Université de Neuchâtel
Faculté des sciences économiques

SOCIAL MEDIA AND TEXTUAL DATA: THREE EMPIRICAL STUDIES

Thèse présentée à la Faculté des sciences économiques

Institut d'analyse financière

Université de Neuchâtel

Pour l'obtention du grade de docteur en Finance

par

Arnaud Thierry MAÎTRE

Acceptée par le jury de thèse :

Prof. Florian Weigert, Université de Neuchâtel, directeur de thèse

Prof. Carolina Salva, Université de Neuchâtel, Présidente du jury

Prof. Sebastian Stöckl, University of Liechtenstein, Liechtenstein

Prof. Tim A. KRÖNCKE, Fachhochschule Nordwestschweiz

Soutenue le 21 Novembre 2025

Université de Neuchâtel
Faculté des sciences économiques

IMPRIMATUR POUR LA THÈSE

Social Media and Textual Data : Three Empirical Studies

Arnaud MAÎTRE

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES ÉCONOMIQUES

La Faculté des sciences économiques,

sur le rapport des membres du jury

Prof. Florian WEIGERT, Université de Neuchâtel, directeur de thèse

Prof. Carolina SALVA, Université de Neuchâtel

Prof. Tim A. KROENCKE, FHNW

Prof. Sebastian STÖCKL, Universität Liechtenstein

autorise l'impression de la présente thèse.

Neuchâtel, le 9 décembre 2025

Le doyen
Adrian Holzer

Faculté des sciences économiques
Avenue du 1er-Mars 26
CH-2000 Neuchâtel
Email: secretariat.seco@unine.ch
Tel: +41 32 718 15 00
www.unine.ch/seco

Acknowledgments

Here I am after four years of hard but rewarding work. First of all, I would like to thank my thesis supervisor, Florian Weigert, and my wife, Emma Pagnon, for their unwavering support throughout this adventure. Without their help, this thesis would not have been possible.

I am also grateful to my current and former colleagues at the IAF. It was thanks to the professors at the IAF that my interest in research was awakened. I am grateful to my fellow doctoral students and Kira Facchinetti, who have always supported and accompanied me.

Although I am happy to have completed my studies, I feel melancholic. I know that I will always look back on my doctoral years with a smile.

Bientôt, le premier marteau résonna contre le coin de bois ferré qui enfonçait un cercle sur la partie renflée d'un tonneau, une varlope gémit dans un noeud de bois, et l'une des scies, lancées par Esposito, démarra avec un grand bruit de lames froissées. Saïd, à la demande, apportait des douelles, ou allumait les feux de copeaux sur lesquels on plaçait des tonneaux pour les faire gonfler dans leur corset de lames ferrées. Quand personne ne le réclamait, il rivait aux établis, à grands coups de marteau, les larges cercles rouillés. L'odeur des copeaux brûlés commençait de remplir le hangar. Yvars, qui rabotait et ajustait les douelles taillées par Esposito, reconnut le vieux parfum et son coeur se desserra un peu. Tous travaillaient en silence, mais une chaleur, une vie renaissaient peu à peu dans l'atelier. Par les grands vitrages, une lumière fraîche remplissait le hangar. Les fumées bleuissaient dans l'air doré...

— Albert Camus, *Les muets*

Abstract

In this thesis, I investigate how unstructured data and alternative data sources can be used to answer new questions and to revisit old problems in the finance literature. As our world becomes increasingly digitalized, the research opportunities available to researchers are widening as well. Textual data is extremely rich and therefore valuable. Since causal analyses are not always possible, texts can be used to understand why companies or investors take certain actions. Despite being on separate topics, the three chapter of this PhD thesis share this ideal to leverage the latest technology to advance research a tiny bit.

The first chapter investigates how investors' abnormal attention affects the cross-section of cryptocurrency returns in the period from 2018 to 2022. We capture abnormal attention using the (log) number of Twitter posts on individual cryptocurrencies on the current day minus a 30-day average. Our results reveal that abnormal attention is positively associated with contemporaneous and one-day ahead crypto performance. Among the different Twitter tweets, return predictability arises due to *Ticker*-tweets from investors, but not due to tweets from the cryptocurrency channel. These *Official*-tweets, however, are able to forecast technological innovations on the blockchain.

In the second chapter, I use negotiation transcripts to capture advisor efforts during the private negotiation of US M&A deals. I show that the average effect of advisor services on deal outcomes is beneficial both for the target and the acquirer. Further sentiment analysis of the negotiation transcripts suggests that firms mainly benefit from the ability of their advisors to negotiate better merger terms. I document that advisors are skilled at identifying differences in negotiation power between the merger candidates. When differences in bargaining power are larger, advisor effects become larger and their negotiation strategy becomes more aggressive.

The third chapter show that firms use business combinations to manage their exposure to some of their business risks. Using mandatory corporate risk disclosures, I find that companies with increasing exposure to financial risks are more likely to become the target of a potential acquirer, and become less likely to acquire another company within twelve of the 10-K filing date. In contrast,

firms with increasing cash-flow uncertainty are more likely to acquire another firm. Analyses of disclosed deal objectives and merger negotiations confirm that firms are indeed merging to manage their risks exposures. At the industry level, I show that variations in similarity between firms filings predict negatively the future number of M&A deals announced. As industry risks become more concentrated, companies have less incentives to merge. Overall, the results are consistent with the mechanism whereby firms decide to merge in part to mitigate some of their business risks.

Keywords: cryptocurrencies, Bitcoin, Twitter attention, textual sentiment, mergers and acquisition, investment banks, acquisition premium, textual analysis, machine learning, finBERT, announcement returns, takeover probability, SEC filings

Résumé

Dans cette thèse, j'étudie comment les données non structurées et les sources de données alternatives peuvent être utilisées pour répondre à de nouvelles questions et revisiter d'anciens problèmes dans la littérature financière. À mesure que notre monde se numérise, les possibilités de recherche s'élargissent également pour les chercheurs. Les données textuelles sont extrêmement riches, et donc précieuses. Comme les analyses causales ne sont pas toujours possibles, les textes peuvent être analysés pour comprendre pourquoi les entreprises ou les investisseurs entreprennent certaines actions. Bien qu'ils traitent de sujets distincts, les trois chapitres de cette thèse de doctorat partagent cet idéal qui consiste à tirer parti des dernières technologies pour faire avancer la recherche d'un petit pas.

Le premier chapitre étudie comment l'attention anormale des investisseurs affecte la section transversale des rendements des crypto-monnaies entre 2018 et 2022. Nous capturons l'attention anormale en utilisant le nombre (log) de messages Twitter sur les crypto-monnaies individuelles le jour en cours moins une moyenne de 30 jours. Nos résultats révèlent que l'attention anormale est positivement associée à la performance contemporaine et à celle à un jour des crypto-monnaies. Parmi les différents tweets, la prévisibilité des rendements est due aux tweets *Ticker* des investisseurs, mais pas aux tweets du canal des crypto-monnaies. Ces tweets *Officiel* sont toutefois capables de prévoir les innovations technologiques sur la blockchain.

Dans le deuxième chapitre, j'utilise des transcriptions de négociations pour mesurer les efforts déployés par les conseillers lors des négociations privées d'opérations de fusion-acquisition aux États-Unis. Je montre que l'effet moyen des services des conseillers sur les résultats des transactions est bénéfique tant pour la cible que pour l'acquéreur. Une analyse plus approfondie des transcriptions des négociations suggère que les entreprises bénéficient principalement de la capacité de leurs conseillers à négocier de meilleures conditions de fusion. Je montre que les conseillers exploitent les différences de pouvoir de négociation entre les candidats à la fusion. Lorsque les différences de pouvoir de négociation sont plus importantes, l'effet des conseillers est plus marqué et leur stratégie de négociation devient plus agressive.

Le troisième chapitre montre que les entreprises ont recours aux fusions et acquisitions d'entreprises pour gérer leur exposition à certains risques commerciaux. À partir des informations obligatoires sur les risques divulguées par les entreprises, je constate que celles qui sont de plus en plus exposées aux risques financiers sont plus susceptibles d'être la cible d'un acquéreur potentiel et moins susceptibles d'acquérir une autre entreprise dans les douze mois suivant la date de dépôt du formulaire 10-K. À l'inverse, les entreprises inquiètent de l'incertitude grandissante de leurs futurs cash-flows sont plus susceptibles d'acquérir une autre entreprise. L'analyse des objectifs déclarés de ces fusions confirme que les entreprises fusionnent effectivement pour gérer leur exposition aux risques. Au niveau sectoriel, je montre que les variations dans la similitude entre les déclarations des entreprises permettent de prédire négativement le nombre futur d'opérations de fusion-acquisition annoncées. À mesure que les risques sectoriels se concentrent, les entreprises ont moins d'intérêts à fusionner. Dans l'ensemble, les résultats sont cohérents avec le mécanisme selon lequel les entreprises décident de fusionner en partie pour diversifier et atténuer certains de leurs risques commerciaux.

Mots-clés: cryptomonnaies, Bitcoin, attention de Twitter, sentiment textuel, fusions et acquisitions, banques d'investissement, premium d'acquisition, analyse textuelle, apprentissage automatique, finBERT, rendements à lors de l'annonce, probabilité de reprise, documents de la SEC.

Contents

List of Tables

List of Figures

Chapter 1

Social Media-Based Attention and the Cross-Section of Cryptocurrency Returns

with Nikolay Pugachyov and Florian Weigert
Published in the Journal of Banking & Finance[1]

1.1. Introduction

Social media has significant effects on financial markets, as illustrated by the involvement of the WallStreetBets subreddit in the Gamestop short squeeze and the infamous "to the moon" tweets, leading to subsequent Dogecoin price hikes.[2] Internet platforms are particularly important for cryptocurrencies, as they constitute the preferred medium of information exchange between cryptocurrency market participants, with Twitter arguably being the most important platform.[3] To put the importance of social media, and especially Twitter, for cryptocurrencies in context, about 90% of cryptocurrencies have Twitter accounts, whereas this proportion is only 50% for US public firms (Hosseini et al., 2020).

Given the size of the cryptocurrency market and the significance of Twitter as an information source, studying their relationship is essential for understanding the cross-section of expected cryptocurrency returns.[4] Recent literature provides mixed evidence on the link between Twitter activity and cryptocurrency prices. In the cross-section, Benedetti and Kostovetsky (2021) provide evidence supportive of an overreaction channel consistent with Barber and Odean (2008) and Da, Engelberg, and Gao (2011), while Borri et al. (2022) identify a negative risk premium for investor attention. In the time-series, Liu and Tsyvinski (2021) and Borri et al. (2022) report a positive relationship between the number of tweets and future cumu-

[1] `https://doi.org/10.1016/j.jbankfin.2025.107518`

[2] The topic has received wide attention in the empirical finance literature. Previous research shows that social media content has predictive power over expected stock returns and expected earnings (Chen et al., 2014; Bartov, Faurel, and Mohanram, 2018; Broadstock and Zhang, 2019; Gu and Kurov, 2020).

[3] The first Bitcoin transaction was arranged on a forum `https://techcrunch.com/2016/01/02/why-bitcoin-matters/` followed by the first-ever Bitcoin-related tweet as early as January 11$^{\text{th}}$, 2009 `https://twitter.com/halfin/status/1110302988?refsrctwsrc5Etfw`.

[4] According to Forbes, August 2024, the global cryptocurrency market cap is US$ 2.25 trillion.

lative cryptocurrency returns. Altogether, the link between Twitter-based investor attention and cryptocurrency expected returns remains unclear.

Inspired by Da, Engelberg, and Gao (2011), we inspect the relation between investor attention and cryptocurrency returns and capture abnormal investor attention computed as the (log) number of tweets during the current day minus the (log) mean number of tweets during the previous 30 days. The novelty of our paper is that we collect several different samples of tweets for each cryptocurrency. We consider both tweets written by the organization developing the cryptocurrency (*Official*-tweets) and tweets written by all other users, i.e., user-generated tweets. We further separate user-generated tweets into tweets on the cryptocurrency's ticker (*Ticker*-tweets) and tweets sent to the cryptocurrency's official account (*Mention*-tweets).

Our motivation to categorize tweets is that different categories have unique features, characteristics, and usage motivations. For instance, Barber et al. (2022) find that unique features of the Robinhood app partly drive attention-induced trading among the platform's investors. Likewise, *Ticker*-tweets possess the unique feature of being clickable and are commonly employed for discussing trading strategies. Once a user clicks on these tweets, they are presented with the latest tweets related to the associated financial security, making *Ticker*-tweets easily seen by users who are not following the tweet's author. Therefore, *Ticker*-tweets reach a wider audience than the other types of tweets. *Mention*-tweets can be viewed as public messages, as the account tagged by this type of tweet receives a notification. *Official*-tweets, on the other hand, represent corporate announcements. It is particularly interesting to investigate the possible impact of these features on crypto investors' trading behavior through the lens of attention-induced trading.

In our empirical analysis, we first consider the aggregated set of tweets, i.e., *All*-tweets. For each cryptocurrency, an abnormal attention measure similar to Da, Engelberg, and Gao (2011) is created. We show that abnormal attention is positively related to contemporaneous returns and next-day returns. On average, a one cross-sectional standard deviation rise in Twitter-based abnormal attention is associated with an increase in cryptocurrency excess returns by 0.70% contemporaneously and 0.11% on the following day. By utilizing our Twitter samples separately, we show that this effect is driven entirely by the user-generated tweets, with *Ticker*-tweets showing a stronger impact than *Mention*-tweets. Thus, unlike Benedetti and Kostovetsky (2021), we do not find a link between *Official*-tweets and excess returns in the cross-section, suggesting that the predictability of Twitter attention arises mainly from user-generated content.

Although the existing literature offers several possible explanations, we show that tweets written by Twitter users predict returns consistently in line with the continued overreaction channel. Boosts in tweet posting activity grab the attention of retail investors who, due to limited cognitive ability, become more likely to purchase these cryptocurrencies (Barber and Odean, 2008; Da, Engelberg, and Gao, 2011). This mechanism results in a temporary positive price pressure. Usually, price rises due to overreaction tend to correct over the following days after the surge in attention, which is not the case in our empirical data. In this regard, Schmeling, Schrimpf, and Todorov (2023) point out that limits-to-arbitrage are more important in the crypto market than for other asset classes. Hence, we hypothesize that, e.g., the difficulty of short-selling cryptocurrencies is likely to explain the absence of reversal in returns.

To better understand this difference in predictability power, we create a new lexicon to capture the textual characteristics of user-generated content following the methodology of Renault (2017). We apply this new lexicon to compute a score on the cryptocurrency-day level and show that this score predicts the cross-section of expected returns in the same fashion as abnormal attention. We interpret this finding as evidence that the predictability of user-generated content arises in part from its unique textual content. Furthermore, we show that our results are not driven by the GameStop short-squeeze in 2021, which was surrounded by high social media activity and positive performance for cryptocurrencies.

Given that social media is frequently utilized to report bugs, hacks, or technical problems with blockchain technology, a concurrent explanation of our results is that Twitter abnormal attention predicts returns through its link with future development activities. This hypothesis is motivated by the results of Cong, Li, and Wang (2021); Liu, Sheng, and Wang (2022) who reveal that cryptocurrency valuations and ICO success are linked to the quality of their underlying technology. By using the daily number of commits on GitHub as a proxy for technological innovation, we show that *Official*-tweets is the only Twitter sample predicting future technological improvements in the cross-section. As *Official*-tweets do not have predictive power for future returns, the technological innovation channel is not supported by our findings. This difference in predictability between the different Twitter samples supports the notion that social media content is indeed heterogeneous. It also highlights the need to carefully select the appropriate social media data that is best suited for the desired application.

Our paper contributes to two strands of literature. First and foremost, we add to quantify the impact of influential social media users on their followers (Benetton et al., 2024; Pedersen, 2022). We document that the posting activity of influential users increases the magnitude of the continued overreaction effect. Our results reveal that even small influencers can have a significant impact on cryptocurrency prices (Benetton et al., 2024). We argue that, similar to media coverage (Hillert, Jacobs, and Müller, 2014), influential tweets exacerbate behavioral biases. Therefore, our paper contributes to the nascent literature that explores behavioral biases induced by social media activity.[5]

Second, we add to the emerging literature that studies the impact of information salience on investor behavior (Bose et al., 2022; Kumar, Ruenzi, and Ungeheuer, 2021; Frydman and Wang, 2020). For instance, Barber et al. (2022) find that Robinhood users are more likely to purchase stocks displayed in the Top-Movers list than stocks with similar returns absent from the list. In our context, we document that *Ticker*-tweets are seen on average by a wider audience than the other types of tweets after controlling for their number of likes and retweets. Despite this increased visibility of *Ticker*-tweets over *Mention*-tweets, both Twitter samples predict returns similarly in the cross-section. This suggests that our results are not driven by information salience alone. Instead, we provide evidence that the difference in predictability between user-generated content and tweets written by cryptocurrencies is also driven by content differences.

Sections 1.2 and 1.3 describe the data and methods used in the paper. We investigate the drivers of Twitter-based attention in Section 1.4. Section 1.5 discusses the interplay between aggregate attention and cryptocurrency returns by breaking down attention into several components using the specificity of Twitter. The robustness of our main findings is addressed in Section 1.6. Finally, Section 1.7 concludes the paper.

1.2. Data and Sample Construction

We apply several datasets to construct our sample. Our primary datasets are data on cryptocurrency returns from `https://coinmarketcap.com/`CoinMarketCap and data on cryptocurrency-related posts on Twitter. We explain the construction of our Twitter sample in detail in Section 1.2.2. Additionally, we gather data on each cryptocurrency from GitHub, the leading platform for software development and project collaboration. We use data sampled at a daily

[5]For example, social media users tend to self-expose to information in line with their beliefs (Cookson, Engelberg, and Mullins, 2023) and are influenced by investment returns experienced by users they are following (Bailey et al., 2018; Pedersen, 2022).

frequency in our empirical tests, as Twitter effects documented by the literature are generally short-lived (Benedetti and Kostovetsky, 2021; Gu and Kurov, 2020).

1.2.1. Cryptocurrency data

CoinMarketCap (henceforth CMC) is a widely accepted data source for cryptocurrency market data. However, the data from the CMC website is subject to survivorship bias because it only provides information on currently listed cryptocurrencies. To address this, we use an API to download survivorship bias-free data from CMC, following the methodology proposed by Ammann et al. (2022) that corrects the bias.[6]

Our dataset is at a daily frequency for the period from 2018 to 2022. We drop assets with missing volume data and exclude cryptocurrencies with erroneously reported data. Additionally, we drop stablecoins, which are cryptocurrencies whose value is pegged to other assets such as USD or gold. Our final dataset includes both cryptocoins and tokens.[7] Figure 1.1 displays the average monthly number of cryptocurrencies meeting our criteria.

[Insert Figure 1.1 here]

1.2.2. Twitter data

Given the download restrictions imposed by Twitter, we limit our sample to the 165 largest cryptocurrencies as of the end of 2017.[8] We choose this year because it includes a large number of cryptocurrencies meeting our criteria. Our sample excludes cryptocurrencies that have changed names during our sample period or those that do not have a Twitter account.[9] For each selected cryptocurrency, we separately collect tweets written by the organization developing the cryptocurrency (henceforth *Official*-tweets) and tweets written by all other users, i.e., user-generated tweets. While

[6] We access the API through the R package named ***crypto2***, which can be found at `https://CRAN.R-project.org/package=crypto2`. The authors of this package are Sebastian Stöckl and Jesse Vent. More information about the package can be found on the author's personal website: `https://www.sebastianstoeckl.com/`

[7] Crypto coins primarily act as a medium of exchange and a store of value. Some well-known examples include Bitcoin (BTC), Ethereum (ETH), and Dogecoin (DOGE). Crypto tokens, on the other hand, are created to serve various purposes, such as utility, ownership, and governance rights. For instance, the Edgeless token can be used to play in an online casino.

[8] Due to the abrupt stop of the Twitter academic API in 2023, we restrict the Twitter samples to the 165 cryptocurrencies for which data could be fully collected.

[9] While it is true that excluding cryptocurrencies that changed their names during the sample period introduces a look-ahead bias, we believe that this bias is unlikely to have a substantial impact on our study's findings. The difference in mean returns between the two samples is not statistically significant from zero.

collecting *Official*-tweets is straightforward (simply by requesting tweets posted by the developers' verified accounts), collecting user-generated tweets is more complex. The complexity arises from the fact that users have several ways to signal that their tweet is about a specific cryptocurrency. For instance, users can utilize the name (e.g., #*bitcoin* or #*BTC*), the ticker (e.g., $*BTC*), or tag (mention) the official account of the cryptocurrency (e.g., @*Bitcoin*). We choose to collect tweets on the cryptocurrency's ticker (henceforth *Ticker*-tweets) and tweets sent to the cryptocurrency's official account (henceforth *Mention*-tweets) to exclude tweets with ambiguous hashtags.[10] Ultimately, we categorize tweets into three categories in our Twitter sample: *Official*-tweets, *Mention*-tweets, and *Ticker*-tweets.

Our motivation to categorize tweets is that different categories contain unique features, characteristics, and usage motivations. For instance, Barber et al. (2022) find that unique features of the Robinhood app partly drive attention-induced trading among the platform's investors. *Ticker*-tweets, commonly employed for discussing trading strategies, possess the unique feature of being clickable. When a user clicks on these tweets, they can view the latest tweets related to the associated financial security. Therefore, they have a potentially higher reach than other types of tweets, as they can be easily seen by users who are not following the tweet's author, thanks to the clickable feature. It is particularly interesting to investigate the possible impact of this feature on crypto investors' trading behavior through the lens of attention-induced trading. *Mention*-tweets can be viewed as public messages, as the account tagged by this type of tweet receives a notification. *Official*-tweets, on the other hand, represent corporate announcements.

Table 1.1 summarizes the information about the sample size and the characteristics of each sample. The number of cryptocurrencies in the *Official* sample is lower than in the other samples because we could not retrieve any *Official*-tweets for eleven cryptocurrencies. We chose to leave the history of *Official*-tweets as missing for those eleven cryptocurrencies because it is hard to tell whether those eleven cryptocurrencies had never posted any tweets during the sample period or had retroactively deleted their tweets.

[Insert Table 1.1 here]

Figure 1.2 reports the number of tweets posted per sample per day. In all our samples, the number of Twitter posts published per day and per cryptocurrency corresponds to the actual number of tweets published on that date. Twitter activity has increased over our sample period, with the only exception being the number of

[10]For instance, consider two cryptocurrencies named ICON (ICX) and TRON (TRX). Both #*icon* and #*tron* can be associated with different meanings, not just cryptocurrencies.

Official-tweets, which has remained stable over time. For all samples except the *Official* sample, we only download the 100 most relevant tweets per day over the period from 2018 to 2022.[11] The *Official* sample includes every tweet posted by the cryptocurrency's official account during the sample period.

[Insert Figure 1.2 here]

Unfortunately, as Twitter provides us with tweets based on textual matching of keywords, *Ticker*-tweets may contain measurement error, as tickers are generally not unique. We do not expect this bias to be substantial for two reasons. First, we only consider tweets written in English, which limits the number of ticker homonyms in our sample. Second, as cryptocurrencies are a popular topic on social media, we expect the *Ticker* sample to be primarily composed of tweets about cryptocurrencies.[12] Our Twitter samples are also subject to survivorship bias, as tweets and Twitter accounts can be deleted by their creators. This problem is more severe for the *Official* sample, as it is impossible to retrieve tweets from deleted accounts. Consequently, we miss *Official* tweets from cryptocurrencies with deleted Twitter accounts, which are likely to be defunct. However, also in this case, the bias is assumed to be small, as we could retrieve *Official* tweets from 35 out of the 44 defunct cryptocurrencies included in our Twitter samples.

1.2.3. GitHub data

To better identify the channels through which Twitter attention predicts the cross-section of cryptocurrency returns, we also collect data on each cryptocurrency from GitHub, the leading platform for software development collaboration. Specifically, we collect the list of all historical contributions (commits) made by developers on all repositories owned by the organization developing the cryptocurrency. Similar to *Official*-tweets, GitHub data is also subject to survivorship bias, as cryptocurrencies with missing data are likely defunct. Our GitHub data covers 143 cryptocurrencies and includes data for 32 out of the 44 defunct cryptocurrencies in our sample.[13]

[11] As determined by Twitter, the exact methodology is not disclosed but considers, among other factors, the degree of keyword matching, tweet engagement, and the author's popularity.

[12] We also report the results if ambiguous tickers are removed from our sample as a robustness check. We judge a ticker as ambiguous if a stock present in the CRSP database has a similar ticker during our sample period. For example, consider a cryptocurrency named Neo (ticker: $\$NEO$; `https://x.com/neo_blockchain`) and a publicly traded company on NASDAQ, NeoGenomics, Inc. (ticker: $\$NEO$; `https://x.com/NeoGenomics`).

[13] We miss data on 22 cryptocurrencies. Three cryptocurrencies use other platforms than GitHub to share their code. We could not find information about 18 cryptocurrencies, and one cryptocurrency's repository contains no commits.

1.3. Empirical Methodology

1.3.1. Abnormal Twitter-based attention

Inspired by Da, Engelberg, and Gao (2011), we define our main measure of investor attention as abnormal attention, which we compute as the (log) number of tweets during the current day minus the (log) mean number of tweets during the previous 30 days.[14] Specifically,

$$Abn\,Attention_{i,t} = Log(NT_t) - Log(E[NT_{t-30}, NT_{t-29}, ..., NT_{t-1}]), \tag{1.1}$$

where NT_t is the number of tweets on day t. Abnormal attention has the advantage of being less sensitive to large spikes in the number of tweets than the (log) number of tweets because it takes into account the average level of attention through its rolling mean component.

1.3.2. Twitter lexicon

The literature has not reached a consensus on which type of social media content is more relevant to empirical studies. Some authors focus on *Ticker*-tweets or other social media platforms oriented towards investing to capture the attention and sentiment of financially savvy users (Chen et al., 2014; Renault, 2017; Ardia and Bluteau, 2024; Gu and Kurov, 2020), whereas others look at content published by companies or aggregated content (Benedetti and Kostovetsky, 2021; Da, Engelberg, and Gao, 2011; Borri et al., 2022). Despite most studies focusing on one type of social media data, there is limited evidence documenting the effects of this decision. To fill this gap, we construct a new lexicon aimed at capturing the differences in textual content between *Ticker*-tweets and other tweets.

We construct our lexicon using a methodology similar to Renault (2017). First, we clean and process the text following common practices. We remove stop words from the tweets and replace words with bigrams when possible, using our own list of bigrams supplemented by the list of Renault (2017).[15] We remove all punctuation. Email addresses, web links, and emojis are replaced by keywords (mailtag, linktag, emojipos, or emojineg).[16] Following Renault (2017), we also add a prefix (negtag) to

[14] Da, Engelberg, and Gao (2011) rely on Google's Search Volume Index (SVI) to measure investor attention. Their main variable is abnormal SVI, which is defined as the (log) SVI during the current week minus the (log) median SVI during the previous eight weeks.

[15] For forming the bigrams, we use 300,000 randomly selected tweets over the complete sample.

[16] We manually classify the most frequent emojis as either being positive or negative.

tokens directly following a negation. Finally, we create a training sample composed of 180,000 tweets equally distributed across *Ticker*-tweets, *Mention*-tweets, and *Official*-tweets.

Importantly, the training sample does not contain any retweets. To avoid any look-ahead bias, the training sample is only composed of tweets published in 2017. We also filter out irrelevant unigrams and bigrams by restricting our sample to n-grams appearing in at least 0.01% of tweets. To prevent our final lexicon from being too influenced by prolific authors, cryptocurrencies, or seasonal topics, we require n-grams to be used by at least 10 different authors, to be used in discussions about at least 10 different cryptocurrencies, and to be used in at least three distinct months. Our last filtering step is to remove tickers and user mentions from the selected tokens to prevent our lexicon from loading on terms that are specific to one sample of tweets.

To spot unigrams and bigrams that are specific to *Ticker*-tweets, we compute a coefficient for each token i which is defined as:

$$Coefficient_i = \frac{Occurence_i^T - \max(Occurence_i^M; Occurence_i^O)}{Occurence_i^T + \max(Occurence_i^M; Occurence_i^O)}, \tag{1.2}$$

where T, M, and O stand for Ticker, Mention, and Official, respectively.

Like Renault (2017), we only keep in the lexicon n-grams with a coefficient in the first or the last quintile. The resulting lexicon is composed of 2,833 n-grams. Table 1.2 presents a selection of the lexicon's vocabulary. We observe that the lexicon loads positively on terms related to trading such as "bull–flag" or "sold–all" regardless of their sentiment and negatively on terms related to technology or news such as "support–team" or "fully-decentralized".

The lexicon confirms the intuition that different types of users use different types of tweets. Users tend to use mainly *Ticker*-tweets to discuss cryptocurrency investing. In contrast, *Official*-tweets and *Mention*-tweets are used for corporate announcements and to signal and solve issues related to the cryptocurrency. In the rest of the paper, we compute the lexicon score per tweet as the equal-weighted average coefficient of all n-grams present in the tweet that belong to the lexicon. We then compute the cryptocurrency-day lexicon score as the average score of all tweets published about that cryptocurrency on a given date. To illustrate our methodology, the hypothetical tweet *"@username please contact support team"* would get a score of $-0.8189 = \frac{-0.7353 + -0.9024}{2}$ where -0.7353 and -0.9024 are respectively the score assigned by the lexicon for the n-grams "contact" and "support-team".

[Insert Table 1.2 here]

1.3.3. *Variables description and summary statistics*

As in our Twitter data, we observe the number of likes, text, and author identity for a subset of the tweets posted per cryptocurrency-date; however, we are not able to identify the complete activity of tweet authors. Consequently, we choose to measure influential user activity in an indirect manner by checking if at least one influential user has tweeted on this cryptocurrency-date. We construct a dummy variable ($Popular\,Tweet_{i,t}^{All}$) equal to one when at least one tweet published per cryptocurrency-date has an aggregated number of likes and replies equal to or above 100 across all of our Twitter samples.[17] The number 100 corresponds to the 93$^{\text{th}}$ percentile of the distribution of the sum of the number of comments and likes. Classifying tweets with as few as 100 likes and/or comments as influential might seem optimistic. However, the actual number of users reached by a tweet is often much larger than its number of likes and replies. Using the number of views of each tweet, we find that the median number of views for tweets with at least 100 likes is 17,523 views and this number does not include the views obtained by the activity induced by popular tweets.[18] We choose to compute this variable across all Twitter samples simultaneously because we are interested in how influential users' activity, regardless of their tweeting preferences, impacts other Twitter users.

Sentiment- and lexicon-based variables are constructed at the tweet level and then averaged at the cryptocurrency-date level. To compute sentiment, we use a bag-of-words approach and consider the dictionary of Renault (2017), which is the main lexicon used for quantifying the sentiment of texts from social media in the financial academic literature.[19]

We follow Liu, Tsyvinski, and Wu (2022) to construct the cryptocurrency market factor. Throughout the paper, we control for a wide range of variables impacting asset prices. We control for illiquidity and liquidity fluctuations using the measure of Amihud (2002); Chordia, Subrahmanyam, and Anshuman (2001) and measures of codependence with the cryptocurrency market using co-skewness (Kraus and Litzenberger, 1976; Harvey and Siddique, 2000) and co-kurtosis (Fang and Lai, 1997; Jondeau and Rockinger, 2006). In addition, we also include controls related to volatility, higher-order moments of returns, and tail risk using variables such as idiosyncratic volatility (Ang et al., 2006), skewness, kurtosis, and value-at-risk.

[17] The number of retweets is excluded from this total. We do so because the number of retweets is shared across the retweets and the original tweet. Therefore, using the number of retweets to spot popular tweets can be misleading.

[18] We do not directly use the number of views to spot popular tweets in our analysis because this variable is only available for a limited number of tweets.

[19] Our results do not change if we use the lexicon of Loughran and McDonald (2011) instead.

Momentum is included in our set of controls, as momentum strategies are profitable for a wide range of asset classes (Asness, Moskowitz, and Pedersen, 2013) and are related to investor attention (Daniel, Hirshleifer, and Subrahmanyam, 1998; Hillert, Jacobs, and Müller, 2014). Hillert, Jacobs, and Müller (2014) find that media coverage of larger firms attracts investor attention (Solomon, Soltes, and Sosyura, 2014) and intensifies investors' overconfidence and self-attribution biases (Daniel, Hirshleifer, and Subrahmanyam, 1998), which result in temporary improvements in momentum returns. However, evidence from Liu and Tsyvinski (2021) suggests that the relationship between momentum returns and investor attention is different between cryptocurrencies and stocks. The authors find that the return predictability of attention and momentum does not encompass each other.

In addition, we add a control variable capturing the "MAX effect" (Bali, Cakici, and Whitelaw, 2011; Fong and Toh, 2014) which is related to investor sentiment and gambling preferences (Fong and Toh, 2014). The GameStop short-squeeze attracted a significant number of retail speculators and triggered a surge in investor attention (Lyócsa, Baumöhl, and Vỳrost, 2022). This event suggests that the "Max effect" and investor attention could be related. Table 1.3 provides a description of all variables used in our paper.

[Insert Table 1.3 here]

Given the suspicious returns documented by Ammann et al. (2022), we follow their approach and trim the daily cryptocurrency returns at the 99% level. The remaining variables are winsorized at the 1% level. Table 1.4 displays descriptive statistics for the variables used in our paper. In our data, abnormal attention is negative on average and is negatively skewed across all samples. Furthermore, we can see that tweets tend to have positive sentiment and to focus on terms related to trading as indicated by the positive mean of sentiment and lexicon scores. In terms of popularity on Twitter, cryptocurrencies differ significantly. For instance, Bitcoin gets a median number of tweets of 11,236 per day, while the median across all cryptocurrencies is 60 tweets. The low number of tweets for some assets is not purely driven by size; the smallest cryptocurrency we consider as a market cap of $47 million as of the start of our sample. Even if the median number of tweets seems low, one has to remember that the actual number of people reached by tweets is larger. The median number of views per tweet is about 118 and it does not include the number of times that tweets were collected by web-scrapping algorithms. Anecdotal evidence suggests that automatized data collection efforts can be sizeable. For instance, 27% of hedge funds surveyed by Ernst and Young in their "EY Global Hedge Fund and Investor Survey 2017" reports that they are using or planning to use social media data as part of their investment process.[20] Furthermore, the low popularity of some cryptocurrencies goes against us finding any return predictability for Twitter attention.

[Insert Table 1.4 here]

1.4. Characterizing Twitter-Based Investor Attention

1.4.1. Drivers of Twitter-based investor attention

We start our empirical analysis by investigating the drivers of Twitter-based investor attention. For this purpose, Table 1.5 reports the results on contemporaneous relationships between the different attention measures. We use changes in attention rather than abnormal attention because we are interested in what causes variation in attention more generally. The panel regression models control for the variables defined in Panel A of Table 1.3.

Table 1.5 reveals that attention variables are positively related to each other at the 1% or 5% significance level. Daily excess returns are also significantly linked to contemporaneous changes in attention for *Mention*-tweets and *Ticker*-tweets at the 1% level, but do not appear to be linked to the change in the number of *Official*-tweets. The presence of $Popular\,Tweet^{All}_{i,t}$ is strongly positively related to changes in the number of tweets across all three samples. This is in line with the results of Benetton et al. (2024) who document that celebrity tweeting activity influences their followers' behavior. Variables derived from volume, $\Delta\,Volume_{i,t}$ and $Volume\,Vola_{i,t}$, are significantly linked with the dependent variables. Changes in volume have a similar regression coefficient to excess returns, and volatility of volume is negatively linked with changes in the number of tweets from each sample. Interestingly, momentum is negatively related to attention changes, which suggests that the association between attention and momentum is less strong for cryptocurrencies compared to equities (Liu and Tsyvinski, 2021). We do not find strong links between change in tweeting activity and the other control variables described in Panel A of Table 1.3.

[Insert Table 1.5 here]

[20] https://eyfinancialservicesthoughtgallery.ie/wp-content/uploads/2017/12/ey-how-will-you-embrace-innovation.pdf

We run similar tests using future change in tweeting activity in Table 1.6. Future changes in the different changes in attention measures are inversely related to their own one-day lags at the 1% significance level. Interestingly, we observe differences in cross-one-day auto-correlations between attention variables. For instance, *Official*-tweets and *Ticker*-tweets positively predict future changes in the number of tweets for the other sample at the 1% significance level. On the other hand, *Mention*-tweets are negatively related to future changes in *Official*-tweets. As in Table 1.5, daily cryptocurrency returns and volume changes are strong predictors of the number of tweets, even for *Official*-tweets. Furthermore, our results reveal that coefficient estimates of $Popular\,Tweet_{i,t}^{All}$ become negative. We interpret these results as indicative that attention tends to revert to a normal level following large changes; Holding other factors equal, a *Popular*-tweet posted at time t stimulates the number of tweets posted at time t and reinforces the attention reversal at time $t+1$. Similarly to Table 1.5, momentum is negatively related to future changes in tweeting activity in all models.

Taken together, Table 1.5 and Table 1.6 both suggest that tweeting activity tends to spike and revert partially over the following day. As expected, the link between past returns and Twitter attention is sizeable (Da, Engelberg, and Gao, 2011; Liu and Tsyvinski, 2021) and is the strongest for *Ticker*-tweets, both in terms of statistical and economic significance. Intuitively, this observation makes sense given that *Ticker*-tweets are mainly used for discussing trading and investments as shown in Table 1.2. Despite Twitter users reacting to past and contemporaneous returns, change in attention is inversely related to momentum as documented by Liu and Tsyvinski (2021). The authors argue that if momentum arises from under-reaction to news (Hong and Stein, 1999), then momentum and change in attention should be negatively correlated as observed in our data.

While the different attention variables are positively related contemporaneously, we observe different patterns at $t+1$, especially for *Mention*-tweets, which become negatively related to the other samples. One could wonder why *Official*-tweets and *Ticker*-tweets are positively associated with future tweeting activity in other samples. We posit that those two samples of tweets are on average more visible and therefore induce more people to tweet. *Ticker*-tweets have a unique clickable feature which should increase their visibility, hence their impact. Alternatively, *Official*-tweets are posted by a recognized Twitter user and constitute an important source of information for cryptocurrency investors. We investigate our hypothesis in the following section.

[Insert Table 1.6 here]

1.4.2. Differences between the Twitter samples

The three categories of tweets used in this study have different characteristics. Their most important difference is how they are disseminated on Twitter. By default, a tweet is displayed on the profile page of the author and in the timeline of users following the author.[21] However, both *Ticker*-tweets and *Mention*-tweets differ from this default behavior. *Ticker*-tweets also appear in the respective financial security timeline. *Mention*-tweets can either follow the default behavior or be customized such that these tweets only appear in the timeline of users following both the author and the user being tagged. Therefore, *Ticker*-tweets should be more visible than classic tweets, whereas *Mention*-tweets should be less visible.

Several studies show that information salience impacts investor behavior (Barber and Odean, 2008; Barber et al., 2022). For instance, Tan, Wang, and Zhou (2015) find that better readability helps investors to better incorporate new information. Therefore, it is reasonable to expect that the Twitter samples predict returns differently in the cross-section based on their visibility. As a first step, we test in Table 1.7 whether the category of the tweet indeed impacts its visibility and other tweet metrics. The different models use date and cryptocurrency fixed effects to control for unobserved invariant characteristics and cluster standard errors by authors. In addition, we exclude retweets from each estimated model. As original tweets and their retweets all share the same retweet count, keeping retweets in the sample may lead to spurious relationships.

Our results reveal that being a *Ticker*-tweet or an *Official*-tweet positively relates to the number of views obtained by the tweet. This effect is statistically significant at the 1% and 5% levels for *Official*-tweets and *Ticker*-tweets, respectively. In addition, *Ticker*-tweets are also significantly associated with a larger number of retweets. In contrast, we find a negative relationship between *Mention*-tweets and the number of views obtained by the tweet at the 1% level. Textual sentiment is positively related to the number of likes, replies, and retweets at the 1% significance level, but negatively correlated with the number of views at the 1% level. The lexicon score, which captures trading jargon, is positively linked with the number of views at the 1% significance level, but negatively associated with the other dependent variables at the 1% significance level.

The results support our intuition that *Ticker*-tweets are more visible than classic tweets thanks to their clickable feature. In addition, we find similar results for *Official*-tweets, which is in line with our hypothesis that investors value the news announced by *Official*-tweets. In contrast, the regression coefficients of *Mention*-

[21] `https://help.twitter.com/en/using-x/types-of-posts`

tweets in the last model are negative. As expected, *Mention*-tweets are generally written with the intent to reach a specific set of users. Interestingly, $Lexicon_{i,n,t}$ is positively associated with the number of views and negatively associated with the number of retweets and likes. Social media users do not seem to actively share tweets with high lexicon scores, but they still search for them as indicated by the positive correlation with the number of views. We observe similar patterns for the effects of tweets with negative sentiment on tweet visibility. One potential interpretation could be that Twitter users actively search for bad news and other users' opinions as part of their investment process. Such behavior would be in line with the "Do your own research" (DYOR) advice which is frequently addressed to new cryptocurrency investors on social media.[22] Overall, the evidence contained in Table 1.7 confirms that the different characteristics of tweets matter and impact their salience on Twitter, thus providing evidence that the choice of social media is not innocuous.

[Insert Table 1.7 here]

1.5. Twitter-Based Investor Attention and the Cross-Section of Cryptocurrency Returns

1.5.1. Twitter-based abnormal attention

This section studies the link between abnormal attention and the cross-section of cryptocurrency returns. In Table 1.8, we present the results of panel regressions of excess returns on abnormal attention and various controls as defined in Panel A of Table 1.3. Model (1) investigates the association between Twitter attention and returns contemporaneously. The other models replace contemporaneous returns with future returns using different time horizons. One potential issue with *Ticker*-tweets is that their trading symbol is not unique, as trading symbols are attributed by the exchange where assets are traded. Given that cryptocurrencies are traded on their own exchanges, they can possibly share their ticker with other financial securities traded elsewhere such as US stocks. Therefore, for robustness, we replicate the same regressions with a different sample in Panel B; when doing so, we drop cryptocurrencies that share their ticker with a firm covered by the CRSP database during our sample period.[23]

[22]The unigram "dyor" has a score of 0.5673 in the lexicon.

[23]The cryptocurrencies that are dropped in Panel B include both large and small cryptocurrencies. For instance, both Bitcoin and Ethereum are dropped in Panel B.

Our results reveal that the regression coefficient of $Abn\,Attention_t^{All}$ is statistically significant and positive in models (1) and (2) at the 1% significance level. Evaluating its economic significance, the effect is large; a one standard-deviation increase in $Abn\,Attention_t^{All}$ is related to an increase of 0.72% in contemporaneous daily excess returns and an increase in future daily excess returns of 0.11%. The association between abnormal attention at day t and excess returns at $t+1$ hence amounts to an annual 39%. The predictability of Twitter activity is robust to a wide range of popular predictors used in the literature. Among those predictors, only past returns, size, and change in volume are all significantly related to contemporaneous and future cryptocurrency returns. Volume change predicts returns in a similar way as abnormal attention. Past returns predict future returns up to two periods ahead, and size is negatively related to returns in all model specifications, consistent with previous literature. Furthermore, we document similar results in Panel B where cryptocurrencies with ambiguous trading symbols have been removed, suggesting that our results are not driven by tweets that should not be included in our sample.

The positive connection between $Abn\,Attention_t^{All}$ and excess returns is consistent with the results of Liu and Tsyvinski (2021), who also document a positive relationship between investor attention and cryptocurrency returns. However, we need to be cautious when interpreting the results of Table 1.8, as several theories could explain why attention and (expected) returns are positively correlated. For instance, as Twitter is used to discuss the technology underlying cryptocurrencies, Twitter attention could predict technological improvements and therefore returns (Cong, Li, and Wang, 2021; Lyandres, Palazzo, and Rabetti, 2022; Liu, Sheng, and Wang, 2022). Another possibility is that attention predicts increases in the user base, which would have positive network externalities for existing cryptocurrency users (Cong, Li, and Wang, 2021; Sockin and Xiong, 2023).

Alternatively, the results could also be explained by an overreaction channel (Barber and Odean, 2008; Da, Engelberg, and Gao, 2011). Increases in tweet posting activity grab the attention of retail investors, who, due to limited cognitive ability, become more likely to purchase those attention-grabbing cryptocurrencies. This mechanism results in temporary positive price pressure. Usually, price rises caused by overreaction tend to correct over the following days after the surge in attention. However, we do not observe such price correction (as revealed in models (3) and (4)). In this regard, Schmeling, Schrimpf, and Todorov (2023) point out that limits-to-arbitrage are important in the cryptocurrency market and prevent some arbitrage strategies, which are commonly used for other asset classes, from being profitable for cryptocurrencies. The difficulty of short-selling cryptocurrencies could therefore explain the absence of reversal.[24]

As documented in Tables 1.2 and 1.7, each Twitter sample has some unique properties; Tweets are not equally salient between our samples, nor do they share similar textual content. Moreover, the samples also exhibit qualitative differences. For instance, *Official*-tweets are mainly constituted of announcements about their associated cryptocurrency and thus differ fundamentally from tweets written by the crowd. As a result, *Official*-tweets represent a more credible source of information than the other types of tweets for cryptocurrency investors. Therefore, in the following section, we analyze whether the three Twitter samples similarly predict returns.

[Insert Table 1.8 here]

1.5.2. Refinements of Twitter-based abnormal attention

As each category of tweets differs in terms of types of authors, functionalities, or reach, we expect that the relationship between expected returns and Twitter-based attention may change depending on the Twitter sample being used. Compared to general tweets, *Mention*-tweets have the particularity of triggering a notification for the recipient of the mention. *Ticker*-tweets, upon being clicked, display the most recent *Ticker*-tweets about the corresponding financial asset.

We now study if the qualitative and quantitative differences between our Twitter samples translate into different relationships in the cross-section of cryptocurrency expected returns. In Table 1.9, we estimate panel regressions of excess returns on the Twitter samples while controlling for our set of control variables. *Mention*-tweets are positively linked with excess returns in t and $t+1$ at the 1% and 10% significance levels, respectively. In contrast, we do not find any link between *Official*-tweets and returns. The regression coefficients of *Ticker*-tweets are positive and statistically significant at the 1% significance level in both models (1) and (2). The results of both panels are similar.

The results of Table 1.9 illustrate that the return predictability of Twitter activity is mainly derived from user-generated content and not from the announcements made by cryptocurrencies. This evidence provides additional support for

[24] A risk-based explanation of our results is also possible. Andrei and Hasler (2015) find that variation in investor attention affects the volatility through its impact on the speed of information incorporation into asset prices. Therefore, variation in attention is compensated with similar risk premia variation. We provide additional tests in the appendix showing that our results are best explained by behavioral reasons rather than risk-based mechanisms.

a behavioral-based explanation of our results, as we would expect *Official*-tweets to be associated with returns if our results were driven by a technological innovation channel (Cong, Li, and Wang, 2021). Lastly, we note that *Ticker*-tweets are more strongly related to excess returns than *Mention*-tweets. As the reason behind this difference in predictability power is not clear, we identify two non-exclusive candidate explanations. *Ticker*-tweets could predict returns better because they are more visible and therefore reach more users than *Mention*-tweets. A relatively larger number of users reached would correspond to larger (future) returns in both an overreaction channel (Barber and Odean, 2008; Da, Engelberg, and Gao, 2011) and a network growth channel (Sockin and Xiong, 2023; Cong, Li, and Wang, 2021). Alternatively, investors who use Twitter to inform their investment decisions could choose to primarily consume *Ticker*-tweets as their content is geared toward cryptocurrencies' financial characteristics. The idea that *Ticker*-tweets predict returns because they are more representative of investor attention would be more consistent with an overreaction channel. Our rationale is that the attention of all types of users should matter in a network growth channel and not just investors' attention.

[Insert Table 1.9 here]

1.5.3. Twitter textual content and investor attention

In this section, we examine why user-generated content-based attention is able to forecast returns in the cross-section of expected returns, whereas cryptocurrency-generated content does not show predictability. As both *Mention*-tweets and *Ticker*-tweets are able to predict returns despite their different visibility, it seems that investors are concerned about the tweets' content. To test this hypothesis, we now investigate the relationship between the lexicon score computed over the aggregated sample of tweets and daily excess cryptocurrency returns. We also control for textual sentiment to verify that abnormal attention doesn't predict returns because it originates from positive fundamental news. In addition, we also include an interaction term to test whether the return predictability is stronger when an attention-grabbing tweet is posted. We report the regression results in Table 1.10.

As displayed in Table 1.8, the regression coefficients of $Abn\,Attention_{i,t}^{All}$ are statistically significant in models (1) and (2) at the 1% significance level. $Sentiment_{i,t}^{All}$, $Lexicon_{i,t}^{All}$, and $Popular\,Tweet_{i,t}^{All}$ are also positively linked with excess returns at time t and $t+1$ at the 1% significance level. The interaction term between Twitter activity and popular tweets is positively connected with returns in (1) and (2) at the 1% and 5% significance levels, respectively.

We find $Sentiment_{i,t}^{All}$ to be positively associated with contemporaneous and next-day returns, which is in line with previous literature on Twitter sentiment (Gu and Kurov, 2020; Jiang et al., 2023). The effect of popular users posting on (future) returns is consistent with Benetton et al. (2024); Pedersen (2022), who document that influential users have a strong impact on the investment behavior of their followers and on asset returns. Compared to Benetton et al. (2024), which focus on a set of 75 celebrities, our results show that even small influencers can have a significant impact on asset prices. When we interact Twitter abnormal attention with $Popular\,Tweet_{i,t}^{All}$, the effect of Twitter activity on returns is even stronger. This finding also echoes with the results of Hillert, Jacobs, and Müller (2014), who find that media coverage can amplify behavioral biases. In the context of our study, any users with sufficient popularity on social media seem to be able to exacerbate behavioral biases through their influence on their followers. This latter observation provides more support for the overreaction interpretation of our results.

The regression coefficients of $Abn\,Attention_{i,t}^{All}$ are robust to the inclusion of the new independent variables. We interpret this result as indicative that the number of tweets posted is not subsumed by the textual content nor the sentiment of tweets. This provides support to our explanation that *Ticker*-tweets predict better returns than the other Twitter samples due to their better visibility. Given that lexicon score also forecasts returns, we conclude that both tweet volume and textual content matter in grabbing users' attention. In addition, the positive predictability of the lexicon on (future) returns casts doubts on the interpretation that abnormal attention predicts returns because of its relationship with future user base's growth (Cong, Li, and Wang, 2021; Sockin and Xiong, 2023). In fact, the lexicon loads negatively on tokens which we expect to be frequently used by newcomers, such as terms related to assistance or troubleshooting.

[Insert Table 1.10 here]

1.5.4. Technological innovations and investor attention

In this section, we test whether Twitter-based attention predicts expected returns through a technological innovation channel. Twitter content could potentially forecast future returns through its predictability of technological improvement in the blockchain. This channel is plausible as social media is frequently used by developers to exchange ideas about potential improvements or to signal cybersecurity breaches. In the literature, Cong, Li, and Wang (2021) theoretically show that cryptocurrencies valuations are influenced by technological improvements. Empirically, technological innovation is linked negatively to delisting probability (Liu, Sheng, and Wang, 2022) and positively to ICO success (Lyandres, Palazzo, and Rabetti, 2022).

To proxy for technological improvements, we utilize the log-difference of the number of commits published for each date and cryptocurrency on GitHub + 1. We believe that this proxy for technological improvements is appropriate because commits capture all code revisions made by developers. Therefore, any new feature or improvement made on the underlying technology used by a specific cryptocurrency will be reflected in its commit history.

Regression results are reported in Table 1.11. We find that contemporaneous and future cryptocurrency's technological development is strongly associated with *Official*-tweets. The contemporaneous relationship is positive and statistically significant at the 1% level. The regression coefficients of *Official*-tweets in the other models are all negative and statistically significant at the 1% or 10% level. We do not observe any links between logarithmic change in daily commit and the other Twitter samples. We interpret the results as indicative that *Ticker*-tweets and *Mention*-tweets do not predict returns through a technology innovation channel. In contrast, the significant association between the number of *Official*-tweets and technological innovation further confirms that tweets posted by cryptocurrencies can be interpreted as news.

[Insert Table 1.11 here]

1.6. Additional Empirical Tests

1.6.1. Robustness

To provide robustness of our results between abnormal attention and future cryptocurrency returns, we estimate panel regression models with slight modifications compared to Table 1.8. The robustness checks are reported in Table 1.12. In column one, we consider raw attention in the regression setup which is defined as:

$$Raw\,Attention_{i,t} = Log(1 + Number\,Of\,Tweets_{i,t}) \tag{1.3}$$

In the other columns, we use abnormal attention as our main independent variable as in Table 1.8. In models (2) and (3), we filter the sample to keep only cryptocurrencies classified as coins or as tokens, respectively, to verify that our results are not driven by characteristic differences between coins and tokens. We also check if our results are influenced by the rally of meme stocks led by R/WallStreetBets in 2021 which also affected some cryptocurrencies. For this purpose, we restrict our sample in models (4) and (5) to observations occurring before and after the first three months of 2021. In models (6) and (7), we explore if our results hold for both large and small assets by restricting our sample to assets with a market capitalization, respectively, above or below the median market capitalization. Finally, model (8) excludes pump and dump events to verify that our results are not driven by deliberate price manipulation schemes. We use data from Ardia and Bluteau (2024) to identify and filter out pump and dump events from our sample.[25] The pump and dump data lists both successful and unsuccessful events. By conservatism, we choose to remove the 3237 cryptocurrency-weeks in our sample concerned by such an event.

In all model specifications of Panel A, the regression coefficients of Twitter attention are significantly positive which is consistent with our main model specification presented in Table 1.8. When regressing expected returns at $t + 1$ in Panel B, we see that the regression coefficients are statistically significant for all specifications, except when small coins are removed from the sample. This observation is consistent with a limits-to-arbitrage explanation, as larger cryptocurrencies are easier to short-sell than smaller assets.

[Insert Table 1.12 here]

1.6.2. Additional tests and alternative explanation of the results

In this section, we conduct complementary tests to better characterize the empirical links between Twitter abnormal attention and cryptocurrency market variables.

To provide additional evidence that our results are indeed driven by an overreaction narrative, we investigate how Twitter-based attention predicts contemporaneous and future change in trading volume in a panel regression setting. Results are reported in Table 1.13. *Mention*-tweets and *Ticker*-tweets are both strongly linked with contemporaneous and future change in volume up to $t + 3$. Both types of tweets are positively linked with contemporaneous volume at the 1% significance level and negatively linked with future trading volume. The signs of the regression coefficients of *Ticker*-tweets and *Mention*-tweets make intuitive sense, as an overreaction channel is characterized by an increased buying pressure that decreases over the subsequent days.

[25] The data has been made publicly available by the authors at https://doi.org/10.5281/zenodo.12019080

[Insert Table 1.13 here]

To verify that the absence of reversal is plausibly due to limits-to-arbitrage (Schmeling, Schrimpf, and Todorov, 2023), we investigate the relationship between abnormal attention and cryptocurrency squared returns. Schmeling, Schrimpf, and Todorov (2023) argue that due to cryptocurrency futures exchanges rules about maximum losses on futures positions, even small price fluctuations can easily trigger the liquidation of an entire future position. This special set of rules impedes the ability of sophisticated investors to implement common arbitrage strategies, as highlighted by Schmeling, Schrimpf, and Todorov (2023). To align with our results, rises in abnormal attention need to predict increases in volatility to discourage sophisticated investors from arbitraging away the short term positive price pressure. This is indeed what we find in Table 1.14. Abnormal attention based on *Ticker*-tweets strongly predicts future squared returns up to $t + 4$ making it risky for arbitrageurs to bet on price decreases. Furthermore, price increases seem to be driven by behavioral factors.

[Insert Table 1.14 here]

Our results could also be explained by the model Andrei and Hasler (2015), who document similar relationships between attention, expected returns, and volatility as in this paper. The authors argue and show that increases in investor attention accelerate the incorporation of new information into asset prices, which leads to stronger price variations and risk premia. To shed light on whether our results indeed relate to Andrei and Hasler (2015), we investigate how abnormal attention relates to idiosyncratic squared returns. Idiosyncratic squared returns have the advantage of capturing large changes in returns that are not driven by variation in risk factors. The results are reported in Table 1.15.

We document a strong relationship between daily idiosyncratic squared returns and abnormal attention measured on *Mention*-tweets and *Ticker*-tweets. The regression coefficients are statistically significant at the 5% or 1% in all model specifications. As the dependent variable accounts for risk factor variations, Twitter attention's relationship with expected returns is not purely driven by variation in priced risk. This finding provides support for the overreaction channel to be the most likely mechanism behind our results.

[Insert Table 1.15 here]

1.7. Conclusion

Given the size of the cryptocurrency market and the importance of Twitter as a source of information, studying their interplay is essential for understanding the cross-section of cryptocurrency returns. In the literature, the evidence on the link between Twitter activity and cryptocurrency prices is mixed. Consistent with Benedetti and Kostovetsky (2021), we show that Twitter impacts the cross-section of cryptocurrency expected returns through an overreaction channel (Barber and Odean, 2008; Da, Engelberg, and Gao, 2011). However, we do not document any price reversals, which is plausibly due to limits-to-arbitrage, as already noted by (Schmeling, Schrimpf, and Todorov, 2023). We further find that the tweeting activity of popular users exacerbates behavioral biases and therefore the overreaction effect (Hillert, Jacobs, and Müller, 2014). We interpret this evidence as a warning sign about the ability of influential users to manipulate asset prices.

To better understand the relationship of Twitter attention with expected returns, we refine our abnormal attention by utilizing different samples of tweets. We show that the return predictability power of Twitter activity mainly arises from *Ticker*-tweets. In contrast, we find no association between tweets posted by official cryptocurrency channels and future returns, despite *Official*-tweets being able to forecast future innovations in the implementation code of each cryptocurrency (unlike the other Twitter samples). This empirical finding does not align with the predictions of the theoretical model of Cong, Li, and Wang (2021) that cryptocurrency valuations should be linked with technological improvements. Lastly, we document that the return predictability of *Ticker*-tweets is partly due to their salience and unique textual content, which caters to the preferences of retail investors. Overall, our results emphasize the heterogeneity of social media content, highlighting the need for researchers and practitioners to carefully consider which types of social media content best suit their needs.

References

Amihud, Y. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5:31–56.

Ammann, M., T. Burdorf, L. Liebi, and S. Stöckl. 2022. Survivorship and delisting bias in cryptocurrency markets. Unpublished Working Paper.

Andrei, D., and M. Hasler. 2015. Investor attention and stock market volatility. *Review of Financial Studies* 28:33–72.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61:259–99.

Ardia, D., and K. Bluteau. 2024. Twitter and cryptocurrency pump-and-dumps. *International Review of Financial Analysis* 95:103479–.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen. 2013. Value and momentum everywhere. *Journal of Finance* 68:929–85.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel. 2018. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy* 126:2224–76.

Bali, T. G., N. Cakici, and R. F. Whitelaw. 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99:427–46.

Barber, B. M., X. Huang, T. Odean, and C. Schwarz. 2022. Attention-induced trading and returns: Evidence from robinhood users. *Journal of Finance* 77:3141–90.

Barber, B. M., and T. Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21:785–818.

Bartov, E., L. Faurel, and P. S. Mohanram. 2018. Can twitter help predict firm-level earnings and stock returns? *Accounting Review* 93:25–57.

Benedetti, H., and L. Kostovetsky. 2021. Digital tulips? returns to investors in initial coin offerings. *Journal of Corporate Finance* 66:101786–.

Benetton, M., W. Mullins, M. Niessner, and J. Toczynski. 2024. Celebrity influencers: This is not financial advice. Unpublished Working Paper.

Borri, N., D. Massacci, M. Rubin, and D. Ruzzi. 2022. Crypto risk premia. Unpublished Working Paper.

Bose, D., H. Cordes, S. Nolte, J. C. Schneider, and C. F. Camerer. 2022. Decision weights for experimental asset prices based on visual salience. *Review of Financial Studies* 35:5094–126.

Broadstock, D. C., and D. Zhang. 2019. Social-media and intraday stock returns: The pricing power of sentiment. *Finance Research Letters* 30:116–23.

Chen, H., P. De, Y. J. Hu, and B.-H. Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies* 27:1367–403.

Chordia, T., A. Subrahmanyam, and V. R. Anshuman. 2001. Trading activity and expected stock returns. *Journal of financial Economics* 59:3–32.

Cong, L. W., Y. Li, and N. Wang. 2021. Tokenomics: Dynamic adoption and valuation. *Review of Financial Studies* 34:1105–55.

Cookson, J. A., J. E. Engelberg, and W. Mullins. 2023. Echo chambers. *Review of Financial Studies* 36:450–500.

Da, Z., J. Engelberg, and P. Gao. 2011. In search of attention. *Journal of Finance* 66:1461–99.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam. 1998. Investor psychology and security market under-and overreactions. *Journal of Finance* 53:1839–85.

Fang, H., and T.-Y. Lai. 1997. Co-kurtosis and capital asset pricing. *Financial Review* 32:293–307.

Fong, W. M., and B. Toh. 2014. Investor sentiment and the max effect. *Journal of Banking & Finance* 46:190–201.

Frydman, C., and B. Wang. 2020. The impact of salience on investor behavior: Evidence from a natural experiment. *Journal of Finance* 75:229–76.

Gu, C., and A. Kurov. 2020. Informational role of social media: Evidence from twitter sentiment. *Journal of Banking & Finance* 121:105969–.

Harvey, C. R., and A. Siddique. 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 55:1263–95.

Hillert, A., H. Jacobs, and S. Müller. 2014. Media makes momentum. *Review of Financial Studies* 27:3467–501.

Hong, H., and J. C. Stein. 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance* 54:2143–84.

Hosseini, A., G. Jostova, A. Philipov, and R. Savickas. 2020. The social media risk premium. Unpublished Working Paper.

Jiang, H., P. Liu, A. F. Roch, and X. Zhou. 2023. Social media sentiment and bitcoin price dynamics: Evidence from 36 million postings. Unpublished Working Paper.

Jondeau, E., and M. Rockinger. 2006. The copula-garch model of conditional dependencies: An international stock market application. *Journal of International Money and Finance* 25:827–53.

Kraus, A., and R. H. Litzenberger. 1976. Skewness preference and the valuation of risk assets. *Journal of Finance* 31:1085–100.

Kumar, A., S. Ruenzi, and M. Ungeheuer. 2021. Daily winners and losers. Unpublished Working Paper.

Liu, Y., J. Sheng, and W. Wang. 2022. Technology and cryptocurrency valuation: Evidence from machine learning. Unpublished Working Paper.

Liu, Y., and A. Tsyvinski. 2021. Risks and returns of cryptocurrency. *Review of Financial Studies* 34:2689–727.

Liu, Y., A. Tsyvinski, and X. Wu. 2022. Common risk factors in cryptocurrency. *Journal of Finance* 77:1133–77.

Loughran, T., and B. McDonald. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance* 66:35–65.

Lyandres, E., B. Palazzo, and D. Rabetti. 2022. Initial coin offering (ico) success and post-ico performance. *Management Science* 68:8658–79.

Lyócsa, Š., E. Baumöhl, and T. Vỳrost. 2022. Yolo trading: Riding with the herd during the gamestop episode. *Finance Research Letters* 46:102359–.

Pedersen, L. H. 2022. Game on: Social networks and markets. *Journal of Financial Economics* 146:1097–119.

Renault, T. 2017. Intraday online investor sentiment and return patterns in the us stock market. *Journal of Banking & Finance* 84:25–40.

Schmeling, M., A. Schrimpf, and K. Todorov. 2023. Crypto carry. Unpublished Working Paper.

Sockin, M., and W. Xiong. 2023. A model of cryptocurrencies. *Management Science* 69:6684–707.

Solomon, D. H., E. Soltes, and D. Sosyura. 2014. Winners in the spotlight: Media coverage of fund holdings as a driver of flows. *Journal of Financial Economics* 113:53–72.

Tan, H.-T., E. Y. Wang, and B. Zhou. 2015. How does readability influence investors' judgments? consistency of benchmark performance matters. *The Accounting Review* 90:371–93.

Figures

Figure 1.1: **Number of Cryptocurrencies, 2018-2022**

This figure shows the evolution of the number of cryptocurrencies, coins, and tokens over time at a monthly frequency that meet our criteria. The increase in the number of cryptocurrencies is mainly driven by an increase in tokens. The number of coins has even slightly decreased over the sample period.

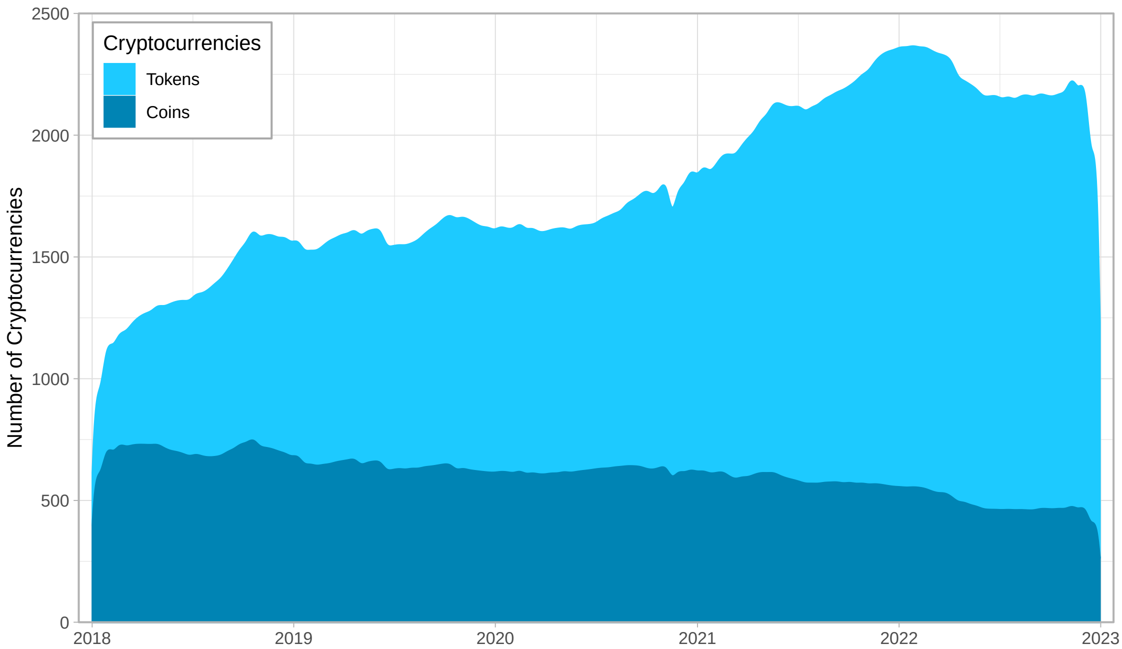
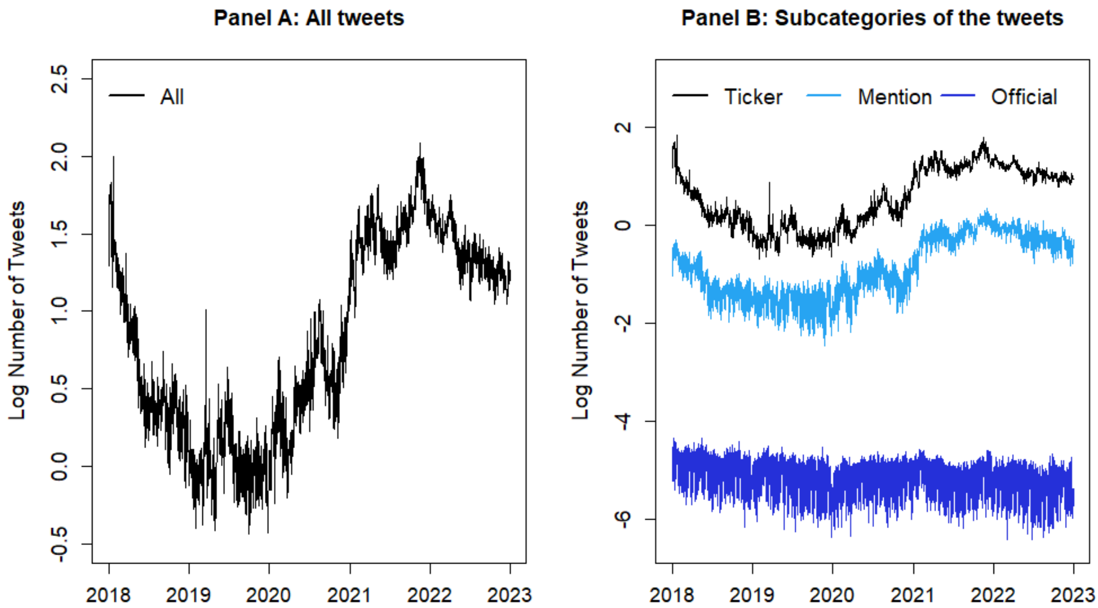


Figure 1.2: **Average Number of Tweets, 2018-2022**

This figure displays the evolution of the average number of tweets for cryptocurrencies over time. The average number of tweets is scaled by the number of cryptocurrencies in our sample at each point in time. For scaling purposes, the series is shown on a logarithmic scale. Panel A displays the evolution of the total number of tweets. Panel B shows the number of tweets separated across the different subcategories.



Tables

Table 1.1: **Description of Twitter Samples**

This table presents the different samples of tweets used in the paper along with their number of constituents. The column Description describes which types of tweets are contained in each sample. Note that the Twitter samples have some overlap and are not mutually exclusive, since users can tweet simultaneously about several cryptocurrencies and can mix the types of tweets.

Sample	Description	Number of Cryptos	Defunct Cryptos
Ticker	Any tweet that contains the ticker of the cryptocurrency.	165	44
Mention	Any tweet that tags (mentions) the official account of the crypto.	165	44
Official	Tweets posted by the official account.	154	35
All	The three samples above aggregated.	165	44

Table 1.2: **Selected N-grams from Lexicon**

This table contains selected n-grams from the lexicon along with their number of occurrences across the different Twitter samples.

N-gram count:	*Mention*	*Official*	*Ticker*	Coefficient
answering–questions	2	16	0	-1.00
full–stack	6	12	0	-1.00
sorry–inconvenience	5	35	1	-0.94
patch	9	32	1	-0.94
thank–patience	4	30	1	-0.93
assistance	16	99	4	-0.92
support–team	7	39	2	-0.90
pull–request	8	33	2	-0.89
token–sale	53	371	29	-0.86
opensource	49	122	13	-0.81
pleased–announce	22	42	5	-0.79
source–code	13	24	6	-0.60
vulnerabilities	6	11	3	-0.57
fully–decentralized	7	11	3	-0.57
withdrawing	4	11	3	-0.57
trading	522	669	1626	0.42
sold–all	11	0	28	0.44
take–profits	7	0	26	0.58
sell–orders	4	6	24	0.60
still–cheap	14	0	59	0.62
buy–hold	15	2	66	0.63
resistance	46	29	394	0.79
daily–chart	2	0	21	0.83
overbought	2	0	24	0.85
pattern	11	6	166	0.88
buys–numbertag	2	2	36	0.89
stop–loss	4	0	83	0.91
bull–flag	1	0	46	0.96
made–returntag	10	3	692	0.97
break–above	0	0	25	1.00

Table 1.3: **Variables Definition**

This table contains a description of the variables used in the paper. CMC stands for CoinMarket-Cap's website. "Author Homepage" indicates that the data described in the variable definition can be found on the website of the respective authors. KF is Kenneth French's website.

Variable	Definition	Source
Panel A: Cryptocurrency returns and characteristics		
$Excess\,Return_{i,t}$	Excess return on day t for cryptocurrency i.	CMC, KF
$Size_{i,t}$	Logarithmic market capitalization on day t for cryptocurrency i.	CMC
$\Delta Volume_{i,t}$	Logarithmic daily change in trading volume on day t for cryptocurrency i.	CMC
	The variables listed below until the start of Panel B are all computed over a rolling window of 60 days with a minimum of 30 days of non-missing observations.	
$Beta_{i,t}$	Regression coefficient of daily excess return of cryptocurrency i on the daily cryptocurrency market excess return.	CMC
$Momentum_{i,t}$	Compounded return of cryptocurrency i.	CMC
$Volatility_{i,t}$	Standard deviation of returns of cryptocurrency i.	CMC
$Idio\,Vola_{i,t}$	Standard deviation of the residuals when daily excess returns of cryptocurrency i are regressed on daily cryptocurrency market excess returns.	CMC
$Max\,Ret_{i,t}$	Average of the five highest daily excess return of cryptocurrency i.	CMC
$Volume\,Vola_{i,t}$	Standard deviation of Log transformed trading volume of cryptocurrency i.	CMC
$Illiquidity_{i,t}$	Ratio of illiquidity of cryptocurrency i, see Amihud (2002). $Amihud = \frac{1}{T}\Sigma_t^T \frac{\lvert r_{i,t}\rvert}{Volume_{i,t}}$	CMC
$Skewness_{i,t}$	Skewness of cryptocurrency i daily excess returns.	CMC
$Kurtosis_{i,t}$	Kurtosis of cryptocurrency i daily excess returns.	CMC
$Co-skewness_{i,t}$	The Co-Skewness of cryptocurrency i daily excess returns with daily cryptocurrency market excess return. $Coskew = \frac{E[(R_i-\mu_i)(R_m-\mu_m)^2]}{\sigma_{r_i}\sigma_{r_m}^2}$	CMC
$Co-kurtosis_{i,t}$	The Co-Kurtosis of cryptocurrency i daily excess returns with daily cryptocurrency market excess return. $Cokurt = \frac{E[(R_i-\mu_i)(R_m-\mu_m)^3]}{\sigma_{r_i}\sigma_{r_m}^3}$	CMC
$VaR_{i,t}$	The 5% percentile of daily cryptocurrency i excess returns.	CMC
Panel B: Attention and sentiment measures		
$Abn\,Attention_{i,t}^{All}$	Difference between the (log) number of *All*-tweets of cryptocurrency i at time t and the (log) mean number of *All*-tweets of cryptocurrency i during the previous 30 days. See equation 1.1. Abnormal attention for other Twitter samples is defined analogously.	Twitter
$\Delta Attention_{i,t}^{All}$	Difference between the (log) total number of *All*-tweets of cryptocurrency i at time t and the (log) number of *All*-tweets of cryptocurrency i at time $t-1$.	Twitter
$Sentiment_{i,t}^{All}$	Sentiment of the tweets published at date t on cryptocurrency i. We use the Renault (2017) lexicon to compute the sentiment of tweets. Sentiments of individual tweets are then averaged to get a sentiment score at a daily frequency. Sentiment is computed using the sample *All*-tweets.	Twitter, Author Homepage
$Popular\,Tweet_{i,t}^{All}$	Dummy variable equals one if at least one tweet from any sample published about cryptocurrency i at time t has an aggregated number of likes and replies equal to or higher than 100 across all Twitter samples. $Popular\,Tweet_{i,t}^{All}$ is computed using the sample *All*-tweets.	Twitter
$Lexicon_{i,t}^{All}$	Lexicon score of the tweets published at date t on cryptocurrency i. The lexicon is made to capture terms that are specific to *Ticker*-tweets. Lexicon score is computed using the sample *All*-tweets.	Twitter

Table continued on next page

Table continued

Panel C: Tweet characteristics		
$RT_{i,n,t}$	(Log) sum of the number of retweets and the number of quotes + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
$PM_{i,n,t}$	(Log) sum of the number of likes and the number of bookmarks + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
$Views_{i,n,t}$	(Log) number of times the tweet has been seen on Twitter + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
$LengthTweet_{i,n,t}$	(Log) length of the tweet + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
Panel D: Technology improvement measures		
$\Delta Commit_{i,t}$	(Log) daily change of the number of commits + 1 published on GitHub at time t for each repository of the organization developing the respective cryptocurrency i.	GitHub

Table 1.4: **Summary Statistics**

This table contains the summary statistics of the variables defined in Table 1.3. All variables are winsorized at the 1% level, except for $Excess\,Returns_t$ which is trimmed at the 99% level. The variables are expressed in decimal points. All variables are at a daily frequency.

	Mean	25%	Median	75%	StdDev
Panel A: Cryptocurrency returns and characteristics					
$Excess\,Return_{i,t}$	−0.0006	−0.038	−0.002	0.032	0.092
$Beta_{i,t}$	0.93	0.76	0.96	1.14	0.35
$Size_{i,t}$	17.43	15.75	17.12	18.84	2.53
$Momentum_{i,t}$	0.008	−0.422	−0.161	0.175	0.780
$Volatility_{i,t}$	0.080	0.052	0.069	0.096	0.041
$Idio\,Vola_{i,t}$	0.064	0.035	0.051	0.079	0.044
$Max\,Ret_{i,t}$	0.161	0.096	0.134	0.199	0.094
$\Delta Volume_{i,t}$	−0.005	−0.308	−0.023	0.255	0.731
$Volume\,Vola_{i,t}$	0.783	0.462	0.689	0.968	0.481
$Illiquidity_{i,t}$	0.002	0	0	0	0.021
$Skewness_{i,t}$	0.375	−0.216	0.280	0.872	1.025
$Kurtosis_{i,t}$	6.173	3.612	4.723	7.042	4.073
$Co\text{-}Skewness_{i,t}$	−0.345	−0.575	−0.279	−0.034	0.515
$Co\text{-}Kurtosis_{i,t}$	3.108	1.701	2.569	3.667	2.509
$VaR_{i,t}$	−0.113	−0.131	−0.099	−0.076	0.062
Panel B: Attention and sentiment measures					
$Abn\,Attention_{i,t}^{All}$	−0.210	−0.571	−0.179	0.176	0.675
$Abn\,Attention_{i,t}^{Mention}$	−0.315	−0.790	−0.210	0.111	0.834
$Abn\,Attention_{i,t}^{Official}$	−0.119	−0.383	−0.065	0.034	0.487
$Abn\,Attention_{i,t}^{Ticker}$	−0.201	−0.546	−0.162	0.166	0.665
$Sentiment_{i,t}^{All}$	0.041	0	0.044	0.084	0.062
$Popular\,Tweet_{i,t}^{All}$	0.222	0	0	0	0.416
$Lexicon_{i,t}^{All}$	0.097	−0.014	0.099	0.236	0.184
$\#\,Tweets_{i,t}^{All}$	812	16	60	233	5127
Panel C: Tweet characteristics					
$RT_{i,n,t}$	1.594	0	1.099	2.565	1.657
$PM_{i,n,t}$	1.004	0	0	1.609	1.380
$Views_{i,n,t}$	4.547	2.773	4.771	6.528	2.934
$Length\,Tweet_{i,n,t}$	2.613	2.197	2.639	3.044	0.553
Panel D: Technology improvement measures					
$\Delta Commit_{i,t}$	−0.0008	−0.29	0	0.24	0.90

Table 1.5: **Contemporaneous Determinants of Refined Twitter-based Attention**

The dependent variable is the change in attention using several Twitter samples. The regression spans the period from 2018 to 2022 for a sample of 154 cryptocurrencies. Control variables are defined in Panel A of Table 1.3. The regression coefficients are expressed in percentage points. Standard errors are clustered along time and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1)	(2)	(3)
	$\Delta Attention_t^{Mention}$	$\Delta Attention_t^{Official}$	$\Delta Attention_t^{Ticker}$
$\Delta Attention_{i,t}^{Mention}$		22.95^{***}	10.34^{***}
		(26.33)	(16.75)
$\Delta Attention_{i,t}^{Official}$	46.46^{***}		1.3^{**}
	(21.75)		(2.54)
$\Delta Attention_{i,t}^{Ticker}$	20.84^{***}	1.3^{**}	
	(15.33)	(2.53)	
$Excess\ Return_{i,t}$	24.39^{***}	-2.23	53.75^{***}
	(8.44)	(-1.55)	(10.93)
$Popular\ Tweet_{i,t}^{All}$	24.77^{***}	12.05^{***}	14.38^{***}
	(12.01)	(9.23)	(13.38)
$Beta_{i,t}$	-0.46	0.22	-0.44
	(-1.35)	(0.82)	(-1.07)
$Size_{i,t}$	-1.55^{***}	-0.72^{***}	-0.96^{***}
	(-5.5)	(-4.13)	(-4.87)
$Momentum_{i,t}$	-0.32^{**}	-0.34^{***}	-0.69^{***}
	(-2.12)	(-3.54)	(-3.43)
$Volatility_{i,t}$	-18.81	-24.29	14.55
	(-0.85)	(-1.55)	(0.6)
$Idio\ Vola_{i,t}$	10.07	25.32^{*}	3.93
	(0.59)	(1.88)	(0.2)
$Max\ Ret_{i,t}$	-2.81	1.01	-9.37^{**}
	(-0.77)	(0.51)	(-2.11)
$\Delta Volume_{i,t}$	4^{***}	-0.39^{*}	10.36^{***}
	(9.9)	(-1.95)	(12.03)
$Volume\ Vola_{i,t}$	-0.9^{***}	-0.37^{***}	-0.91^{***}
	(-3.33)	(-2.71)	(-3.93)
$Illiquidity_t$	-7.15	-5.65	-1.02
	(-0.8)	(-1.52)	(-0.14)
$Skewness_{i,t}$	0.21	0	0.24^{**}
	(1.26)	(0.04)	(2.12)
$Kurtosis_{i,t}$	0	-0.01	-0.06^{**}
	(0.15)	(-0.45)	(-2.16)
$Co-Skewness_{i,t}$	-0.79^{**}	-0.21	0.14
	(-2.52)	(-0.79)	(0.45)
$Co-Kurtosis_{i,t}$	0.2^{*}	0.05	0.25^{**}
	(1.84)	(0.6)	(2.36)
$VaR_{i,t}$	-7.16^{*}	2.34	2.56
	(-1.68)	(1.04)	(0.68)
Time FE	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes
Number of observations	246957	246957	246957

Table 1.6: **Predictive Determinants of Refined Twitter-based Attention**

The dependent variables are changes in attention using several Twitter samples. The regression spans the period from 2018 to 2022 for a sample of 154 cryptocurrencies. Control variables are defined in Panel A of Table 1.3. The regression coefficients are expressed in percentage points. Standard errors are clustered along time and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1)	(2)	(3)
	$\Delta Attention_{t+1}^{Mention}$	$\Delta Attention_{t+1}^{Official}$	$\Delta Attention_{t+1}^{Ticker}$
$\Delta Attention_{i,t}^{Mention}$	-36.6^{***}	-1.58^{***}	-0.08
	(-82.21)	(-9.04)	(-0.47)
$\Delta Attention_{i,t}^{Official}$	10.33^{***}	-41.78^{***}	1.31^{***}
	(15.02)	(-78.76)	(5.4)
$\Delta Attention_{i,t}^{Ticker}$	2.44^{***}	0.66^{***}	-32.52^{***}
	(6.22)	(2.83)	(-66.33)
$Excess\ Return_{i,t}$	19.56^{***}	2.81^{**}	25.93^{***}
	(6.91)	(2.19)	(8.06)
$Popular\ Tweet_{i,t}^{All}$	-7.38^{***}	-9.97^{***}	-2.99^{***}
	(-7.52)	(-11.26)	(-6.58)
$Beta_{i,t}$	-0.35	0.08	-0.66^{*}
	(-0.97)	(0.24)	(-1.76)
$Size_{i,t}$	-0.05	0.51^{***}	-0.36^{**}
	(-0.31)	(2.93)	(-2.27)
$Momentum_{i,t}$	-0.73^{***}	-0.2^{*}	-1.31^{***}
	(-4.13)	(-1.85)	(-5.95)
$Volatility_{i,t}$	44.97^{**}	18.2	33.2
	(2.06)	(0.99)	(1.6)
$Idio\,Vola_{i,t}$	-50.16^{***}	-15.43	-35.68^{**}
	(-2.99)	(-0.93)	(-2.09)
$Max\ Ret_{i,t}$	-2.96	0.73	-2.68
	(-0.64)	(0.34)	(-1.03)
$\Delta Volume_{i,t}$	1.9^{***}	0.48^{***}	1.57^{***}
	(6.95)	(2.7)	(7.24)
$Volume\,Vola_{i,t}$	0.09	0.18	-0.22
	(0.58)	(1.24)	(-0.96)
$Illiquidity_t$	7.73	7.37^{**}	1.84
	(1.51)	(2.01)	(0.32)
$Skewness_{i,t}$	0.11	-0.1	0.08
	(0.63)	(-1.09)	(0.7)
$Kurtosis_{i,t}$	-0.07^{***}	-0.02	-0.09^{***}
	(-2.62)	(-0.99)	(-4.12)
$Co-Skewness_{i,t}$	0.08	0.19	0.61^{*}
	(0.27)	(0.74)	(1.82)
$Co-Kurtosis_{i,t}$	0.1	-0.06	0.22^{**}
	(0.98)	(-0.68)	(2.01)
$VaR_{i,t}$	-1.06	3.07	-0.36
	(-0.24)	(1.36)	(-0.1)
Time FE	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes
Number of observations	246844	246844	246844

Table 1.7: **Visibility of Tweets**

The dependent variables are the number of retweets, the number of likes and replies, and the number of views for each tweet. The dependent variables are log-transformed. The sample covers the period from December 15, 2022, to December 31, 2022, for a sample of 151 cryptocurrencies. The regression coefficients are expressed in percentage points. Standard errors are clustered by authors. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1)	(2)	(3)
	$RT_{i,n,t}$	$PM_{i,n,t}$	$Views_{i,n,t}$
$Mention\,Tweet_{i,n,t}$	0.32	0.04	-34.12^{***}
	(0.08)	(0.01)	(-6.09)
$Official\,Tweet_{i,n,t}$	105.59^{***}	7.45	50.21^{***}
	(10.84)	(1.19)	(6.64)
$Ticker\,Tweet_{i,n,t}$	31.75^{***}	3.5	9.89^{**}
	(8.62)	(1.05)	(2.05)
$RT_{i,n,t}$		41.24^{***}	-65.53^{***}
		(63.18)	(-64.82)
$PM_{i,n,t}$	77.03^{***}		158.85^{***}
	(59.06)		(104.52)
$Views_{i,n,t}$	-29.47^{***}	38.25^{***}	
	(-50.86)	(118.99)	
$Length\,Tweet_{i,n,t}$	21.07^{***}	-4.64^{***}	25.87^{***}
	(15.64)	(-2.7)	(7.66)
$Sentiment_{i,n,t}$	20.41^{***}	16.04^{***}	-79.69^{***}
	(5.5)	(4.71)	(-11.84)
$Lexicon_{i,n,t}$	-13.72^{***}	-10.62^{***}	28.64^{***}
	(-8.08)	(-5.11)	(7.95)
Time FE	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes
Number of observations	63221	63221	63221

Table 1.8: **Investor Attention and Cryptocurrency Returns, 2018-2022**

The dependent variables are daily excess returns ($Excess\,Return_t$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 165 (103) cryptocurrencies. Control variables are defined in Panel A of Table 1.3. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1)	(2)	(3)	(4)
		$Daily\,Excess\,Return_i$		
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\ Attention^{All}_{i,t}$	1.06***	0.16***	0.01	0.02
	(18.57)	(4.29)	(0.37)	(0.89)
$Excess\ Return_{i,t}$		−18.94***	−2.65***	−0.19
		(−16.85)	(−5.26)	(−0.45)
$Excess\ Return_{i,t-1}$	−18.55***	−6.11***	−0.62	−0.46
	(−16.46)	(−9.14)	(−1.5)	(−0.89)
$Excess\ Return_{i,t-2}$	−5.28***	−1.93***	−0.79	−0.25
	(−7.66)	(−4.04)	(−1.39)	(−0.57)
$Excess\ Return_{i,t-3}$	−1.65***	−1.04**	−0.18	0.71
	(−3.77)	(−2.06)	(−0.42)	(1.45)
$Beta_{i,t}$	0.04	0.04	0.15	0.08
	(0.34)	(0.34)	(1.18)	(0.63)
$Size_{i,t}$	−0.34***	−0.3***	−0.23***	−0.21***
	(−10.51)	(−10.22)	(−8.99)	(−8.29)
$Momentum_{i,t}$	0	−0.06	−0.08	−0.13**
	(0.01)	(−0.84)	(−1.4)	(−2.11)
$Volatility_{i,t}$	0.42	0.14	−3.21	−0.87
	(0.04)	(0.01)	(−0.55)	(−0.13)
$Idio\,Vola_{i,t}$	−7.73	−7.35	−2.96	−4.65
	(−1.6)	(−1.58)	(−0.64)	(−1.03)
$Max\ Ret_{i,t}$	2.92	2.63	2.1	2.41
	(0.63)	(0.71)	(1.36)	(1.16)
$\Delta Volume_{i,t}$	1.81***	0.26***	−0.04	−0.01
	(12.43)	(6.7)	(−1.18)	(−0.32)
$Volume\,Vola_{i,t}$	−0.06	−0.08	−0.08	−0.05
	(−0.91)	(−1.13)	(−1.4)	(−0.78)
$Illiquidity_{i,t}$	−0.45	1.37	1.89	1.3
	(−0.31)	(0.9)	(1.53)	(1.15)
$Skewness_{i,t}$	0.02	0.01	0.02	0.02
	(0.22)	(0.09)	(0.42)	(0.47)
$Kurtosis_{i,t}$	−0.01	−0.01	−0.01	−0.01
	(−0.48)	(−0.63)	(−1.19)	(−1.36)
$Co\text{-}Skewness_{i,t}$	−0.03	−0.01	0.01	−0.04
	(−0.35)	(−0.14)	(0.14)	(−0.58)
$Co\text{-}Kurtosis_{i,t}$	0.02	0.02	0.01	0.02
	(0.63)	(0.63)	(0.54)	(0.74)
$VaR_{i,t}$	−2.24	−1.49	0.67	1.32
	(−1.25)	(−1.02)	(0.45)	(0.78)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	255684	254850	254431	254206
Panel B: Filtered ticker				
$Abn\ Attention^{All}_{i,t}$	1.08***	0.14***	0.03	0.04
	(15.58)	(2.96)	(1.13)	(1.21)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	160712	160246	159991	159846

Table 1.9: **Refined Investor Attention and Cryptocurrency Returns, 2018-2022**

The dependent variables are daily excess returns ($Excess\,Return_t$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table 1.3. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1)	(2)	(3)	(4)
	Daily Excess Return$_i$			
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\,Attention^{Mention}_{i,t}$	0.31***	0.04*	0.01	−0.02
	(10.85)	(1.92)	(0.32)	(−1.3)
$Abn\,Attention^{Official}_{i,t}$	−0.02	0.02	−0.02	0.04
	(−0.42)	(0.67)	(−0.61)	(1.22)
$Abn\,Attention^{Ticker}_{i,t}$	1.11***	0.17***	0	0.03
	(16.9)	(4.35)	(0.16)	(0.92)
$Excess\,Return_{i,t}$		−18.86***	−2.44***	−0.11
		(−16.09)	(−4.77)	(−0.27)
$Excess\,Return_{i,t-1}$	−18.82***	−5.94***	−0.55	−0.58
	(−16.22)	(−8.55)	(−1.3)	(−1.07)
$Excess\,Return_{i,t-2}$	−5.3***	−1.93***	−0.93	−0.24
	(−7.44)	(−3.95)	(−1.56)	(−0.52)
$Excess\,Return_{i,t-3}$	−1.72***	−1.13**	−0.12	0.55
	(−3.85)	(−2.15)	(−0.26)	(1.08)
$Beta_{i,t}$	0.07	0.05	0.16	0.09
	(0.55)	(0.42)	(1.21)	(0.67)
$Size_{i,t}$	−0.35***	−0.31***	−0.24***	−0.23***
	(−10.56)	(−10.04)	(−9.08)	(−8.29)
$Momentum_{i,t}$	0	−0.05	−0.08	−0.12**
	(−0.02)	(−0.76)	(−1.35)	(−2.05)
$Volatility_{i,t}$	−2.21	−0.57	−2.98	−0.91
	(−0.18)	(−0.06)	(−0.49)	(−0.14)
$Idio\,Vola_{i,t}$	−7.09	−7.17	−2.72	−4.2
	(−1.44)	(−1.51)	(−0.58)	(−0.91)
$Max\,Ret_{i,t}$	3.74	2.99	2.15	2.5
	(0.78)	(0.78)	(1.36)	(1.17)
$\Delta Volume_{i,t}$	1.8***	0.26***	−0.03	−0.01
	(11.85)	(6.69)	(−1.06)	(−0.4)
$Volume\,Vola_{i,t}$	−0.06	−0.07	−0.08	−0.05
	(−0.88)	(−0.94)	(−1.34)	(−0.83)
$Illiquidity_{i,t}$	−1.23	0.57	1.2	0.79
	(−0.68)	(0.32)	(0.76)	(0.58)
$Skewness_{i,t}$	0	−0.01	0.01	0.02
	(0)	(−0.09)	(0.28)	(0.4)
$Kurtosis_{i,t}$	0	−0.01	−0.01	−0.01
	(−0.3)	(−0.53)	(−1.24)	(−1.39)
$Co\text{-}Skewness_{i,t}$	0	0.01	0.02	−0.01
	(0)	(0.16)	(0.3)	(−0.18)
$Co\text{-}Kurtosis_{i,t}$	0.01	0.01	0.01	0.01
	(0.34)	(0.4)	(0.35)	(0.68)
$VaR_{i,t}$	−2.57	−1.53	0.94	1.71
	(−1.37)	(−1.02)	(0.62)	(1)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243633	243260	243061
Panel B: Filtered ticker				
$Abn\,Attention^{Mention}_{i,t}$	0.32***	0.02	0	−0.01
	(9.84)	(0.76)	(0.03)	(−0.24)
$Abn\,Attention^{Official}_{i,t}$	−0.01	0.04	−0.01	0
	(−0.22)	(0.93)	(−0.41)	(0.13)
$Abn\,Attention^{Ticker}_{i,t}$	1.12***	0.17***	0.03	0.04
	(14.27)	(3.56)	(1.06)	(1.06)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156682	156442	156304

Table 1.10: **Investor Attention, Sentiment and Cryptocurrency Returns, 2018-2022**

The dependent variables are daily excess returns ($Excess\,Return_t$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 165 (103) cryptocurrencies. Control variables are defined in Panel A of Table 1.3. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)
	$Daily\,Excess\,Return_i$			
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\,Attention^{All}_{i,t}$	0.82^{***}	0.11^{***}	0.01	0.04
	(17.02)	(2.92)	(0.52)	(1.42)
$Sentiment^{All}_{i,t}$	3.84^{***}	1.25^{***}	0.28	-0.56^{*}
	(9.69)	(4.07)	(0.91)	(−1.73)
$Abn\,Attention^{All}_{i,t} X$	1.39^{***}	0.19^{**}	−0.06	−0.05
$Popular\,Tweet^{All}_{i,t}$	(8.92)	(2.19)	(−0.86)	(−0.83)
$Popular\,Tweet^{All}_{i,t}$	0.32^{***}	0.13^{***}	0.02	−0.02
	(5.24)	(2.68)	(0.44)	(−0.57)
$Lexicon^{All}_{i,t}$	2.15^{***}	0.41^{***}	0.06	−0.13
	(13.92)	(3.27)	(0.56)	(−1.06)
$Excess\,Return_{i,t}$		-19.02^{***}	-2.65^{***}	−0.16
		(−16.96)	(−5.24)	(−0.39)
$Excess\,Return_{i,t-1}$	-18.89^{***}	-6.18^{***}	−0.62	−0.44
	(−17.08)	(−9.26)	(−1.48)	(−0.84)
$Excess\,Return_{i,t-2}$	-5.53^{***}	-1.98^{***}	−0.79	−0.24
	(−8.15)	(−4.14)	(−1.38)	(−0.54)
$Excess\,Return_{i,t-3}$	-1.8^{***}	-1.07^{**}	−0.18	0.72
	(−4.17)	(−2.11)	(−0.41)	(1.46)
$Beta_{i,t}$	0.02	0.04	0.15	0.08
	(0.14)	(0.32)	(1.18)	(0.63)
$Size_{i,t}$	-0.36^{***}	-0.31^{***}	-0.23^{***}	-0.21^{***}
	(−10.55)	(−10.43)	(−9.01)	(−8.04)
$Momentum_{i,t}$	−0.02	−0.06	−0.08	-0.12^{**}
	(−0.31)	(−0.93)	(−1.41)	(−2.07)
$Volatility_{i,t}$	1.51	0.24	−3.24	−0.92
	(0.13)	(0.02)	(−0.55)	(−0.14)
$Idio\,Vola_{i,t}$	-9.53^{**}	-7.66^{*}	−2.95	−4.52
	(−1.96)	(−1.65)	(−0.64)	(−1)
$Max\,Ret_{i,t}$	3.14	2.68	2.1	2.39
	(0.68)	(0.72)	(1.36)	(1.15)
$\Delta Volume_{i,t}$	1.78^{***}	0.25^{***}	−0.04	−0.01
	(12.44)	(6.62)	(−1.19)	(−0.29)
$Volume\,Vola_{i,t}$	−0.07	−0.08	−0.08	−0.05
	(−1.04)	(−1.17)	(−1.41)	(−0.78)
$Illiquidity_{i,t}$	−0.38	1.32	1.87	1.33
	(−0.27)	(0.87)	(1.52)	(1.17)
$Skewness_{i,t}$	0.02	0.01	0.02	0.02
	(0.2)	(0.09)	(0.42)	(0.47)
$Kurtosis_{i,t}$	−0.01	−0.01	−0.01	−0.01
	(−0.55)	(−0.64)	(−1.18)	(−1.36)
$Co\text{-}Skewness_{i,t}$	−0.02	−0.01	0.01	−0.05
	(−0.24)	(−0.13)	(0.15)	(−0.59)
$Co\text{-}Kurtosis_{i,t}$	0.02	0.02	0.01	0.02
	(0.65)	(0.67)	(0.56)	(0.72)
$VaR_{i,t}$	−2.42	−1.56	0.66	1.35
	(−1.35)	(−1.07)	(0.44)	(0.8)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	255684	254850	254431	254206
Panel B: Filtered ticker				
$Abn\,Attention^{All}_{i,t}$	0.82^{***}	0.08^{*}	0.04	0.06
	(14.21)	(1.76)	(1.21)	(1.63)
$Sentiment^{All}_{i,t}$	3.73^{***}	1.22^{***}	0.08	-0.72^{*}
	(7.95)	(3.17)	(0.22)	(−1.82)
$Abn\,Attention^{All}_{i,t} X$	1.54^{***}	0.24^{**}	−0.07	−0.04
$Popular\,Tweet^{All}_{i,t}$	(8.23)	(2.2)	(−0.82)	(−0.64)
$Popular\,Tweet^{All}_{i,t}$	0.32^{***}	0.14^{***}	0.04	−0.04
	(5.17)	(2.83)	(0.94)	(−0.61)
$Lexicon^{All}_{i,t}$	2.19^{***}	0.43^{***}	0.11	−0.09
	(11.26)	(2.59)	(0.87)	(−0.63)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	160712	160246	159991	159846

Table 1.11: **Refined Investor Attention and GitHub Commits**

The dependent variables are daily log-changes in the number of commits ($\Delta Commit_t$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 136 (86) cryptocurrencies. Control variables are defined in Panel A of Table 1.3. Standard errors are clustered along weeks and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)
	$\Delta Commit_i$			
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\ Attention_{i,t}^{Mention}$	−0.33	−0.41	−0.08	−0.12
	(−1.07)	(−1.04)	(−0.24)	(−0.71)
$Abn\ Attention_{i,t}^{Official}$	5.61***	−6.62***	−5.71***	−0.81*
	(5.37)	(−6.5)	(−5.28)	(−1.7)
$Abn\ Attention_{i,t}^{Ticker}$	0.06	−0.04	−0.49	0.13
	(0.16)	(−0.12)	(−1.43)	(0.59)
$Excess\ Return_{i,t}$	1.84	−0.6	0.93	1.95
	(0.8)	(−0.25)	(0.43)	(0.68)
$Excess\ Return_{i,t-1}$	−1.71	0.05	3.45	−3.33
	(−0.75)	(0.02)	(1.27)	(−1.59)
$Excess\ Return_{i,t-2}$	−1.26	3.69	−3.8*	−0.34
	(−0.58)	(1.36)	(−1.75)	(−0.15)
$Excess\ Return_{i,t-3}$	2.91	−3.3	−0.73	0.98
	(1.06)	(−1.55)	(−0.3)	(0.48)
$Beta_{i,t}$	0.17	0.13	−0.01	−0.19
	(0.27)	(0.2)	(−0.02)	(−0.29)
$Size_{i,t}$	0.07	−0.19	−0.14	−0.07
	(0.23)	(−0.59)	(−0.45)	(−0.23)
$Momentum_{i,t}$	−0.07	0.19	0.26	0.06
	(−0.2)	(0.71)	(0.9)	(0.23)
$Volatility_{i,t}$	1.83	−4.38	−1.2	11.64
	(0.06)	(−0.14)	(−0.04)	(0.35)
$Idio\,Vola_{i,t}$	8.54	3.06	−10.48	−17.62
	(0.29)	(0.11)	(−0.36)	(−0.59)
$Max\ Ret_{i,t}$	−3	−2.42	1.24	0.31
	(−1.26)	(−0.96)	(0.6)	(0.1)
$\Delta Volume_{i,t}$	0.34	0.31	−0.54*	0.16
	(0.98)	(0.95)	(−1.77)	(0.64)
$Volume\,Vola_{i,t}$	−0.21	−0.09	−0.15	−0.15
	(−1.09)	(−0.44)	(−0.73)	(−0.75)
$Illiquidity_{i,t}$	4.73	2.19	−2.11	−4.26
	(0.72)	(0.36)	(−0.32)	(−0.66)
$Skewness_{i,t}$	−0.04	−0.02	−0.09	−0.01
	(−0.27)	(−0.19)	(−0.65)	(−0.07)
$Kurtosis_{i,t}$	−0.02	0.01	0.02	0.02
	(−0.33)	(0.2)	(0.42)	(0.61)
$Co\text{-}Skewness_{i,t}$	0.34	0.56**	0.42	0.21
	(0.85)	(1.97)	(1.25)	(0.81)
$Co\text{-}Kurtosis_{i,t}$	−0.02	−0.04	−0.04	−0.04
	(−0.11)	(−0.23)	(−0.22)	(−0.23)
$VaR_{i,t}$	2.28	−2.68	−5.37	−4.34
	(0.55)	(−0.56)	(−1.25)	(−1)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	218873	218769	218668	218568
Panel B: Filtered ticker				
$Abn\ Attention_{i,t}^{Mention}$	−0.34	−0.93*	−0.21	−0.12
	(−1.02)	(−1.69)	(−0.47)	(−0.48)
$Abn\ Attention_{i,t}^{Official}$	6.56***	−7.47***	−6.78***	−1.14*
	(5.06)	(−5.9)	(−4.62)	(−1.79)
$Abn\ Attention_{i,t}^{Ticker}$	−0.13	0.14	−0.6*	0.19
	(−0.34)	(0.38)	(−1.67)	(0.63)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	139313	139244	139178	139113

Table 1.12: **Robustness Checks**

The dependent variables are daily excess returns ($Excess\, Return_t$). The regression spans the time period 2018 to 2022 for a sample of 165 cryptocurrencies. Control variables are defined in Panel A of Table 1.3. The first model (1) uses raw attention instead of abnormal change as the main independent variable. Models (2) and (3) restrict the sample to coins only or tokens only, respectively. Models (4) and (5) restrict the sample respectively to observations occurring before or after the GameStop short-squeeze. For models (6) and (7), we restrict the sample to assets with a market cap below or above the median, respectively. Finally, model (8) excludes pump and dump events from the sample. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Raw Attention	*Coins Only*	*Tokens Only*	*Before GameStop*	*After GameStop*	*Large Only*	*Small Only*	*No P&D Events*
Panel A: Contemporaneous relationship								
	$Daily\, Excess\, Return_{i,t}$							
$Abn\, Attention^{All}_{i,t}$	0.58***	1.02***	1.09***	1.11***	0.88***	1.49***	0.62***	1.05***
	(10.29)	(14.18)	(12.25)	(17.19)	(11.31)	(19.41)	(11.64)	(18.38)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	255684	147494	108190	165359	79856	129704	125160	252490
Panel B: Predictive relationship								
	$Daily\, Excess\, Return_{i,t+1}$							
$Abn\, Attention^{All}_{i,t}$	0.11***	0.15***	0.15***	0.17***	0.13**	0.03	0.13***	0.17***
	(3.68)	(3.19)	(2.58)	(3.84)	(2.41)	(0.68)	(2.89)	(4.42)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	254850	147023	107827	164984	79440	129570	124461	251659

Table 1.13: **Refined Investor Attention and Changes in Volume**

The dependent variables are daily logarithmic changes in trading volume. The regression spans the period 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table 1.3. Standard errors are clustered along weeks and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)
	Daily Δ *Volume*$_i$			
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\ Attention^{Mention}_{i,t}$	1.26***	−1.35***	−0.73***	−0.23
	(5.99)	(−7.23)	(−3.8)	(−1.52)
$Abn\ Attention^{Official}_{i,t}$	0.31	0.84***	−0.54*	0.47*
	(1.03)	(2.89)	(−1.75)	(1.73)
$Abn\ Attention^{Ticker}_{i,t}$	7.31***	−3.42***	−1.8***	−0.65***
	(13.03)	(−9.97)	(−6.96)	(−3.34)
$Excess\ Return_{i,t}$	169.03***	44.07***	−46.82***	−13.45***
	(10.53)	(9.95)	(−9.64)	(−4.44)
$Excess\ Return_{i,t-1}$	3.93	−57.62***	−26.09***	−5.53*
	(0.86)	(−9.25)	(−8.9)	(−1.67)
$Excess\ Return_{i,t-2}$	−60.01***	−46.92***	−12.48***	−7.42***
	(−9.76)	(−12.07)	(−3.32)	(−3)
$Excess\ Return_{i,t-3}$	−29.21***	−14.8***	−6.87***	−1.37
	(−9.31)	(−4.8)	(−2.67)	(−0.51)
$Beta_{i,t}$	−1.34***	−1.79***	−0.63	−0.87**
	(−2.94)	(−3.01)	(−1.34)	(−2.28)
$Size_{i,t}$	−0.2	−0.79***	−0.71***	−0.4***
	(−1.54)	(−5.28)	(−5.06)	(−3.3)
$Momentum_{i,t}$	−1.28***	−1.14***	−0.56***	−0.56***
	(−5.28)	(−4.26)	(−2.77)	(−3.11)
$Volatility_{i,t}$	44.57	75.76**	26.27	21.26
	(1.31)	(2.21)	(0.91)	(0.77)
$Idio\ Vola_{i,t}$	−82.84***	−113.26***	−47.35**	−39.21**
	(−4.09)	(−4.75)	(−2.39)	(−2.45)
$Max\ Ret_{i,t}$	12.13	7.97	1.78	1.53
	(1.04)	(0.83)	(0.28)	(0.21)
$\Delta Volume_{i,t}$		−34.06***	−8.12***	−1.92***
		(−67.28)	(−19.27)	(−4.22)
$Volume\ Vola_{i,t}$	0.89**	0.98**	0.49	0.49
	(2.36)	(2.27)	(1.49)	(1.55)
$Illiquidity_{i,t}$	35.42***	36.73***	20.1**	−0.17
	(3.47)	(2.82)	(2)	(−0.02)
$Skewness_{i,t}$	−0.09	0.1	0.14	0.2
	(−0.35)	(0.37)	(0.74)	(1.03)
$Kurtosis_{i,t}$	−0.07*	−0.14***	−0.08**	−0.07*
	(−1.69)	(−3.08)	(−2.25)	(−1.74)
$Co\text{-}Skewness_{i,t}$	0.29	0.48	0.4	−0.13
	(0.82)	(1.05)	(1.12)	(−0.4)
$Co\text{-}Kurtosis_{i,t}$	0.13	0.33**	0.22	0.17
	(1.08)	(2)	(1.53)	(1.53)
$VaR_{i,t}$	7.08	2.76	−3.95	−6.53
	(1.16)	(0.3)	(−0.66)	(−1.5)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243970	243571	243362
Panel B: Filtered ticker				
$Abn\ Attention^{Mention}_{i,t}$	1.16***	−1.62***	−0.55**	−0.34*
	(4.3)	(−7)	(−2.26)	(−1.72)
$Abn\ Attention^{Official}_{i,t}$	0.39	1***	−0.75**	0.81**
	(1.02)	(2.59)	(−2.08)	(2.32)
$Abn\ Attention^{Ticker}_{i,t}$	7.48***	−3.91***	−1.7***	−0.57**
	(10.75)	(−9.31)	(−5.32)	(−2.45)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156868	156616	156475

Table 1.14: **Refined Investor Attention and Squared Returns**

The dependent variables are daily squared excess returns. The regression spans the time period 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table 1.3. Standard errors are clustered along week and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)
	Daily Squared R_i^e			
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\ Attention_{i,t}^{Mention}$	0.06***	0.03***	0.02**	0.01
	(6.28)	(3.49)	(2.26)	(1.5)
$Abn\ Attention_{i,t}^{Official}$	−0.01	−0.02*	−0.01	0.02
	(−1.16)	(−1.73)	(−0.77)	(1.24)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.09***	0.08***	0.06***
	(8.35)	(5.68)	(4.8)	(3.44)
$Excess\ Return_{i,t}$	12.98***	0.28	0.58***	0.34
	(17.16)	(0.87)	(2.96)	(1.54)
$Excess\ Return_{i,t-1}$	2.84***	0.7***	0.51**	0.34
	(9.01)	(3.13)	(2.09)	(1.21)
$Excess\ Return_{i,t-2}$	1.6***	0.59**	0.36	0.21
	(7.67)	(2.5)	(1.37)	(1.44)
$Excess\ Return_{i,t-3}$	0.74***	0.18	0.04	0.22
	(4.05)	(0.9)	(0.29)	(1.23)
$Beta_{i,t}$	−0.21***	−0.23***	−0.23***	−0.21***
	(−2.91)	(−3)	(−2.8)	(−2.83)
$Size_{i,t}$	0.01	−0.05***	−0.05***	−0.06***
	(0.67)	(−3.87)	(−3.43)	(−4.04)
$Momentum_{i,t}$	0.07**	0.07**	0.02	0.01
	(2.33)	(2.29)	(0.65)	(0.29)
$Volatility_{i,t}$	16.71***	16.28***	11.03**	9.46***
	(4.62)	(4.86)	(2.5)	(2.66)
$Idio\,Vola_{i,t}$	9.66***	6.39***	5.26**	6***
	(4.18)	(2.94)	(2.37)	(2.96)
$Max\ Ret_{i,t}$	−6.31***	−5.08***	−2.68*	−2.56**
	(−4.65)	(−4.15)	(−1.73)	(−2.34)
$\Delta Volume_{i,t}$	0.24***	0.08***	0.02	−0.01
	(10.35)	(5.77)	(1.44)	(−0.83)
$Volume\,Vola_{i,t}$	0.02	0.04	0.04	0.05
	(0.6)	(0.9)	(0.93)	(1.14)
$Illiquidity_{i,t}$	1	0.04	0.75	0.63
	(0.63)	(0.03)	(0.49)	(0.48)
$Skewness_{i,t}$	0.04	0.05	0	0
	(1.26)	(1.59)	(0.11)	(−0.17)
$Kurtosis_{i,t}$	−0.02***	−0.03***	−0.02***	−0.02***
	(−4.08)	(−5.66)	(−3.63)	(−4.06)
$Co\text{-}Skewness_{i,t}$	0.09*	0.11**	0.09*	0.12**
	(1.85)	(2.35)	(1.92)	(2.52)
$Co\text{-}Kurtosis_{i,t}$	0.06***	0.07***	0.06***	0.06***
	(3.86)	(4.33)	(3.59)	(3.81)
$VaR_{i,t}$	−3.28***	−3.22***	−3.89***	−3.95***
	(−2.59)	(−2.72)	(−2.75)	(−3.06)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243633	243260	243061
Panel B: Filtered ticker				
$Abn\ Attention_{i,t}^{Mention}$	0.05***	0.02	0.02	0.01
	(4.23)	(1.51)	(1.41)	(0.88)
$Abn\ Attention_{i,t}^{Official}$	−0.01	−0.01	−0.01	0.02
	(−1.01)	(−0.58)	(−0.94)	(1.27)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.1***	0.09***	0.06***
	(7.73)	(5.14)	(4.94)	(3.32)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156682	156442	156304

Table 1.15: **Refined Investor Attention and Idiosyncratic Squared Returns**

The dependent variables are daily idiosyncratic squared excess returns. The regression spans the time period 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table 1.3. Standard errors are clustered along week and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)
	Daily Idiosyncratic $(R_i^e)^2$			
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\ Attention_{i,t}^{Mention}$	0.07***	0.03***	0.03***	0.02**
	(6.9)	(4.1)	(2.84)	(2.07)
$Abn\ Attention_{i,t}^{Official}$	−0.02	−0.02**	−0.01	0.02
	(−1.51)	(−1.97)	(−0.72)	(1.47)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.09***	0.08***	0.06***
	(8.32)	(5.57)	(4.85)	(3.55)
$Excess\ Return_{i,t}$	12.45***	0.41	0.62***	0.28
	(17.86)	(1.32)	(3.06)	(1.25)
$Excess\ Return_{i,t-1}$	2.88***	0.81***	0.49**	0.4
	(9.29)	(3.71)	(2.04)	(1.45)
$Excess\ Return_{i,t-2}$	1.65***	0.6***	0.42	0.27*
	(7.89)	(2.58)	(1.6)	(1.72)
$Excess\ Return_{i,t-3}$	0.73***	0.22	0.1	0.27
	(3.91)	(1.08)	(0.63)	(1.57)
$Beta_{i,t}$	−0.27***	−0.28***	−0.29***	−0.27***
	(−3.69)	(−3.74)	(−3.47)	(−3.55)
$Size_{i,t}$	−0.01	−0.07***	−0.07***	−0.08***
	(−0.7)	(−5.24)	(−4.62)	(−5.32)
$Momentum_{i,t}$	0.07**	0.06**	0.01	0
	(2.2)	(2.06)	(0.27)	(−0.1)
$Volatility_{i,t}$	14.65***	14.79***	10.41**	9.29***
	(4.08)	(4.54)	(2.46)	(2.7)
$Idio\ Vola_{i,t}$	11.16***	7.59***	6.42***	7.12***
	(4.9)	(3.46)	(2.88)	(3.39)
$Max\ Ret_{i,t}$	−6.06***	−4.9***	−2.75*	−2.78***
	(−4.56)	(−4.09)	(−1.82)	(−2.6)
$\Delta Volume_{i,t}$	0.23***	0.07***	0.02	−0.01
	(9.97)	(5.43)	(1.33)	(−0.61)
$Volume\ Vola_{i,t}$	0.02	0.04	0.05	0.05
	(0.69)	(1.04)	(1.09)	(1.27)
$Illiquidity_{i,t}$	1.03	−0.12	0.54	0.42
	(0.63)	(−0.07)	(0.36)	(0.32)
$Skewness_{i,t}$	0.04	0.04	0	0
	(1.11)	(1.44)	(0.03)	(−0.11)
$Kurtosis_{i,t}$	−0.02***	−0.03***	−0.02***	−0.02***
	(−4.06)	(−5.71)	(−3.71)	(−4.36)
$Co\text{-}Skewness_{i,t}$	0.11**	0.12***	0.11**	0.14***
	(2.39)	(2.75)	(2.28)	(2.82)
$Co\text{-}Kurtosis_{i,t}$	0.07***	0.08***	0.06***	0.06***
	(4.03)	(4.54)	(3.78)	(4.08)
$VaR_{i,t}$	−3.4***	−3.27***	−3.8***	−3.85***
	(−2.81)	(−2.88)	(−2.71)	(−2.94)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243633	243260	243060
Panel B: Filtered ticker				
$Abn\ Attention_{i,t}^{Mention}$	0.06***	0.02*	0.02*	0.01
	(4.89)	(1.92)	(1.65)	(1.27)
$Abn\ Attention_{i,t}^{Official}$	−0.02*	−0.01	−0.01	0.02
	(−1.72)	(−1.1)	(−0.78)	(1.21)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.1***	0.08***	0.07***
	(7.85)	(4.99)	(4.9)	(3.66)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156682	156442	156304

Chapter 2

Effects of Financial Advisors on Deal Outcomes: Evidence from SEC Filings

2.1. Introduction

What are the average effects of investment banks on merger and acquisition deal outcomes is a largely unexplored question. Due to limited data, existing research tends to examine the relationships of some specific financial advisor characteristics with deal premiums or announcement returns. Although this approach can identify which consultant characteristics matter, it provides limited insights about the effects of the average advisor on deals, which is an important question given the large risks and costs that M&A deals represent for the merging firms' stockholders.

I automate the collection of mandatory SEC filings on M&A deals to constitute a large database which allows me to quantify financial advisors efforts during private negotiations. Unlike other approaches relying on investment bank characteristics, this new variable allows me to investigate if firms do in average extract benefits from hiring financial advisors. I find that larger targets' (acquirers)' advisors' participation in the negotiation process translates into larger (lower) deal premiums. An one standard deviation increase in advisor involvement is related to a increase (decrease) of 1.76% (2.4%) in premiums, which is economically sizeable considering the large deal values paid in M&A transactions. As I do not observe any relationship with cumulative abnormal announcement return, I conclude that consultants improve deals mainly thanks to their negotiation expertise. Consistent with this interpretation, I find that advisors are generally more active in complex transactions and that their effects on premiums are larger when their clients possess a relatively larger bargaining power. Then, I investigate concerns that results could be driven by advisors' deal picking ability. My rationale is that parties possessing valuable information about target fundamentals will alter their negotiation strategy in consequence, which will be reflected in the textual sentiment of the negotiations.

Using FinBERT, a state-of-the-art model, to compute the textual sentiment of each document, I include sentiment among the control variables and rule out consultants' superior deal picking ability. Lastly, my results show that firms are more likely to re-employ advisors when they participate more in the negotiations. This indicates that firms are aware of the negotiation gains obtained by their investment banks. Thus, my findings suggest that advisor services are in average profitable for both targets and acquirers.

The literature on the effects of advisors on M&A deals finds empirical support for two conflicting hypotheses. The *Deal completion* hypothesis states that investment banks are incentivized to close as many deals as possible to maximize their ranking in the league table and their market share (Derrien and Dessaint, 2018). Under this hypothesis, the terms under which deals are closed are of reduced importance (Rau, 2000; McLaughlin, 1992; Nguyen and Tsai, 2024; Rajamani et al., 2017). In contrast, the *Superior deal* hypothesis posits that bank reputation is self-reinforcing and is used to charge higher fees. More prestigious banks are more skilled and are therefore able to negotiate better deals for their clients which in turn reinforces their position of leaders (Walter, Yawson, and Yeung, 2008; Hunter and Jagtiani, 2003; Ismail, 2010; Golubov, Petmezas, and Travlos, 2012). By using data on advisor participation in the private negotiations, I document that consultants' behavior is in average more consistent with the *Superior deal* hypothesis.

My results contribute to the literature studying the effects of financial advisors on deal outcomes in M&A transactions (McLaughlin, 1992; Golubov, Petmezas, and Travlos, 2012; Derrien and Dessaint, 2018; Bao and Edmans, 2009; Ismail, 2010). Past literature documents weak relationships between contract incentives of financial advisors and deal outcomes (McLaughlin, 1992). Golubov, Petmezas, and Travlos (2012) find a stronger relationship by showing that reputational costs mitigation is an important source of incentives for bulge bracket banks. My results differ by showing that advisor services benefit their clients in most cases; The benefits are not limited to a specific type of deal, investment bank category or side of the transaction.

By using the "Background of the Merger" sections from the mandatory SEC filings, I contribute to a new literature that exploits SEC filings about merger and acquisitions (Liu et al., 2022; Schubert, 2020; Eaton, Liu, and Officer, 2021; Masulis and Simsir, 2018; Edwards et al., 2023). For instance, Guo et al. (2020) look at how the risk factors disclosed by firms impact deal outcome. Heitzman and Klasa (2021) exploit private negotiation disclosures to study informed trading. I contribute by showing that the number of meetings where advisors are involved is a valid proxy of advisor efforts. I further document that this variable increases with deal complexity. Consistent with my findings, past literature documents that advisors tend to provide more value to their clients when information asymmetries are larger (Graham et al., 2017; Wang, Xie, and Zhang, 2022). Furthermore, the data collected for this paper is freely available to encourage further research using M&A negotiation data.[1]

Section 2.2 reviews the literature on M&A advisors. Section 2.3 describes the data used in the study. Section 2.4 examines the drivers of advisor efforts and their effects on merger outcomes. Section 2.5 presents robustness checks and section 2.6 concludes.

2.2. Literature Review

Advisors can undertake a wide range of critical tasks essential to the success of M&A transactions. The potential tasks include representing clients during negotiations, finding potential targets or acquirers, conducting the due diligence process, assessing the terms of the proposed merger and providing financing among others. Given the existence of a market for those services, advisor efforts should help forming better mergers. Yet, there is no consensus on the link between investment bank activities and deal outcomes in the literature. Two competing views posit how consultant efforts and deal outcome should relate. The *deal completion* hypothesis states that advisors maximize their revenues by closing as many deals as possible (McLaughlin, 1992; Rau, 2000; Derrien and Dessaint, 2018). This strategy is profitable, as closing deals is a sufficient condition for earning (partial) contingent fees (McLaughlin, 1992), and that investment bank market share is a predictor of future deal flow (Derrien and Dessaint, 2018). Under this hypothesis, the link between advisor efforts and negotiation outcomes should be negative, advisors' primary goal is to close the maximum number of deals as quickly as possible at the expense of deals' relevancy. Alternatively, the *superior deal* hypothesis posits that advisors reputation is an asset that investment banks use to attract more deals and to charge higher fees. Banks build their reputation by closing deals that provide value to their clients (Ismail, 2010; Golubov, Petmezas, and Travlos, 2012). Under this hypothesis, the relationship between advisor efforts and negotiation outcomes should be positive as banks benefit from the quality of the terms negotiated. Prior research documents the superior ability of some advisors to form better deals based on some advisor characteristics (Ismail, 2010; Golubov, Petmezas, and Travlos, 2012; Graham et al., 2017; Song, Wei, and Zhou, 2013). But, prior work is silent on what is the average

[1]The data is available at `https://doi.org/10.5281/zenodo.14988895`

effect of advisor efforts on deal outcomes. As highlighted by Welch et al. (2020), little is known about the factors influencing advisor behavior. The lack of evidence on the effects of advisors services is understandable given that most of advisor work happen in private meetings that are not easily observable by researchers.

2.3. Methodology and Data

2.3.1. Sample

Data on private negotiations among the merging firms are retrieved from the "Background of the Merger" sections of the S-4, SC 14D1, SC 14D9, PREM14A, DEFM14A, DEFM14C or PREM14C filings which are downloadable from the SEC through the EDGAR platform. The type of the form needed to be filed depends on the acquisition method. For instance, tender offers are reported using SC 14D1 and SC 14D9 forms. The filings can be found either on the EDGAR page of the acquirer or the target depending on the form. Although "Background of the Merger" sections differ in terms of comprehensiveness, firms must provide a minimum level of details which can be verified by the SEC.[2]

Stock market and accounting data are obtained from CRSP and COMPUSTAT. I obtain data on merger and acquisition from SDC Platinum. In this study, I only consider deals where both companies are US public firm. I further restrict my sample to deals announced between 1994 and 2022. 1994 corresponds to the year where filings first started to be numerically uploaded on EDGAR. In line with previous literature, I apply other filters to ensure that only deals where there is a transfer of control are kept. Namely, I exclude repurchases and keep deals only if the acquirer is purchasing more than 50% of the target shares. Table 2.1 summarizes the exact filters applied to the universe of M&A deals.

[Insert Table 2.1 here]

Importantly, M&A related filings are not available for all deals. Filings are only filed with the SEC if private negotiations results in a public tender offer or in a merger agreement signed by both firms. Furthermore, as described by Li, Liu, and Wu (2018) firms can manipulate the consideration structure paid

[2] Division of Corporation Finance, Office of Technology. (2021, September 9). *Re: Ikonics corp. amendment no. 2 to registration statement on Form S-4 filed August 11, 2021 file no. 333-258335.* U.S. Securities and Exchange Commission. `https://www.sec.gov/Archives/edgar/data/1083301/000000000021011003/filename1.pdf`

to target shareholders to avoid having to file any documents with the SEC. As a consequence, data collected from EDGAR does not cover most of the failed deals and misses some of the deals including acquirer stock as part of the deal structure. Fortunately, the number of missing files is low. Out of the 728 missing files missing in EDGAR, 367 are missing because the negotiations did not reach a definitive merger agreement. The majority of the remaining missing deals were announced in, or before, 2000. This is mainly driven by two reasons, which are unrelated with firms' decision to manipulate deal structure. As EDGAR was gradually implemented for more and more firms until May 1996, some fillings are simply not existing on EDGAR (Gao and Huang, 2020). In addition, COMPUSTAT does not provide linking tables between the CIK codes used by EDGAR and PERMNO identifiers prior to 2007. Therefore, I have to rely on fuzzy name matching for the earliest part of my sample, which results in a loss of observations. [3]

2.3.2. Methodology

To measure advisor activities, I use the number of meetings in which they are involved during the private negotiations. My rationale is that if an advisor provides services to its client that are material to the negotiation process, their advices should be disclosed in the SEC filings in order to comply with the SEC directives concerning the "Background of the Deal" sections:

> The disclosure should provide shareholders with an understanding of how, when, and why the material terms of your proposed transaction evolved and why this transaction is being recommended as opposed to any alternatives. (U.S. Securities and Exchange Commission, 2021, p. 5)[4]

In practice, firms indeed provide significant details on the negotiations leading to the proposed merger. The average number of words in the "Background of the Deal" sections is 2947. As an illustration of what the SEC filings look like, I report below an excerpt from the negotiations between Galileo International and Cendant:

[3] I use the python package "name_matching" developed by De Nederlandsche Bank to perform the fuzzy merger on company names. The package is available on PyPi and on GiHub at the following link: `https://github.com/DeNederlandscheBank`

[4] Division of Corporation Finance, Office of Technology. (2021, September 9). *Re: Ikonics corp. amendment no. 2 to registration statement on Form S-4 filed August 11, 2021 file no. 333-258335.* U.S. Securities and Exchange Commission. `https://www.sec.gov/Archives/edgar/data/1083301/000000000021011003/filename1.pdf`

> JPMorgan advised Cendant that the Galileo board would be unlikely to approve a transaction where the expected consideration to be received by Galileo stockholders was $32.50. In response, Cendant agreed to increase its offer to provide expected consideration to Galileo stockholders of $33 per Galileo share, but that such price was its final and best offer. (Cendant Corporation, 2001, p. 42)[5]

As shown in the example above, the report tends to truthfully represent the contribution of each party. The Galileo board accepts to increase the offered price after being suggested to do so by its advisors.

Advisor efforts are also sizeable across the full sample. In Figure 2.1, I display the average log-transformed number of meetings in which advisors are involved per year in blue. The number of meetings where targets' (acquirers') advisors are participating are represented using a solid (dotted) line. Both time-series are increasing through time. The rising number of meetings reflects the steady increase in the number of days between deal initiation and deal announcement since the 1980s (Liu, Mulherin, and Brown, 2022). Therefore, I consider the number of advisor meetings scaled by the total number of meetings in the reminder of the analysis. The scaled variable has the advantage to be relatively stable throughout my sample as illustrated by the two black lines. The next section describes in greater details the steps I take to process and clean SEC fillings in order to create the variables from the negotiation summaries.

[Insert Figure 2.1 here]

2.3.3. SEC Variables

Summarizing large textual reports in structured data is challenging. Fortunately, "Background of the Deal" sections follow some conventions which make the automatization of the data collection process possible. For instance, the majority of the SEC filings use simple and clear titles such as "Background of the Deal", "Background of the offer" or "Background of the merger". Abbreviations and the affiliation of the people involved in the negotiations are clearly indicated; Paragraphs are generally organized by meeting or by date. Therefore, processing the

[5]Cendant Corporation. (2001, July 6). *Form S-4 registration statement under the Securities Act of 1933.* U.S. Securities and Exchange Commission. `https://www.sec.gov/Archives/edgar/data/723612/0000950130-01-502903.txt`

reports can be completed by using regular expressions to extract "Background of the Deal" sections from every filing.

Firstly, I use the different titles within the document to retrieve the section containing the negotiation summary. Then, I take care to group paragraphs that should belong together; It is frequent for meeting descriptions to be separated by line or page breaks due to document formatting. Some documents also directly insert copies of letters that were exchanged between negotiators or press releases in the text. To avoid that those attachments get counted as multiple meetings, I group letters and press releases into one paragraph by searching for common letter and press release beginnings and endings. Some filings use bullet points to list the different topics that were addressed during some meetings; Such examples are also combined. In the following step, I search every date contained in the text to break the paragraphs into meetings. I take care to ignore dates that are used to express the expiration date of an offer or used to date some letters or meetings.[6] Having identified the different meetings, I then use the names of the companies and financial advisors reported by SDC platinum to identify which parties were involved in each meeting. Furthermore, I complete the list of advisor and firm names with acronyms and abbreviations introduced in the text.[7] As SDC does not always report the names of all advisors involved in deals, I also collect the names of the advisors hired during the private negotiations by spotting sentences about hiring financial advisors and by searching for advisor names. Moreover, investment banks are sometimes referred to using relational naming. For this reason, I retrieve textual patterns such as "target's financial advisor" among others. As a given firm can employ several advisors, I aggregate the number of meetings involving any financial advisors to measure advisor efforts.

To ensure validity of the information scrapped from the "Background of the Deal" sections, I conduct extensive automatic and manual checks throughout the data processing steps. I control that both target and acquiring firms can be found in the text. I verify that the "Background of the Deal" sections end with a conclusion to the negotiation and check that the reports contain several valid dates among

[6]As an illustration, the dates contained in "the May 3rd letter" or "the offer is valid until April 4, 2001" are ignored. But, dates such as "On March 23, 2009, an interested bidder approached our CEO." are processed.

[7]For instance, target companies are frequently referred to as "The Company" in the "Background of the Deal" sections. Aliases and abbreviations are generally introduced at the beginning of filings using formulations such as "Activision Blizzard, Inc., which we refer to as 'Activision Blizzard'" (Activision Blizzard Incorporated, 2022).
Activision Blizzard Incorporated. (2022, March 21). *Schedule 14A.* U.S. Securities and Exchange Commission. `https://www.sec.gov/Archives/edgar/data/718877/000110465922036155/tm225196-4_defm14a.htm`

others. Furthermore, some negotiation summaries start with an introduction describing the historical business relationships between the target and the acquirer. As those introductions do not concern the current negotiations directly, I drop them using several tests. For instance, I look for keywords suggesting that one of the firms has manifested its interest in pursuing a business combinations and keep meetings occurring after.[8] As some texts start with a description of prior talks that did not ultimately lead to a merger, I also verify that the number of days elapsed in between two consecutive meetings is below 365 days. Otherwise, I truncate the negotiation summary until the last large time gap present in the text.

I also use the SEC filings to retrieve known predictors of M&A outcomes (Masulis and Simsir, 2018; Schubert, 2020). Masulis and Simsir (2018) document that negotiated premiums tend to be lower when the deal is initiated by the target. The rationale is that targets are more likely to initiate deals when they possess bad private information. I control for deals initiated by target by examining for each deal which party made the first contact with potential acquirers or targets. In addition, I also quantify the level competition among bidders, as this variable is related to merger outcomes (Schubert, 2020; Dittmar, Li, and Nain, 2012). Schubert (2020) find that higher competition between bidders is linked with larger deal premiums and lower announcement returns for the acquirer consistent with a winner's curse. To control for this effect, I count the number of unique rival bidders, excluding acquirer, that appear throughout the private negotiations.[9] I use bidder count to define a dummy variable equal to one if there was at least one bidder besides the acquirer during the negotiations (Graham et al., 2017).

2.3.4. SDC Variables

Deciding on how to measure deal premium is not straightforward. There is a trade-off between choosing a date sufficiently early to capture most of the premium, but sufficiently late to reduce the level of noise in the estimated premium. Evidence from Sanders and Zdanowicz (1992) suggests that targets' share price start to move considerably already during private negotiations. Following Eaton, Liu, and Officer (2021), I compute premiums using the price paid by the acquiring firm and the

[8] Examples include for instance "merger", "sale", "offer", "acquisition", "acquire".

[9] To preserve anonymity of rival bidders, they are often referred to using generic names along with some numbering. Examples of anonymized names include "Company A", "Strategic Entity 2", or "Interested Buyer J" among others.

stock price of the target firm from 105 days before the announcement date.

$$Premium = \frac{Price\,Paid - Reference\,Price}{Reference\,Price} \tag{2.1}$$

In the equation above, I replace *Price Paid* by the final price paid reported by SDC if the initial offer price is missing. For robustness, deal premiums are also computed using reference prices from 63 days before the announcement date. Cumulative abnormal returns ($CAR_{t-1,t+1}$) are measured over the three days period surrounding the announcement similarly to Ahern (2012). I compute abnormal returns using the Fama-French three factors model estimated with a rolling window of 200 observations. I use the estimated regression coefficients from 105 business days before deal announcement to compute abnormal returns.

Variables measuring advisor skills are likely correlated with both premiums paid and advisor participation. Therefore, I include as control variables some advisor characteristics that are known to influence M&A outcomes. Namely, I follow Golubov, Petmezas, and Travlos (2012) and check if firms employ at least one bulge bracket firm as a financial advisor for the deal. Golubov, Petmezas, and Travlos (2012) show that firm are more likely to employ bulge bracket advisors when the deals are more complex and that the use of those top-tier advisors are positively related with cumulative announcement returns. I also measure investment bank skill by estimating the average CAR obtained in deals announced in the five previous years (Bao and Edmans, 2009). As some advisors can have different experience levels advising target and acquirer, I estimate an investment bank average CAR separately depending on the side of its client. I also control for factors influencing the choice to hire or not an advisor (Servaes and Zenner, 1996; Golubov, Petmezas, and Travlos, 2012) by defining *In-House M&A Experience* as the number of deals in which the firm did not employ any advisors over the last five years.

In addition, the literature has documented that investment banks can provide value to their client by reducing information asymmetry between negotiators (Chang et al., 2016; Graham et al., 2017; Wang, Xie, and Zhang, 2022). As targets possess better information about their own fundamental value than bidders, they have an advantage during the negotiations. The asymmetry increases the risk for bidders to overpay in M&A transactions. Given that bidders aim to decrease the acquisition price during negotiations, it is likely that the degree of advisor expertise in target's industry correlates with their negotiation involvement. Therefore, I follow Graham et al. (2017) and compute the difference between the relative importance of each industry for each advisor and the relative importance of each industry across all advisors. I compute this measure using the deals announced by the advisor over the five years before deal announcement and classify an advisor as a specialist in target (acquirer) industry if the variable is above 0. I also incorporate the variables developed by Chang et al. (2016) who find that acquirer (target) can also reduce information asymmetry by hiring the past advisors of target (acquirer). Given that my sample of deals is small after applying the filtering steps outlined in Table 2.1, I compute advisor specialization in a given industry and identify advisor-firm relationships using all deals involving at least one public firm. The other filters are left unchanged.

I also add variables that are likely to influence the incentives and the range of services offered by financial advisors. In this regard, I use dummy variables to control for the different mandates that a financial advisor might endorse. The four main tasks are providing advisory services, providing fairness opinion, representing stakeholders such as the board of directors or majority shareholders and arranging financing. Controlling for the different duties assumed by advisors ensures that the results are not driven by the underlying roles of advisors instead of financial advisor efforts.

Finally, I control for common predictors of deal premiums and or announcement returns used in the M&A literature. Consistent with Jennings and Mazzeo (1993); Betton, Eckbo, and Thorburn (2009), I control for toehold as this variable affects negotiation dynamics by acting as a tool to deter competition from other bidders. Similarly to Moeller, Schlingemann, and Stulz (2004), I use a dummy variable equals to one if the acquirer has a toehold greater or equal than 5% of the target stocks. Other controls include whether the deal is friendly or not. Schwert (2000) argue that hostile deals differ from friendly deals in the negotiation strategies being used. Bidder run-ups are added in the regression models, as they are correlated with bidder announcement returns (Rosen, 2006). I also consider acquirer's sigma which is linked with announcement returns and proxies for information asymmetry between target and acquirer (Moeller, Schlingemann, and Stulz, 2007). Lastly, I control for whether acquirer's stock are included as part of the transaction; Myers and Majluf (1984); Travlos (1987) document that the form of payments used reflects manager private signal about their firm valuation. Please refer to Table 2.2 for more details on variable construction.

[Insert Table 2.2 here]

2.3.5. Summary Statistics

Table 2.3 provides summary statistics for the main variables used in this paper after having merged the different datasets. Details about variable construction can be found in Table 2.2. All variables are winsorized to reduce the impact of extreme values on regression results. Except, premiums which are treated as missing values if they fall outside the interval $[-0.5, 2]$ following Song, Wei, and Zhou (2013).

The main variables of interests are reported in Panel A and Panel B. In Panel A, the average log-transformed number of meetings per deal in my sample is 3.55 which is equivalent to 35 meetings in absolute count. As reported in Panel B, target average participation rate in negotiation meetings is 87% where the same rate for acquirer is lower with a percentage of 70%. The average participation rates for targets' and acquirers' financial advisors follow a similar trend with average rates of 34% and 14% respectively. Consistent with intuition, deals with several bidders tend to have a higher average number of meetings. The difference in means is statistically significant at the 1% level.[10] Stronger competition among bidders translates into a larger number of meetings (Schubert, 2020).

The rest of the variables in Panel A are deal characteristics. In average, acquirer have a run-up of 12% with an average sigma of 2%. The average log-transformed acquirer's size and deal value are respectively 21.22 and 19.53 which leads to a mean relative size of 47%. The mean premiums in my sample are about 41% which in line with premiums documented by Golubov, Petmezas, and Travlos (2012). Consistent with the fact that negotiation need to reach an advanced stage before filing proxy statements, the average (median) completion probability is 93% (100%). Deals are in general filled and initiated by acquirer in approximately 65% and 60% of cases respectively. As an accuracy check of my automated data collection approach, this proportion of acquirer initiated deals is close to the one documented by Masulis and Simsir (2018). 98% of deals are friendly. In my sample, the proportion of tender offers is 14%. The average percentage of deals where the acquirer holds at least 5% of target shares on deal announcement's date is 2% (Moeller, Schlingemann, and Stulz, 2004) and most deals (62%) have at least one rival bidder to the acquirer.

Panel B displays variables that are measured either from target's or from acquirer's perspective. For instance, the mean Book-to-market ratio is 0.68 (0.58), the average Free Cash Flow is 0% (5%) and leverage is in average 21% (21%) approximately for target (acquirer). The average cumulative abnormal return around announcement date for the target (acquirer) is 10% (−1%) which is close

[10] $T\text{-}statistic = \frac{3.7741-3.1976}{\sqrt{0.0002+0.0002}} = 30.9783$

to the average past CAR of advisors of 10% (0%). Advisor can endows several non-exclusive mandates. 95% (91%) of target (acquirer) advisors have an advisory duty, 67% (33%) issue fairness opinion, 3% (1%) represent some specific party during the negotiations and 0% (4%) provide financing to the buyer. For both target and acquirer, the mean proportion of deals employing bulge bracket firms as advisors is between 34% and 42%, and the average fraction of advisors that specialized in either target or acquirer industry ranges from 54% to 61% (Graham et al., 2017).

[Insert Table 2.3 here]

2.4. Results

2.4.1. Predictors of Advisors Participation

If *Meetings Advisor* correctly reflects the involvement of investment bankers in negotiations, one would expect this variable to be partly determined by advisor mandates. I test this relationship using panel regressions where the dependent variable is the fraction of meetings in which financial advisors are involved and the independent variables are various deal and advisor characteristics. As the determinants of targets and acquirers advisors efforts may differ, I report the results for target and acquirer separately in Table 2.4. All models include fixed effects for the year and the target industry. Standard errors are clustered by acquirer, and t-statistics are reported in parentheses.

Consistent with intuition, I find that the inclusion of advisory, fairness opinion and representation duties are all positively related with the participation of either targets' advisors, acquirers' advisors or both. For instance, advisory duties are statistically significantly linked with targets' and acquirers' advisors participation at the 1% significance level. Representative duties predict higher acquirers' advisors efforts at the 1% significance level in models (4)-(6). Therefore, counselors are more likely to be active in the negotiation if it is planned in their contract. However, advisor participation is only weakly linked with advisor characteristics studied in the literature. Bulge bracket and large CAR advisors are in average not more active than the other advisors (Golubov, Petmezas, and Travlos, 2012; Bao and Edmans, 2009). In contrast, the link between advisor industry specialization and advisor efforts is significant for target (Graham et al., 2017). Deal characteristics constitute more important drivers. More complex deals with more lengthy negotiation processes tend to mobilize investment bankers more heavily, as indicated by the significant positive regression coefficients of friendly deal and the proportion of meetings involving target or acquirer. Interestingly, advisor current efforts are statistically significant and strongly related to their past average efforts. This suggests that this variable captures some degree of advisor specialization or advisor human capital (Chemmanur, Ertugrul, and Krishnan, 2019) that is persistent across deals.

Given that SEC filings are filed by one of the two firms, one may fear that advisor efforts get misrepresented by the filer. Misrepresentations could bias the measure of advisor participation if they correlate with deal characteristics. As SEC filings are filed in average 53 days after deal announcement, firms and investment banks have the opportunity to adapt the text based on investors reaction. Such feedback scenario would lead to a simultaneity bias in all models investigating merger outcomes and consultant efforts. I advance two reasons why this scenario is not likely. Firstly, the objective of the "Background of the Deal" sections is to provide investors a fair representation of the negotiation process which limits the extent of potential manipulations. Furthermore, the primary objective of filings is to inform and convince shareholders that the board of directors' recommendation is in stockholders' bests interests. Therefore, managers and directors have an incentive to over-disclose to show that they have respected their fiduciary duty. Similarly, consultants have limited incentives to exaggerate or undermine their contribution to the deal; The main determinant of an investment bank future deal flows is its league table ranking (Derrien and Dessaint, 2018). However, measurement errors are likely present in my sample. For instance, misrepresentations could occur if the filler does not observe or record the meetings arising in between the non-filler and its advisors. In Table 2.4, I observe that when documents are filed by target, involvement of target's (acquirer's) advisors is statistically significantly larger (lower) at the 1% level in all models. This evidence is consistent with the hypothesis that non-filer advisor efforts are not always captured in the filings. As a sanity check, I test if my latter findings still hold after restricting my sample to fillings that have been filled by targets or acquirers to verify the absence of problematic filer bias (not reported). I do not find any significant differences and conclude that filer effects on advisor participation are not problematic.

[Insert Table 2.4 here]

2.4.2. *Effects of Advisors Participation on Deal Outcomes*

Understanding the average effects of advisor efforts on deal outcomes is of key interest to researchers and firms alike. In Table 2.5, I report the results of panel regressions of deal premiums on financial advisor activity. For robustness, I consider two different premium variables using respectively target's stock price at $t - 105$, and $t - 63$ as reference prices. As in previous section, I study separately the effects of acquirer and target advisors on deal premiums. All models include target industry and year fixed effects. Standard errors are clustered by acquirer. The main independent variable of interest is $Meetings\,Advisor$. For better readability, this table doesn't provide the regression coefficients for the control variables. The interested reader can find them in Table 2.A1.

I find a significant positive relationship between deal premium and target advisor activity in models (1) to (3). However, the significance of the results is weaker when control variables are included in models (2) and (4). Alternatively, acquirer advisor efforts predict negatively premiums in models (5) to (8). Regression coefficients are statistically significant at the 1% significance level. Compared to models (1) and (5), the weaker predictability observed in models (3) and (7) could be caused by the shorter time period used to compute premiums. Eaton, Liu, and Officer (2021) highlight that premiums tend to be underestimated when prices from $t - 63$ are used to compute premiums. Taken together, the evidence supports the hypothesis that advisors services are beneficial to their clients during negotiations which is consistent with the *superior deal* hypothesis. Putted in context, a one standard deviation increase in $Meetings\,Advisor$ is associated with a 1.76% (2.4%) increase (decrease) in premium in model (2) ((6)). Given that the average premium in absolute terms is about \$866 millions, the estimated dollar gain amounts to approximately \$15.25 (\$20.80) millions for target (acquirer) which is significantly above the average total fees paid to target's (acquirer's) advisors of \$9.3 (\$6.6) millions as reported by SDC. However, the estimated dollar gains have to be taken with care, as I do not use any identification strategy to establish causality.

[Insert Table 2.5 here]

As the results of Table 2.5 are not sufficient to determine if premium gains arise from a better identification of good merger counter-parties which lead to larger synergy gains or from an improvement in the negotiated merger terms, I test how advisor efforts relates to shareholders gains. I use panel regressions of $CAR_{t-1,t+1}$

on the same set of regressors as Table 2.5. To investigate if advisor involvement relates to announcement returns beyond their effect on merger premium, I include deal premium among the control variable. The results are reported in Table 2.6.

I find weak predictability of advisor participation rates on target or acquirer shareholders gains. Target advisor efforts relationship with acquirer gains is positive, but is only statistically significant when controls are omitted. In a similar vein, acquirer counselor participation is negatively related with both target and acquirer gains at the 1% level. But, this finding is not robust to the inclusion of controls. Thus, increases in consultant efforts during negotiations seem to improve deals mainly through their impact on the final acquisition price. However, this evidence does not mean that advisors do not help their clients finding better merger candidates. As the "Background of the Deal" sections are written from the point of view of the firms involved, the exact process used by investment banks to select merger counterparties is generally not disclosed, which could explain why I do not find any relationship between shareholders gain and bank efforts. This lack of predictability also provides support for the absence of problematic filer bias in my sample. If investment banks were manipulating their citations throughout the filings conditional on deal outcomes, we would probably observe significant statistical relationships between announcement returns and advisor efforts, which is not the case.

[Insert Table 2.6 here]

2.4.3. Post-Merger Profitability

One reason for the absence of relationship between announcement returns and advisor efforts could be that investment banks are improving merger terms at the expense the strategic fit of the two companies. This hypothesis is plausible given that success fees are based on deal value which therefore create a conflict of interests between advisor's fiduciary duty toward their clients and their compensation structure (Rau, 2000; McLaughlin, 1992). Such mechanism is consistent with a *deal completion* hypothesis. I explore this hypothesis by exploring the link between measures of post-merger success and advisors efforts for deals that are successfully completed. Post-merger success is measured using the average free cash-flow over the five years following deal completion. The results of the panel regressions are reported in Table 2.7.

When control variables are omitted, target (acquirer) advisor efforts are positively (negatively) linked with future free cash-flows at the 5% (1%) significance level. However, the statistical significance disappears after including control variables. Therefore, the results suggest that investment banker improve deal terms without damaging the strategic fit of the merging companies.

[Insert Table 2.7 here]

2.4.4. Are Results Driven by Advisors' Deal Picking Ability

As shown in previous tables, investment banks seem to provide valuable services to their clients. However, one doubt remains considering the results' interpretation. It is not entirely clear whether consultants actively help their client obtaining more advantageous terms, or if investment bankers are just good deal pickers. In the previous tests, some of the included controls already account for the relative negotiation power of targets and acquirers. For instance, Masulis and Simsir (2018) find that targets that are initiating the discussions are more likely to be financially constrained. Furthermore, I also control for the presence of multiple bidders in the negotiations which has important effects on deal premiums (Schubert, 2020). However, those controls do not rule out the risk that financial advisor efforts is simply an advisor characteristics that correlates with advisor deal picking ability. This risk is not remote, as investment banks are specialized in advising mergers and acquisitions. For instance, Guo, Liu, and Tu (2023) show that advisors are skilled in selecting good comparable firms when preparing their valuations. The selected peers tend to perform well in the post-merger period. To investigate this point, I use sentiment analysis on every sentence of the negotiation summaries to compute a deal level sentiment score. My rationale is that if advisors are just picking clients which have an negotiation advantage over the other firm, the predictability of advisor participation rate should drop after accounting for negotiation sentiment. I argue that sentiment should be linked with relative differences in negotiation power, as it is in the firms best interests to reveal or to act based on information that could affect the merger valuation. Even if material information is not revealed per se in the text, the effects of its dissemination should be observable in the text. For instance, positive (negatives) news could encourage (deter) potential bidders to engage in discussions with target.

To compute sentence-level sentiment, I use the FinBERT model of Huang, Wang, and Yang (2023) which is a BERT model (Devlin et al., 2019) that is further trained on financial texts and fine-tuned on human labeled sentences from analyst reports in order to predict sentiment.[11] FinBERT outperforms other machine learning models and popular bag-of-words approaches (Loughran and McDonald, 2011) in quantifying sentiment (Huang, Wang, and Yang, 2023). One of the advantages of FinBERT is its ability to consider the context in which words appear to compute sentiment. This feature makes it especially suited to negotiation data where the use of negation and conditional statements is frequent. The model returns for each sentence a sentiment label which is either neutral, positive or negative along with a confidence score. In Table 2.8, I report nine sentences along with their predicted sentiment out of the approximately $433,072$ total sentences included in my sample. Overall, sentences classified as positive are often about potential synergy gains, positive news or increases in acquisition price. Negative sentences often capture negotiation stops, bad news or concerns about execution or financing risk for instance. In addition, FinBERT effectively captures textual nuances and handles negations appropriately. For example, it correctly identifies mentions of growth synergies as expressing positive sentiment, as shown in the first example of Table 2.8. Similarly, it accurately classifies statements concerning poor growth prospects as negative sentiment as shown in the fourth example.

[Insert Table 2.8 here]

Having collected sentiment for each sentence, I compute sentiment at the deal level using the following formula:

$$Sentiment = \frac{Positive - Negative}{Positive + Negative} \tag{2.2}$$

Where Positive and Negative are the number of positive and negative sentences in the "Background of the Deal" sections respectively. The sentiment score is in the range $[-1, 1]$. Having obtained a sentiment score for each deal, I re-estimate panel regressions of deal premiums on advisor participation rates including sentiment as a new control variable. The results are reported in Table 2.9.

I find similar results as in Table 2.5. Targets' advisor efforts are positively related with acquisition premium in models (1) to (3). Acquirers' advisor efforts are negatively related with the dependent variable in columns (5) to (8) at the 1% significance level respectively. Despite accounting for sentiment, the predictability of advisor participation is untouched which suggests that the results are not

[11] Their FinBERT model is available on huggingface at `https://huggingface.co/yiyanghkust/finbert-tone`.

driven by consultants' deal picking ability. Counselors seem to be able to secure better terms during the negotiations for their clients. Consistent with intuition, sentiment is significantly positively associated with deal premiums in all models but model (4). Thus, positive events occurring during the negotiations indeed have a stimulating effect on premiums paid by the winning bidder. Taken together the evidence suggests that advisor services are in average beneficial to their clients, which is in line with the *superior deal* hypothesis. The positive effects of advisors mainly materialize through better negotiated terms.

[Insert Table 2.9 here]

The negotiation channel can also be observed by examining the time-series of meetings for each deal. In Figure 2.2, I plot average advisor efforts and the average number of meetings where firms negotiate the acquisition price across time. As private negotiations differ in terms of duration, time is normalized. I identify negotiation meetings by searching for words mentioning the acquisition price such as "valuation", "premium", "price" among others in the meeting descriptions. We see that each time-series is increasing as time progresses. This similarity suggests that advisor efforts strongly relate to negotiation efforts within each deal.

[Insert Figure 2.2 here]

2.4.5. Advisors' Effects and Market Valuations

Advisors seem to help their clients. However, some important questions remain. When do advisor efforts matter the most? Are there instances where advisor gains become larger or lower? Intuitively, advisors are more likely to be able to negotiate a higher (lower) acquisition price when aggregate market valuation is high (low). Persuading the other party becomes easier if price fluctuations support your arguments to increase or decrease the acquisition price. Furthermore, as managers tend to use their firm's temporary overvaluation as a means of payment in mergers (Shleifer and Vishny, 2003), larger valuations lead to an increase in the number of potential merger counterparties, which directly enhances (worsens) the negotiation power of the target (acquiror). Consistent with the literature, I document a strong link between market valuations, as measured by the CAPE ratio (Jivraj and Shiller, 2017; Bunn and Shiller, 2014), and the number of deals announced in Figure 2.3.

[Insert Figure 2.3 here]

I empirically test the relationship between the CAPE ratio and the investment bank effects in Table 2.10. Specifically, I split my sample in a large CAPE and in a low CAPE subsample in Panel A and B, respectively. I consider the CAPE from 105 business days before deal announcement as a reference. Target advisors' involvement effects are significantly positive at the 5% or 1% level in Panel A, but become insignificant or weakly significant in Panel B. Similarly, the regression coefficients of acquirer bankers are only strongly significant in Panel A when control variables are not included. The results are, however, statistically significant at the 1% or 5% level in Panel B. Overall, the results confirm my hypothesis. Advisor effects become larger when market valuations align with their goal to negotiate a lower, or larger, acquisition price. Putted differently, consultants are more effective when their clients have more bargaining power. This evidence suggest that investment banks skillfully spot differences in negotiation power between the negotiating parties and act on this information to secure better terms. As most deals are initiated by acquirers (Masulis and Simsir, 2018), this finding explains why I document larger advisor effects for acquirer than for target in Table 2.5. This observation may also help explain why not all firms choose to engage financial advisors, even if their impact is generally positive for clients, since these benefits tend to be concentrated during periods of high (or low) market valuations.

[Insert Table 2.10 here]

2.4.6. Effects of Advisors on Negotiation

In this section, I investigate if consultants are associated with different negotiation dynamics. Negotiation literature suggests that parties with larger relative bargaining power tend to initiate negotiation more often (Kapoutsis, Volkema, and Nikolopoulos, 2013; Kapoutsis, Volkema, and Lampaki, 2017). Given the extensive expertise of M&A counselors in conducting negotiations compared to their client firms, negotiation dynamics are likely impacted by advisors' involvement. As advisor services become more effective when difference in negotiation power are larger and that the costs of reaching a negotiation impasse are lessened when firms can pursue several attractive alternatives, I posit that deals with larger advisor participation are associated with more aggressive negotiation tactics. I quantify the aggressiveness of the negotiations by computing the ratio of meeting initiated by each firm using the following ratio:

$$Proportion\ of\ initiated\ Meeting = \frac{\#Meetings\,initiated\,by\,Firm}{\#Meetings\,initiated\,by\,Target + \#Meetings\,initiated\,by\,Acquiror} \quad (2.3)$$

Where $\#Meetings\,initiated\,by\,Firm$ is replaced by acquirer or target depending on the regression model considered. Table 2.11 shows the result of panel regressions of the proportion of meetings initiated by each firm on advisors involvement. Both target and acquirer advisor efforts are statistically significant at the 1% across all model specifications. Increases in advisor participation translate into a larger proportion of meetings initiated by their client. Consistent with my hypothesis, advisors seem to adapt the negotiation strategy based on the relative bargaining power of their client.

[Insert Table 2.11 here]

2.4.7. Benefits for Advisors

As supported by my results, investment banks act in the best interests of their clients. Golubov, Petmezas, and Travlos (2012) argue that banks follow this strategy to maintain their reputation. Under the *superior deal* hypothesis, reputation is used to gain or to retain existing business relationships with firms. I test in Table 2.12, if advisors are rewarded for their efforts by their clients. Namely, I examine if firms decide to re-employ some of the advisors, they hired in their last deal, for their next future deal. As most target firms are disappearing from my sample following their acquisition, my sample focuses on the acquirer decisions to re-employ or not their previous advisors. However, I impose a minimum delay of 365 days in between the announcement dates of the two deals to ensure that the private negotiation periods do not overlap. My rationale is that firms are better able to assess the performance of their advisors after that the deals have been closed.

In Table 2.12, I document that the relationship between advisor participation and firm decisions to re-employ those advisors in the future is positive and statistically significant at the 10% level. Therefore, firms seem to recognize the value of the services delivered by its advisors, and incorporate this information in their hiring process for future deals. This evidence suggests that investment banks, in average, choose to work in their clients best interests. Doing so helps them to retain their existing clients, which is in line with a *superior deal* hypothesis.

[Insert Table 2.12 here]

2.5. Robustness Checks

In Table 2.13, I tweak my main variable of interest *Meeting Advisor* to see if my results are robust to slight changes in model specifications. Instead of calculating the proportion of meetings involving advisors, I use the log-transformed number of meetings involving advisors plus one without any scaling. As negotiation intensity and advisor implication have similar increasing pattern in Figure 2.2, it raises concerns that advisor efforts is just a proxy for the negotiation efforts of firms. For this reason, I do a second robustness check where I control for the number of meetings where acquirer or target firms are negotiating. I estimate this variable by aggregating the number of meetings where firms are appearing and where the text mentions the acquisition price. This approach is similar to the one used in Figure 2.2. Lastly, I include an additional test where I control for the log-transformation of the total fees earned by advisors. Fees are the main tool that firms can use to align banks' objectives with their goals. Unfortunately, the fees reported by SDC are missing for most deals Derrien and Dessaint (2018). This issue becomes problematic when other control variables are added to the regression models. For this reason, I test here, with a reduced set of controls, if fees cause advisor efforts predictability to vanish. The results of the robustness checks are reported in Table 2.13.

Overall, the results mirror those reported in Table 2.5. In Panel A, the relationship between unscaled target consultant efforts and deal premium becomes slightly weaker for target, but remains strongly significant for acquirer. In Panel B, controlling for firm negotiation efforts does not affect the regression coefficients of the main independent variable of interest. Therefore, advisor participation is unlikely to be a sideshow of the work undertook by their client. In Panel C, including fees as a control variable results in a large loss of observations. However, the link between advisor efforts and premiums remains statistically significant across most models. Therefore, omitting advisor fees from the other tables do not threaten the interpretation of the main results.

[Insert Table 2.13 here]

2.6. Conclusion

What is the average effect of services provided by financial advisors on deal outcomes is not a settled question in the literature. With this paper, I answer this question by using textual analysis tools on negotiation data. I find that larger advisor participation during the private negotiations is linked with improved deal premiums for both target and acquirer, which is consistent with a *superior deal* hypothesis. Clients recognize the bank efforts by retaining active advisors to assist them in future deals.

I then explore how exactly the consultants impact the deals. By controlling for the textual sentiment of the negotiation reports, I document that advisors seem to actually improve the negotiated price rather than identifying deals with already good proposed terms. Thanks to advisors' advices, acquirer is able to purchase target at a smaller price and target is able to obtain a higher selling price. In addition, the negotiation gains obtained by investment banks tend to be larger when differences in bargaining power are larger. Lastly, I find that counselors affect the negotiation dynamics between the target and the acquirer. When their advisor involvement is larger, firms are more likely to lead the talks by initiating meetings. Overall, my results contribute to a nascent branch of research that use SEC filings and textual analysis tools to answer new questions about M&A deals.

References

Ahern, K. R. 2012. Bargaining power and industry dependence in mergers. *Journal of Financial Economics* 103:530–50.

Bao, J., and A. Edmans. 2009. Do investment banks have skill?: Performance persistence of m & a advisors. Unpublished Working Paper.

Betton, S., B. E. Eckbo, and K. S. Thorburn. 2009. Merger negotiations and the toehold puzzle. *Journal of Financial Economics* 91:158–78.

Bunn, O. D., and R. J. Shiller. 2014. Changing times, changing values: A historical analysis of sectors within the us stock market 1872-2013. Working Paper, National Bureau of Economic Research.

Chang, X., C. Shekhar, L. H. Tam, and J. Yao. 2016. The information role of advisors in mergers and acquisitions: Evidence from acquirers hiring targets' ex-advisors. *Journal of Banking & Finance* 70:247–64.

Chemmanur, T. J., M. Ertugrul, and K. Krishnan. 2019. Is it the investment bank or the investment banker? a study of the role of investment banker human capital in acquisitions. *Journal of Financial and Quantitative Analysis* 54:587–627.

Derrien, F., and O. Dessaint. 2018. The effects of investment bank rankings: Evidence from m&a league tables. *Review of Finance* 22:1375–411.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. Unpublished Working Paper.

Dittmar, A., D. Li, and A. Nain. 2012. It pays to follow the leader: Acquiring targets picked by private equity. *Journal of Financial and Quantitative Analysis* 47:901–31.

Eaton, G. W., T. Liu, and M. S. Officer. 2021. Rethinking measures of mergers & acquisitions deal premiums. *Journal of Financial and Quantitative Analysis* 56:1097–126.

Edwards, A., S. Heitzman, S. Klasa, and M. Todtenhaupt. 2023. Private merger negotiation responses to ongoing tax legislation. Unpublished Working Paper.

Gao, M., and J. Huang. 2020. Informing the market: The effect of modern information technologies on information production. *Review of Financial Studies* 33:1367–411.

Golubov, A., D. Petmezas, and N. G. Travlos. 2012. When it pays to pay your investment banker: New evidence on the role of financial advisors in m&as. *Journal of Finance* 67:271–311.

Graham, M., T. S. Walter, A. Yawson, and H. Zhang. 2017. The value-added role of industry specialist advisors in m&as. *Journal of Banking & Finance* 81:81–104.

Guo, F., T. Liu, T. Shu, and X. Yan. 2020. Risks in mergers and acquisitions. Unpublished Working Paper.

Guo, F., T. Liu, and D. Tu. 2023. Neglected peers in merger valuations. *Review of Financial Studies* 36:3257–310.

Heitzman, S., and S. Klasa. 2021. Informed trading reactions to new private information: Evidence from nonpublic merger negotiations. *Management Science* 67:2630–56.

Huang, A. H., H. Wang, and Y. Yang. 2023. Finbert: A large language model for extracting information from financial text. *Contemporary Accounting Research* 40:806–41.

Hunter, W. C., and J. Jagtiani. 2003. An analysis of advisor choice, fees, and effort in mergers and acquisitions. *Review of Financial Economics* 12:65–81.

Ismail, A. 2010. Are good financial advisors really good? the performance of investment banks in the m&a market. *Review of Quantitative Finance and Accounting* 35:411–29.

Jennings, R. H., and M. A. Mazzeo. 1993. Competing bids, target management resistance, and the structure of takeover bids. *Review of Financial Studies* 6:883–909.

Jivraj, F., and R. J. Shiller. 2017. The many colours of cape. Unpublished Working Paper.

Kapoutsis, I., R. Volkema, and A. Lampaki. 2017. Mind the first step: The intrapersonal effects of affect on the decision to initiate negotiations under bargaining power asymmetry. *Frontiers in Psychology* 8:1313–.

Kapoutsis, I., R. J. Volkema, and A. G. Nikolopoulos. 2013. Initiating negotiations: the role of machiavellianism, risk propensity, and bargaining power. *Group Decision and Negotiation* 22:1081–101.

Li, K., T. Liu, and J. Wu. 2018. Vote avoidance and shareholder voting in mergers and acquisitions. *Review of Financial Studies* 31:3176–211.

Liu, T., J. H. Mulherin, and W. O. Brown. 2022. The development of the takeover auction process: The evolution of property rights in the modern wild west. Unpublished Working Paper.

Liu, T., T. Shu, E. Towery, and J. Wang. 2022. The role of external regulators in mergers and acquisitions: Evidence from sec comment letters. *Review of Accounting Studies* 1–42.

Loughran, T., and B. McDonald. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance* 66:35–65.

Masulis, R. W., and S. A. Simsir. 2018. Deal initiation in mergers and acquisitions. *Journal of Financial and Quantitative Analysis* 53:2389–430.

McLaughlin, R. M. 1992. Does the form of compensation matter?: Investment banker fee contracts in tender offers. *Journal of Financial Economics* 32:223–60.

Moeller, S. B., F. P. Schlingemann, and R. M. Stulz. 2004. Firm size and the gains from acquisitions. *Journal of Financial Economics* 73:201–28.

———. 2007. How do diversity of opinion and information asymmetry affect acquirer returns? *Review of Financial Studies* 20:2047–78.

Myers, S. C., and N. S. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13:187–221.

Nguyen, H. T. D., and H. Tsai. 2024. Relationship advisors guiding financially weak targets in m&a deals: Conflict of interest and reputation concerns. *Finance Research Letters* 61:105013–.

Rajamani, A., M. Van der Poel, A. De Jong, and S. Ongena. 2017. The international diversification of banks and the value of their cross-border m&a advice. *Management Science* 63:2211–32.

Rau, P. R. 2000. Investment bank market share, contingent fee payments, and the performance of acquiring firms. *Journal of Financial Economics* 56:293–324.

Rosen, R. J. 2006. Merger momentum and investor sentiment: The stock market reaction to merger announcements. *Journal of Business* 79:987–1017.

Sanders, R. W., and J. S. Zdanowicz. 1992. Target firm abnormal returns and trading volume around the initiation of change in control transactions. *Journal of Financial and Quantitative Analysis* 27:109–29.

Schubert, R. 2020. Pre-public m&a negotiations: How does bidding competition affect deal outcomes? Unpublished Working Paper.

Schwert, G. W. 2000. Hostility in takeovers: in the eyes of the beholder? *Journal of Finance* 55:2599–640.

Servaes, H., and M. Zenner. 1996. The role of investment banks in acquisitions. *Review of Financial Studies* 9:787–815.

Shleifer, A., and R. W. Vishny. 2003. Stock market driven acquisitions. *Journal of Financial Economics* 70:295–311.

Song, W., J. D. Wei, and L. Zhou. 2013. The value of "boutique" financial advisors in mergers and acquisitions. *Journal of Corporate Finance* 20:94–114.

Travlos, N. G. 1987. Corporate takeover bids, methods of payment, and bidding firms' stock returns. *Journal of Finance* 42:943–63.

Walter, T. S., A. Yawson, and C. P. Yeung. 2008. The role of investment banks in m&a transactions: Fees and services. *Pacific-Basin Finance Journal* 16:341–69.

Wang, C., F. Xie, and K. Zhang. 2022. Expert advice: Industry expertise of m&a advisors and acquirer shareholder returns. *Journal of Financial and Quantitative Analysis* 57:599–628.

Welch, X., S. Pavićević, T. Keil, and T. Laamanen. 2020. The pre-deal phase of mergers and acquisitions: A review and research agenda. *Journal of Management* 46:843–78.

Figures

Figure 2.1: **Evolution Through Time of the Number of Meeting with Advisors**

This figure displays the average number of meetings per deal announcement year. The blue lines represents the average log-transformed number of meetings involving advisors trough time. The black lines represents the the same number of meetings but scaled by the total number of meetings. Dotted lines represent acquirers' advisors and full lines represent targets' advisors.

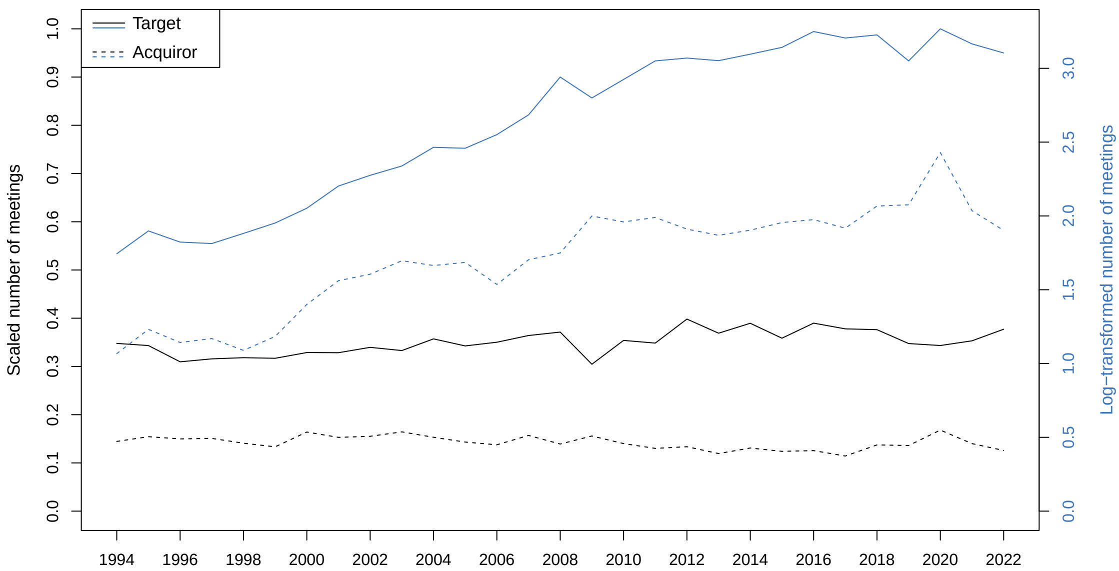


Figure 2.2: **Evolution Through Time of Advisor Involvement and Negotiation Intensity**

In the first two panels, this figure displays the average involvement of advisor through the normalized span of the private negotiations. The last panel shows the average frequency of meetings where terms related to the acquisition price are mentioned.

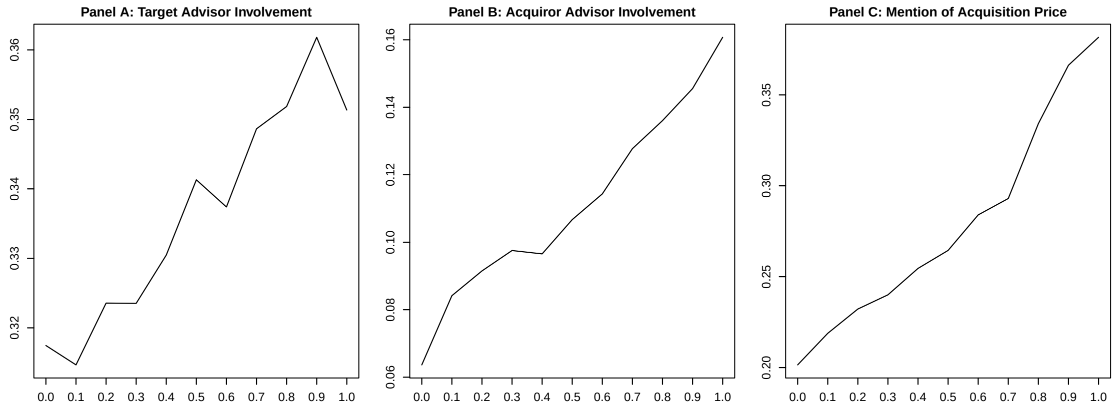
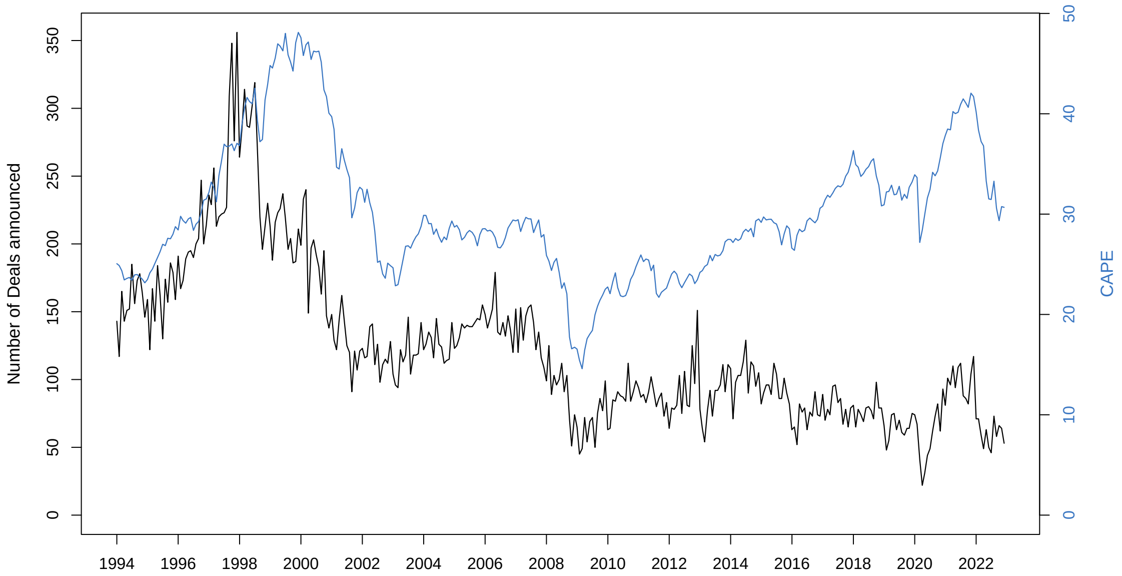


Figure 2.3: **Evolution Through Time of the Number of Deals and CAPE**

This figure displays the average number of deals per month in black alongside the total return CAPE ratio (Bunn and Shiller, 2014; Jivraj and Shiller, 2017) in blue. The CAPE ratio is obtained through Robert Shiller's website using the following link `https://shillerdata.com/`.



Tables

Table 2.1: **M&A Deal screening**

This table describes the different filters applied to the universe of M&A deals contained in SDC to obtain the final sample of M&A deals used in this study.

Filters Applied	# of Deals remaining
M&A deals announced between 1994 and 2022 among US firms	290,279
Exclude repurchase or restructuring	260,886
Both target and acquirer are public firms	9,535
Deal status is completed or withdrawn	9,001
acquirer is seeking to buy more than 50% of target shares	7,602
Deal value is at least $1 million	6,809
Both target and acquirer have a valid CRSP identifier	4,960
EDGAR data is potentially available	4,232
Filling was successfully parsed	3,771

Table 2.2: **Variables definition**

This table contains a description of the variables used in the paper. Variables with the source SEC have been collected using Edgar. M&A filings considered on Edgar include PREM14A, DEFM14A, PREM14C, DEFM14C, S-4, SC 14D1, SC 14D9 forms and their amended versions. The data source *FF* refers to the website of Kenneth French (`https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`)

Panel A: Deal variables

Variable	Definition	Source
Run-up	Market-adjusted acquirer's cumulative stock return during the period $t-205$ to $t-6$ where t is the deal announcement date.	CRSP & (Moeller, Schlingemann, and Stulz, 2007)
Sigma	Market-adjusted acquirer's stock return standard deviation during the period $t-205$ to $t-6$ where t is the deal announcement date.	CRSP & (Moeller, Schlingemann, and Stulz, 2007)
Book-to-Market	Ratio of firm's book equity from the fiscal year prior the deal announcement and firm's market equity measured at 105 days before deal announcement.	COMPUSTAT & CRSP
Free Cash Flow	firm's free cash-flow measured using accounting data from the year prior deal announcement. This variable is computed as EBITDA minus interests and income taxes plus changes in deferred taxes and investment tax credits minus dividends paid to common shares and preferred shares.	COMPUSTAT
Leverage	Sum of short-term and long-term debt scaled by total assets.	COMPUSTAT
Acquiror Size	Log-transformed total acquirer's market capitalization 105 days before deal announcement.	CRSP
Deal Value	Log-transformed total consideration paid by acquirer to purchase the shares of target.	SDC
Relative Size	Deal value divided by acquirer's market capitalization.	SDC & CRSP
BB Advisor	Dummy variable equal to one if at least one of the financial advisors is a bulge bracket firm.	SDC & (Golubov, Petmezas, and Travlos, 2012)
In-House M&A Experience	Number of deals where the firm has not used any financial advisors in the five years before deal announcement.	SDC & (Golubov, Petmezas, and Travlos, 2012)
Specialist Target Industry	Dummy variable equal to one if at least one of the financial advisors is a specialist of target's industry. Industry specialization is computed over the deals closed by the advisor in the last five years.	SDC & (Graham et al., 2017)
Specialist Acquiror Industry	Dummy variable equal to one if at least one of the financial advisors is a specialist of acquirer's industry.	SDC & (Graham et al., 2017)
Hire Ex Advisor	Dummy variable equal to one if target(acquirer) hires at least one past advisor of acquirer(target) with which the firm has worked with over the last five years.	SDC & (Chang et al., 2016)
Has Ex Advisor	Dummy variable equal to one if acquirer(target) has worked with any financial advisors over the last five years.	SDC & (Chang et al., 2016)
Deal Is Completed	Dummy variable equals to one if deal is completed.	SDC
Friendly	Dummy variable equals to one if deal is classified as friendly or solicited.	SDC
Toehold	Dummy variable equals to one if the percentage of shares owned by the acquirer at announcement date is greater or equal than 5%.	SDC & (Moeller, Schlingemann, and Stulz, 2004)
Stock Deal	Dummy variable equals to one if acquirer's stock is proposed as part of the merger consideration.	SDC
Has ADV Duty	Dummy variable equals to one if at least one advisor has an advisory duty toward its client.	SDC
Has REP Duty	Dummy variable equals to one if at least one advisor has an representation duty toward its client.	SDC

Table continued on next page

Table continued

$Has\,FIN\ Duty$	Dummy variable equals to one if at least one advisor provides financing to its client.	SDC
$Has\,FO\ Duty$	Dummy variable equals to one if at least one advisor provides a fairness opinion to its client.	SDC
$Premium_{t-X,t}$	Premium paid for the acquisition where the reference date is either 105 or 63 business days before deal announcement.	SDC & SEC
$CAR_{t-1,t+1}$	Cumulative abnormal returns computed from $t-1$ to $t+1$ around announcement date t. Abnormal returns are computed using Fama French 3 factor model. The betas used to compute abnormal returns are estimated in $t-105$ with a 200 days window.	CRSP & FF
$Past\,Advisor\,CAR$	Average CAR of the deals advised by advisors over the last five years.	CRSP & FF
	Panel B: Edgar variables	
$\#Meetings$	Log-transformed 1 + number of meetings during private negotiations.	SEC
$Meetings\,Advisor$	Number of meetings during private negotiations in which financial advisors are involved. This variable is scaled by the total number of meetings.	SEC
$Meetings\,Firm$	Number of meetings during private negotiations in which firms are involved. This variable is scaled by the total number of meetings.	SEC
$Multiple\,Bidders$	Dummy variable equal to one if acquirer has at least one rival bidder during the negotiation process.	SEC & (Graham et al., 2017)
$Initiated\,By\,Target$	Dummy variable equals to one if target initiated contacts with potential bidders.	SEC
$Is\,Filed\,By\,Target$	Dummy variable equals to one if target has filled the SEC filling.	SEC
$Tender\,Offer$	Dummy variable equals to one if the deal is a tender offer.	SEC

Table 2.3: **Summary Statistics**

This table contains summary statistics of variables defined in Table 2.2. Variables are reported in decimal points. Panel A reports deal variables that are common for both target and acquirer. Panel B and Panel C display the variables concerning respectively target and acquirer. Please refer to Table 2.2 for additional information on the computation details of each variable.

Panel A: Deal variables

	$Mean$	25^{th}	50^{th}	75^{th}	SD
$\#Meetings$	3.55	3.04	3.56	4.11	0.77
$Deal\,Value$	19.53	18.1	19.4	20.92	1.92
$Acquiror\,Size$	21.22	19.82	21.23	22.73	1.94
$Relative\,Size$	0.47	0.08	0.24	0.62	0.61
$Run\text{-}up$	0.12	-0.09	0.06	0.25	0.37
$Sigma$	0.02	0.01	0.02	0.03	0.01
$Premium_{t-105,t}$	0.42	0.16	0.37	0.62	0.42
$Premium_{t-63,t}$	0.41	0.17	0.36	0.59	0.38
$Deal\,Is\,Completed$	0.93	1	1	1	0.25
$Initiated\,By\,Target$	0.4	0	0	1	0.49
$Is\,Filed\,By\,Target$	0.35	0	0	1	0.48
$Friendly$	0.98	1	1	1	0.14
$Tender\,Offer$	0.14	0	0	0	0.35
$Stock\,Deal$	0.6	0	1	1	0.49
$Toehold$	0.02	0	0	0	0.15
$Multiple\,Bidders$	0.62	0	1	1	0.49

	Panel B: Target					**Panel C: Acquirer**				
	$Mean$	25^{th}	50^{th}	75^{th}	SD	$Mean$	25^{th}	50^{th}	75^{th}	SD
$CAR_{t-1,t+1}$	0.1	0.05	0.11	0.14	0.08	-0.01	-0.05	-0.01	0.02	0.06
$Past\,Advisor\,CAR$	0.09	0.08	0.09	0.1	0.03	0	-0.01	0	0.01	0.02
$Book\text{-}to\text{-}Market$	0.68	0.33	0.56	0.85	0.53	0.58	0.29	0.49	0.73	0.43
$Free\,Cash\,Flow$	0	0.01	0.03	0.09	0.19	0.05	0.02	0.05	0.1	0.11
$Leverage$	0.21	0.03	0.15	0.33	0.2	0.21	0.07	0.17	0.32	0.18
$Meetings\,Firm$	0.87	0.83	0.9	0.95	0.13	0.7	0.57	0.73	0.85	0.2
$Meetings\,Advisor$	0.34	0.22	0.33	0.46	0.18	0.14	0.04	0.12	0.21	0.13
$Has\,ADV\,Duty$	0.95	1	1	1	0.21	0.91	1	1	1	0.29
$Has\,FO\,Duty$	0.67	0	1	1	0.47	0.33	0	0	1	0.47
$Has\,REP\,Duty$	0.03	0	0	0	0.17	0.01	0	0	0	0.1
$Has\,FIN\,Duty$	0	0	0	0	0.05	0.04	0	0	0	0.19
$BB\,Advisor$	0.34	0	0	1	0.47	0.42	0	0	1	0.49
$In\text{-}House\,M\&A\,Experience$	0.41	0	0	1	0.49	0.69	0	1	1	0.46
$Hire\,Ex\,Advisor$	0.02	0	0	0	0.13	0.07	0	0	0	0.26
$Has\,Ex\,Advisor$	0.22	0	0	0	0.42	0.45	0	0	1	0.5
$Specialist\,Acquiror\,Industry$	0.61	0	1	1	0.49	0.55	0	1	1	0.5
$Specialist\,Target\,Industry$	0.61	0	1	1	0.49	0.54	0	1	1	0.5

Table 2.4: **Drivers of Financial Advisor Efforts**

This table contains the results of panel regressions. The dependent variable are the number of meetings in which the financial advisors of the target and the acquirer are involved. Regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per bidder. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: *Meetings Advisor*					
	Independent variables from Target			*Independent variables from Acquirer*		
	(1)	(2)	(3)	(4)	(5)	(6)
Deal Value	0.1 (0.51)	0.1 (0.49)	0.04 (0.18)	−0.07 (−0.42)	−0.2 (−1.18)	−0.2 (−1.06)
Relative Size	−2.71*** (−6.01)	−2.48*** (−5.43)	−2.38*** (−4.51)	2.42*** (5.45)	2.55*** (5.97)	2.03*** (4.12)
Specialist Target Industry	−0.57 (−0.72)	−0.43 (−0.55)	0.34 (0.39)	0.38 (0.57)	0.56 (0.86)	0.63 (0.85)
Specialist Acquiror Industry	2.29*** (2.82)	2.1** (2.57)	0.96 (1.11)	−0.13 (−0.19)	−0.21 (−0.32)	−0.69 (−0.9)
BB Advisor	−0.46 (−0.74)	−0.24 (−0.39)	0.79 (1.14)	0.14 (0.28)	0.28 (0.59)	0.17 (0.31)
In-House M&A Experience	−0.72 (−1.06)	−0.68 (−1)	−0.64 (−0.82)	−0.85 (−1.48)	−0.53 (−0.96)	−0.38 (−0.63)
Hire Ex Advisor	0.55 (0.29)	1.01 (0.54)	1.5 (0.73)	−0.74 (−0.96)	−0.9 (−1.2)	−1.01 (−1.24)
Has Ex Advisor	−1.96** (−2.48)	−1.86** (−2.37)	−1.84** (−2.05)	−1.01* (−1.83)	−1.17** (−2.22)	−0.91 (−1.57)
Has Advisor	22.12*** (7.78)	22.94*** (7.93)	14.92*** (3.15)	9.07*** (6.68)	9.3*** (6.7)	8.4*** (4.13)
Has ADV Duty	7.23*** (5.16)	7.19*** (5.15)	5.2*** (3.35)	5.93*** (4.93)	5.08*** (4.29)	3.37** (2.26)
Has FO Duty	1.9*** (2.98)	1.42** (2.21)	1.25 (1.6)	4.36*** (8.43)	2.95*** (5.7)	3.14*** (5.36)
Has REP Duty	1.54 (1.16)	1.79 (1.34)	2.26 (1.57)	11.21*** (4.73)	11.29*** (4.76)	11.8*** (4.84)
Has FIN Duty	−0.16 (−0.03)	−2 (−0.31)	0.94 (0.14)	−2.28** (−2.41)	−1.68* (−1.79)	−1.28 (−1.36)
Initiated By Target	6.61*** (11.64)	6.25*** (10.95)	5.77*** (8.63)	−2.79*** (−6.51)	−1.08** (−2.46)	−0.34 (−0.71)
Is Filed By Target	1.69*** (2.76)	2.38*** (2.95)	3.11*** (3.31)	−4.19*** (−8.46)	−1.78*** (−2.78)	−1.9*** (−2.65)
Past Advisor CAR	−4 (−0.33)	−1.35 (−0.11)	−5.31 (−0.34)	−28.88* (−1.88)	−16.24 (−1.1)	−20.25 (−0.94)
Friendly		8.81*** (5.13)	5.94*** (2.98)		7.15*** (6.23)	2.9* (1.96)
Tender Offer		−3.57*** (−3.78)	−2.21** (−2.12)		−0.53 (−0.82)	−0.9 (−1.38)
Stock Deal		−1.48** (−2.1)	−1.66* (−1.93)		1.53*** (2.75)	1.16* (1.83)
Toehold		−0.96 (−0.58)	1.35 (0.64)		−0.81 (−0.64)	0.09 (0.06)
#Meetings		−0.03 (−0.05)	0.22 (0.34)		0.21 (0.49)	0.53 (1.13)
Meetings Firm		6.02** (2.53)	3.55 (1.28)		16.55*** (13.45)	15.72*** (11.86)
Past Meetings Advisor			23.35*** (4.86)			13.01** (2)
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	3808	3800	2765	3389	3381	2509

Table 2.5: **Premium and Advisors' Activity**

This table contains the results of panel regressions. Dependent variables are three measures of deal premium. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. The regression coefficients for the full set of regressor is shown in Table 2.A1. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: *Premium*							
	Independent variables from Target				*Independent variables from Acquirer*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	t-105 $\rightarrow$ t	t-105 $\rightarrow$ t	t-63 $\rightarrow$ t	t-63 $\rightarrow$ t	t-105 $\rightarrow$ t	t-105 $\rightarrow$ t	t-63 $\rightarrow$ t	t-63 $\rightarrow$ t
Meetings Advisor	13.94***	9.8*	8.42**	4.03	−25.35***	−18.25***	−24.46***	−19.77***
	(3.38)	(1.88)	(2.27)	(0.83)	(−4.79)	(−2.71)	(−4.89)	(−2.9)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	3477	2197	3561	2230	2985	2163	3058	2197

Table 2.6: **CAR and Advisors' Activity**

This table contains the results of panel regressions. Dependent variables are cumulative abnormal returns of acquirer, target and deal weighted-average respectively. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. The regression coefficients for the full set of regressor is shown in Table 2.A2. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: $CAR_{t-1,t+1}$							
	Independent variables from Target				*Independent variables from Acquirer*			
	(1) *Target*	(2) *Target*	(3) *Acquirer*	(4) *Acquirer*	(5) *Target*	(6) *Target*	(7) *Acquirer*	(8) *Acquirer*
Meetings Advisor	0.75	−0.58	1.01**	−0.42	−4.19***	−0.99	−3.13***	0.06
	(0.94)	(−0.52)	(2.02)	(−0.54)	(−3.46)	(−0.64)	(−3.67)	(0.04)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	3783	2197	4280	2196	3236	2163	3609	2162

Table 2.7: **Post-Merger Profitability and Advisors' Activity**

This table contains the results of panel regressions. Dependent variables are average free cash-flows over the five years following deal consummation date. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. The regression coefficients for the full set of regressor is shown in Table 2.A3. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: *Post Merger FCF*			
	Independent variables from Target		*Independent variables from Acquirer*	
	(1)	(2)	(3)	(4)
Meetings Advisor	1.49**	−0.01	−2.8**	0.12
	(1.97)	(−0.01)	(−2.26)	(0.1)
Controls	No	Yes	No	Yes
Target Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# Deals	2422	1411	1989	1402

Table 2.8: **FinBERT Sentence Sentiment Examples**

This table provides some example of sentences found in "Background of the Deals" sections along with their sentiment. To this end, I use the FinBERT model of Huang, Wang, and Yang (2023). The sources of the SEC filings are listed below the examples.

Sentence	*Sentiment*
At the meeting, Mr. Hell discussed the potential benefits of a combination of the DivX and Sonic businesses, including the perceived complementary nature and relative strengths of each business and the operational, marketing and growth synergies that might be achieved by combining the two companies. `https://www.sec.gov/Archives/edgar/data/916235/000119312510157443/ds4.htm`	Positive
Starwood Financial consulted with its financial advisor, Bear Stearns, and identified TriNet as a potential merger partner because of the complementary nature of TriNet's and Starwood Financial's credit tenant lease businesses, as well as Starwood Financial's recent success in structuring investments in the sector. `https://www.sec.gov/Archives/edgar/data/831972/000104746999033670/0001047469-99-033670.txt`	Positive
These 2005 earnings per share forecasts, which were prepared by Maytag management, were significantly higher than the 2005 earnings per share forecasts prepared in connection with presentations to the Maytag board on May 11 and 19, 2005, of $0.56 to $0.88 (see the section entitled "—Financial Projections" beginning on page 66 of this proxy statement/prospectus) and did not reflect subsequent declines in Maytag's business. `https://www.sec.gov/Archives/edgar/data/106640/000104746905023630/a2163113zs-4.htm`	Positive
The Company concluded that the future growth prospects of Party B, whose stock price had declined significantly over the preceding 12 months, were not attractive. `https://www.sec.gov/Archives/edgar/data/1651235/000119312519204131/d751132dprem14a.htm`	Negative
The Board and management have also considered various challenges that we have faced as a public company, including the challenges currently facing the retail sector, including the impact of the COVID-19 pandemic and government and private business responses to the pandemic. `https://www.sec.gov/Archives/edgar/data/761648/000119312522095654/d328333dprem14a.htm`	Negative
The MSLO Board of Directors discussed the strategic rationale of a potential transaction and determined that consideration of a potential transaction and various other strategic alternatives available to MSLO would be in MSLO's best interests at the time given certain challenges facing MSLO, including difficulty developing the business internationally, declining domestic revenues, a lack of a succession plan for Ms. Stewart and industry-wide shifts in the publishing business. `https://www.sec.gov/Archives/edgar/data/1648428/000114420415044975/v415933_s4.htm`	Negative
At a regular meeting of the Healthaxis board of directors held on August 8, 2007, Healthaxis management recommended and the Healthaxis board of directors approved the engagement of Ansley Securities, LLC ("Ansley") as financial advisor to assist in the process of selling Healthaxis. `https://www.sec.gov/Archives/edgar/data/1015920/000104746908011129/a2188609zprem14a.htm`	Neutral
During its March 14, 2002 meeting, the EEX board heard presentations from EEX management and Morgan Stanley concerning the four proposals which had been received and ongoing discussions with two other potential purchasers. `https://www.sec.gov/Archives/edgar/data/912750/000095013402007526/h97760sv4.txt`	Neutral
On July 29, 2021, Desktop Metal and ExOne entered into a supplement to their May 1, 2020 non-disclosure agreement so that their respective legal and financial advisors could review diligence material without having to disclose competitively sensitive information to their respective business teams. `https://www.sec.gov/Archives/edgar/data/1754820/000110465921115991/dm-20210915xs4a.htm`	Neutral

Table 2.9: **Premium, Advisors' Activity and Sentiment**

This table contains the results of panel regressions. Dependent variables are three measures of deal premium. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. The regression coefficients for the full set of regressor is shown in Table 2.A4. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: *Premium*							
	Independent variables from Target				*Independent variables from Acquirer*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	t-105 → t	t-105 → t	t-63 → t	t-63 → t	t-105 → t	t-105 → t	t-63 → t	t-63 → t
Meetings Advisor	14.77***	10.16*	9.23**	4.43	−27.37***	−19.45***	−25.88***	−20.28***
	(3.56)	(1.95)	(2.46)	(0.91)	(−5.11)	(−2.89)	(−5.15)	(−2.99)
Sentiment	3.78***	3.06**	2.43**	1.85	2.68**	3.88***	1.83*	2.29*
	(3.3)	(2.09)	(2.41)	(1.49)	(2.14)	(2.61)	(1.69)	(1.86)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	3467	2193	3551	2226	2977	2159	3050	2193

Table 2.10: **CAPE and Advisors' Activity**

This table contains the results of panel regressions. Dependent variables are two measures of deal premium. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

Panel A: Large CAPE

	Dependent variable: *Premium*							
	Independent variables from Target				*Independent variables from Acquirer*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	t-105 → t	t-105 → t	t-63 → t	t-63 → t	t-105 → t	t-105 → t	t-63 → t	t-63 → t
Meetings Advisor	16.44***	20.73***	11.97**	15.59**	−21.22***	−11.05	−22.71***	−17.58*
	(2.71)	(2.63)	(2.26)	(2.22)	(−2.85)	(−1.13)	(−3.16)	(−1.74)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	1721	1082	1772	1103	1515	1072	1562	1093

Panel B: Low CAPE

Meetings Advisor	9.64*	−1.74	3.24	−7.8	−28.39***	−20.83**	−26.03***	−22.51**
	(1.76)	(−0.24)	(0.62)	(−1.11)	(−3.76)	(−2.16)	(−3.7)	(−2.36)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	1746	1110	1779	1122	1463	1086	1489	1099

Table 2.11: **Meeting Initiation and Advisors' Activity**

This table contains the results of panel regressions. Dependent variables are the proportions of meetings initiated by target or acquiring firm. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. The regression coefficients for the full set of regressor is shown in Table 2.A5. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: *Proportion of initiated Meeting*			
	Independent variables from Target		*Independent variables from Acquirer*	
	(1)	(2)	(3)	(4)
Meetings Advisor	33.86*** (12.8)	19.33*** (5.47)	29.31*** (8.85)	15.48*** (3.3)
Controls	No	Yes	No	Yes
Target Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# Deals	4360	2284	3693	2254

Table 2.12: **Future Business Relationship and Advisors' Activity**

This table contains the results of panel regressions. Dependent variables are dummy variables equal to one if some of the advisors employed are re-used in the client's next deal. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. The regression coefficients for the full set of regressor is shown in Table 2.A6. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: *Will Advisors Be Reused* *Independent variables from Acquirer*	
	(1)	(2)
Meetings Advisor	10.83 (1.17)	25.82* (1.92)
Controls	No	Yes
Target Industry FE	Yes	Yes
Year FE	Yes	Yes
# Deals	1648	1086

Table 2.13: **Robustness Check**

This table contains the results of panel regressions. Dependent variables are three measures of deal premium. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per year and industry. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

Panel A: Unscaled meeting count

	Dependent variable: *Premium*							
	Independent variables from Target				*Independent variables from Acquirer*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	t-105 → t	t-105 → t	t-63 → t	t-63 → t	t-105 → t	t-105 → t	t-63 → t	t-63 → t
#Meetings Advisor	−0.22 (−0.24)	3.34** (1.98)	−0.74 (−0.86)	1.75 (1.18)	−4.38*** (−5.86)	−2.74** (−2.44)	−4.8*** (−6.79)	−3.83*** (−3.42)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	3477	2197	3561	2230	2985	2163	3058	2197

Panel B: Controlling for negotiation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Meetings Advisor	13.94*** (3.38)	9.2* (1.76)	8.42** (2.27)	3.93 (0.8)	−25.35*** (−4.79)	−18.24*** (−2.71)	−24.46*** (−4.89)	−19.8*** (−2.91)
Meetings Firm X Negotiation		14.34* (1.87)		2.43 (0.34)		14.12* (1.65)		6.18 (0.81)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	3477	2197	3561	2230	2985	2163	3058	2197

Panel C: Controlling for fees

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Meetings Advisor	13.94*** (3.38)	11.19** (2.43)	8.42** (2.27)	6.84 (1.61)	−25.35*** (−4.79)	−24.79*** (−3.07)	−24.46*** (−4.89)	−13.45* (−1.77)
Total Fees		0.71 (0.85)		−0.31 (−0.41)		−0.55 (−0.47)		−1.25 (−1.08)
#Meetings		−5.62*** (−3.63)		−3.85*** (−2.82)		−3.27 (−1.47)		−2.42 (−1.29)
Meetings Firm		6.93 (1.02)		5.19 (0.92)		−0.13 (−0.02)		−0.86 (−0.15)
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	3477	2684	3561	2732	2985	1262	3058	1286

Appendix

For better readability, some of the tables report only the results for the main independent variables of interest. This section provides the complete set of regression coefficients obtained for the truncated tables.

Table 2.A1: **Premium and Advisors' Activity**

This table contains the results of panel regressions. Dependent variables are three measures of deal premium. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: *Premium*							
	Independent variables from Target				*Independent variables from Acquirer*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	t-105 → t	t-105 → t	t-63 → t	t-63 → t	t-105 → t	t-105 → t	t-63 → t	t-63 → t
Meetings Advisor	13.94***	9.8*	8.42**	4.03	−25.35***	−18.25***	−24.46***	−19.77***
	(3.38)	(1.88)	(2.27)	(0.83)	(−4.79)	(−2.71)	(−4.89)	(−2.9)
Target Book-to-Market		18.56***		17.51***		17.49***		16.67***
		(7.14)		(6.81)		(6.76)		(6.67)
Target Free Cash Flow		−9.24		−6.6		−9.86		−6.89
		(−1.06)		(−0.91)		(−1.1)		(−0.99)
Target Leverage		1.12		0.32		1.48		0.95
		(0.19)		(0.06)		(0.25)		(0.18)
Run-up		23.6***		13.16***		22.95***		12.12***
		(8.12)		(4.87)		(7.65)		(4.42)
Sigma		−466.11***		−347.01**		−440.36***		−349.56**
		(−3.19)		(−2.51)		(−3.01)		(−2.51)
Acquiror Book-to-Market		2.8		−0.34		2.12		−0.96
		(1.13)		(−0.15)		(0.84)		(−0.42)
Acquiror Free Cash Flow		11.29		4.79		10.33		4.77
		(0.93)		(0.43)		(0.84)		(0.44)
Acquiror Leverage		−11.05*		−5.97		−11.78**		−8.19
		(−1.92)		(−1.15)		(−2.03)		(−1.59)
Acquiror Size		0.53		1.19		−0.27		0.3
		(0.49)		(1.15)		(−0.23)		(0.27)
Deal Value		1.08		−0.52		0.48		−0.3
		(0.92)		(−0.43)		(0.41)		(−0.26)
Relative Size		1.32		1.51		1.8		2.17
		(0.59)		(0.74)		(0.78)		(1.05)
Specialist Target Industry		1.44		4.05		0.66		−1.78
		(0.57)		(1.63)		(0.25)		(−0.74)
Specialist Acquiror Industry		−1.13		0.31		−2.48		0.9
		(−0.45)		(0.13)		(−0.95)		(0.39)
BB Advisor		−2.89		−0.02		2.44		1.21
		(−1.39)		(−0.01)		(1.22)		(0.66)
Past Advisor CAR		−151.1**		−45.68		−35.27		−56.72
		(−2.41)		(−0.81)		(−0.88)		(−1.49)
In-House M&A Experience		2.31		1.12		0.51		−1.54
		(1.05)		(0.52)		(0.21)		(−0.67)
Hire Ex Advisor		−2.3		4.39		7.69**		2.95
		(−0.56)		(1.05)		(2.27)		(0.97)
Has Ex Advisor		−6.53***		−6.47***		−0.23		2.08
		(−2.71)		(−2.79)		(−0.1)		(0.95)
Has Advisor		0.65		−24.78		5.18		−4.92
		(0.04)		(−1.38)		(0.79)		(−0.8)
Has ADV Duty		−3.89		−0.55		−2.8		4.1
		(−0.59)		(−0.1)		(−0.55)		(0.98)
Has FO Duty		1.02		−0.8		−5.94***		−4.97***
		(0.5)		(−0.43)		(−2.99)		(−2.75)
Has REP Duty		−4.3		−6.98**		−0.71		−4.13
		(−0.95)		(−2.09)		(−0.11)		(−0.91)
Has FIN Duty		−23.23***		−14.77***		1.34		−0.13
		(−4.61)		(−2.94)		(0.42)		(−0.05)
Initiated By Target		−2.78*		−5.11***		−2.58		−5.62***
		(−1.67)		(−3.39)		(−1.5)		(−3.55)
Is Filed By Target		2.45		3.54		2.77		3.42
		(0.91)		(1.48)		(1.01)		(1.43)
Friendly		4.4		4.72		10.46*		9.32**
		(0.83)		(1.19)		(1.95)		(2.32)
Tender Offer		−0.63		2.34		−0.57		3.62
		(−0.21)		(0.8)		(−0.19)		(1.23)
Stock Deal		−5.8**		−3.85*		−4.37*		−2.05
		(−2.42)		(−1.8)		(−1.76)		(−0.95)
Toehold		−2.69		−5.24		−2.26		−4.9
		(−0.48)		(−0.9)		(−0.4)		(−0.81)
Multiple Bidders		−2.28		0.07		−2.51		−0.04
		(−1.19)		(0.04)		(−1.27)		(−0.02)
#Meetings		−2.75		−3.56**		−2.22		−3.86**
		(−1.57)		(−2.16)		(−1.27)		(−2.3)
Meetings Firm		16.93***		10.88*		1.77		−0.7
		(2.6)		(1.87)		(0.37)		(−0.16)
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	3477	2197	3561	2230	2985	2163	3058	2197

Table 2.A2: **CAR and Advisors' Activity**

This table contains the results of panel regressions. Dependent variables are cumulative abnormal returns of acquirer, target and deal weighted-average respectively. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: $CAR_{t-1,t+1}$							
	Independent variables from Target				*Independent variables from Acquirer*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Target	*Target*	*Acquirer*	*Acquirer*	*Target*	*Target*	*Acquirer*	*Acquirer*
Meetings Advisor	0.75	−0.58	1.01**	−0.42	−4.19***	−0.99	−3.13***	0.06
	(0.94)	(−0.52)	(2.02)	(−0.54)	(−3.46)	(−0.64)	(−3.67)	(0.04)
Target Book-to-Market		0.39		0		0.37		−0.01
		(0.8)		(0)		(0.79)		(−0.04)
Target Free Cash Flow		1.73		−0.81		2.16		−0.76
		(1)		(−0.71)		(1.28)		(−0.65)
Target Leverage		0.74		−0.53		0.5		−0.28
		(0.69)		(−0.67)		(0.48)		(−0.35)
Run-up		−1.35**		−0.81*		−1.06*		−0.73
		(−2.45)		(−1.72)		(−1.94)		(−1.53)
Sigma		−35.77		−13.38		−34.25		−17.52
		(−1.3)		(−0.57)		(−1.26)		(−0.75)
Acquiror Book-to-Market		−0.43		0.42		−0.56		0.33
		(−1.06)		(1.12)		(−1.39)		(0.87)
Acquiror Free Cash Flow		3.93*		4.68**		4.19*		5.06**
		(1.7)		(2.24)		(1.8)		(2.28)
Acquiror Leverage		−1.47		2.58**		−1.07		2.83***
		(−1.34)		(2.56)		(−0.96)		(2.82)
Acquiror Size		0.21		0.22		0.19		0.17
		(0.96)		(1.22)		(0.8)		(0.92)
Deal Value		−0.33		−0.53***		−0.31		−0.57***
		(−1.42)		(−2.84)		(−1.24)		(−3.1)
Relative Size		−0.6		−0.33		−0.6		−0.14
		(−1.41)		(−0.81)		(−1.36)		(−0.35)
Specialist Target Industry		0.04		0.57		0.26		−0.59
		(0.09)		(1.54)		(0.55)		(−1.45)
Specialist Acquiror Industry		−0.2		−0.49		0.32		−0.04
		(−0.38)		(−1.33)		(0.65)		(−0.1)
BB Advisor		−0.38		−0.25		−0.21		0.51
		(−0.96)		(−0.75)		(−0.54)		(1.62)
Past Advisor CAR		−6.57		−0.72		1.83		1.53
		(−0.52)		(−0.07)		(0.21)		(0.24)
In-House M&A Experience		−0.62		−0.39		−0.42		0.01
		(−1.4)		(−1.09)		(−0.84)		(0.02)
Hire Ex Advisor		1.53		−1.2		0.29		0.87**
		(1.48)		(−1.43)		(0.46)		(2.02)
Has Ex Advisor		0.25		−0.1		0.11		−0.3
		(0.52)		(−0.27)		(0.23)		(−0.8)
Has Advisor		−2.99		−4.84		−0.24		1.96*
		(−0.85)		(−1.52)		(−0.18)		(1.79)
Has ADV Duty		2.17**		1.74**		−0.76		−1.57*
		(2.08)		(2.08)		(−0.75)		(−1.87)
Has FO Duty		0.28		0.31		−0.21		−0.33
		(0.69)		(0.97)		(−0.5)		(−0.95)
Has REP Duty		−0.64		0.84		−4.01**		−2.42*
		(−0.73)		(1.18)		(−2.45)		(−1.86)
Has FIN Duty		−1.76		0.58		0.78		−3.06***
		(−1.31)		(0.19)		(1.1)		(−4.42)
Initiated By Target		−0.83**		−0.33		−0.9**		−0.41
		(−2.27)		(−1.25)		(−2.42)		(−1.55)
Is Filed By Target		0.24		1.86***		0.09		1.8***
		(0.46)		(4.27)		(0.16)		(4.08)
Friendly		−1.93**		0.79		−1.77**		1.04
		(−2.21)		(0.88)		(−2.04)		(1.16)
Tender Offer		1.32***		−0.66		1.42***		−0.71*
		(2.63)		(−1.54)		(2.82)		(−1.68)
Stock Deal		−0.66		−1.33***		−0.83*		−1.38***
		(−1.39)		(−3.38)		(−1.73)		(−3.41)
Toehold		0.57		0.2		0.42		0.06
		(0.48)		(0.22)		(0.34)		(0.06)
Multiple Bidders		−0.81**		−0.71**		−0.69*		−0.69**
		(−2.02)		(−2.23)		(−1.71)		(−2.13)
$Premium_{t-105,t}$		3.04***		−0.61		3.03***		−0.61*
		(6.31)		(−1.63)		(6.25)		(−1.66)
#Meetings		−0.95***		0.12		−0.86**		0.17
		(−2.75)		(0.43)		(−2.48)		(0.64)
Meetings Firm		1.47		−1.45		1.08		0.2
		(1.06)		(−1.44)		(1.12)		(0.28)
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	3783	2197	4280	2196	3236	2163	3609	2162

Table 2.A3: **Post-Merger Profitability and Advisors' Activity**

This table contains the results of panel regressions. Dependent variables are average free cash-flows over the five years following deal consummation date. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: *Post Merger FCF*			
	Independent variables from Target		*Independent variables from Acquirer*	
	(1)	(2)	(3)	(4)
Meetings Advisor	1.49^{**} (1.97)	−0.01 (−0.01)	-2.8^{**} (−2.26)	0.12 (0.1)
Target Book-to-Market		−0.09 (−0.25)		−0.35 (−0.97)
Target Free Cash Flow		0.96 (0.64)		1.28 (0.9)
Target Leverage		−0.39 (−0.52)		−0.23 (−0.29)
Run-up		0.29 (0.56)		0.13 (0.25)
Sigma		−29.07 (−1.07)		−33.87 (−1.29)
Acquiror Book-to-Market		0.4 (0.85)		0.28 (0.61)
Acquiror Free Cash Flow		30.91^{***} (6.78)		31.42^{***} (7.01)
Acquiror Leverage		0.82 (0.77)		0.35 (0.33)
Acquiror Size		0.49^{***} (2.64)		0.42^{**} (2.36)
Deal Value		-0.27^{*} (−1.69)		-0.44^{***} (−2.78)
Relative Size		0.06 (0.22)		0.15 (0.53)
Specialist Target Industry		−0.05 (−0.13)		−0.4 (−1.03)
Specialist Acquiror Industry		-0.7^{*} (−1.96)		0.6 (1.52)
BB Advisor		−0.2 (−0.66)		0.86^{***} (2.7)
Past Advisor CAR		6.09 (0.66)		−3.71 (−0.48)
In-House M&A Experience		−0.43 (−1.19)		−0.29 (−0.81)
Hire Ex Advisor		0.46 (0.73)		−0.12 (−0.35)
Has Ex Advisor		0 (0.01)		0.48 (1.36)
Has Advisor		−0.98 (−0.38)		0.17 (0.15)
Has ADV Duty		−0.43 (−0.3)		−0.11 (−0.12)
Has FO Duty		−0.07 (−0.27)		-0.52^{*} (−1.78)
Has REP Duty		−0.09 (−0.18)		−0.34 (−0.23)
Has FIN Duty		−7.96 (−1.4)		0.11 (0.16)
Initiated By Target		−0.17 (−0.7)		−0.06 (−0.26)
Is Filed By Target		0.4 (1.05)		0.44 (1.07)
Friendly		−0.77 (−1.05)		−0.72 (−1.04)
Tender Offer		−0.6 (−1.53)		-0.77^{*} (−1.93)
Stock Deal		0.08 (0.23)		−0.11 (−0.32)
Toehold		-1.91^{*} (−1.79)		−1.76 (−1.52)
Multiple Bidders		−0.01 (−0.03)		0.09 (0.32)
#Meetings		0.06 (0.23)		−0.05 (−0.2)
Meetings Firm		0.2 (0.27)		1.9^{***} (2.78)
Target Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# Deals	2422	1411	1989	1402

Table 2.A4: **Premium, Advisors' Activity and Sentiment**

This table contains the results of panel regressions. Dependent variables are three measures of deal premium. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: *Premium*							
	Independent variables from Target				*Independent variables from Acquirer*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	t-105 → t	t-105 → t	t-63 → t	t-63 → t	t-105 → t	t-105 → t	t-63 → t	t-63 → t
Meetings Advisor	14.77*** (3.56)	10.16* (1.95)	9.23** (2.46)	4.43 (0.91)	−27.37*** (−5.11)	−19.45*** (−2.89)	−25.88*** (−5.15)	−20.28*** (−2.99)
Sentiment	3.78*** (3.3)	3.06** (2.09)	2.43** (2.41)	1.85 (1.49)	2.68** (2.14)	3.88*** (2.61)	1.83* (1.69)	2.29* (1.86)
Target Book-to-Market		18.73*** (7.19)		17.61*** (6.83)		17.74*** (6.84)		16.81*** (6.7)
Target Free Cash Flow		−9.08 (−1.04)		−6.57 (−0.91)		−9.56 (−1.07)		−6.8 (−0.98)
Target Leverage		1.81 (0.31)		0.7 (0.14)		2.32 (0.39)		1.45 (0.28)
Run-up		23.78*** (8.15)		13.27*** (4.89)		23.12*** (7.7)		12.21*** (4.45)
Sigma		−458.48*** (−3.13)		−342.25** (−2.48)		−430.58*** (−2.93)		−343.96** (−2.46)
Acquiror Book-to-Market		2.89 (1.16)		−0.29 (−0.13)		2.24 (0.89)		−0.91 (−0.39)
Acquiror Free Cash Flow		12.34 (1.01)		5.36 (0.49)		11.55 (0.94)		5.37 (0.49)
Acquiror Leverage		−10.92* (−1.9)		−5.86 (−1.13)		−11.58** (−2)		−8.09 (−1.57)
Acquiror Size		0.59 (0.54)		1.23 (1.18)		−0.27 (−0.23)		0.31 (0.28)
Deal Value		0.85 (0.71)		−0.66 (−0.54)		0.31 (0.27)		−0.4 (−0.35)
Relative Size		1.35 (0.59)		1.5 (0.74)		1.82 (0.78)		2.16 (1.05)
Specialist Target Industry		1.51 (0.59)		4.06 (1.62)		0.89 (0.34)		−1.74 (−0.73)
Specialist Acquiror Industry		−1.19 (−0.47)		0.31 (0.13)		−2.68 (−1.04)		0.77 (0.33)
BB Advisor		−2.79 (−1.34)		0.04 (0.02)		2.44 (1.22)		1.15 (0.63)
Past Advisor CAR		−141.28** (−2.23)		−42.08 (−0.74)		−31.23 (−0.78)		−54.78 (−1.44)
In-House M&A Experience		2.51 (1.15)		1.21 (0.57)		0.63 (0.26)		−1.45 (−0.63)
Hire Ex Advisor		−2.23 (−0.54)		4.37 (1.04)		7.71** (2.29)		2.94 (0.97)
Has Ex Advisor		−6.68*** (−2.77)		−6.54*** (−2.82)		−0.41 (−0.18)		2 (0.91)
Has Advisor		0.41 (0.02)		−24.93 (−1.39)		4.95 (0.75)		−4.94 (−0.8)
Has ADV Duty		−3.89 (−0.59)		−0.52 (−0.1)		−3 (−0.59)		3.95 (0.95)
Has FO Duty		1 (0.49)		−0.78 (−0.42)		−6.23*** (−3.13)		−5.18*** (−2.84)
Has REP Duty		−4.21 (−0.91)		−6.98** (−2.06)		−0.74 (−0.11)		−4.27 (−0.92)
Has FIN Duty		−23.19*** (−4.34)		−14.73*** (−2.85)		1.3 (0.4)		−0.14 (−0.05)
Initiated By Target		−2.7 (−1.62)		−5.08*** (−3.37)		−2.54 (−1.47)		−5.6*** (−3.54)
Is Filed By Target		2.69 (1)		3.73 (1.55)		2.87 (1.04)		3.55 (1.47)
Friendly		5.12 (0.96)		5.16 (1.3)		11.76** (2.18)		10.11** (2.51)
Tender Offer		−0.98 (−0.33)		2.1 (0.71)		−0.9 (−0.29)		3.36 (1.14)
Stock Deal		−6.09** (−2.51)		−4* (−1.85)		−4.68* (−1.86)		−2.19 (−1)
Toehold		−2.79 (−0.48)		−5.08 (−0.87)		−2.14 (−0.37)		−4.54 (−0.74)
Multiple Bidders		−2.08 (−1.09)		0.14 (0.08)		−2.34 (−1.19)		0.01 (0.01)
#Meetings		−2.55 (−1.46)		−3.39** (−2.04)		−1.96 (−1.12)		−3.66** (−2.17)
Meetings Firm		17*** (2.6)		10.98* (1.88)		0.78 (0.16)		−1.25 (−0.28)
Target Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Deals	3467	2193	3551	2226	2977	2159	3050	2193

Table 2.A5: **Meeting Initiation and Advisors' Activity**

This table contains the results of panel regressions. Dependent variables are the proportions of meetings initiated by target or acquiring firm. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: *Proportion of initiated Meeting*			
	Independent variables from Target		*Independent variables from Acquirer*	
	(1)	(2)	(3)	(4)
Meetings Advisor	33.86*** (12.8)	19.33*** (5.47)	29.31*** (8.85)	15.48*** (3.3)
Target Book-to-Market		0.79 (0.73)		−0.54 (−0.52)
Target Free Cash Flow		−3.8 (−1.51)		2.52 (0.98)
Target Leverage		0.21 (0.08)		2.39 (0.91)
Run-up		−0.21 (−0.17)		0.61 (0.51)
Sigma		−32.59 (−0.5)		32.14 (0.5)
Acquiror Book-to-Market		0.84 (0.77)		−1.2 (−1.1)
Acquiror Free Cash Flow		3.55 (0.74)		−6.04 (−1.28)
Acquiror Leverage		−0.41 (−0.14)		1.01 (0.34)
Acquiror Size		0.56 (0.86)		−0.67 (−1)
Deal Value		−0.38 (−0.55)		0.54 (0.82)
Relative Size		1.18 (0.92)		−1.4 (−1.08)
Specialist Target Industry		0.06 (0.05)		−2.63** (−2.08)
Specialist Acquiror Industry		−1.74 (−1.36)		3.51*** (2.83)
BB Advisor		−0.4 (−0.4)		−0.4 (−0.39)
Past Advisor CAR		44.69 (1.39)		−8.42 (−0.37)
In-House M&A Experience		1.82 (1.57)		−0.05 (−0.04)
Hire Ex Advisor		−3.87 (−1.56)		−1.04 (−0.58)
Has Ex Advisor		−0.38 (−0.3)		−0.77 (−0.64)
Has Advisor		−0.76 (−0.13)		−11.01** (−2.48)
Has ADV Duty		2.68 (0.94)		3.81 (1.12)
Has FO Duty		−0.96 (−0.82)		0.88 (0.78)
Has REP Duty		0.4 (0.18)		−0.66 (−0.14)
Has FIN Duty		−14.78*** (−2.87)		−1.8 (−0.97)
Initiated By Target		12.15*** (11.49)		−10.68*** (−9.96)
Is Filed By Target		2.51* (1.68)		−0.04 (−0.03)
Friendly		−0.34 (−0.11)		−5.09* (−1.72)
Tender Offer		−7.01*** (−4.7)		5.87*** (4.11)
Stock Deal		0.88 (0.59)		−1.6 (−1.08)
Toehold		0.14 (0.05)		0.86 (0.34)
Multiple Bidders		4.44*** (4.05)		−3.15*** (−2.81)
#Meetings		−4.81*** (−3.92)		5.39*** (4.19)
Meetings Firm		12.1** (2.12)		22.62*** (5.95)
Target Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# Deals	4360	2284	3693	2254

Table 2.A6: **Future Business Relationship and Advisors' Activity**

This table contains the results of panel regressions. Dependent variables are dummy variables equal to one if some of the advisors employed are re-used in the client's next deal. Independent variables are variables extracted from the negotiation contained in the "Background of the Deal" section from SEC filings. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per acquirer. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable: *Will Advisors Be Reused*	
	Independent variables from Acquirer	
	(1)	(2)
Meetings Advisor	10.83 (1.17)	25.82* (1.92)
Target Book-to-Market		5.04 (1.47)
Target Free Cash Flow		12.58 (1.57)
Target Leverage		12.03 (1.37)
Run-up		−1.4 (−0.42)
Sigma		−50.7 (−0.27)
Acquiror Book-to-Market		−2.65 (−0.62)
Acquiror Free Cash Flow		−2.49 (−0.15)
Acquiror Leverage		−11.59 (−1.28)
Acquiror Size		4.78** (2)
Deal Value		−3.05 (−1.32)
Relative Size		8.09* (1.81)
Specialist Target Industry		3.31 (0.7)
Specialist Acquiror Industry		0.09 (0.02)
BB Advisor		9.44*** (3.22)
Past Advisor CAR		−125.76** (−2.05)
In-House M&A Experience		−2.94 (−0.79)
Hire Ex Advisor		−3.18 (−0.58)
Has Ex Advisor		4.89 (1.33)
Has ADV Duty		13.47*** (2.86)
Has FO Duty		−2.97 (−0.91)
Has REP Duty		−0.56 (−0.04)
Has FIN Duty		1.11 (0.17)
Initiated By Target		0.57 (0.19)
Is Filed By Target		1.66 (0.35)
Friendly		−34.49*** (−3.52)
Tender Offer		−4.44 (−0.94)
Stock Deal		−1.32 (−0.31)
Toehold		−22.43*** (−3.9)
Multiple Bidders		−4.66 (−1.4)
#Meetings		0.48 (0.16)
Meetings Firm		−5.91 (−0.68)
Controls	No	Yes
Target Industry FE	Yes	Yes
Year FE	Yes	Yes
# Deals	1648	1086

Chapter 3

Takeover Probability and Firms' Risk Disclosures

3.1. Introduction

M&A deal announcements have a large impact on asset prices. When information is released to the market, target firms generally experience a large increase in their share prices to reflect part of the acquisition premium. Therefore, takeover probabilities are strongly linked to the cross-section of stock returns (Cremers, Nair, and John, 2009). However, the determinants of why companies engage in mergers are not yet well understood. Mergers can be undertaken to achieve valuable product synergies (Hoberg and Phillips, 2010) or to reduce cash-flow uncertainty through vertical integrations (Garfinkel and Hankins, 2011), but the benefits of merging are not limited to these examples. Acquisitions can be used to manage or mitigate various business risks. For instance, M&A deals can help the target company raise more capital using the acquirer's resources, or help the acquiring firm penetrate new markets. I investigate if firms are strategically using M&A deals to manage their risk exposure as disclosed on their 10-K forms filled with the SEC.

I use state-of-the-art topic modeling tools to assess firms' exposure to each risk topic. This new methods enables me to assign each risk disclosed in the item 1.A section to a topic with greater precision than a method based on word count. With this new approach, I am able to preserve the order in which risks are ranked. Investor guidance published by the SEC suggests that the order of risks is informative. Indeed, I find ample evidence supporting this assertion. By creating a measure of risk factor exposure that takes risk ranking into account, I show that increases in corporate exposure to financing risks are positively related to the probability that the firm will be targeted by an acquirer within 12 months of the filing date. Putted into context, the unconditional takeover probability increases by 34% if the filing firm adds financial risks to the first place of its mandatory risk disclosures. In contrast, I find that financial risks and cash flow uncertainty have a negative and positive impact, respectively, on the likelihood of acquiring another company. By investigating the link between corporate disclosures and deal objectives, I confirm that financially constrained firms do indeed merge to raise more capital or pay down debt. When the acquirer is concerned about cash-flow volatility, these concerns are typically addressed during private merger negotiations. The findings therefore suggest that companies merge to manage their business risks.

Having shown that companies mitigate some of their business risks by engaging in mergers and acquisitions. A natural follow-up question is whether the concentration of risks at the industry level affects the occurrence of merger waves. If companies are increasingly exposed to similar risks, the risk diversification benefits of merging decrease. As a result, companies' incentives to merge are also likely to diminish, resulting, all other things being equal, in fewer deals being announced. Using a similarity measure from the textual analysis literature, I show that this is indeed the case. As the risk factor sections of companies from the same industry-year become increasingly similar, the number of future transactions in that industry decreases. To the best of my knowledge, this finding is new. Past literature shows that variation in aggregate valuation or industry shocks can lead to merger waves (Harford, 2005; Maksimovic and Phillips, 2001; Rhodes-Kropf and Viswanathan, 2004). However, no other papers studies the impact of risk factor exposure similarity on future M&A activity.

In line with previous literature, my results show that financing issues constitute a strong motive for firms to merge. The idea being that the firms with better access to financing can help its merger counterpart securing better financing terms (Cornaggia and Li, 2019). Unlike (Cornaggia and Li, 2019) who study how acquirer purchase target with good access to bank lending supply, I document that financially constrained firms are more likely to be acquired in the near future and less likely to make acquisitions. Masulis and Simsir (2018) shows that financially constrained targets are more likely to initiate negotiations, but they do not establish any relationship with future merger activity. In addition, I document that cash flow uncertainty can also incentivize firms to acquire other firms (Garfinkel and Hankins, 2011). Lastly, I do not find strong empirical support for recurrent losses, R&D, competition and growth risks to be important drivers of future M&A activity unlike past literature (Phillips and Zhdanov, 2013; Flannery et al., 2023; Humphery-Jenner, Powell, and Zhang, 2019). Overall, I contribute to the literature examining why firms merge by showing that business risk factors are intuitive and effective predictors of future corporate mergers.

I contribute to the literature analyzing merger waves under a risk management perspective (Ahern and Harford, 2014; Garfinkel and Hankins, 2011). Garfinkel and Hankins (2011) show that merger waves can arise due to industry shocks. If industry shocks are causing cash-flow uncertainty, firms will vertically integrate to reduce their cash-flow volatility. I contribute by showing that cash-flow uncertainty is not the sole business risk that can contribute to merger waves. After all, ongoing industry consolidation can constitute a threat to other businesses by itself in various ways. For instance, companies frequently mention recent mergers and acquisitions among other firms to justify their concerns about their lack of competitiveness or loosing key suppliers or clients. By searching the mandatory risk disclosures for mentions of ongoing industry consolidation to quantify firms' concerns about sector concentration, I find that companies mentioning industry consolidation in their description of risk factors are more likely to be acquired in the future. As explained, firms perceive industry consolidation as a threat. Therefore, they decide to merge to preserve their competitive position or existing business relationships. These findings help explaining why mergers can cluster in time. Merger waves function as chain reactions: while mergers may reduce business risks for the participating firms, they tend to exacerbate risks for non-participating firms.

Prior evidence suggests that firms disclose risk factors truthfully in 10-K forms. For instance, Hanley and Hoberg (2019) document that bank risk disclosures start to signal greater risks in the housing sector long before the global financial crisis. Similarly, Celeny and Maréchal (2024) show that firms' exposure to cyber-security risk can be observed from corporate disclosures. Long-short portfolios formed on risk disclosures yield positive risk-adjusted returns (Celeny and Maréchal, 2024; Ross, 2019). More generally, the wealth of information readily available in 10-K forms is not timely incorporated by investors, which is creating profitable investment strategies (Cohen, Malloy, and Nguyen, 2020). I contribute by showing that one can use 10-Ks to predict takeover probability. Furthermore, my results emphasize that both content and formatting of 10-Ks contain valuable information. Yet, textual formatting is to my knowledge not exploited in other research papers, despite its strong predictability in my analyses. I show that ignoring changes in risk ordering leads to a severe underestimation of the impact of the 2008 financial crisis and the COVID-19 pandemic on corporate disclosures.

The paper is organized as follows: Section 3.2 presents the different data sources used in this study. Section 3.3 describes how the data is prepared for the empirical analysis. Section 4 contains the results. I provide results on concentration risk in Section 3.5 and Section 3.6 concludes.

3.2. Data

I obtain stock market data from CRSP and accounting data from COMPUSTAT. Following Hanley and Hoberg (2019), I retrieve the 10-K, 10-K405, 10KSB and 10KSB40 forms from EDGAR. Consistent with the literature (Hanley and Hoberg, 2019; Ross, 2019), I do not consider amended filings. The sample starts with the 2004 fiscal year and finishes in 2022. My sample starts one year earlier than Ross (2019); Hanley and Hoberg (2019). This extra year of data is used to train my BERTopic model. The choice of 2004 as the starting date for the sample has several advantages. During this period, most companies began using HTML tags to structure their documents, which facilitates the data collection process. In addition, the sections devoted to risk factors became more standardized during this period (Ross, 2019; Hanley and Hoberg, 2019). I drop trading firms from my sample based on FF industry classification.[1] In addition, information about firms' M&A activity is obtained through SDC. I consider any deals involving at least one US public firm. I impose the deal value to be equal or above $1 million. The filtered deals include partial or total sales of businesses to take into account business restructurings. Deal rumors and share repurchases are filtered out of the sample.

3.3. Creation of Risk Factor Topics

3.3.1. SEC Filings Processing

To obtain the risk factors for each company, I download all 10-K forms published between 2004 and 2023 from EDGAR. For each company, I take into account the first 10-K form published after the end of each of their fiscal years. Risk factors are often discussed under item 1.A of financial statements. Unfortunately, this is not always the case. [2] Furthermore, some firms also disclose a summary of the risk factors section towards the beginning of the 10-K. Therefore, I rely on the degree of textual correspondence with key words ("risk factors", "Item 1.A") and on the plausibility of finding the risk factors at a given location in the financial reports to correctly locate the start of the risk factor section.[3] As described in Maître (2025), I use regular expressions to spot titles in the filings. Next, I truncate the text up to

[1]More precisely, I filter out of my sample firms from industry 48 or with a missing industry classification.

[2]For instance, some firms include risk factors in item 7.

[3]It is generally the case that risk factors summaries are located before the traditional risk factor section.

the title of the next section, which is generally item 1.B. Then, I split the selected text into smaller chunks by exploiting HTML tags and document formatting. In most cases, the risk factors are structured as a series of titles with small paragraphs in between to provide further details on each risk. [4]

The approach outlined above is laborious to automate, but it works well as most reports tend to follow a similar structure. Alternatively, one could use topic modeling tools such as LDA (Blei, Ng, and Jordan, 2003) or dictionary-based methods (Loughran and McDonald, 2011) directly on the full risk factor sections. However, such approach totally ignores the order in which risk factors are listed. As suggested by the SEC, the order is likely informative about the severity of each risk. "'Risk Factors' includes information about the most significant risks that apply to the company or to its securities. Companies generally list the risk factors in order of their importance." (U.S. Securities and Exchange Commission, 2011, p. 2).[5] For this reason, I take into account the rank of each risk when creating my measures of risk exposures. Despite my best efforts to perfectly process each report, errors can remain in my sample. For this reason, I filter out of my sample firm-year with less than 5 risk factors, which represents the $1.2th$ percentile. In addition, I do not consider risks beyond the $40th$ position in each report, as it is uncertain whether these risks are material. Only 16% of 10-Ks disclose more than 40 risks.

In Figure 3.1, I show the average number of risks disclosed by firms through time. Firms disclose in average 19.34 risks in 2005 and 29.22 risks in 2022. With the exception of two fiscal years, the average number of risks listed increased continuously over the sample period. This trend is not limited to the risk factor sections. Cohen, Malloy, and Nguyen (2020) show that 10-K filings have become more detailed with time.

[Insert Figure 3.1 here]

3.3.2. Topic Modeling

To observe how disclosed risk factors relate to future mergers, one needs to reduce the dimensionality of this textual data. Therefore, I identify common risk factors across SEC filings using topic modeling. I define a company as being exposed to a specific risk if this risk was disclosed in their last 10-K publicly available. This

[4] In rare cases, firms do not use headings to structure their report. In this case, I use a variety of approaches to retrieve the main risk names. Depending on document formatting, I can use the first sentence from each new paragraph or use bullet points as risk names and treat the remaining text as part of the risk description, if applicable.

[5] Office of Ivestor Education and Advocacy. (2011, September). *Investor bulletin: How to read a 10-K.* U.S. Securities and Exchange Commission. `https://www.sec.gov/files/reada10k.pdf`

approach has the merit of being simple and intuitive; Each disclosed risk is mapped to a unique topic.

I choose BERTopic (Grootendorst, 2022) to create risk factor topics. This model combines several state-of-the-art machine learning models to create meaningful textual clusters. Unlike LDA (Blei, Ng, and Jordan, 2003), which relies on word counts, BERTopic uses BERT embeddings as input, enabling it to capture the underlying semantics of the text, an especially important advantage when dealing with short texts. Furthermore, BERTopic supports by default soft-clustering, which the model incorporates by adding an additional topic to classify any ambiguous text. Thanks to this feature, the clusters obtained are easier to interpret than when hard clustering models such as K-Means are used. For instance, observations do not have to belong to a cluster if they are overly specific. To avoid any look-ahead bias, I fit the model using the 10-Ks corresponding to the 2004 fiscal year, and I conduct my analyses using the remaining fiscal years. Moreover, I only use the titles of risk factors to create topics; I do not take into account the description of each risk, as the level of detail provided in financial reports varies considerably.

The output is a list of 46 topics. The list is presented in Table 3.1, along with examples of actual risk factors published by companies. Topics are ordered by decreasing frequency. For instance, risks related to product liability and intellectual property are among the most frequently cited risks in 10-K filings after the "Other" topic, which contains texts that the model was unable to classify. The clusters range from industry specific risks, such as uncertainty related to clinical trials, to more general topics such as competition, currency risks, or credit risks. Overall, the different risks are meaningful and are effectively related with their associated example. For instance, topic 17 relates to risks related to energy commodities and effectively matches texts about petroleum prices.

[Insert Table 3.1 here]

Some business weaknesses can be partly mitigated through M&A. Cornaggia and Li (2019) show that financially constrained acquirers create financial synergies and can decrease financing costs by purchasing targets with greater access to bank credit supply. Similarly, newly acquired firms are better able to reach their ideal capital structure after merging (Flannery et al., 2023). In addition, vertical integrations are successfully used by firms to reduce cash-flow uncertainty (Garfinkel and Hankins, 2011). Divestitures can also improve future firm performance by disposing of underperforming firm divisions (Humphery-Jenner, Powell, and Zhang, 2019). Lastly, firms can profit from the M&A market to acquire firms with strong patent portfolio (Phillips and Zhdanov, 2013). Based on these previous findings, I identify six key risks that can be attenuated by business combinations, namely financing risk, recurring losses, worsening competitiveness, business uncertainties, innovation risk and growth challenges. These six merger motives closely align with the five most common deal objectives disclosed by SDC, namely raising financing, business restructuring, target undervaluation, strategic combination and R&D acquisition. Taken together, 95.85% of the M&A deals available on SDC cite at least one of the five motives listed above.[6]

However, not all risk topics created by BERTopic are related to the six key risks that can be mitigated through M&A. For instance, some clusters are industry specific. Therefore, I manually group individual BERTopic risks in one of these six key risks, where applicable. The matching is reported in Table 3.2. Intuitively, I link raising financing through business combination as a way to mitigate risks related to severe indebtedness and financing issues. As divestitures, or corporate restructuring, are taken in response of poor performing units within firms (Brickley and Van Drunen, 1990; Humphery-Jenner, Powell, and Zhang, 2019), I relate this corporate action with topics about recurring operating losses. Mergers driven by competitiveness concerns can be linked to risks about specific parts of the the supply chain. For instance, strategic business combinations are frequently used to attenuate operating risks by strengthening the competitiveness of firms' products (Hoberg and Phillips, 2010). I associate the cash-flow uncertainty risk (Garfinkel and Hankins, 2011) with the topic focused on gross margins and selling prices. I identify patents and risks related to technological changes, as belonging to R&D risks. Lastly, I relate growth challenges to the cluster focused on growth management issues.

Despite my best efforts to relate individual BERTopic risks with the six risks identified above, the resulting mapping is arbitrary by nature. However, I do not consider this arbitrariness to be problematic, given that poor risk grouping will simply result in noisy independent variables. Alternatively, I could use zero-shot topic modeling to directly form the desired group of risks using a set of desired cluster names. In this case, the user only needs to provide strings that are representative of the desired clusters. Risk factors are then assigned to a group based on the similarity of their embeddings with the defined cluster labels. However, this approach also requires the researcher's judgment, as names of desired risk groups must be provided. Furthermore, the success of zero-shot topic modeling

[6]Less common deal motives not mentioned here include mergers and acquisitions made possible by changes in foreign ownership regulations and bids made in response to public takeover bids.

depends on the user's ability to find risk labels that accurately capture all aspects of the desired topics. For this reason, I prefer to use a classic clustering approach, and then group individual topics into broader topics. This method mitigates the risk of omitting certain topics due to judgment errors.

[Insert Table 3.2 here]

3.3.3. Variable Creation

The final empirical challenge is to determine how to measure firms' risk exposures? A naive approach would be to simply use dummy variables to measure which firms are exposed to which risks. However, this approach has several weaknesses. First, it ignores the information contained in the ranking of risk factors. Furthermore, exposure to risk factors are likely strongly correlated with firm and industry characteristics.

The problem arising from ignoring the ranking of risk factors is illustrated in Figure 3.2. This plot shows how the similarity of firms' risk factor sections between successive 10-K filings changes over time. The black line aggregates similarity scores per year. The gray lines represent the aggregated scores per industry-year. In the first Panel of Figure 3.2, the similarity measure is based on the number of risks shared between successive reports, disregarding any information contained in their ordering. When computing similarity, I truncate the number of elements in each report to prevent time-series variation from being driven by changes in disclosure behavior, as shown in Figure 3.1. The resulting time-series is relatively stable, except in 2020 due to the COVID-19 pandemic. In the second graph, I use the Damerau-Levenshtein edit distance measure to account for differences in the composition of risk factors and their order of importance. As this distance measure was initially developed to measure similarity between words, I first encode the risk factor exposure of each firm in a string using all BERTopics topics defined in Table 3.1.[7] When using the Damerau-Levenshtein distance, the aggregate similarity score is much more responsive to the financial crises of 2008 and the COVID-19 pandemic, both at the aggregate and industry levels. This finding suggests that companies are effectively following the SEC's guidelines by disclosing material risks first. At first glance, it is surprising that both time-series were much more affected by the COVID-19 pandemic than by the 2008 financial crisis. However,

[7] To this end, each risk topic is mapped to an unique ASCII character. The characters are then assembled using the same order than in their respective 10-K reports to form a string.

the COVID-19 pandemic is easier to observe, as it affected both the composition of risks and their order. In contrast, the 2008 financial crisis only affected the order of risks, as economic and financial risks were already frequently disclosed by companies before 2008 (Hanley and Hoberg, 2019).

[Insert Figure 3.2 here]

I illustrate the correlation of risk factor citations and industry classification in Figure 3.3. The plot shows the average frequency a risk mitigable through M&A is mentioned per industry. Darker colors indicate a higher frequency. As expected, business risks are strongly related with industry classification. Financing risks are frequently cited by financial firms and by highly leveraged industries such as utilities and mining, while R&D risks are important for pharmaceutical firms and the technological industry. Competition risks, on the other hand, hold roughly the same level of importance across all industries, except for agriculture, utilities and oil.

[Insert Figure 3.3 here]

The two issues raised above can be resolved by calculating a score that takes into account the way risk factor sections are structured, and by taking the first difference of this exposure measure to limit its correlation with firm fixed characteristics. Namely, I compute a score based on the rank of the risk as:

$$Score_{r,i,t} = \begin{cases} \frac{1}{Rank_{r,i,t}} & \text{if } r \in \phi_{i,t} \\ 0 & \text{Otherwise} \end{cases} \tag{3.1}$$

$$\Delta\, Score_{r,i,t} = Score_{r,i,t} - Score_{r,i,t-1} \tag{3.2}$$

Where $Rank_{r,i,t}$ is the rank of a risk r that is mitigable through M&A in the list $\phi_{i,t}$ containing the risks disclosed by firm i in year t. Please note that $\phi_{i,t}$ can include the same risks r at multiple positions if some risks are repeated, or if they belong to the same bigger risk topic defined in Table 3.2. In this case, I use their maximum $Rank_{r,i,t}$ to compute the score. $Score_{r,i,t}$ is in [0,1], and is equal to 1 for the first risk mentioned on the 10-K filing, and quickly decreases for subsequent risks. Non-cited risks get a score of 0. I choose this scoring method, because it puts more emphasis on the first risks, and is unaffected by differences in the number of risks disclosed across companies. Given that companies may be incentivized to over-disclose risks in order to comply with SEC requirements, I find it appropriate to create a score that focuses on the most material risks.

In Table 3.3, I present some descriptive statistics for firms' risk factor exposures and M&A activity in Panel A, SDC merger deal purposes in Panel B and firms' characteristics in Panel C. Among the retained risks, financial, competition, R&D and growth risks are listed in more than half of the filings, with competition issues being frequently cited in the top two places, as shown by its high 75^{th} percentile score. In comparison, recurrent losses and CF uncertainty are rarely cited in 10-Ks. In this paper, the variables $Is\,Target$ and $Is\,Acquirer$ are dummy variables equal to one when the filing company becomes a target or an acquirer in at least one M&A deal announced in the 12 months following the 10-K filing date. In my sample, 6.43% firms are targeted by at least one acquirer during the year. In comparison, 16.78% of firms are interested in acquiring another company each year. Panel B shows deal purpose statistics for announced business combinations. The main motivation behind mergers is strategic combination, which is listed in approximately 83% of deals. Attractive target valuation and interest in targets' patents are also mentioned in a significant number of deals with 22.35% and 20.04% average citation rates, respectively. Lastly, restructuring and raising financing motives are only cited in 4.43% and 3.96% of deals, respectively. I display in Panel C descriptive statistics for firms' characteristics, which are used as control variables in the subsequent tests. I use accounting data from the same reporting period as risk factor data. *Return* is the cumulative six months stock return before filing date. FCF is defined as EBITDA minus tax and interest expenses scaled by total assets. *Leverage* is defined as current and long-term debt scaled by total assets. $Firm\,Size$ is the log-transformed market capitalization as of six months prior to the filing date. $Cash$ is the log-transformation of 1+ cash and short-term investments, and I define $Gross\,Margin$ as the difference between revenues and cogs scaled by revenues.

[Insert Table 3.3 here]

3.4. Do Firms mitigate their cited risk factors?

3.4.1. Analysis at the Firm-level

I first investigate if firms exposed to risks mitigable through M&A are more likely to be (successfully) acquired. As 10-Ks are posted yearly, I compute the dependent variables over the 12 months following the filing date. I display the results of panel regressions in Table 3.4 using firm and year fixed effects and standard errors clustered by firms. We see that $\Delta\, Score\, Fin$ is significantly related with future takeover probability and probability of deal completion across all models (Cornaggia and Li, 2019). Despite accounting for financial leverage and other firm variables in most models, financing risk disclosures remain statistically significant. Consequently, textual data appears to provide valuable information on the severity of the financial risks facing the company, beyond accounting measures. In models (4), (5), (9) and (10), I control for industry-year shocks with fixed effects and change in exposure to financing risks are still significantly related to the dependent variables. In addition, the results are similar when logit models are used instead of OLS models despite differences in sample size; The MLE estimation of logit models require each fixed effect group to have within grouped variations in order to be kept in the estimation sample. The economical impact of $\Delta\, Score\, Fin$ is large. Based on model (5), an one unit change is related with a 34% increase in the unconditional takeover probability over the next twelve months. Changes in scores about recurring losses topic are also positively related with the dependent variables in a significant manner when controls and two-way fixed effects are included. However, I do not observe any significant patterns, once I control for time-varying industry shocks. The regression coefficients of the R&D topic are negatively related with future takeover probability, but the results are not robust to the inclusion of control variables and two-way fixed effects. The remaining risk topics are not statistically significant. Therefore, I conclude that financing risks seem to be the main business risk pushing target firms to seek a buyer.

[Insert Table 3.4 here]

In Table 3.5, I repeat a similar analysis as in Table 3.4. This time, I investigate whether the likelihood that the filing firm will announce the (successful) acquisition of another company within twelve months of the filing date is influenced by changes in its risk factor exposures. As documented above, changes in financial risk exposure are again significantly associated with the dependent variables across all models. In this case, however, the relationship found is negative. The economic impact is less important than for takeover probability. Based on model (5), firms that were not mentioning financial risks in their previous filing, and are now listing this risk in the first place of their latest 10-K report become 19% less likely to try acquire another company in the following 12 months compared to the unconditional probability. Consistent with Garfinkel and Hankins (2011), $\Delta\, Score\ Uncertainty$ is positively related to the outcome variables in most model specifications. An one-unit increase in cash-flow uncertainty is linked with a 25.94% increase relative to the base likelihood that a firm announces the acquisition of another entity. Other risks hedgeable through business combinations are not strongly related with the probability to make acquisitions in the foreseeable future.

[Insert Table 3.5 here]

In Table 3.6, I now test whether changes in risk factor exposures also predict expected returns. It is unclear what is the expected relationship between these variables. On the one hand, predictors of takeover probability should predict positively expected returns, as target's price generally increase on deal announcement date. Similarly, factors influencing firms' decision to make acquisitions should be negatively related with expected returns all else being equal. However, positive changes in risk exposures are likely caused by bad news. For instance, large changes in risk exposures could reveal profound industry shock or poor risk management processes within the filing firm. In this regard, Cohen, Malloy, and Nguyen (2020) show that shorting firms who changed significantly the content of their 10-K compared to their previous filing offers significant positive alpha.

As rises in risk exposures may correlate with other events occurring around the filing period, I measure post-filing returns starting from $t+5$ after the filing date at time t. For the same reason, I drop the last five days of the post-filing period when measuring yearly returns. Overall, I observe negative predictability of returns for all risks, with the exception of growth risks. The negative regression coefficients are generally statistically significant for financial, competition, uncertainty and R&D risks. The found predictability is stronger when expected returns are measured over the next six months compared to when returns are yearly sampled. In this case, only changes in financial or R&D risk exposures remain strong expected return predictors. The results confirm the view that changes in financial statements generally constitute bad news (Cohen, Malloy, and Nguyen, 2020).

[Insert Table 3.6 here]

As I do not establish causality, the predictability of future deals by the financing and uncertainty topics could also be driven by confounding factors. For this reason, I now test whether changes in risk ordered citations also predict the underlying reasons of the announced deals. To this end, I use dummy variables for each deal purpose defined in Panel B of Table 3.3 as dependent variables. The results of the panel regressions are presented in Table 3.7. Please note that the independent variables are measured from the perspective of the target company, and I treat all transactions announced within 12 months of the filing date as separate observations. Panel A reports the results of OLS regressions, whereas Panel B displays the results of Logit regressions. In Panel A, I find that change in ordered risk exposure to financial issues is positively related with the mention of "raising financing" as one of the deal purposes. This result is robust to the inclusion of control variables and to the use of logit regressions in Panel B. This finding validates the hypothesis that firms effectively attempt to solve their financing risks by merging. The results are consistent with those of Cornaggia and Li (2019), who show that financially constrained firms are indeed merging to improve their access to financing. The authors find that those firms are able to reduce their interest expenses after merging with a company that has good access to bank loans. Interestingly, the presence of growth risks in targets' filings is positively correlated with whether or not the acquirer considers the target to be an attractive investment. Similarly, companies with increased exposure to recurrent losses are more likely to be involved in deals listing business restructuring as one of their deal motives. Although these variables are not predictors of future M&A activity, their significant relationships with closely related deal purposes suggest than the risk grouping proposed in Table 3.2 works as intended.

[Insert Table 3.7 here]

I repeat a similar analysis in Table 3.8. The difference being that independent variables are now measured from the acquirer's perspective. As expected, $\Delta\, Score\ Fin$ is not positively associated with raising financing purposes in models (1) and (2) across both panels. This finding confirms that financing problems are preventing companies from making acquisitions. Uncertainty shocks about future cash flows are, however, not linked significantly with any deal purposes. This could either indicate that confounding factors drive the predictability of $\Delta\, Score\ Uncertainty$ in previous tests, or that the deal purpose categories disclosed by SDC are not sufficiently detailed to confirm the predictability of future M&A activity. As in Table 3.7, some changes in risk exposures are linked with their related deal purposes. For instance, firms with increased sensitivity to R&D risks are more likely to cite the acquisition of innovations as a motivation for a transaction when a deal is announced. Despite not predicting future acquisition activity in my sample, this finding suggests that BERTopic clusters capture reasonably well exposure to innovation risks.

[Insert Table 3.8 here]

3.4.2. Merger Negotiations and Risk Disclosures

From the previous analyses, it is not entirely clear whether cash-flow uncertainties relate directly to future mergers and acquisitions activity as the results from deal purposes tests were inconclusive. To assess whether firms strategically hedge their business risks through mergers and acquisitions, one could examine whether the mandatory risk disclosures change following a deal. However, such an analysis may be misleading, as changes in firms' exposure to risk factors tend to be negatively auto-correlated over time.

For this reason, I now test whether there is a link between firms' disclosures and the private negotiations of M&A deals. If mergers are indeed undertook due to specific risks, these risks are likely to be mentioned in the discussions among the target and the acquirer. For instance, Masulis and Simsir (2018) show that financially constrained firms are more likely to initiate deals. I apply the trained BERTopic model to the dataset from Maître (2025) to examine whether firm disclosures are associated with the topics discussed during private negotiations.[8] Specifically, I create dummy variables equal to one if the topics related to financing issues and/or cash-flow uncertainties appear in the negotiation transcripts. I report in Table 3.9 the results of panel regressions. The dependent variables are the two dummy variables described above, and the independent variables are targets' risk disclosures with control variables. I find results consistent with the analyses of deal motives. When the target company is more exposed to financial risks, debt mentions in the private negotiations become more likely. The results are unchanged when logit models are used instead of OLS regressions.[9] Consistent with prior tests, financing risks seem to be an important motive for firms to merge.

[8] As meeting descriptions and risk factors titles have different lengths and wording, I begin by applying a few cleaning steps to the negotiation summaries. I remove people's names and dates, and split longer sentences using the symbols and words “,” “;”, "and" or “or”. These simple pre-processing steps improve the accuracy of the BERTopic model's classification by making the meeting descriptions more similar to the risk factor titles.

[9] The results remain similar when I use the degree of semantic similarity between topic embeddings and meeting transcripts, instead of relying on dummy variables, as the dependent variables.

[Insert Table 3.9 here]

I repeat a similar test using acquirers' mandatory risk disclosures in Table 3.10. I document a positive link between uncertainty risk exposure and mentions of uncertainty during the negotiations. The relationship is statistically significant in 3 out of 4 model specifications. This evidence seems to confirm that uncertainty risks are indeed incentivizing firms to merge.

[Insert Table 3.10 here]

3.4.3. Analysis at the Industry-level

I show in the preceding section that firms do hedge some of their business risks through business combinations to some extent. However, if firms are indeed using M&A to hedge their main risk factors, aggregate risk exposure similarity should impact the future number of deals announced. My rationale is that if firms' disclosed risks become increasingly concentrated in a few risks, the incentives to merge are decreasing, as the diversification benefits of merging are decreasing. Therefore, I posit that increases in risk factor similarity within an industry should predict negatively future M&A activity in that industry. To measure risk factor similarity at the industry-year level, I take the average of the Damerau-Levenshtein edit distance for every firm pair in that industry-year. Having defined an industry-year similarity measure, I then take the simple difference of this measure to capture changes in aggregate risk factor exposure similarity. I also create various control variables at the industry-year level by averaging individual yearly firm data. All variables are measured on July 1 of each year, based on the latest published data available on that date. I test in Table 3.11 with panel regressions if my hypothesis is correct. I document a negative and statistically significant regression coefficient for my variable Δ *RF Similarity* in both model specifications. Hence, this finding confirms my intuition that greater risk concentration in an industry leads to lower business combinations activity in the following year. While changes in industry risk concentration can be induced by industry shocks, the existing literature generally shows a positive correlation between industry shocks and the occurrence of future merger waves (Harford, 2005; Maksimovic and Phillips, 2001), which mitigates concerns that industry shocks explain my results.

[Insert Table 3.11 here]

3.5. Additional Results

3.5.1. Consolidation Risk

The analyses above do not utilize the risk factor descriptions in any way to assess firms risk exposures. However, the additional details contained in these descriptions often explain the underlying causes for which the listed risks are considered material by senior management. For instance, changes in economic conditions andor changes in regulation are frequently mentioned in the risk descriptions. Interestingly, companies also cite trends towards industry consolidation or mergers between competitors to support their concerns about future competitiveness:

> In addition, we may face greater competition due to consolidation of companies in the technology sector, through strategic mergers or acquisition. Consolidation activity may result in new competitors with greater scale, a broader footprint, or offerings that are more attractive than ours. We believe that this competition could have a negative effect on our ability to compete for new work and skilled professionals. (Perficient Incorporated, 2020, p. 10)[10]

Business combinations between firms' suppliers or clients also pose a risk, as critical reliance on key third-parties increases:

> In addition, our customers continue to experience ongoing industry consolidation, particularly in the sports specialty sector. As this consolidation continues, it increases the risk that if any one customer significantly reduces their purchases of our products, we may be unable to find sufficient alternative customers to continue to grow our net revenues, or our net revenues may decline. (Under Armour Incorporated, 2020, p. 11)[11]

One of the way, firms can mitigate this concentration risk is by participating themselves to this consolidation or integration trend. Ahern and Harford (2014) show that merger waves propagate from an industry to another through supply-chain links. Therefore, I posit that the effects of merger waves propagation is directly observable at the firm-level. Namely, firms worrying about the effects of

[10] Perficient Incorporated. (2020, February 25). *Form 10-K.* U.S. Securities and Exchange Commission. `https://www.sec.gov/Archives/edgar/data/1085869/000108586920000010/perficientinc10-k20192.htm`

[11] Under Armour Incorporated. (2020, February 26). *Form 10-K.* U.S. Securities and Exchange Commission. `https://www.sec.gov/Archives/edgar/data/1336917/000133691720000010/ua-20191231.htm`

ongoing industry consolidation on their business activities are more likely to engage in business combinations in the future.

I use a bag-of-words approach to determine whether firms are discussing ongoing merger and acquisition among their peers in their business risk descriptions. Using regular expressions, I spot every possible combination of keywords about industry consolidation.[12] Examples of these concerns about ongoing business integrations are provided in Table 3.A1, which is located in the appendix. The dictionary-based method identifies sentences about increasing industry concentration more accurately than LDA or BERTopic. Since only specific portions of the textual descriptions may pertain to consolidation risk, applying clustering models to the full texts is likely to yield a high number of false negatives. After having scanned each disclosed risk, I create exposure variables using Equation 3.1 and Equation 3.2.

Consistent with the two examples listed above, consolidation concerns are generally expressed alongside issues of competitiveness, excessive dependence on key customers, supply chain problems and uncertainty about future cash-flows. In Table 3.12, I report the results of panel regressions of future M&A events on $\Delta\,Score\ Consolidation$. This variable measures the evolution of companies' concerns regarding consolidation between each of their 10-K forms. In Panel A, I find that variations in exposure to consolidation risk is positively and significantly related with the likelihood to be (successfully) acquired. Consistent with Ahern and Harford (2014), I document that merger waves do propagate within an industry and through the supply-chain. However, the results show that this effect also holds at the firm-level and not just at the industry-level as previously found. I test the same relationship in Panel B, where dependent variables captures the future acquiring activity of the filing firm. I do not document any strong statistical relationships between the dependent variables and the key independent variable of interest. The results obtained also align with the results of Garfinkel and Hankins (2011), who show that the desire of firms to mitigate their cash-flow uncertainty trough vertical integrations can lead to merger waves. However, my findings suggest that uncertainty about cash-flows is not the only factor influencing merger waves. Indeed, firms mention ongoing consolidation when discussing a wide range of different risks. My results instead explain why mergers are concentrated in time rather than why merger waves start. The findings suggest that when companies merge, other market players become increasingly concerned about the effects of these mergers on their businesses and therefore decide to merge themselves.

[12] Among other, I match texts such as "consolidation of competitors", "if our suppliers merge" or "increase in mergers and acquisitions activity". In addition, I take care of removing matches where the filing firm is mentioning its own recent business combinations.

[Insert Table 3.12 here]

3.6. Conclusion

In this paper, I investigate whether firms do engage in business combinations to mitigate the business risks mentioned in their 10-K mandatory risk disclosures. I document that firms that are becoming more exposed to financing risks are more likely to seek an acquirer within the next 12 months after the 10-K's filing date. In addition, raising financing is more likely to be cited among the factors motivating the deal when target is severely exposed to financial risks. In contrast, I document that firms facing rising financing risks are less likely to acquire other firms. Companies with increasing exposure to cash flow uncertainty risk are more likely to acquire other firms. Using the mandatory risk disclosures, I am able to identify the propagation of merger waves directly at the firm level. Companies that are concerned about the impact of industry consolidation on their business are more likely to be acquired. Lastly, I document that changes in aggregate industry risk similarity predict future M&A activity negatively. In other words, when the risks mentioned by firms in the same industry become more similar, the number of future business combinations in that industry drops. This is line with the idea that firms merge partly to mitigate their business risks. As business risks become more similar, the benefits to merge are decreasing. Overall, the findings suggest that accounting for risk factor ordering is important; The way risks are ordered is informative in itself. Furthermore, the results support the hypothesis that firms utilize the M&A market to hedge some of their business risks.

References

Ahern, K. R., and J. Harford. 2014. The importance of industry links in merger waves. *Journal of Finance* 69:527–76.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3:993–1022.

Brickley, J. A., and L. D. Van Drunen. 1990. Internal corporate restructuring: An empirical analysis. *Journal of Accounting and Economics* 12:251–80.

Celeny, D., and L. Maréchal. 2024. Cyber risk and the cross-section of stock returns. Unpublished Working Paper.

Cohen, L., C. Malloy, and Q. Nguyen. 2020. Lazy prices. *Journal of Finance* 75:1371–415.

Cornaggia, J., and J. Y. Li. 2019. The value of access to finance: Evidence from m&as. *Journal of Financial Economics* 131:232–50.

Cremers, K. M., V. B. Nair, and K. John. 2009. Takeovers and the cross-section of returns. *Review of Financial Studies* 22:1409–45.

Flannery, M. J., J. Hanousek, A. Shamshur, and J. Tresl. 2023. M&a activity and the capital structure of target firms. *Journal of Financial and Quantitative Analysis* 58:2064–95.

Garfinkel, J. A., and K. W. Hankins. 2011. The role of risk management in mergers and merger waves. *Journal of Financial Economics* 101:515–32.

Grootendorst, M. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. Unpublished Working Paper.

Hanley, K. W., and G. Hoberg. 2019. Dynamic interpretation of emerging risks in the financial sector. *Review of Financial Studies* 32:4543–603.

Harford, J. 2005. What drives merger waves? *Journal of Financial Economics* 77:529–60.

Hoberg, G., and G. Phillips. 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies* 23:3773–811.

Humphery-Jenner, M., R. Powell, and E. J. Zhang. 2019. Practice makes progress: Evidence from divestitures. *Journal of Banking & Finance* 105:1–19.

Loughran, T., and B. McDonald. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance* 66:35–65.

Maître, A. T. 2025. Effects of financial advisors on deal outcomes: Evidence from sec filings. Unpublished Working Paper.

Maksimovic, V., and G. Phillips. 2001. The market for corporate assets: Who engages in mergers and asset sales and are there efficiency gains? *Journal of Finance* 56:2019–65.

Masulis, R. W., and S. A. Simsir. 2018. Deal initiation in mergers and acquisitions. *Journal of Financial and Quantitative Analysis* 53:2389–430.

Phillips, G. M., and A. Zhdanov. 2013. R&d and the incentives from merger and acquisition activity. *Review of Financial Studies* 26:34–78.

Rhodes-Kropf, M., and S. Viswanathan. 2004. Market valuation and merger waves. *Journal of Finance* 59:2685–718.

Ross, L. 2019. Are item 1a risk factors priced? Unpublished Working Paper.

Figures

Figure 3.1: **Number of Risk Factors disclosed Through Time**

This figure shows the average number of risk factors disclosed by firms for each fiscal year. The number of risks in firms 10-Ks are truncated between 5 and 40 to alleviate the impact of outliers and data collection errors.

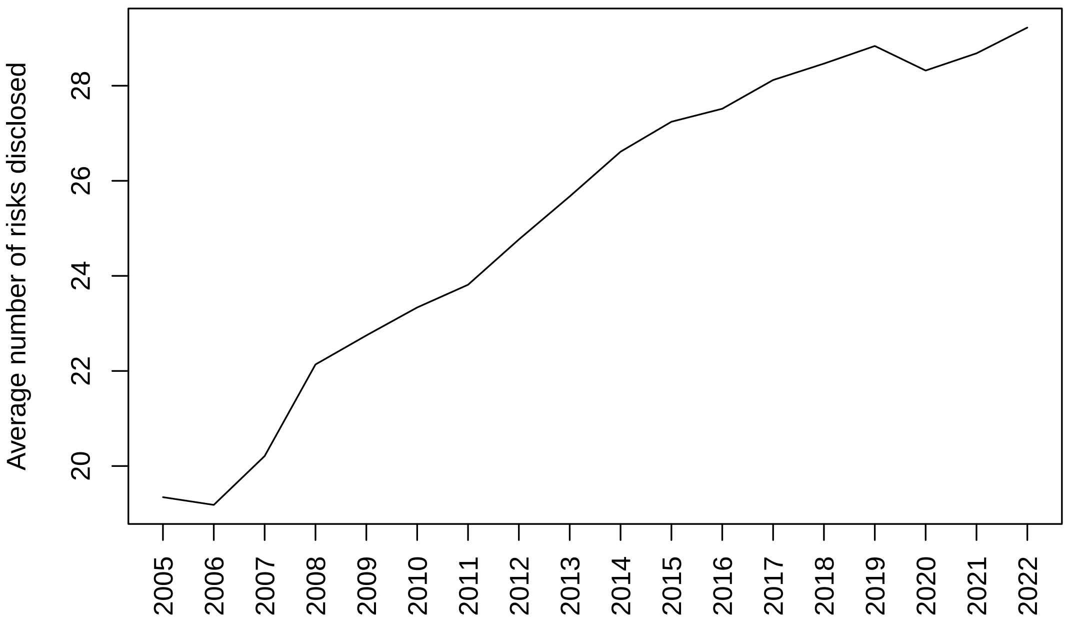


Figure 3.2: **Risk Factor Similarity Through Time**

This figure shows the similarity between the sections on risk factors disclosed by companies from one year to the next. The black line show the yearly average of similarity scores, while the gray lines display the average similarity score per industry and year. In the first picture, I use the proportion of risks that are shared by the current and the previous 10-K to measure risk exposure similarity. In the second picture, I use the Damerau-Levenshtein edit distance to take into account differences in risk composition, as well as differences in risk factor ordering.

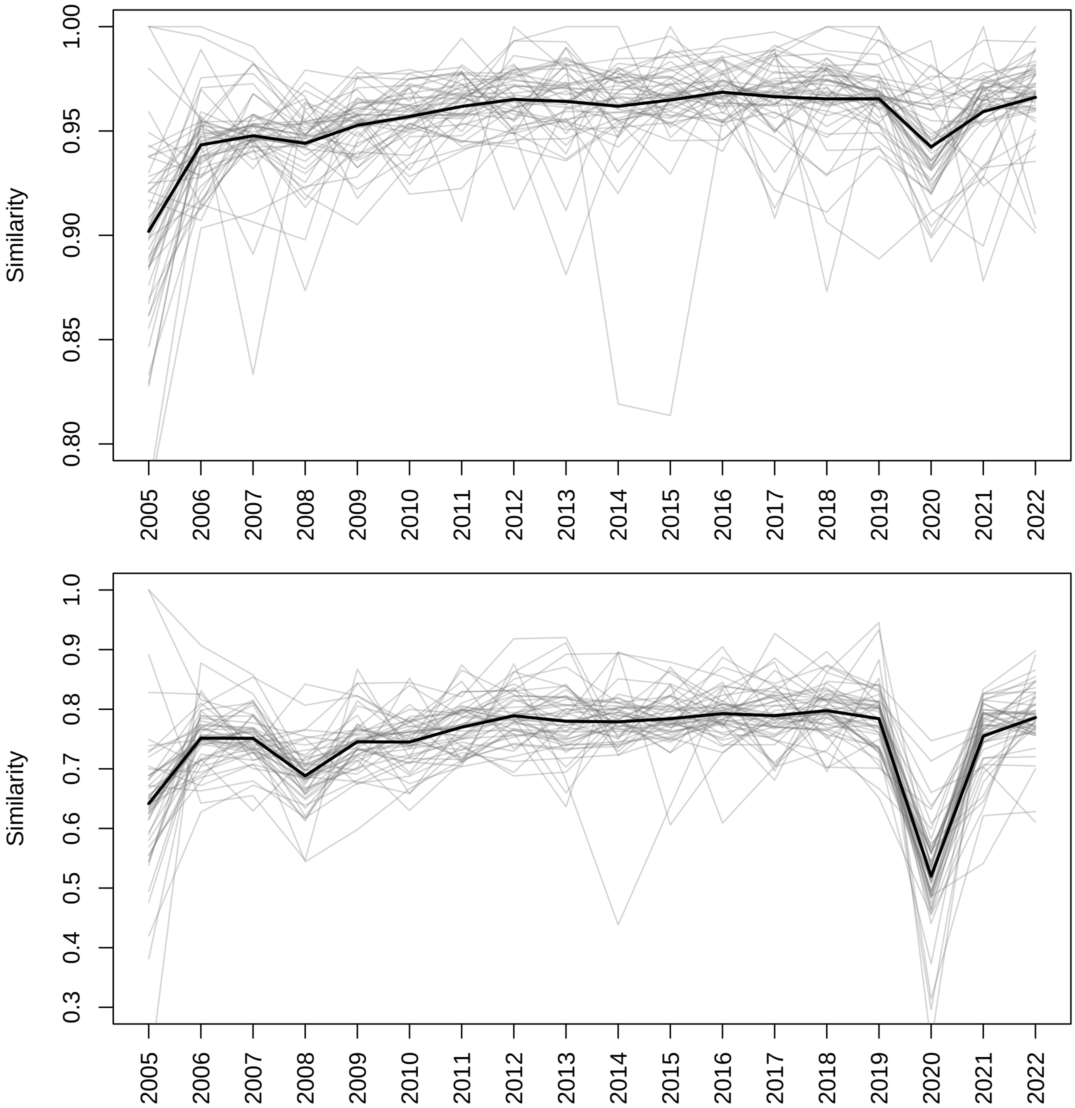
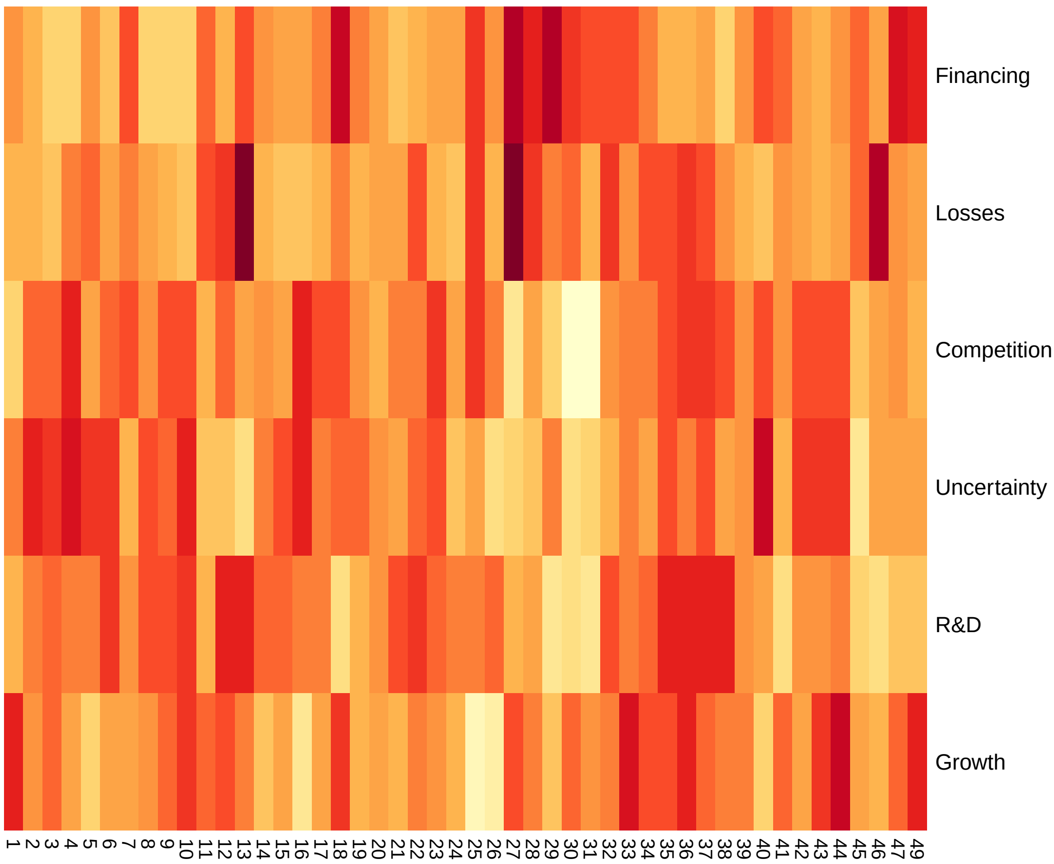


Figure 3.3: **Risk Factor Prevalence per Industry**

This figure shows the frequency with which certain selected risk factors are cited in each industry. The plotted topics are defined in Table 3.2. Industry definitions follow FF classification. Brighter (darker) colors indicate low (high) risk prevalence in the corresponding industry, and frequencies are scaled by topic.



Tables

Table 3.1: **Risk Factor Topics**

This table contains the resulting topics from BERTopic along with an example of risk factor per topic. The model is fitted over the 10-Ks corresponding to the 2004 fiscal year. Risk topics are ordered by frequency in decreasing order. The topic "Other" contains all disclosed risks that could not be classified in a risk topic. To enhance readability, the source corresponding to each example is listed in Table 3.A2.

	Topic	*Examples*
1	*Other*	industry capacity may adversely affect our business
2	*Product liability & Insurance coverage*	our business is subject to risks of litigation
3	*Intellectual property & Proprietary rights*	our intellectual property rights are subject to risks, including
4	*Environmental laws & Government regulation*	government regulations could adversely affect our business
5	*Key personnel & Senior management*	we may be unable to attract and retain key personnel
6	*Quarterly operating & Sales cycle*	our quarterly and annual operating results will likely fluctuate
7	*New products & Rapid technological*	rapid technological change may render our products obsolete or non-competitive
8	*Clinical trials & Regulatory approval*	our medicines could be subject to regulatory limitations following approval
9	*Future acquisitions & Successfully integrate*	acquisitions may not be successful
10	*Highly competitive & Intense competition*	we operate in competitive markets
11	*International operations & Risks associated*	risks associated with the company's global operations
12	*Effectively manage & Growth strategy*	our management may be unable to manage rapid growth effectively
13	*Credit facility & Substantial indebtedness*	we have indebtedness which could adversely affect our financial position
14	*Additional capital & Financing*	we may have problems financing our future growth
15	*Source suppliers & Key components*	dependence on third party vendors
16	*Anti takeover & Delaware law*	anti-takeover provisions may discourage a change of our control
17	*Natural gas & Oil*	our profitability could be adversely affected by high petroleum prices
18	*Operating losses & Maintain profitability*	we are generating negligible revenue and expect future losses
19	*Principal stockholders & Significant influence*	control by existing stockholder
20	*Internal controls & Financial reporting*	our internal control over financial reporting may need enhancement
21	*Common stock & Market price*	the market price for our common shares is particularly volatile
22	*Party payors & Health care*	changes in healthcare policy could adversely affect our business
23	*Economic conditions & Semiconductor industry*	we may particularly be affected by general economic conditions
24	*Key customers & Small number*	loss of any additional personnel could adversely affect our business
25	*Income tax & Taxes*	we may have additional tax liabilities
26	*Future sales & Common stock*	future sales of our common stock may cause dilution
27	*Telecommunications industry & Continued growth*	changes in telecommunications regulation and tariffs could harm our business
28	*Fixed price & Government contracts*	we may lose contracts through competitive bidding or early termination
29	*Currency exchange & Rate fluctuations*	currency fluctuations may negatively affect our results of operations
30	*Risk factors & Following*	risks related to commercial, regulatory and other business matters
31	*Accounting standards & Reported results*	changes in accounting standards, policies and practices

Table continued on next page

Table continued

	Topic	*Examples*
32	*License* & *Licensed* & *Licensing*	our licensees' conduct could harm our business
33	*Information systems* & *Failures*	business interruptions could adversely affect our operations
34	*Products contain* & *Market acceptance*	defects in our solutions could harm our reputation and business
35	*Pay dividends* & *Foreseeable future*	we may never pay any dividends to shareholders
36	*Terrorist attacks* & *Adversely impact*	terrorist attacks or national disasters may adversely affect our business
37	*Significant portion* & *Small number*	the company's sales and dependence on major customers
38	*Natural disasters* & *Catastrophic*	a natural disaster could harm our business
39	*Credit risk* & *Real estate*	we face other risks related to the current credit crisis
40	*Rate fluctuations* & *Increases*	changes in interest rates would affect our profitability
41	*Selling prices* & *Gross margins*	our revenue and our revenue growth rate may decline
42	*Intangible assets* & *Goodwill*	potential for goodwill impairment
43	*Stock options* & *Accounting rules*	changes in accounting standards for stock option plans
44	*Stock price* & *Highly volatile*	we may experience volatility in our stock price
45	*Security* & *Breaches* & *Confidential*	security and privacy breaches may harm our business
46	*Distribution channels* & *Distributors*	our revenues are dependent on selling through distributors

Table 3.2: **Risk Factor Topics Grouping**

This table summarizes how BERTopic clusters are matched with the risks that can be mitigated through business combinations.

	Grouped Topics / RF Topics	*Financing*	*Losses*	*Competition*	*Uncertainty*	*R&D*	*Growth*
3	*Intellectual property & Proprietary rights*					✓	
7	*New products & Rapid technological*					✓	
10	*Highly competitive & Intense competition*			✓			
12	*Effectively manage & Growth strategy*						✓
13	*Credit facility & Substantial indebtedness*	✓					
14	*Additional capital & Financing*	✓					
15	*Source suppliers & Key components*			✓			
18	*Operating losses & Maintain profitability*		✓				
24	*Key customers & Small number*			✓			
37	*Significant portion & Small number*			✓			
41	*Selling prices & Gross margins*				✓		
46	*Distribution channels & Distributors*			✓			

Table 3.3: **Summary Statistics**

This table contains summary statistics for risk factor exposures, merger related variables and control variables. All coefficients in Panel A and Panel B are reported in percentage points, otherwise statistics are reported in decimal points.

			Risk Factor Exposure		
Panel A	*Mean*	25^{th}	50^{th}	75^{th}	*SD*
Score Fin	16.71	0	7.14	20	25.01
Score Loss	10.35	0	0	3.57	26.07
Score Competition	30.28	9.09	20	50	30.44
Score Uncertainty	8.83	0	0	8.33	20
Score R&D	13.34	0	6.25	16.67	21.73
Score Growth	12.61	0	3.85	12.5	23.06
Δ *Score Fin*	0.15	−0.13	0	0	16.33
Δ *Score Loss*	−0.48	0	0	0	13.94
Δ *Score Competition*	−0.52	−0.33	0	0	17.99
Δ *Score Uncertainty*	−0.03	0	0	0	11.97
Δ *Score R&D*	−0.22	0	0	0	12.6
Δ *Score Growth*	−0.28	0	0	0	15.28
Is Target	6.43	0	0	0	24.53
Is Acquirer	16.78	0	0	0	37.37
Panel B			*Deal Purposes*		
Raise Financing	3.96	0	0	0	19.51
Business Restructuring	4.43	0	0	0	20.59
Attractive Investment	22.35	0	0	0	41.66
Strategic Combination	82.99	100	100	100	37.57
Acquire Innovation	20.04	0	0	0	40.03
Panel C			*Control Variables*		
Return	0.05	−0.15	0.03	0.21	0.53
FCF	−0.01	0.01	0.05	0.1	0.45
Leverage	0.22	0.04	0.18	0.35	0.21
Firm Size	13.45	11.96	13.42	14.87	2.09
Cash	4.42	3.04	4.39	5.74	2.03
Gross Margin	0.42	0.24	0.39	0.61	0.26

Table 3.4: **Future M&A activity and Risk Factor Disclosures**

This table contains the results of panel regressions. The dependent variables are two dummy variables equals to one if the filing company will receive an offer to be acquired by another company within the next twelve months following the filing date and if the deal is successful. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per firm. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable:									
	Is Target					*Is Completed*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Δ *Score Fin*	2.91***	2.17**	43.89**	2.06**	29.52**	2.65***	1.82**	46.36**	1.5*	25.94*
	(3.47)	(2.36)	(2.53)	(2.18)	(2.15)	(3.42)	(2.16)	(2.36)	(1.72)	(1.7)
Δ *Score Loss*	2.57**	2.53*	47.34*	0.56	7.38	2.45**	2.86**	62.8**	0.49	7.53
	(2.23)	(1.92)	(1.95)	(0.43)	(0.4)	(2.3)	(2.35)	(2.34)	(0.4)	(0.38)
Δ *Score Competition*	0.65	0.45	14.88	−0.01	0.11	1.05*	0.84	29.07*	0.37	7.32
	(1)	(0.67)	(1.04)	(−0.01)	(0.01)	(1.74)	(1.34)	(1.79)	(0.56)	(0.58)
Δ *Score Uncertainty*	1.81	2.07*	43.88	0.79	13.05	0.93	1.03	27.11	−0.28	−5.89
	(1.51)	(1.68)	(1.6)	(0.61)	(0.61)	(0.82)	(0.89)	(0.89)	(−0.23)	(−0.25)
Δ *Score R&D*	−2.55**	−1.73	−16.45	−2.32*	−35.44*	−2.46**	−1.72	−12.51	−2.22*	−40.03*
	(−2.25)	(−1.43)	(−0.69)	(−1.84)	(−1.8)	(−2.36)	(−1.55)	(−0.47)	(−1.89)	(−1.87)
Δ *Score Growth*	0.09	0.44	8.45	−1.27	−18.66	0.37	0.91	26.31	−0.82	−13.75
	(0.09)	(0.43)	(0.4)	(−1.2)	(−1.16)	(0.4)	(0.94)	(1.09)	(−0.81)	(−0.75)
Return		−2.86***	−68.27***	−2.75***	−45.23***		−2.43***	−81.81***	−2.22***	−43.82***
		(−7.22)	(−7.37)	(−6.74)	(−6.44)		(−6.77)	(−7.63)	(−5.96)	(−5.71)
FCF		2.52*	60.25*	−0.4	−0.75		2.11*	66.86	−0.1	3.16
		(1.72)	(1.65)	(−0.4)	(−0.05)		(1.65)	(1.63)	(−0.12)	(0.22)
Leverage		3.62***	84.92***	5.02***	77.03***		3.16***	112.5***	4.18***	75.96***
		(3.16)	(2.8)	(7.33)	(7.66)		(2.98)	(2.94)	(6.83)	(7.14)
Firm Size		−1.16***	−24.16***	−0.87***	−14.69***		−0.99***	−31.85***	−0.61***	−12.12***
		(−5.2)	(−4.1)	(−7.91)	(−7.95)		(−4.91)	(−4.49)	(−6.25)	(−6.18)
Cash		−0.36**	−10.3**	0.35***	5.78***		−0.39**	−15.52***	0.18*	3.39*
		(−2.08)	(−2.3)	(3.24)	(3.12)		(−2.43)	(−2.91)	(1.88)	(1.73)
Gross Margin		−0.56	−44.36	1.18*	18.43*		−0.11	−66.48**	1.7***	31.94***
		(−0.46)	(−1.61)	(1.81)	(1.8)		(−0.1)	(−2.06)	(2.9)	(2.93)
Logit	No	No	Yes	No	Yes	No	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Industry X Year FE	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Observations	51,911	47,852	19,228	47,852	44,188	51,911	47,852	16,512	47,852	43,359

Table 3.5: **Future M&A activity and Risk Factor Disclosures**

This table contains the results of panel regressions. Dependent variables are two dummy variables equals to one if the filing company will make an offer to acquire another company within the next twelve months and if the deal is successful. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per firm. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable:									
	Is Acquirer					*Is Completed*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Δ *Score Fin*	−2.16**	−2.08*	−27.58**	−2.09*	−21.62**	−2.14**	−1.67	−23.66*	−1.63	−18.82*
	(−2.14)	(−1.83)	(−2.21)	(−1.84)	(−2.16)	(−2.26)	(−1.54)	(−1.87)	(−1.51)	(−1.82)
Δ *Score Loss*	0.52	1.47	24.78	1.4	14.08	0.69	1.15	23.07	1.28	14.02
	(0.39)	(0.94)	(1.41)	(0.92)	(1.04)	(0.55)	(0.77)	(1.29)	(0.89)	(1.01)
Δ *Score Competition*	0.88	0.31	4.06	0.68	4.98	0.81	0.23	3.93	0.67	5.39
	(0.95)	(0.32)	(0.46)	(0.71)	(0.72)	(0.91)	(0.25)	(0.43)	(0.73)	(0.75)
Δ *Score Uncertainty*	3.02*	2.85	28.14*	2.66	23.07*	3.33**	3.23*	34.04**	2.83*	26.29**
	(1.77)	(1.62)	(1.73)	(1.54)	(1.8)	(2.01)	(1.88)	(2.01)	(1.69)	(1.98)
Δ *Score R&D*	2.26	2.28	22.21	2.29	15.92	1.85	1.43	15.96	1.66	12.5
	(1.43)	(1.28)	(1.41)	(1.34)	(1.29)	(1.24)	(0.85)	(1.01)	(1.01)	(1)
Δ *Score Growth*	0.1	−0.43	−4.17	0.88	6.77	0.38	−0.05	−0.21	1.14	9.19
	(0.08)	(−0.31)	(−0.32)	(0.64)	(0.64)	(0.3)	(−0.04)	(−0.02)	(0.89)	(0.85)
Return		4.95***	56.09***	5.71***	50.35***		4.58***	56.21***	5.45***	52.16***
		(8.84)	(9.84)	(10.48)	(11.77)		(8.53)	(9.56)	(10.47)	(11.88)
FCF		3.82**	93.91***	6.02***	74.79***		4.31**	106.3***	6.07***	79.89***
		(2.06)	(3.23)	(4.45)	(4.85)		(2.37)	(3.41)	(4.74)	(5.01)
Leverage		−19.2***	−176.6***	1.14	8.55		−19.15***	−189.6***	0.64	4.79
		(−10.5)	(−9.52)	(0.88)	(0.85)		(−10.91)	(−9.82)	(0.53)	(0.46)
Firm Size		0.87**	13.98***	3.74***	28.83***		0.54	11.76***	3.55***	29.59***
		(2.51)	(3.99)	(16.58)	(16.6)		(1.63)	(3.26)	(16.72)	(16.8)
Cash		1.08***	10.14***	−0.62***	−5.8***		0.94***	9.72***	−0.7***	−7***
		(4.04)	(4.16)	(−2.8)	(−3.52)		(3.63)	(3.85)	(−3.36)	(−4.17)
Gross Margin		9.29***	95.98***	5.57***	42.4***		9.01***	105.2***	5.24***	42.93***
		(5.63)	(4.68)	(4.31)	(4.16)		(5.52)	(4.63)	(4.29)	(4.11)
Logit	No	No	Yes	No	Yes	No	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Industry X Year FE	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Observations	51,911	47,852	30,750	47,852	47,339	51,911	47,852	29,718	47,852	47,233

Table 3.6: **Future expected returns and Risk Factor Disclosures**

This table contains the results of panel regressions. Dependent variables are yearly excess returns and abnormal returns over the twelve months following filing date. Abnormal returns are computed by deducting returns attributable to market exposure which is measured using 100 observations before filing date. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per firm. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable:					
	$Post\,Filing\,Return^{6\,months}$			$Post\,Filing\,Return^{1\,year}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta\,Score\,Fin$	−1.29	−2.78**	−3.01**	−2.31	−6***	−5.61***
	(−1.09)	(−2.28)	(−2.54)	(−1.42)	(−3.59)	(−3.41)
$\Delta\,Score\,Loss$	0.27	−1.24	−1.54	1.42	−0.54	0.41
	(0.16)	(−0.69)	(−0.92)	(0.59)	(−0.22)	(0.18)
$\Delta\,Score\,Competition$	−2.85***	−2.48***	−2.28***	−2.76**	−1.64	−1.96*
	(−3.24)	(−2.89)	(−2.7)	(−2.25)	(−1.4)	(−1.67)
$\Delta\,Score\,Uncertainty$	−3.54**	−3.15**	−2.92*	−1.62	−1	−0.67
	(−2.3)	(−2.06)	(−1.94)	(−0.74)	(−0.48)	(−0.32)
$\Delta\,Score\,R\&D$	−5.5***	−4.2**	−3.85**	−7.49***	−4.38*	−5.04**
	(−3.17)	(−2.46)	(−2.3)	(−3.17)	(−1.92)	(−2.2)
$\Delta\,Score\,Growth$	−1.3	−1.09	−0.75	0.01	0.89	0.35
	(−1.04)	(−0.88)	(−0.6)	(0.01)	(0.53)	(0.2)
$Return$		−9***	−6.07***		−22.11***	−7.72***
		(−14.08)	(−10.07)		(−25.46)	(−9.49)
FCF		3.21	14.99***		2.71	23.36***
		(1.39)	(10.44)		(0.82)	(11.71)
$Leverage$		1.61	1.33		−2.66	2.79**
		(0.93)	(1.61)		(−0.98)	(2.35)
$Firm\,Size$		−9.56***	0.39***		−18.15***	0.07
		(−26.05)	(2.98)		(−33.02)	(0.36)
$Cash$		−0.45*	−0.24*		−0.41	0.32*
		(−1.94)	(−1.94)		(−1.21)	(1.84)
$Gross\,Margin$		8.88***	1		17.15***	1.74
		(5.11)	(1.31)		(6.18)	(1.59)
$Firm\,FE$	Yes	Yes	No	Yes	Yes	No
$Year\,FE$	Yes	Yes	No	Yes	Yes	No
$Industry\ X\ Year\,FE$	No	No	Yes	No	No	Yes
Observations	51,362	47,821	47,821	51,362	47,821	47,821

Table 3.7: **Deal Purpose and Target Risk Factor Disclosures**

This table contains the results of panel regressions. Dependent variables are deal purpose categories. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per industry. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Target's point of view									
	Raise Financing		*Business Restructuring*		*Attractive Investment*		*Strategic Combination*		*Acquire Innovation*	
Panel A: OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Δ *Score Fin*	16.01***	12.65***	1.1	−0.22	−0.76	2.36	−2.5	−3.38	0.34	1.9
	(4.01)	(2.77)	(0.32)	(−0.06)	(−0.14)	(0.41)	(−0.53)	(−0.57)	(0.1)	(0.49)
Δ *Score Loss*	−4.08	−1.37	−2.13	3.26*	16.2***	6.96	10.63	8.48	1.43	0.83
	(−0.68)	(−0.27)	(−0.61)	(1.68)	(2.72)	(1.3)	(1.52)	(1.32)	(0.23)	(0.16)
Δ *Score Competition*	−0.13	−0.53	0.96	0.69	5.88	4.68	2.06	1.18	−3.35	−2.95
	(−0.04)	(−0.17)	(0.29)	(0.23)	(1.22)	(0.95)	(0.4)	(0.23)	(−1.07)	(−0.96)
Δ *Score Uncertainty*	−3.25	−3.61	−1.77	−3.22	−11.51	−11.39	10.14	8.94	−5.51	−2.88
	(−0.74)	(−0.83)	(−0.52)	(−0.93)	(−1.26)	(−1.19)	(1.08)	(0.99)	(−0.84)	(−0.46)
Δ *Score R&D*	−0.65	−0.49	−5.31	−1.46	6.68	4.66	15.55	8.89	0.35	−0.62
	(−0.14)	(−0.1)	(−1.27)	(−0.39)	(1.22)	(0.91)	(1.53)	(0.95)	(0.08)	(−0.11)
Δ *Score Growth*	0.49	−0.43	−2.94	−2.43	14.64**	16.8***	1.05	1.77	−2.06	−2.18
	(0.17)	(−0.14)	(−1.25)	(−1.14)	(2.56)	(2.59)	(0.17)	(0.27)	(−0.55)	(−0.56)
Return		−4.79**		−5.4***		2.74		9.55***		3.79**
		(−2.26)		(−2.99)		(0.86)		(4.03)		(2.15)
FCF		−5.87		2.05		14.99**		−3.47		−5.25
		(−1.37)		(0.66)		(2.04)		(−0.48)		(−0.87)
Leverage		14.9***		13.92***		0.13		−11.11*		−6.99**
		(4.62)		(3.79)		(0.02)		(−1.86)		(−2.12)
Firm Size		−1.49**		−0.61		2.99***		3.31***		0.8
		(−2.49)		(−0.97)		(3.03)		(3.66)		(1.11)
Cash		1.57***		0.78*		0.71		−2.76***		0.09
		(2.8)		(1.72)		(0.85)		(−3.85)		(0.14)
Gross Margin		−7.76**		1.09		2.22		4.79		4.73
		(−2.25)		(0.44)		(0.44)		(0.79)		(1.23)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,779	2,616	2,779	2,616	2,779	2,616	2,779	2,616	2,779	2,616
Panel B: Logit	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Δ *Score Fin*	169.3***	109.8**	27.61	−5.52	−4.19	11.03	−11.94	−16.14	−1.55	13.3
	(5.17)	(2.45)	(0.39)	(−0.08)	(−0.16)	(0.37)	(−0.49)	(−0.5)	(−0.05)	(0.38)
Δ *Score Loss*	−83.06	−28.76	−59.77	186.1*	77.77***	36.04	56.69	47.74	4.01	−0.07
	(−0.85)	(−0.29)	(−0.58)	(1.73)	(2.67)	(1.36)	(1.53)	(1.37)	(0.07)	(0)
Δ *Score Competition*	−11.82	−31.14	31.97	16.85	27.22	22.07	13.51	7.98	−28.54	−25.79
	(−0.18)	(−0.51)	(0.33)	(0.2)	(1.21)	(0.95)	(0.49)	(0.28)	(−1.05)	(−0.94)
Δ *Score Uncertainty*	−63.89	−62.26	−37.88	−59.5	−53.56	−53.48	52.62	46.59	−51.26	−29.54
	(−0.92)	(−0.84)	(−0.44)	(−0.59)	(−1.27)	(−1.18)	(1.07)	(0.94)	(−1.05)	(−0.63)
Δ *Score R&D*	−33.45	−24.05	−127.3	−18.76	29.36	20.4	81.68	47.45	−11.5	−20.5
	(−0.4)	(−0.21)	(−1.2)	(−0.17)	(1.15)	(0.85)	(1.57)	(0.92)	(−0.34)	(−0.48)
Δ *Score Growth*	−5.23	−29.83	−91.85	−68.93	69.21**	81.55**	4.57	9.24	−20.62	−20.65
	(−0.1)	(−0.45)	(−1.45)	(−1.11)	(2.49)	(2.46)	(0.13)	(0.25)	(−0.71)	(−0.68)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,601	2,419	2,519	2,370	2,777	2,614	2,765	2,602	2,737	2,582

Table 3.8: **Deal Purpose and Acquirer Risk Factor Disclosures**

This table contains the results of panel regressions. Dependent variables are deal purpose categories. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per industry. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Acquirer's point of view									
	Raise Financing		*Business Restructuring*		*Attractive Investment*		*Strategic Combination*		*Acquire Innovation*	
Panel A: OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Δ *Score Fin*	0.34	0.35	0.42	1.44	0.03	−1.86	−3.9	−3.95	1.36	2.97
	(0.21)	(0.22)	(0.18)	(0.55)	(0.01)	(−0.33)	(−1.18)	(−1.19)	(0.3)	(0.62)
Δ *Score Loss*	4**	3.05*	5.14*	4.62	0.08	−0.67	−6.9	−5.14	6.64	1.36
	(2.25)	(1.67)	(1.66)	(1.36)	(0.01)	(−0.09)	(−1.01)	(−0.71)	(0.94)	(0.2)
Δ *Score Competition*	−1	−0.9	−2.57*	−2.52	2.43	2.86	0.74	0.38	−2.27	−2.86
	(−0.98)	(−0.87)	(−1.65)	(−1.59)	(0.98)	(1.17)	(0.33)	(0.17)	(−0.66)	(−0.8)
Δ *Score Uncertainty*	−1.42	−1.36	−0.33	−0.22	2.85	2.18	−1.64	−0.79	1.41	1.53
	(−0.76)	(−0.71)	(−0.13)	(−0.09)	(0.74)	(0.57)	(−0.33)	(−0.16)	(0.27)	(0.29)
Δ *Score R&D*	0.11	−0.24	2.86	3.32	−5.57	−6.43*	−0.02	0.29	11.41***	10.59***
	(0.07)	(−0.15)	(1.34)	(1.58)	(−1.63)	(−1.87)	(0)	(0.05)	(3.54)	(2.94)
Δ *Score Growth*	−0.11	0.47	−0.11	0.15	2.26	2.59	−0.54	−0.56	−6.35*	−6.84*
	(−0.08)	(0.36)	(−0.06)	(0.09)	(0.64)	(0.69)	(−0.13)	(−0.12)	(−1.66)	(−1.77)
Return		−0.78		−0.67		−0.71		4.17***		0.1
		(−1.17)		(−0.8)		(−0.5)		(2.75)		(0.06)
FCF		−3		−5.27**		−0.21		16.85***		−2.38
		(−1.3)		(−2.38)		(−0.04)		(2.86)		(−0.3)
Leverage		4.07***		4.6**		1.76		−5.87*		−5.32**
		(3.68)		(2.57)		(0.69)		(−1.92)		(−2.38)
Firm Size		0.02		0.18		0.69		0.95**		0.06
		(0.09)		(1.22)		(1.27)		(2.15)		(0.17)
Cash		0.3*		0.46*		0.37		−1.25***		0.49
		(1.75)		(1.86)		(0.75)		(−3.74)		(1.45)
Gross Margin		−2.49**		−1.75		−5.32**		−0.26		−0.52
		(−2.2)		(−1.13)		(−2.1)		(−0.09)		(−0.17)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,204	7,933	8,204	7,933	8,204	7,933	8,204	7,933	8,204	7,933

Panel B: Logit	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Δ *Score Fin*	18.1	27.96	13.48	43.9	1.49	−11.17	−32.62	−32.91	8.61	21.42
	(0.36)	(0.63)	(0.27)	(0.8)	(0.04)	(−0.28)	(−1.2)	(−1.23)	(0.25)	(0.6)
Δ *Score Loss*	272.8***	251.2**	179.1	191.3	−1.66	−9.41	−56.22	−39.29	39.41	4.56
	(2.66)	(2.26)	(1.48)	(1.34)	(−0.03)	(−0.14)	(−0.95)	(−0.63)	(0.79)	(0.1)
Δ *Score Competition*	−37.01	−27.53	−57.62	−52.73	20.7	24.15	5.66	2.93	−15.22	−18.75
	(−0.91)	(−0.69)	(−1.57)	(−1.5)	(1.07)	(1.26)	(0.31)	(0.15)	(−0.71)	(−0.84)
Δ *Score Uncertainty*	−60.58	−56.93	−7.4	−7.28	25.03	18.01	−10.5	−2.33	6.48	6.88
	(−0.88)	(−0.82)	(−0.13)	(−0.13)	(0.78)	(0.56)	(−0.27)	(−0.06)	(0.2)	(0.21)
Δ *Score R&D*	3.41	−25.77	72.04	81.49	−46.96*	−56.36**	0.87	2.99	66.82***	61.67***
	(0.04)	(−0.32)	(1.37)	(1.51)	(−1.65)	(−1.99)	(0.02)	(0.07)	(3.5)	(2.92)
Δ *Score Growth*	−6.54	9.82	−0.95	3.35	18.68	20.65	−2.36	−2.85	−42.61*	−45.69*
	(−0.12)	(0.2)	(−0.02)	(0.08)	(0.69)	(0.69)	(−0.07)	(−0.08)	(−1.75)	(−1.87)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,066	7,734	8,059	7,800	8,197	7,926	8,197	7,926	8,165	7,895

Table 3.9: **Merger Negotiation and Target Risk Factor Disclosures**

This table contains the results of panel regressions. Dependent variables are dummy variables equal to one if topics about "raising financing" or "cash-flow uncertainty" are discussed during the private merger negotiations. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per industry. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Target's point of view							
	Financial Issues Mention				*Cash-Flow Uncertainty Mention*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ *Score Fin*	13.75**	75.27**	13.21*	79.07*	14.08	97.99	18.19*	113.7*
	(2.28)	(2.08)	(1.75)	(1.68)	(1.62)	(1.54)	(1.73)	(1.69)
Δ *Score Loss*	−7.09	−41.99	−20.44*	−117.8*	−7.8	−54.24	−10.43	−69.65
	(−0.64)	(−0.72)	(−1.66)	(−1.81)	(−0.96)	(−0.79)	(−1.3)	(−1.08)
Δ *Score Competition*	−1.13	−5.26	2.44	12.13	−6.82	−51.58	−2.27	−10.09
	(−0.14)	(−0.12)	(0.32)	(0.27)	(−0.88)	(−0.78)	(−0.28)	(−0.15)
Δ *Score Uncertainty*	28.12*	160.5**	26.18*	158.7*	−6.76	−55.71	−5.58	−71.82
	(1.92)	(2.02)	(1.76)	(1.85)	(−0.33)	(−0.36)	(−0.3)	(−0.52)
Δ *Score R&D*	4.48	24.15	7.09	37.48	4.77	66.27	9.24	106.1
	(0.41)	(0.4)	(0.46)	(0.47)	(0.42)	(0.62)	(0.75)	(0.9)
Δ *Score Growth*	1.22	7.09	9.58	50.02	9.02	64.7	17.11*	136.3*
	(0.13)	(0.13)	(0.9)	(0.87)	(0.95)	(0.87)	(1.77)	(1.7)
Return			−0.69	−5.88			−8.84*	−67.36
			(−0.14)	(−0.18)			(−1.86)	(−1.62)
FCF			−4.91	−33.63			−6.5	−40.37
			(−0.64)	(−0.68)			(−0.58)	(−0.52)
Leverage			19.6***	119.2***			5.38	43.92
			(2.85)	(2.64)			(0.67)	(0.74)
Firm Size			−0.29	−2.37			−3.43**	−27.23**
			(−0.12)	(−0.16)			(−2.51)	(−2.53)
Cash			−0.1	−0.18			1.43	11.63
			(−0.05)	(−0.01)			(1.25)	(1.32)
Gross Margin			−14.41*	−85.81*			−21.02**	−170.8**
			(−1.93)	(−1.91)			(−2.54)	(−2.52)
Logit	No	Yes	No	Yes	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	944	923	895	876	944	923	895	876

Table 3.10: **Merger Negotiation and Acquirer Risk Factor Disclosures**

This table contains the results of panel regressions. Dependent variables are dummy variables equal to one if topics about "raising financing" or "cash-flow uncertainty" are discussed during the private merger negotiations. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per industry. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Acquirer's point of view							
	Financial Issues Mention				*Cash-Flow Uncertainty Mention*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ *Score Fin*	11.01 (1.28)	60.46 (1.36)	9.31 (1.14)	53.43 (1.21)	−16.01 (−1.53)	−146.2 (−1.31)	−15.13 (−1.29)	−119 (−1)
Δ *Score Loss*	18.74 (0.98)	85.4 (0.95)	19.76 (0.97)	94.46 (0.92)	11.96 (0.84)	107.5 (0.82)	8.63 (0.64)	55.28 (0.46)
Δ *Score Competition*	16.4** (2.14)	81.28** (2.06)	15.31** (2.06)	73.98* (1.95)	5.83 (0.96)	41.04 (0.8)	4.51 (0.7)	29.32 (0.52)
Δ *Score Uncertainty*	−21.32 (−1.22)	−116.8 (−1.2)	−22.3 (−1.22)	−123.4 (−1.23)	19.05** (2.12)	131.3* (1.71)	17.1* (1.91)	112.6 (1.5)
Δ *Score R&D*	9.11 (0.59)	45.25 (0.56)	14.39 (0.85)	71.83 (0.83)	1.4 (0.09)	7.52 (0.07)	7.69 (0.55)	46.55 (0.43)
Δ *Score Growth*	4.59 (0.49)	17.99 (0.36)	2.28 (0.22)	11.76 (0.21)	−3.43 (−0.27)	−34.3 (−0.33)	−3.87 (−0.32)	−41.69 (−0.42)
Return			−7.88 (−1.22)	−39.26 (−1.08)			−1.72 (−0.31)	−11.3 (−0.26)
FCF			−1.37 (−0.09)	−16.07 (−0.21)			−9.58 (−0.61)	−91.1 (−0.79)
Leverage			21.11** (2.26)	113.2** (2.24)			8.53 (0.87)	68.3 (0.99)
Firm Size			−2.04 (−1.43)	−10.08 (−1.21)			−0.86 (−0.85)	−6.07 (−0.78)
Cash			−0.32 (−0.21)	−2.02 (−0.22)			0.14 (0.15)	1.46 (0.2)
Gross Margin			−12.48 (−1.55)	−66.63 (−1.58)			−12.52* (−1.7)	−94.49 (−1.63)
Logit	No	Yes	No	Yes	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,170	1,157	1,133	1,121	1,170	1,123	1,133	1,090

Table 3.11: **Merger Wave and Risk Factor Exposure Similarity**

This table contains the results of panel regressions. Dependent variables are the log-transformed number of M&A deals announced over the following year per industry. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per industry. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable:	
	Number of Deals	
	(1)	(2)
Δ *RF Similarity*	-282.3^{**} (-2.11)	-299.9^{**} (-2.12)
Number of Deals		4.89 (1.03)
Return		6.36 (0.52)
FCF		-11.93 (-0.4)
Leverage		36.28 (0.7)
Firm Size		-12.34^{**} (-2.27)
Cash		-0.07 (-0.01)
Gross Margin		78.3^{*} (1.65)
Industry FE	Yes	Yes
Year FE	Yes	Yes
Observations	810	810

Table 3.12: **Future M&A Activity and Concerns about ongoing Consolidation**

This table contains the results of panel regressions. In Panel A, the dependent variables are two dummy variables equals to one if the filing company will receive an offer to be acquired by another company within the next twelve months following the filing date and if the deal is successful. Panel B has similar dependent variables, except that the dependent variables capture whether the filing firm announces a (successful) M&A deal. The regression coefficients are reported in percentage points. T-statistics are reported within parenthesis below the regression coefficients. Standard errors are clustered per firm. *, **, *** indicate statistical significance respectively at the 10%, 5% and 1%.

	Dependent variable:									
	Is Target					*Is Completed*				
Panel A: Target	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Δ Score Consolidation	6.13*	6.35*	162.6**	8.05**	136.9**	5.43*	6.05*	186.6**	8.18**	162.3**
	(1.87)	(1.91)	(2.2)	(2.3)	(2.39)	(1.72)	(1.89)	(2.24)	(2.45)	(2.57)
Return		−2.93***	−70.29***	−2.83***	−46.54***		−2.49***	−84.08***	−2.27***	−44.87***
		(−7.41)	(−7.59)	(−6.94)	(−6.64)		(−6.95)	(−7.86)	(−6.11)	(−5.85)
FCF		2.32	55.95	−0.49	−2.51		1.91	63.2	−0.16	1.79
		(1.59)	(1.55)	(−0.49)	(−0.18)		(1.5)	(1.55)	(−0.18)	(0.13)
Leverage		3.73***	89.21***	5.07***	77.95***		3.25***	118***	4.2***	76.59***
		(3.25)	(2.96)	(7.38)	(7.75)		(3.07)	(3.1)	(6.87)	(7.21)
Firm Size		−1.18***	−24.8***	−0.88***	−14.88***		−1***	−32.15***	−0.62***	−12.23***
		(−5.3)	(−4.23)	(−8.03)	(−8.06)		(−4.95)	(−4.55)	(−6.31)	(−6.25)
Cash		−0.35**	−10.1**	0.35***	5.86***		−0.38**	−15.42***	0.18*	3.38*
		(−2.04)	(−2.25)	(3.31)	(3.17)		(−2.4)	(−2.89)	(1.9)	(1.73)
Gross Margin		−0.66	−47.55*	1.14*	17.7*		−0.17	−68.72**	1.68***	31.63***
		(−0.54)	(−1.73)	(1.75)	(1.73)		(−0.16)	(−2.14)	(2.88)	(2.91)
Logit	No	No	Yes	No	Yes	No	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Industry X Year FE	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Observations	52,012	47,941	19,270	47,941	44,272	52,012	47,941	16,543	47,941	43,441

	Dependent variable:									
	Is Acquirer					*Is Completed*				
Panel B: Acquirer	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Δ Score Consolidation	0.03	−0.19	5.94	3.93	26.72	−2.25	−2.22	−12.77	0.6	3.85
	(0.01)	(−0.04)	(0.14)	(0.75)	(0.75)	(−0.45)	(−0.43)	(−0.29)	(0.12)	(0.1)
Return		5***	56.19***	5.75***	50.6***		4.63***	56.31***	5.48***	52.33***
		(8.97)	(9.89)	(10.62)	(11.9)		(8.64)	(9.6)	(10.58)	(11.98)
FCF		3.91**	94.78***	6.1***	75.24***		4.37**	107.2***	6.13***	80.25***
		(2.11)	(3.26)	(4.52)	(4.88)		(2.41)	(3.43)	(4.8)	(5.05)
Leverage		−19.41***	−178.6***	1.04	7.97		−19.33***	−191.5***	0.55	4.16
		(−10.63)	(−9.63)	(0.81)	(0.79)		(−11.04)	(−9.93)	(0.45)	(0.4)
Firm Size		0.91***	14.31***	3.76***	28.94***		0.58*	12.11***	3.56***	29.69***
		(2.64)	(4.09)	(16.67)	(16.67)		(1.75)	(3.36)	(16.81)	(16.88)
Cash		1.08***	10.14***	−0.62***	−5.8***		0.94***	9.76***	−0.7***	−7***
		(4.02)	(4.17)	(−2.8)	(−3.54)		(3.62)	(3.87)	(−3.35)	(−4.18)
Gross Margin		9.27***	95.82***	5.56***	42.36***		8.99***	104.7***	5.23***	42.86***
		(5.63)	(4.7)	(4.31)	(4.16)		(5.52)	(4.63)	(4.29)	(4.11)
Logit	No	No	Yes	No	Yes	No	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Industry X Year FE	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Observations	52,012	47,941	30,811	47,941	47,427	52,012	47,941	29,792	47,941	47,321

Appendix

Table 3.A1: **Matched Texts about Consolidation Risks**

This table contains examples of texts classified as being about industry consolidation. I use a mix of regular expressions and bag-of-words approaches to correctly classify texts. The sources of the respective examples are provided directly below each row.

	Risk Title	*Risk Description*
1	the consolidation of retailers may adversely affect anheuser-busch.	the retail industry in the united states and in other countries in which anheuser-busch operates continues to consolidate. large retailers may seek to improve profitability and sales by reducing the prices or increasing the promotional activities for anheuser-busch products. although retailers purchase products not from anheuser-busch, but from its wholesalers including in a limited number of markets, the anheuser-busch wholesaler operations, the efforts of retailers could result in reduced profitability for the beer industry as a whole and indirectly adversely affect our financial results.
	`https://www.sec.gov/Archives/edgar/data/310569/000106880008000087/ab.htm`	
2	consolidation among our competitors may negatively impact our business.	in recent years, several large auto parts chains have merged. we do not know the impact these mergers will have upon competition in the retail automotive aftermarket. if our competitors are able to achieve efficiencies in their mergers, then there may be greater competitive pressures in the markets in which they are stronger.
	`https://www.sec.gov/Archives/edgar/data/866787/000114420405032748/v027481.htm`	
3	continued consolidation of distributor networks in the pharmaceutical industry as well as increases in retailer concentration may limit our ability to profitably sell our products.	we sell most of our products to large pharmaceutical wholesalers, who in turn sell to hospitals, surgery centers and retail pharmacies. the distribution network for pharmaceutical products has become increasingly consolidated in recent years. further consolidation or financial difficulties could also cause our customers to reduce the amounts of our products that they purchase, which would materially and adversely impact our business, financial condition and results of operations.
	`https://www.sec.gov/Archives/edgar/data/1087294/000162828015001629/a2014-10k.htm`	
4	industry consolidation may have a negative impact on our business model.	the telecommunications industry is currently undergoing consolidation. as customers combine businesses, they may require less colocation space, and there may be fewer networks available to choose from. given the competitive and evolving nature of this industry, further consolidation of our customers and/or our competitors may present a risk to our network neutral business model and have a negative impact on our revenues. in addition, increased utilization levels industry-wide could lead to a reduced amount of attractive expansion opportunities available to us.
	`https://www.sec.gov/Archives/edgar/data/1101239/000119312506056358/d10k.htm`	
5	consolidation in the healthcare industry could result in greater competition and reduce our revenues and harm our business.	many healthcare industry companies are consolidating to create new companies with greater market power. as the healthcare industry consolidates, competition to provide products and services to industry participants will become more intense. these industry participants may try to use their market power to negotiate price reductions for our products or may undertake additional vertical integration or supplier diversification initiatives. if we are forced to reduce our prices, our revenues would decrease and our operating results would suffer.
	`https://www.sec.gov/Archives/edgar/data/1114483/000111448322000002/gb-20211231.htm`	
6	consolidation in the retail and foodservice industries could affect our sales and profitability.	if our retail and foodservice customers continue to grow larger due to consolidation in their respective industries, they may demand lower pricing and increased promotional programs. meeting these demands could adversely affect our sales and profitability.
	`https://www.sec.gov/Archives/edgar/data/1128928/000119312514058624/d634175d10k.htm`	
7	the consolidation of distribution channels within the rv industry could have a material negative effect on revenues and profitability.	over the last several years, several large-scale recreational vehicle dealers have grown to represent a significant presence in the industry. the expansion of large-scale dealers and the continued consolidation of dealerships among large players may result in increased pricing pressures in the industry in general. such pressure exerted by the distribution channel may have a material adverse effect on our revenues and profitability.
	`https://www.sec.gov/Archives/edgar/data/910655/000110465907024474/a07-5812_110k.htm`	

Table 3.A2: **Sources for Matched Texts about Consolidation Risks**

This table contains the sources of the examples provided in Table 3.1. Elements share the same ordering between the two tables.

	Link
1	https://www.sec.gov/Archives/edgar/data/318996/000095012909000678/h65786e10vk.htm
2	https://www.sec.gov/Archives/edgar/data/1220754/000119312507056902/d10k.htm
3	https://www.sec.gov/Archives/edgar/data/918580/000114420419015943/tv515018_10k.htm
4	https://www.sec.gov/Archives/edgar/data/809012/000101795104000245/k604.htm
5	https://www.sec.gov/Archives/edgar/data/737468/000073746811000017/form10-k20101231.htm
6	https://www.sec.gov/Archives/edgar/data/1036044/000095013409006515/f51973e10vk.htm
7	https://www.sec.gov/Archives/edgar/data/946644/000094664415000008/heb1231201410k.htm
8	https://www.sec.gov/Archives/edgar/data/874015/000087401522000079/form10k.htm
9	https://www.sec.gov/Archives/edgar/data/105132/000119312508214535/d10k.htm
10	https://www.sec.gov/Archives/edgar/data/911583/000104746911002131/a2202616z10-k.htm
11	https://www.sec.gov/Archives/edgar/data/804753/000095013705003217/c93199e10vk.htm
12	https://www.sec.gov/Archives/edgar/data/886128/000088612816000032/fcel-103115x10k.htm
13	https://www.sec.gov/Archives/edgar/data/1675820/000167582017000006/tivocorp12311610-k.htm
14	https://www.sec.gov/Archives/edgar/data/866734/000119312506081325/d10k.htm
15	https://www.sec.gov/Archives/edgar/data/879465/000089109204004564/e19136_10k.txt
16	https://www.sec.gov/Archives/edgar/data/701853/000070185314000003/bxs-20131231x10k.htm
17	https://www.sec.gov/Archives/edgar/data/1028215/000119312512423466/d354175d10k.htm
18	https://www.sec.gov/Archives/edgar/data/1078099/000119312506070559/d10ksb.htm
19	https://www.sec.gov/Archives/edgar/data/1054303/000093041307001010/c46580_10ksb.txt
20	https://www.sec.gov/Archives/edgar/data/876043/000095013707000369/c11132ke10vk.htm
21	https://www.sec.gov/Archives/edgar/data/1090061/000109006118000010/omn-20171130x10k.htm
22	https://www.sec.gov/Archives/edgar/data/911216/000165495421010495/ptn_10k.htm
23	https://www.sec.gov/Archives/edgar/data/1085770/000095012905002265/v06627e10vk.htm
24	https://www.sec.gov/Archives/edgar/data/1302707/000095012311029824/g26402ke10vk.htm
25	https://www.sec.gov/Archives/edgar/data/1019671/000119312514131604/d652929d10k.htm
26	https://www.sec.gov/Archives/edgar/data/1158780/000121390020026105/f10k2020_pluristemtherap.htm
27	https://www.sec.gov/Archives/edgar/data/1388133/000114036114035330/form10k.htm
28	https://www.sec.gov/Archives/edgar/data/1730346/000162828019003506/chra-1231201810xk.htm
29	https://www.sec.gov/Archives/edgar/data/1034054/000103405417000003/sbac-20161231x10k.htm
30	https://www.sec.gov/Archives/edgar/data/1808665/000180866521000019/asrt-20201231.htm
31	https://www.sec.gov/Archives/edgar/data/225051/000143774911001418/arden_10k-010111.htm
32	https://www.sec.gov/Archives/edgar/data/913241/000162828018002589/shoo-20171231x10k.htm
33	https://www.sec.gov/Archives/edgar/data/1001902/000156459015000735/d835534d10k.htm
34	https://www.sec.gov/Archives/edgar/data/1108967/000143774922007887/oeg20211231_10k.htm
35	https://www.sec.gov/Archives/edgar/data/1090514/000101968706000688/aobo_10k-123105.txt
36	https://www.sec.gov/Archives/edgar/data/1116463/000119312510057039/d10k.htm
37	https://www.sec.gov/Archives/edgar/data/275858/000095013306001618/w19013e10vk.htm
38	https://www.sec.gov/Archives/edgar/data/1617242/000156459018022401/krny-10k_20180630.htm
39	https://www.sec.gov/Archives/edgar/data/794172/000095012309066890/c93071e10vk.htm
40	https://www.sec.gov/Archives/edgar/data/949721/000121390017003860/f10k2016_meridianwaste.htm
41	https://www.sec.gov/Archives/edgar/data/1596532/000159653220000024/anet20191231-10k.htm
42	https://www.sec.gov/Archives/edgar/data/1049108/000104910808000082/form10k123107.htm
43	https://www.sec.gov/Archives/edgar/data/836106/000120677405000346/d16545_10-k.htm
44	https://www.sec.gov/Archives/edgar/data/917491/000091749119000012/faro-123118x10k.htm
45	https://www.sec.gov/Archives/edgar/data/948708/000119312514088577/d657850d10k.htm
46	https://www.sec.gov/Archives/edgar/data/872448/000095013408003856/f38449e10vk.htm