. i don't have the numbers off the top of my head. i would say approximately 75

Q: "And i – if i – i think the logical inference from the way you're describing a little bit of a change in business model there is that we'll see, what? greater s easonality, if you will, in – or maybe greater volatility, a wider range of margins through the year in international that gives you the ability to do more when you're doing well, but you'll have some extra staffing costs in softer quarters?"

A: "Well, it could be. we'll have to wait and see, yes."

A.4 Cross-sectional Fama-MacBeth regression

Table A2: Fama-MacBeth Regressions

	$FF3 - CAR_{0;1}$				$FF5 - CAR_{0;1}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$NonAnswer$	-0.122**	-0.098***			-0.110**	-0.091***		
	(-2.35)	(-4.19)			(-2.40)	(-4.09)		
$NonAnswer^{\phi}$			-0.053***	-0.106*			-0.048***	-0.102*
			(-3.86)	(-1.87)			(-3.94)	(-1.83)
$Negativity$		-0.470***		-0.467***		-0.428***		-0.427***
		(-6.08)		(-6.01)		(-6.13)		(-6.05)
$Uncertainty$		-0.318		-0.014		-0.324		-0.028
		(-0.95)		(-0.19)		(-1.00)		(-0.39)
$Numbers$		-0.095		-0.096		-0.083		-0.084
		(-1.23)		(-1.18)		(-1.16)		(-1.12)
$Complexity$		0.203		0.414**		0.169		0.375**
		(1.39)		(2.62)		(1.23)		(2.46)
Constant	0.011**	0.015*	0.010***	0.012**	0.010**	0.014*	0.009***	0.011**
	(2.07)	(1.87)	(2.97)	(2.18)	(2.10)	(1.91)	(3.01)	(2.25)
Observations	21191	21191	21191	21191	21191	21191	21191	21191
R^2	0.007	0.097	0.007	0.098	0.006	0.096	0.007	0.097
FirmControls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Fama-MacBeth cross-sectional regressions for Equation (5). The dependent variable is the abnormal returns over the Fama-French three (1993) and five (2015) factor model returns cumulated from the day of the earnings call to the day after it, $FF3 - CAR_{0;1}$ ($FF5 - CAR_{0;1}$). $NonAnswer$ ($NonAnswer^{\phi}$) is the ratio of trigrams in our non-answer glossary (weighted by loadings) to the total words. $Negativity$ and $Uncertainty$ are the ratio of negative and uncertain words to the total words. List of negative and uncertain words are from Loughran and McDonald (2011) word-lists. $Numbers$ is the share of numbers to the sum of total words and numbers calculated as in Zhou (2018). $Complexity$ is the complexity measure derived from the Loughran and McDonald (2019) complexity word list. Firm controls include $EarnSurp$, BTM, $ln(Assets)$ and Tobin's Q. t-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

A.5 Distribution of Non-Answers and Non-Answers over time

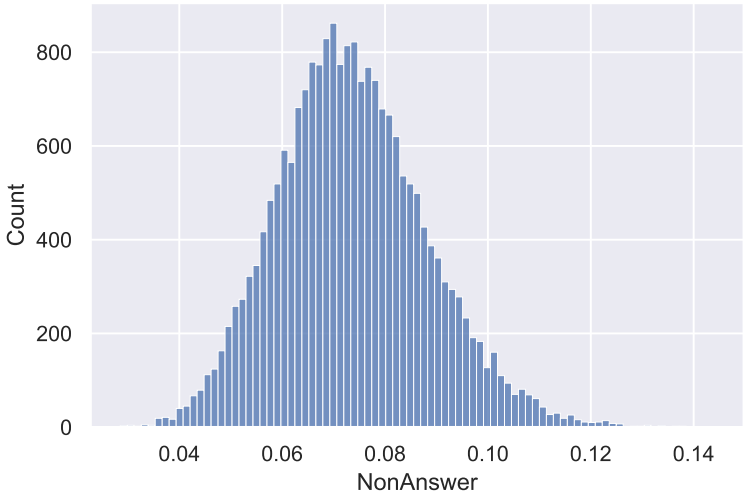


Figure A1: Distribution of *NonAnswer* for the validation set of S&P 500 earnings calls.

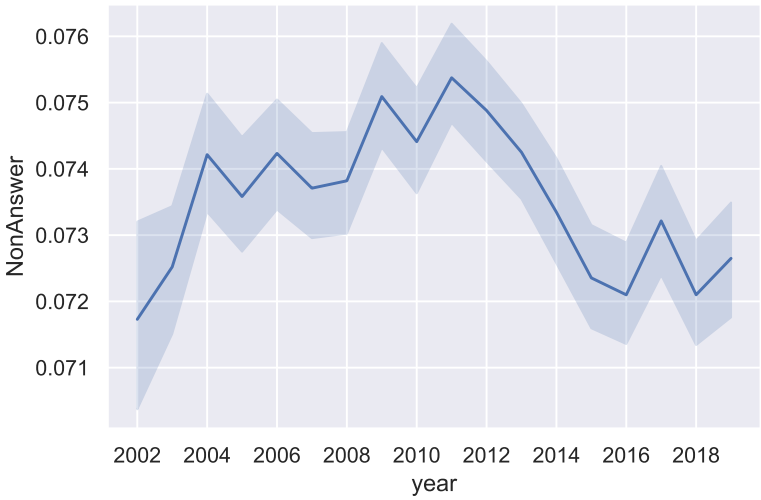


Figure A2: Average *NonAnswer* over time for the validation set of S&P 500 earnings calls.

A.6 Alternative Q&A settings

A non-answer does not require a specific context. As both our method and glossary are free of financial context, we believe that the measure is applicable to other fields with a question and answers setup. In order to corroborate this claim, consider Mark Zuckerberg's responses to the US Senate during the Cambridge Analytica hearing as anecdotal evidence. His response "Senator, [...] I can certainly have my team get back to you on any specifics there that I don't know, sitting here today." is clearly a non-answer, and would have been identified as such by the glossary method.

In this Appendix, we briefly explore textual data of presidential interviews as another structured Q&A setting. Starting in 1864 with an interview with Abraham Lincoln, we analyze the answers of roughly 900 presidential interviews, which were collected by UCSB's American Presidency Project, see www.presidency.ucsb.edu.

The presidential data shows substantial variation in $NonAnswer$ (c.f. Figure A3), a necessary condition for the measure to be informative. Particularly high non-answer scores are found in interviews with President Clinton at the time when sexual assault allegations surfaced that later became the basis for an impeachment charge of perjury.

As an example for a high $NonAnswer$ presidential response, consider telephone interview with Morton Kondracke, in which Bill Clinton responded to the question: *"Okay. Let me just ask you one more question about this. You said in a statement today that you had no improper relationship with this intern. What exactly was the nature of your relationship with her?"* with the words *"Well, let me say, the relationship's not improper, and I think that's important enough to say. But because the investigation is going on and because I don't know what is out—what's going to be asked of me, I think I need to cooperate, answer the questions, but I think it's important for me to make it clear what is not. And then, at the appropriate time, I'll try to answer what is. But let me answer, it is not an improper relationship, and I know what the word means."*.

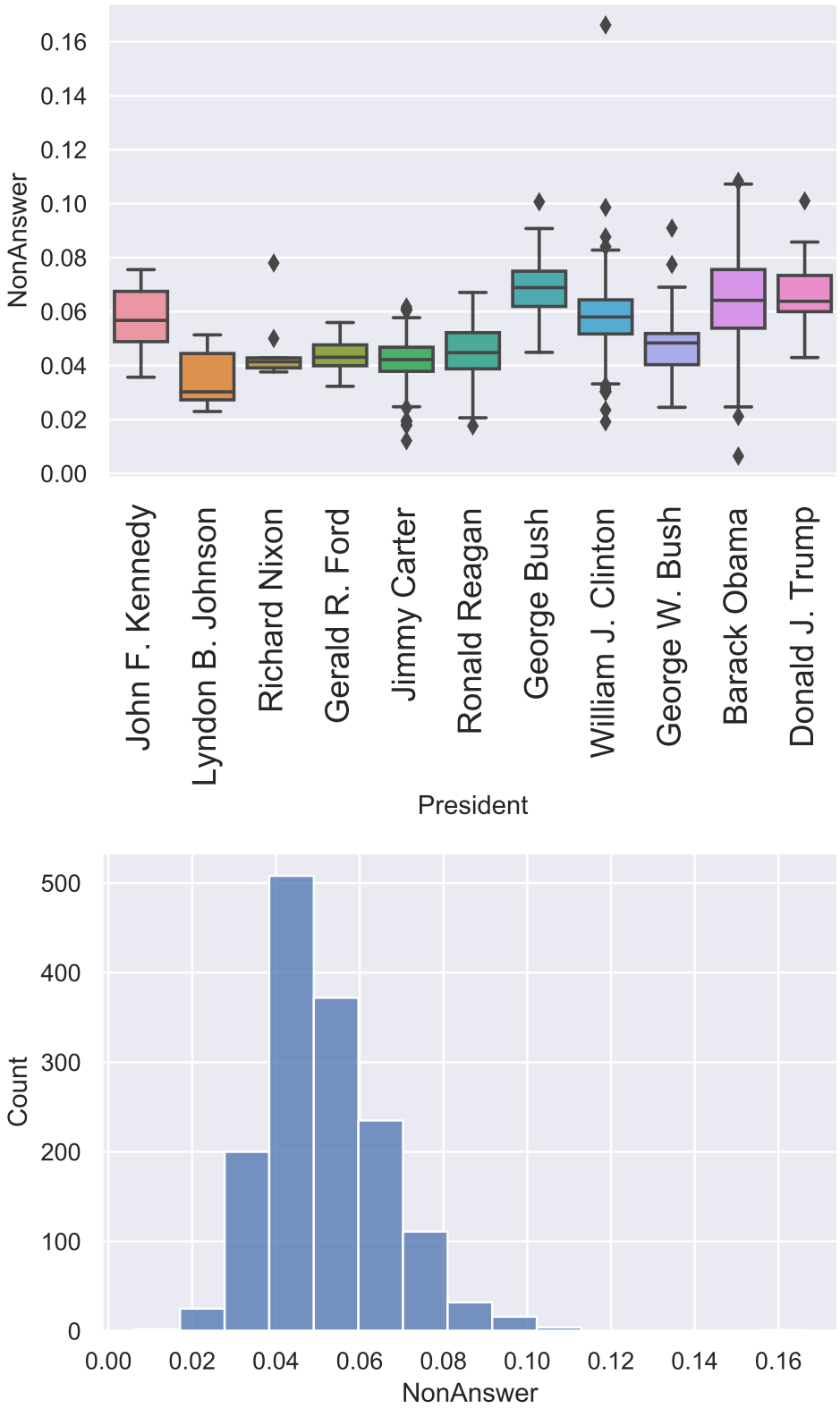


Figure A3: Presidents

Chapter III

Does firm's silence drive media's attention away?

Does firm's silence drive media's attention away?*

Sasan Mansouri†

May 30, 2021

ABSTRACT

In this study, using a comprehensive dataset on business media coverage and textual analysis of the discussions in firms' quarterly earnings conference calls, we show that when management fails to satisfy the demand for information, ceteris paribus, their firms receive less media coverage. Poor information environment hurts the information-creation capacity of the media, while such an environment does not show a similar association with the media's information-dissemination role. Furthermore, this association is more prominent for professional business media, compared to their non-professional counterparts such as blogs and alternative articles. Our results add nuance to the literature on media coverage bias by showing that supply-side factors, i.e. the factors affecting the suppliers of the coverage, mainly drive the coverage of firms, not the demand.

Keywords: non-answers; conference calls; media coverage; non-professional business media.

JEL-Classification: D82, G14, G30.

*This manuscript benefited from feedback by Andreas Barth, Martin Brown, Roland Füss, Alexander Hillert, Gerard Hoberg, Stephan Hollander, Beatriz Garcia Osma, Steven Ongena, and Mark Wahrenburg. I declare that I have no relevant or material financial interests that relate to the research described in this paper. All errors are my own.

†Goethe University Frankfurt ✉: mansouri@finance.uni-frankfurt.de

1 Introduction

Among different information intermediaries, media platforms enjoy the broadest audience (Zingales, 2000). This enables media to play a crucial role in the financial markets in several ways. Media coverage can cause stock prices to move (Fang and Peress, 2009; Hillert et al., 2014), force decisions in corporate governance (Dyck et al., 2008) and firm behavior (Baloria and Heese, 2018), and act as a whistle-blower of corporate wrongdoings (Dyck et al., 2010; Miller, 2006). Despite the abundance of literature on the effects of media attention, we know less about the driving forces behind media coverage.

In this paper, we examine the media coverage of big corporations and test two competing hypotheses regarding media attention towards firms. We use the variations in the quality of information environment around the firms, and ask if the media coverage is associated with the richness of this information environment. Specifically, media attention is ''demand-driven'' if business media respond to the stakeholders' demand for more information/analysis about firms with less robust information environments. For firms with less robust information environments, it is more challenging for media sources to gather enough publishable materials. In other words, a firm's media exposure is ''supply-driven'', if the media sources, as suppliers of information, reduce their coverage of firms that are more difficult to cover.

To measure the quality of the information environment surrounding a firm, we rely on the literature on the informativeness of their quarterly earnings conference calls and advancements in computational linguistics. The Q&A sessions of earnings conference calls offer a unique setting for our study. In these calls, investors can glean information about the company by questioning senior managers directly. When responding to a question, the management decides whether to fulfill this need for information or leave the demand for information "non-answered". We quantify the level of non-answers in a call using the non-answer score proposed by Barth et al. (2020). This metric is trained on a set of Q&As between equity analysts and management during earnings calls, and is calculated using a bag-of-words approach with a glossary of 1,364 trigrams like '[let me get] back to you', '[I] do not know', '[it's] hard to predict', '[let's] wait and see', '[it's] too early to', etc..

These trigrams are frequently used to refrain from factually answering a question – either a direct rejection, i.e. refusal to answer a question, as documented by Gow et al. (2019), or a less noticeable symptom of non-answers, i.e. beating around the bush by blathering, as outlined in Barth et al. (2019).

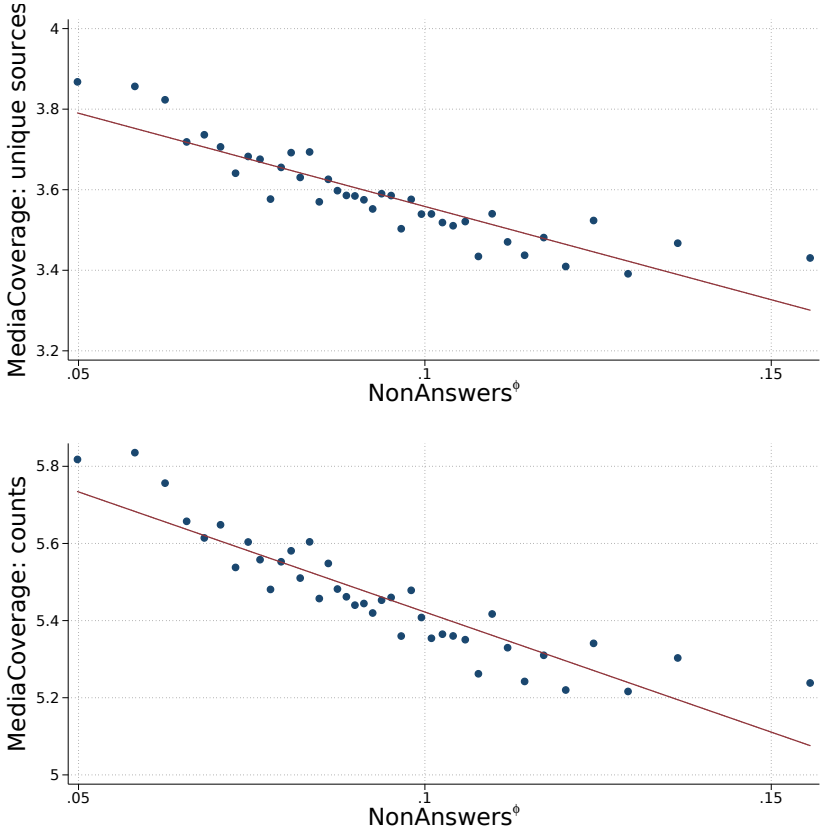
We follow a survivorship bias-free sample of firms appearing in the S&P 500 index for the period 2007 to 2019, and analyze the transcripts from their quarterly earnings conference calls to measure the non-answer score in the management responses. Furthermore, we collect the media coverage information before and after each earnings call using Ravenpack (RP). RP includes millisecond-timestamped media coverage data from tens of thousands of news sources like Dow Jones Newswires, the Wall Street Journal, the Financial Times, Bloomberg, Reuters, Seeking Alpha, as well as blogs like Zero Hedge and The Motley Fool. Our analysis window spans from the day after the earnings call until 60 days later to capture the period between the current call and a company's next call.[1] We measure the media coverage by considering the number of unique news sources that publish content during our analysis window, i.e. *Sources*, and the number of pieces published about the firm during the same window, *Counts*.

Our analysis first examines the association of non-answers in the calls with the media coverage that the firm receives in the next quarter. Figure 1 illustrates binned scatter plots of our media coverage variables versus the non-answers in earnings calls. We scrutinize the negative correlation shown in Figure 1 using a regression analysis framework and find that, in line with the supply-driven media coverage hypothesis, the more a firm turns down the demand for information in its earnings call, the less media attention it receives in the coming quarter. In other words, while the demand for firm-specific information increases due to non-answers, media also supply less content due to supply-side difficulties of acquiring and providing content.

Next, we investigate if the supply-driven decrease in coverage is more substantial for content that is more difficult to create. Ravenpack categorizes the coverage as either a full

[1] While it seems natural to consider a 90-day window between two calls, in 25% of the cases, companies hosted two calls less than 90 days apart from one another. Less than 1% of companies held two calls in less than 60 days. The window was also fixed at 60 days to allow for the media coverage measures to be comparable across firms.

Figure 1: Media coverage and non-answers in earnings calls
This figure displays binned scatter plots for the main analysis of this paper.



article, a (hot-)newsflash, or tabular material. We differentiate between different types of coverage based on their production cost for the media sources, i.e. full articles versus other types, and repeat our analysis. The results confirm the supply-driven coverage hypothesis for full articles only. Contrarily, the firm's initiated press releases mainly explain the variation in the non-full articles' coverage. Overall, this suggests that the poor information environment around the firm only curtails the information-production role of the media and does not hurt their information-dissemination role.

Business media exist on a spectrum of professionalism, and both professional and non-professional media play significant roles in shaping investors' opinions (Chen et al., 2014; Drake et al., 2017).[2] Our analysis continues by asking how non-answer earnings calls associate with the extent of (non-)professional media coverage. First, in line with the previous results, we find there are fewer media sources of both categories that publish content for the firms with less robust information environment. The professional media,

[2]In this analysis, we instead divide the media sources into two categories as defined in Subsection 3.2.

however, publish significantly fewer articles, which results in a lower ratio of professional coverage for the non-answering firms.

All of the above-mentioned results are robust to controlling for several confounding factors. First, we control for common factors that previous literature find to drive media attention, such as size, profitability and book-to-market ratio. Second, we control for the common language measures of the management answers that contain value-relevant information for the stock market, e.g., tone, uncertainty and complexity. Furthermore, we include the difference between the market (analyst) expectations and the actual quarterly results, i.e. earnings surprise. Third, we use a topic-modeling algorithm to develop and control for 25 news topics to address the topic-specific tendencies in attracting more media coverage. Finally, we absorb several observable and unobservable factors by incorporating several fixed effects; we remove common time trends with quarter fixed effects and control for any time-constant firm-specific factors by firm fixed effects. In some specifications, instead of firm fixed effects, we either include industry-year fixed effects or, more conservatively, firm-year fixed effects, which allow us to study the same firm within a year.

Finally, we verify these findings from the perspective of the media sources' coverage portfolio, and specifically ask if the non-answer earnings calls shift media attention to a firm's peers. We restrict our sample to earnings calls of firms in the same industry that hold their conference calls on the same day, and measure the share of the articles belonging to each of these firms in the media one month before and after the date of the earnings call. We find that media sources shift their coverage from the non-answering firms to the firms with more informative earnings calls. This result is robust to the inclusion of media fixed effects.

Our paper contributes to the existing literature in number of ways. First, we add to the literature on the factors that skew media attention. Previous literature has uncovered several factors such as the advertisement (Reuter and Zitzewitz, 2006), local proximity (Gurun and Butler, 2012), and firm size and reputation (Miller, 2006). We add to this literature by showing that media coverage is inclined toward firms with a better

information environment. Second, the literature of accounting and finance separates the role of media that a) disseminates the currently available information, and b) creates new information through active journalism (Bushee et al., 2010). We add to this literature by showing that the"supply-driven" decrease of media coverage is associated only to the information-creation role of the media. Third, Ravenpack News Analytics provides wonderfully detailed data on media coverage (Miller and Skinner, 2015), and our topic-modeling approach enables researchers to enhance the practicality of the news taxonomy data provided in Ravenpack. Finally, we further demonstrate the importance of non-professional business media for the financial markets (Chen et al., 2014; Drake et al., 2017), and show that they are less susceptible to reducing coverage in poor information environments as compared to their professional counterparts.

The remainder of this paper is organized as follows: section 2 reviews the relevant literature and outlines the testable hypotheses. Section 3, provides a detailed description of the dataset for our empirical analyses. We describe our empirical analyses and show the results in section 4. Section 5 concludes the paper.

2 Background literature and hypotheses

2.1 Information content of earnings calls

Firms voluntarily hold earnings conference calls on a regular basis to fill the information gap among their equity investors (Brown et al., 2004). Compared to firm's other types of disclosures, earnings calls are indeed very informative for market participants (Bushee et al., 2004; Matsumoto et al., 2011), as they contain more forward-looking details about the firm's expected performance and direction (Kimbrough and Louis, 2011). Earnings calls usually begin with the management presenting the previous quarter's earnings results followed by a Q&A session between the management and participants who are mainly equity analysts. While both the presentation and the Q&A session are enlightening for the market, Matsumoto et al. (2011) show that the latter is more informative.

With advances in computational linguistics and development of finance-specific glos-

saries like Loughran and McDonald (2011), a growing body of empirical literature deals with the information content of management responses to the question asked during the earnings calls. Price et al. (2012) show that investors react to the soft information during the call, i.e. the tone of management's answers. While the market digests the hard information (e.g. earnings surprise) in the one-day window of the call, they further show that tone predicts the stock price drift up to 60 days post-call. Furthermore, Dzieliński et al. (2017) and Zhou (2018) show, respectively, that uncertain language from management, as well as the ratio of numeric contents in the management responses, contain value-relevant information.

Although earnings calls are a medium to provide investors with value-relevant information, management can obscure the flow of information in several ways. Most severely, Mayew (2008) and Cohen et al. (2013) provide evidence showing that management discriminates against questions raised by the analysts whose stock recommendations are considered unfavorable. Moreover, management can avoid answering unfavorable questions by ''obfuscating'' through complex language (Bushee et al., 2018), ''blathering'', i.e. beating around the bush (Barth et al., 2019), directly refusing/rejecting (Hollander et al., 2010; Gow et al., 2019), or a mix of all of these techniques (Barth et al., 2020).

Investors react to the discussions in the earnings calls; yet, they also rely on information intermediaries to digest the content of these calls. Sell-side analysts are one of the most studied information intermediaries. Frankel et al. (2006) show that the information content of the analysts' report complements the firms' disclosures. Huang et al. (2018) show that when the management withholds value-relevant information during the call, equity analysts intensify their ''discovery'' role (as opposed to solely ''interpreting'' discussions). In this study, we shed light on the other types of information intermediaries, namely the business media, and verify how they respond to the management withholding information.

2.2 Media as an information intermediary

Media causally affects firms' security prices, corporate governance and investors' attention.[3] The media coverage takes two main roles, namely disseminating/packaging the available information and stale news or creating new information through active journalism practices. A growing body of accounting literature deals with the disentangling of these two roles. Bushee et al. (2010) eliminate the journalists' interpretations from media coverage and show that the further dissemination of available information leads to lower information asymmetry among investors. Drake et al. (2014) confirm the role of media dissemination in incorporating accounting information into stock prices. Blankespoor et al. (2018) similarly show that disseminating information, identified by the introduction of robo-journalism, increases the trading volume and liquidity. Controlling for the available information, Engelberg and Parsons (2011) show that local media coverage of S&P500 companies strongly predicts local trading.

In addition to dissemination, media coverage can influence investors' behavior by creating new content (Dougal et al., 2012; Guest, 2018). Above all, the information-creation role makes the media a watchdog for accounting fraud (Miller, 2006)[4]. More generally, media is one of the most diligent whistle-blowers for corporate fraud (Dyck et al., 2010). Investors also value media's monitoring role. Specifically, Gao et al. (2020) show that the closure of local newspapers results in an increase in municipal borrowing costs of 5 to 11 basis points. We contribute to this literature by investigating whether the media's information-creation ability declines when firm's management withholds value-relevant information.

Media coverage is prone to several biases.[5] Media may engage in ''sensationalism'' by disproportionately covering stories that may be interesting to a broad audience. For example, media tend to cover the CEOs with more option exercises, disproportionately

[3]See Tetlock (2015); Miller and Skinner (2015); Blankespoor et al. (2020) for a comprehensive literature review.

[4]Compared to other information intermediaries, media benefit from a broader audience that enables it to play a governance role by shaping investors' beliefs (Zingales, 2000) by only disseminating the available information (Rogers et al., 2016).

[5]There are many empirical papers showing the existence of media bias in political coverage. See Puglisi and Snyder Jr (2015) for a comprehensive literature review. Here, we only discuss the case of media bias regarding coverage of corporations.

more negatively, and ignore the total salary (Core et al., 2008). Moreover, the media's watchdog role is mostly limited to the cases where the fraudulent activities are related to a famous/large corporation that could be interesting to a wide audience (Miller, 2006). There are also shreds of evidence concerning other sources of bias, e.g., advertisement pressure (Reuter and Zitzewitz, 2006), reciprocity between journalists and corporations (Dyck and Zingales, 2003; Westphal and Deephouse, 2011), and favoritism toward socially-responsible firms (Zavyalova et al., 2012; Cahan et al., 2015).

Finally, we contribute to the discussion on professional versus non-professional business media. Drake et al. (2017) argue that professionalism exists on a spectrum. Their study classifies the sample of media into three groups: professional, semi-professional and non-professional. They show that the coverage by the first two groups has positive capital market effects, while the coverage by the latter contains more noise than real information. Drake et al. (2017) classify Seeking Alpha (SA) as a semi-professional media outlet. SA is a platform where non-professional analysts share their stock recommendations. Chen et al. (2014) show that the articles as well as the commentaries on SA predict future stock returns and earnings surprises. In this study, we compare the coverage behavior of professional and non-professional business media when they face firms with less robust information environment.

2.3 Research question

This section examines two main ''demand-driven'' versus ''supply-driven'' media coverage hypotheses[6].

A demand-driven media coverage hypothesis postulates that several stakeholders in a firm demand information and a media source addresses this demand in its coverage of a firm. Stakeholders in a company rely on the discussions in earnings calls to expand their ''understanding'' of the company (Barker et al., 2012). The management's refusal to provide the requested information in an earnings call keeps the demand for value-relevant information unfulfilled. Therefore, the stakeholders rely on other intermediaries

[6]See Puglisi and Snyder Jr (2011) for the literature review on different supply- versus demand-side factors contributing to the bias of political newspapers.

to acquire the missing information that they demand. Investors are especially receptive to the missing value-relevant information from earnings call discussions (Huang et al., 2018). Since investors' demand for more information and analyses is one of the most important drivers behind equity analysts' decision what to cover (Brown et al., 2015), as an important information intermediary, greater media coverage is expected of firms for which the management rejects investors' demand for information.

Hypothesis ("Demand-driven coverage"). *Firms with more non-answers in their earnings calls receive c.p. coverage from* ***more*** *media sources.*

A supply-driven media coverage hypothesis, on the other hand, suggests that media sources cover firms based on their own preferences rather than the level of demand by their readers/subscribers. Firms deliberately reject the demand for more information because of the proprietary costs associated with providing such information (Gow et al., 2019). These costs affect the firms' disclosure preferences and put them in a poorer information environment (Ellis et al., 2012). Such an environment makes it more difficult for media sources and journalists to acquire enough information to publish news articles about a firm (Guest and Kim, 2020). According to the theoretical model of Bhushan (1989), a poorer information environment shifts the supply curve of media to the left, resulting in less total supply of coverage[7].

Hypothesis ("Supply-driven coverage"). *Firms with more non-answers in their earnings calls receive c.p. coverage from* ***fewer*** *media sources.*

3 Data

For our study, we construct a survivorship bias-free S&P 500 dataset, which resolves any concerns over unobservable factors that may factor into the media coverage of smaller firms. First, the media (Miller, 2006; Hillert et al., 2014) and analysts (Martineau and Zoican, 2020) put S&P500 firms in the spotlight. Second, these firms are all high-volume publicly

[7]Lang and Lundholm (1996) show empirically that more equity analysts tend to follow the firms with more clear disclosure policies.

traded companies, for which investors have a strong demand for information. Finally, the substantial cost of wrongdoing discourages management from using non-answers to conceal potential fraudulent activities.

The following subsections define the main variables used in this paper. Subsection 3.4 provides the descriptive statistics of the main variables in our empirical analyses. Table A1 summarizes the variables used in this study and their corresponding definitions.

3.1 Measurement of management's withholding information

This study uses the "non-answer" score proposed by Barth et al. (2020). to quantify the management's withholding of information. This metric is based on the two symptoms, "rejecting" (Gow et al., 2019) and "blathering" (Barth et al., 2019), by employing a Multinomial Inverse Regression (MNIR) technique (Taddy, 2013). Barth et al. (2020) identify a glossary of 1,364 trigrams such as "back to you", "do not know", "hard to predict", etc., which are frequently used in English Q&As[8] to refrain from answering a question concisely and factually. Figure 2 shows the word cloud of the most important trigrams of this glossary.

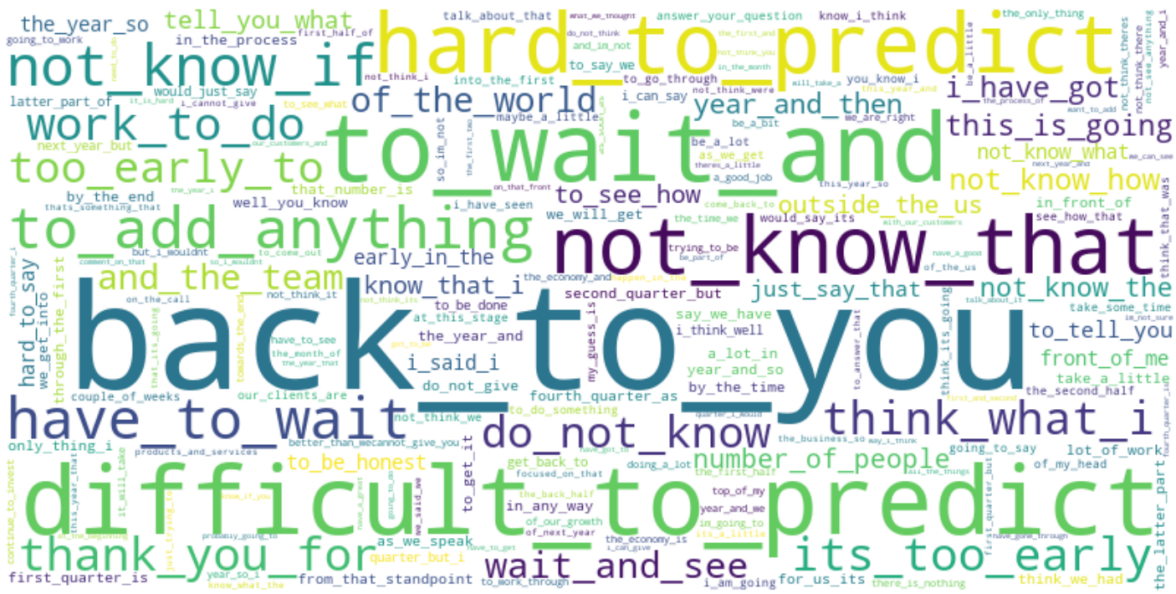
We collect transcripts of every earnings call held by S&P 500 companies from Thomson Reuters' StreetEvents for 2007 to 2019. [9] These calls are released quarterly and usually take place on the same day as the corresponding earnings release. Calls mostly start with the management presenting a (prepared) statement, and then analysts (and investors) are invited to a Q&A session. The spontaneous nature of the Q&A session is a unique laboratory to measure the degree to which the management avoids providing factual responses to analysts' questions. All earnings calls without a Q&A session are excluded from the dataset, since this area is the focus of our study.

We define $NonAnswer^{\phi}$ as the weighted count of the terms in the glossary provided by Barth et al. (2020) for every answer in the earnings call of company i in quarter t:

[8]Barth et al. (2020) show that their glossary measures non-answers in the financial context as well as the political context (e.g. US presidential interviews and US Senate hearings) and sport press conferences.

[9]Data is available starting in 2002, but we cut the data points before 2007 because of the missing data in our media coverage dataset. We use the full data for some of the robustness checks in Appendix D.

Figure 2: Non-answer glossary
The glossary consists of 1,364 trigrams provided by Barth et al. (2020). Larger font size indicates a higher factor load. A machine-readable version of the glossary is available at econlinguistics.org



$$NonAnswer_{it}^{\phi} = \frac{\sum_{k=1}^{K} \phi_k \times Non\text{-}Answer\ Glossary\ Token_{it}^{k}}{Total\ Words_{it}},$$

where ϕ_k is the loading associated with trigram $k \in \{1, 2, ..., K\}$ in the glossary of Barth et al. (2020). We limit our observations to the transcripts where the total management's answers are at least 500 words long. This helps to avoid the extreme values of the non-answer measure due to a small denominator.

3.2 Media coverage

We use RavenPack News Analytics (RPNA) to obtain the media coverage of the firms in our sample. The RavenPack dataset consists of three main editions; Dow Jones Edition provides the historical coverage, since 2000, of Dow Jones Newswires, regional editions of the Wall Street Journal, Barron's and MarketWatch. The Web Edition, starting from January 2007, includes hundreds of thousands of articles a day from leading publishers and web aggregators with more than 22,000 sources (Hafez and Xie, 2014). Finally, the PR Edition that includes the PR-Newswire. News coverage of RPNA is timestamped to the millisecond and includes several sentiment metrics as well as the news taxonomy.

These features make RP an interesting dataset for asset managers, investment bankers, and hedge funds(RavenPack, 2017).

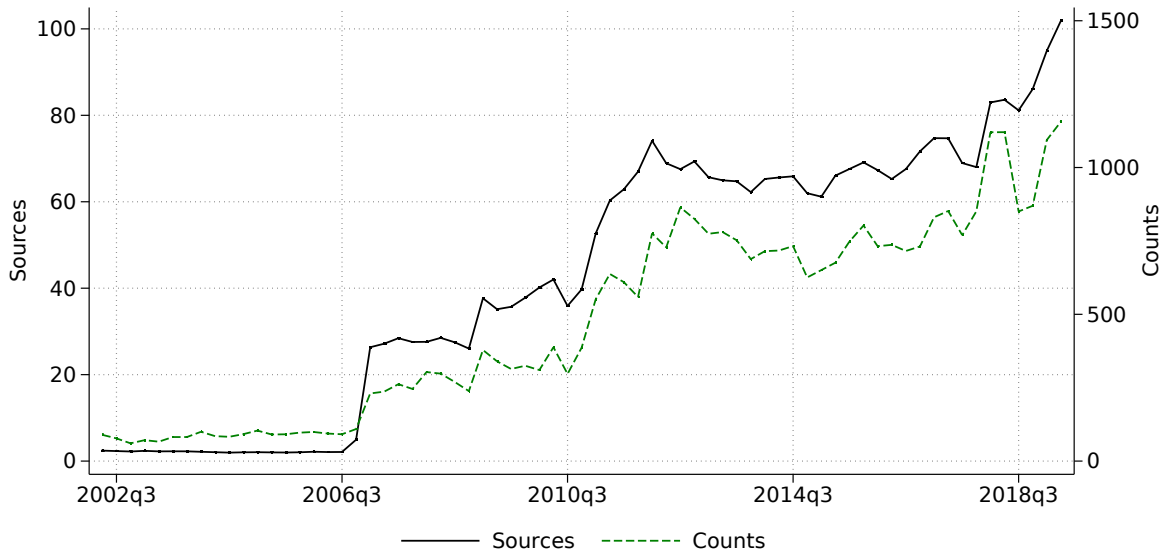
We merge our sample of firms with RP using the 8-digit CUSIP code. For the unmatched sample, we perform a fuzzy match of the company names in Compustat with the company name in RP and manually check if the matching score is less than 95%. We further filter for all the news contents that have a "Relevance" score of at least 90% to the firm.

Coverage *Sources* is the main coverage measure used throughout this paper. It is defined as the natural logarithm of one plus the count of unique media sources (or channels) that publish content – a full article, (hot-)newsflash, and/or tabular material – in a two-month period after a firms' earnings call. Similarly, $Sources^F$ if we filter only for the full articles, and $Sources^{NF}$ if we only consider the (hot-)news-flashes and tabular-materials. Additionally, $Counts$ is defined as the natural logarithm of one plus the count of all the above-mentioned types of content. We exclude the day of the conference call as well as the day after, to avoid overloading our metrics with lots of news regarding the company's quarterly earnings results.

Figure 3 shows the time trends for the average distinct Sources and news Counts for all the firms in a given quarter. The spike of both scores at the beginning of 2007 coincides with the introduction of RavenPack Web Edition, which has continuously added the coverage of many news sources since. Although we can address this problem with quarter fixed effects, we trim the sample and focus only on the observations from 2007, as our analyses require a consistent sample. For the robustness check of the main results, we repeat our analysis with the full sample as well.

Press releases Furthermore, we control for the number of press releases in our analysis window, i.e. PR. To do so according to Bushee and Miller (2007); Core et al. (2008); Bushee et al. (2010) and assume all the articles on the press release wire as well as the entries with ‘‘NEWS_TYPE" as ‘‘PRESS-RELEASE" are firm-initiated disclosures.

Figure 3: Time-trend of news coverage
This figure shows the number of media sources that published at least one piece with more than 90% relevance to the firms in our sample within two months after the firms' quarterly earnings call (left axis) as well as the total count of publications (right axis). From January 2007, RavenPack includes the WEB Edition (RP-WEB) which covers articles from leading (online) publishers & web aggregators. RavenPack is gradually adding news sources to its universe; e.g., the RP-WEB only includes content for "Reuters" after 2011. In all of the analyses in this paper, we include the quarter fixed effects to address the time-trends of news coverage.



Professional versus non-professional business media Inspired by Drake et al. (2017), we divide the sample of sources into either professional or non-professional business media. We consider a media professional if it or its parent company is listed among Barron's (with RavenPack ID: 18A55F), Bloomberg News (208421), Business Insider (C75B8C), CNBC (AA1167), Dow Jones Newswires (B5569E), Entrepreneur (938822), the Financial Times (FD0B00), Forbes (22AC8B), MarketWatch (1E5E35), Morningstar (E04BE4), Reuters (751371) or the Wall Street Journal (AA6E89). By this definition, our list of professional business media includes sources like "FT Alphaville - Hedge funds" (parent company the Financial Times) and "Bloomberg Businessweek" (Bloomberg), for example. This list includes 65 sources as listed in Appendix C. Non-professional media included in this study are mostly blogs and news websites associated with the non-professional analysts/journalists, e.g., Seeking Alpha (B61D8F), Zero Hedge (5E506B) and The Motley Fool (C81722).

Similar to our coverage variables, $Sources_{Pro}(Sources_{N-Pro})$ and $Count_{Pro}(Count_{N-Pro})$

refer to the coverage filter the media sources according to (non-)professional.

Newsworthiness Inspired by Dyck et al. (2008), we define $NewsWorthiness$ as the natural logarithm of one plus the number of articles referring to a company in the Wall Street Journal (RavenPack ID "AA6E89") and the Financial Times (RavenPack ID "FD0B00") during the 6-month period before the earnings call (excluding the day of and immediately before the earnings call).

News contents To control for the news content relevant to the firms in our analysis window, we collect news taxonomy information, i.e. a ‘‘Category’’ field, for all the entries of the firms in our analysis window from one day to 60 days after the call. In total, it sums up to 526 distinct categories. To efficiently control for the news content, we reduce these 526 categories to 25 ‘‘news topics’’ using the LDA topic modeling algorithm of Blei et al. (2003)[10].

In a conventional LDA for language modeling, documents must be tokenized by breaking down the sentences into smaller linguistic units like a word, a bi-gram, or other higher-order n-grams, and forming a ‘‘document-term-matrix’’ that tabulates the counts of the tokens in several documents of the training set. Here we consider each news category in the list of all the categories that appeared in the news following a firm's earnings call as a token in a document. LDA clusters the tokens that co-occur frequently, and forms a topic. One feature that makes the LDA popular is that the topics are formed independent of the researchers' prejudgment. Researchers must only input the number of topics. We select 25 topics as it suggests the best goodness of fit according to the metrics developed in computer science literature (Nikita and Chaney, 2016)[11].

Figure 4 shows the most important news categories associated with each of the news topics. β is the weight of each token in the given topics. For example, Topic 10 (or 19)

[10]LDA stands for Latent Dirichlet Allocation and is an unsupervised machine learning algorithm. The topics resulting from a fitted LDA model are very similar to the factors in a factor analysis model. LDA is a popular topic modeling algorithm for textual analysis in finance and accounting. See Gentzkow et al. (2019) for an overview of topic modeling methodology and Eickhoff and Neuss (2017) for the literature review.

[11]Figure B1 in Appendix B shows the metrics for the goodness of the LDA fit for the different number of topics in the range of 5 to 95.

consists of the news most relevant to acquisitions (or layoffs). The distribution of weights in topic 7, on the other hand, is more homogeneous among the different news related to the technical analysis of the shares.

In the next stage, we measure the distribution of each of 25 topics in the list of all news after an earnings call. γ_{ijt} is the portion of the news that belongs to the topic j after firm i's earnings call in quarter t. We have $\sum_j \gamma_{ijt} = 1$ and each topic has a log-normal distribution among documents. Figure 5 shows the histograms of $log(\gamma)$ for each topic.

3.3 Other variables

Alternative speech characteristics Finance and accounting literature offers several standard metrics to quantify earnings calls' language content. Following the literature, we adopt a dictionary (bag-of-words) approach, in which one calculates the desired sentiment by counting the occurrence of words in a corresponding word list divided by the total words in the document.

We calculate the *Negativity* (as a measure of tone) and *Uncertainty* of management answers using the "negative" and "uncertainty" word lists offered by Loughran and McDonald (2011). We do not consider the "positive" word list for the tone calculations, as suggested by Loughran and Mcdonald (2016).

Additionally, we control for the *Complexity* in the management language using the word-list of Loughran and McDonald (2020). Loughran and McDonald (2020) show that the complexity measure of 10-k filings[12], is associated with the stock returns around the filing date and unexpected earnings.

Earnings surprise We collect analysts' Earnings per Share (EPS) forecasts for the firms in our sample from Institutional Brokers' Estimate System (IBES) database and define *EarnSurp* following Dzieliński et al. (2017). More precisely, we calculate Earnings Surprises as the difference between the actual and consensus forecast earnings, divided

[12]Here, we deviate from Loughran and McDonald (2020) in the sense that they measure complexity by counting the number of unique words, in a given document, that appear in their word list of 374 words. We use a dictionary approach of counting the frequency of the words in their word list divided by the total words.

by the closing share price on the 5th trading day before the earnings announcement in every quarter. We then group the (zero and) positive as well as the negative numbers separately into five quantiles (i.e. quintiles). Finally, we sort these ten categories and label the earnings surprises from 1 (most negative) to 5 (least negative) and from 6 (least positive) to 10 (most positive).

Firm characteristics We use Compustat to obtain quarterly balance sheet data. We control for book-to-market (BTM) ratio, and the firm size as the natural logarithm of total assets($ln(Assets)$). We calculate Tobin's Q as the book value of assets minus book value of common equity plus the market value of common equity, divided by the total book value of assets.

3.4 Descriptive statistics

Figure 4: News categories
The figure shows the 25 topics that resulted from the topic modeling analysis of the 200 most common news categories in the analysis window of the two months after the earnings calls in our sample. β shows the weights of each news category in the topics. For each topic, the two most important news categories are shown here.

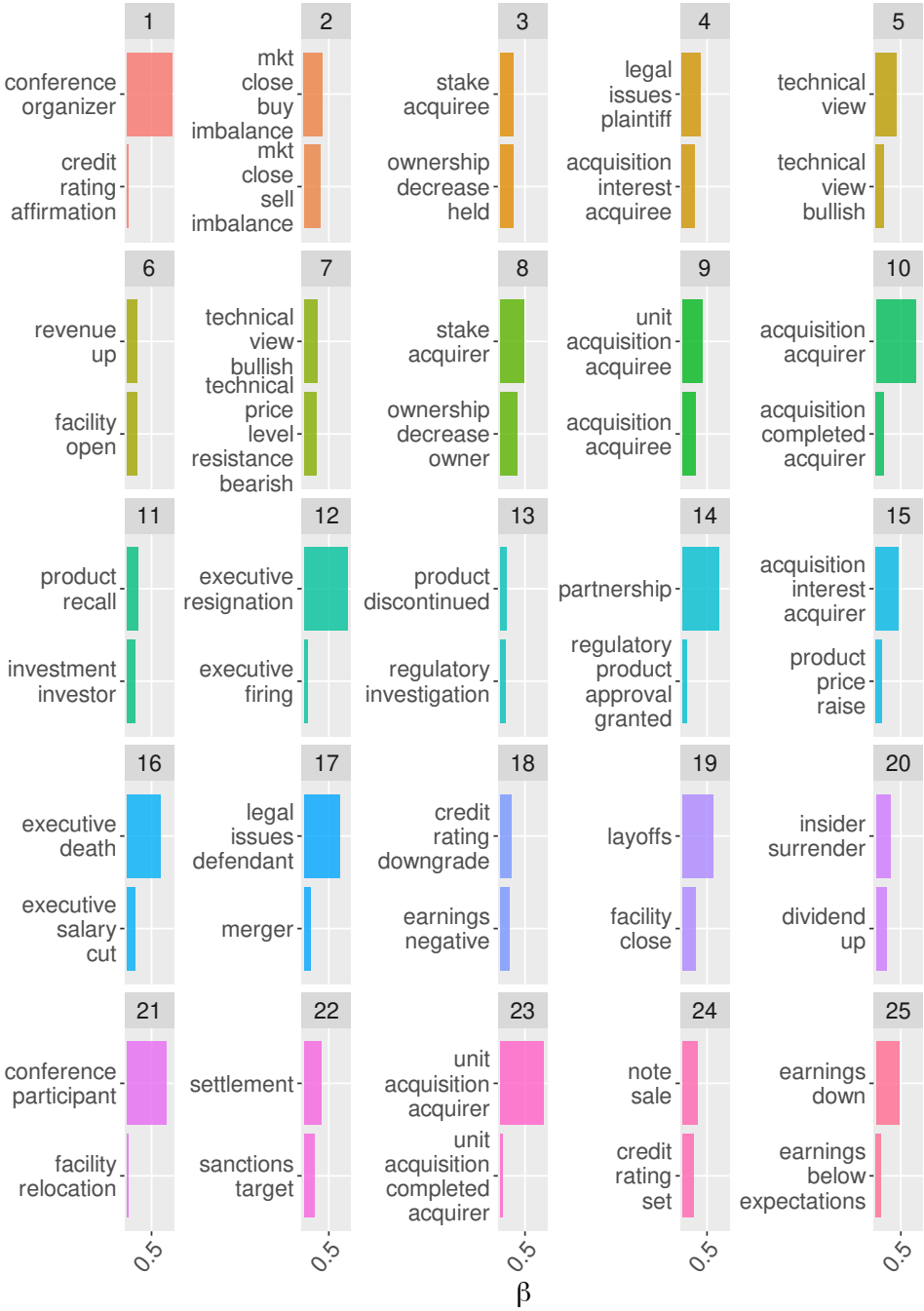


Figure 5: Topic distribution of news categories
The figure shows the distributions of the topic probabilities for each of the 25 topics that resulted from the topic modeling analysis of the 200 most common news categories in the analysis window of two months after the earnings calls in our sample. γ shows the probability that a document belongs to a topic.

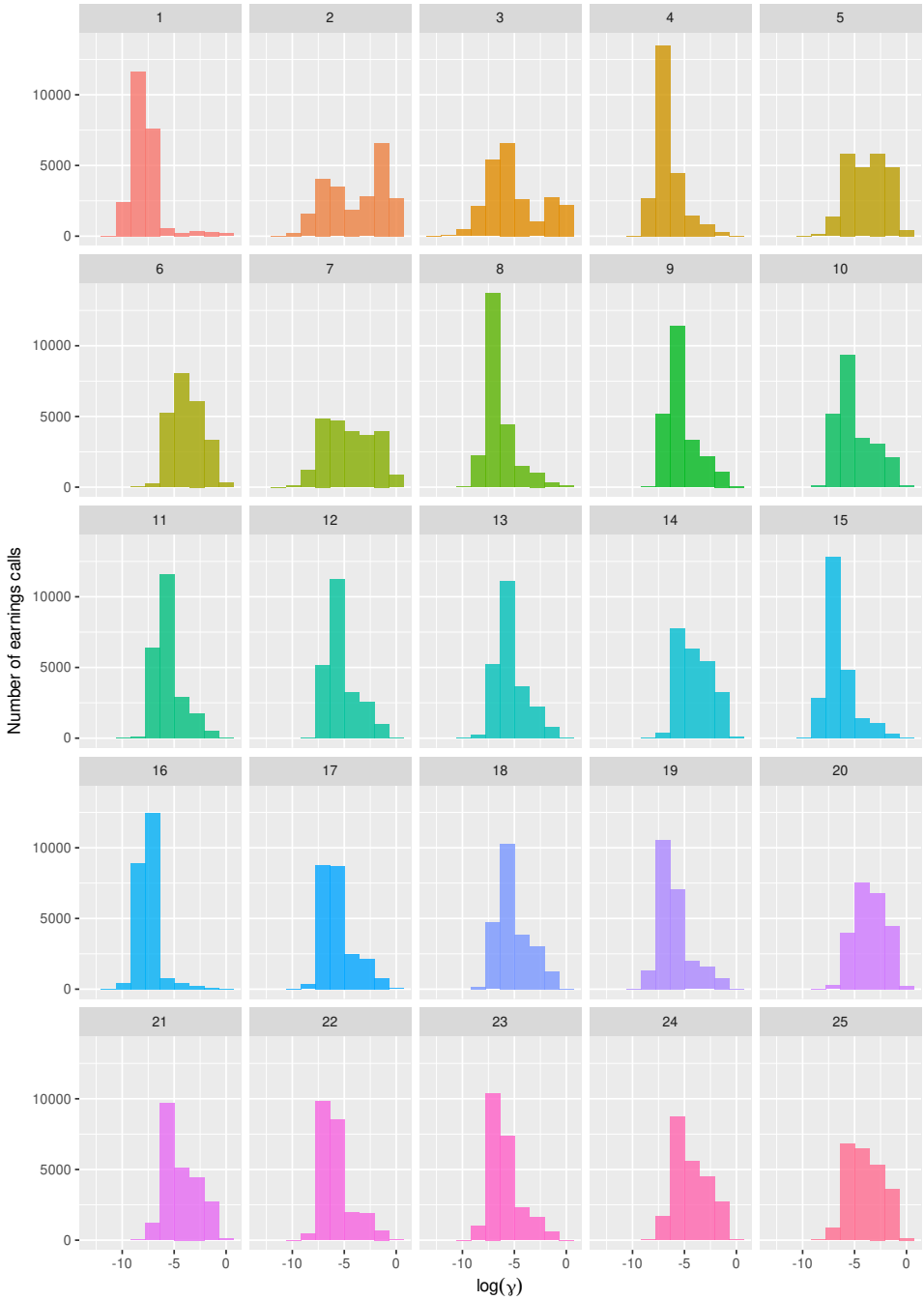


Table 1: Descriptive statistics
Descriptive statistics of the variables used in the analyses. The sample consists of 18,275 earnings calls of the companies in the S&P 500 index between 2007 and 2019. All variables are defined in Table A1.

Variable	Obs.	Mean	Std. Dev.	Min	P10	P50	P90	Max
$NonAnswer^{\phi}$	18,275	.094	.022	.032	.067	.092	.12	.22
$Sources$	18,275	3.6	.9	.69	2.6	3.4	4.8	7.4
$Sources^{F}$	18,275	3.5	.91	.69	2.6	3.4	4.8	7.3
$Sources^{NF}$	18,275	2.1	.75	.69	1.1	1.9	3	6.2
$Sources^{F}_{Pro}$	17,661	2	.45	.69	1.4	1.9	2.6	3.3
$Sources^{F}_{N-Pro}$	17,661	3.3	1	.69	2.1	3.2	4.7	7.3
$Counts$	18,275	5.5	1.1	1.6	4.2	5.3	6.9	11
$Counts^{F}_{Pro}$	17,661	3.9	1.1	.69	2.6	3.8	5.4	8.9
$Counts^{F}_{N-Pro}$	17,661	4.8	1.5	.69	2.8	4.9	6.5	11
$Share - Pro$	17,661	.33	.22	.0029	.087	.27	.67	.99
PR	15,910	1.3	.37	.69	.69	1.4	1.8	2.6
$NewsWorthiness$	18,275	.97	1.3	0	0	.69	2.8	7.9
$Negativity$	18,275	.028	.0071	.0075	.02	.027	.037	.079
$Uncertainty$	18,275	.016	.0056	0	.0089	.015	.023	.056
$EarnSurp$	18,275	5.7	2.9	1	2	6	10	10
BTM	18,275	.44	.41	-3.2	.099	.35	.88	17
$ln(Assets)$	18,275	9.8	1.4	6.2	8.2	9.6	12	15
Q	18,275	2.1	1.4	.63	1	1.7	3.8	36
$Complexity$	18,275	.0071	.0039	0	.0028	.0065	.012	.029

Table 1 presents descriptive statistics for the variables in this analysis. For $NonAnswer^{\phi}$, the magnitude is in line with the distribution presented in Barth et al. (2020). The median firm in our sample receives coverage in 200 published pieces from around 30 media sources, and the firm itself initiates around 4 press releases during our analysis window of two months after the earnings call.

4 Empirical analysis

Section 4.1 analyzes the association of our media coverage measure with the management withholding of information in earnings calls. We further explore this relationship by separating several types of media coverage in section 4.2. Section 4.3 compares the coverage choice of professional versus non-professional business media. Finally, section 4.4 investigates the above-mentioned association from the perspective of media preferences.

4.1 Media coverage

We first investigate if non-answers by management during earnings calls are associated with the media coverage that their firm receives during the following quarter. We model the count of media sources that cover firm i in a window of two months after the call[13] in quarter t, as indicated in the following equation:

$$\begin{aligned} MediaCoverage_{it} = \beta_0 + \beta_1 \cdot NonAnswer_{it}^{\phi} + \beta_2 \cdot PR_{it} \\ + \beta_3 \cdot X_{it} + \alpha_{\mathbb{I}(i)} + \alpha_t + \epsilon_{it}, \end{aligned} \tag{1}$$

$MediaCoverage_{it}$ interchangeably refers to the coverage variables ($Sources$ & $Counts$) as defined in Subsection 3.2. $NonAnswer_{it}^{\phi}$ is the main variable of interest, measuring management's degree of non-answers in the call. According to the ''demand-driven coverage'' hypothesis, $NonAnswer_{it}^{\phi}$ is a proxy for the demand for a firm's information after the earnings call, and therefore, a positive β_1 means that the media answers this demand by increasing their coverage of the firms that withhold more information. On the contrary, the ''supply-driven coverage'' predicts a negative β_1 reflecting less media coverage of the firms for which the cost of acquiring information is higher.

We control for several important variables that contribute to firms' media coverage. PR_{it} is the amount of firm-disseminated news and contributes directly to the firm's coverage because of the dissemination role of the media. We also control for several other firm-quarter observations in X_{it}; Earnings Surprise ($EarningsSurp$) captures the difference between analysts' expectations about earnings and the realized earnings in quarter t. The natural logarithm of total assets ($ln(Assets)$), book-to-market ratio (BTM), and Tobin's Q, as a measure of profitability, are all standard variables to control for determinants of a firm's media coverage (Miller, 2006; Bushee et al., 2010). Furthermore, we control for the information content of the calls via several standard measures available in the literature.

In particular we control for pessimism in the management's response tone (Neg) as

[13]Results for a short window analysis, i.e. the first 48 hours after the call, are consistent with this analysis, and are provided in the appendix Table D1.

Price et al. (2012) show that tone is a significant predictor of abnormal returns and trading volume in the initial reaction window, and also dominates earnings surprises over the 60 trading days following the conference call. Dzieliński et al. (2017) show that investors punish the *Uncertainty* of the management's answers and tone with a lower valuation. We further control for *Complexity* in the answers, as Loughran and McDonald (2020) suggest this measure complements a corporation's size. Complexity also controls for the required informativeness of a firm's disclosure (Guay et al., 2016). Inspired by Dyck et al. (2008), X_{it} includes $NewsWorthiness$ to control for the particular newsworthy timings around companies. Finally, X_{it} also includes the prevalence of each of the 25 news topics during the analysis window to address the possibility of certain news categories (e.g. legal issues or M&A) driving more media coverage.

Additionally, we include various fixed effects to capture several unobservable characteristics of the firms and quarters that contribute to both the media coverage and non-answers in earnings calls. $\alpha_{\mathbb{I}}(\alpha_i)$ denotes industry (firm) fixed effects. In some specifications, instead of firm fixed effects, we control for firm-year dummy variables to absorb time-varying heterogeneity at the firm level. α_t absorbs time trends of coverage. To allow for a potential serial correlation of media coverage within each firm and within each quarter, we employ a two-way clustering of standard errors (Cameron et al., 2011) to the firm and quarter dimensions.

Table 2 summarizes the results of this analysis. In all of the specifications, $NonAnswer^{\phi}$ has a negative coefficient for the media coverage, supporting the "supply-driven coverage" hypothesis. In particular, column (2) shows that an increase of one standard deviation in our $NonAnswer^{\phi}$ score is associated with approximately 3% less media coverage. The results hold in the within-firm-year variation, i.e. comparing the same firm with different $NonAnswer^{\phi}$ in different quarterly earnings calls in the same fiscal year. A positive and statistically significant coefficient for the number of PR releases also confirms the role of media in disseminating firms' press releases. $NewsWorthiness$ positively contributes to media coverage.

4.2 Coverage type

Media sources/channels spend different amounts of time, energy and resources to publish different types of content about a firm. Media disseminate the currently available information mostly via "newsflashes" which include a headline or a link to other sources' coverage. On the contrary, publishing a full article requires the media source to put in more effort by providing editorial content (Drake et al., 2014). In line with the supply-driven coverage hypothesis, we postulate that the firms with higher non-answer scores receive less media coverage in full articles, as this is more costly the poorer the information environment around the firm is. Moreover, we expect to find no significant correlation between the management withholding information and the publication of content other than full articles.

Our dataset can separate the coverage types into full articles and other types of content according to the tags provided by the RavenPack. We then repeat our analysis of the previous section (Equation 1) except that we differentiate between these two types of coverage.

Table 3 shows that the decrease in the coverage only occurs in the case of the full articles. The regression coefficient for the non-answer score is negative and statistically significant in the first three columns. Column (1) shows that an increase of one standard deviation in the within-firm non-answer metric is associated with 0.9% fewer media sources that publish at least one full article about a firm in the two-month window after the earnings call. On the other hand, for other types of coverage, shown in columns (4) to (6), the coefficient of the non-answer variable is not significantly different from zero. Comparing columns (4) and (5), the count of PR releases includes most of the heterogeneity regarding the non-full article coverage.

4.3 Professional and non-professional business media

This section explores whether the source's professionalism influences its reduction in coverage it provides. Non-professional business media play a significant information intermediary role alongside their professional counterparts. Investors react to the tenor of

the articles posted by non-professional analysts and journalists (Chen et al., 2014; Drake et al., 2017).

In light of the analyses in section 4.1, we first verify if professional and non-professional business media reduce their coverage equally, i.e., whether a higher non-answer score is related to a firm receiving less media coverage, regardless of the media type. We identify the professional business media according to the definition in subsection 3.2 and repeat the analysis using equation 1.

Table 4 shows the result of this analysis with several specifications. The negative coefficient of the $NonAnswer^{\phi}$ variable shows that both professional and non-professional business media reduce their coverage.

Next, we ask which of these two media types are more sensitive to the firms' disclosure style. Bloggers and non-professional journalists/analysts have different incentives to cover a company than professional media do. First, skin in the game, i.e. their open positions in the underlying stock, motivates them to follow a specific firm diligently (Campbell et al., 2019). Second, non-professional media are incentivized to signal their quality by covering firms with less available coverage. Third, non-professional media sources are limited in the sense of the alternative choices they have for coverage. Most bloggers/non-professional business journalists are specialized and interested in covering certain firms, whereas professional sources like Bloomberg could easily shift their coverage portfolio. To summarize, although the poorer information environment discourages media sources from covering a firm, non-professional media sources have the incentive to fill the void by publishing more articles about that firm. To clarify this point, we analyze if the number of articles is also negatively associated with the management withholding information from both the professional and non-professional media. We repeat the analysis using equation 1, where the dependent variable is the natural logarithm of one plus the count of news in professional and non-professional media, separately.

Table 5 shows the results of this analysis. Columns (1) & (2) reveal a negative correlation between the number of articles and the non-answer variable for professional coverage. On the other hand, a significant association is absent from columns (3) & (4).

This suggests that although less media sources are covering a company in the analysis window after a non-answer earnings call, the total count of non-professional coverage is not less for the within-firm and within firm-year non-answer.

Furthermore, in columns (5) & (6), the dependent variable is the share of the full articles written by the professional media to total full articles. The negative coefficient of the non-answer variable suggests that management withholding information is associated with a lower ratio of professional coverage.

4.4 Coverage shift within the industry

This subsection verifies the implications of the supply-driven coverage hypothesis for media level observations, and asks if media sources shift their coverage of a firm with a poor information environment to a peer firm that provides factual responses to questions during the earnings calls. Unlike the previous analyses, here we limit our focus to the media sources that already have a set of peers in their coverage portfolio, which allows us to compare the change in the coverage weights before and after an earnings call. To do so, we restrict our sample to earnings calls of the firms, within an industry, held on the same day.

First, we define our coverage shift score and then verify the association of coverage shift with management non-answers in a regression analysis.

Coverage shift For the firm i of industry $\mathbb{I}$ on the earnings call day t, we define the "prior" coverage weight of ω_{mit} for a source m as:

$$\omega_{mit} = \frac{\sum_d Counts_{mid}}{\sum_{i \in \mathbb{I}} \sum_d Counts_{mid}}, \quad d \in [t-60, ..., t-1] \tag{2}$$

where $Counts_{mid}$ counts the number of full articles, (hot-)newsflashes and tabular material published on the day d. By definition we have $\omega_{mit} \geq 0$ and $\sum_{i \in \mathbb{I}} \omega_{mit} = 1$. Similar to the prior weights in Equation 2, we define the "posterior" coverage weights,

ω'_{mit}, where $d \in [t+1, ..., t+60]$.

We define the coverage shift as the distance between the posterior and prior coverage weights. More specifically, the coverage shift after the earnings call of firm i on earnings call of date t is

$$\Delta\omega_{mit} = ln(\frac{1+\omega'_{mit}}{1+\omega_{mit}}). \tag{3}$$

Here we clarify this definition by an example. On 14 April 2015 (t), J.P. Morgan (JPM) and Wells Fargo & Co (WFC) held their earnings calls with $NonAnswer^{\phi}$ of 0.12 and 0.07, respectively. These firms both belong to industry 44 based on the 48 Fama-French industry classification ($\mathbb{I}$). Here, we verify the shifts in the coverage of WSJ (m) for these two firms. In the two months before t, WSJ published 23 articles for JPM and 2 for WFC. This translates to the prior weights of $\omega_{m,JPM,t} = 23/(1+(2+23)) = 0.885$ and $\omega_{m,WFC,t} = 2/(1+25) = 0.077$. In the two months after t, WSJ published 14 pieces in total for these two banks, 12 of which were for JPM and 2 were for WFC. So, the posterior coverage weights are $\omega'_{m,JPM,t} = 12/(1+14) = 0.8$ and $\omega'_{m,WFC,t} = 2/(1+14) = 0.133$. Finally, we can use Equation 3 to calculate the shift in coverage, which is $\Delta\omega_{m,JPM,t} = ln(1+0.8/1+0.885) = -0.046$ and $\Delta\omega_{m,WFC,t} = ln(1+0.133/1+0.077) = +0.051$

Analysis We model the shift in the firm's coverage weight within industries with the following linear equation:

$$\Delta\omega_{mit} = \beta_0 + \beta_1 \cdot NonAnswer_{it} + \beta_2 \cdot X_{it} + \alpha_m + \alpha_{\mathbb{I}t} + \epsilon_{mit} \tag{4}$$

where $NonAnswer_{it}$ is interchangeably the $NonAnswer^{\phi}$ or the non-answer rank of firm i among its industry peers on date t. Firm i achieves rank $1(N)$ if it has the highest (lowest) $NonAnswer^{\phi}$ among all the other N firms in the same industry that hold the earnings call on the date t.

X_{it} controls for firm-quarter level observations as in Equation 1. We include α_m, media fixed effects, to exploit within-media variations. We also control for $\alpha_{\mathbb{I}t}$ dummies to

absorb common characteristics of an industry's earnings calls on a given date[14]. Finally, we cluster the standard errors at the earnings call level.

Table 6 shows the results of this analysis. Results in columns (1) & (2) confirm that the top-ranked non-answer firms witness a statistically significant reduction in coverage weight in media level as compared to their peers. Columns (3) & (4) show the same results using the level of $NonAnswer^{\phi}$. In other words, ceteris paribus, media shifts its coverage to firms that obstruct the flow of information less. Similarly, the language complexity of an earnings call is negatively associated with the shift in the coverage weight of the firms. These results hold after controlling for the media fixed effects.

[14]For example, firms tend to shift announcements of bad news to the weekend (Damodaran, 2015) or macro news (Hirshleifer and Sheng, 2019) and policy/regulators announcements may alter the attention to the firm and industry level news.

Table 2: Management withholding information and media coverage

Notes: OLS regressions for Equation (1). In columns (1) to (5) the dependent variable is the natural logarithm of one plus the number of distinct news agencies that published at least one article about the firm from the day after the earnings call until 60 days later. Firm controls include EarningsSurprise, BTM, ln(Assets), and Tobin's Q. News categories refer to news about a company in the analysis window being associated with the 25 groups of news classified using the topic modeling algorithm described in Subsection 3.2. Industry classification is based on the 48 Fama-French industries. All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	Sources				
	(1)	(2)	(3)	(4)	(5)
$NonAnswer^{\phi}$	-4.630***	-1.411***	-0.369**	-0.498***	-0.501***
	(-3.73)	(-3.97)	(-2.20)	(-3.14)	(-3.19)
PR		0.544***	0.264***	0.208***	0.208***
		(18.70)	(18.14)	(14.21)	(14.05)
$Negativity$		-3.074***	0.689	0.491	0.510
		(-2.71)	(1.17)	(1.01)	(1.05)
$Uncertainty$		-5.780***	-0.473	-0.413	-0.444
		(-5.60)	(-0.78)	(-0.70)	(-0.76)
$Complexity$		0.804	1.156	0.959	0.942
		(0.38)	(1.20)	(0.97)	(0.96)
$NewsWorthiness$		0.295***	0.082***		
		(22.65)	(10.06)		
Observations	18275	15763	15903	15459	15459
R^2	0.013	0.778	0.895	0.944	0.944
Firm Controls	No	Yes	Yes	Yes	Yes
News Categories	No	No	No	No	Yes
QuarterYear FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Implied	Implied	Implied
Firm FE	No	No	Yes	Implied	Implied
FirmYear FE	No	No	No	Yes	Yes

Table 3: Management withholding information and different types of media coverage

Notes: OLS regressions for Equation (1). In columns (1) to (3) the dependent variable is the natural logarithm of one plus the number of distinct news agencies that published at least one full article about the firm from the day after the earnings call until 60 days later; columns (4) to (6) show the same metric for all the coverage types except for full articles. Firm controls include EarningsSurprise, BTM, ln(Assets) and Tobin's Q. News categories refer to news about a company in the analysis window being associated with the 25 groups of news classified using the topic modeling algorithm described in Subsection 3.2. Industry classification is based on the 48 Fama-French industries. All specifications include dummies for industry multiply by the date of earnings calls. All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	$Sources^F$			$Sources^{NF}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$NonAnswer^{\phi}$	-0.447***	-0.344**	-0.521***	-0.280	-0.264	-0.082
	(-2.74)	(-2.07)	(-3.15)	(-1.59)	(-1.47)	(-0.40)
PR		0.275***	0.221***		0.209***	0.170***
		(18.18)	(14.42)		(12.15)	(12.10)
$Negativity$	0.566	0.846	0.542	0.146	0.340	0.472
	(0.94)	(1.39)	(1.04)	(0.22)	(0.53)	(0.78)
$Uncertainty$	-0.251	-0.442	-0.449	-0.812	-1.003	-0.711
	(-0.40)	(-0.71)	(-0.77)	(-1.22)	(-1.45)	(-0.87)
$Complexity$	1.186	0.874	0.687	-0.378	-0.686	0.386
	(1.26)	(0.87)	(0.68)	(-0.33)	(-0.58)	(0.35)
$NewsWorthiness$	0.080***	0.079***		0.103***	0.102***	
	(9.05)	(9.41)		(10.40)	(9.97)	
Observations	18269	15903	15459	18269	15903	15459
R^2	0.882	0.892	0.942	0.769	0.784	0.870
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
News Categories	No	No	Yes	No	No	Yes
QuarterYear FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Implied	Yes	Yes	Implied
FirmYear FE	No	No	Yes	No	No	Yes

Table 4: Management withholding information and the professional vs. non-professional media coverage

Notes: OLS regressions for Equation (1). In columns (1) and (2) the dependent variable is the natural logarithm of one plus the number of distinct professional news agencies that published at least one full article about the firm from the day after the earnings call until 60 days later; columns (3) and (4) show the same metric for non-professional media coverage. Firm controls include EarningsSurprise, BTM, ln(Assets) and Tobin's Q. News categories refer to news about a company in the analysis window being associated with the 25 groups of news classified using the topic modeling algorithm described in Subsection 3.2. Industry classification is based on the Fama-French 48 industries. All specifications include dummies for industry multiply by the date of earnings calls. All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	$Sources^F_{Pro}$		$Sources^F_{N-Pro}$	
	(1)	(2)	(3)	(4)
$NonAnswer^{\phi}$	-0.273**	-0.327**	-0.398**	-0.504**
	(-2.05)	(-2.28)	(-2.04)	(-2.66)
PR		0.130***		0.234***
		(10.37)		(12.88)
$Negativity$	0.878**	0.726*	0.409	0.405
	(2.26)	(1.78)	(0.61)	(0.71)
$Uncertainty$	0.335	0.033	-0.347	-0.467
	(0.75)	(0.06)	(-0.51)	(-0.74)
$Complexity$	0.808	0.562	1.285	1.164
	(1.14)	(0.63)	(1.11)	(0.94)
$NewsWorthiness$	0.041***		0.091***	
	(7.71)		(8.44)	
Observations	17653	15007	17653	15007
R^2	0.709	0.831	0.884	0.943
Firm Controls	Yes	Yes	Yes	Yes
News Categories	No	Yes	No	Yes
QuarterYear FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Implied	Yes	Implied
FirmYear FE	No	Yes	No	Yes

Table 5: Management withholding information and (non-)professional media attention

Notes: OLS regressions for Equation (1). In columns (1) and (2) the dependent variable is the natural logarithm of one plus the count of full articles that professional business media published about the firm from the day after the earnings call until 60 days later. In columns (3) and (4) the dependent variable is the same metric for the non-professional media coverage. In columns (5) and (6) the dependent variable is the proportion of full articles by the professional media to the total number of full articles available in the RavenPack for the same analysis window. Firm controls include EarningsSurprise, BTM, ln(Assets) and Tobin's Q. News categories refer to news about a company in the analysis window being associated with the 25 groups of news classified using the topic modeling algorithm described in Subsection 3.2. Industry classification is based on the Fama-French 48 industries. All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	$Counts^F_{Pro}$		$Counts^F_{N-Pro}$		$Share-Pro$	
	(1)	(2)	(3)	(4)	(5)	(6)
$NonAnswer^{\phi}$	-0.709**	-1.152***	0.233	-0.440	-0.160**	-0.114**
	(-2.22)	(-3.88)	(0.76)	(-1.67)	(-2.15)	(-2.16)
PR	0.400***	0.348***	0.420***	0.351***	-0.005	-0.002
	(15.33)	(12.62)	(15.08)	(16.18)	(-0.95)	(-0.39)
$Negativity$	1.664	0.114	0.984	0.685	0.102	-0.145
	(1.64)	(0.12)	(0.99)	(0.93)	(0.50)	(-0.88)
$Uncertainty$	-0.482	-0.924	-0.355	-0.623	0.058	-0.001
	(-0.44)	(-0.86)	(-0.33)	(-0.68)	(0.25)	(-0.01)
$Complexity$	1.132	1.270	0.124	-0.743	0.029	0.194
	(0.70)	(0.80)	(0.06)	(-0.50)	(0.10)	(0.75)
$NewsWorthiness$	0.120***		0.134***		-0.002	
	(8.62)		(7.09)		(-0.82)	
Observations	15499	15007	15499	15007	15499	15007
R^2	0.806	0.895	0.907	0.956	0.840	0.925
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
News Categories	No	Yes	No	Yes	No	Yes
QuarterYear FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Implied	Yes	Implied	Yes	Implied
FirmYear FE	No	Yes	No	Yes	No	Yes

Table 6: Management withholding information and media coverage shift

Notes: OLS regressions for Equation (4). The sample is limited to earnings calls held on dates, on which at least two firms from the same Fama-French industry classification hold their calls. Observations are at the media source level. The dependent variable is a firm's change in weight in a media source's portfolio compared to their peers that hold earnings call on the same day. In columns (1) and (2), the non-answer score is the rank of the non-answer scores of the firms holding their calls on the same date. Firm controls include EarningsSurprise, BTM, ln(Assets) and Tobin's Q. News categories refer to news about a company in the analysis window being associated with the 25 groups of news classified using the topic modeling algorithm described in Subsection 3.2. Industry classification is based on the 48 Fama-French industries. All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered in the earnings call level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	$\Delta\omega$			
	(1)	(2)	(3)	(4)
$NonAnswer^{rank}$	0.002***	0.002***		
	(2.69)	(2.68)		
$NonAnswer^{\phi}$			-0.078*	-0.078*
			(-1.89)	(-1.88)
$Negativity$	-0.212	-0.212	-0.210	-0.210
	(-1.48)	(-1.47)	(-1.46)	(-1.45)
$Uncertainty$	0.269*	0.269*	0.259	0.259
	(1.70)	(1.69)	(1.62)	(1.61)
$Complexity$	0.148	0.148	0.153	0.153
	(0.56)	(0.56)	(0.58)	(0.57)
Observations	1487854	1487854	1487854	1487854
R^2	0.028	0.037	0.028	0.037
Firm Controls	Yes	Yes	Yes	Yes
News Categories	Yes	Yes	Yes	Yes
Source FE	No	Yes	No	Yes

5 Conclusion

Firm's earnings calls are a type of disclosure designed to reduce information asymmetries among investors, shareholders and market participants. The management must assure that this information is broadly available (Bushee et al., 2004). During the Q&A section of these conference calls, management should respond to participants' questions directly. Faced with a question, the management can faithfully addressing the request for information or a ''non-answer'', i.e., either in the form of a direct rejection like "we cannot provide this information" or indirectly blathering and beating around the bush. Non-answers leave the request for information unmet. This leads stakeholders to rely on other information intermediaries to provide them with the missing value-relevant knowledge.

This paper examines how the business media cover companies with different information environment richness. We state two competing hypotheses to verify if 1) the media coverage increases when a higher demand for information exists (a 'demand-driven' media coverage hypothesis), or 2) the media coverage decreases as supplying information about firms in poorer information environments is more challenging (a 'supply-driven' media coverage hypothesis).

Using data on media coverage of the S&P 500 firms between 2007 and 2019, we show that the management's decision to provide non-answers to questions during quarterly earnings calls, in line with the supply-driven coverage hypothesis, is associated with significantly less media attention on the firm in the next quarter. Separating the types of coverage, we show this decline is mainly due to fewer media sources publishing full articles about a firm, which require more (editorial) effort and resources to prepare than other types of coverage. Moreover, any drop in the number of articles published by professional media sources, like Dow Jones Newswire and Bloomberg, is more severe compared to that of the non-professional media like Seeking Alpha or stock blogs. Finally, we verify these findings in the media-level observations and show that within a given industry, media sources rank the firms based on the level of their non-answers and shift their attention to those who provide more answers in their earnings calls.

The recent increase in the body of the literature regarding media coverage is mainly

due to the emergence of new datasets, such as RavenPack (Tetlock, 2014), which is still evolving and increasing the scope of its data availability (Miller and Skinner, 2015). Future research will benefit from an extended timeline of coverage provided by ever-growing media sources and focusing on more firms, which will pave the way for deeper media-level analyses in the literature.

References

Baloria, V. P. and Heese, J. (2018). The effects of media slant on firm behavior. Journal of Financial Economics, 129(1):184--202.

Barker, R., Hendry, J., Roberts, J., and Sanderson, P. (2012). Can company-fund manager meetings convey informational benefits? exploring the rationalisation of equity investment decision making by uk fund managers. Accounting, Organizations and Society, 37(4):207 -- 222.

Barth, A., Mansouri, S., and Woebbeking, F. (2020). ''let me get back to you'' - a machine learning approach to measuring non-answers. Available at SSRN 3567724.

Barth, A., Mansouri, S., Woebbeking, F., and Zörgiebel, S. (2019). How to talk down your stock returns. Working paper.

Bhushan, R. (1989). Collection of information about publicly traded firms: Theory and evidence. Journal of Accounting and Economics, 11(2-3):183--206.

Blankespoor, E., deHaan, E., and Marinovic, I. (2020). Disclosure processing costs, investors' information choice, and equity market outcomes: A review. Journal of Accounting and Economics, page 101344.

Blankespoor, E., deHaan, E., and Zhu, C. (2018). Capital market effects of media synthesis and dissemination: Evidence from robo-journalism. Review of Accounting Studies, 23(1):1--36.

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

Brown, L. D., Call, A. C., Clement, M. B., and Sharp, N. Y. (2015). Inside the "black box" of sell-side financial analysts. Journal of Accounting Research, 53(1):1--47.

Brown, S., Hillegeist, S. A., and Lo, K. (2004). Conference calls and information asymmetry. Journal of Accounting and Economics, 37(3):343--366.

Bushee, B. J., Core, J. E., Guay, W., and Hamm, S. J. (2010). The role of the business press as an information intermediary. Journal of accounting research, 48(1):1--19.

Bushee, B. J., Gow, I. D., and Taylor, D. J. (2018). Linguistic complexity in firm disclosures: Obfuscation or information? Journal of Accounting Research, 56(1):85--121.

Bushee, B. J., Matsumoto, D. A., and 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.

Bushee, B. J. and Miller, G. S. (2007). Investor relations, firm visibility, and investor following. Firm Visibility, and Investor Following (August 2007).

Cahan, S. F., Chen, C., Chen, L., and Nguyen, N. H. (2015). Corporate social responsibility and media coverage. Journal of Banking & Finance, 59:409--422.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. Journal of Business & Economic Statistics, 29(2):238--249.

Campbell, J. L., DeAngelis, M. D., and Moon, J. R. (2019). Skin in the game: Personal stock holdings and investors' response to stock analysis on social media. Review of Accounting Studies, 24(3):731--779.

Chen, H., De, P., Hu, Y. J., and Hwang, B.-H. (2014). Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. The Review of Financial Studies, 27(5):1367--1403.

Cohen, L., Lou, D., and Malloy, C. (2013). Playing favorites: How firms prevent the revelation of bad news. Technical report, National Bureau of Economic Research.

Core, J. E., Guay, W., and Larcker, D. F. (2008). The power of the pen and executive compensation. Journal of financial economics, 88(1):1--25.

Damodaran, A. (2015). The Weekend Effect in Information Releases: A Study of Earnings and Dividend Announcements. The Review of Financial Studies, 2(4):607--623.

Dougal, C., Engelberg, J., García, D., and Parsons, C. A. (2012). Journalists and the Stock Market. The Review of Financial Studies, 25(3):639--679.

Drake, M. S., Guest, N. M., and Twedt, B. J. (2014). The media and mispricing: The role of the business press in the pricing of accounting information. The Accounting Review, 89(5):1673--1701.

Drake, M. S., Thornock, J. R., and Twedt, B. J. (2017). The internet as an information intermediary. Review of Accounting Studies, 22(2):543--576.

Dyck, A., Morse, A., and Zingales, L. (2010). Who blows the whistle on corporate fraud? The journal of finance, 65(6):2213--2253.

Dyck, A., Volchkova, N., and Zingales, L. (2008). The corporate governance role of the media: Evidence from russia. The Journal of Finance, 63(3):1093--1135.

Dyck, A. and Zingales, L. (2003). The media and asset prices. Technical report, Working Paper, Harvard Business School.

Dzieliński, M., Wagner, A. F., and Zeckhauser, R. J. (2017). Straight talkers and vague talkers: The effects of managerial style in earnings conference calls. National Bureau of Economic Research, No. w23425.

Eickhoff, M. and Neuss, N. (2017). Topic modelling methodology: its use in information systems and other managerial disciplines.

Ellis, J. A., Fee, C. E., and Thomas, S. E. (2012). Proprietary costs and the disclosure of information about customers. Journal of Accounting Research, 50(3):685--727.

Engelberg, J. E. and Parsons, C. A. (2011). The causal impact of media in financial markets. The Journal of Finance, 66(1):67--97.

Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. The Journal of Finance, 64(5):2023--2052.

Frankel, R., Kothari, S., and Weber, J. (2006). Determinants of the informativeness of analyst research. Journal of Accounting and Economics, 41(1-2):29--54.

Gao, P., Lee, C., and Murphy, D. (2020). Financing dies in darkness? the impact of newspaper closures on public finance. Journal of Financial Economics, 135(2):445--467.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. Journal of Economic Literature, 57(3):535--74.

Gow, I. D., Larcker, D. F., and Zakolyukina, A. A. (2019). Non-answers during conference calls. Chicago Booth Research Paper 19-01.

Guay, W., Samuels, D., and Taylor, D. (2016). Guiding through the fog: Financial statement complexity and voluntary disclosure. Journal of Accounting and Economics, 62(2-3):234--269.

Guest, N. M. (2018). Do journalists help investors analyze firms' earnings news? Available at SSRN 3067387.

Guest, N. M. and Kim, J. (2020). On the interplay among information intermediaries: How does analyst coverage affect media coverage? Available at SSRN 3337116.

Gurun, U. G. and Butler, A. W. (2012). Don't believe the hype: Local media slant, local advertising, and firm value. The Journal of Finance, 67(2):561--598.

Hafez, P. and Xie, J. (2014). Web news analytics enhance stock portfolio returns. Available at SSRN 2423362.

Hillert, A., Jacobs, H., and Müller, S. (2014). Media makes momentum. The Review of Financial Studies, 27(12):3467--3501.

Hirshleifer, D. A. and Sheng, J. (2019). The attention trigger effect: macro news and efficient processing of micro news. Available at SSRN 2890864.

Hollander, S., Pronk, M., and Roelofsen, E. (2010). Does silence speak? an empirical analysis of disclosure choices during conference calls. Journal of Accounting Research, 48(3):531--563.

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

Kimbrough, M. D. and Louis, H. (2011). Voluntary Disclosure to Influence Investor Reactions to Merger Announcements: An Examination of Conference Calls. The Accounting Review, 86(2):637--667.

Lang, M. H. and Lundholm, R. J. (1996). Corporate disclosure policy and analyst behavior. Accounting review, pages 467--492.

Loughran, T. and McDonald, B. (2011). When is a Liability not a Liability? Journal of Finance, 66(1):35 -- 65.

Loughran, T. and Mcdonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. Journal of Accounting Research, 54(4):1187--1230.

Loughran, T. and McDonald, B. (2020). Measuring firm complexity. Working paper.

Martineau, C. and Zoican, M. (2020). Crowded analyst coverage. In Proceedings of Paris December 2019 Finance Meeting EUROFIDAI-ESSEC.

Matsumoto, D., Pronk, M., and 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.

Mayew, W. J. (2008). Evidence of management discrimination among analysts during earnings conference calls. Journal of Accounting Research, 46(3):627--659.

Miller, G. S. (2006). The press as a watchdog for accounting fraud. Journal of Accounting Research, 44(5):1001--1033.

Miller, G. S. and Skinner, D. J. (2015). The evolving disclosure landscape: How changes in technology, the media, and capital markets are affecting disclosure. Journal of Accounting Research, 53(2):221--239.

Nikita, M. and Chaney, N. (2016). Tuning of the latent dirichlet allocation models parameters.

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

Puglisi, R. and Snyder Jr, J. M. (2011). Newspaper coverage of political scandals. The Journal of Politics, 73(3):931--950.

Puglisi, R. and Snyder Jr, J. M. (2015). Empirical studies of media bias. In Handbook of media economics, volume 1, pages 647--667. Elsevier.

RavenPack (2017). Why investment firms choose ravenpack.

Reuter, J. and Zitzewitz, E. (2006). Do ads influence editors? advertising and bias in the financial media. The Quarterly Journal of Economics, 121(1):197--227.

Rogers, J. L., Skinner, D. J., and Zechman, S. L. (2016). The role of the media in disseminating insider-trading news. Review of Accounting Studies, 21(3):711--739.

Taddy, M. (2013). Multinomial inverse regression for text analysis. Journal of the American Statistical Association, 108(503):755--770.

Tetlock, P. C. (2014). Information transmission in finance. Annu. Rev. Financ. Econ., 6(1):365--384.

Tetlock, P. C. (2015). The role of media in finance. In Handbook of media Economics, volume 1, pages 701--721. Elsevier.

Westphal, J. D. and Deephouse, D. L. (2011). Avoiding bad press: Interpersonal influence in relations between ceos and journalists and the consequences for press reporting about firms and their leadership. Organization Science, 22(4):1061--1086.

Zavyalova, A., Pfarrer, M. D., Reger, R. K., and Shapiro, D. L. (2012). Managing the message: The effects of firm actions and industry spillovers on media coverage following wrongdoing. Academy of Management Journal, 55(5):1079--1101.

Zhou, D. (2018). Do numbers speak louder than words? Technical report.

Zingales, L. (2000). In search of new foundations. The journal of Finance, 55(4):1623--1653.

Appendix A

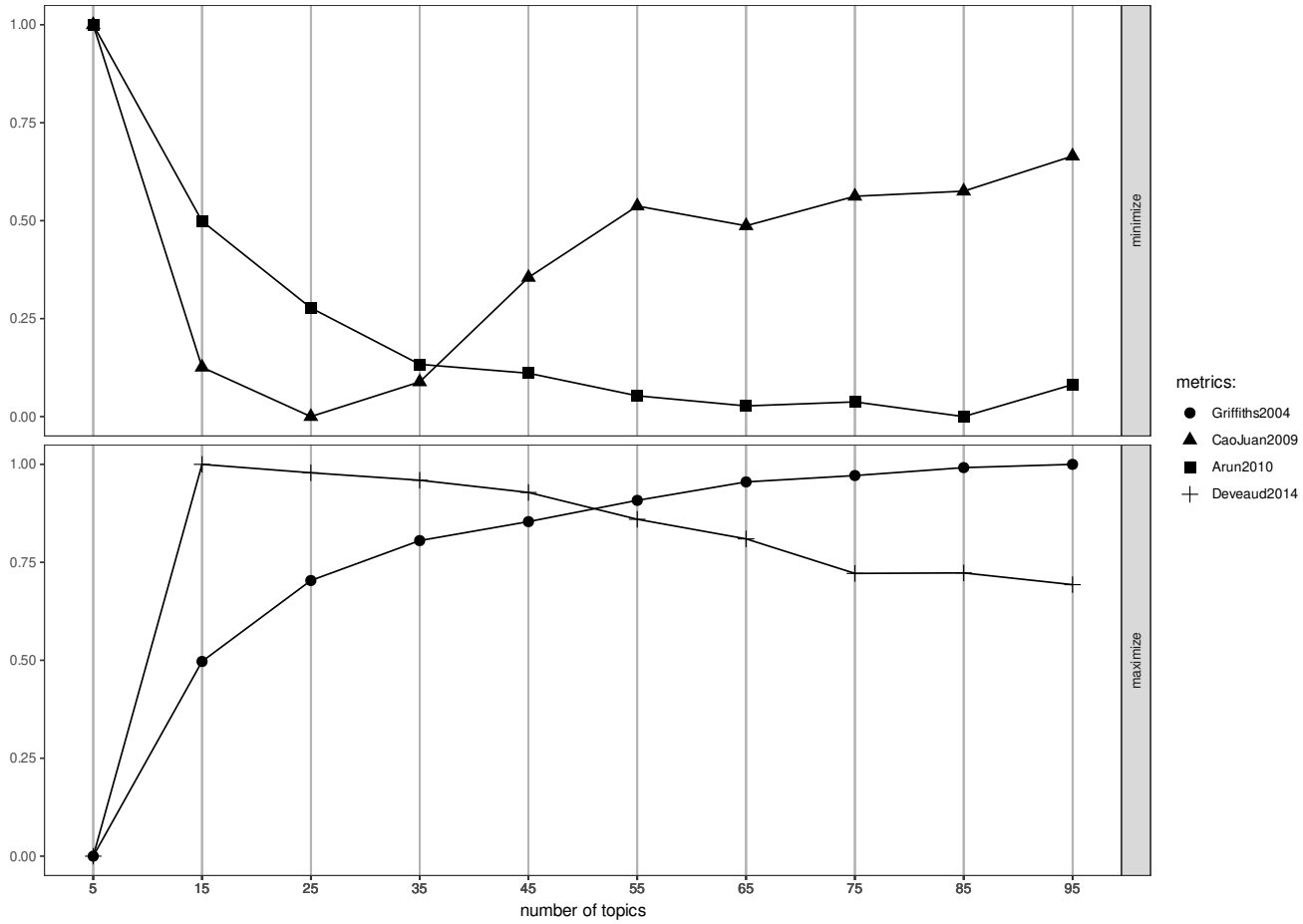
Table A1: Definition of variables

Variable	Definition
BTM	book-to-market ratio; defined as the total common/ordinary equity divided by the market value of equity (from Compustat)
$Complexity$	the ratio of the complex words to the total words in the answers provided by the management. "Complexity" word list provided by Loughran and McDonald (2020)
$EarningsSurp$	represents the grouping of all firms in deciles of earnings surprise following Dzieliński et al. (2017) (defined as the difference between the actual and the consensus forecast earnings (from I/B/E/S) as a ratio to the share price 5 trading days before the announcement)
$ln(Assets)$	the natural logarithm of total assets. (from Compustat)
$Counts$	the natural logarithm of one plus the total number of news pieces in a two-month period after firms' earnings calls
$Sources$	the natural logarithm of one plus the number of unique news sources that publish a piece in a two-month period after firms' earnings calls (From Ravenpack)
$Sources^F_{N-Pro}$	the $Sources^F$ variable filtered for the sources we identify as non-professional.
$Sources^F_{Pro}$	the $Sources$ variable filtered for the sources we identify as professional. We label a source as professional if it is associated with Dow Jones Newswire, Barron's, Marketwatch, Bloomberg News, Thomson Reuters, the Wall Street Journal, the Financial Times, Entrepreneur, Business Insider, CNBC and Forbes.
$Negativity$	the ratio of the negative words to the total words in the answers provided by the management. "negative" word list provided by Loughran and McDonald (2011)
$NewsWorthiness$	inspired by Dyck et al. (2008), defined as the natural logarithm of one plus the number of articles in the Wall Street Journal and the Financial Times that refer to a company during the six-month period before its earnings call (excluding the day of earnings call and one day before it)
$NonAnswer^{\phi}$	the ratio of trigrams weighted by their corresponding loadings in the non-answer glossary of Barth et al. (2020) to the total words
PR	The natural logarithm of one plus the firm's initiated press releases (from Ravenpack PR-edition)

Q	the Tobin's Q; the book value of assets minus book value of common equity plus the market value of common equity, divided by the total book value of assets (from Compustat)
$Uncertainty$	the ratio of the uncertain words to the total words in the management answers. ''Uncertainty'' word list provided by Loughran and McDonald (2011)

Appendix B Topic modeling of news

Figure B1: News categories
The figure shows LDA's goodness of fit for our data of news categories by different number of topics ranging from 5 to 95 according to scores provided in the R-package "ldatuning" of Nikita and Chaney (2016).



Appendix C List of the professional media

RavenPack ID	Name
18A55F	BARRONS
8B4BD7	BARRONS.COM BLOG
97AF0A	BARRONS.COM ONLINE
6DFE43	BLOOMBERG BUSINESSWEEK
5F78A9	BLOOMBERG GOVERNMENT
208421	BLOOMBERG NEWS
CAF003	BLOOMBERG VIEW
FA7478	BLOOMBERG-QUINT
DCD6DA	BREAKINGVIEWS
8B7199	BRUSSELS BLOG
87C2EA	BUSINESS BLOG
C75B8C	BUSINESS INSIDER
E38558	CAPITOL REPORT - MARKET WATCH
AA1167	CNBC
B5569E	DOW JONES NEWSWIRES
A89221	DOW JONES ONLINE

RavenPack ID	Name
0FDF1E	ECONOMISTS' FORUM
F67012	EMERGING MARKETS DAILY
DF7445	ENCORE
2143AD	ENERGY TICKER
938822	ENTREPRENEUR
931400	FAITHWORLD
0B0728	FAST MONEY
D335E4	FELIX SALMON
35913F	FELIX SALMON - ALL POSTS
90CE21	FINANCIAL REGULATORY FORUM
FD0B00	FINANCIAL TIMES
75B2CD	FINDLAW
DE57D6	FOCUS ON FUNDS
22AC8B	FORBES.COM
BF0799	FROM REUTERS.COM
0EB1B3	FT ADVISER
8E9A55	FT ADVISER MONEY MANAGEMENT

RavenPack ID	Name
766283	FT ALPHAVILLE
895559	FT DATA
6DDC36	FT INVESTMENT ADVISER
EF1ABD	FT.COM - CREDIT SQUEEZE
F1D2BD	GAVYN DAVIES
D3E2F8	GLOBAL INVESTING
8A7104	HUGO DIXON
358DFE	INCOME INVESTING
B4A99C	INDIA INSIGHT
A0588D	JAMES SAFT
AF0676	MACROSCOPE
A0099A	MAD MONEY WITH JIM CRAMER
1E5E35	MARKETWATCH
C325FC	MARKETWATCH (ONLINE)
E04BE4	MORNINGSTAR
A92D7D	PHOTOGRAPHERS BLOG

RavenPack ID	Name
751371	REUTERS
806C8E	SILICON ALLEY INSIDER
C76E42	SMART MONEY
3F9DB7	STOCKS TO WATCH
499718	TECH BLOG
F707BB	TECH CHECK WITH JIM GOLDMAN
9BA337	TECH TRADER DAILY
13A271	THE A-LIST
ADFD64	THE BANKER
DE348B	THE GREAT DEBATE
C0AB6A	THE TELL
F73069	THE WORLD
3EA04F	THOMSON REUTERS FOUNDATION NEWS
AA6E89	WALL STREET JOURNAL
9AE635	WALL STREET JOURNAL (ONLINE)
0BBE7B	WESTMINSTER BLOG

Appendix D Robustness

Table D1: Management withholding information and media coverage -- The first 48 hours after the conference call

Notes: OLS regressions for Equation (1). In columns (1) to (3) the dependent variable is the natural logarithm of one plus the number of distinct news agencies that published at least one news piece in the first 48 hours after the earnings call. In columns (4) to (6) the dependent variable is the natural logarithm of one plus the count of total pieces published about the firm in the same window. Firm controls include EarningsSurprise, BTM, ln(Assets) and Tobin's Q. News categories refer to news about a company in the analysis window being associated with the 25 groups of news classified using the topic modeling algorithm described in Subsection 3.2. Industry classification is based on the 48 Fama-French industries. All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	Sources			*Counts*		
	(1)	(2)	(3)	(4)	(5)	(6)
$NonAnswer^{\phi}$	-1.093***	-0.380**	-0.377*	-1.530***	-0.340*	-0.337*
	(-2.83)	(-2.04)	(-2.00)	(-3.24)	(-1.79)	(-1.77)
PR	0.334***	0.141***	0.141***	0.480***	0.242***	0.242***
	(11.59)	(8.61)	(8.55)	(11.10)	(11.24)	(11.30)
$Negativity$	-1.643	1.473***	1.534***	-2.203	2.988***	3.007***
	(-1.46)	(3.23)	(3.29)	(-1.47)	(4.28)	(4.28)
$Uncertainty$	-4.769***	-0.752	-0.766	-5.669***	-0.833	-0.832
	(-4.78)	(-1.25)	(-1.25)	(-4.09)	(-1.06)	(-1.05)
$Complexity$	-3.424*	-0.600	-0.660	-3.076	-2.023	-2.067
	(-1.71)	(-0.55)	(-0.60)	(-1.12)	(-1.61)	(-1.63)
$NewsWorthiness$	0.257***			0.322***		
	(18.69)			(19.73)		
Observations	11036	10208	10208	11036	10208	10208
R^2	0.764	0.936	0.937	0.760	0.938	0.938
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
News Categories	No	No	Yes	No	No	Yes
QuarterYear FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Implied	Implied	Yes	Implied	Implied
Firm FE	No	Implied	Implied	No	Implied	Implied
FirmYear FE	No	Yes	Yes	No	Yes	Yes

Table D2: Management non-answer and media coverage

Notes: OLS regressions for Equation (1). In columns (1) to (5) the dependent variable is the natural logarithm of one plus the number of distinct news agencies that published at least one article about the firm from the day after the earnings call until 60 days later. Firm controls include EarningsSurprise, BTM, ln(Assets) and Tobin's Q. News categories refer to the association of the companies' news in the analysis window with the 25 groups of news classified using the topic modeling algorithm described in Subsection 3.2. Industry classification is based on the 48 Fama-French industries. All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered in the firm and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	Sources				
	(1)	(2)	(3)	(4)	(5)
$NonAnswer^{\phi}$	-5.496***	-1.602***	-0.541**	-0.336*	-0.345**
	(-4.08)	(-4.66)	(-2.47)	(-1.99)	(-2.06)
PR		0.484***	0.235***	0.179***	0.179***
		(14.56)	(11.68)	(11.37)	(11.28)
$Negativity$		-3.288***	0.095	0.570	0.572
		(-3.14)	(0.12)	(1.24)	(1.26)
$Uncertainty$		-4.433***	-0.182	-0.366	-0.417
		(-4.20)	(-0.27)	(-0.69)	(-0.79)
$Complexity$		-0.737	1.134	0.329	0.295
		(-0.37)	(0.91)	(0.39)	(0.35)
$NewsWorthiness$		0.268***	0.084***		
		(18.00)	(8.87)		
Observations	22604	18923	19066	18513	18513
R^2	0.009	0.873	0.924	0.970	0.970
Firm Controls	No	Yes	Yes	Yes	Yes
News Categories	No	No	No	No	Yes
QuarterYear FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Implied	Implied	Implied
Firm FE	No	No	Yes	Implied	Implied
FirmYear FE	No	No	No	Yes	Yes

Chapter IV

ICO analysts

ICO Analysts *

Andreas Barth, Valerie Laturnus, Sasan Mansouri and Alexander F. Wagner†

May 30, 2021

ABSTRACT

Initial Coin Offerings (ICOs) provide a clean opportunity and rich data to study the contribution of analysts to the functioning of capital markets. The assessments of freelancing ICO analysts vary in quality and exhibit biases due to the reciprocal interactions of analysts with ICO team members. Ratings predict ICO success, but imperfectly. Even favorably rated ICOs tend to fail when a greater portion of their ratings reciprocate prior ratings. Failure despite strong ratings is also frequent when analysts have a history of optimism, and when reviews strike a particularly positive tone. These findings suggest that information about the track record of analysts and their potentially conflicting activities is valuable to investors.

JEL classification: G14, G24, L26, D82, D83

Keywords: Analysts, Asymmetric Information, FinTech, Initial Coin Offering (ICO)

*We are grateful to Christian Leuz, Evgeny Lyandres, and Arthur Petit-Romec for comments. Wagner acknowledges financial support from the University of Zurich Research Priority Program "Financial market regulation". We declare that we have no relevant or material financial interests that relate to the research described in this paper.

†Barth, Laturnus and Mansouri are with Goethe University Frankfurt, Wagner is with University of Zurich, CEPR, ECGI and SFI. ✉: Andreas.Barth@finance.uni-frankfurt.de ;Laturnus@finance.uni-frankfurt.de ;mansouri@finance.uni-frankfurt.de ; alexander.wagner@bf.uzh.ch

1 Introduction

The question of how analysts contribute to the functioning of capital markets has been on the agenda of accounting research for years (Bradshaw et al., 2017). This paper uses the setting of Initial Coin Offerings (ICOs) to investigate determinants and consequences of the quantitative and qualitative aspects of investment ratings issued by human experts (henceforth ICO analysts).[1] Ratings predict the success of ICOs, but imperfectly. Even among ICOs with an average rating in the top quartile, more than 50% fail. As a novel contribution, this analysis addresses potential conflicts of interest in ICO analyst ratings. We find that ICO analysts tend to reciprocate favorable ratings for their own ventures; however, the results also suggest that investors place lower emphasis on reciprocating ratings.

Studying ICOs and their analysts is of interest for at least four reasons. First, the ICO environment provides a clean setup to investigate how analysts contribute to capital markets. The market is particularly interesting to study the role of intermediaries because its regulation has only recently begun to clarify.[2]

Second, like financial analysts, ICO analysts potentially suffer from conflicts of interests.[3] However, the conflicts in this case (i) are potentially more extreme and (ii) can be more directly identified than in the case of the typical security analysts. As for (i), ICO analysts do not only *provide ratings* for ICOs, but may also *run their own* ICOs. Thus, whenever an ICO analyst i provides a rating for an ICO j, there is a chance that a team member of this ICO j will rate for the ICO of analyst i at a later stage. As for (ii), most of the literature on financial analysts classifies analysts as "affiliated" (and thus potentially conflicted) if they belong to a bank that has an underwriting relationship with the firms on which they are reporting. The potential bias of revolving door equity

[1] Initial Coin Offerings (ICOs) are token sale events on an own or existing blockchain that facilitate financing for an entrepreneurial venture.

[2] The U.S. Securities and Exchange Commission (SEC) does consider tokens to be securities, and it recently won a landmark lawsuit against Telegram for a violation of federal securities laws in the course of its ICO. See `https://www.sec.gov/news/press-release/2020-146`. European economies tend to be significantly less stringent in their regulatory approaches, see Kaal (2018); Tiwari et al. (2019); thinkBLOCKtank (2019).

[3] See, for example, Lin and McNichols (1998), Michaely and Womack (1999), and Chan et al. (2007) for evidence of biased financial analysts.

analysts can only be identified ex post that is largely hidden information. By contrast, the ICO setting presents a situation where linkages are more direct and where investors can be aware of potential biases right away.

Third, non-professional analysts and their crowd forecasts have been shown to be important information intermediaries for equity investors (Chen et al., 2014; Jame et al., 2016; Drake et al., 2017; Campbell et al., 2019; Farrell et al., 2020; Da and Huang, 2020). However, we know little about the potential conflicts of interest that such analysts face and whether market participants consider the differential credibility and informativeness of these analyses in their investment decisions.

Fourth, tokens offerings are a potentially powerful instrument for new ventures to obtain crowdfunding-like resources (Goldstein et al., 2019; Lyandres, 2019; Chod and Lyandres, 2020; Lee and Parlour, 2020; Li and Mann, 2020; Lyandres et al., 2020; Gryglewicz et al., 2020). Therefore, understanding the workings of this relatively new market for the funding of companies and products is important.

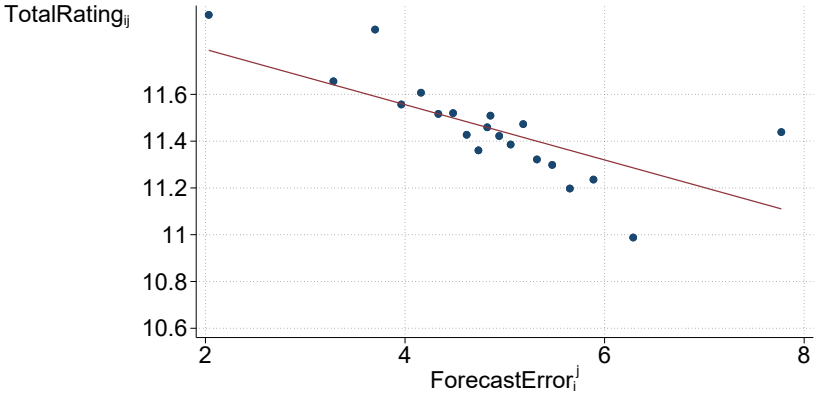
We collect data on 5,384 ICOs between 2015 and 2020 from the platform ICObench.com. We identify 531 experts who issued a total of 13,834 ratings. Figure 1 illustrates some main results using binned scatter plots.

We begin by investigating determinants of analysts' ratings. We first find that analysts who, in the past, had issued very positive ratings for ICOs that did not succeed (i.e., analysts with large forecast errors) provide, on average, lower ratings in the future (see Panel (a) of Figure 1). ICObench.com also rank ICO analysts, which gives an equivalent setting to all-star financial analysts (Leone and Wu, 2002). We observe that "star analysts" are less optimistic and their ratings are, on average, lower. In addition to quantitative ratings, we also consider the length and linguistic tone of the reviews that accompany the evaluation, i.e., the qualitative nature of ICO analyst ratings. We observe that lower ratings often accompany longer reviews with a more negative tone. In all these analyses, we compare different ratings for the same ICO, which helps to rule out that these results are purely driven by the self-selection of analysts to certain ICOs.

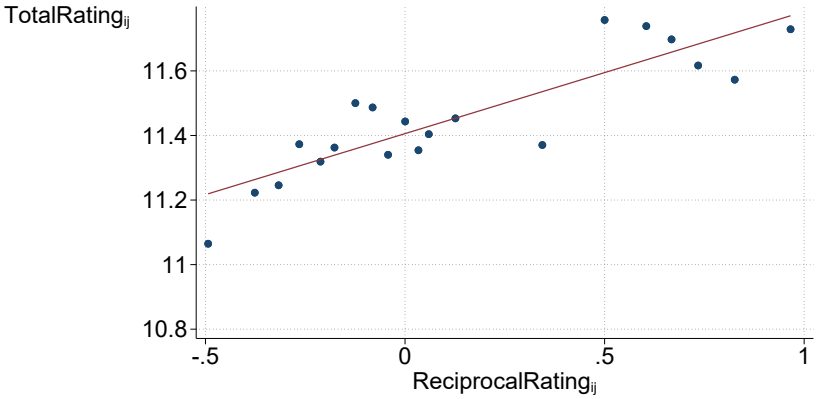
Importantly, reciprocal ratings are special (see Panel (b) of Figure 1): the total rating

Figure 1: Summary of the main results
The figure shows binned scatter plots summarizing the main results. Panels (a) and (b) use within ICO variation, i.e., ICO fixed effects are absorbed. All variables are defined in Table A1.

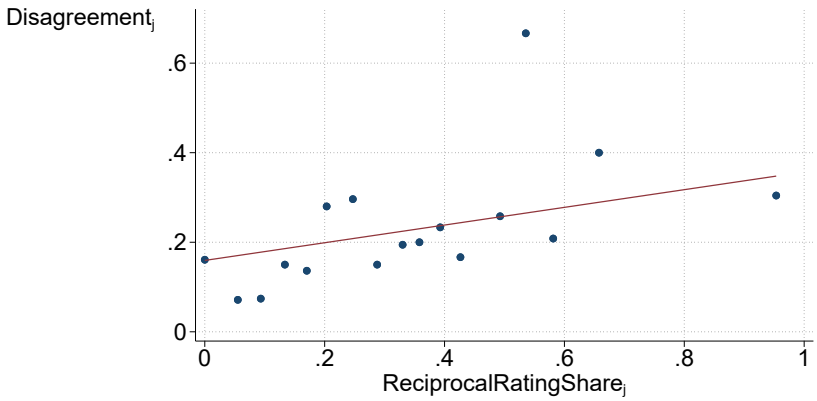
(a) Optimistic analysts become more careful



(b) Reciprocal ratings are more favorable



(c) Even ICOs with high average ratings fail frequently, and especially so when many ratings are reciprocal



score an analyst gives to an ICO j is higher if she received a rating in the past for her own ICO by any team member of coin j. This effect is stronger the higher the prior received

rating was. These effects continue to hold when we compare analysts providing a rating to the same ICO in a given month. Comparing different assessments of the same analyst and for the same ICO allows also to rule out that the optimistic assessment is due to the high difficulty of forecasting tasks or due to a non-random match between founders of good ICOs that also serve as analysts.

Next, we analyze the explanatory power of ICO analyst ratings for the success of an ICO campaign, and for the failure of an ICO campaign despite strong analyst endorsement. We first confirm the result of prior work that investors appear to value the fact that a human analyst provided a rating for the ICO.[4] Moreover, a better average quantitative ICO analyst rating translates into a higher success probability for the respective ICO.

However, while the unconditional failure rate is about 64%, even 53.6% of ICOs with an average rating in the top quartile fail. Our main interest is in the characteristics of analysts or the ICO itself that lead to such disagreement between analysts' advice and the market outcome.

The share or reciprocal ratings is an important determinant of failure despite high ratings: If ICO j receives a rating from many reciprocal analysts, i.e., analysts whose rating is a response to a rating they received from a team member of ICO j, the market is more likely to disregard analyst recommendations; see Panel (c) of Figure 1. There are two possible interpretations of this result. First, it is conceivable that, even though we control for a wide variety of factors presumably capturing variation in ICO quality, reciprocal ratings occur with "objectively" bad ICOs, i.e., they pick up some additional variation in quality. Second, investors may trust ICOs with more reciprocal ratings less (even when they may potentially be worth funding).[5] Either way, the findings imply that investors do not blindly pile capital into highly rated ICOs.

Interesting patterns also emerge for the linguistic measures of the rating. The length and linguistic tone of the reviews that accompany the evaluation explain only little of the variation in the success of ICOs. However, the likelihood that an ICO fail despite

[4]This is in line with several studies that document the benefits of analyst coverage (Sufi, 2009; Demiroglu and Ryngaert, 2010; Mola et al., 2013; Crawford et al., 2012).

[5]Several studies discuss whether or not investors are sophisticated enough to detect biased ratings (Ellis, 1998; Baker and Mansi, 2002; Livingston et al., 2010; Hirth, 2014; Badoer et al., 2019).

high ratings increases with the positivity of the tone and complexity of the language in the reviews.

Finally, the quantitative and qualitative ratings by human analysts do not systematically differ on average for ICOs that prove to be fraudulent. A higher share of reciprocal ratings is not associated with a higher fraud probability, suggesting that criminal intentions do not typically drive reciprocity. ICOs exhibiting fraud do show a larger dispersion of both rating scores and tone of rating reviews among analysts.

Overall, the results suggest that investors value the access to information about the track record and potentially conflicting activities of analysts. Having the chance to easily collect these information allows ICO investors to respond to qualitative differences among analyst ratings in a differentiated way. Thus, even though ICO analyst activity is relatively unregulated, the investors' ability to access information about analysts seems to at least partially rectify the problems that could arise from conflicts of interest or otherwise biased analyst recommendations.[6] This suggests that information intermediaries and platforms collecting data about ICO analysts play an important role in the functioning of markets.

These results add to the literature in four important ways. First, the literature on financial analysts suggests that a close link between analysts and firms leads to superior information and better assessments (Bradley et al., 2017; Bae et al., 2008), but also highlights the problem of conflicts of interest in a similar spirit of "affiliated" analysts (e.g. Lin and McNichols, 1998; Michaely and Womack, 1999; O'Brien et al., 2005; Malmendier and Shanthikumar, 2007; Agrawal and Chen, 2008; Kadan et al., 2009) or revolving door analysts (Lourie, 2019; Kempf, 2020).[7] However, there is scarce data on direct interactions of analysts with the firms they analyze. The data on ICO analysts provide

[6]Transparency about the background, characteristics, or track records of information providers and intermediaries has been identified as critically important in other settings. For example, Law and Mills (2019) highlight the importance of the transparency provided by the Financial Industry Regulatory Authority (FINRA) about brokers' (criminal) backgrounds. In academia, too, market discipline based on transparent disclosure can work. For example, Leuz et al. (2021) find that medical scholars cite papers reporting research sponsored by drug companies less frequently. For an overview of the literature on the economic consequences of disclosure regulation, see Leuz and Wysocki (2016).

[7]A similar conflict of interest is present for rating agencies (e.g. Bolton et al., 2012; Bar-Isaac and Shapiro, 2013; Baghai and Becker, 2017; Chu and Rysman, 2019).

distinct advantages in that respect, and by showing that investors do take differences among analysts into account we highlight that these data are of value to investors.

Second, the paper complements the literature on semi-professional analysts in equity markets (Chen et al., 2014; Drake et al., 2017). That literature recognizes the possibility of conflicts of interest if the semi-professional analyst is holding positions on the stock themselves, resulting in a subjective, distorted analysis (Campbell et al., 2019).[8] While these studies focus on equity markets in which semi-professional analysts complement the information produced by professional analysts, one particular advantage of the ICO market, besides very detailed and structured information, is the absence of professional analysts.[9]

Third, the paper adds to the growing literature on the relationship between machine-generated evaluations and human expert ratings.[10] In addition to human evaluations, many platforms set up machine-generated ratings. These ratings do not evaluate the content of an ICO, but are based on observable factors such as features of the ICO's campaign and team.[11] Importantly, we show that both ratings are informative about ICO success. However, many ICOs fail despite high ratings, which is why we analyze this discrepancy.

Finally, ICOs are a potentially powerful way to fund new ventures, not least because of the underlying distributed ledger-based technology and the platform's special features (Bakos and Halaburda, 2019; Biais et al., 2019; Cong and He, 2019; Easley et al., 2019; Hinzen et al., 2020). This paper advances our knowledge of why and for which ICO investors are willing to provide funding. Usually, the sale of tokens or ICOs appear at a very early planning stage of a product's or a firm's life cycle and suffer from severe

[8]Campbell et al. (2019) use non-professional analysts' disclosures of stock positions as an indicator of the analyst's position, which may not be reported truthfully.

[9]There are of course many stocks that professional analysts do not cover. This lack of coverage is the is the analysts' choice, however, and as such provides information to the market.

[10]For example, Aubry et al. (2019) use data on paintings auctioned to study the accuracy and usefulness of valuations generated by using a pricing algorithm based on neural networks. With data from a leading startup accelerator, Catalini et al. (2018) show that artificial intelligence can help humans to screen and evaluate information when there is an information overload.

[11]Automated algorithms that simply count disclosed information are usually applied. For example, a high number of social media platforms on which an ICO is present or being listed on several rating websites automatically improves the rating for the respective ICO (Boreiko and Vidusso, 2019).

information asymmetry and adverse selection problems (Malinova and Park, 2018; Chod and Lyandres, 2020; Chod et al., 2020; Gan et al., 2020). As such, tokens have no intrinsic value at the time of the investment. Instead, they rather derive value from trust in future usage (Conley, 2017). Hence, the literature has investigated both the supply side, i.e., choices by ICO entrepreneurs (Adhami et al., 2018; Amsden and Schweizer, 2018; Benedetti and Kostovetsky, 2018; Cerchiello and Toma, 2018; Roosenboom et al., 2020; Howell et al., 2019; Fisch, 2019; Ernst and Young, 2018; Chakraborty and Swinney, 2020; PwC, 2019; Deng et al., 2018; Davydiuk et al., 2019), and the demand side, i.e. choices by investors (Fahlenbrach and Frattaroli, 2020; Fisch et al., 2019; Fisch and Momtaz, 2020).[12] Little attention has been paid to the information providing intermediaries in between supply and demand, however, and the literature largely focuses on the governance role of whitepapers provided by the ICO team (Adhami et al., 2018; Feng et al., 2019; Giudici and Adhami, 2019; Zhang et al., 2019; Samieifar and Baur, 2020; Florysiak and Schandlbauer, 2019).

To the best of our knowledge, only three previous papers examine ICO analysts (Aggarwal et al., 2019; Bourveau et al., 2019; Lee et al., 2019).[13] All three document that ICOs with higher expert assessments are more successful. Our baseline results confirm this finding, but we focus on the striking fact more than 50% of the ICOs with the highest quartile of ratings fail. We show that accounting for the heterogeneity among analysts is important. For example, we exploit the specific feature of the market that ICO analysts provide ICO ratings, while also often running their own ICOs. We show that reciprocal ratings are biased, but also that investors discount such reciprocal ratings. We also uncover several other predictors of failure despite praise from analysts.

The rest of the paper is organized as follows. Section 2 presents the data and descriptive statistics. Section 3 describes the results, and Section 4 concludes.

[12]There is also a literature on price dynamics of tokens (Cong et al., 2020a,b; Lee and Parlour, 2020; Li and Mann, 2020) as well as studies of asset pricing properties of coins on secondary markets and post-ICO performance (Dittmar and Wu, 2019; Hu et al., 2019; Fisch and Momtaz, 2020; Lyandres et al., 2020). See Li and Mann (2019) for a review of recent literature advances in ICO research.

[13]Momtaz (2020) uses information on ICO analysts, too. However, the paper does not focus on ICO analysts per se, but uses their evaluations as a proxy for one dimension of project quality.

2 Data and descriptive statistics

2.1 Sample and data source

We collect data on ICOs, ICO ratings and ICO experts from the platform ICObench.com. Our sample consists of 5,384 ICOs (of which 2,378 were rated by at least one expert) and spans the time period from the first ICOs in August 2015 to February 2020. ICOs in our sample were launched in 127 different countries, of which the USA, Singapore and the UK have the highest market shares.[14]

2.2 ICO analysts

In contrast to regulated financial analysts, ICO analysts are not certified. However, they have to apply for expert status on a platform, in our case ICObench.com. In their application, experts are required to describe their level of experience in crypto assets and motives to rating ICOs. The platform confirms the analysts after reviewing their credentials. The selection is relatively stringent. As of March 2020, the ICObench.com platform hosts more than 111,000 community members of which only 531 have expert status and thus the ability to provide ratings.

ICObench.com ranks the analysts based on several factors like profile completeness and analysts' consistency in contributions to the platform.[15]This provides an analogy to the widely used all-star rankings of financial analysts. We collect these rankings over time and flag whether an analyst is among the top 30 analysts, i.e., within approximately the top 5%. The dummy variable $StarAnalysts_{ij}$ equals one if analyst i is listed among the top 30 list prior to evaluating ICO j.[16]

[14]The compiled dataset is of comparable size to data used in other empirical ICO studies. For example, Benedetti and Kostovetsky (2018) use a sample of 4,003 ICO campaigns from five websites. Most information was retrieved from ICObench.com and ICOrating.com. Florysiak and Schandlbauer (2019) analyze 4,053 ICOs from ICObench.com. Deng et al. (2018) hand-collected a sample of 4,489 ICOs. Recently, Lyandres et al. (2020) cover the largest data set from the ICO universe with 7,514 ICO projects merged from various websites. Note that our sample period also covers the time after the collapse of the ICO market.

[15]The expert weight is calculated based on a profile score, a rating score, a time score, an acceptance score, and a contribution score. See https://icobench.com/faq for a detailed description.

[16]We find than star analysts were not only active on ICObench, but also among the most active users on other ICO websites, e.g., ICOholder.com.

Interestingly, many ICO analysts are involved in one or more ICO campaigns themselves.[17] This dependent network structure of analysts offers a unique setting for investigating the role of human experts in crowdfunding markets (see Section 2.4).

2.3 Ratings

We identify 531 experts on ICObench.com who rate for 2,378 ICOs. Each analyst rate an average of 29.64 ICOs, resulting in a total number of 13,834 ratings. Experts can provide a rating for three dimensions of an ICO - team, vision and product - with each dimension being scored from 1 (poor) to 5 (best). The $TotalRating$ is defined as the sum of these three individual ratings i.e., an integer in the interval $[3, 15]$.

For all ratings, we collect the date when the analyst issued the rating. The main analysis only considers ratings issued before ICO completion (or cancellation), which helps prevent look-ahead bias. However, our findings also hold when we include all ratings. Analysts have the opportunity to modify their ratings: when this happens, users can only see the updated rating score as well as two dates - the date of the first rating and the date of the update, but not the full history.[18] This paper considers the modification date as the date for the rating and flag a modified rating by analyst i to ICO j with a dummy variable $Modified_{ij}$.

When issuing a rating, the analyst gives a score and, typically, justifies the decision by writing a review. We collect all reviews and calculate linguistic measures from these texts. Based on the Loughran and McDonald (2011) dictionaries, we calculate the tone of the language, defined as the difference between positive and negative words to total words, as well as the uncertainty of the language, defined as the count of uncertain words divided by total words. We further control for the complexity of the reviews, measured by the Gunning (1952) Fog index, which is a function of the number of words per sentence (length of a sentence) and the share of complex words (words with more

[17] Note that if ICO analysts become part of the ICO project by advising the team members, they lose the ability to rate their own ICO. We found that 4 analysts rated an ICO project before becoming an advisory team member.

[18] There is a well-documented phenomenon of "walking down" forecasts in the literature on sell-side analysts. The absence of access to the rating history of analysts on ICObench.com prevents us from studying this phenomenon in the ICO context.

than two syllables) relative to total words.[19]

For some analyses, we aggregate the analyst-ICO information to the ICO level. More precisely, we count the total number of analysts who rated ICO j in $\#\,Analysts_j$. We further aggregate all analyst ratings to ICO j in the variable $TotalRating_j$ by averaging all ratings that ICO j received from all analysts that cover this ICO. Finally, we proxy the lack of consensus among analysts that provide a rating for ICO j with $AnalystDispersion_j$, defined as the standard deviation of all ratings for ICO j.

Figure 2 presents the number of ratings in a given month over time of the newly announced ICOs, the number of ratings by analysts who registered in the same month, as well as the Bitcoin price in US dollar. While the number of new ratings went up hand-in-hand with the number of ICOs to the peak of the Bitcoin price in January 2018, the number of ratings exploded thereafter and only recently has converged again to the number of announced ICOs. Figure 2 further shows that the surging demand for information about crypto assets was met by the increase in the supply of analysts.

Figure 2: Number of ratings in a month and the Bitcoin price in $
This figure presents the number of ratings in a given month over time, the number of the newly announced ICOs, the number of ratings by analysts who registered in the same month, as well as the Bitcoin price in $.

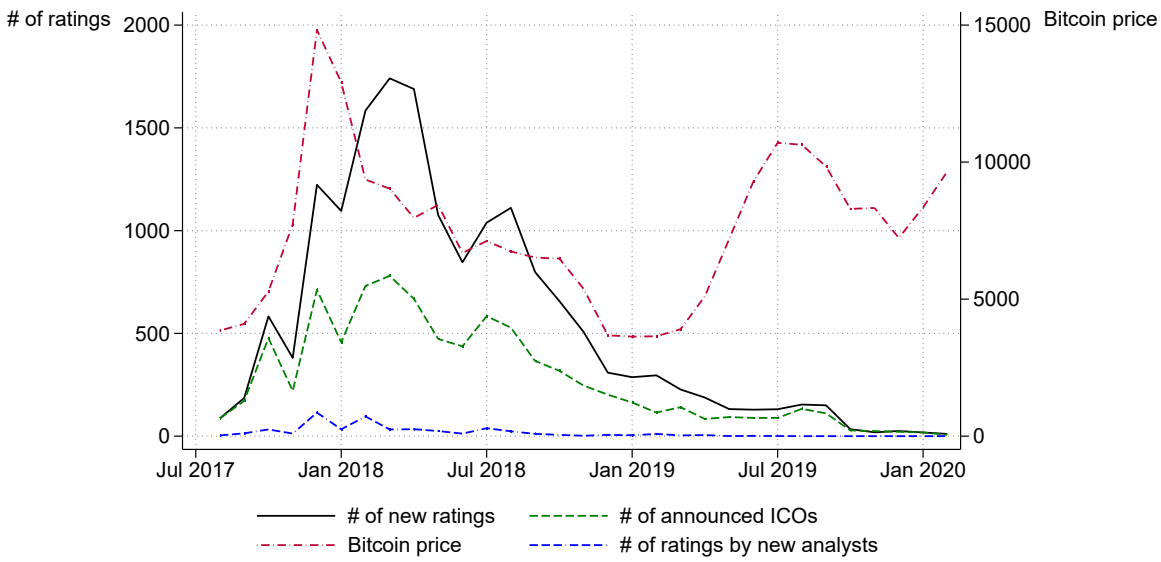
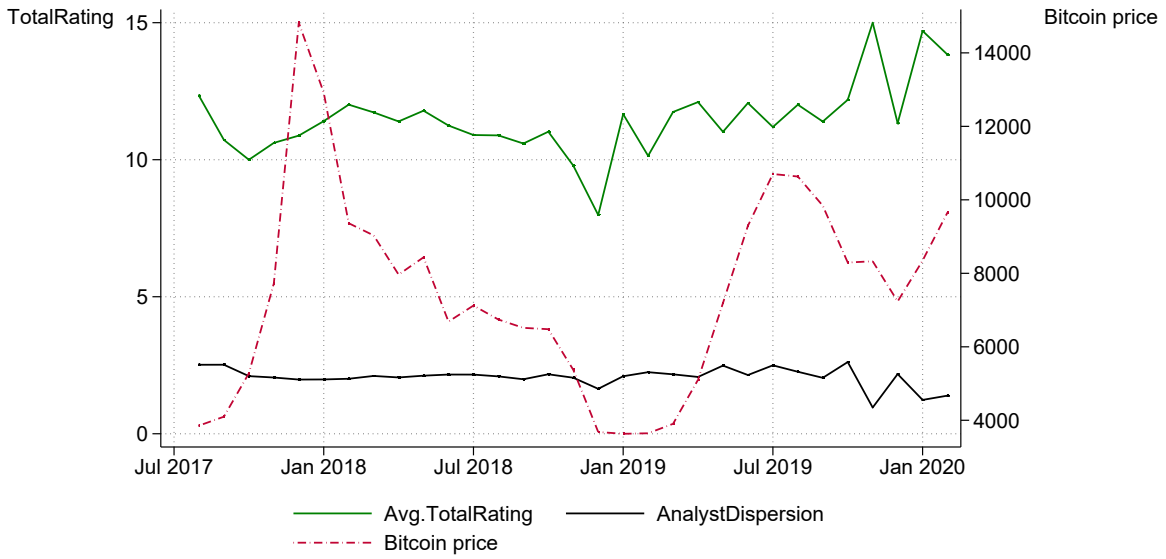


Figure 3 shows the monthly averaged $TotalRating$ as well as the analysts' rating

[19]The Gunning (1952) Fog index is defined as $Fog = 0.4 \cdot \left(\frac{TotalWords}{TotalSentences} + \frac{ComplexWords}{TotalWords}\right)$.

dispersion, measured by the standard deviation of ratings within an ICO in a given month. We observe that the average total rating of experts is overall very positive with a small decrease in the rating score around the Bitcoin price drop in 2018. Analyst dispersion remains at a relatively constant level over the sample period. It only slightly increases around the time when the Bitcoin price was low at the end of 2018, but decreases as the Bitcoin price rises at the end of 2019 again.

Figure 3: ICO analyst ratings and rating dispersion over time
This figure plots average total rating and analysts' rating dispersion (left axis) as well as the Bitcoin price in $ (right axis).



Complementing the assessment of human experts, many platforms have set up machine-generated ratings. Instead of evaluating an ICO's quality directly, these ratings based on the availability of information *about* the ICO. The idea is that more transparency indicates higher trustworthiness and quality of the ICO. For every ICO in our database, we collect the machine-generated rating by ICObench.com, which is called "Benchy". The Benchy bot provides a higher rating for higher transparency on team and event information. Moreover, Benchy uses factual data such as "presence of the social media links" and "the level of activity on them", see `https://icobench.com/faq`. Benchy re-evaluates each ICO profile at least once daily and issues a rating ranging between 1 (poor) and 5 (best). Only the most recent evaluation is observable, not the history of Benchy ratings.

While all ICOs listed on the platform ICObench.com automatically receive a machine-generated rating from the Benchy bot, 2,378 out of 5,384 ICOs listed on this website were also rated by ICO analysts. On average, the ICOs with(out) an analyst rating have a Benchy rating of 3.2 (2.7) out of 5.

2.4 Reciprocal ratings

A specific feature of the market is that ICO analysts also participate in ICOs. We identify those experts that are involved in one or several ICO projects by collecting each expert's self-description of experiences and achievements from the 'About'-section of their profile page on ICObench.com. Table 1 shows the distribution of ICO projects among analysts. Out of the 531 experts in our sample, 329 have been involved in at least one ICO, with some analysts being very active in launching ICOs.

Table 1: ICO affiliation of analysts from the platform ICObench.com
This table tabulates the distribution of ICO projects among analysts. The total number of analysts in our sample is 531. The list of associated ICOs for each analyst is available on their webpage in ICObench.com

Number of associated ICOs	Count
0	230
1	127
2	52
3	29
4	24
5	15
6	9
7	13
8	3
9	8
>=10	49
Total number of analysts	531

We use this information to flag whether a rating of analyst i to ICO j is a response to a rating that analyst i received from any team member of coin j at any point in time, and generate the indicator variable $ReciprocalRating_{ij}$ as follows:

$$ReciprocalRating_{ij} = \begin{cases} 1, & \exists\, TotalRating^{ij}_{j'i'} \text{ where } \mathbf{i} \in \Omega_{i'},\, j' \in \Omega_{\mathbf{j}} \\ 0, & \text{else} \end{cases}$$

where $\Omega_{\mathbf{j}}$ refers to the set of all team members of the ICO j. Table 2 represents a hypothetical illustration of how we define this variable. $ReciprocalRating_{ij}$ thus flags whether any member of ICO j has provided a rating of any ICOs that expert i is associated with. Reciprocal ratings are not directly flagged by ICObench.com, but users can easily obtain the information given the available links to the analyst's associated ICOs and the timeline of the ratings provided on ICObench.com.

Whenever $ReciprocalRating_{ij}$ indicates reciprocity, we additionally identify the level of the reciprocal rating, i.e., the $TotalRating$, as well as the three components $TeamRating$, $VisionRating$ and $ProductRating$ by any member of ICO j to the ICO with which expert i is associated. The level of the reciprocal rating is labeled $ReceivedTotalRating_{ij}$.

2.5 ICO outcome variables

We generate a dummy variable $Success$, which takes the value of 1 if the ICO-related coin successfully completes the offering and receives funding. For these ICOs, we collect information on the dollar amount raised during the campaign from ICObench.com, ICOmarks.com, tokendata.io and ICOdata.io. Tokens were classified as failed when we could not find the amount raised nor any success information on the above-mentioned web pages. In total, we identify 1,932 successful ICOs.

Figure 4 shows the time trend of successful ICOs. ICOs became popular at the beginning of 2017. While only 29 ICO tokens were on sale before then, the number increased to 1,127 ICOs within one year with around 94 offerings per month and a 53% success rate. The market peaked in 2018, with 3,360 ICOs in total and a success rate of 33%. In 2019, around 64 ICO offerings were sold per month, of which 25% were successful on average. Thus, flow of ICOs continues, albeit at a lower level, even after the sharp decline of cryptocurrency prices and the corresponding decline of enthusiasm towards ICOs.

Table 2: ICO analyst networks: An example
This table presents a hypothetical example of our data set. In Panel A, we show the team members of the three ICOs in the sample, namely, "A-Tokens" where Adam and Ashley are among the team members, "Bethereum" where the team includes Barbara and Benjamin, and "CryptoPay" with Cora and Chris in the team. In Panel B, we outline a hypothetical rating history. For example, in October 2017, Ashley (member of A-Tokens) provides a rating of 12 for Bethereum. In December 2017, Chris (member of CryptoPay) provides a rating of 15 for Bethereum. For this rating, we set *ReciprocalRating* equal to one because, in a month before that, in November 2017, Benjamin (member of Bethereum) gave a rating of 14 for CryptoPay, with which Chris is affiliated. Hence, we consider the rating given in December 2017 as a reply to the rating received in November 2017.

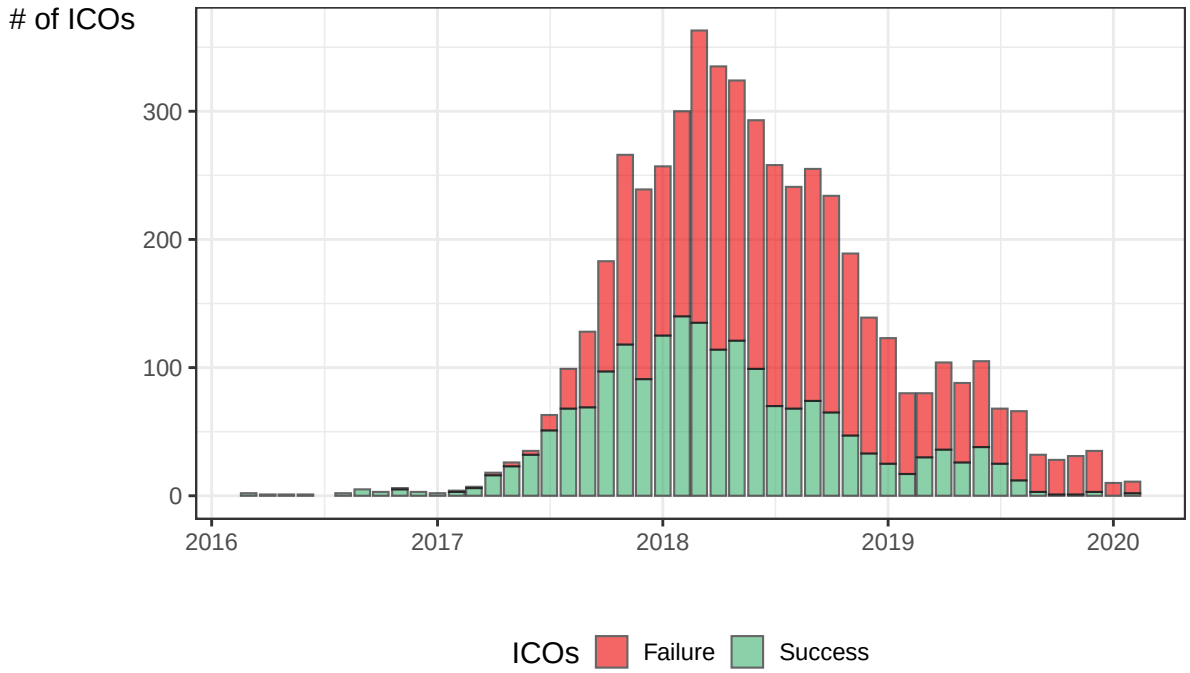
A. ICOs and members:

A-Tokens	Bethereum	CryptoPay	...
Adam	Barbara	Cora	
Ashley	Benjamin	Chris	

B. Ratings:

Date	Analyst	provides a rating for:	TotalRating	ReciprocalRating	ReceivedTotalRating if $ReciprocalRating = 1$
1)Oct 2017	Ashley	Bethereum	12	0	-
2)Nov 2017	Benjamin	CryptoPay	14	0	-
3)Dec 2017	Chris	Bethereum	15	1	14
4)Jan 2018	Adam	CryptoPay	9	0	-
...					

Figure 4: Successful and unsuccessful ICOs over time
The figure shows the number of ICOs over time, distinguishing between successful and failed ICOs. An ICO is labeled successful if the related coin successfully completes the offering and receives funding. In total, we identify 5,384 ICOs of which 1,932 ICOs succeeded.



In addition to measures of success, we collect information about scams, i.e., ICOs that were launched with the intention to defraud investors. To do so, we use the marker 'Scam or Other Issues' for dead coins listed on Coinopsy.com, as well as information from Deadcoins.com, a message board where users post about scams. Some of these ICOs can also be found in the U.S. Securities and Exchange Commission (SEC) press releases, especially when they fine ICO companies for fraudulent practices.[20] With this (likely conservative) method, 234 ICOs were flagged as scams in our data.

2.6 Forecast errors

Combining the ICO success variable and the analyst rating score allows us to construct, an ex-post forecast error measure for each rating. As the outcome of an ICO is either success or failure, we define the forecast error of a rating as the distance to the highest (lowest) possible rating in case of success (failure):

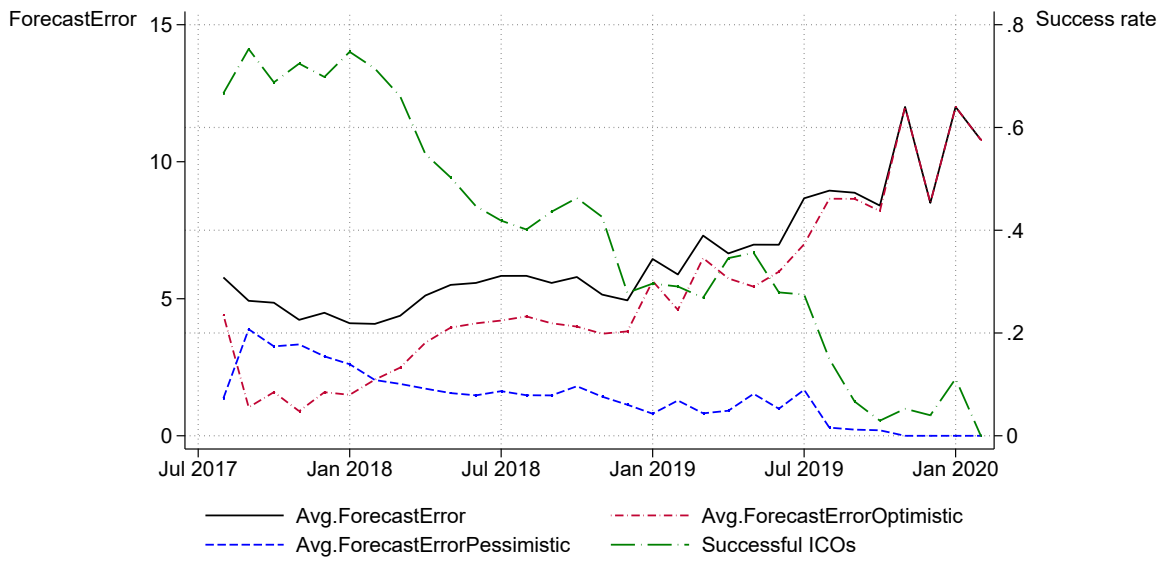
[20]See, e.g., `https://www.sec.gov/news/press-release/2019-259`.

$$ForecastError_{ij} = \begin{cases} 15 - TotalRating_{ij}, & \text{if ICO succeeded} \\ TotalRating_{ij} - 3, & \text{if ICO failed} \end{cases}$$

If an analyst gives a successful ICO a total rating of 15, their rating was fully precise, resulting in a $ForecastError$ measure of 0. If the ICO had failed, however, the forecast error of this rating would flag a 12, such that $ForecastError \in \{0, ..., 12\}$.[21]

Figure 5 shows the monthly averaged $TotalRating$, the average forecast error and the number of successful ICOs (as a share of total ICOs) over time. In addition, we plot the monthly average forecast error separately for ratings where analysts were too optimistic and too pessimistic, respectively.[22] Interestingly, ratings become less precise over time. This is driven by overoptimistic analysts.

Figure 5: Forecast error over time
This figure shows the average forecast error and the number of successful ICOs (as a share of total ICOs) in a given month. The average forecast error is further split into $ForecastErrorOptimistic_i$ and $ForecastErrorPessimistic_i$ to capture the monthly averaged forecast error separately for the ratings when the analyst was too optimistic and pessimistic, respectively.



In our regression analysis, we use an analyst-specific measure of the forecast error

[21] While this $ForecastError$ measure is not immediately available for investors on ICObench.com, one can easily view the entire timeline of an analysts' ratings with a link to detailed information on the rated ICO.

[22] An analyst is too optimistic (pessimistic) if her rating of a failing (successful) ICO was larger (smaller) than 3 (15). Thus, we calculate the monthly average optimistic (pessimistic) forecast error over the lower (upper) cases of the $ForecastError_{ij}$ definition.

that takes the entire history of an analyst's ICO-specific $ForecastError_{ij}$ into account. We recursively average the $ForecastError_{ij}$ of analyst i over all of their issued ratings up to ICO j using an expanding window. We denote this variable $ForecastError_i^j$.

2.7 ICO characteristics

For every ICO in the sample, we collect data on the campaign characteristics that have been found in the literature to indicate the perceived quality of an ICO by investors (Amsden and Schweizer, 2018; Burns and Moro, 2018; Howell et al., 2019; Roosenboom et al., 2020). For many characteristics, we generate binary indicators that flag whether an ICO exhibits the respective feature. The dummy variable $Presale$ equals one if an ICO offers coins at pre-sale stage and zero otherwise. The $Bonus$ and $Bounty$ dummies equal 1 if there were discounts on the token sale or incentives to boost social media presence, respectively. The dummy MVP flags the availability of a minimum viable product or whether a product prototype was in place. The dummy KYC equals one if investors need to validate their identity by signing up to a whitelist to get access to the token sale.[23] The dummy IEO indicates the use of a centralized token launch platform provided by a cryptocurrency exchange. The $RetentionRatio$ is the retained share of total token supply (in percent); it captures the "skin in the game" of ICO team members. We collect the information whether the ICO has its own webpage on the forum Bitcointalk.org and/or Facebook to discuss and promote their project idea. Finally, we include $LengthWhitePaper$, the natural logarithm of (1 + total words of the white paper) as a proxy for the informative value of the ICO whitepaper. We set this variable to zero if no white paper could be found. In addition to these campaign-specific characteristics, we collect the year-month information of the date when the ICO was launched.

[23]Note that ICObench.com provides information on two different KYC procedures. One KYC symbol means the identity verification of ICObench.com profiles, while the second flags the identification and registration process of investors to receive access to the token sale. We use the second KYC throughout the paper.

2.8 Descriptive statistics

Table 3 shows descriptive statistics of the key variables of rating and ICO characteristics. All variables are defined in Table A1.

In our sample, the analyst's average $TotalRating$ is 11, where the $ProductRating$ is slightly more pessimistic than the other two dimensions $TeamRating$ and $VisionRating$. Of all ratings, 12% are flagged $ReciprocalRating$, which are found to be somewhat more positive with an average $ReceivedTotalRating_{ij}$ of 13. On the ICO dimension, we observe a success rate of 36%. In terms of dollar amount raised, EOS, Telegram, and Bitfinex were the most successful ICOs in our sample. The scam rate is 4.3%. Each ICO is covered by 2.6 analysts, on average, and 44% of all ICOs are covered by at least one analyst. The ICOs for Sharpay (94), Truegame (82) and WePower (64) had the largest number of analysts covering them.

Table 3: Descriptive statistics
This table shows descriptive statistics of the variables used in the analysis. The variables are sorted alphabetically within each panel. The sample consists of 5,384 ICOs listed in ICObench.com, of which, 2,378 received 13,834 ratings in total. All variables are defined in Table A1.

	N	Min	P25	Mean	P50	P75	Max	Std. Dev.
A. Rating characteristics								
$ForecastError_i^j$	12,460	0	4.1	4.8	4.8	5.6	12	1.5
$Modified_{ij}$	13,834	0	0	.13	0	0	1	.33
$OrderRank_{ij}$	7,639	1	6	14	11	18	94	12
$ProductRating_{ij}$	13,834	1	3	3.6	4	5	5	1.1
$ReceivedProductRating_{ij}$	1,754	1	4	4.1	4	5	5	.74
$ReceivedTeamRating_{ij}$	1,754	1	4	4.3	4	5	5	.67
$ReceivedTotalRating_{ij}$	1,754	3	12	13	12	14	15	1.8
$ReceivedVisionRating_{ij}$	1,754	1	4	4.2	4	5	5	.7
$ReciprocalRating_{ij}$	13,834	0	0	.13	0	0	1	.33
$ReviewLength_{ij}$	9,165	1.1	3.4	3.8	3.9	4.3	7.9	.97
$ReviewTone_{ij}$	9,165	-.75	-.043	-.014	-.0086	.018	.67	.075
$StarAnalyst_{ij}$	13,834	0	0	.27	0	1	1	.44
$TeamRating_{ij}$	13,834	1	3	3.9	4	5	5	1.1
$TotalRating_{ij}$	13,834	3	10	11	12	14	15	3
$VisionRating_{ij}$	13,834	1	3	3.9	4	5	5	1.1
B. ICO characteristics								
$AmountRaised_j$	5,384	0	0	5.4	0	14	22	7.3
$AnalystDispersion_j$	1,638	0	1.2	2	1.9	2.6	8.5	1.3
$Benchy_j$	5,384	.1	2.4	2.9	2.9	3.5	5	.75
$Bitcointalk_j$	5,384	0	0	.57	1	1	1	.5
$Bonus_j$	5,384	0	0	.14	0	0	1	.35
$Bounty_j$	5,384	0	0	.28	0	1	1	.45
$Disagreement_j$	2,378	0	0	.17	0	0	1	.38
$Facebook_j$	5,384	0	1	.78	1	1	1	.41
$ForecastError_j$	2,322	0	4.3	5	4.9	5.7	11	1.1
IEO_j	5,384	0	0	.051	0	0	1	.22
KYC_j	5,384	0	0	.48	0	1	1	.5
$LengthWhitePaper_j$	5,384	0	0	1.2	0	0	11	3.1
MVP_j	5,384	0	0	.19	0	0	1	.4
$Presale_j$	5,384	0	0	.52	1	1	1	.5
$PreviousRatings_j$	2,322	3	11	11	11	12	15	1.4
$ReciprocalRatingShare_j$	2,378	0	0	.072	0	0	1	.19
$RetentionRatio_j$	4,238	0	30	46	45	60	100	21
$ReviewComplexity_j$	1,883	4.6	11	12	12	14	59	3.2
$ReviewLength_j$	1,883	1.1	3.6	4	4	4.5	7.3	.81
$ReviewToneDispersion_j$	1,240	0	.03	.057	.046	.073	.55	.046
$ReviewTone_j$	1,883	-.67	-.04	-.02	-.015	.003	.29	.056
$ReviewUncertainty_j$	1,883	0	0	.016	.011	.021	.33	.022
$Scam_j$	5,384	0	0	.043	0	0	1	.2
$StarAnalysts_j$	2,378	0	0	.31	.22	.5	1	.34
$Success_j$	5,384	0	0	.36	0	1	1	.48
$\#\,Analysts_j$	5,384	0	0	2.6	0	2	94	6.1

3 Empirical Analysis

Section 3.1 analyzes the determinants of an analysts' rating (both quantitative and qualitative) of an ICO. Section 3.2 in turn considers whether investors consider differences in the reliability of analyst ratings.

3.1 What determines analyst ratings?

3.1.1 Baseline results

We model the rating of analyst i for ICO j as a function of analyst characteristics, as indicated in the following equation:

$$\begin{aligned} Rating_{ij} = \beta_0 + \beta_1 \cdot Benchy_j + \beta_2 \cdot StarAnalysts_{ij} + \beta_3 \cdot ForecastError_i^j \\ + \beta_4 \cdot Modified_{ij} + \beta_5 \cdot X_j + Month_{ij} + \alpha_j + \epsilon_{ij}, \end{aligned} \tag{1}$$

$Rating_{ij}$ denotes the respective rating score that analyst i gives to ICO j for the different rating categories team, vision and product (on a scale from 1–5), as well as the $TotalRating_{ij}$ score as the sum of the three categories (on a scale from 3–15). The vector X_j contains the ICO campaign characteristics as described in Subsection 2.7. Time trends of ratings are absorbed by $Month_{ij}$ dummies. α_j denotes ICO fixed effects. We allow for a potential serial correlation of ratings within each analyst and within each ICO and employ two-way clustering of standard errors (Cameron et al., 2011) at the analyst and ICO dimension.

Table 4 summarizes the results of this analysis. Column (1) shows that machine-generated and human expert ratings point in the same direction qualitatively, i.e., ICOs with higher machine-generated ratings receive a higher rating score by human analysts on average. Moreover, in the cross-section of analysts, columns (2) and (3), we find a statistically significant negative coefficient on $ForecastError_i^j$, implying that analysts with historically higher forecast errors give on average lower ratings. The negative relationship between the past forecast errors and the rating remains also in the within ICO estimation, as column (4) shows. Analysts listed within the top 30 analysts on ICObench.com are

Table 4: Rating determinants
This table presents linear regression results for Equation 1. The dependent variable is the total rating score that an analyst gave an ICO. All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered at the ICO and analyst level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Dep. variable:	$TotalRating_{ij}$				
	(1)	(2)	(3)	(4)	(5)
$Benchy_j$	1.518***	1.598***		0.996***	
	(10.43)	(11.35)		(7.01)	
$StarAnalyst_{ij}$		-0.720***	-0.699***	-0.701***	-0.454**
		(-3.58)	(-3.41)	(-3.47)	(-2.47)
$ForecastError_i^j$		-0.158***	-0.156***	-0.165***	-0.081**
		(-3.61)	(-3.70)	(-4.12)	(-2.31)
MVP_j			0.264*	-0.084	
			(1.79)	(-0.58)	
IEO_j			0.850***	0.623***	
			(3.98)	(2.71)	
$Presale_j$			0.311**	0.242**	
			(2.40)	(1.99)	
KYC_j			0.912***	0.548***	
			(5.36)	(3.42)	
$Bounty_j$			0.258**	0.179	
			(1.98)	(1.45)	
$Bonus_j$			0.145	0.148	
			(1.20)	(1.29)	
$RetentionRatio_j$			0.006**	0.005*	
			(2.04)	(1.72)	
$LengthWhitePaper_j$			0.012	-0.007	
			(0.69)	(-0.39)	
$Bitcointalk_j$			0.070	-0.026	
			(0.36)	(-0.14)	
$Facebook_j$			0.312	0.302	
			(1.26)	(1.24)	
Observations	13834	12460	11257	11257	11698
R^2	0.121	0.146	0.067	0.099	0.528
MonthRating Dummies	No	No	No	No	Yes
ICO FE	No	No	No	No	Yes

more critical and issue lower ratings on average.

Furthermore, in line with the literature, the coefficients of the control variables suggest that analysts consider the characteristics of the underlying ICO (Deng et al., 2018; Bourveau et al., 2019; Roosenboom et al., 2020). In general, we find that ICOs with a pre-sale event, with a KYC feature and an IEO feature receive better ratings. Moreover, analysts perceive it as a good signal when founders retain a higher share of the tokens themselves.

3.1.2 Reciprocal ratings

When ICO analysts issue new ratings, do these ratings depend on ratings that their own affiliated ICOs previously received? To answer this question, we run regressions, as specified in the following equation:

$$\begin{aligned} Rating_{ij} = \beta_0 + \beta_1 \cdot ReciprocalRating_{ij} + Analyst \times Month_{\ ij} \\ + ICO \times Month_{\ ij} + \epsilon_{ij}, \end{aligned} \tag{2}$$

where $ReciprocalRating_{ij}$ indicates a dummy that flags whether analyst i received a rating from a team member of ICO j. We include $Analyst \times Month$ and $ICO \times Month$ dummies to exploit only the analyst and ICO pairing within the month of the rating. These fixed effects detect the variation previously established in Table 4, and they help to rule out that the results were driven by a non-random match between founders of good ICOs that also serve as analysts. Comparing different assessments of the same analyst and for the same ICO allows also to differentiate whether analysts behave in a deliberately optimistic biased manner or whether the optimistic assessment is due to the high difficulty of forecasting tasks. For reciprocal ratings, we also analyze whether the level of the prior rating predicts the level of the reciprocal rating.[24] We again employ two-way clustering at the analyst and ICO dimension.

Table 5 shows that ratings indeed contain a reciprocal element. Column (1) indi-

[24]We use $Analyst \times Quarter$ and $ICO \times Quarter$ dummies for the analysis because of the restricted sample size.

Table 5: Reciprocal ratings

This table presents linear regression results for Equation 2. The dependent variable is the total rating score that an analyst gave an ICO. In columns (1), (3), (5) and (7), regressions include all the ratings in the sample. In columns (2), (4), (6) and (8), we restrict the sample to the reciprocal ratings ($ReciprocalRating = 1$). All specifications include *Analyst* and *ICO* fixed effects multiplied by dummies for the time of the rating (i.e., $Analyst \times Month$ and $ICO \times Month$ fixed effects in odd columns and $Analyst \times Quarter$ and $ICO \times Quarter$ fixed effects in even columns). All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered at the ICO and analyst level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Dep. variable:	$TotalRating_{ij}$		$TeamRating_{ij}$		$VisionRating_{ij}$		$ProductRating_{ij}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$ReciprocalRating_{ij}$	0.252**		0.062*		0.074*		0.117***	
	(2.47)		(1.80)		(1.76)		(2.85)	
$ReceivedTotalRating_{ij}$		0.080*						
		(1.74)						
$ReceivedTeamRating_{ij}$				0.117***				
				(3.13)				
$ReceivedVisionRating_{ij}$						0.065		
						(1.14)		
$ReceivedProductRating_{ij}$								0.000
								(0.01)
$Modified_{ij}$		-1.114***		-0.381***		-0.298**		-0.427***
		(-3.54)		(-3.02)		(-2.53)		(-3.46)
Observations	10354	1302	10354	1302	10354	1302	10354	1302
R^2	0.757	0.682	0.717	0.621	0.692	0.666	0.708	0.647
$Analyst \times Time_{ij}$ Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$ICO \times Time_{ij}$ Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

cates a positive association between the total rating an analyst gives to an ICO and the $ReciprocalRating_{ij}$ dummy. More specifically, the total rating score is around 0.25 points higher when the analyst is in a position to respond to a prior rating. Additionally, within the sample of reciprocal ratings, column (2) shows that ratings are more positive, the higher the prior received rating was. In other words, analysts reciprocate positive ratings. For example, column (2) shows that each one-unit (one standard deviation) increase for the previous rating leads an analyst to issue a total rating around 0.08 (0.15) higher. Note that this result holds within ICO-time and analyst-time combinations, i.e., comparing ratings by two (otherwise) identical analysts, where one analyst previously received a rating by a team member of coin j and the other one did not. This reciprocal rating behavior is similar to the *quid pro quo* between hedge funds and sell-side equity analysts described in Klein et al. (2019).

Columns (3)–(8) analyze the three different rating categories team, vision and product separately. The coefficient for the $ReciprocalRating_{ij}$ dummy is positive and significant for all three categories, indicating that, on average, analysts give a higher rating for the team, vision and product of ICO j if any team member of ICO j rated them. However, the actual score is significant only for the team dimension. That is, an analyst rates the team component of ICO j higher if she received a more favorable team rating from a team member of ICO j. By contrast the scores for vision and product are not significant. These findings are intuitive, as the team category constitutes a "soft factor". The results also indicate a relatively personal nature of the reciprocity.

3.1.3 Linguistic characteristics of rating reviews

When issuing ratings, analysts often justify the rating scores with written reviews. We next analyze whether more optimistic ratings are special in terms of the linguistic nature of the written review. The literature on earnings conference calls uses the number of words spoken by analysts as a proxy for the question difficulty, so analysts who ask lengthier questions are regarded as more critical (Merkley et al., 2017). Correspondingly, we investigate whether the rating score correlates with the length of the written text or

with the linguistic tone of the review. Moreover, we investigate whether the relationship between the rating score and the review length and tone differs for reciprocal versus non-reciprocal ratings. This idea follows Cohen et al. (2020), who document that biased analysts ask easier questions. We run regressions for the overall sample as well as for reciprocal and non-reciprocal ratings separately as specified in the following equation:

$$\begin{aligned} Linguistic\,Measure_{ij} = \beta_0 + \beta_1 \cdot TotalRating_{ij} + Analyst \times Time_{\ ij} \\ + ICO \times Time_{\ ij} + \epsilon_{ij}, \end{aligned} \tag{3}$$

where $Linguistic\,Measure_{ij}$ indicates interchangeably the length of the rating review measured by the (natural logarithm of the) number of words and the ratio of positive words minus negative words to total words in the review. As before, we employ two-way clustering by analysts and ICOs.

Table 6 shows the results. In Panel A column (1), we find a negative relationship between the rating score and the length of the review, suggesting that more negative ratings come with a more detailed explanation. In column (2), we include $Analyst \times Month$ and $ICO \times Month$ dummies to rule out that the results are driven by a non-random match between analyst characteristics (e.g. mood) and the quality of the rated ICO. For the review tone in Panel B, columns (1) and (2), we find that analysts use more positive terminology when reviewing an ICO that they score higher.

When investigating whether the relationship between the rating score and the review length and tone differs for reciprocal versus non-reciprocal ratings, we find that lower rating scores are justified with even lengthier reviews for reciprocal ratings, with a statistically significant difference to the coefficient for non-reciprocal ratings.[25] The relationship between review tone and rating score does not differ noticeably between reciprocal and non-reciprocal ratings.

[25]Because of the restricted sample size for the sample of reciprocal ratings, we use $Analyst \times Quarter$ and $ICO \times Quarter$ dummies for these analyses.

Table 6: Linguistic nature of rating reviews
This table presents linear regression results for Equation 3. The dependent variable in Panel A is *ReviewLength*, defined as the natural logarithm of the total number of words in a review, and in Panel B *ReviewTone*, defined as the ratio of positive words minus negative words to total words in the review. We restrict the sample to reciprocal ratings ($ReciprocalRating = 1$) in column (3) and to non-reciprocal ratings ($ReciprocalRating = 0$) in column (4). We include *Analyst* and *ICO* fixed effects multiplied by dummies for the month of ratings (i.e., $Analyst \times Month$ and $ICO \times Month$ fixed effects) in column (2). As in Table 5, we can only include the interaction of ICO (analyst) and quarter dummies when restricting the sample to reciprocal ratings in column (3), i.e., $Analyst \times Quarter$ and $ICO \times Quarter$ fixed effects. In order to compare the coefficients for reciprocal and non-reciprocal ratings, we also include $Analyst \times Quarter$ and $ICO \times Quarter$ fixed effects in column (4). All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered at the ICO and analyst level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	Panel A			
Dep. Variable:	$ReviewLength_{ij}$			
	(1)	(2)	(3)	(4)
$TotalRating_{ij}$	-0.056***	-0.041***	-0.090***	-0.036***
	(-5.42)	(-5.41)	(-5.10)	(-5.37)
Observations	9165	6206	866	6119
R^2	0.033	0.825	0.800	0.786
$Analyst \times Time_{ij}$ *Dummies*	No	Yes	Yes	Yes
$ICO \times Time_{ij}$ *Dummies*	No	Yes	Yes	Yes

	Panel B			
Dep. Variable:	$ReviewTone_{ij}$			
	(1)	(2)	(3)	(4)
$TotalRating_{ij}$	0.006***	0.006***	0.009***	0.005***
	(10.09)	(7.51)	(3.34)	(8.17)
Observations	9165	6206	866	6119
R^2	0.062	0.537	0.522	0.477
$Analyst \times Time_{ij}$ *Dummies*	No	Yes	Yes	Yes
$ICO \times Time_{ij}$ *Dummies*	No	Yes	Yes	Yes

3.1.4 Order of ratings

The literature on security analysts has documented herding behavior among analysts and shows that heir buy or sell recommendations have a significant positive influence on the next analysts' recommendations (Welch, 2000). Thus, reciprocal analysts' scores may impact investors as well as other analysts when they cover the ICO at an early stage. We therefore analyze whether analysts provide reciprocal ratings faster and move earlier for ICOs where they issue more positive ratings. We generate a variable that counts the rank of rating arrival per ICO j from analyst i, i.e., whether analyst i was the first, second, third, ... last analyst that rated for ICO j. We relate the order of the rating coverage to the $ReciprocalRating_{ij}$ dummy, as indicated in the following equation:

$$OrderRank_{ij} = \beta_0 + \beta_1 \cdot TotalRating_{ij} + \beta_2 \cdot ReciprocalRating_{ij} + \beta_3 \cdot StarAnalysts_j + \beta_4 \cdot ForecastError_i^j + Month_{ij} + \alpha_i + \alpha_j + \epsilon_{ij}. \quad (4)$$

We again absorb any ICO and analyst characteristics with ICO and analyst fixed effects and control for time trends by $Month_{ij}$. As before, we use two-way clustered standard errors at the analyst and ICO dimension.

The results are shown in Table 7. In line with the literature on the analyst coverage of stocks (Demiroglu and Ryngaert, 2010), we first find that analysts who give favorable ratings tend to issue their rating early. Second, star analysts tend to move first and rate the same ICO earlier than their less experienced peers. Third, reciprocal ratings tend to be issued early. In particular, in the chronological sequence of ratings given to an ICO j, a reciprocal analyst appears to issue her rating on average 1.3 positions earlier than a non-reciprocal analyst.

Table 7: Order of rating issuance
This table presents linear regression results for Equation 4. The dependent variable is the order rank of the rating for an ICO. A lower value of the variable indicates that analyst i issued the rating for ICO j earlier. All specifications include month dummies of the analyst rating. The sample is restricted to ICOs with more than ten ratings. All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered at the ICO and analyst level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Dep. variable:	$OrderRank_{ij}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$TotalRating_{ij}$	-0.104		-0.085	-0.182***		-0.177***
	(-1.65)		(-1.36)	(-2.82)		(-2.75)
$ReciprocalRating_{ij}$		-1.764***	-1.725***		-1.249**	-1.213**
		(-3.17)	(-3.08)		(-2.56)	(-2.48)
$StarAnalyst_{ij}$	-1.109***	-0.855***	-0.886***			
	(-3.00)	(-2.60)	(-2.69)			
$ForecastError_i^j$	-0.028	-0.027	-0.036	-0.130	-0.133	-0.122
	(-0.32)	(-0.31)	(-0.42)	(-1.18)	(-1.22)	(-1.12)
Observations	6829	6829	6829	6767	6767	6767
R^2	0.672	0.674	0.674	0.709	0.709	0.709
MonthRating Dummies	Yes	Yes	Yes	Yes	Yes	Yes
ICO FE	Yes	Yes	Yes	Yes	Yes	Yes
Analyst FE	No	No	No	Yes	Yes	Yes

3.2 Are ICOs with higher ratings more successful, and which ICOs fail despite high ratings?

So far, we have established several important determinants of ICO analyst ratings, with analyst-specific factors, such as experience, prior forecasting ability and reciprocal status, playing a major role in addition to objective differences among the ICOs. Now, we investigate whether ICOs with higher ratings are indeed more successful. First, we establish baseline results for (unconditional) ICO success, but our main interest is in explaining when investors deviate from the ICO analyst consensus, that is, the ICO success probability conditional on an extreme positive (or negative) rating outcome. We also consider whether the factors that explain such deviations predict scams.

3.2.1 Ratings and ICO success

Table 8 presents descriptive statistics for the relationship between ratings and ICO success. Panel A indicates that ICOs are more likely to be successful when it motivates analysts to rate it. In Panel B, we tabulate success statistics for groups of the quantitative rating score. The probability of receiving funding and the average dollar amount raised is higher for ICOs with more positive ratings, though the relationship is not strictly monotonic.

While these results highlight that successful ICOs have, on average, higher ratings, there are numerous cases in which ICOs were either unsuccessful despite positive ratings or successful despite negative ratings. To quantify this phenomenon, we define for each ICO j a $Disagreement_j$ dummy as a conditional success outcome. More precisely, the $Disagreement_j$ dummy equals one if (i) the average $TotalRating_j$ of an ICO is greater than or equal to 13 but the ICO is unsuccessful, or (ii) the average total rating is less than or equal to 5 and the ICO is successful. In our sample, this $Disagreement_j$ dummy is one in 413 of 2,378 rated ICOs (17%).[26] While the unconditional failure rate of ICOs is about 64%, 53.6% of ICOs with an average rating in the top quartile fail.

[26]Note that disagreement most often concerns the case where the rating is high but the ICO fails. We focus mainly on these cases when analyzing disagreement. There are only very few cases of successful ICOs with an average poor rating (N=44).

These mismatches between ratings and ICO success do not occur randomly. To illustrate, in Panel C, we tabulate the disagreement dummy against the occurrence of reciprocal ratings. We observe that the ICO outcome does not correspond to what one would expect given the ratings level if reciprocal analysts cover the ICO. ICOs that receive very favorable recommendations fail much more frequently if the reciprocal rating share is positive than if none of the ratings is reciprocal.

Table 8: Ratings and ICO success: Descriptive evidence
This table presents descriptive statistics for the relationship between ratings and ICO success. Panel A shows the success of ICOs that human analysts did or did not cover. Panel B links ICO success to the quantitative rating score. Panel C shows investor disagreement for ICOs with or without any reciprocal rating.

Panel A

Analyst Coverage	Total #	Funded #	Funded in %	AmountRaised avg. in $	Ln(AmountRaised) avg. in $
No	3,006	899	29.91	18,600,000	4.40
Yes	2,378	1,033	43.44	12,700,000	6.60
Total	5,384	1,932	35.88	15,500,000	5.38

Panel B

TotalRating Score	Total #	Funded #	Funded in %	AmountRaised avg. in $	Ln(AmountRaised) avg. in $
3	217	52	23.96	20,600,000	3.38
4–6	253	55	21.74	6,006,378	3.09
7–9	487	175	35.93	14,200,000	5.54
10–12	1,076	549	51.02	12,100,000	7.77
13–15	767	405	52.80	31,200,000	8.27

Panel C

Reciprocal Rating	Total #	Disagreement #	Disagreement in %	Disagreement with Avg. Rating ≥ 13 #	Disagreement with Avg. Rating ≥ 13 in %
Yes*	415	97	23.37	96	49.23
No**	1,963	316	16.10	272	23.65

*$ReciprocalRatingShare_j > 0$
** $ReciprocalRatingShare_j = 0$

In order to formally analyze ICO success in a regression framework, we first explain the unconditional success of ICO j by characteristics of participating analysts and other

variables in a logit regression:

$$\begin{aligned} Success_j = \beta_0 + \beta_1 \cdot TotalRating_j + \beta_2 \cdot \#Analysts_j + \beta_3 \cdot ReciprocalRatingShare_j \\ + \beta_4 \cdot StarAnalysts_j + \beta_5 \cdot PreviousRatings_j + \beta_6 \cdot AnalystDispersion_j \\ + \beta_7 \cdot Benchy_j + \beta_8 \cdot Z_j + \beta_9 \cdot X_j + Month_j + \epsilon_j. \end{aligned} \tag{5}$$

$Success_j$ indicates the success dummy as described in Section 2.[27] X_j again represents the controls as in Equation 1. Additionally, we control for linguistic measures with Z_j, which contains the average tone, uncertainty and complexity as well as the length of all rating reviews written about ICO j. We further include a dummy for each month of the sample, $Month_j$ to absorb time trends common to all ICOs. Note that the sample size drops when including the $AnalystDispersion_j$ measure, as it is only defined for ICOs with at least two ratings.

In addition to the unconditional success of ICOs, we investigate the success conditional on having received very high or very low ratings. Thus, we run the following logit regression on the ICO level:

$$\begin{aligned} Disagreement_j = \beta_0 + \beta_1 \cdot \#Analysts_j + \beta_2 \cdot ReciprocalRatingShare_j \\ + \beta_3 \cdot StarAnalysts_j + \beta_4 \cdot PreviousRatings_j \\ + \beta_5 \cdot AnalystDispersion_j + \beta_6 \cdot Benchy_j \\ + \beta_7 \cdot Z_j + \beta_8 \cdot X_j + Month_j + \epsilon_j, \end{aligned} \tag{6}$$

where X_j is the same set of controls as in Equation 1 and $Month_j$ dummies absorb common time trends. We again control for some linguistic measures with Z_j, which contains the average tone, uncertainty, complexity and length of all rating reviews written on ICO j.

As a baseline result, the regressions confirm that ratings are predictive on average.

[27] In the Appendix, we alternatively use the dollar amount raised during the ICO as a measure of success.

Table 9: Ratings and ICO success
This table presents marginal effects of logit regressions for Equation 5. The dependent variable is the *Success* dummy. The controls for which coefficients are not shown for space reasons include the retention ratio and dummies for pre-sale, bonus/bounty options, KYC, Bitcointalk, and Facebook. All specifications include month dummies. All variables are defined in Table A1. t-statistics based on robust standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Dep. variable:	$Success_j$		
	(1)	(2)	(3)
$TotalRating_j$	0.110***	0.130***	0.072*
	(5.47)	(3.99)	(1.65)
$\#\,Analysts_j$	0.026***	0.020**	0.025***
	(3.51)	(2.56)	(2.76)
$StarAnalysts_j$		-0.105	-0.165
		(-0.40)	(-0.46)
$ReciprocalRatingShare_j$		-0.110	-0.111
		(-0.35)	(-0.31)
$PreviousRatings_j$		-0.020	-0.067
		(-0.27)	(-0.75)
$AnalystDispersion_j$		-0.003	-0.060
		(-0.07)	(-0.97)
$Benchy_j$	0.682***	0.774***	0.852***
	(8.15)	(7.15)	(5.37)
$ReviewToneDispersion_j$			-0.626
			(-0.30)
$ReviewTone_j$			0.952
			(0.48)
$ReviewUncertainty_j$			-5.779
			(-1.11)
$ReviewComplexity_j$			-0.001
			(-0.02)
$ReviewLength_j$			0.014
			(0.11)
MVP_j			-0.374**
			(-2.11)
IEO_j			0.752**
			(2.42)
$LengthWhitePaper_j$			0.040*
			(1.87)
Observations	2372	1612	1111
Pseudo R^2	0.146	0.154	0.186
Controls	No	No	Yes
$Month_j$ *Dummies*	Yes	Yes	Yes

Table 10: ICO outcomes that deviate from what ratings predict
This table presents marginal effects of logit regressions for Equation 6. The dependent variable is the *Disagreement* dummy which equals one if (i) analysts give an average $TotalRating_j \geq 13$ and the ICO fails, or if (ii) analysts give an average $TotalRating_j \leq 5$ and the ICO succeeds. In column (3), we restrict the sample to cases where the reciprocal ratings are on average greater than or equal to the average of non-reciprocal ratings for the same ICO. In column (4), we restrict the sample to ICOs where the average reciprocal rating is lower than the average of non-reciprocal ratings. All analyst variables are average values of every analyst that rates the ICO. Control variables for which coefficients are not shown for space reasons include the retention ratio and dummies for pre-sale, bonus/bounty options, *KYC*, Bitcointalk, and Facebook. All specifications include month dummies. All variables are defined in Table A1. *t*-statistics based on robust standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Dep. variable:	$Disagreement_j$			
	(1)	(2)	(3)	(4)
$\#\,Analysts_j$	0.007	0.008	0.004	-0.001
	(0.83)	(0.91)	(0.20)	(-0.03)
$StarAnalysts_j$	-0.762**	-0.609	0.092	-2.598
	(-2.11)	(-1.55)	(0.08)	(-1.56)
$ReciprocalRatingShare_j$	1.028***	0.885**	2.697*	0.900
	(2.80)	(2.18)	(1.87)	(0.46)
$PreviousRatings_j$	0.295***	0.211**	0.753**	1.495**
	(3.24)	(2.08)	(2.46)	(2.22)
$AnalystDispersion_j$	-0.458***	-0.474***	-0.698***	-0.504
	(-7.07)	(-6.37)	(-2.82)	(-1.45)
$Benchy_j$	0.055	-0.124	-1.065	0.166
	(0.48)	(-0.83)	(-1.53)	(0.22)
$ReviewTone_j$		8.626***	3.343	9.535
		(4.09)	(0.59)	(0.85)
$ReviewUncertainty_j$		-0.010	32.704	9.264
		(-0.00)	(1.61)	(0.32)
$ReviewComplexity_j$		0.064*	0.139	-0.247
		(1.90)	(1.08)	(-1.50)
$ReviewLength_j$		-0.072	-0.421	0.873
		(-0.49)	(-0.82)	(1.26)
MVP_j		0.106	0.544	0.486
		(0.53)	(0.86)	(0.63)
IEO_j		-0.722**	-0.549	0.450
		(-2.29)	(-0.39)	(0.38)
$LengthWhitePaper_j$		-0.053**	-0.060	-0.111
		(-2.11)	(-0.78)	(-1.55)
Observations	1615	1222	188	116
Pseudo R^2	0.194	0.201	0.235	0.270
Controls	No	Yes	Yes	Yes
$Month_j$ *Dummies*	Yes	Yes	Yes	Yes

The likelihood of an ICO being successful is higher if that ICO motivates analysts to rate it, even after controlling for a wide variety of ICO characteristics.[28] As Table 9 shows, however, the existence of experts' ratings is is important, but the level of that rating matters as well. A more positive $TotalRating_j$ is associated with a higher probability of success. The machine-generated rating Benchy is predictive as well, indicating that ICOs are, on average, more likely to be successful the more easily publicly available information there is about it.[29] Importantly, the human ratings remain significant determinants of success throughout, indicating that investors use the two rating types for different information. As for control variables, we observe similar results as the prior literature.[30]

Our main interest is in the heterogeneity among analysts, and in what predicts ICO failure despite high ratings. First, reciprocal ratings do not explain the outcome of an ICO unconditionally (see Table 9), but they do correlate significantly with failure (success) conditional on high (low) ratings (see Table 10). Table A3 shows that the effect emerges largely from failed ICOs despite high ratings. Thus, if ICO j receives favorable ratings from many analysts, whose ratings are a response to any team member of ICO j, the ultimate outcome of the ICO is more likely to deviate from that recommendation. There are two possible interpretations of this result. First, we note that we control for a wide variety of factors presumably capturing variation in ICO quality. However, it is still possible that reciprocal ratings are correlated with some additional unobserved variation in ICO quality. The second interpretation is that, a matter of principle, investors trust ICOs with more reciprocal ratings less, even when these ratings do not suffer from a conflict of interest.

[28]These results are in line with the general literature on analysts and rating agencies, which indicates that the market appreciates analyst coverage (Demiroglu and Ryngaert, 2010) and the existence of ratings (Sufi, 2009).

[29]This result is in line with the finding that investors value the dissemination of corporate news releases via robots, even when that information in principle is already available (Blankespoor et al., 2018).

[30]For example, we find a positive coefficient for Bitcointalk and negative coefficients for the Bounty and Bonus dummies. Successful ICOs tend to have longer white papers. For reasons of space, we only tabulate two control variables that received little prior attention in the literature: MVPs and IEOs. The use of crypto-exchange launchpads for Initial Exchange Offerings (IEOs) positively correlates with the two success variables. Somewhat surprisingly, ICOs with a minimum viable product (MVP) feature a lower probability of success and were only able to attract a lower dollar amount of funding. This unexpected result might be due to a non-regulated definition of minimum viable products. For example, drafts of codes on GitHub.com that are open to a discussion by other GitHub users were classified as MVP.

While these interpretations are not mutually exclusive, an additional test provides further insight. For each ICO, we calculate the difference between the average reciprocal and non-reciprocal ratings. We then divide the sample into cases where the average reciprocal rating is higher than or equal to the average of non-reciprocal ratings, and cases where reciprocal ratings were lower than the non-reciprocal ones. In the former case reciprocal ratings influence the overall ICO rating to a large extent, whereas in the latter case reciprocal ratings are less likely to bias the overall ICO rating. If investors dislike reciprocal ratings in general, we would expect the reciprocal rating share to be a significant determinant of disagreement in both cases. Columns 3 and 4 of Table 10 present the results. It is noteworthy that the share of reciprocal analysts matters only for the conditional success for those cases where reciprocal ratings are at least as positive as non-reciprocal ratings. A caveat is that these regressions are based on relatively small samples (because they are only available for the subsample with reciprocal ratings). That said, they provide some suggestive evidence that investors are not concerned with reciprocal ratings per se, but rather that positive reciprocal ratings provide an additional signal of poor quality of an ICO.

Consider next the linguistic measures of the rating. This feature also does not predict ICO success per se (at least once controlling for the quantitative rating). Nevertheless, linguistic measures still provide insight. Table 10 shows that the likelihood of failure for a highly-rated ICO increases as the positivity of the tone and complexity of the language increase. Similarly, ICO failure despite high average ratings occurs more frequently when the analysts were more positive in ratings prior to their rating of ICO j.

Table 9 suggests that star analyst coverage is not predictive for ICO success. Again, however, Table 10 reveals that highly rated ICOs fail less frequently when many star analysts cover them.

Finally, analyst dispersion is also relevant for ICO success only if the average view of analysts is very positive. Interestingly, and at first surprisingly, when analysts' rating are highly dispersed, ICOs are less likely to fail. Intuitively, the combination of high average ratings and high dispersion occurs when there are several extremely positive and some

negative views. The very positive ratings then carry the day. This is similar in spirit to the apparent anomaly that stocks with high dispersion of analyst opinions have high prices and, thus, lower future returns (Diether et al., 2002).[31].

Overall, the results highlight that even when a characteristic is not related unconditionally to ICO success, it is not irrelevant for understanding ICO success. Specifically, factors such as the reciprocity of ratings, analyst dispersion and the presence of star analysts explain deviations from the outcome given a very high level of ratings. Thus, while it is perhaps unsurprising that average ratings predict the success of an ICO campaign, our key result is that the detailed characteristics of the ICO ratings and those who provide them contain important additional information.

3.2.2 Ratings and ICO scams

We have established that analyst ratings help predict ICO success, but that investors tend to disregard reciprocal ratings. Does the latter result occur because ICOs with a higher fraction of reciprocal ratings are more likely to be fraudulent? To answer, we rerun the regression from Equation 5, but replace the success dummy with a dummy that equals one if the ICO was detected to intentionally defraud investors.

As Table 11 shows, we find no correlation between the share of reciprocal analysts and fraudulent ICOs. Also, the level of machine-generated ratings or the level of human analyst ratings do not help identify fraudulent ICOs. It still pays for investors to consider the human analyst assessments, however. In particular, ICOs with more dispersion among analysts both in the quantitative and in the qualitative rating tend to be fraudulent.

[31]This interpretation of analyst dispersion has been challenged in equity markets. For example, Avramov et al. (2009) show that the analyst dispersion anomaly is driven by a small fraction of firms with very high credit risk.

Table 11: ICO scams
This table presents marginal effects of logit regressions analogous to Equation 5, with the dependent variable being the *Scam* dummy. Control variables for which coefficients are not shown for space reasons include the retention ratio and dummies for pre-sale, bonus/bounty options, KYC, Bitcointalk, and Facebook. All specifications include month dummies. All variables are defined in Table A1. t-statistics based on robust standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Dep. variable:	$Scam_j$		
	(1)	(2)	(3)
$TotalRating_j$	0.059	-0.009	0.066
	(1.31)	(-0.13)	(0.55)
$\#\,Analysts_j$	0.027***	0.027***	0.031**
	(2.82)	(2.73)	(2.18)
$StarAnalysts_j$		-0.272	-0.918
		(-0.42)	(-0.91)
$ReciprocalRatingShare_j$		-0.157	-0.066
		(-0.22)	(-0.08)
$PreviousRatings_j$		0.248	0.517**
		(1.61)	(2.04)
$AnalystDispersion_j$		0.274**	0.564***
		(2.34)	(4.09)
$Benchy_j$	-0.217	-0.386**	-0.208
	(-1.20)	(-2.05)	(-0.58)
$ReviewToneDispersion_j$			5.808**
			(2.01)
$ReviewTone_j$			-1.859
			(-0.53)
$ReviewUncertainty_j$			-9.649
			(-1.11)
$ReviewComplexity_j$			-0.063
			(-1.02)
$ReviewLength_j$			-0.132
			(-0.41)
MVP_j			0.615
			(1.62)
IEO_j			-0.606
			(-0.80)
$LengthWhitePaper_j$			-0.075
			(-1.26)
Observations	2127	1466	966
Pseudo R^2	0.057	0.066	0.191
Controls	No	No	Yes
$Month_j$ *Dummies*	Yes	Yes	Yes

4 Conclusion

The intersection of new technologies and financial markets (FinTech) holds great promise. One relatively recent phenomenon in this space is the opportunity for new ventures to engage in Initial Coin Offerings (ICOs), a new form of financing. Yet, the problem of asymmetric information looms large in these markets. This paper studies information intermediaries, human experts, that may help ameliorate this asymmetric information problem.

While the ICO setting is interesting in its own right, ICO analysts show many interesting parallels to equity analysts or rating agencies. Particularly noteworthy are potential conflicts of interest, and how investors interpret them. The advantage of the ICO setting is that detailed data on links between analysts and securities they rate are available. For example, we document that an ICO analyst i, when rating an ICO j, tends to issue a rating that depends on the rating that their own affiliated ICO had previously received from team members of ICO j. However, there is a higher probability that an ICO fails, even when it has very favorable ratings, when more of those ratings are reciprocal. Moreover, such disagreement between investors and analysts is also more likely in ICOs with a large share of ratings by analysts with a history of very positive ratings.

Thus, while the prior literature shows that human experts' average ratings predict the success of an ICO campaign (over and above machine-generated ratings), our key result is that understanding ICO success requires looking past averages and instead studying the detailed characteristics of the ratings and those who provide them. After all, failure is frequent even among the most highly rated ICOs. Reciprocal ratings and highly positive reviews may be correlates of problems of highly rated ICOs not reflected in the many characteristics we control for; alternatively, investors may trust ICOs with more reciprocal and optimistic analysts less despite them being worth funding. Either way, the findings suggest that investors do not blindly pile capital into highly rated ICOs.

A necessary precondition for such investors to take such a differentiated approach to investments is the availability of information about the track record and potentially conflicting activities of analysts. Thus, the availability of information appears to support

the ICO market's allocative role in society. While these results obtain on this largely unregulated market, the general insight that investors seem to value information about analysts is likely to be relevant for other markets as well.

References

Adhami, S., Giudici, G., and Martinazzi, S. (2018). Why do businesses go crypto? An empirical analysis of Initial Coin Offerings. *Journal of Economics and Business*, 100(C):64–75.

Aggarwal, R., Hanley, K. W., and Zhao, X. (2019). The role of information production in private markets: Evidence from Initial Coin Offerings. Working paper.

Agrawal, A. and Chen, M. A. (2008). Do analyst conflicts matter? evidence from stock recommendations. *The Journal of Law & Economics*, 51(3):503–537.

Amsden, R. and Schweizer, D. (2018). Are blockchain crowdsales the new 'gold rush'? Success determinants of Initial Coin Offerings. Working paper.

Aubry, M., Kräussl, R., Manso, G., and Spaenjers, C. (2019). Machine learning, human experts, and the valuation of real assets. Working paper.

Avramov, D., Chordia, T., Jostova, G., and Philipov, A. (2009). Dispersion in analysts' earnings forecasts and credit rating. *Journal of Financial Economics*, 91(1):83–101.

Badoer, D. C., Demiroglu, C., and James, C. M. (2019). Ratings quality and borrowing choice. *The Journal of Finance*, 74(5):2619–2665.

Bae, K.-H., Stulz, R. M., and Tan, H. (2008). Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts. *Journal of Financial Economics*, 88(3):581–606.

Baghai, R. and Becker, B. (2017). Non-rating revenue and conflicts of interest. *Journal of Financial Economics*, 127.

Baker, H. K. and Mansi, S. A. (2002). Assessing credit rating agencies by bond issuers and institutional investors. *Journal of Business Finance & Accounting*, 29(9-10):1367–1398.

Bakos, Y. and Halaburda, H. (2019). The role of cryptographic tokens and ICOs in fostering platform adoption. Working paper.

Bar-Isaac, H. and Shapiro, J. (2013). Ratings quality over the business cycle. *Journal of Financial Economics*, 108(1):62 – 78.

Benedetti, H. and Kostovetsky, L. (2018). Digital tulips? Returns to investors in Initial Coin Offerings. Working paper.

Biais, B., Bisière, C., Bouvard, M., and Casamatta, C. (2019). The blockchain folk theorem. *Review of Financial Studies*, 32(5):1662–1715.

Blankespoor, E., deHaan, E., and Zhu, C. (2018). Capital market effects of media synthesis and dissemination: Evidence from robo-journalism. *Review of Accounting Studies*, 23(1):1–36.

Bolton, P., Freixas, X., and Shapiro, J. (2012). The credit ratings game. *The Journal of Finance*, 67(1):85–111.

Boreiko, D. and Vidusso, G. (2019). New blockchain intermediaries: Do ICO rating websites do their job well? *Journal of Alternative Investments*, 21(4):67–79.

Bourveau, T., De George, E. T., Ellahie, A., and Macciocchi, D. (2019). Information intermediaries in the crypto-tokens market. Working paper.

Bradley, D., Gokkaya, S., and Liu, X. (2017). Before an analyst becomes an analyst: Does industry experience matter? *The Journal of Finance*, 72(2):751–792.

Bradshaw, M., Ertimur, Y., O'Brien, P., et al. (2017). Financial analysts and their contribution to well-functioning capital markets. *Foundations and Trends (R) in Accounting*, 11(3):119–191.

Burns, L. and Moro, A. (2018). What makes an ICO successful? An investigation of the role of ICO characteristics, team quality and market sentiment. Working paper.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business and Economic Statistics*, 29(2):238–249.

Campbell, J., DeAngelis, M., and Moon, J. (2019). Skin in the game: personal stock holdings and investors' response to stock analysis on social media. *Review of Accounting Studies*, 24(C):731–779.

Catalini, C., Foster, C., and Nanda, R. (2018). Machine intelligence vs. human judgement in new venture finance. Working paper.

Cerchiello, P. and Toma, A. M. (2018). ICOs success drivers: A textual and statistical analysis. Working paper.

Chakraborty, S. and Swinney, R. (2020). Signaling to the crowd: Private quality information and rewards-based crowdfunding. *Manufacturing & Service Operations Management*, 1(C):1–15.

Chan, L. K. C., Karceski, J., and Lakonishok, J. (2007). Analysts' conflicts of interest and biases in earnings forecasts. *Journal of Financial and Quantitative Analysis*, 42(4):893–913.

Chen, H., De, P., Hu, Y. J., and Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, 27(5):1367–1403.

Chod, J. and Lyandres, E. (2020). A theory of ICOs: Diversification, agency, and information asymmetry. *Management Science*, forthcoming.

Chod, J., Trichakis, N., and Yang, A. S. (2020). Platform tokenization: Financing, governance, and moral hazard. Working paper.

Chu, C. S. and Rysman, M. (2019). Competition and strategic incentives in the market for credit ratings: Empirics of the financial crisis of 2007. *American Economic Review*, 109(10):3514–55.

Cohen, L., Lou, D., and Malloy, C. J. (2020). Casting conference calls. *Management Science*, forthcoming.

Cong, L. W. and He, Z. (2019). Blockchain disruption and smart contracts. *The Review of Financial Studies*, 32(5):1754–1797.

Cong, L. W., Li, Y., and Wang, N. (2020a). Token-based platform finance. Working paper.

Cong, L. W., Li, Y., and Wang, N. (2020b). Tokenomics: Dynamic adoption and valuation. *Review of Financial Studies*, forthcoming.

Conley, J. P. (2017). Blockchain and the economics of crypto-tokens and Initial Coin Offerings. Working paper.

Crawford, S. S., Roulstone, D. T., and So, E. C. (2012). Analyst initiations of coverage and stock return synchronicity. *The Accounting Review*, 87(5):1527–1553.

Da, Z. and Huang, X. (2020). Harnessing the wisdom of crowds. *Management Science*, 66(5):1847–1867.

Davydiuk, T., Gupta, D., and Rosen, S. (2019). De-crypto-ing signals in Initial Coin Offerings: Evidence of rational token retention. Working paper.

Demiroglu, C. and Ryngaert, M. (2010). The first analyst coverage of neglected stocks. *Financial Management*, 39(2):555–584.

Deng, X., Lee, Y. T., and Zhong, Z. (2018). Decrypting coin winners: Disclosure quality, governance mechanism and team networks. Working paper.

Diether, K. B., Malloy, C. J., and Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance*, 57(5):2113–2141.

Dittmar, R. and Wu, D. A. (2019). Initial Coin Offerings hyped and dehyped: An empirical examination. Working paper.

Drake, M. S., Thornock, J. R., and Twedt, B. J. (2017). The internet as an information intermediary. *Review of Accounting Studies*, 22(2):543–576.

Easley, D., O'Hara, M., and Basu, S. (2019). From mining to markets: The evolution of bitcoin transaction fees. *Journal of Financial Economics*, 134(1):91–109.

Ellis, D. (1998). Different sides of the same story: Investors? And issuers? Views of rating agencies. *The Journal of Fixed Income*, 1:35–45.

Ernst and Young (2018). Initial Coin Offerings (ICOs): The class of 2017 – one year later. EY study.

Fahlenbrach, R. and Frattaroli, M. (2020). ICO Investors. *Financial Markets and Portfolio Management*, pages 1–59.

Farrell, M., Jame, R., and Qiu, T. (2020). The cross-section of non-professional analyst skill. Working paper.

Feng, C., Li, N., Wong, M., and Zhang, M. (2019). Initial Coin Offerings, blockchain technology, and white paper disclosures. Working paper.

Fisch, C. (2019). Initial Coin Offerings (ICOs) to finance new ventures. *Journal of Business Venturing*, 34(1):1–22.

Fisch, C., Masiak, C., Vismara, S., and Block, J. (2019). Motives and profiles of ICO investors. *Journal of Business Research*, forthcoming.

Fisch, C. and Momtaz, P. P. (2020). Institutional investors and post-ICO performance: An empirical analysis of investor returns in Initial Coin Offerings (ICOs). *Journal of Corporate Finance*, 64:101679.

Florysiak, D. and Schandlbauer, A. (2019). The information content of ICO white papers. Working paper.

Gan, R. J., Tsoukalas, G., and Netessine, S. (2020). Initial Coin Offerings, speculators, and asset tokenization. Working paper.

Giudici, G. and Adhami, S. (2019). The impact of governance signals on ICO fundraising success. *Journal of Industrial and Business Economics*, 46(2):283–312.

Goldstein, I., Gupta, D., and Sverchkov, R. (2019). Initial Coin Offerings as a commitment to competition. Working paper.

Gryglewicz, S., Mayer, S., and Morellec, E. (2020). Optimal financing with tokens. Working paper.

Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill, New York.

Hinzen, F. J., John, K., and Saleh, F. (2020). Bitcoin's fatal flaw: The limited adoption problem. Working paper.

Hirth, S. (2014). Credit rating dynamics and competition. *Journal of Banking & Finance*, 49(C):100 – 112.

Howell, S. T., Niessner, M., and Yermack, D. (2019). Initial Coin Offerings: Financing growth with cryptocurrency token sales. *The Review of Financial Studies*.

Hu, A. S., Parlour, C. A., and Rajan, U. (2019). Cryptocurrencies: Stylized facts on a new investible instrument. *Financial Management*, 48(4):1049–1068.

Jame, R., Johnston, R., Markov, S., and Wolfe, M. C. (2016). The value of crowdsourced earnings forecasts. *Journal of Accounting Research*, 54(4):1077–1110.

Kaal, W. A. (2018). Initial Coin Offerings: The top 25 jurisdictions and their comparative regulatory responses (as of May 2018). *Stanford Journal of Blockchain Law & Policy*, 1(1):41–63.

Kadan, O., Madureira, L., Wang, R., and Zach, T. (2009). Conflicts of Interest and Stock Recommendations: The Effects of the Global Settlement and Related Regulations. *The Review of Financial Studies*, 22(10):4189–4217.

Kempf, E. (2020). The job rating game: Revolving doors and analyst incentives. *Journal of Financial Economics*, 135(1):41–67.

Klein, A., Saunders, A., and Wong, Y. T. F. (2019). Is there a quid pro quo between hedge funds and sell-side equity analysts? *The Journal of Portfolio Management*, 45(6):117–132.

Law, K. K. F. and Mills, L. F. (2019). Financial gatekeepers and investor protection: Evidence from criminal background checks. *Journal of Accounting Research*, 57(2):491–543.

Lee, J., Li, T., and Shin, D. (2019). The wisdom of crowds in FinTech: Evidence from Initial Coin Offerings. Working paper.

Lee, J. and Parlour, C. A. (2020). Consumers as financiers: Consumer surplus, crowdfunding, and Initial Coin Offerings. Working paper.

Leone, A. J. and Wu, J. (2002). What does it take to become a superstar? Evidence from institutional investor rankings of financial analysts. Working paper.

Leuz, C., Jakab, L., Malani, A., and Muhn, M. (2021). Do conflict of interests disclosures work? evidence from citations in medical journals. Working paper.

Leuz, C. and Wysocki, P. D. (2016). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of accounting research*, 54(2):525–622.

Li, J. and Mann, W. (2019). Initial Coin Offering: Current research and future directions. In *Handbook of Alternative Finance*. Palgrave-MacMillan, forthcoming.

Li, J. and Mann, W. (2020). Initial Coin Offerings and platform building. Working paper.

Lin, H. and McNichols, M. F. (1998). Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics*, 25(1):101–127.

Livingston, M. B., Wei, J. D., and Zhou, L. (2010). Moody's and S&P ratings: Are they equivalent? Conservative ratings and split rated bond yields. *Journal of Money, Credit and Banking*, 42(7).

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? *Journal of Finance*, 66(1):35–65.

Lourie, B. (2019). The revolving door of sell-side analysts. *The Accounting Review*, 94(1):249–270.

Lyandres, E. (2019). Product market competition with crypto tokens and smart contracts. Working paper.

Lyandres, E., Palazzo, B., and Rabetti, D. (2020). ICO success and post-ICO performance. Working paper.

Malinova, K. and Park, A. (2018). Tokenomics: When tokens beat equity. Working paper.

Malmendier, U. and Shanthikumar, D. (2007). Are small investors naive about incentives? *Journal of Financial Economics*, 85(2):457–489.

Merkley, K., Michaely, R., and Pacelli, J. (2017). Does the scope of the sell-side analyst industry matter? An examination of bias, accuracy, and information content of analyst reports. *The Journal of Finance*, 72(3):1285–1334.

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

Mola, S., Rau, P. R., and Khorana, A. (2013). Is there life after the complete loss of analyst coverage? *The Accounting Review*, 88(2):667–705.

Momtaz, P. P. (2020). Initial Coin Offerings. *PLOS ONE*, 15(5):1–30.

O'Brien, P., McNichols, M. F., and Hsiou-Wei, L. (2005). Analyst impartiality and investment banking relationships. *Journal of Accounting Research*, 43(4):623–650.

PwC (2019). 5th ICO / STO report. PwC study.

Roosenboom, P., van der Kolk, T., and de Jong, A. (2020). What determines success in Initial Coin Offerings? *Venture Capital*, 22(2):1–23.

Samieifar, S. and Baur, D. G. (2020). Read me if you can! An analysis of ICO white papers. *Finance Research Letters*, page 101427.

Sufi, A. (2009). The real effects of debt certification: Evidence from the introduction of bank loan ratings. *The Review of Financial Studies*, 22(4):1659–1691.

thinkBLOCKtank (2019). The regulation of tokens in Europe. Parts A & B: The EU legal and regulatory framework. thinkBLOCKtank study.

Tiwari, M., Gepp, A., and Kumar, K. (2019). The future of raising finance–A new opportunity to commit fraud: A review of Initial Coin Offering (ICOs) scams. *Crime, Law and Social Change*, pages 1–25.

Welch, I. (2000). Herding among security analysts. *Journal of Financial Economics*, 58(3):369–396.

Zhang, S., Aerts, W., Lu, L., and Pan, H. (2019). Readability of token whitepaper and ICO first-day return. *Economics Letters*, 180(C):58–61.

Appendix

Table A1: Variable definitions

Variable	Definition
$\#\,Analysts_j$	Number of analysts that rate an ICO.
$AmountRaised_j$	Natural logarithm of one plus $ amount raised by an ICO.
$AnalystDispersion_j$	Standard deviation of ratings within an ICO.
$Benchy_j$	Machine-generated rating created by ICObench.com.
$Bitcointalk_j$	Dummy variable that equals one if the ICO is discussed on the forum bitcointalk.org.
$Bonus_j$	Dummy variable that equals one for ICOs with a quantity discount at the token sale or a discount program for early-bird investors.
$Bounty_j$	Dummy variable that equals one for ICOs with incentives to promote social media presence.
$Disagreement_j$	Dummy variable that equals one if (i) on average, analysts recommend buying ($TotalRating_j \geq 13$) and the ICO fails, or (ii) on average, analysts recommend selling ($TotalRating_j \leq 5$) and the ICO succeeds.
$Facebook_j$	Dummy variable that equals one if an ICO has a Facebook page.
$ForecastError_{ij}$	The distance of the total rating from the highest (lowest) possible rating in the case of ICO success (failure).
$ForecastErrorOptimistic_i$	The distance of the highest possible rating score to the actual total rating of analyst i, defined as $15 - RatingTotal$, if the ICO was unsuccessful, and averaged over all ICOs j.
$ForecastErrorPessimistic_i$	The distance of the total rating of analyst i to the lowest possible rating score, defined as $RatingTotal - 3$, if the ICO was successful, and averaged over all ICOs j.
$ForecastError_i^j$	A recursive average of all previous forecast errors for any analyst i up to the rating issuance date for ICO j.
$ForecastError_j$	A recursive average of the previous forecast errors of all analysts covering ICO j up to the rating issuance date.
IEO_j	Dummy variable that equals one for ICOs conducted on the platform of a cryptocurrency exchange (Initial Exchange Offerings).
KYC_j	Dummy variable that equals one for ICOs where investors are required to sign up to a whitelist using their wallet address to receive access to the ICO sale (Know Your Customer).
$LengthWhitePaper$	The natural logarithm of (1 + total words of the white paper), set to 0 if no white paper could be found.
$Modified_{ij}$	Dummy variable that equals one if the rating for ICO j was modified by analyst i at any point in time.

$Month_j$	Dummy variable for each month, indicating the month when an ICO was launched.
$Month_{ij}$	Dummy variable for each month, indicating the month when a rating was given.
MVP_j	Dummy variable that equals one for ICOs with a prototype. This can be a version of a new product with sufficient features to satisfy early adopters (minimum viable product) or drafts of code on Github.com that are open to discussion by other GitHub users.
$Presale_j$	Dummy variable that equals one if an ICO features a token sale event that runs prior to the official ICO campaign.
$OrderRank_{ij}$	The order rank of the rating by analyst i issued for ICO j in a given month.
$PreviousRatings_j$	Average past TotalRating of all analysts that provide a rating for ICO j
$ReceivedTeamRating_{ij}$/ $ReceivedVisionRating_{ij}$/ $ReceivedProductRating_{ij}$/ $ReceivedTotalRating_{ij}$	Level of the rating when ReciprocalRating dummy equals 1, i.e., level of rating that the analyst of ICO j received for their own ICO from any team member of ICO j prior to the rating issuance date.
$ReciprocalRating_{ij}$	Dummy variable that equals one for reciprocal ratings. A rating is reciprocal when the corresponding analyst was a team member of another ICO project that previously received a rating by one of the team members of this new ICO. Table 2 represents a hypothetical illustration of our variable composition.
$ReciprocalRatingShare_j$	Share of reciprocal analysts that provide a rating for ICO j.
$RetentionRatio_j$	The percentage of token supply that is retained by the ICO members, and not available for sale.
$ReviewComplexity_j$	The complexity of an analyst's review text, measured by the Gunning (1952) Fog index, and averaged together on ICO level.
$ReviewLength_{ij}$	The natural logarithm of the number of total words in an analyst review. For the $ReviewLength_j$, we measure the natural logarithm of the average review text lengths for ICO j.
$ReviewTone_{ij}$	The tone of the analyst review text. Using the Loughran and McDonald (2011) *Positive* and *Negative* word-lists, the tone of a text is defined as the difference between the count of positive and negative words divided by the total words.
$ReviewTone_j$	The tone averaged across all analysts' review texts for ICO j.
$ReviewToneDispersion_j$	The standard deviation of $ReviewTone_{ij}$ within an ICO.

$ReviewUncertainty_j$	The uncertainty of the analysts' review texts, averaged together on ICO level. Using the Loughran and McDonald (2011) *Uncertainty* word-list, the uncertainty of a text is defined as the count of uncertain words divided by the total words.
$Scam_j$	Dummy variable that equals one for ICO projects that intentionally defraud investors.
$Success_j$	Dummy variable that equals one for ICOs that completed the token sale and collected (at least $1) funding.
$StarAnalysts_{ij}$	Dummy variable that equals one when ICO j was rated by one of the top 30 analysts i according to a ranking on ICObench.com.
$StarAnalysts_j$	Share of the top 30 analysts that provide a rating for ICO j.
$TeamRating_{ij}$/ $VisionRating_{ij}$/ $ProductRating_{ij}$	Rating score for team/ vision/ product of an ICO, ranging from 1 (lowest) to 5 (highest).
$TotalRating_{ij}$	The sum of team, vision and product ratings for the respective ICO, ranging from 3 to 15.
$TotalRating_j$	Average rating of ICO j by all analysts.

Table A2: Ratings and ICO success: An alternative success measure
This table presents linear regression results for Equation 5. The dependent variable is the natural logarithm of the amount raised by an ICO in columns. The controls for which coefficients are not shown for space reasons include the retention ratio and dummies for pre-sale, bonus/bounty options, KYC, Bitcointalk, and Facebook. All specifications include month dummies. All variables are defined in Table A1. t-statistics based on robust standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Dep. variable:	$AmountRaised_j$		
	(1)	(2)	(3)
$TotalRating_j$	0.325***	0.390***	0.246**
	(6.11)	(4.74)	(2.14)
$\#\,Analysts_j$	0.097***	0.072***	0.078***
	(5.11)	(3.57)	(3.52)
$StarAnalysts_j$		-0.371	-0.302
		(-0.45)	(-0.28)
$ReciprocalRatingShare_j$		-0.402	-0.340
		(-0.39)	(-0.30)
$PreviousRatings_j$		-0.103	-0.201
		(-0.48)	(-0.75)
$AnalystDispersion_j$		-0.017	-0.140
		(-0.12)	(-0.80)
$Benchy_j$	1.858***	2.062***	2.182***
	(8.53)	(7.38)	(5.58)
$ReviewToneDispersion_j$			-2.660
			(-0.45)
$ReviewTone_j$			2.451
			(0.45)
$ReviewUncertainty_j$			-16.625
			(-1.05)
$ReviewComplexity_j$			0.016
			(0.15)
$ReviewLength_j$			-0.042
			(-0.10)
MVP_j			-1.252**
			(-2.39)
IEO_j			1.985**
			(2.32)
$LengthWhitePaper_j$			0.117*
			(1.85)
Observations	2378	1629	1128
R^2	0.190	0.208	0.247
Controls	No	No	Yes
$Month_j$ Dummies	Yes	Yes	Yes

Table A3: ICO outcomes that deviate from what ratings predict
This table presents marginal effects of logit regressions for Equation 6. The dependent variable is the *Disagreement* dummy, which equals one if analysts recommend buying (average $TotalRating_j \geq 13$) and the ICO fails. All analyst variables are average values over all analysts that rate the ICO. Control variables for which coefficients are not shown for space reasons include the retention ratio and dummies for pre-sale, bonus/bounty options, KYC, Bitcointalk, and Facebook. All specifications include month dummies. All variables are defined in Table A1. t-statistics based on robust standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Dep. variable:	$Disagreement_j$	
	(1)	(2)
$\#\,Analysts_j$	0.042***	0.056***
	(2.99)	(3.30)
$StarAnalysts_j$	-1.346***	-0.886*
	(-3.01)	(-1.83)
$ReciprocalRatingShare_j$	2.033***	1.799***
	(4.34)	(3.21)
$PreviousRatings_j$	0.532***	0.452***
	(4.25)	(3.35)
$AnalystDispersion_j$	-0.678***	-0.733***
	(-7.67)	(-7.13)
$Benchy_j$	0.597***	0.277
	(3.76)	(1.28)
$ReviewTone_j$		17.694***
		(4.27)
$ReviewUncertainty_j$		-10.057
		(-1.02)
$ReviewComplexity_j$		0.135***
		(3.01)
$ReviewLength_j$		-0.015
		(-0.08)
MVP_j		0.017
		(0.07)
IEO_j		-0.540
		(-1.23)
$LengthWhitePaper_j$		-0.039
		(-1.10)
Observations	776	640
Pseudo R^2	0.263	0.308
Controls	No	Yes
$Month_j$ *Dummies*	Yes	Yes

Chapter V

Financing sustainable entrepreneurship

Financing Sustainable Entrepreneurship: ESG Measurement, Valuation, and Performance in Token Offerings*

Sasan Mansouri† Paul P. Momtaz‡§

July 17, 2021

Abstract

Sustainable Entrepreneurship (SE) seeks to attain profitability *and* sustainability goals. A major research gap concerns the economic attractiveness of SE for entrepreneurs and investors. The question is ambiguous because sustainability orientation creates costly constraints, while startups cannot fully appropriate their positive externalities. We relate startups' Environment, Society and Governance (ESG) properties obtained from a machine-learning approach (www.SustainableEntrepreneurship.org) to SE valuation and performance in token offerings. Startups with salient ESG goals are able to raise financing at more favorable valuations, incentivizing entrepreneurs to adopt ESG goals in the first place. However, their post-funding performance is weaker than in conventional startups, suggesting that investors incur a relative *financial* loss for backing sustainability-oriented entrepreneurs. Both valuation and post-funding performance are weaker in ESG startups with pre-existing binding constraints.

Keywords: Sustainable Entrepreneurship, Sustainability, ESG, Token Offering, Initial Coin Offering (ICO), Crowdfunding, Entrepreneurial Finance, Machine Learning

JEL Codes: L26, M13, Q01, Q56

*We thank seminar participants at the Entrepreneurial Finance Association 2021 Annual Meeting for helpful comments.

†House of Finance, Goethe-University Frankfurt, Theodor-W.-Adorno-Platz 3. 60629 Frankfurt, Germany. mansouri@finance.uni-frankfurt.de

‡(Corresponding author.) UCLA Anderson School of Management, Los Angeles, CA 90095, momtaz@ucla.edu

§House of Finance, Goethe-University Frankfurt, Theodor-W.-Adorno-Platz 3. 60629 Frankfurt, Germany.

1 Introduction

Sustainable Entrepreneurship (SE) is a rapidly growing literature (for excellent recent reviews, see Anand et al., 2021; Johnson and Schaltegger, 2020).[1] SE is characterized by profit-seeking entrepreneurial activity that embraces the broader (non-financial) Environment, Society and Governance (ESG) goals of our time.[2] A common theme in the literature is that it evokes Schumpeter's (1942) notion of 'creative destruction' to explain how SE may affect sustainable change (e.g., Cohen and Winn, 2007; Hall and Vredenburg, 2003; S. L. Hart and Christensen, 2002; S. L. Hart and Milstein, 1999; Senge et al., 2001; Shepherd and Patzelt, 2017; and, for a general discussion of Schumpeterian logic applied to SE, Hockerts and Wüstenhagen, 2010; York and Venkataraman, 2010). The literature's tenet is that market failure to solve ESG challenges creates entrepreneurial opportunities.

An important research gap is whether ESG-driven opportunities are *economically* attractive for entrepreneurs in the first place. Schumpeter (1934, 1942) assumes that technological innovations provide entrepreneurs with a *business case* (often associated with more cost-efficient production than incumbents), which is the underlying force behind unfolding 'creative destruction' dynamics. It is ambiguous, however, whether such a business case exists for SE for at least two reasons: (i) ESG goals impose binding restrictions upon entrepreneurs that limit the scope of viable routes to (economic) success, and (ii) entrepreneurs largely fail to internalize ESG rents because they come as positive externalities. Uncertainty about the economic appeal of SE is ubiquitous in the literature. For example, Hall et al. (2010) refer to SE as a "controversial" field with "major gaps in our knowledge of whether and how this process [i.e., SE] will actually unfold", partly because opportunities for SE "lie beyond the pull of existing markets" (p. 439). Our paper represents a first step towards addressing this important gap by posing the following research question:

How (economically) attractive is SE for entrepreneurs and investors?[3]

This question is fundamental for SE scholars and policy-makers alike because a potential lack of economic incentives would suggest that entrepreneurs need government

[1] For earlier reviews, see Bischoff and Volkmann, 2018; Dean and McMullen, 2007; Gast et al., 2017; Kraus et al., 2018; Muñoz and Cohen, 2018; Sarango-Lalangui et al., 2018; Schaefer et al., 2015; Shepherd and Patzelt, 2011; Terán-Yépez et al., 2020.

[2] SE's profit orientation is the key distinguishing factor from social and environmental entrepreneurship that focuses on socio-ecological returns as its primary goal, as discussed by, e.g., Kraus et al. (2018)

[3] While we are not the first to examine the relation between sustainability-orientation and funding success (see also, e.g., Guzmán et al., 2020; Hörisch, 2015; Vismara, 2019), we are the first to provide a holistic view on the question, in particular by extending the narrow focus from the funding campaign to the post-funding performance of SE startups.

subsidies to act as ESG "change agents" (Anand et al., 2021, p. 2), and that SE scholars potentially need to adopt a different lens than Schumpeter's (1942).[4]

We argue, both theoretically and empirically, that a sufficient condition for SE to effect sustainable change is that sustainability-oriented startups obtain enough funding at sufficiently high valuations relative to conventional startups. The literature on financing SE is very limited, with the notable exception of Vismara (2019).[5] Therefore, reflecting the "multidisciplinary character" of SE (Anand et al., 2021, p. 1), we also borrow from signaling (Ahlers et al., 2015; O. Colombo, 2021; Fisch, 2019), non-economic utility (Barber et al., 2021; Cornell, 2021), and financial markets theory (Fama and French, 2007; Pástor et al., 2020; and, for a review in the ESG context, Gillan et al., 2021) to develop two hypotheses related to the valuation and performance of sustainable startups.

The prospect of non-economic utility is the key feature distinguishing SE from Conventional Entrepreneurship (CE) in entrepreneurial finance markets (Vismara, 2019). In the hypothetical scenario that SE and CE share the same business case, SE should receive higher valuations, with the differential being attributable to investors' ESG-related utility. An '*ESG premium*' on startup value is even in line with Friedman's (1970) famous claim that "the social responsibility of business is to make profits." As long as entrepreneurs have a competitive advantage to achieve economic and ESG goals together, then investors should delegate ESG goals to entrepreneurs with specialized skills (O. Hart and Zingales, 2017). For example, it is more efficient for investors to delegate their ESG goals to three specialized startups – one that targets E-goals, another for S-goals, and a third for G-goals – than to tackle all ESG goals themselves. Therefore, under our '***Valuation Premium Hypothesis***' (***VPH***), SE receives higher valuations than CE. As further discussed in section 3, the *VPH* can be connected to existing evidence that SE is associated with, inter alia, (i) better risk management (Knight, 1997; Kraus et al., 2018), (ii) trust-creating altruism (Momtaz, 2020c; Tilley and Young, 2009), (iii) first-mover advantages (Hockerts and Wüstenhagen, 2010; Lieberman and Montgomery, 1988), and (iv) personal characteristics that are correlated with signals of entrepreneurial quality, such as human, social and intellectual capital (Ahlers et al., 2015; O. Colombo, 2021; Egri and Herman, 2000; Fisch, 2019; Spence et al., 2011; Vega and Kidwell, 2007).

The flip side of delegated philanthropy is that SE may (economically) underperform in the long run, which is at the core of much controversy in the SE literature (Hall et al., 2010; Kraus et al., 2018). We label this prediction the '***Post-Funding Underperformance Hypothesis***' (***PFUH***). Financial equilibrium theory argues that investors' greater

[4]It is important to note that our focus is on the financial rents of SE, as there is no consensus about how to measure non-financial rents. In the SE context, Anand et al. (2021, p. 12), discuss that "concerns regarding 'how to measure sustainability' emerge as one of the major challenges."

[5]See also Cumming et al., 2016; Cumming et al., 2017; Guzmán et al., 2020; Hörisch, 2015.

"willingness-to-pay" (Barber et al., 2021, p. 1), which is a source of the ESG premium in the first place, has to be followed by lower expected (financial) returns (Fama and French, 2007; Gillan et al., 2021). Two important aspects deserve elaboration. Lower financial returns (i.e., underperformance) do not eliminate incentives for entrepreneurs or investors to get involved in SE. The former benefit from the ESG premium during the funding stage, while the latter sacrifice financial returns for the sake of ESG returns. In aggregate, i.e., after adding ESG to financial returns, investors may be better off depending on their personal preferences for sustainability goals. Therefore, it is helpful to draw a distinction between 'investor value' and 'investor welfare,' only the latter referring to combined economic and ESG rents. To our knowledge, our study is the first to examine the long-term economic performance of SE.[6]

Empirically, we employ a machine-learning (ML) approach to quantify startups' ESG properties, using information disclosed in ICO whitepapers. Specifically, we use Mikolov et al.'s (2013) semi-supervised word-embedding approach, which trains a neural network to learn the meaning of words and phrases in their respective context. Our approach to finding the ESG-related words and phrases is inspired by Li et al. (2020) in the sense that we define a set of "seed" words/phrases in the first step, and then use the trained word embedding model to find the closest terms to our seeds. We manually create seed word lists of Environmental (E), Social (S) and Governance (G)-related terminologies by collecting every *Financial Times* article tagged as "ESG investing" or "Moral Money", and focusing on the words and phrases used most frequently in them. Our procedure yields a total of 1,495 ESG-related terms consisting of 508, 463, and 524 terms for E, S and G, respectively (additional analyses based on more parsimonious lists demonstrate the robustness of our results). We then measure startups' E, S and G intensities by measuring the unique counts of the terms from the respective word list in the whitepaper. The sum of the three E, S and G intensities gives a startup's aggregate ESG score. Manual inspection suggests that our ESG scores identify the startups with the most salient ESG properties well. For replication purposes and as an aid for future SE research, our source code is available at our easy-to-use web application www.SustainableEntrepreneurship.org.

Our results support both the VPH and the PFUH. We examine a large sample of 1,043 token offerings over the period 2016-2020.[7] Token offerings are blockchain-

[6]Our study estimates financial underperformance of SE, i.e., investor value. In contrast, investor welfare cannot be observed directly, as sustainability preferences are heterogenous across investors and private. Nevertheless, our study can be understood as an upper bound to ESG rents, acknowledging the fact that ESG rents are fully responsible for SE underperformance relative to CE, as moral hazard in ESG signaling may also explain part of the underperformance (Momtaz, 2020a; Spence et al., 2011). Other studies that focus on SE outcomes, but with a different focus are Dickel (2017), Djupdal and Westhead (2015), Gregori et al. (2019), Hoogendoorn et al. (2019), Jahanshahi and Brem (2017), Kraus et al. (2017), Lans et al. (2014), Muñoz, Cacciotti, et al. (2018), Mupfasoni et al. (2018), Testa et al. (2019), and Volkmann et al. (2021).

[7]Our paper is fully replicable. The data come from the Token Offerings Research Database (TORD),

based crowdfunding campaigns, in which smart contracts govern the exchange of fiat money for tokens between investors and entrepreneurs (Amsden and Schweizer, 2018; Bellavitis et al., 2020; Bellavitis et al., 2021; Fisch, 2019; Giudici and Adhami, 2019; Howell et al., 2020; Huang et al., 2020; Momtaz, 2019, 2020b). Token offerings are an ideal laboratory to examine the economics of SE. They largely tap pools of individual investors that may be more motivated by non-financial goals than institutional investors (Fisch et al., 2019). Startups with salient ESG properties benefit from substantially higher valuations, supporting the *VPH*. A one-standard-deviation increase in the ESG metric is associated with a 28% increase in the funding amount, which corresponds to around $4.2million (relative to the mean funding amount of $15.2million in our sample). Consistent with the *PFUH*, startups with pronounced ESG properties underperform in the first year after which a token was listed on an exchange platform. A one-standard-deviation increase in the ESG metric is associated at least with a 16% decrease in the first 12-month buy-and-hold abnormal (equally weighted relative to a composite market index) token price performance after the crowdfunding event. Relative to financial utility, non-financial (ESG-related) utility for SE investors amounts to 16-31% of total utility.[8] Both main results are robust to endogeneity concerns related to observed and unobserved heterogeneity.

Given these results, an important next question for entrepreneurs and investors alike in moving forward with SE is whether and how the negative effect on financial performance can be mitigated (Parrish, 2010, for a general discussion of organizational design differences between SE and CE). The excellent review by Kraus et al. (2018) synthesizes the literature, concluding that a high degree of formalization may drive poor SE performance. Formalization refers to organizational structure, such as "control systems and reporting procedures, as well as the formal style of tracking the progress" (Kraus et al., 2018, p. 8). Therefore, pre-existing binding constraints should negatively moderate the empirical patterns predicted by the *VPH* and *PFUH*. Consistent with this reasoning, technological, network and governance aspects associated with startup formalization all hurt SE success. This marks a stark contrast to CE. While typical technology startup attributes, such as open-source code, a large social network and venture capital backing, are typically associated with entrepreneurial success (Fisch, 2019; Fisch and Momtaz, 2020), they can be detrimental in ESG startups. The finding highlights the need for future research to better understand how organizational design can promote, rather than hurt, sustainability-oriented venturing.

The remainder of the paper is organized as follows. Section 2 reviews the existing and multidisciplinary literature on SE and section 3 derives empirical predictions. Sec-

see www.paulmomtaz.com/data/tord, and the machine-learning algorithm to quantify ESG properties of our sample startups is made available along with this publication.

[8]We view this estimate as an upper bound on ESG-related utility. See also footnote 5.

tion 4 discusses our machine-learning approach to quantify startups' ESG properties. Section 5 describes our sample and section 6 presents our empirical results. Finally, Section 7 provides a discussion, highlights limitations and potential avenues for future research, and concludes the paper.

2 Related Literature

2.1 Sustainable Entrepreneurship

A consensual definition of sustainable entrepreneurship does not yet exist. However, Anand et al. (2021) and Johnson and Schaltegger (2020) provide excellent recent reviews of the literature.[9] Early studies draw on the concept of "sustainable development," which the *United Nations' World Commission on Environment and Development* (*WCED*) introduced in 1987 (e.g., Cohen and Winn, 2007; Dean and McMullen, 2007; Hall et al., 2010). According to the WCED, sustainable development refers to society striving to satisfy its needs without compromising the ability of future generations to satisfy their needs. Some studies draw a strict demarcation line between SE and social and environmental entrepreneurship along the entrepreneurs' distinct objective functions. As reviewed in Kraus et al. (2018), SE's primary goal is to create *positive financial returns* while not harming society and the environment (i.e., *non-negative non-financial returns*), whereas social and environmental entrepreneurship's primary goal is to create *positive non-financial returns*. Furthermore, in contrast to the broader ESG literature in management and economics, SE has thus far focused on E and S goals, thus neglecting G goals. For example, Dean and McMullen (2007) define SE as "the role entrepreneurs can play in creating a more socially and environmentally sustainable economy" (p. 53). For the purpose of our paper, we propose an inclusive definition of SE that embraces all ESG aspects and highlights the dual objective function, as follows:[10]

> *SE encompasses all entrepreneurial activity that, in addition to positive financial returns, aims to generate non-negative non-financial returns related to environmental, social and governance aspects..*

Existing work on SE is "truly multidisciplinary" (Hall et al., 2010, p. 441). In terms of the entrepreneurial lifecycle, a substantial and rapidly growing literature with het-

[9]Other very helpful reviews of sustainable entrepreneurship include Bischoff and Volkmann (2018), Dean and McMullen (2007), Gast et al. (2017), Kraus et al. (2018), Muñoz and Cohen (2018), Sarango-Lalangui et al. (2018), Schaefer et al. (2015), Shepherd and Patzelt (2011), and Terán-Yépez et al. (2020).

[10]In this sense, our definition abstracts from Cohen and Winn's (2007, p. 35) that focuses mainly on opportunities from environmental degradation, which itself is based on Venkataraman (2019): "*how opportunities to bring into existence 'future' goods and services are discovered, created, and exploited, by whom, and with what economic, psychological, social, and environmental consequences*."

erogeneous perspectives has emerged, dealing with antecedents of SE, SE opportunity recognition and execution, and SE outcomes, although outcomes are the least studied aspect of SE (Anand et al., 2021).[11]

Antecedents of SE can be distinguished at the individual and the contextual level (Anand et al., 2021; Kraus et al., 2018). *Individual antecedents* include the entrepreneur's personal intent and characteristics (Kimuli et al., 2020; Kuckertz and Wagner, 2010), with the consensus that sustainability-oriented entrepreneurs have salient moral and altruistic preferences (Ploum et al., 2018; Vuorio et al., 2018), display self-efficacy (Muñoz, Janssen, et al., 2018), sustainability-oriented values, beliefs and motivations (Jahanshahi and Brem, 2017; Mupfasoni et al., 2018; Spence et al., 2011), education and capabilities (Obrecht, 2011, 2016) and, in particular, prior knowledge (Mupfasoni et al., 2018). *Contextual antecedents* include environmental regulations, consumer awareness and demand (Hooi et al., 2016), other institutional enablers, such as social norms and market incentives (Meek et al., 2010; Pacheco et al., 2010; Shepherd and Patzelt, 2011), as well as local embeddedness, stakeholder involvement and collaborations (Schaltegger et al., 2018).

SE opportunity-identification processes are often analyzed through the lens of business model choices (e.g., hybrid (Davies and Chambers, 2018), transformative (Binder and Belz, 2017; Hahn et al., 2018), and sustainability-focused (Breuer et al., 2018) business models; and for an excellent overview, see Schaltegger et al., 2016). Sustainable business model studies often investigate trade-offs between financial and non-financial ESG-related returns, although the evidence is mixed (Anand et al., 2021; Schaltegger et al., 2016). In a widely-cited contribution, Parrish (2010) interviews 32 individuals and concludes that sustainability-oriented entrepreneurs have to employ "perpetual reasoning" to "succeed in a competitive market context" while conventional entrepreneurs can employ "exploitative reasoning," which leads to implications about organizational design choices for SE that "diverge in important ways from the conventional principles of entrepreneurship" (p. 510).

Finally, SE outcomes are arguably the least studied and most segmented field in the literature. SE outcomes refer to the performance of sustainability-oriented ventures

[11] Johnson and Schaltegger (2020) propose an alternative classification of the literature - SE processes, and SE challenges and opportunities. SE processes can span macro-social and global contexts, such as reducing economic inequality, fighting poverty and climate change (Muñoz and Cohen, 2017; Stål and Bonnedahl, 2016; Yunus et al., 2010)), within and between markets, such as counteracting the degradation of natural resources (Cohen and Winn, 2007; Dean and McMullen, 2007), and the timeline of venture development, such as SE formation, execution and managing the "triple bottom line" (Binder and Belz, 2017; Choi and Gray, 2008; Parrish, 2010; Stubbs, 2017). SE challenges and opportunities can be summarized at the macro level, such as poverty and climate change (Mair and Marti, 2009; Shepherd and Patzelt, 2011), at the meso level, such as helping local communities, e.g., with micro-financing or with ideas to reverse environmental degradation (Cohen and Winn, 2007; Dean and McMullen, 2007), and at the micro level, such as the resource mobilization and joint venturing initiatives (Desa, 2012; York et al., 2016).

in terms of the 'triple bottom line' (i.e., people, planet, profit). Although Anand et al. (2021) stress that there "is a need to engage more closely with the outcomes of SE activity" (p. 15), there are a few studies that tackle the outcome question. These studies fall broadly into two areas: *ESG impact* and *SE financing and investing performance*.

First, the *'ESG impact'* area is concerned with the contributions SE makes to ESG goals (e.g., Dickel, 2017; Djupdal and Westhead, 2015; Hoogendoorn et al., 2019; Jahanshahi and Brem, 2017; Kraus et al., 2017; Lans et al., 2014; Muñoz, Cacciotti, et al., 2018; Mupfasoni et al., 2018; Testa et al., 2019; Volkmann et al., 2021). The literature is limited in two important ways. First, the very nature of ESG goals (i.e., very long-term, partly subjective and context-dependent, and highly inter-dependent) make researchers confront the "major challenge" of coming to a consensus on "how to measure sustainability" (Anand et al., 2021, p. 12). Second, given that SE's historical emergence is tied to entrepreneurial opportunities that result from market failure to prevent environmental degradation (e.g., Cohen and Winn, 2007; Dean and McMullen, 2007), most of the work on ESG impact is limited to environmental impact (Anand et al., 2021).[12]

Second, and most important for the focus of our study, the *'SE financing and investing performance'* area "has a relatively short history" (Böckel et al., 2020, p. 433). The reason is that traditional players in the entrepreneurial finance market are often exclusively interested in financial rents (Block et al., 2018; Vismara, 2016), and thus "the lack of financing is a key obstacle that keeps the potential of sustainable entrepreneurship from being unleashed" but "crowdfunding is expected [...] to remove this obstacle" (Böckel et al., 2020, p. 435). A number of studies looks at the financing of SE, but aggregate evidence on the subject is rather limited. Cumming et al. (2016) find a positive relationship between venture capital activity and oil prices in the alternative energy sector ('cleantech'); Cumming et al. (2017) find that reward-based crowdfunding campaigns for cleantech projects on *Indiegogo* are more successful if the projects are not-for-profit and have a video pitch, whereas, whereas, using an overlapping sample from the same crowdfunding platform, Hörisch (2015) finds no relationship between environmental orientation and crowdfunding success; Calic and Mosakowski (2016) find some support for a positive relation between sustainability orientation and reward-based crowdfunding success in technology and film/video projects on *Kickstarter*; finally, Vismara (2019) shows that sustainability-oriented equity-based crowdfunding campaigns are less likely to attract professional investors. Overall, the literature on SE financing is relatively nascent, and a comprehensive analysis may help address several important voids in the literature, such as the "research gap related to the post-funding phase" (Böckel et al.,

[12]Böckel et al. (2020) contest the environment-bias argument in Anand et al. (2021), and argue that the society bias is more pronounced. Nevertheless, both reviews have in common that the governance aspect is entirely missing from the SE literature.

2020, p. 433).

2.2 ESG Investing

Sustainable (or impact) investing describes the practice of investors considering ESG when making investment and portfolio decisions. Sustainable investing is experiencing soaring growth (Gillan et al., 2021; Pástor et al., 2020). This is mainly driven by large net capital inflows that investment funds experience from institutional investors. For example, in 2019, mutual funds received more than $20 billion in net capital inflows, which increased fourfold over the previous year. Further, the Principles of Responsible Investments (PRI) initiative had $86 trillion assets under management in 2019 (up from $6.5 trillion in 2006), and more than 3,0000 institutional players in the financial market committed to the initiative in 2019.[13] Accordingly, most S&P500 firms recognize the increased demand of ESG, and 86% published separate sustainability or responsibility reports in 2018 (up from 20% in 2011) (Gillan et al., 2021).[14]

The ESG literature is limited in a number of important ways. First, several data providers offer ESG metrics, and the between-provider correlation is very low (Berg et al., 2020). Thus, there is substantial disagreement as to how ESG is measured, and different components are weighted to arrive at a composite measure. Second, the time horizon over which ESG activities should be measured is also unclear. Most ESG activities are long-term, however, to observe a significant impact of ESG metrics may take longer than a lifetime (e.g., activities aimed at stopping climate change). Third, as ESG is a relatively recent phenomenon and the market may be transitioning to a new equilibrium, it is unclear whether current studies measure a new steady state or simply a transitory, temporary state during the dynamic adjustment process. Fourth, the direction of causality is unclear, i.e., whether the underlying mechanism is 'doing well by doing good' or 'doing good by doing well.' Finally, ESG has been studied in many asset classes, such as stocks, bonds, bank loans and real estate. However, ESG is still largely missing from the entrepreneurial finance literature (with some notable exceptions, such as Cumming et al., 2016; Cumming et al., 2017; Vismara, 2019).

2.3 Token Offerings

Token offerings or Initial Coin Offerings (ICOs) are blockchain-based crowdfunding campaigns, in which investors wire fiat money or other cryptocurrencies via the blockchain and receive tokens from the fundraising venture. The transaction is fully

[13] https://www.morningstar.com/articles/961765/sustainable-fund-flows-in-2019-smash-previous-records

[14] https://www.ga-institute.com/press-releases/article/flash-report-86-of-sp-500-indexR-companies-publish-sustainability-responsibility-reports-in-20.html

automated by a smart contract, often on the Ethereum blockchain via the ERC20 standard. Tokens are often categorized in three ways: (i) cryptocurrency tokens, such as *Bitcoin*, are mere mediums of exchange, (ii) utility tokens are payment instruments that investors can redeem for an issuing venture's product or service once it is developed and on the market, and (iii) security tokens are equity-like instruments that give investors control rights. Shortly after the offering, projects typically list their tokens on liquid exchange platforms, enabling investors to trade tokens with one another (Adhami et al., 2018; Bellavitis et al., 2020; Fisch, 2019; Momtaz, 2020a, 2020b). To our knowledge, we are among the first to look at token offerings to examine the funding success and post-funding performance of SE vs. CE projects.[15]

Token offerings are an ideal playing field to shed more light on the financing of sustainability-oriented startups for at least two reasons. First, individual investors with simultaneous financial and non-financial investment goals predominantly populate the market for token offerings (Fisch et al., 2019). Like in crowdfunding (e.g., Giudici et al., 2018), token offerings were born out of disappointment with the fairness of traditional financial markets (Fisch et al., 2020; Howell et al., 2020; Nakamoto, 2019). Therefore, investors in token offerings may be particularly sensitive to the sustainability orientation of potential investment objects. Second, unlike any other entrepreneurial finance mechanism, institutional features surrounding token offerings facilitate a quantitative analysis of ESG and startup valuation and performance. Specifically, (i) it is standard practice that projects in token offerings publish extensive whitepapers disclosing important information, such as how they aim to solve ESG challenges, and (ii) listing tokens on exchange platforms post-offering enables the project's financial performance to be tracked on a daily basis and in a transparent way by observing equilibrium prices formed by supply-and-demand dynamics in liquid markets. As Böckel et al. (2020) discuss, the post-funding performance of sustainability-oriented startups is an important "research gap" (p. 433). Thus, fair prices obtained from liquid token exchange markets that provide a transparent measure of post-funding performance can help close this gap.

3 Hypotheses

3.1 ESG and Funding

Like their conventional counterparts, sustainable entrepreneurs identify an entrepreneurial opportunity[16] and tap entrepreneurial finance markets for funding. Un-

[15] See, also, Guzmán et al. (2020), for a concurrent study with a narrower focus on global warming.

[16] Patzelt and Shepherd (2011) highlight the role of entrepreneurial knowledge of natural and communal environments for sustainability-related opportunity recognition.

like conventional entrepreneurs, however, sustainable entrepreneurs' funding success is not only determined by the expected cash flows that investors may receive in the future but also by the expected non-financial utility (Block et al., 2021; Vismara, 2019). The literature offers two potential reasons as to why sustainable entrepreneurs may benefit from higher valuations during the funding stage: the economics of delegated philanthropy and the signaling value associated with ESG properties.

The economics of delegated philanthropy. Friedman's (1970) famous proclamation that 'the social responsibility of business is to increase its profits' is often used as an argument against ESG/CSR[17] initiatives. However, Friedman's (1970) theoretical argument is based on sophisticated assumptions: (i) markets are competitive, (ii) the regulatory framework is able to internalize external costs, (iii) companies do not have a competitive advantage vis-à-vis their shareholders to do good, and (iv) companies cannot influence regulation. Under these assumptions, corporate ESG initiatives do not add investor value. However, these assumptions are usually violated in reality, potentially providing ESG strategies with a business case.

For example, if investors also have ESG preferences and financially profitable activities cannot be perfectly separated from ESG-detrimental ones (i.e., a violation of Friedman's third assumption), then companies should indeed maximize investor "welfare" (as compared to "value") (O. Hart and Zingales, 2017). In these situations, and in line with Friedman (1970), companies should augment the business objective and include ESG goals in addition to the financial return. An example for such 'delegated philanthropy' would be a startup involved in the production of 3-D printers that enable customers to produce assault rifles. Assuming that investors prefer anti-gun legislation, the startup could pay investors a dividend, which they themselves could then donate to anti-gun initiatives. However, it would be more efficient if the startup would not sell its 3-D printers to facilitate the production of guns in the first place. While this hurts profits, it serves the greater social goal of the anti-gun movement, and could maximize total (financial and non-financial) shareholder utility (O. Hart and Zingales, 2017).[18]

Empirical evidence suggests that 'doing well by doing good' can work. Traditional financial market theory assumes that investor preferences for future consumption determine the financial market equilibrium (Fama and French, 2007). However, if investors incorporate ESG preferences into their utility models (for evidence of this, see Shepherd et al., 2009; Shepherd et al., 2013), then valuations and expected returns can deviate from the equilibrium suggested by the standard models (Cornell, 2021; Pástor et al., 2020). Therefore, investors with ESG preferences drive up demand for ESG assets, which increases their prices, lets cost of capital decrease, and ultimately

[17] CSR stands for Corporate Social Responsibility.

[18] See also the works by Cheng et al. (2014), Eccles et al. (2014), and Ioannou and Serafeim (2015) that relate ESG/CSR to access to finance, performance, and analyst recommendations.

makes it cheaper to invest in ESG projects. If consumers incorporate ESG considerations into their 'willingness-to-pay' models (Barber et al., 2021) as well, ESG companies would also profit from higher cash inflows. Additionally, Edmans (2011) finds that employee satisfaction (a measure of G in ESG) increases corporate productivity, and Lins et al. (2017) and Albuquerque et al. (2020) report that ESG policies create trust and loyalty among customers, which acts as insurance during economic downturns. Therefore, the economics of delegated philanthropy suggests that, under certain assumptions, sustainability-oriented entrepreneurs may receive higher valuations thanks to the add-on non-financial utility they generate for investors with pronounced ESG preferences.

ESG-related signaling. Several papers establish the importance of signaling venture quality for the funding success in token offerings (e.g., Fisch, 2019; Momtaz, 2020b, 2021a). Building on these findings, we explore additional signaling dimensions that are proprietary to sustainable entrepreneurship. There are at least five such arguments. First, a key concern for investors in token offerings is moral hazard in signaling (Momtaz, 2020a) and outright fraud (Hornuf et al., 2021). Sustainability orientation on the part of the entrepreneur may signal non-financial motives, which reduces investor concerns and creates trust (Kraus et al., 2018). Second and similarly, the ESG orientation signals management team's awareness for broader issues than just the narrow business scope, which may help foresee and prevent adverse events. Thus, sustainability orientation may be correlated with broad awareness for strategic developments, and therefore valuable from a risk management perspective (Kraus et al., 2018). Third, given that younger generations are well represented on crypto markets and empirical studies show that these generations have pronounced ESG orientations (more so than older generations), sustainable entrepreneurs may create a sense of identification among these younger investment groups (Fisch et al., 2019; Kraus et al., 2018; Spence et al., 2011). Fourth, sustainability orientation may act as an insurance mechanism. Given the highly dynamic and competitive token offerings market, a key risk for entrepreneurs and investors is early project competition (or imitation). The ESG profile of sustainable entrepreneurs may help preserve the USP and help retain customer base or growth share (when a similar but non-ESG competitor threatens), thereby reducing this source of risk (Anand et al., 2021; Johnson and Schaltegger, 2020). Finally, and very importantly, ESG awareness has been shown to be correlated with human, social and intellectual capital, which are first-order determinants of funding success in startup financing (Ahlers et al., 2015; Fisch, 2019; Spence et al., 2011).

To summarize, the above rationale may theoretically justify a sustainability premium for high-ESG ventures. SE financing may be positively influenced in three ways: (i) SE may receive more funds thanks to an expanded market size (i.e., high-ESG otherwise non-investors), (ii) SE may steal investors away from CE but otherwise similar ven-

tures, and (iii) SE may benefit from increased willingness-to-invest among high-ESG investors thanks to the non-financial utility they may receive. Such a sustainability premium could be particularly pronounced in the context of token offerings, which is arguably populated by investors with salient non-financial preferences (Fisch et al., 2019; Schückes and Gutmann, 2020).

H1: *The relationship between ESG properties and startup firm valuation is positive.* (The Valuation Premium Hypothesis, VPH)

3.2 ESG and Post-Funding Performance

How does sustainable entrepreneurship perform after the fundraising campaign compared to conventional entrepreneurship? As Hall et al. (2010) discuss, "while the case for entrepreneurship as a panacea for transitioning towards a more sustainable society is alluring, there remain major gaps in our knowledge of whether and how this process will actually unfold." The financial performance is one such "major gap," as we are not aware of any study that has examined the relationship between ESG and long-term financial performance in the entrepreneurial context. Indeed, Böckel et al.'s (2020, p. 433) recent review of studies at the intersection of crowdfunding and sustainability concludes that a major "research gap related to the post-funding phase" exists. Even more generally, the post-financing performance of token offerings and crowdfunding is probably the "least explored" topic (Vanacker et al., 2019, p. 237), not even considering the question of sustainability.

Not many, but a few notable studies look at the post-funding performance of crowdfunded startups. Mollick and Kuppuswamy (2014) report that reward-based crowdfunded ventures on *Kickstarter* in the period 2009-2012 added on average 2.2 new employees (with a standard deviation greater than 9) and 32% of the firms had revenues in excess of $100,000. Iyer et al. (2016) studies lending-based crowdfunding and reports a post-funding default rate of 30%, which clearly exceeds the average return, indicating that lending-based crowdfunding campaigns underperform traditional lending markets. Signori and Vismara (2018) look at 212 crowdfunding campaigns and show that only 3 of them exited successfully through an acquisition. Walthoff-Borm et al. (2018) provide very interesting findings by comparing equity-based crowdfunding campaigns on *Seedrs* and *Crowdcube* in the UK. They report lower financial performance, measured as returns on assets, relative to non-crowdfunded startups. Importantly, they compare the returns in ventures, in which investors become direct shareholders to those in which they become indirect shareholders (i.e., *Seedrs* uses a nominee structure in which the platform holds and manages the shares). They find that direct shareholdings, which is more comparable to our token offerings context, are more likely to lose and less likely

to invest in intangibles. Thus, Walthoff-Borm et al. (2018) is the only crowdfunding study that may suggest that startups with salient ESG attributes (i.e., intangible goals) may underperform post-funding.

Studies on the long-term performance of token offerings are also rare. Momtaz (2021c) studies the performance of cryptocurrencies issued in token offerings over a three-year holding period, and reports that larger ventures underperform. To the possible extent that the sustainability premium (which inflates venture size via the sustainability-related valuation premium) contributes, this finding may suggest that sustainable entrepreneurs are more likely to underperform. Fisch and Momtaz (2020) study the involvement of institutional investors on ventures' post-ICO performance, and find that the relationship is positive. Given that institutional investors focus on financial performance and shy away from ESG startups (Vismara, 2019), the finding may also indicate that ventures focusing on ESG may underperform. However, we have to attest to the lack of work on the post-funding performance of crowdfunded startups, and acknowledge that the existing work in entrepreneurial finance is *at best* vaguely indicative of SE underperformance.

Given this lack of prior work to build upon, we draw on the broader ESG investing literature (for a review, see Gillan et al., 2021).[19] The overarching tenet is that ESG commitment poses a *binding constraint* that may restrict managerial agility and therefore depress financial performance (Barber et al., 2021; Cornell, 2021). This may be of particular importance in the entrepreneurial context, where product market-related hypothesis testing and frequently changing directions is of paramount importance (Johnson and Schaltegger, 2020; Kraus et al., 2018). Thus, many studies argue that sustainability is at odds with capitalist societies (e.g., Balakrishnan et al., 2003).

In the ESG investing literature, the consensus is clear: High-ESG investments *under*perform (Gibson et al., 2020; Liang and Renneboog, 2017; Renneboog et al., 2008). This is because ESG commitment creates a binding restriction on portfolio choice, which leads to under-diversification and, in turn, hurts the risk-return trade-off. Moreover, equilibrium asset pricing theory suggests that high valuations are mechanically related to lower expected returns (Campbell et al., 2012; Fama and French, 2007). For these reasons, if there is a sustainability premium, as hypothesized in ***H1***, then sustainable entrepreneurs should have a negative long-term performance.

[19] It is important to note that there is also no consensus on the performance question in the sustainability literature itself (Anand et al., 2021). Here are two examples: First, sustainability-oriented ventures may perform better thanks to an increased market size (i.e., gaining high-ESG preference customers), yet they may also underperform because they lose low-ESG groups that are, e.g., unwilling to pay more for ESG products (Hörisch, 2015; Kraus et al., 2018). Second, some argue that ESG provides intrinsic motivation to entrepreneurs that may boost performance, while others argue that ESG-driven entrepreneurs are "dreamers" and unlikely to be successful businessmen (Edmans, 2011). In both cases, the net effect of sustainability orientation is not clear, which makes the topic "controversial" (Hall et al., 2010, p. 439).

H2: *The relation between ESG properties and post-funding performance is negative.* (The Post-Funding Underperformance Hypothesis, PFUH)

3.3 Technological, Network-, and Governance-Related Formalization

Prior work on sustainable entrepreneurship shows that integrating sustainability aspects into the venture model creates a high degree of formalization (i.e., binding constraints) (Kraus et al., 2018). For example, control systems, reporting procedures, process disclosure requirements and policies that track behavior, such as efforts to become CO_2 neutral or obtain green certificates from environmental non-profit organizations, all counteract with entrepreneurial flexibility to foster intuitive management styles that help manage the start-up process in an agile manner to reduce execution risk (Kraus et al., 2018; Spence et al., 2011).

The high degree of formalization is "counterintuitive and potentially disadvantageous" and can "potentially be hazardous" (Kraus et al., 2018, p. 8) to the success and survival of ventures for several reasons. First, formalization requires venture teams to adhere to policies and standards put in place, which has a prolonging effect on the time horizon after which the venture can start harvesting the fruits of its efforts. In particular, in the highly dynamic and competitive startup space, when sustainable entrepreneurs have more uncertain and long-term goals than conventional entrepreneurs, the high degree of formalization poses the risk that the venture may 'die along the way' (Spence et al., 2011). Second, prior work shows that sustainable entrepreneurs are more risk averse (Weerawardena and Mort, 2006), which reflects an attitude that is often associated with lower entrepreneurial success, because entrepreneurial exploration inherently requires substantial risk-taking (S. P. Kerr et al., 2017; W. R. Kerr et al., 2014). The high degree of formalization in sustainable ventures may amplify an entrepreneur's risk aversion, as policies and standards may provide a narrative that justifies not taking additional risk to some extent (Spence et al., 2011). Thus, various dimensions of formalization, such as technical, network and governance formalization, may be negatively related to the relationships between sustainability orientation and ventures' valuations and performances.

H3a: *The proposed positive relationship in **H1** is less pronounced when the sustainable entrepreneur's degree of technological, network and governance-related formalization is high.*

H3b: *The proposed negative relationship in **H2** is more pronounced when the sustainable entrepreneur's degree of technological, network and governance-related formalization*

is high.

4 Quantifying Startups' ESG Properties

4.1 ESG Measurement in Existing Studies

Existing studies measure startups' ESG properties relatively ad-hoc, and a unified framework is missing so far from the literature. For example, Vismara (2019) regresses the dummy variable "sustainability orientation" on the funding amount in crowdfunding campaigns, which is based on whether the projects' descriptions include at least one of the following terms: "sustainability," "sustainable," "ecological," "eco-innovation," "eco-efficient," "eco-effective," "eco-design," "ecology," "environmental," "green," "renewable," "cradle to cradle," "dematerialization," "backcasting," "biomimicry," "jugaad innovation," circular economy," and "closed-loop production;" Hörisch (2015) uses entrepreneurs' self-classification as "environmentally oriented" on the crowdfunding platform *Indigogo*; and Guzmán et al. (2020) regress the global frequency of Google searches with the search term "global warming" without any concrete reference to their specific sample.

We hope to address this problem by offering an integrated machine-learning approach that quantifies startups' ESG properties from text data (e.g., press releases, whitepapers, *Github* documentation, text on their own website as well as on others, such as *Crunchbase*, among others). Broad adoption of our approach would increase the comparability of results across ESG studies (Gentzkow et al., 2019; Li et al., 2020; Loughran and McDonald, 2020a), and reduce subjectivity of ESG measurement in the literature (Berg et al., 2020; Dimson et al., 2020).

4.2 ESG Measurement: A Machine-Learning Approach

Our goal is to measure startups' ESG properties in a relatively objective way from text data (i.e., the information disclosed by startups during their fundraising campaigns). Our approach is in the spirit of the broader "text as data" literature in economics, as reviewed in Gentzkow et al. (2019), which relies on word counts based on topic-specific dictionaries (or word lists). Therefore, our task involves two steps:

1. Creating an ESG-specific dictionary in the startup context
2. Measuring the (normalized) prevalance of ESG cues for each startup ("ESG scores")

For brevity, we defer a comprehensive discussion of our machine-learning approach to the Appendix A, as well as to our source code website on *GitHub*. Here, we summarize the main tenets relevant for understanding our approach and interpreting the results reported in section 6.

4.2.1 ESG Dictionary

An important motivation for creating a novel ESG dictionary using a machine-learning approach comes from the observation that existing ESG ratings are highly subjective, leading to very low correlations between different ratings (Berg et al., 2020; Dimson et al., 2020). Additionally, given the non-standardized nature of startup information disclosure, existing (corporate) ESG ratings cannot be reliably applied to startups. Therefore, our machine-learning approach both (i) helps mitigate the subjectivity bias in ESG ratings and (ii) introduces a replicable "text as data"-based method that derives reliable ESG ratings for startups.

In a first step, we use the Stanford CoreNLP pipeline (Manning et al., 2014) to obtain a dependency representation of each sentence in every whitepaper to help the machine learn the grammatical structure of information that startups typically publish in whitepapers. In particular, we teach the machine to identify "collocations," such as *initial_coin_offering*, which treats conjugate terms as one term. These collocations become important in our second step, as we use a one-hidden-layer neural network (i.e., *word2vec* based on Mikolov et al., 2013[20]) to train the model to predict neighboring collocations, which help to quantify language by creating vectors of real numbers for any dictionary word. For example, using this approach, one could find the closest vector for "ICO" as follows: ICO $=$ STO $-$ security token $+$ utility token.

Following Li et al. (2020), we provide seed words as initial starting points to help the machine to create the ESG dictionary. Specifically, we collect all available *Financial Times* (*FT*) articles with the tags "ESG Investing" or "Moral Money." We follow a standard bag-of-words approach and extract the most frequent bi-grams and tri-grams (two and three-word combinations) that appeared in the pre-selected *FT* corpus. Then, we manually go through these bi-/tri-grams and map them to the best-fitting E, S, or G dimension of ESG. Given *FT*'s focus on larger corporations, we manually add terms like 'kyc' and 'whitelist' (as examples for the G dimension). For replication purposes (and potential modification in future studies), the full list of seed words is available in the Appendix Table A.1. Our seed words consist of 70 E, 38 S, and 46 G-related terms. We also test the sensitivity of our main results to ESG scores obtained from dictionaries with other seed words, and find robust results.

For any term t of the seed words in any of the ESG dimensions j, we obtain a

[20]For a critical discussion of *word2vec*, see Nissim et al. (2020).

vector representation with the size of 300 (the size of the hidden layer in our *word2vec* model) as $V^t_{j\in\{E,S,G\}} = [x^t_1, x^t_2, ..., x^t_{300}]$. We then calculate the average vector for each {E, S, G} dimensions as $\bar{V}^{j\in\{E,S,G\}} = \frac{1}{N}\sum_1^N [x^t_1, x^t_2, ..., x^t_{300}]$ where N is the size of seed words for the dimension j. This leaves us with three vectors of $\bar{V}^E$, $\bar{V}^S$, and $\bar{V}^G$. Finally, we perform a cosine similarity between $\bar{V}^j$ and the vector of all the terms in our whitepapers database, which leaves us with a total of 1,495 ESG-related terms consisting of 508, 463 and 524 terms in the respective ESG dimensions. Figure 1 illustrates the word clouds corresponding to the E, S and G word lists.

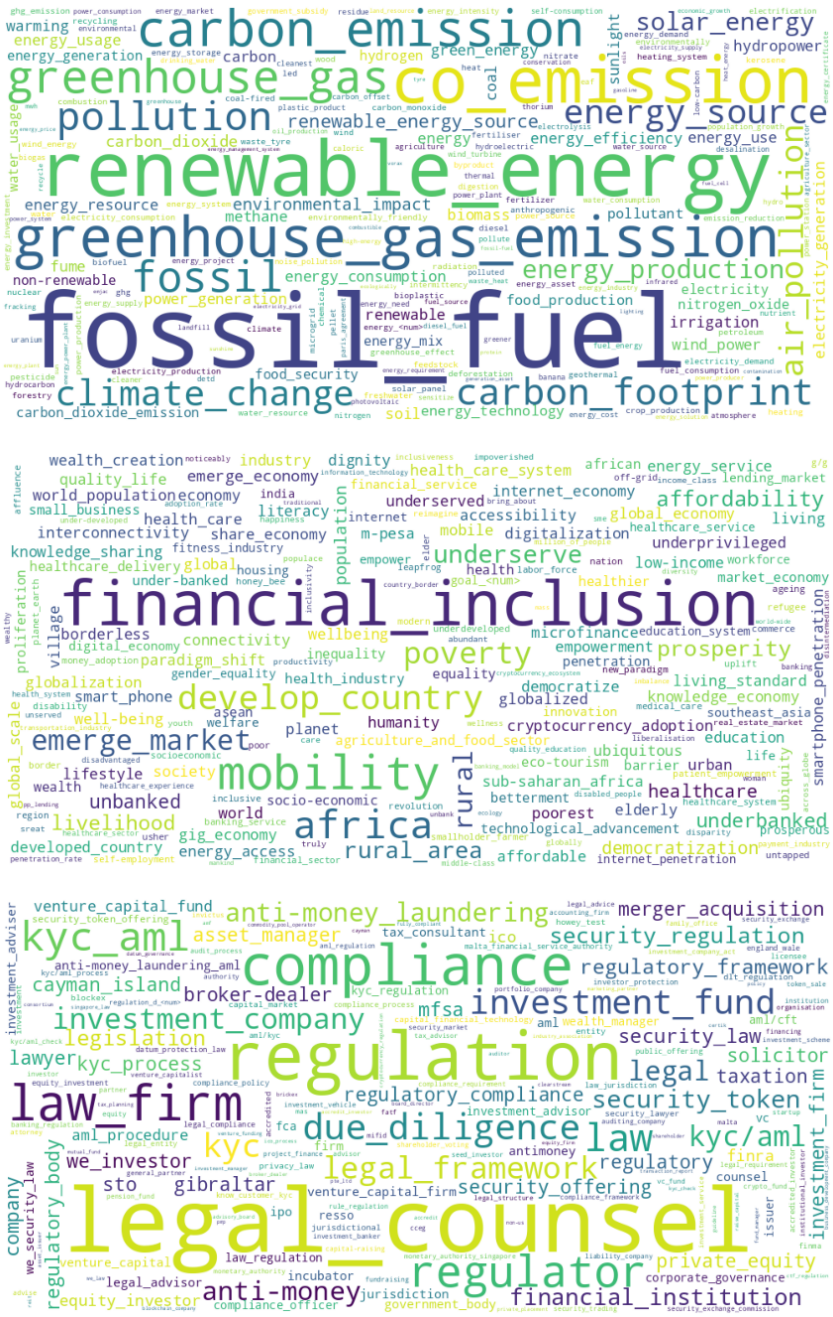


Figure 1: Relative Importance of Terms in ESG Dictionary by {E,S,G} Dimension

4.2.2 ESG Score

We use our ESG dictionary to quantify the E, S and G dimensions by counting the number of distinct occurrences of our dictionary words in whitepapers, normalized to the size of the word list. Specifically, for token offering i, we measure each dimension ζ of ESG as:

$$\zeta_i = \frac{\sum_t 1_{c(t)_i > 0}}{c(n)} \text{ for } \zeta \in \{\text{E, S, G}\} \tag{1}$$

where $c(t)_i$ denotes the count of term t in whitepaper i and $c(n)$ is the size of the corresponding word list. Thus, our approach adapts that of Loughran and McDonald (2020a) to account for the non-standardized nature of whitepapers relative to the highly standardized and regulated use of language in corporate disclosure reports analyzed by Loughran and McDonald (2020a). The aggregate ESG score of startup i is then simply described by the sum of its components:

$$ESG_i = \sum_{\{E_i, S_i, G_i\}} \zeta_i \tag{2}$$

4.2.3 Sanity Checks

We perform manual sanity checks to make sure that our approach identifies startups' ESG properties reliably. The results are reconfirming. For example, the startup with the highest environmental score in our sample is *WPP Energy* (funding amount: $59M). *WPP Energy* is "a Swiss Company that over the last decade has established itself as a repository for disruptive energy and environmental technologies through exclusive global licenses." Similarly, the second-highest environmental score in our sample belongs to *Greencoin* (funding amount: $6M), which is "the first decentralized platform based on sustainable green systems to solve real problems in the world, connecting green systems manufacturers and local Installation companies or certified individuals directly with buyers." Careful examinations of *WPP Energy*'s and *Greencoin*'s whitepapers show that these startups are indeed concerned with addressing salient environmental problems. Similarly, we confirm that our approach correctly identifies the S (e.g., the startups *HARA* and *Ubricoin*) and G (e.g., the startups *SMART VALOR* and *Chainium*) dimensions of ESG.

4.2.4 Web Application

In an effort to facilitate the use of our ESG machine-learning approach in future research, we created a web app that computes ESG scores for text data based on our Python code via simple copy and paste:

www.SustainableEntrepreneurship.org

The Python source code as well as comprehensive and relatively technical documentation of our machine-learning approach for measuring ESG in the entrepreneurial context is provided in the Appendix A, as well as on our *GitHub* project page.

5 Methods

5.1 Data Sources

Our sample is based on the *Token Offerings Research Database* (*TORD*).[21] *TORD* offers the most comprehensive publicly available token offerings database, and therefore addresses some of the key concerns about data limitations in regards to token offerings (for a comprehensive discussion of these concerns, see section 6.4 in Momtaz, 2020a). Our empirical analyses exclude Security Token Offerings (STOs) and Initial Exchange Offerings (IEOs) to avoid biases from various confounding factors that would relate to the governance of these alternative token and offering types, and therefore only sample from utility-token ICOs. For these ICOs, we manually collect whitepapers from the firms' websites, *ICObench*, and the internet archive via the *Wayback Machine* (https://web.archive.org/). Finally, we collect post-ICO token prices from *CoinMarketCap*. We only include token offerings with a complete set of variables, as described in section 5.2, in our final sample. Our final sample consists of 1,043 token offerings.

5.2 Variables

Our independent variables are the startups' ESG properties, which we derive from their whitepapers using a machine-learning approach. We describe the independent variable construction in detail in section 4 and in the Appendix A.[22] Below, we focus on the definitions of our dependent and control variables.

5.2.1 Dependent Variables

Our two dependent variables are the *valuation* of the startup during the funding stage and the *post-funding financial performance*.

Funding valuation. Following existing studies on ICO performance (e.g., Fisch, 2019), we operationalize startup valuation as the logarithmic funding amount in $ million acquired during the token offering.

Post-funding performance. We operationalize the post-funding performance with the 12-month Buy-and-Hold Abnormal Returns (BHARs), following Fisch and Momtaz (2020) and Momtaz (2021c). Specifically, we compute the 12-month return for each

[21] We use Version 1 of the TORD, retrieved on April 1, 2021 at www.paulmomtaz.com/data/tord.

[22] We normalize the ESG scores (mean = 0, standard deviation = 1) so that they are easy to interpret.

startup with regard to the listing date and subtract the performance of an equally-weighted market benchmark for the same investment period. The equally-weighted market benchmark is based on all tokens that are tracked on *Coinmarketcap*. The equally-weighted market benchmark has the important advantage that it deals with the size anomaly in market returns associated with the Bitcoin and Ether-related dominance in value-weighted market benchmarks, as first described in Momtaz (2021c).

5.2.2 Control Variables: Venture Characteristics

We control for the following venture characteristics: Whitepaper length, team size, rating, technical experience, minimum viable product, open source code, and # industries.

Whitepaper length. The natural logarithm of total words in any given whitepaper, which is often used as a proxy for the total information available about a project (e.g., Fisch, 2019).

Team size. The number of team members, which is a first-order determinant of success in token offerings (Fisch, 2019; Momtaz, 2020b).

Rating. The overall project rating based on the consensus of industry experts on ICObench, and is an important predictor of success in token offerings (Bellavitis et al., 2020; Fisch, 2019; Momtaz, 2020b). The scale runs from 1 ("low quality") to 5 ("high quality").

Technical experience. This is the percentage of team members with a technical background. The variable is hand-collected from team members' professional network profiles, such as *LinkedIn*.

Minimum viable product. This is a dummy variable for whether a startup has a minimum viable product available.

Open source code. Coded as a dummy variable for whether the startup discloses its code on *Github*, which is often used as a proxy for a venture's technological sophistication (Fisch, 2019).

 # Industries. We use *ICObench* industry classifications to measure the potential industries the focal venture targets as the logarithm of one plus the number of the industries, which is a proxy for diversification (Fisch and Momtaz, 2020).

5.2.3 Control Variables: Offering Characteristics

We control for the following offering characteristics: Soft and hard caps, pre-sale, whitelist, bonus, bounty and ERC20.

Soft cap. A dummy variable for whether the startup has announced a soft cap in its token offering. A soft cap is the minimum funding amount at which the offering is deemed successful, and funding campaigns that fail to reach the soft cap typically redeem investor money and end the project.

Hard cap. A dummy variable for whether the startup has announced a hard cap in a token offering. A hard cap is the maximum funding amount that a startup accepts. If the hard cap is reached, the offering will end and excess funding will be returned to investors.

Pre-sale. A dummy variable indicating if the actual token offering was preceded by a pre-sale event.

Whitelist. A dummy indicating if the token offering has an active whitelist.

Bonus. A dummy variable for whether the startup offers a bonus structure, which typically involves discounted or free tokens if individual wallet addresses invest above and beyond a certain pre-determined investment amount.

Bounty. A dummy variable for whether the token offering offers a bounty program, which rewards individuals (mostly in the form of free tokens) for marketing activity that promotes the offering and the startup.

ERC20. A dummy variable for whether the token offering relies on the technical ERC20 standard.

5.2.4 Control Variables: Market Characteristics

We control for whether token offerings are launched during bull or bear markets, with market stagnation serving as the base case.

Bull market. A dummy variable for whether the token offering took place during a bull market, i.e., prior to the so-called "crypto winter."

Bear market. A dummy variable for whether the token offering took place during a bear market, i.e., during the "crypto winter."

5.3 Summary Statistics

Table 1 provides summary statistics and bivariate correlations for all of our main variables. The average startup in our sample raises $15.2 million during the token offering with a team of 12.9 people and an average rating of 3.4 (out of 5). More than 4 out of the 12 team members have a technical background. Two-thirds of all startups publish code on *Github*, but only one-fifth of all startups has a minimum viable product at the time of the token offering. These sample statistics resemble those in related studies (e.g., Bellavitis et al., 2020; Fisch, 2019; Howell et al., 2020; Huang et al., 2020; Momtaz, 2020a).

The bivariate correlations indicate that the *aggregate* ESG score is positively correlated with the funding amount ($\rho = 0.238$) and negatively with the post-funding performance ($\rho = -0.107$). The *disaggregated* ESG scores shed further light on what sustainability aspects matter for the funding amount and the post-funding performance. The funding amount is positively correlated with E ($\rho = 0.098$), S ($\rho = 0.213$) and G

(ρ = 0.240), indicating that, of all the ESG aspects, environmental aspects are the least correlated with the funding amount. The post-funding performance is negatively correlated with E ($\rho = -0.045$), S ($\rho = -0.094$) and G ($\rho = -0.110$), with E again being the most weakly correlated ESG aspect with the post-funding performance. Overall, these correlations are in line with our two overarching hypotheses. It is also reconfirming that all disaggregated ESG scores are consistent in terms of their correlation coefficients' signs. For brevity, we note that the remaining correlations are largely consistent with those reported in existing studies (e.g., Fisch and Momtaz, 2020).

Table 2 shows means for the full sample in the first column and differences for subsamples with above-mean ESG scores relative to the full sample in the remaining columns. In line with our two main hypotheses, the average funding amount is higher in high-ESG startups, with the difference being statistically significant at the 1% level; and, the post-funding underperformance, measured as the 12-month holding period return adjusted by an equally-weighted market benchmark, is 16% higher, although the difference is not statistically significant in the univariate comparison.

We also shed some light on whether there is "selection on observables" in our sample by comparing the means for our control variables in the full sample with those in the subsamples. Indeed, we find some statistically significant differences between the full sample and the highly sustainable subsamples. For example, high-ESG startups have, on average, more than two additional team members, with the difference being highly statistically significant. Moreover, high-ESG startups set higher soft and hard caps, and are more likely to conduct a pre-sale and have a whitelist. Furthermore, they are less likely to conduct the token offering during a bull market and more likely to conduct it during a bear market, possibly indicating that sustainable startups are less sensitive to market opportunism.

Overall, these significant differences between low and high-ESG startups suggest that we need to control for selection issues in our sample. Next, we discuss two ways in which we control for selection based on observed and unobserved heterogeneity.

5.4 Econometric Approach

Our goal is to estimate the causal effect startups' ESG properties have on their funding success and post-funding performance. In addition to OLS models, we rely on several two-stage approaches.[23] These models control for observed and/or unobserved heterogeneity, which is often pronounced in entrepreneurial finance.[24]

Specifically, we are interested in the causal effect that startup i's ESG score, ESG_i,

[23]The techniques used in our study have been employed before in similar contexts (e.g., Bertoni et al., 2011; M. G. Colombo and Grilli, 2010; Fisch and Momtaz, 2020).

[24]For example, Momtaz (2021b) finds that unobserved heterogeneity in startups' time-to-funding by venture capitalists is so pronounced that it severely biases common time-to-event models.

Table 1: Descriptive statistics and correlations

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.
Mean	0.000	0.000	0.000	0.000	15.158	-0.534	8.100	3.393	12.924	0.254	0.203	0.661	1.291	0.619	0.884	0.540	0.312	0.008	0.313	0.802	0.325	0.688
SD	1.000	1.000	1.000	1.000	1.912	1.163	0.661	0.587	7.952	0.202	0.403	0.474	0.489	0.486	0.320	0.499	0.463	0.087	0.464	0.399	0.469	0.463
Q1	-0.726	-0.478	-0.668	-0.768	14.215	-1.012	7.775	3.000	7.000	0.091	0.000	0.000	0.693	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
Median	-0.072	-0.277	-0.044	-0.083	15.429	-0.346	8.151	3.400	12.000	0.250	0.000	1.000	1.099	1.000	1.000	1.000	0.000	0.000	0.000	1.000	0.000	1.000
Q3	0.586	0.058	0.579	0.655	16.524	-0.085	8.496	3.900	17.000	0.375	0.000	1.000	1.609	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000
Key variables:																						
1. ESG Score (normalized)																						
2. E-Score (normalized)	0.648																					
3. S-Score (normalized)	0.894	0.419																				
4. G-Score (normalized)	0.815	0.217	0.651																			
Dependent variables:																						
5. Funding amount, in $m	0.238	0.098	0.213	0.240																		
6. BHAR, 12-mo (equally weighted)	-0.107	-0.045	-0.094	-0.110	0.058																	
Control variables: Venture characteristics:																						
7. Whitepaper length, in (log-words)	0.657	0.309	0.655	0.562	0.226	-0.071																
8. Expert rating	0.219	0.055	0.212	0.232	0.112	-0.079	0.292															
9. Team size, in # FTE	0.297	0.068	0.292	0.316	0.169	-0.034	0.289	0.397														
10. Technical background, in %	0.001	-0.010	0.010	0.000	0.086	0.025	0.069	-0.034	0.050													
11. Minimum viable product (dummy)	0.058	0.064	0.040	0.038	-0.100	-0.138	0.054	0.344	0.180	-0.043												
12. Open source (dummy)	0.054	0.033	0.082	0.011	-0.072	-0.114	0.140	0.363	0.146	0.037	0.221											
13. # Industries (log)	0.064	0.072	0.045	0.040	-0.065	-0.180	0.070	0.240	0.160	-0.016	0.213	0.106										
Control variables: Offering characteristics:																						
14. Soft cap (dummy)	0.141	0.097	0.122	0.114	-0.109	-0.107	0.078	0.219	0.144	-0.120	0.219	0.160	0.169									
15. Hard cap (dummy)	0.126	0.066	0.110	0.119	-0.016	-0.027	0.130	0.225	0.131	-0.038	0.131	0.126	0.093	0.363								
16. Pre-sale (dummy)	0.123	0.083	0.094	0.115	-0.050	-0.094	0.108	0.237	0.179	-0.054	0.117	0.102	0.174	0.207	0.176							
17. Whitelist (dummy)	0.180	0.090	0.145	0.186	0.055	-0.150	0.175	0.238	0.229	0.018	0.195	0.084	0.156	0.161	0.121	0.094						
18. Bonus (dummy)	-0.030	-0.033	-0.007	-0.035	-0.005	0.028	-0.019	0.008	0.024	-0.001	-0.017	0.017	0.032	-0.022	0.032	-0.007	0.036					
19. Bounty (dummy)	0.067	0.079	0.059	0.025	-0.119	-0.163	0.062	0.258	0.153	-0.062	0.430	0.160	0.215	0.222	0.167	0.183	0.203	0.012				
20. ERC-20 standard (dummy)	0.050	0.002	0.033	0.077	-0.063	-0.147	0.030	0.102	0.088	-0.017	0.108	0.034	0.099	0.080	0.067	0.057	0.080	0.016	0.102			
Control variables: Market characteristics:																						
21. Bull market (dummy)	-0.150	-0.119	-0.131	-0.107	0.131	0.226	-0.093	-0.261	-0.204	0.112	-0.305	-0.116	-0.203	-0.341	-0.234	-0.205	-0.396	-0.014	-0.380	-0.235		
22. Bear market (dummy)	0.149	0.097	0.113	0.141	-0.022	-0.245	0.083	0.183	0.200	0.012	0.149	0.051	0.177	0.266	0.202	0.168	0.314	-0.036	0.293	0.210	-0.638	

Table 2: Are ESG startups different?

	Sample mean for all startups	*Differences in subsamples: Δ All startups − ...*			
		...high-ESG[1]	*...high-E*	*...high-S*	*...high-G*
Key variables:					
ESG Score (normalized)	0.0	0.773***	0.594***	0.749***	0.689***
E-Score (normalized)	0.0	0.388***	0.493***	0.333***	0.207***
S-Score (normalized)	0.0	0.707***	0.549***	0.821***	0.57***
G-Score (normalized)	0.0	0.705***	0.375***	0.578***	0.806***
Dependent variables:					
Funding amount, in $m	15.158	15.559***	15.411***	15.517***	15.546***
BHAR, 12-mo (equally weighted)	-0.534	-0.658	-0.622	-0.599	-0.602
Control variables:					
Venture characteristics:					
Whitepaper length, in (log-words)	8.1	8.453***	8.419***	8.467***	8.416***
Expert rating	3.393	3.513***	3.5***	3.529***	3.521***
Team size, in # FTE	12.924	15.063***	14.199***	15.137***	15.084***
Technical background, in %	25.438	25.203	24.853	25.273	25.047
Minimum viable product (dummy)	0.203	0.216	0.22	0.228	0.22
Open source (dummy)	0.661	0.658	0.691	0.685	0.663
# Industries (log)	1.291	1.319	1.322	1.328	1.305
Offering characteristics:					
Soft cap (dummy)	0.619	0.669***	0.667***	0.679**	0.673**
Hard cap (dummy)	0.884	0.918**	0.907	0.911***	0.914***
Pre-sale (dummy)	0.54	0.6**	0.571	0.574	0.6**
Whitelist (dummy)	0.312	0.388***	0.352	0.376**	0.396***
Bonus (dummy)	0.008	0.004	0.004	0.006	0.004
Bounty (dummy)	0.313	0.335	0.346	0.347	0.329
ERC-20 standard (dummy)	0.802	0.824	0.797	0.814	0.825
Market characteristics:					
Bull market (dummy)	0.325	0.266**	0.268**	0.267**	0.269**
Bear market (dummy)	0.688	0.759***	0.736**	0.739**	0.758***

[1] High-ESG = Startups with above-median ESG score.

has on the dependent variable, $DV_i \in \{\text{Valuation}_i, \text{Performance}_i\}$, controlling for a vector of independent variables, Ω_i:

$$DV_i = \beta ESG_i + \Omega_i \gamma + \varepsilon_i, \quad DV_i \in \{\text{Valuation}_i, \text{Performance}_i\} \tag{3}$$

To address the potential endogeneity problem associated with $E[\Omega_i, \varepsilon_i] \neq 0$, our first stage explicitly models the selection of startups into their ESG commitment. Specifically, we model the probability that startup i has a high ESG score above the median, $hiESG_i$, by a vector of exogenous control variables that possibly influence the selection mechanism, $\Omega_i^{(s)}$:

$$hiESG_i = \Omega_i^{(s)} \delta + \xi_i \tag{4}$$

We use the results from equation 4 to control for observed and unobserved heterogeneity in two distinct ways.

First, we compute inverse Mills ratios for each startup i's selection based on observable factors (IMR_i):

$$IMR_i = \frac{\phi\left(\frac{\Omega_i^{(s)}\delta}{\sigma_\xi}\right)}{\Phi\left(\frac{\Omega_i^{(s)}\delta}{\sigma_\xi}\right)} \tag{5}$$

We then use IMR_i in the second step to construct the following IMR estimator, where λ tests the null hypothesis that there is no selection effect:

$$DV_i^{IMR} = \beta ESG_i + \lambda IMR_i + \Omega_i \gamma + v_i, \quad DV_i^{IMR} \in \{\text{Valuation}_i^{IMR}, \text{Performance}_i^{IMR}\} \tag{6}$$

Second, we use Generalized Residuals (GRs) as instrumental variables for startups' ESG scores (Gourieroux et al., 1987) to control for unobserved heterogeneity by explicitly modeling any endogeneity in the error term. Consistent with Gourieroux et al. (1987), we define the generalized residual as:

$$GR_i = hESG_i \times \frac{\phi\left(-\Omega_i^{(s)}\delta\right)}{1 - \Phi\left(-\Omega_i^{(s)}\delta\right)} + (1 - hESG_i) \times \frac{-\phi\left(\Omega_i^{(s)}\delta\right)}{\Phi\left(-\Omega_i^{(s)}\delta\right)} \tag{7}$$

where $\phi(.)$ and $\Phi(.)$ denote the probability density and the cumulative density functions of the standard normal distribution, respectively. We restrict the standard deviation of the error term for startups with above-median ESG scores ($\sigma_{\varepsilon,\ hiESG=1}$) to be equal to that of startups with below-median ESG scores ($\sigma_{\varepsilon,\ hiESG=0}$). The restriction ensures that GR_i can be added as an instrumental variable to equation 3.

6 Results

6.1 ESG and Funding

Table 3 shows the main results for the VPH. All models include quarter-year and country fixed effects to absorb both time-related and geographical variation. All reported standard errors are robust. The R^2 in all of our models exceeds 30%, which is slightly higher than in previous studies (e.g., Fisch, 2019).

Our baseline (OLS) regression results are in column (1), with log of the funding amount in $ million as the dependent variable. The coefficient of the normalized ESG score is 0.25, with a p-value $< 1\%$, suggesting that a one standard deviation increase in the ESG score increases the average funding amount of $15.2 million by $4.2 million, or 28%. This strongly supports the VPH that there is a sustainability-related valuation premium in token offerings.

The coefficients of the control variables are largely consistent with those reported in prior studies (e.g., Bellavitis et al., 2020; Fisch, 2019; Huang et al., 2020; Momtaz, 2020a). Specifically, we find that the (i) whitepaper length, (ii) expert rating, (iii) team size, and (iv) presence of a whitelist are significantly positively related to the funding amount, while (v) open source code has a negative association with the funding amount. For sensitivity checks, we show a control model excluding the normalized ESG score in column (2). Both the signs and the magnitudes of the coefficients are similar in columns (1) and (2).

Given the evidence of startups' selection into ESG levels, we perform a two-stage approach in columns (3)-(5). Column (3) contains a first-stage Heckman selection model, which predicts the conditional probability that a startup chooses to have an above-median ESG score. Whitepaper length, team size and bear markets positively predict token offerings of high-ESG startups, while open source code's marginal effect is negative. We use the first-stage results to obtain IMRs and GRs, as described in section 5.4. We include IMRs as an additional control in column (4). The coefficient of the normalized ESG score is almost unchanged (0.250 in column (1) vs. 0.251 in column (4)). We also find that the IMR is statistically insignificant (unreported), indicating that "selection on observables" is not biasing the marginal effect of the normalized ESG score on the log of the funding amount. Finally, we use the GR as an instrumental variable for the normalized ESG score in column (5) to also address concerns about "selection on unobservables." This reduces the coefficient of the normalized ESG score to 0.211. Thus, unobserved heterogeneity may inflate the sustainability-related valuation premium in token offerings to some extent. Nevertheless, the valuation premium is still economically very significant in the IV model in column (5). In particular, an increase in the ESG score by one standard deviation increases the average funding amount of

$15.2 million by $3.6 million, corresponding to a relative effect of 23%. Overall, our baseline result is very robust to controlling for both observed and unobserved heterogeneity, and therefore provides strong support for the VPH (Hypothesis 1).

Our machine-learning approach to ESG measurement can also disaggregate the ESG score into its components E, S and G. Table 4 shows the regression results with the disaggregated ESG scores. Column (1) reprints the ESG coefficient from our baseline model in column (1) of Table 3 for comparison. Columns (2), (3), and (4) report regression coefficients for the disaggregated and normalized E, S and G scores, respectively. All disaggregated scores are statistically significant at least at the 5% level in these models. The E score coefficient is 0.137 (p-value < 0.01), the S score coefficient is 0.179 (p-value < 0.05), and the G score coefficient is 0.162 (p-value < 0.01). However, testing the effect of the three disaggregated scores simultaneously in column (5) shows that only the E (0.115) and the G (0.126) score are statistically significant at least at the 10% level. Therefore, ceteris paribus increases by one standard deviation in E and G are associated with 12% and 13% increases in the average funding amount, respectively.

Table 4 also reports Variance Inflation Factors (VIFs). All VIFs for the ESG variables are below 3, with the highest VIF being 2.95 for the S score in the simultaneous model in column (5). Additionally, the VIFs for all other control variables are clearly below 5, which is a commonly agreed threshold (e.g., Leitterstorf and Rau, 2014), indicating that multicollinearity is not a concern in our analyses.

Our final tests repeat the analyses in Tables 3 and 4 for Propensity Score Matched (PSM) samples. The rationale is that the PSM approach improves on the IMR-based "selection on observables" control approach if the selection process does not follow a normal distribution. This is because the *conditional independence assumption*[25] inherent in the IMR approach would be violated (e.g., Dehejia and Wahba, 2002; Rosenbaum and Rubin, 1983). Our PSM approach employs a one-to-one nearest-neighbor matching with two different selection cutoffs: 80% and 70%. That is, the PSM samples are based on selection models that predict whether a startup's ESG score is higher than the 80th and 70th percentiles, leading to different subsample sizes of 627 and 939 observations, respectively.

Table 5 presents the results for the PSM samples. Panels A and B regress on the aggregate and disaggregated ESG scores, respectively. Columns (1)-(2) and (3)-(4) report results for the baseline model and for the IV model, respectively. For brevity, we note that the results are consistent. The marginal effect of the aggregate and normalized ESG score ranges between 0.186 and 0.222, with a p-value always lower than 5%. The marginal effects of the disaggregated and normalized E score ranges between 0.105 and 0.126, with a p-value always below 10%. The coefficients for the S and G scores

[25] That is, conditional on IMR_i, the ESG scores must be independent of the other control variables.

Table 3: The Sustainability Premium

Explanations: These are results from regressions of startup valuation on the ESG score. The dependent variable is natural logarithm of the funding amount (in $ million). Our econometric identification approach is detailed in Section 5.4. In column (3), the dependent variable is a dummy indicating whether the token offering has an above-median ESG score. Column 4 shows the results for the Inverse Mills Ratio (IMR) approach. Column 5 uses the generalized residuals as an Instrumental Variable (IV) for the ESG score. Control variables are defined in Section 5.2. The sample consists of 1,043 token offerings between 2016 and 2020. All specifications include country and quarter-year fixed effects. Huber-White robust standard errors are in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Column	(1)	(2)	(3)	(4)	(5)
Model:	Main	Control	Selection	IMR[2]	IV[3]
Dependent variable:	Valuation[1]	Valuation	$\mathbb{1}_{\text{High-ESG}}$	Valuation	Valuation
Key variables:					
ESG Score (normalized)	0.250***			0.251***	0.211**
	(0.067)			(0.067)	(0.103)
Venture characteristics:					
Whitepaper length, in (log-words)	0.251**	0.492***	0.391***	0.242**	0.289**
	(0.122)	(0.115)	(0.035)	(0.122)	(0.147)
Expert rating	0.490***	0.485***	0.022	0.484***	0.490***
	(0.112)	(0.114)	(0.029)	(0.113)	(0.105)
Team size, in # FTE	0.031***	0.034***	0.006***	0.031***	0.031***
	(0.008)	(0.008)	(0.002)	(0.008)	(0.008)
Technical background, in %	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.004)	(0.004)	(0.001)	(0.004)	(0.003)
Minimum viable product (dummy)	0.001	-0.014	-0.061	0.007	-0.001
	(0.181)	(0.183)	(0.042)	(0.182)	(0.170)
Open source (dummy)	-0.351***	-0.385***	-0.125***	-0.328**	-0.356***
	(0.128)	(0.128)	(0.031)	(0.131)	(0.122)
# Industries (log)	-0.143	-0.147	-0.003	-0.151	-0.143
	(0.127)	(0.129)	(0.031)	(0.128)	(0.119)
Offering characteristics:					
Soft cap (dummy)	-0.206	-0.182	0.023	-0.206	-0.203
	(0.134)	(0.135)	(0.034)	(0.135)	(0.126)
Hard cap (dummy)	-0.035	-0.051	-0.021	-0.051	-0.037
	(0.199)	(0.201)	(0.049)	(0.200)	(0.187)
Pre-sale (dummy)	-0.149	-0.134	0.040	-0.155	-0.147
	(0.122)	(0.122)	(0.029)	(0.123)	(0.115)
Whitelist (dummy)	0.223*	0.246*	0.035	0.227*	0.227*
	(0.130)	(0.131)	(0.035)	(0.130)	(0.122)
Bonus (dummy)	0.111	0.087	-0.109	0.137	0.107
	(0.603)	(0.600)	(0.110)	(0.607)	(0.565)
Bounty (dummy)	-0.175	-0.179	-0.009	-0.177	-0.176
	(0.150)	(0.151)	(0.035)	(0.151)	(0.141)
ERC-20 standard (dummy)	-0.183	-0.189	0.003	-0.186	-0.184
	(0.139)	(0.141)	(0.035)	(0.140)	(0.131)
Market characteristics:					
Bull market (dummy)	-0.010	-0.010	0.029	-0.024	-0.010
	(0.189)	(0.189)	(0.058)	(0.190)	(0.177)
Bear market (dummy)	0.107	0.149	0.114*	0.094	0.113
	(0.233)	(0.232)	(0.061)	(0.233)	(0.218)
Observations	1043	1043	1043	1039	1043
R^2	0.313	0.306	0.408	0.315	0.313
IMR[2]	✗	✗	✗	✓	✗
IV[3]	✗	✗	✗	✗	✓
Quarter-year FEs	✓	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓	✓

[1] Valuation = Funding amount (log.).
[2] IMR = Inverse Mills Ratio
[3] IV = Instrumental Variable

Table 4: Decomposing the Sustainability Premium
Explanations: These are results from regressions of startup valuation on the ESG score and its components. The dependent variable is natural logarithm of the funding amount (in $ million). Our econometric identification approach is detailed in Section 5.4. In column (3), the dependent variable is a dummy indicating whether the token offering has an above-median ESG score. Column 4 shows the results for the Inverse Mills Ratio (IMR) approach. Column 5 uses the generalized residuals as an Instrumental Variable (IV) for the ESG score. Control variables are defined in Section 5.2. The sample consists of 1,043 token offerings between 2016 and 2020. All specifications include country and quarter-year fixed effects. Huber-White robust standard errors are in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. The table also shows Variance Inflation Factors (VIFs) for the ESG score and its components, as well as the highest VIF for the control variables. Control variables are similar to those reported in Table 3 and therefore suppressed here for brevity.

Column	(1)	(2)	(3)	(4)	(5)
Dependent variable: Valuation = funding amount (log.)					
ESG Score (normalized)	0.250***				
	(0.067)				
E-Score (normalized)		0.137***			0.115**
		(0.051)			(0.056)
S-Score (normalized)			0.179**		0.074
			(0.071)		(0.084)
G-Score (normalized)				0.162***	0.126*
				(0.062)	(0.069)
Observations	1043	1043	1043	1043	1043
R^2	0.313	0.310	0.310	0.309	0.314
VIF* [ESG]	2.16				
VIF [E]		1.27			1.41
VIF [S]			2.15		2.95
VIF [G]				1.82	2.28
VIF [argmax(controls)]	4.63	4.63	4.63	4.62	4.64
Controls	✓	✓	✓	✓	✓
Quarter-year FEs	✓	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓	✓

* VIF = Variance Inflation Factor.

are not consistently statistically significant. Therefore, the environmental component largely drives the sustainability-oriented valuation premium in token offerings.

6.2 ESG and Post-Funding Performance

Table 6 displays the tests of the PFUH (Hypothesis 2). The dependent variable in all of the models is the 12-month BHAR relative to an equally-weighted market benchmark. Panel A regresses on the aggregate normalized ESG score, while Panel B regresses on the disaggregated normalized E, S and G scores. Both panels contain the baseline model and the IMR model. Only Panel A contains the IV model (because GRs cannot be simultaneously calculated for each of the three ESG dimensions in Panel B).

The evidence supports the PFUH that startups with salient ESG properties underper-

Table 5: Propensity Score Matched (PSM) Samples
Explanations: These are results from regressions of startup valuation on the ESG score (Panel A) and its components (Panel B) based on Propensity Score Matched (PSM) samples. Models (1) and (3) match the 20% highest ESG score startups, and models (2) and (4) match the 30% highest ESG score startups to expand the PSM sample size. Columns (1) and (2) are baseline OLS regressions, while columns (3) and (4) are based on our instrumental variable (IV) model in Panel A and on the inclusion of the generalized residual (GR) in Panel B. Everything else is as in the regressions in Table 3 and Table 4.

Column	(1)	(2)	(3)	(4)
Model:	PSM		IV/GR[1]	
Selection cutoff:	80%ile	70%ile	80%ile	70%ile
Dependent variable: Funding amount (log)				
Panel A: ESG composite				
ESG Score (normalized)	0.195***	0.186***	0.222**	0.262***
	(0.065)	(0.061)	(0.103)	(0.100)
Controls	✓	✓	✓	✓
Quarter-year FEs	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓
IV	✗	✗	✓	✓
Observations	627	939	627	939
R^2	0.296	0.253	0.296	0.252
Panel B: ESG decomposition				
E-Score (normalized)	0.126**	0.110**	0.124**	0.105*
	(0.059)	(0.056)	(0.060)	(0.056)
S-Score (normalized)	0.059	0.050	0.054	0.034
	(0.091)	(0.084)	(0.095)	(0.089)
G-Score (normalized)	0.100	0.126*	0.094	0.109
	(0.074)	(0.068)	(0.082)	(0.075)
Controls	✓	✓	✓	✓
Quarter-year FEs	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓
Generalized Residual (GR)	✗	✗	✓	✓
Observations	627	939	627	939
R^2	0.299	0.256	0.299	0.256

[1] Columns (3) and (4) are based on the IV model in panel A and on the inclusion of the GR as a simple control in panel B.

form the market. Columns (1) and (2) show a one-standard-deviation increase in the aggregate ESG score is associated with a 16.3% underperformance over the first year of token trading. Interestingly, the estimated underperformance in column (3) is clearly higher, with a marginal effect of ESG on BHAR of −37.3%, suggesting that unobserved heterogeneity attenuates true underperformance.

In contrast to the dominance of the *environmental* component in the valuation premium (Hypothesis 1), Panel B of Table 6 shows that the *governance* component

drives the post-funding underperformance. Only the disaggregated G component is consistently statistically significant at least at the 10% level in Panel B. An increase by one standard deviation in the G dimension is associated with 19.2% (column 1) to 19.6% (column 2) post-funding underperformance. The E and S dimensions are not statistically significant and also economically insignificant, with coefficients ranging from -2.7% to -0.2%. Overall, these results support our second hypothesis that sustainability-oriented startups underperform the market post-funding, with the effect being mostly attributable to the governance dimension in ESG.

6.3 The Moderating Effect of Formalization

The results so far suggest that entrepreneurs benefit from sustainability-orientation in the form of a premium during the funding stage and investors incur a relative financial loss post- funding. Our third hypothesis posits that formalization (i.e., binding constraints) has a negative moderating effect on both funding and post-funding performance. The rationale is that sustainability-orientation already imposes binding constraints onto the startup, the effects of which may be magnified by other constraints. Binding constraints reduce entrepreneurial flexibility and the scope of experimentation (e.g., March, 1991), therefore, ***H3a*** and ***H3b*** posits that formalization is associated negatively with the ESG funding and performance relations articulated in the ***VPH*** and ***PFUH***.

Table 7 presents the results of the moderation tests. We use three proxies for the formalization. For technological formalization, proxy 1 is a dummy indicator for whether the startup open-sourced its code on *GitHub*. For network formalization, proxy 2 is the log of the number of followers on *Twitter*. For governance formalization, proxy 3 is a dummy equal to one if the startup is backed by a VC. These variables have been introduced before in the token offerings literature (e.g., Fisch, 2019; Fisch and Momtaz, 2020).

Columns (1)-(3) and (4)-(6) regress the log of the funding amount and the 12-month BHAR, respectively. For the valuation models, we find that all formalization proxies have a strong direct effect on startups' valuations, as well as negative moderating effects, with the interactions with the network and governance proxies being statistically significant at least at the 10% level. For the performance models, we find only partial support for our hypothesis. Only the governance-related formalization proxy has a statistically significant direct effect on the 12-month BHAR (p-value $< 1\%$), while only the technology-related formalization proxy has a statistically significantly negative moderating effect. In particular, ceteris paribus, if the startup increases its ESG score by one standard deviation while having open-sourced some of its platform code, then the post-funding underperformance increases by 33.2%. Overall, the results in Table 7

Table 6: The Performance of Sustainable Entrepreneurs
Explanations: These are results from regressions of the long-run performance on the ESG score in Panel A and on its components in Panel B. The dependent variable is the 12-month Buy-and-Hold Abnormal Return (BHAR) after the token listing date relative to an equally-weighted composite crypto-market benchmark. Model (1) is the baseline model, model (2) conditions on observable heterogeneity, and model (3) controls for unobserved heterogeneity. Control variables are defined in Section 5.2. The sample size is reduced because either tokens have not been listed and therefore we do not observe token price performance or because tokens have been listed less than 12 months ago. All specifications include country and quarter fixed effects. Huber-White robust standard errors are in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Column	(1)	(2)	(3)
Model:	Main	IMR	IV/GR
Dependent variable: BHAR, 12 months			
Panel A: ESG composite			
ESG Score (normalized)	-0.163*	-0.163*	-0.373**
	(0.091)	(0.092)	(0.155)
Controls	✓	✓	✓
Quarter-year fixed effects	✓	✓	✓
Country fixed effects	✓	✓	✓
IMR	✗	✓	✗
IV	✗	✗	✓
Observations	302	300	302
R^2	0.368	0.377	0.357
Panel B: ESG decomposition			
E-Score (normalized)	-0.024	-0.027	.
	(0.083)	(0.084)	.
S-Score (normalized)	-0.015	-0.007	.
	(0.138)	(0.142)	.
G-Score (normalized)	-0.192*	-0.196*	.
	(0.110)	(0.111)	.
Controls	✓	✓	.
Quarter-year fixed effects	✓	✓	.
Country fixed effects	✓	✓	.
IMR	✗	✓	.
GR	✗	✗	.
Observations	302	300	.
R^2	0.372	0.382	.

provide partial support for the moderating effects posited by ***H3a*** and ***H3b***.

For brevity, Table 7 does not report the coefficients of our control variables as they resemble those shown in Table 3. Note that the R^2 increases significantly compared to the unmoderated specifications in Tables 3 and 6.

Our robustness checks are available in the Appendix for the sake of brevity. Specifi-

Table 7: The Moderating Effect of the Degree of Formalization
Explanations: These are moderation results from regressions of startup valuation on the ESG score, interacted with proxies for technology, network, and governance formalization. The dependent variable is natural logarithm of the funding amount (in $ million) in columns (1)-(3) and the 12-month BHAR in columns (4)-(6). Proxy 1 is the dummy indicator for whether the token offering firm open-sourced its code on *GitHub*. Proxy 2 is the log-number of followers in Twitter. Proxy 3 is a dummy that equals one if the token offering is backed by a venture capital fund. Control variables are defined in Section 5.2. The sample consists of 1,043 token offerings between 2016 and 2020. All specifications include country and quarter-year fixed effects. Huber-White robust standard errors are in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Column	(1)	(2)	(3)	(4)	(5)	(6)
Model	Valuation			Performance		
Dependent variable:	Funding amount (log.)			BHAR (12 months)		
ESG	0.255**	0.776**	0.312***	0.087	0.461	-0.196*
	(0.129)	(0.333)	(0.070)	(0.148)	(0.396)	(0.101)
Formalization (Proxy 1)	-0.351***			-0.197		
	(0.128)			(0.160)		
Formalization (Proxy 2)		0.166***			0.035	
		(0.039)			(0.043)	
Formalization (Proxy 3)			1.092***			0.594***
			(0.137)			(0.186)
ESG × Formalization (Proxy 1)	-0.007			-0.332*		
	(0.141)			(0.174)		
ESG × Formalization (Proxy 2)		-0.068*			-0.068	
		(0.039)			(0.046)	
ESG × Formalization (Proxy 3)			-0.533***			0.005
			(0.116)			(0.176)
Controls	✓	✓	✓	✓	✓	✓
Quarter-year fixed effects	✓	✓	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓	✓	✓
Observations	1043	1008	1043	302	290	302
R^2	0.313	0.332	0.350	0.378	0.376	0.401

cally, in Appendix B, we test the sensitivity of our results to the inclusion of additional control variables, which has the advantage of absorbing more variation, while the limited availability of additional variables reduces our sample size substantially. Our main results do not qualitatively change with these specifications. For example, controlling for the percentage of tokens distributed in the token offering (token retention often serves as a signal for project quality, see Leland and Pyle, 1977), does not affect the ESG-valuation relation. Importantly, we also report robustness tests for different ESG scores, by altering the initial seed words we provide the machine in order to compile the ESG dictionary. Again, our results are very robust. All of these tests are discussed in detail in Appendix C.

7 Discussion and Concluding Remarks

7.1 Summary of Main Results

This paper tests two overarching hypotheses. The ***Valuation Premium Hypothesis*** (***VPH***) posits that Sustainable Entrepreneurship (SE) achieves higher valuations in entrepreneurial finance markets than Conventional Entrepreneurship (CE) does. The ***Post-Funding Underperformance Hypothesis*** (***PFUH***) posits that SE (financially) underperforms CE in the long run. The empirical context is utility token offerings or Initial Coin Offerings (ICOs). Token offerings provide an ideal laboratory to test these hypotheses because (i) the information disclosed in whitepapers can be used to quantify startups' ESG properties, and (ii) tokens are often listed on exchange platforms after the offering, providing a transparent measure of financial performance (Fisch and Momtaz, 2020). Examining a sample of 1,043 token offerings in the period 2016-2020, we find support for both the *VPH* and *PFUH*. For the *VPH*, we find that a one-standard-deviation increase in our ESG metric is associated with a 28% increase in the funding amount. This corresponds to $4.2 million (relative to the mean funding amount of $15.2 million in our sample). For the *PFUH*, we find that a one-standard-deviation increase in our ESG metric is associated with a 16% decrease in the first 12-month buy-and-hold abnormal (equally weighted relative to a composite market index) token price performance. Relative to financial utility, non-financial (ESG-related) utility for SE investors amounts to 16-31% of total utility.

Additional analyses investigate the moderating role of technology, network, and governance-related formalization at the startup level on the relationships articulated in the *VPH* and *PFUH*. Formalization refers broadly to all organizing that imposes dependencies and constraints onto the startup (Kraus et al., 2018). The contingency effects of formalization are important. Technological, network and governance aspects associated with startup formalization all hurt the valuation and performance of sustainability-oriented startups in our sample. These findings are consistent with predictions in the SE literature, as synthesized in Kraus et al. (2018), albeit puzzling. Attributes, such as open-source code, large social networks and venture capital backing, that are associated with success in conventional startups (Fisch, 2019; Fisch and Momtaz, 2020), actually impede success in sustainability-oriented startups, suggesting that organizing SE may need to meet fundamentally different requirements than CE (Parrish, 2010).

Our results are robust to endogeneity concerns related to observed and unobserved heterogeneity in our sample, and insensitive to various modifications of our empirical baseline model, as well as to modifications of our machine-learning approach to measure startups' ESG properties.

7.2 Theoretical Contributions and Practical Implications

Our study contributes to the SE literature in several important ways. First, the sustainability-related valuation premium suggests that entrepreneurs have an economic incentive to launch sustainability-oriented projects or to introduce ESG aspects to existing ones. The existence of the sustainability premium also implies that Schumpeterian logic may apply (Schumpeter, 1934, 1942), and that the demand for ESG creates entrepreneurial opportunity, potentially leading to a replacement of conventional businesses with sustainability-oriented ones. As such, entrepreneurs may act as "change agents" for sustainability-oriented change (Anand et al., 2021, p. 2). Our finding thus addresses a major question around SE, potentially helping to resolve much of the "controversy" around the incentives of SE in the literature (Hall et al., 2010, p. 439).

Second, our study helps close the "research gap related to the post-funding phase" of SE (Böckel et al., 2020, p. 433). Financial underperformance by SEs suggests that investors in ESG startups are willing to pay for non-financial sustainability-related returns. Viewing the financial underperformance as an upper bound for the non-financial utility from ESG, our study suggests that non-financial utility constitutes 17-27% of total utility in sustainability-oriented entrepreneurial activity. It is worth noting that even after one year of post-funding underperformance, the average ESG startup still trades at a valuation premium of up to 10%. Therefore, despite the underperformance and the opportunity to exit the investment anytime in liquid token markets, investors remain invested in ESG startups, again highlighting the importance of non-financial utility for SE investors.

Finally, our study contributes to the emerging SE literature by highlighting the importance of binding constraints (Pástor et al., 2020; Renneboog et al., 2008). That technology, network and governance-related formalization negatively influence both the valuation premium and the post-funding performance underscores the importance of delegated philanthropy in solving problems associated with SE's binding constraints (Kraus et al., 2018). O. Hart and Zingales (2017) relax Friedman's (1970) assumptions to show that sustainability likely has a business case outside of neoclassical models, and that delegated philanthropy can reduce SE execution risk associated with the rigidity of binding constraints. The more specialized the delegation of ESG problems to various ventures, the more successful these ventures are in terms of funding and post-funding performance because granular delegation reduces binding constraints for individual startups. Nevertheless, delegation, underperformance, and moderating formalization all suggest that SE is not a "no brainer," and that more work, particularly on organizing SE (Parrish, 2010), is necessary to understand when SE is beneficial for entrepreneurs and investors, and why.

Our study reveals three distinct practical implications. First, from a public policy

perspective, that entrepreneurs have an economic incentive for sustainability-oriented venturing and receive funding suggests that Schumpeter's (1942) notion of 'creative destruction' seems applicable to the SE context. Thus, the SE market should sustain itself without the need for government subsidies. Second, SE investors need to expect financial losses relative to CE. Thus, SE may only attract investors whose personal ESG preferences can compensate financial underperformance. Third, and arguably most importantly, entrepreneurs need to cautiously weigh the pros and cons of various organizational designs, and consider that organizing that is optimal for CE may not be optimal for SE (Parrish, 2010).

7.3 Avenues for Further Research

Our study represents a first step towards understanding the relevant financial aspects of SE activity that matter for entrepreneurs and investors alike. Given the vast and growing interest in SE, as evidenced by the large number of recent reviews (for a recent overview, see Anand et al., 2021, chapter 2.1), and the necessarily high level of abstraction in our analysis, it seems very likely that a vivid literature around the financial aspects of sustainability-oriented venturing will soon emerge. Below, we suggest some avenues for potentially fruitful further research:

1. *Financial returns to SE.* Our study focuses on token offerings, a market predominantly populated by relatively young generations with strong ESG preferences (Fisch et al., 2019; Kraus et al., 2018; Spence et al., 2011). Also, given the recency of token offerings, our analysis of 'long-run' returns is limited to a one-year period. This gives rise to a number of interesting questions. First, do our results of an ESG premium followed by underperformance hold in other contexts, in particular in those with institutional investors, e.g., venture capitalists, whose limited partnership agreements often require them to focus exclusively on financial returns? Second, how long do investors bear SE underperformance, and is there a point of financial loss at which investors abandon sustainability-oriented startups? Third, our study is set in a period when demand for ESG is relatively high. This raises the question of how our results would change with less aggregate demand for ESG.

2. *ESG returns to SE.* Our study estimates the financial rents associated with SE for entrepreneurs and investors. While our underperformance measure can be viewed as an upper bound for investors' ESG rents, it leaves a number of questions unanswered. For example, relative to financial rents, how important are ESG rents for investors in sustainability-oriented startups, and to what extent are investors willing to sacrifice financial rents for ESG goals? Of course, as Anand et

al. (2021, p. 12) correctly observe, a "major challenge" here is "how to measure sustainability". This is owed partly to the subjectivity of many ESG rents (e.g., normative dimensions of ESG, such as relative economic equality), partly to the longevity of many ESG goals (e.g., climate change), and partly to the difficulty associated with quantifying ESG rents, among others. Moving forward, we believe that case studies offer the best possibility to understand cause-and-effect in SE.

3. *Disaggregating ESG.* Similarly, our study employs a machine-learning approach to quantify ESG properties of startups. We also decompose ESG into E, S and G. However, an even more granular approach may help unveil contingency aspects of SE (e.g., E is composed of many grand challenges itself, such as climate change, air and water pollution, solar energy and other renewable energies, and carbon footprints of new and old technologies). Future research can easily modify our machine-learning algorithm, which we publish as open source along with this paper, to measure more granular components of ESG.

4. *Organizing SE.* Our study of how a high degree of formalization associated with technology, network and governance aspects, which are all associated with success in conventional startups, can be detrimental in sustainability-oriented startups raises the important question of the optimal organizational design in SE. Parrish (2010) discusses how organizational design in CE and SE may be fundamentally different, employing an inductive approach based on 32 qualitative interviews. Yet, the research on concrete, practically implementable forms of startup structure conducive to SE success is very limited.

References

Adhami, S., Giudici, G., & Martinazzi, S. (2018). Why do businesses go crypto? an empirical analysis of initial coin offerings. *Journal of Economics and Business*, *100*, 64–75.

Ahlers, G. K., Cumming, D., Günther, C., & Schweizer, D. (2015). Signaling in equity crowdfunding. *Entrepreneurship Theory and Practice*, *39*(4), 955–980.

Albuquerque, R., Koskinen, Y., Yang, S., & Zhang, C. (2020). Resiliency of environmental and social stocks: An analysis of the exogenous covid-19 market crash. *The Review of Corporate Finance Studies*, *9*(3), 593–621.

Amsden, R., & Schweizer, D. (2018). Are blockchain crowdsales the new'gold rush'? success determinants of initial coin offerings. *Success determinants of initial coin offerings (April 16, 2018)*.

Anand, A., Argade, P., Barkemeyer, R., & Salignac, F. (2021). Trends and patterns in sustainable entrepreneurship research: A bibliometric review and research agenda. *Journal of Business Venturing*, *36*(3), 106092.

Balakrishnan, U., Duvall, T., & Primeaux, P. (2003). Rewriting the bases of capitalism: Reflexive modernity and ecological sustainability as the foundations of a new normative framework. *Journal of Business Ethics*, *47*(4), 299–314.

Barber, B. M., Morse, A., & Yasuda, A. (2021). Impact investing. *Journal of Financial Economics*, *139*(1), 162–185.

Bellavitis, C., Cumming, D., & Vanacker, T. (2020). Ban, boom, and echo! entrepreneurship and initial coin offerings. *Entrepreneurship Theory and Practice*, 1042258720940114.

Bellavitis, C., Fisch, C., & Wiklund, J. (2021). A comprehensive review of the global development of initial coin offerings (icos) and their regulation. *Journal of Business Venturing Insights*, *15*, e00213.

Berg, F., Koelbel, J. F., & Rigobon, R. (2020). *Aggregate confusion: The divergence of esg ratings*. MIT Sloan School of Management.

Bertoni, F., Colombo, M. G., & Grilli, L. (2011). Venture capital financing and the growth of high-tech start-ups: Disentangling treatment from selection effects. *Research Policy*, *40*(7), 1028–1043.

Binder, J. K., & Belz, F.-M. (2017). Developing sustainable opportunities: Merging insights from social identity and structuration theory. *Academy of Management Proceedings*, *2017*(1), 14167.

Bischoff, K., & Volkmann, C. K. (2018). Stakeholder support for sustainable entrepreneurship-a framework of sustainable entrepreneurial ecosystems. *International Journal of Entrepreneurial Venturing*, *10*(2), 172–201.

Block, J. H., Colombo, M. G., Cumming, D. J., & Vismara, S. (2018). New players in entrepreneurial finance and why they are there. *Small Business Economics*, *50*(2), 239–250.

Block, J. H., Hirschmann, M., & Fisch, C. (2021). Which criteria matter when impact investors screen social enterprises? *Journal of Corporate Finance*, *66*, 101813.

Böckel, A., Hörisch, J., & Tenner, I. (2020). A systematic literature review of crowdfunding and sustainability: Highlighting what really matters. *Management Review Quarterly*, 1–21.

Breuer, H., Fichter, K., Lüdeke-Freund, F., & Tiemann, I. (2018). Sustainability-oriented business model development: Principles, criteria and tools. *International Journal of Entrepreneurial Venturing*, *10*(2), 256–286.

Calic, G., & Mosakowski, E. (2016). Kicking off social entrepreneurship: How a sustainability orientation influences crowdfunding success. *Journal of Management Studies*, *53*(5), 738–767.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (2012). *The econometrics of financial markets*. Princeton University Press.

Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate social responsibility and access to finance. *Strategic management journal*, *35*(1), 1–23.

Choi, D. Y., & Gray, E. R. (2008). The venture development processes of "sustainable" entrepreneurs. *Management Research News*.

Cohen, B., & Winn, M. I. (2007). Market imperfections, opportunity and sustainable entrepreneurship. *Journal of Business Venturing*, *22*(1), 29–49.

Colombo, M. G., & Grilli, L. (2010). On growth drivers of high-tech start-ups: Exploring the role of founders' human capital and venture capital. *Journal of Business Venturing*, *25*(6), 610–626.

Colombo, O. (2021). The use of signals in new-venture financing: A review and research agenda. *Journal of Management*, *47*(1), 237–259.

Cornell, B. (2021). Esg preferences, risk and return. *European Financial Management*, *27*(1), 12–19.

Cumming, D., Henriques, I., & Sadorsky, P. (2016).

cleantech'venture capital around the world. *International Review of Financial Analysis*, *44*, 86–97.

Cumming, D., Leboeuf, G., & Schwienbacher, A. (2017). Crowdfunding cleantech. *Energy Economics*, *65*, 292–303.

Davies, I. A., & Chambers, L. (2018). Integrating hybridity and business model theory in sustainable entrepreneurship. *Journal of Cleaner Production*, *177*, 378–386.

Dean, T. J., & McMullen, J. S. (2007). Toward a theory of sustainable entrepreneurship: Reducing environmental degradation through entrepreneurial action. *Journal of Business Venturing*, *22*(1), 50–76.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, *84*(1), 151–161.

Desa, G. (2012). Resource mobilization in international social entrepreneurship: Bricolage as a mechanism of institutional transformation. *Entrepreneurship Theory and Practice*, *36*(4), 727–751.

Dickel, P. (2017). The impact of protectability and proactiveness on the environmental performance of new ventures. *Corporate Governance: The International Journal of Business in Society*.

Dimson, E., Marsh, P., & Staunton, M. (2020). Divergent esg ratings. *The Journal of Portfolio Management*, *47*(1), 75–87.

Djupdal, K., & Westhead, P. (2015). Environmental certification as a buffer against the liabilities of newness and smallness: Firm performance benefits. *International Small Business Journal*, *33*(2), 148–168.

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, *60*(11), 2835–2857.

Edmans, A. (2011). Does the stock market fully value intangibles? employee satisfaction and equity prices. *Journal of Financial Economics*, *101*(3), 621–640.

Egri, C. P., & Herman, S. (2000). Leadership in the north american environmental sector: Values, leadership styles, and contexts of environmental leaders and their organizations. *Academy of Management Journal*, *43*(4), 571–604.

Fama, E. F., & French, K. R. (2007). Disagreement, tastes, and asset prices. *Journal of Financial Economics*, *83*(3), 667–689.

Fisch, C. (2019). Initial coin offerings (icos) to finance new ventures. *Journal of Business Venturing*, *34*(1), 1–22.

Fisch, C., Masiak, C., Vismara, S., & Block, J. (2019). Motives and profiles of ico investors. *Journal of Business Research*.

Fisch, C., Meoli, M., & Vismara, S. (2020). Does blockchain technology democratize entrepreneurial finance? an empirical comparison of icos, venture capital, and reits. *Economics of Innovation and New Technology*, 1–20.

Fisch, C., & Momtaz, P. P. (2020). Institutional investors and post-ico performance: An empirical analysis of investor returns in initial coin offerings (icos). *Journal of Corporate Finance*, *64*, 101679.

Friedman, M. (1970). The social responsibility of business is to increase its profits. *The New York Times Magazine*, *(13 September)*, 32–33.

Gast, J., Gundolf, K., & Cesinger, B. (2017). Doing business in a green way: A systematic review of the ecological sustainability entrepreneurship literature and future research directions. *Journal of Cleaner Production*, *147*, 44–56.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–74.

Gibson, R., Glossner, S., Krueger, P., Matos, P., & Steffen, T. (2020). Responsible institutional investing around the world. *Swiss Finance Institute Research Paper*, (20-13).

Gillan, S. L., Koch, A., & Starks, L. T. (2021). Firms and social responsibility: A review of esg and csr research in corporate finance. *Journal of Corporate Finance*, 101889.

Giudici, G., & Adhami, S. (2019). The impact of governance signals on ico fundraising success. *Journal of Industrial and Business Economics*, *46*(2), 283–312.

Giudici, G., Guerini, M., & Rossi-Lamastra, C. (2018). Reward-based crowdfunding of entrepreneurial projects: The effect of local altruism and localized social capital on proponents' success. *Small Business Economics*, *50*(2), 307–324.

Gourieroux, C., Monfort, A., Renault, E., & Trognon, A. (1987). Generalised residuals. *Journal of Econometrics*, *34*(1-2), 5–32.

Gregori, P., Wdowiak, M. A., Schwarz, E. J., & Holzmann, P. (2019). Exploring value creation in sustainable entrepreneurship: Insights from the institutional logics perspective and the business model lens. *Sustainability*, *11*(9), 2505.

Guzmán, A., Pinto-Gutiérrez, C., & Trujillo, M.-A. (2020). Attention to global warming and the success of environmental initial coin offerings: Empirical evidence. *Sustainability*, *12*(23), 9885.

Hahn, T., Figge, F., Pinkse, J., & Preuss, L. (2018). A paradox perspective on corporate sustainability: Descriptive, instrumental, and normative aspects. *Journal of Business Ethics*, *148*(2), 235–248.

Hall, J., Daneke, G., & Lenox, M. (2010). Sustainable development and entrepreneurship: Past contributions and future directions. *Journal of Business Venturing*, *25*(5), 439–448.

Hall, J., & Vredenburg, H. (2003). The challenges of innovating for sustainable development. *MIT Sloan Management Review*, *45*(1), 61–68.

Hart, O., & Zingales, L. (2017). Companies should maximize shareholder welfare not market value. *ECGI-Finance Working Paper*, (521).

Hart, S. L., & Christensen, C. M. (2002). The great leap: Driving innovation from the base of the pyramid. *MIT Sloan Management Review*, *44*(1), 51.

Hart, S. L., & Milstein, M. B. (1999). Global sustainability and the creative destruction of industries. *MIT Sloan Management Review*, *41*(1), 23.

Hockerts, K., & Wüstenhagen, R. (2010). Greening goliaths versus emerging davids—theorizing about the role of incumbents and new entrants in sustainable entrepreneurship. *Journal of Business Venturing*, *25*(5), 481–492.

Hoogendoorn, B., Van der Zwan, P., & Thurik, R. (2019). Sustainable entrepreneurship: The role of perceived barriers and risk. *Journal of Business Ethics*, *157*(4), 1133–1154.

Hooi, H. C., Ahmad, N. H., Amran, A., & Rahman, S. A. (2016). The functional role of entrepreneurial orientation and entrepreneurial bricolage in ensuring sustainable entrepreneurship. *Management Research Review*.

Hörisch, J. (2015). Crowdfunding for environmental ventures: An empirical analysis of the influence of environmental orientation on the success of crowdfunding initiatives. *Journal of Cleaner Production*, *107*, 636–645.

Hornuf, L., Kück, T., & Schwienbacher, A. (2021). Initial coin offerings, information disclosure, and fraud. *Small Business Economics*, 1–19.

Howell, S. T., Niessner, M., & Yermack, D. (2020). Initial coin offerings: Financing growth with cryptocurrency token sales. *The Review of Financial Studies*, *33*(9), 3925–3974.

Huang, W., Meoli, M., & Vismara, S. (2020). The geography of initial coin offerings. *Small Business Economics*, *55*(1), 77–102.

Ioannou, I., & Serafeim, G. (2015). The impact of corporate social responsibility on investment recommendations: Analysts' perceptions and shifting institutional logics. *Strategic Management Journal*, *36*(7), 1053–1081.

Iyer, R., Khwaja, A. I., Luttmer, E. F., & Shue, K. (2016). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, *62*(6), 1554–1577.

Jahanshahi, A. A., & Brem, A. (2017). Sustainability in smes: Top management teams behavioral integration as source of innovativeness. *Sustainability*, *9*(10), 1899.

Johnson, M. P., & Schaltegger, S. (2020). Entrepreneurship for sustainable development: A review and multilevel causal mechanism framework. *Entrepreneurship Theory and Practice*, *44*(6), 1141–1173.

Kerr, S. P., Kerr, W. R., & Xu, T. (2017). Personality traits of entrepreneurs: A review of recent literature.

Kerr, W. R., Nanda, R., & Rhodes-Kropf, M. (2014). Entrepreneurship as experimentation. *Journal of Economic Perspectives*, *28*(3), 25–48.

Kimuli, S. N. L., Orobia, L., Sabi, H. M., & Tsuma, C. K. (2020). Sustainability intention: Mediator of sustainability behavioral control and sustainable entrepreneurship. *World Journal of Entrepreneurship, Management and Sustainable Development*.

Knight, G. A. (1997). Cross-cultural reliability and validity of a scale to measure firm entrepreneurial orientation. *Journal of Business Venturing*, *12*(3), 213–225.

Kraus, S., Burtscher, J., Niemand, T., Roig-Tierno, N., & Syrjä, P. (2017). Configurational paths to social performance in smes: The interplay of innovation, sustainability, resources and achievement motivation. *Sustainability*, *9*(10), 1828.

Kraus, S., Burtscher, J., Vallaster, C., & Angerer, M. (2018). Sustainable entrepreneurship orientation: A reflection on status-quo research on factors facilitating responsible managerial practices. *Sustainability*, *10*(2), 444.

Kuckertz, A., & Wagner, M. (2010). The influence of sustainability orientation on entrepreneurial intentions—investigating the role of business experience. *Journal of Business Venturing*, *25*(5), 524–539.

Lans, T., Blok, V., & Wesselink, R. (2014). Learning apart and together: Towards an integrated competence framework for sustainable entrepreneurship in higher education. *Journal of Cleaner Production*, *62*, 37–47.

Leitterstorf, M. P., & Rau, S. B. (2014). Socioemotional wealth and ipo underpricing of family firms. *Strategic Management Journal*, *35*(5), 751–760.

Leland, H. E., & Pyle, D. H. (1977). Informational asymmetries, financial structure, and financial intermediation. *The Journal of Finance*, *32*(2), 371–387.

Li, K., Mai, F., Shen, R., & Yan, X. (2020). Measuring corporate culture using machine learning. *The Review of Financial Studies*.

Liang, H., & Renneboog, L. (2017). On the foundations of corporate social responsibility. *The Journal of Finance*, *72*(2), 853–910.

Lieberman, M. B., & Montgomery, D. B. (1988). First-mover advantages. *Strategic Management Journal*, *9*(S1), 41–58.

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.

Loughran, T., & McDonald, B. (2020a). Measuring firm complexity. *Available at SSRN 3645372*.

Loughran, T., & McDonald, B. (2020b). Textual analysis in finance. *Annual Review of Financial Economics*, *12*(1), 357–375. https://doi.org/10.1146/annurev-financial-012820-032249

Mair, J., & Marti, I. (2009). Entrepreneurship in and around institutional voids: A case study from bangladesh. *Journal of Business Venturing*, *24*(5), 419–435.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *Association for Computational Linguistics (ACL) System Demonstrations*, 55–60. http://www.aclweb.org/anthology/P/P14/P14-5010

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87.

Meek, W. R., Pacheco, D. F., & York, J. G. (2010). The impact of social norms on entrepreneurial action: Evidence from the environmental entrepreneurship context. *Journal of Business Venturing*, *25*(5), 493–509.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *arXiv preprint arXiv:1310.4546*.

Mollick, E. R., & Kuppuswamy, V. (2014). After the campaign: Outcomes of crowdfunding. *UNC Kenan-Flagler Research Paper*, (2376997).

Momtaz, P. P. (2019). Token sales and initial coin offerings: Introduction. *The Journal of Alternative Investments*, *21*(4), 7–12.

Momtaz, P. P. (2020a). Entrepreneurial finance and moral hazard: Evidence from token offerings. *Journal of Business Venturing*, 106001.

Momtaz, P. P. (2020b). Initial coin offerings. *Plos one*, *15*(5), e0233018.

Momtaz, P. P. (2020c). Initial coin offerings, asymmetric information, and loyal ceos. *Small Business Economics*, 1–23.

Momtaz, P. P. (2021a). Ceo emotions and firm valuation in initial coin offerings: An artificial emotional intelligence approach. *Strategic Management Journal*, *42*(3), 558–578.

Momtaz, P. P. (2021b). Econometric models of duration data in entrepreneurship with an application to start-ups' time-to-funding by venture capitalists (vcs). *Journal of Applied Statistics*, 1–22.

Momtaz, P. P. (2021c). The pricing and performance of cryptocurrency. *The European Journal of Finance*, *27*(4-5), 367–380.

Muñoz, P., Cacciotti, G., & Cohen, B. (2018). The double-edged sword of purpose-driven behavior in sustainable venturing. *Journal of Business Venturing*, *33*(2), 149–178.

Muñoz, P., & Cohen, B. (2017). Mapping out the sharing economy: A configurational approach to sharing business modeling. *Technological Forecasting and Social Change*, *125*, 21–37.

Muñoz, P., & Cohen, B. (2018). Entrepreneurial narratives in sustainable venturing: Beyond people, profit, and planet. *Journal of Small Business Management*, *56*, 154–176.

Muñoz, P., Janssen, F., Nicolopoulou, K., & Hockerts, K. (2018). Advancing sustainable entrepreneurship through substantive research. *International Journal of Entrepreneurial Behavior & Research*.

Mupfasoni, B., Kessler, A., & Lans, T. (2018). Sustainable agricultural entrepreneurship in burundi: Drivers and outcomes. *Journal of Small Business and Enterprise Development*.

Nakamoto, S. (2019). *Bitcoin: A peer-to-peer electronic cash system* (tech. rep.). Manubot.

Nissim, M., van Noord, R., & van der Goot, R. (2020). Fair is better than sensational: Man is to doctor as woman is to doctor. *Computational Linguistics*, *46*(2), 487–497.

Obrecht, J.-J. (2011). On the relevance of the concept of entrepreneurial capabilities. *International Journal of Entrepreneurship and Small Business*, *13*(2), 107–125.

Obrecht, J.-J. (2016). Sustainable entrepreneurship education: A new field for research in step with the'effectual entrepreneur'. *International Journal of Entrepreneurship and Small Business*, *29*(1), 83–102.

Pacheco, D. F., Dean, T. J., & Payne, D. S. (2010). Escaping the green prison: Entrepreneurship and the creation of opportunities for sustainable development. *Journal of Business Venturing*, *25*(5), 464–480.

Parrish, B. D. (2010). Sustainability-driven entrepreneurship: Principles of organization design. *Journal of Business Venturing*, *25*(5), 510–523.

Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2020). Sustainable investing in equilibrium. *Journal of Financial Economics*.

Patzelt, H., & Shepherd, D. A. (2011). Recognizing opportunities for sustainable development. *Entrepreneurship Theory and Practice*, *35*(4), 631–652.

Ploum, L., Blok, V., Lans, T., & Omta, O. (2018). Toward a validated competence framework for sustainable entrepreneurship. *Organization & Environment*, *31*(2), 113–132.

Renneboog, L., Ter Horst, J., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance*, *32*(9), 1723–1742.

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.

Sarango-Lalangui, P., Santos, J. L. S., & Hormiga, E. (2018). The development of sustainable entrepreneurship research field. *Sustainability*, *10*(6), 2005.

Schaefer, K., Corner, P. D., & Kearins, K. (2015). Social, environmental and sustainable entrepreneurship research: What is needed for sustainability-as-flourishing? *Organization & Environment*, *28*(4), 394–413.

Schaltegger, S., Beckmann, M., & Hockerts, K. (2018). Collaborative entrepreneurship for sustainability. creating solutions in light of the un sustainable development goals. *International Journal of Entrepreneurial Venturing*, *10*(2), 131–152.

Schaltegger, S., Hansen, E. G., & Lüdeke-Freund, F. (2016). Business models for sustainability: Origins, present research, and future avenues.

Schückes, M., & Gutmann, T. (2020). Why do startups pursue initial coin offerings (icos)? the role of economic drivers and social identity on funding choice. *Small Business Economics*, 1–26.

Schumpeter, J. A. (1934). *The theory of economic development*. Harvard University Press, Cambridge.

Schumpeter, J. A. (1942). *Capitalism, socialism and democracy*. Harper, New York.

Senge, P. M., Carstedt, G., & Porter, P. L. (2001). Next industrial revolution. *MIT Sloan Management Review*, *42*(2), 24–38.

Shepherd, D. A., Kuskova, V., & Patzelt, H. (2009). Measuring the values that underlie sustainable development: The development of a valid scale. *Journal of Economic Psychology*, *30*(2), 246–256.

Shepherd, D. A., & Patzelt, H. (2011). The new field of sustainable entrepreneurship: Studying entrepreneurial action linking "what is to be sustained" with "what is to be developed". *Entrepreneurship Theory and Practice*, *35*(1), 137–163.

Shepherd, D. A., & Patzelt, H. (2017). Researching entrepreneurships' role in sustainable development. *Trailblazing in entrepreneurship* (pp. 149–179). Springer.

Shepherd, D. A., Patzelt, H., & Baron, R. A. (2013).
i care about nature, but...": Disengaging values in assessing opportunities that cause harm. *Academy of Management Journal*, *56*(5), 1251–1273.

Signori, A., & Vismara, S. (2018). Does success bring success? the post-offering lives of equity-crowdfunded firms. *Journal of Corporate Finance*, *50*, 575–591.

Spence, M., Gherib, J. B. B., & Biwolé, V. O. (2011). Sustainable entrepreneurship: Is entrepreneurial will enough? a north–south comparison. *Journal of Business Ethics*, *99*(3), 335–367.

Stål, H. I., & Bonnedahl, K. (2016). Conceptualizing strong sustainable entrepreneurship. *Small Enterprise Research*, *23*(1), 73–84.

Stubbs, W. (2017). Sustainable entrepreneurship and b corps. *Business Strategy and the Environment*, *26*(3), 331–344.

Terán-Yépez, E., Marın-Carrillo, G. M., del Pilar Casado-Belmonte, M., & de las Mercedes Capobianco-Uriarte, M. (2020). Sustainable entrepreneurship: Review of its evolution and new trends. *Journal of Cleaner Production*, *252*, 119742.

Testa, S., Nielsen, K. R., Bogers, M., & Cincotti, S. (2019). The role of crowdfunding in moving towards a sustainable society. *Technological Forecasting and Social Change*, *141*, 66–73.

Tilley, F., & Young, W. (2009). Sustainability entrepreneurs. *Greener Management International*, (55).

Vanacker, T., Vismara, S., & Walthoff-Borm, X. (2019). What happens after a crowdfunding campaign? *Handbook of research on crowdfunding*. Edward Elgar Publishing.

Vega, G., & Kidwell, R. E. (2007). Toward a typology of new venture creators: Similarities and contrasts between business and social entrepreneurs. *New England Journal of Entrepreneurship*.

Venkataraman, S. (2019). The distinctive domain of entrepreneurship research. *Seminal ideas for the next twenty-five years of advances*. Emerald Publishing Limited.

Vismara, S. (2016). Equity retention and social network theory in equity crowdfunding. *Small Business Economics*, *46*(4), 579–590.

Vismara, S. (2019). Sustainability in equity crowdfunding. *Technological Forecasting and Social Change*, *141*, 98–106.

Volkmann, C., Fichter, K., Klofsten, M., & Audretsch, D. B. (2021). Sustainable entrepreneurial ecosystems: An emerging field of research. *Small Business Economics*, *56*(3), 1047–1055.

Vuorio, A. M., Puumalainen, K., & Fellnhofer, K. (2018). Drivers of entrepreneurial intentions in sustainable entrepreneurship. *International Journal of Entrepreneurial Behavior & Research*.

Walthoff-Borm, X., Vanacker, T. R., & Collewaert, V. (2018). Equity crowdfunding, shareholder structures, and firm performance. *Corporate Governance: An International Review*, *26*(5), 314–330.

Weerawardena, J., & Mort, G. S. (2006). Investigating social entrepreneurship: A multidimensional model. *Journal of world business*, *41*(1), 21–35.

York, J. G., O'Neil, I., & Sarasvathy, S. D. (2016). Exploring environmental entrepreneurship: Identity coupling, venture goals, and stakeholder incentives. *Journal of Management Studies*, *53*(5), 695–737.

York, J. G., & Venkataraman, S. (2010). The entrepreneur–environment nexus: Uncertainty, innovation, and allocation. *Journal of business Venturing*, *25*(5), 449–463.

Yunus, M., Moingeon, B., & Lehmann-Ortega, L. (2010). Building social business models: Lessons from the grameen experience. *Long range planning*, *43*(2-3), 308–325.

A Quantifying Startups' ESG Properties

A Machine-Learning Approach: Detailed Description

Textual analysis in economics, finance and accounting literature is mainly applied using a dictionary count approach, in which researchers rely on predefined word lists to extract information from textual data (Gentzkow et al., 2019). Should we rely on humans or machines to create these word lists?[26] The advantage of human wisdom in creating these word lists comes at the cost of subjectivity and requires substantial transparency. In the context of ESG measurement, subjectivity plays a crucial role as many of the available ESG scoring databases provide inconsistent ratings (Berg et al., 2020; Dimson et al., 2020).

In this paper, we choose a middle ground. On the one hand, the machine relies on itself in detecting the meaningful phrases in the context of startups' whitepapers (i.e., word embedding via word2vec), and on the other hand, we guide the machine to come up with the terminologies that are most relevant for our purpose, based on a set of seed words.

Our procedure to create the ESG-relevant lexicon is methodologically close to Li et al. (2020). First, we collect whitepaper documents and parse their textual contents. Second, we clean the text by performing standard prepossessing procedures and define the set of words and context-specific phrases. Third, we do word embedding using the word2vec method (Mikolov et al., 2013) to obtain vector representation of all the words and phrases that have appeared in the corpus of whitepapers. Fourth, we define a set of seed words that represent each of the three pillars of ESG. Fifth, we use our trained word2vec model and generate our word lists by finding the closest word and/or phrase to our seed words. In training our word2vec model, we accept all standard assumptions for the hyper-parameter tuning of the model. Specifically, we use the Python package provided by Li et al. (2020), which implements all the previous steps. Finally, we calculate the ESG intensity of each whitepaper using the generated word lists.

A.1 Text prepossessing

Before we feed the corpus (universe of texts) to our word2vec model, we apply standard text prepossessing procedures to ensure the efficiency of our ML training process.

First, we remove line breaks from the text and replace numbers/emails/URLs/phone numbers with the respective tags, i.e., "<num>", "<email>", "<url>", "<phone>".

Second, we employ the Stanford CoreNLP pipeline (Manning et al., 2014) to gen-

[26] See (Loughran & McDonald, 2020b) for a comprehensive discussion

erate a dependency representation of each sentence.[27] Figure A.1 shows the dependency representation for an example sentence: "The basis for the distribution of GNC in the domain of real economic activity is the loyalty system, this is the most important and central tool of the platform." This helps the machine to better understand the grammatical structure of the sentences, and enables it to form collocations, i.e., a collection of more than one word that tends to appear frequently together, like "initial_coin_offering". We treat these collocations as single words in the following steps.

Third, we remove the stop words, i.e., words that do not add much meaning to a sentence like 'the', 'as', 'of', etc., as well as punctuation marks. Note that this step must follow the creation of collocations, as they could consist of some stop words, such as in "as_well_as".

A.2 Word Embedding

Word embedding is a way to mathematically represent words and enables the machine to compare the semantic similarity of the words. Our word embedding approach relies on the revolutionary word2vec method developed by Mikolov et al. (2013). The idea behind word2vec is to use a shallow (only one hidden layer) neural network, which is trained to predict words in the neighborhood of an input word, by exploring all the sentences in the corpus. In other words, during the training phase, an input word is translated to a vector in the hidden layer (a), and then this vector should predict the neighboring word (b). After the training, the trained weights of the neural network for (a) would be able to create a vector of real numbers for any input word of the corpus.[28] If trained on a vast corpus, the results of this seemingly simple algorithm would be very precise. A famous example of a trained word2vec model would be that one could find the vector closest to the vector of word 'Queen' by subtracting the vector of 'man' from the vector of the word 'King' and add the results to the vector of the word 'woman' (i.e., $King - Man + Woman = Queen$).[29]

A.3 Seed Words

As the starting point for measuring ESG intensity of the startup whitepapers, we collect all the available Financial Times (FT) articles with the tag of "ESG Investing" or

[27] The CoreNLP pipeline incorporates several steps. The most important steps include 1) tokenization, i.e., breaking down the text to smaller language units like words, 2) lemmatization, i.e., converting a word to its base form (e.g., "coins" to "coin"), and 3) entity chunking, i.e., replacing the entities' names with a proper tag.

[28] The size of this vector is the same as the size of the hidden layer in the neural network. We use the same settings as in Li et al. (2020) and consider a vector of size 300 for the word representations.

[29] Like any other ML framework, word2vec has its limitations. See Nissim et al. (2020) for a discussion on interesting and humorous examples of word2vec predictions.

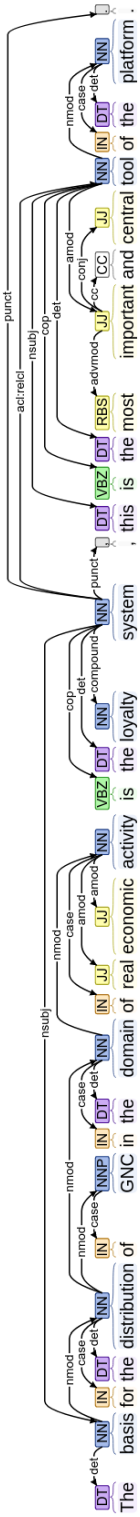


Figure A.1: Example of a dependency representation

"Moral Money". We follow a standard bag-of-words approach and extract bi-grams and tri-grams[30] that appear most frequently in the FT corpus. We then manually go through these n-grams and decide if they belong to the E, S or G dimensions of the ESG. As the

[30]Please note that bi-grams and tri-grams are two and three-word combinations of the words that appear in a neighborhood, and are not necessarily a collocation.

FT mostly covers the corporate world, it may not necessarily include the governance terms that are important for our context of ICO whitepapers. Therefore, we manually add terms like 'kyc', 'whitelist', 'blockchain', 'utility', 'security_token', etc. for the governance dimension. The full list of our seed words (available in Table A.1) consists of 70 Environmental, 38 Social, and 46 Governance related words/n-grams.

Table A.1: Seed Words

E	S	G
climate_change	moral_money	pension_funds
green_bonds	responsible_investing	investment_management
fossil_fuel	development_goals	supply_chain
green_bond	sustainable_development_goals	task_force
carbon_emission	impact_investment	investment_managers
carbon_footprint	social_issues	chief_investment_officer
renewable_energy	uns_sustainable	governance_issues
global_warming	social_impact	private_sector
greenhouse_gas	positive_impact	hedge_funds
climate_risk	essential_forward_thinking	managing_director
energy_source	gender_diversity	shareholder_proposals
green_finance	developing_countries	due_diligence
greenhouse_gas_emissions	decentralized	stakeholder_capitalism
carbon_footprint	defi	retail_investors
paris_climate	democratize	annual_meetings
climate_change_meets	democratization	esg_disclosure
paris_agreement	disintermediation	law_firm
fuel_companies	africa	global_advisors
fossil_fuel_companies	poor	board_members
climate_crisis	catching_up	investors_looking
natural_gas	india	passive_managers
environmental_impact	mobile	institutional_investors
thermal_coal	mobility	advisors
force_climaterelated_disclosures	cell_phone	bounty
green_bond_market	smart_phone	kyc
climaterelated_risks	access	whitelist
green_energy	geography	blockchain
low_carbon	dispersion	utility
oil_gas_companies	microfinance	security_token
environmental_issues	micro_finance	token_distribution
carbon_dioxide	impact_investing	intermediary
zero_emissions	equality	law
indispensable_energy	inequality	regulation
bn_green	care	policy
carbon_pricing	income	regulator
green_deal	responsible_investment	token_retention
carbon_neutral	impact_investing	airdrop
fight_climate_change	csr	founder
carbon_price		partner
coal_power		compliance
green_bonds		howey_test
fossil_fuel		sec
tackle_climate		equity
lowcarbon_economy		venture_capital
co_emissions		VC
risks_climate		incubator
zero_carbon		
green_investment		
risks_climate_change		
green_credentials		
reduce_carbon		
action_climate		
save_planet		
green_debt		
greenhouse_gases		

coal_projects		
away_fossil		
climate_accord		
carbon_credits		
first_green		
environmental_standards		
un_climate		
new_green		
netzero_carbon		
solar_wind		
renewable_energy		
global_warming		
sustainable_investing		
sustainable_investment		
sustainable_development		

A.4 ESG wordlists

For any term t of the seed words in any of the ESG dimensions j, we obtain a vector representation with the size of 300 (the size of the hidden layer in our word2vec model) as $V^t_{j\in\{E,S,G\}} = [x^t_1, x^t_2, ..., x^t_{300}]$. We then calculate the average vector for each of the ESG dimensions as $\bar{V}^{j\in\{E,S,G\}} = \frac{1}{N}\sum_1^N [x^t_1, x^t_2, ..., x^t_{300}]$ where N is the size of seed words for the dimension j. This leaves us with three vectors of $\bar{V}^E$, $\bar{V}^S$, and $\bar{V}^G$.

Next, we perform a cosine similarity between $\bar{V}^j$ and the vector of all of the terms in our whitepaper corpus and select the 500 most similar terms for each dimension. If a term appears in more than one dimension, then it is only considered for the dimension that has a higher cosine similarity.[31] Furthermore, some of our seed words have never appeared in the corpus of whitepapers.[32] We did not remove them from our word lists, though not affecting our results at all, as these terms could be relevant for future out-of-sample studies. This leaves us with a total of 1,495 ESG-related terms consisting of 508, 463 and 524 terms in the respective ESG dimensions.

A.5 ESG Score

We quantify the E, S and G dimensions using a dictionary-based approach, by counting the number of distinct occurrences of our respective word list in the ICOs whitepapers, normalized to the size of the word list. Specifically, for ICO i we measure each dimension of the ESG as:

$$E[S\,or\,G]_i = \frac{\sum_t 1_{c(t)_i>0}}{c(n)}, \tag{8}$$

[31]This is the reason why some dimensions could have a word list smaller than 500.
[32]This is the reason why some dimensions could have a word list greater than 500.

Where $c(t)_i$ is the count of term t in the whitepaper of ICO i and $c(n)$ is the size of the corresponding *word list*.

According to Loughran and McDonald (2020a), this approach slightly deviates from the norm in accounting and finance literature, where researchers count the total frequency of the words in a word list and normalize it to the total words in the document. In our context, however, this will lead to biases. Unlike corporate disclosures, ICO whitepapers are neither standardized nor regulated, and they vary substantially in length, format and content. Moreover, some ICOs have the words like 'green' or 'human' in their titles, which leads to bias in measuring the environmental or social score if a traditional frequency count method is applied.

Furthermore, we measure the total ESG score of the startup i by adding the three dimensions' intensity, i.e. $ESG_i = E_i + S_i + G_i$.

B Additional Controls

In this section, we check the robustness of our findings by including additional control variables to our baseline model. We control for the following additional controls: # investors, KYC, ICO duration, fiat accepted, % distributed in ICO, Twitter followers, LinkedIn, and crypto experience.

 # Investors. The logarithm of the number of institutional investors, as listed on the *CryptoFundResearch* list.

KYC. A dummy variable that equals one if the firm has a Know-Your-Customer (KYC) procedure, and zero otherwise.

ICO duration. The difference in days between the start and end of the ICO.

Fiat accepted. A dummy variable that equals one if the ICO accept fiat currencies.

Distributed in ICO. The percentage of tokens distributed in the token offering (i.e., 1 - "Distributed in ICO" is the token retention ratio).

Twitter followers. The logarithm of the number of the firm's Twitter followers.

Linkedin. A dummy variable that equals one if the ICO has a Linkedin page.

Crypto experience. The percentage of the team members who have experience in the crypto environments.

Table A.2 reports the results of this analysis. Adding the additional controls reduces our observations from 1043 in column (1) to 808 in column (5), which has the highest number of control variables. Our main results do not qualitatively change in these specifications. In all specifications, the coefficient on the normalized ESG score remains statistically significant at least at 5%.

Table A.2: Robustness Tests: Additional Controls

Column	(1)	(2)	(3)	(4)	(5)
	Dependent variable: Valuation, Funding amount (log.)				
ESG	0.276***	0.234***	0.223**	0.198**	0.197**
	(0.085)	(0.088)	(0.097)	(0.099)	(0.100)
GR	-0.356	-0.151	0.192	0.121	0.161
	(0.755)	(0.761)	(0.849)	(0.872)	(0.872)
$Investors$		0.690***	0.605***	0.554***	0.547***
		(0.093)	(0.113)	(0.114)	(0.112)
KYC		0.186	0.257	0.256	0.261
		(0.165)	(0.180)	(0.184)	(0.183)
$ICO\ Duration$		0.002	0.002	0.002	0.002
		(0.001)	(0.002)	(0.002)	(0.002)
$Fiat\ Accepted$			0.427	0.509	0.505
			(0.304)	(0.318)	(0.317)
$Distributed\ in\ ICO$			-0.676*	-0.608	-0.633
			(0.406)	(0.412)	(0.411)
$TwitterFollowers$				0.103**	0.096**
				(0.042)	(0.042)
$Linkedin$					0.058
					(0.180)
$CryptoExperience$					0.507*
					(0.302)
Observations	1043	1039	835	808	808
R^2	0.314	0.345	0.368	0.369	0.372
Controls	✓	✓	✓	✓	✓
Quarter-year FEs	✓	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓	✓

C Other Seed Words

In this section, we address potential concerns that our results could be driven by our manual selection of the seed words. To this end, we repeat the steps in generating our word lists with the exception that we consider only two or three seed words for each dimension of the ESG. Specifically, we set the seed words to be ['environmental', 'climate'] for the E dimension, ['society', 'social_responsibility'] for the S dimension, and ['governance', 'white_paper', 'token'] for the G dimension. Figure A.2 illustrates the resulting word lists, and it shows that we are able to capture the most relevant terms needed to construct our ESG word lists with only two or three words.

C.1 Other Seed Words and Funding

To test the validity of the word lists created with the small set of seed words, we repeat our baseline (OLS) regression with the log of the funding amount in $ million as the dependent variable, on the ESG score as well as its components derived from these word lists.

Table A.3 shows the results of this analysis. In column (1), the coefficient of the normalized ESG score is 0.34, with a p-value $<$ 1%, suggesting that a one standard deviation increase in the ESG score increases the average funding amount of $15.2 million by $6.1 million, or 40%. Columns (2), (3) and (4) report regression coefficients for the disaggregated and normalized E, S and G scores, respectively. All disaggregated scores are statistically significant at the 1% level in these models. The E score coefficient is 0.138 (p-value $<$ 0.01), the S score coefficient is 0.212 (p-value $<$ 0.01), and the G score coefficient is 0.321 (p-value $<$ 0.01). However, testing the effect of the three disaggregated scores simultaneously in column (5) shows that only the E (0.123) and the G (0.301) score are statistically significant at least at the 5% level. Thus, ceteris paribus increases by one standard deviation in E and G are associated with 13% and 35% increases in the average funding amount, respectively. These results are in line with the paper's analysis and strongly support the VPH that there is a sustainability-related valuation premium in token offerings.

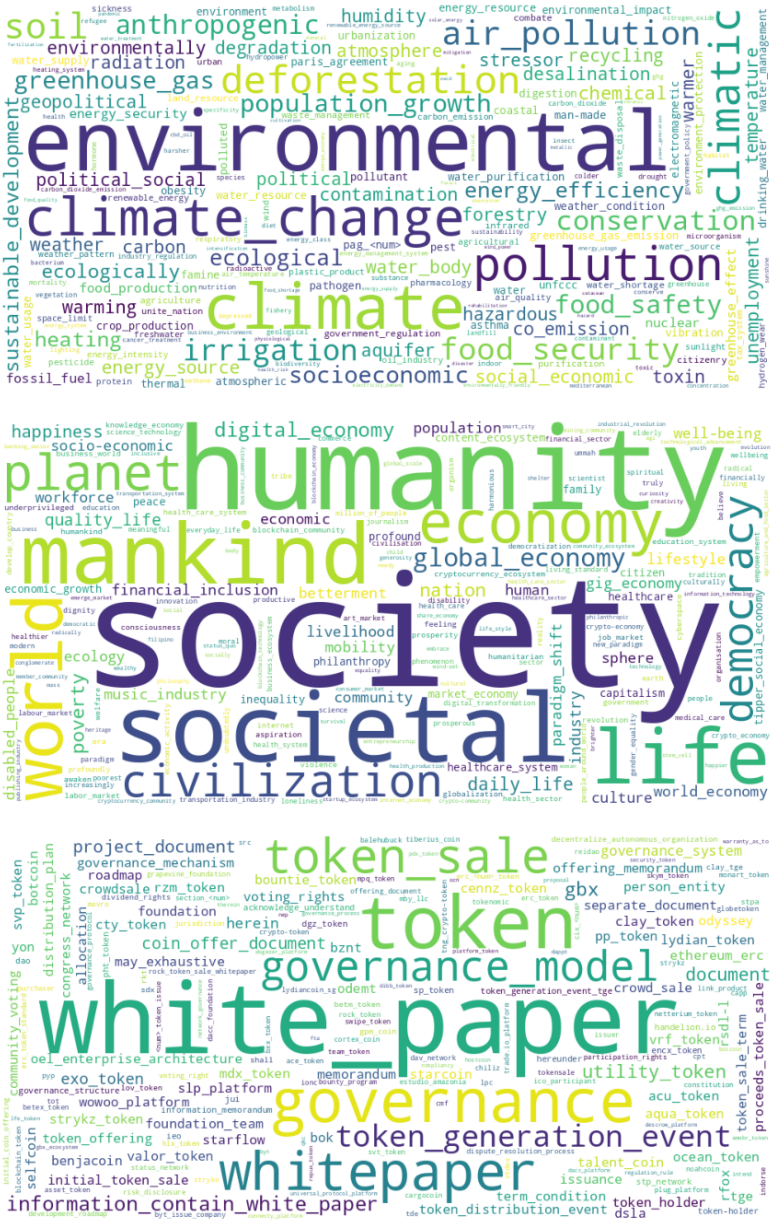


Figure A.2: Robustness Tests: ESG word lists with alternative seed words

Table A.3: Robustness - Seed words

Column	(1)	(2)	(3)	(4)	(5)
Dependent variable: Valuation, Funding amount (log.)					
ESG	0.338***				
	(0.070)				
Environmental		0.138***			0.123**
		(0.052)			(0.059)
Social			0.212***		0.037
			(0.074)		(0.091)
Governance				0.321***	0.301***
				(0.077)	(0.088)
Observations	1043	1043	1043	1043	1043
R^2	0.318	0.310	0.311	0.318	0.322
Controls	✓	✓	✓	✓	✓
Quarter-year FEs	✓	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓	✓

Curriculum Vitae

Sasan Mansouri | Curriculum Vitæ

Lyober Straße 52, B1.1.01, – 60528, Frankfurt a.M., Germany
(+49) 015781284258 • Mansouri@uni-frankfurt.de • www.sasanm.de
Born on 11th September 1991, Sabzevar, Iran

Education

Goethe Universität Frankfurt

Dr. rer. pol. in Economics — *Expected January 2022*
Research assistant at the Chair of Banking and Finance (Lehrstuhl für Bankbetriebslehre)
Thesis Title: Essays on Information Transmission & Machine Learning in Finance
Supervisors: Prof. Dr. Mark Wahrenburg & Prof. Dr. Alexander Hillert

Goethe Universität Frankfurt

Master of Science in Quantitative Economics — *2017*
Graduate School of Economics, Finance & Management(GSEFM)

Sharif University of Technology

Bachelor of Science in Petroleum Engineering — *2014*
Tehran, Iran

Publications (SSRN)

Corporate Culture and Banking (joint with Andreas Barth) - Journal of Economic Behavior & Organization, June 2021

Working Papers

Financing Sustainable Entrepreneurship: ESG Measurement, Valuation, and Performance in Token Offerings (joint with Paul P. Momtaz)

ICO Analysts (joint with Andreas Barth, Valerie Laturnus, and Alexander F. Wagner)

Does firm's silence drive media's attention away?

"Let me get back to you" - A machine learning approach to measuring non-answers (joint with Andreas Barth, and Fabian Woebbeking) - In Revision

How to talk down your stock performance (joint with Andreas Barth, Fabian Woebbeking, and Severin Zoergiebel) - In Revision

Experiences

Teaching Experiences - Master

2017-21,7 Semesters: TA of Master Course Mergers and Acquisitions, Prof. Dr.Mark Wahrenburg & Mr. Jan P. Weidner, Goethe University Frankfurt
This course placed as the best-ranked master course of the Faculty of Economics and Business in SoSe 2017, SoSe2019

WS 2019-20: TA of Master Course, Managerial Economics - Game Theory, Prof. Dr. Jenny Kragl & Elena Jarocinska, PhD, EBS universität wiesbaden - Evaluation Score 1.8 (1 = best; 5 = worst)

Teaching Experiences - Bachelor

WS 2020-21: TA of Bachelor Course, Mathematics 1, Dr. Laura Turrini, EBS universität wiesbaden - Evaluation Score 1.7 (1 = best; 5 = worst)

SS 2020: TA of Bachelor Course, BSC Microeconomics 1 & 2, Dr. Clemens Buchen, EBS universität wiesbaden - Evaluation Score 1.6 (1 = best; 5 = worst)

2013-14, 3 Semesters: TA of Bachelor Course Computer Programming Python, Dr.Sharareh Alipour, Sharif University of Technology Tehran Campus and International Campus

Conferences, Seminars & Summer schools

July. 2021: The 2021 Australasian Meeting of the Econometric Society (ESAM) - Presentation

July. 2021: Annual Accounting Finance Association of Australia and New Zealand (AFAANZ) Conference - Presentation

June. 2021: Annual Conference of the Canadian Academic Accounting Association - Presentation & discussion

May. 2021: European Accounting Association Virtual Annual Congress - Presentation

May. 2021: 37th International Conference of the French Finance Association (AFFI) - Presentation & discussion

Mar. 2021: Annual Conference of the Swiss Society for Financial Market Research, Zurich - Presentation

Dec. 2020: EUROFIDAI Paris December 2020 Finance Meeting - Presentation & discussion

April. 2020: Finance Brown Bag Seminar at Goethe University Frankfurt - Presentation

Mar. 2020: Digitale Transformation - 82. Jahrestagung des VHB - Presentation & discussion

Dec. 2019: 7th Paris Financial Management Conference (PFMC-2019) - Presentation & discussion

Sep. 2019: Summer School in Machine Learning in Economics and Businuss with Prof. Dr. Johannes Binswanger - Johannes Gutenberg University Mainz

Jul. 2018: International Finance and Banking Society (IFABS 2018 Porto Conference), Porto - Presentation

Apr. 2018: 21st Annual Conference of the Swiss Society for Financial Market Research, Zurich - Presentation & discussion

Nov. 2017: Universitätsübergreifendes Doktorandenseminar, Eltville - Presentation

Teamwork Experiences

Macroeconomic Measures-Early Warning Indicators — **SAFE Research Center**
Developmer at SAFE Systemic Financial Risk Platform(SFRP), University Frankfurt am Main — *2015-16*

Fixed-income and Hybrid indices Development Team — **Concerto Financial Solutions GmbH**
Internship, Frankfurt am Main — *Apr-Jul 2016*

20thNational Food Conference(20th NFC) — **Sharif University of Technology**
Vice-chair, Tehran, Iran — *2011*

Honors and Awards

2021: Lazaridis Institute Prize - "Let me get back to you: a machine learning approach to measuring non-answers" has been selected as the best paper on accounting issues relevant to technology firms

2020: Best conference paper award for "Econlinguistics", Digitale Transformation - 82. Jahrestagung des VHB

2015: Germany Scholarship (**Deutschlandstipendium**) 2015/16

2010: Ranked among the top **0.1** percent students in the Iranian University Entrance Exam

Computer skills

Programming: **Python** Data science with Pandas, Numpy, Data Visualisation (Matplotly, Seaborn, & Plotly), Textual Analysis (NLTK & Genism), Web Crawling (Beautifulsoup & Selenium), and Deep-learning (Tensorflow, PyTorch, & Keras), Flask,**R**,**SQL**, **MATLAB**(including GUI), **STATA**, **Html**, **LaTeX**, **Microsoft Office**(including VBA)

Data Works: Thomson Reuters Datastream,Bloomberg Terminal

Personal

Languages

Persian: Native, **English**: Fluent, **German**: Upper-intermediate (Level B2)

Memberships

Lindau Alumni Network - Lindau Nobel Laureate Meetings

Gesellschaft für Bildung und Physikalischer Verein - Frankfurt am Main

Hobbies

Squash, Bouldering, Swimming, Books (↗)

Ehrenwörtliche Erklärung

Ich habe die vorgelegte Dissertation selbst verfasst und dabei nur die von mir angegebenen Quellen und Hilfsmittel benutzt. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind sowie alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Datum und Unterschrift