technische universität
dortmund

Essays in Finance: Initial Public Offerings and Risk Backtesting

Dissertation zur Erlangung des akademischen Grades
Doctor rerum politicarum
der Fakultät für Wirtschaftswissenschaften
der Technischen Universität Dortmund

Philipp Schmidtke

December 14, 2020

Contents

List of Figures

List of Tables

Acknowledgment

I thank my advisor, Professor Peter N. Posch. He contributed to this dissertation's content by offering continuous input, he gave academic freedom by allowing the exploration of different ideas, and he provided substantial resources, which made this work possible. I am deeply grateful to my co-authors, Gerrit Köchling and Professor Axel Bücher, for our close and fruitful collaboration. I am also grateful to the finance research group at TU Dortmund University - Johannes, Daniel U., Janis, Vladislav, Timotheos, Gerrit, Nils, Aydin, Miguel, Daniel N., and Lars - as well as Nicole for contributing to a creative and thoughtful working environment. Similarly, I thank our student research assistants for their engaged and precise work. I also owe the various editors and reviewers of journals and conferences for their improvements of my papers. Finally, I thank my family for their continuous support, which eventually led to this work.

This work has been supported in part by the Collaborative Research Center "Statistical Modeling of Nonlinear Dynamic Processes" (SFB 823, Project A7) of the German Research Foundation (DFG).

1 Introduction

Market prices of publicly traded companies are a vital research object throughout the world of finance and beyond. On the one hand, researchers seek to understand properties of market prices, such as expected returns (Harvey et al., 2016; Fama and French, 2015) or time-series behavior (Engle, 1982; Bollerslev, 1986a). On the other hand, market prices are used to estimate various economic quantities, e.g., systemic risk of financial institutions (Adrian and Brunnermeier, 2016; Acharya et al., 2017), assuming that prices contain meaningful information, which is naturally linked to their hypothesized informational efficiency (Fama, 1970, 1991).

This dissertation's two main focuses are related to this topic. First, Chapters 2 and 3, study the initial public offering (IPO), the process by which previously private firms initially offer their shares to the general public, eventually resulting in a first stock exchange valuation. Chapter 2 proposes a new measure of investor sophistication using the internet log file data set of the United States Securities and Exchange Commission (SEC) Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system and investigates the role of sophisticated and unsophisticated investor attention for IPO pricing. Chapter 3 defines a daily SEC workload measure at the SEC industry office level and performs an analysis of the SEC filing review process for IPOs, including textual analysis of SEC comments. The dissertation uses this to examine the interaction among high SEC workload, filing review outcomes, and IPO pricing.

The second main focus of this dissertation, set out in Chapters 4 and 5, is backtesting of market risk forecasts for financial returns, which is a central concern of risk managers as soon as an asset is traded publicly. Chapter 4 proposes novel Value-at-Risk (VaR) backtests for the independence property of VaR forecasts using the extremal index, which is a concept from extreme value theory. Chapter 5 conducts a critical analysis of volatility forecasting capabilities of a large set of Generalized Autoregressive Conditional Heteroscedasticity (GARCH)-type models for returns on the Bitcoin cryptocurrency, known for its extreme price changes.

IPOs constitute a cesura in the life cycle of many companies. At the cost of providing more informational transparency, issuers can raise capital, and selling stockholders can liquidate investments from earlier financing stages (e.g., venture capital). Underpricing, probably the most prominent feature of IPOs, refers to the observation that the market price at the end of the first trading day is on average considerably higher than the offering price set by the underwriters and issuers. This puzzling and persistent fact has prompted researchers to propose underpricing theories that try to rationalize why issuers and underwriters *leave money on the table*, so to speak, on average. Traditional theories include the Rock model (1986), the litigation risk approach (Tiniç, 1988), and the bookbuilding model (Benveniste and Spindt, 1989).

Chapter 2 focuses on a more recent theory related to underpricing: that attention induces short-term price increases followed by subsequent underperformance (Ljungqvist et al., 2006; Barber and Odean, 2008; Da et al., 2011).[1] This study overcomes some of the limitations of existing attention measures by using the publicly available EDGAR log file data set. For instance, extreme returns as used by Barber and Odean (2008) measure attention only indirectly and Google searches as used in Da et al. (2011) are quite obfuscated and appear problematic in cross-sectional applications. Instead, the EDGAR logs contain IP address-level data of internet user accesses to financial disclosures made available via the SEC's EDGAR system.[2] Based on this detailed and almost plain source of revealed investor attention, this study proposes a measure of financial disclosure experience at the IP address level, which is based on past years' personal access history for each address, i.e., counts of unique daily accesses to filings. Since retrieval of firm-specific information is likely related to gains in knowledge, this study links the extent of research on EDGAR to investor sophistication. For each IPO, pre-IPO-week EDGAR attention conditional on sophistication, i.e., for the two groups of rather sophisticated and unsophisticated EDGAR-users is calculated. Attention from unsophisticated investors is related to more underpricing and to weaker long-term performance by the IPOs when both attention and underpricing were high, i.e., IPOs suspected to be under price pressure. This study does not find the latter effect for sophisticated attention, which supports the attention-induced price pressure theories of

[1] The latter is another stylized fact of IPOs (Loughran and Ritter, 1995).

[2] Since the last octet of each IP address is replaced with letters, the addresses are slightly obfuscated. However, the replacement is constant over time, which makes it possible to track an address over time.

Barber and Odean (2008) and Da et al. (2011).

Chapter 3 examines the interaction among high SEC workload, the SEC filing review process, and IPO pricing. The SEC's Division of Corporation Finance (CF) reviews firm disclosures before an IPO and sends comment letters to issuers to ensure disclosure quality (U.S. Securities and Exchange Commission, 2019b). With this process, the SEC can contribute to the information environment of IPOs by inducing issuers to disclose additional or revised information. However, Gunny and Hermis (2020) and Ege et al. (2020) have recently provided evidence of detrimental workload effects for reviews of periodic filings (10-Ks, 10-Qs). Chapter 3 transfers the workload idea to the IPO setting and constructs a daily workload measure for the CF offices by estimating the number of filings in urgent review for each office. Since the filing volume can fluctuate greatly - e.g., IPOs themselves are known for their wave behavior (Pástor and Veronesi, 2005) - there is also potential workload variation. This study creates a data set of IPO filings and comment letters, which it uses to calculate several filing review measures, such as the number of comments or SEC response times. Further, an exploratory comment clustering using textual analysis, which reveals sizeable clusters of almost identical comments for different IPOs, is performed. While this study finds that a measure of initial SEC concerns based on detrended initial comment counts is related to price changes, usually downward, similar to the literature (Li and Liu, 2017; Lowry et al., 2020), this relation is considerably smaller under high SEC workload. The study finds no evidence of fewer initial comments under high workload, but some evidence of such effects among later letters. However, initial comments similar to recently issued ones for other IPOs are more frequent under high workload. In a duration analysis with proportional hazard models put forth by (Cox, 1972), this study finds a tendency toward quicker SEC replies. All in all, these findings appear to indicate lower quality IPO reviews under high workload. In this case, it would be expected that investors need to produce more information during the bookbuilding (Hanley and Hoberg, 2010). Indeed, this study finds that high workload IPOs are more underpriced.

The field of backtesting market risk forecasts for financial returns, which is the second main focus of this work, is concerned with forecast quality. Two conceptually different approaches are available. The first framework tests the statistical hypothesis that a time-series of forecasts coincides with the true, unknown risk. With the VaR as the risk measure, a prominent example in practice is the Basel "Traffic Light" Approach

(BCBS, 1996b, 2016).[3] In banking regulation, minimal capital requirements can be based on internal risk forecasts. However, they are required to pass backtests, and that emphasizes the relevance of tests of this kind. The tests developed in Chapter 4 belong to this framework for the independence hypothesis. Instead, the Diebold-Mariano framework (Diebold and Mariano, 1995) compares two or more concurrent time-series of forecasts, regardless of their correctness, and aims at testing the relative superiority in terms of a prespecified loss function. This framework is more appropriate in terms of model selection. Chapter 5 offers an empirical analysis in this regard with the volatility as the risk measure.

The VaR backtests proposed in Chapter 4 are based on commonalities between VaR backtesting and the extremal index, a concept from extreme value theory (Leadbetter, 1983; Embrechts et al., 1997). This study focuses on the independence property of VaR forecasts, stating that exceedances of VaR forecasts should happen independently over time. Commonly, forecasts violating this property will yield clustered exceedances. The independence property complements the unconditional coverage property of VaR forecasts, in which the focus is on the number of exceedances. The extremal index quantifies the extent to which extreme observations of a stationary time-series occur in clusters. It is bounded to the interval $(0, 1]$, whereby smaller values indicate more clustering. Hence, an extremal index of 1 means an absence of clustering. This study uses this characteristic, defines a series of relative excess returns as the ratio of realized returns and VaR forecasts, and applies two extremal index estimators (Süveges and Davison, 2010; Northrop, 2015; Berghaus and Bücher, 2018) to test for an extremal index of 1. With Monte Carlo simulations, the study compares the size and power properties of the new tests with existing alternatives (Christoffersen and Pelletier, 2004; Candelon et al., 2011). Depending on the employed estimator, the study often finds improved power to reject unfavorable forecasts. The increased power is achieved partly by using the data more liberal at the cost of losing accuracy for specific data-generating processes.

In the Diebold-Mariano framework, Chapter 5 investigates the volatility forecasting capabilities of a large set of GARCH-type models for returns on the Bitcoin cryptocurrency from 2015 until 2018. Bitcoin is known for substantial price changes and has recently attracted considerable attention. This study combines several robust loss functions

[3] BCBS is an abbreviation for Basel Committee on Banking Supervision.

(Patton, 2011) with various realized volatility estimators (Andersen et al., 2012) and applies the concept of model confidence sets (Hansen et al., 2011). The latter is especially important to account for the large number of models in consideration. Model confidence sets are iteratively constructed by eliminating models with inferior predictive ability. Eventually, this leads to a group of models with statistically equal predictive power, which are the basis of this discussion. The obtained model confidence sets in this study's application are relatively large, and only a few models can consistently be ruled out across the considered scenarios. Hence, unambiguously superior models are hard to detect, a fact that advises caution in the selection of a volatility model for Bitcoin returns.

This dissertation contributes to several literature strands. First, it adds to the well-developed IPO literature (Ljungqvist et al., 2006) by providing comprehensive evidence on the role of attention to IPO pricing as an additional factor for well-known IPO features as underpricing and long-term underperformance. Moreover, it widens the emerging literature on the role of SEC filing reviews in IPO pricing and information production (Li and Liu, 2017; Lowry et al., 2020) by examining SEC workload and SEC comments at a granular level. Second, the dissertation expands research on attention (Da et al., 2011; Ben-Rephael et al., 2017; Drake et al., 2020) as well as the literature on EDGAR log files (Lee et al., 2015; Chen et al., 2020) by proposing a publicly available and easily constructible attention measure that can capture different groups of investors from the same data source. Third, it contributes to the distraction literature. While most research in this regard focuses on investor distraction (e.g., Hirshleifer et al. (2009)), the dissertation adds to the young literature on regulator distraction (Gunny and Hermis, 2020; Ege et al., 2020) and converts this idea to IPOs. Fourth, this work also extends the existing work on SEC filing reviews, typically focusing on periodic filing reviews (Cunningham and Leidner, 2019). In contrast with existing studies, this one does not focus only on initial comment letters and performs a text clustering at the comment level, thus giving unique insights into the nature of the SEC comments. Fifth, this thesis adds to the VaR backtesting literature (Christoffersen, 1998; Candelon et al., 2011) by introducing and analyzing a new set of tests originating from extreme value theory, potentially yielding substantial power improvements. Sixth, it introduces a novel application of concepts from extreme value theory to the financial context similar to, for instance, McNeil and Frey (2000) and Longin (2000), who estimate the VaR applying extreme value theory.[4] Seventh, this dissertation contributes to the literature on risk forecasting for cryptocurrencies (Katsiampa, 2017; Chu et al., 2017) by conducting a critical and conceived analysis of the forecasting capabilities of a large set of GARCH-type models for the Bitcoin volatility. This analysis advises the modeler to be careful about model choice.

[4] See Rocco (2014) for further examples.

1.1 Publication details

Paper I (chapter 2):

Information Acquisition Experience, Investor Sophistication, and IPO Price Pressure

Authors:

Gerrit Köchling, Philipp Schmidtke, Peter N. Posch

Abstract:

We propose a simple measure of investor sophistication based on financial statement experience derived from publicly available EDGAR log data about accounting information acquisition activity. This approach allows us to provide unique empirical evidence for the existence of attention induced price pressure effects in the cross-section of initial public offerings, i.e. that pre-IPO-week attention from likely unsophisticated investors is associated with higher initial returns and subsequent, significant price depreciations. These results are robust to various measures of abnormal post-IPO returns, attention, and sophistication. The proposed direct experience-measurement is easily replicable and potentially useful for many other questions in economics.

Publication details:

Working paper.

Paper II (chapter 3):
SEC Workload, IPO Filing Reviews, and IPO Pricing

Authors:
Gerrit Köchling, Philipp Schmidtke, Peter N. Posch

Abstract:
We analyze the interaction between high workload of the Securities and Exchange Commission (SEC) staff and the information production stimulated by their review process of initial public offerings (IPOs). We find that high workload is associated with more generic comments in the first letter, with fewer overall comments for later letters, and that the SEC answers quicker while being busy. Using a measure of initial SEC concerns based on comment counts, we find, for instance, a positive relation with absolute price revisions from the initial estimate to the final price. If we additionally consider an interaction with high workload, such effects become weaker for high workload IPOs and stronger for non-high workload IPOs. Partly but not entirely, generic comments mediate this effect. Consistent with the view that our findings indicate fewer SEC induced information production under high workload, we find that underpricing is significantly larger for high workload IPOs. This is in line with theories, where investors are compensated for their information production via bookbuilding.

Publication details:
Working paper.

Paper III (chapter 4):
Using the Extremal Index for Value-at-Risk Backtesting

Authors:
Axel Bücher, Peter N. Posch, Philipp Schmidtke

Abstract:

We introduce a set of new Value-at-Risk independence backtests by establishing a connection between the independence property of Value-at-Risk forecasts and the extremal index, a general measure of extremal clustering of stationary sequences. For this purpose, we introduce a sequence of relative excess returns whose extremal index is to be estimated. We compare our backtest to both popular and recent competitors using Monte Carlo simulations and find considerable power in many scenarios. In an applied section, we perform realistic out-of-sample forecasts with common forecasting models and discuss advantages and pitfalls of our approach.

Publication details:

Journal of Financial Econometrics 18 (3) (2020): 556–584.

Paper IV (chapter 5):

Volatility forecasting accuracy for Bitcoin

Authors:

Gerrit Köchling, Philipp Schmidtke, Peter N. Posch

Abstract:

We analyze the quality of Bitcoin volatility forecasting of GARCH-type models applying different volatility proxies and loss functions. We construct model confidence sets and find them to be systematically smaller for asymmetric loss functions and a jump robust proxy.

Publication details:

Economics Letters 191 (2020): 108836

2 Information Acquisition Experience, Investor Sophistication, and IPO Price Pressure

The following is based on Köchling et al. (2020a).

2.1 Introduction

The incorporation of information into asset prices is an extensively studied research area, but information alone is not sufficient to explain many empirical capital market phenomena. Information needs not only to be available, it also needs to draw market participants' attention in order to be processed. Due to heterogeneity in market participants, interpretations and associated processing costs of the perceived information may differ greatly, for instance, depending on the experience and background of each individual.

Disclosure processing plays a crucial role in understanding market outcomes, and it affects all types of investors (Blankespoor et al., 2020). Empirical studies, however, often focus on a specific group of investors. This decision is naturally determined by the underlying research question but also limited by the fact that the existing measures capturing disclosure processing are often attributed to a specific group of investors.[1]

We propose an approach that is able to capture different levels of *financial disclosure experience* and their induced attention solely based on data from the SEC information retrieval system EDGAR, which is the most detailed source for revealed investor attention available. Its unique nature paved the way for numerous papers to study the relation between revealed investor attention and different market outcomes (Chen et al., 2020; Gibbons et al., 2020; Drake et al., 2020; Crane et al., 2018; Bauguess et al., 2018).

[1] For example, measures that use Google search activity capture primarily retail investor attention (Da et al., 2011), while measures that use Bloomberg terminal searches capture mostly institutional attention (Ben-Rephael et al., 2017).

We link our approach to *financial sophistication* by measuring *timely experience* with financial disclosures that we proxy by EDGAR-user's information acquisition activity relative to all other active EDGAR-users. The reasoning behind our approach is simple: Financial disclosures are lengthy, diverse, and possible relevant details easily hide within a boilerplate language. Dyer et al. (2017) document increases in length, boilerplate, stickiness, and redundancy and decreases in specificity, readability, and the relative amount of hard information in companies' annual reports over time. A certain degree of continuity in accessing financial statements implies topic-specific and timely self-education[2] and is hence needed to accurately put the information in company reports in context. This view is supported by actions of the SEC to level the playing field across classes of investors[3] and empirical evidence emphasizing differences in reactions to financial disclosures of investor groups of varying degrees of financial sophistication (Krische, 2019; Elliott et al., 2017; Tan et al., 2014; Miller, 2010).

To proxy for timely financial statement experience, we count the plain number of daily unique filings accessed for each user during the previous year and determine the user's relative position within the distribution of all active EDGAR-users in that year. As a first simple rule, we classify users left from the distribution's median user as less and those on the right side as more sophisticated. We formulate a general notation of our approach, which allows an easy adaptation to other research projects.

Using several versions of the proposed measurement, we provide empirical evidence for the attention induced price pressure hypothesis from Barber and Odean (2008) in a setting of U.S. initial public offerings (IPOs) from 2007 to 2017. The price pressure hypothesis proposes that stocks receiving increased attention - especially from unsophisticated investors - are also facing increased demand. In the case of IPOs, this should yield to inflated prices in the short-run and subsequent price reversals in the long-run (Da et al., 2011), a pattern IPOs generally are known for since initial returns from offer price to first end-of-day market price are large on-average, often followed by significant underperformance (Ibbotson, 1975; Loughran and Ritter, 1995).

In line with Barber and Odean's (2008) price pressure hypothesis, we find pre-IPO-week attention from unsophisticated EDGAR-users to be associated with higher initial

[2] Concepts linking continuity and education date back until Dewey (1938) who proposes a carefully developed theory of experience and its relation to education.

[3] For example, the SEC publishes "A Plain English Handbook" that emphasizes the "need to gauge the financial sophistication of [your] investors" to write understandable disclosures.

returns. Following Da et al. (2011), we proxy IPOs subject to price pressure as those with both large unsophisticated attention and large initial returns and find a pronounced subsequent price depreciation within one year after the IPO. The results are robust against different specifications of sophistication and attention measurement and several variants of matched abnormal post-IPO returns. However, while sophisticated attention is also related to more underpricing, the price pressure long-term return effect dissipates when sophisticated attention is used instead.

Our sophistication measurement based on EDGAR-experience relates to Drake et al. (2020), who propose to classify EDGAR-users from financial institutions like banks or funds as sophisticated and users from internet service providers (ISPs) as unsophisticated. We compare our *experience-based* classifications to the *entity-based* ones by Drake et al. (2020). We find virtually all users from institutions to be also sophisticated by experience, in line with the expectation that users from professional entities acquire more frequently fundamental information of potential investments. For ISPs, we find a broader spectrum of sophistication levels. That is, we classify a considerable amount of IP addresses from ISPs as sophisticated by experience, which differs from Drake et al.'s (2020) approach.

We also compare the sophistication-constrained attention measures from EDGAR with the attention measures based on Google search volume, proposed in Da et al. (2011) as a measure of retail investor attention. Comparing Google and EDGAR measures, we find significant differences that help to shed new light on the economics of retail and sophisticated investors' attention to IPOs. Both measures are explored, and it appears that attention measures based on Google punish firms with high general awareness with particularly low abnormal attention values, unlike EDGAR-based attention measures, which are built on raw, absolute view counts. Our discussion highlights significant limitations of Google as an attention measure in cross-sectional analyses.

Last, we compare the attention determinants for both an unsophisticated and sophisticated group, again in the IPO setting. As expected, both measures have joint drivers, such as IPO size (log of proceeds) and positive trailing earnings. Especially that profitable firms attract both groups is not consistent with previous findings that individuals disregard such publicly available information as in Field and Lowry (2009). Strikingly, we find an attention peak in Google searches within the pre-IPO-week is associated with a 20% increase in direct IPO filing searches by unsophisticated users but only a 10% increase for sophisticated users. This is in line with the generally advocated notion that Google searches reflect mainly unsophisticated attention (Da et al., 2011).

Our contribution to the literature is threefold: First, we introduce a measure of financial statement experience. Due to its simplicity and generalizability, it is easily adaptable to other attention-related questions. As a possible outlook, our measure can help understand whether investors respond differently to variations in financial reports and accounting outcomes (e.g., fogginess, quality) or if firms tailor their disclosure approach based on their readers' experience distribution. A recent study by Blankespoor et al. (2020) underlines the need to understand disclosure processing costs and their implications for a wide array of accounting research.

Second, we use our experience-measurement to introduce a novel sophistication measurement that substantially improves existing ones. Most studies using EDGAR as an attention measure rely mapping the partly obfuscated IP address and the underlying entity (Chen et al., 2020; Gibbons et al., 2020; Drake et al., 2020; Crane et al., 2018). This can be time-consuming, ambiguous, and difficult to replicate because of challenging steps such as combining several databases, assessing historic IP-to-entity mappings, string-matching tasks, and assumptions about the underlying entity.[4] Others rely on proprietary (Ben-Rephael et al., 2017) or aggregated and obfuscated data (Da et al., 2011; Ben-Rephael et al., 2017). The sophistication measure proposed in this paper does not rely on an IP address mapping, builds upon an explicit linkage between a company and its perceived attention due to EDGAR's nature, and allows to extract attention from both unsophisticated and sophisticated investors using the same database. The latter also facilitates drawing inference about one attention group while controlling for the attention of the other.

Third, using a variety of specifications of our sophistication measurement, we provide unique evidence for the existence of attention induced price pressure effects in the cross-section of new equity offerings. Besides Derrien (2005), Cook et al. (2006), Ljungqvist et al. (2006), and Dorn (2009), who also relate to this strand[5], the only study to our

[4] These studies are not primarily intended to measure sophistication but rather to quantify the effects of attention from specific entities, e.g., funds, analysts, etc.

[5] Derrien (2005) develops a model in which the price of IPO shares depends on the information about the intrinsic value of the company and investor sentiment, and finds for a sample of 62 French companies that large individual investors' demand leads to high initial returns and poor long-run performance. Cook et al. (2006) find, among other effects, higher initial returns for IPOs to be related

knowledge providing direct empirical evidence for the existence of price pressure effects is Da et al. (2011) using Google search volume and a sample of 185 IPOs from 2004 to 2007. Standard measures of abnormal Google search volume can be meaningful for *within-firm* analyses but tend to problematic in cross-sectional applications due to Google's scaling of the raw search volume to a range from 0 to 100.[6] Apart from our approach's strengths in effectively capturing cross-sectional individual investor attention, it furthermore allows us to control for overall attention towards an IPO while drawing inference on attention induced by unsophisticated investors, which we believe to be of striking importance for testing the price pressure hypothesis.

The remainder of the paper is organized as follows. In Section 2.2, we introduce our experience-based sophistication measurement with EDGAR logs in detail. Testing the price pressure hypothesis for IPOs is the subject of Section 2.3. In Section 2.4, we compare our approach to alternatives and explore its determinants. Section 2.5 concludes.

2.2 EDGAR Logs, Financial Statement Experience, and Sophistication

The EDGAR log files offer detailed data on demand for company information accurately assignable to a specific firm and further to a specific piece of information via the numerous filings. It is possible to match the IP-level data to specific entities to get even more detailed data on who demands the information. In its entirety, these features are unmatched by any other available attention proxy and have led to a number of insights in various research fields. For instance, Lee et al. (2015) show that the *collective wisdom* of investors allows producing search-based industry classifications outperforming standard industry classifications. Bernard et al. (2020) use the EDGAR log files to predict subsequent mergers and acquisitions as well as how and how much firms invest, relative to rivals. Recently and related to our work, Drake et al. (2020) find that stronger attention from sophisticated investors is more predictive for future firm performance.

Particularly suited for the present study, EDGAR facilitates analyzing the historical

to an investment banker's ability to market an IPO to sentiment investors.

[6] Our discussion relates to a recent paper by DeHaan et al. (2019), who find Google search volume to be a noisy proxy for investor attention.

accounting information activity on the user-level. This allows us to compare the information acquisition behavior of a single EDGAR-user relative to all other active users in a timely manner. A certain degree of continuity in accessing financial statements implies topic-specific and timely self-education, and likely reduces information frictions when accessing future financial statements. This, in turn, may impair not only an investor's ability to handle disclosed information, yet additionally the ability to make accurate investment-related decisions (Blankespoor et al., 2020). Blankespoor et al. (2020) argue that public information can be interpreted as a form of costly private information and make clear that learning from disclosures is an active economic choice. They divide the costs of processing a disclosure into three steps: awareness, acquisition, and integration costs. While the former two are greatly mitigated by disclosure availability, e.g. by data retrieval systems such as EDGAR, the latter refers to putting the information into a (valuation) context and is a largely continuous choice. We believe that the empirical measurement we introduce in this paper strongly relates to the alleviation of these costs and hence relates to investors' sophistication. Our rationale is in line with empirical evidence studying the relation between (topic-specific, timely) financial experience and proper financial decisions (Hilary and Shen, 2013; Calvet et al., 2009; Clement et al., 2007; Mikhail et al., 1997; Bonner and Lewis, 1990).

2.2.1 The Experience-Based Sophistication Measure

We propose our *experience-based* sophistication measure $SEB_{i,t}^{n}$ as follows. For each IP i we denote the set of EDGAR filings accessed on day d by $F_{i,d}(T)$ subject to a set of admissible filings T. We then sum the number of unique filings accessed within each day $|F_{i,d}(T)|$ over the previous n-day period and obtain

$$SEB_{i,t}^{n}(T) = \sum_{d=t-n}^{t-1} |F_{i,d}(T)|, \tag{2.2.1}$$

if the first access day $d_{i,\text{first}} = \min\{d : |F_{i,d}(T)| > 0\}$ of this IP was at least n days before t, that is $d_{i,\text{first}} \leq t - n$. In all other cases, we do not measure the sophistication since we would classify new IPs unsophisticated by default.

Throughout this paper, we use a one-year period, which is $n = 365$, for our sophistication estimates in order to avoid biases due to seasonality. For simplicity, we do not exclude any filings, setting $T = T_\Omega$ where T_Ω contains all filings. Thus, our standard measure is $SEB_{i,t}^{365}(T_\Omega) = SEB_{i,t}$. With T_Ω we measure the broadest kind of EDGAR activity. However, it may be desirable to measure experience more specifically, for instance by restricting the set of admissible filings T to all filings with *fundamental* information such as annual (10-K) or quarterly (10-Q) reports as used by Lee et al. (2015) for example. For robustness, we use a "fundamental" set as well, which yields similar results.

2.2.2 Calculating the Experience-Based Sophistication Measure

The EDGAR log file dataset is downloadable in daily pieces from the EDGAR website at the Security Exchange Commission (SEC), covers the period January 1st, 2003 - June 30th, 2017 with 26,482,889,754 entries, where each row represents an access to EDGAR by an end-user identified by a partly obfuscated IP address to an EDGAR filing identified by an accession number at a specific point in time. To protect the identity of the users underlying the IP addresses the log files provide the first three octets of the IP address with the fourth octet obfuscated with a 3 character string that preserves the uniqueness of the last octet (e.g. 255.255.255.abc).

We delete all entries with unsuccessful delivery by the EDGAR server, self-identified web crawlers, and observations that accessed only an index page but not a filing.[7] In addition to self-identified web crawlers, the data contains accesses likely made by non-self-identified web crawlers. Following Lee et al. (2015); Ryans (2017) we exclude IPs that access more than 25 filings or 3 different companies in a minute, or 500 filings in a single day. The final data set contains 844,369,939 entries.

IP addresses can either be assigned dynamically, as done by internet service providers ISP, or static as most company networks. Since our sophistication measure relies on individuals' research history, it is meaningful to filter out IPs that represent more than one institution. Hence, we define a further heuristic filtering procedure similar to the crawler elimination procedure above that is designed to make it unlikely that a dynamic IP remains in the sample. We require each IP to have

1. at least 5 distinct days with an access ("access day") to an arbitrary filing,

[7] That is, the http status code needs to be 200, the requested site must not be an index page (variable $idx = 0$) and the crawler dummy needs to be 0 in order to be not dropped.

2. at least 10 days between the first and the last access,

3. at least 1 % access days from all days between the first and last access.

This procedure leaves us with 2,214,358 IP addresses. A different choice of parameters in these filtering steps do not affect our results significantly as we demonstrate in robustness tests.

2.2.3 Sophistication Distributions

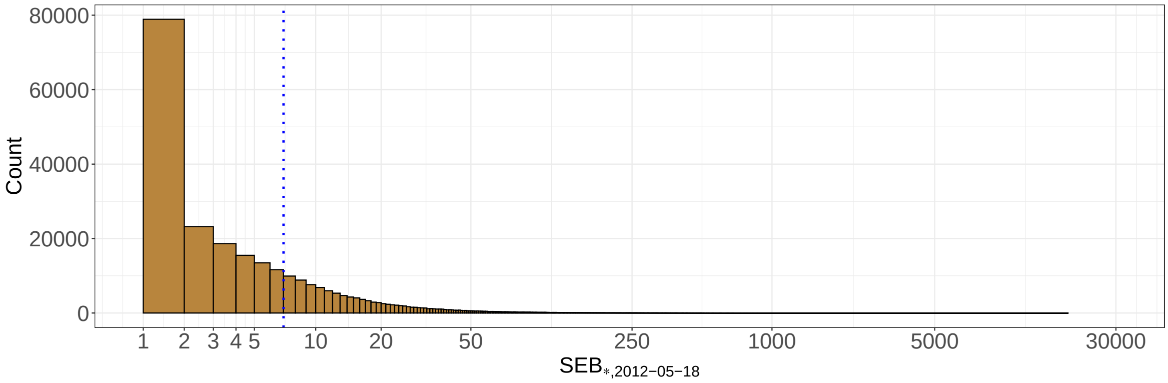
In Figure 2.1 we present an exemplary sophistication distribution showing sophistication levels $SEB_{*,t}$ of all active IP addresses on $t =$18th May, 2012, the day of the IPO of Facebook. An IP address is active if at least one filing was accessed within this period, that is $SEB_{*,t} > 0$. The histogram reveals a wide range of EDGAR activities, ranging from almost 42,000 IP addresses with only one filing accessed in the past year and only a few IP addresses who accessed more than 1,000 filings. The right tail of this distribution is truncated by the robot filtering procedure since it excludes IPs with many accesses in a short time. The maximum of 31,484 filings accessed corresponds to a user with approximately 125 filings accessed each day on average, assuming 250 working days per year. While 125 filings per day are considerable, it seems reasonable for a human charged solely with a data acquisition task.
For each possible day, we calculate a similar one-year sophistication distribution as in Figure 2.1. Figure 2.2 plots five sophistication quantiles q_t^α with $\alpha \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$ over time.[8] Despite a relatively long time period including different market regimes, increasing internet usage, and regulatory changes, also with respect to disclosure, the quantiles are stable over time.

2.2.4 Experience-Based Sophistication in the Context of Existing Sophistication (Attention) Measures

The literature generally distinguishes between direct and indirect measures of attention. Cook et al. (2006) are the first to empirically examine the role of an underwriter's

[8] We exclude all data prior to 11th May 2007 due to data errors from September 25th, 2005 to May 11th, 2006, and our requirement of at least one year of EDGAR activity for sophistication measurement.

Figure 2.1: Exemplary Sophistication Distribution



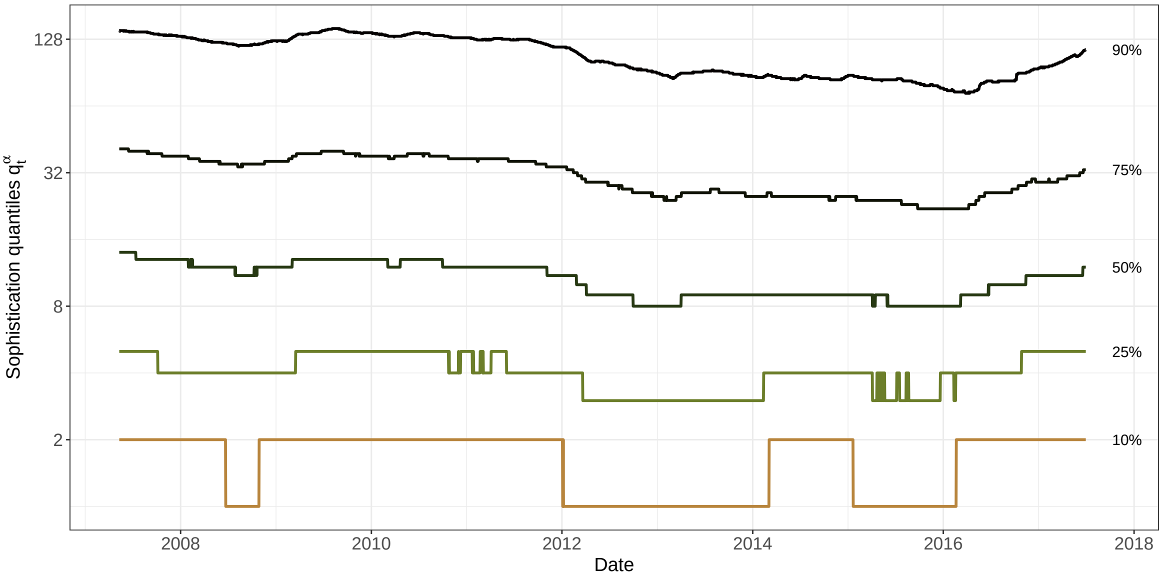
Notes: Exemplary distribution of $SEB_{*,t}$ for t = 18th May, 2012, when Facebook went public. $SEB_{i,t} = SEB^{365}_{i,t}(T_\Omega)$ corresponds to the sum of daily unique EDGAR filings accessed by IP i in the year before t. The dotted blue line shows the median of the distribution (= 9), which is one way to classify EDGAR-users into sophisticated and unsophisticated.

ability to market an IPO and its consequences. They proxy marketing ability by the extent of media coverage prior to an IPO, which indirectly captures investor attention by some degree. Both Cook et al. (2006) and Liu et al. (2014) show that this alternative measure of attention also predicts first-day IPO returns, though they differ in their interpretation of what type of attention from which group of investors they capture.

The possibly most common measure of direct attention builds upon Google's search volume index (SVI) (Da et al., 2011; Drake et al., 2012), and is currently used in over 60 published papers according to a literature review conducted by DeHaan et al. (2019). However, DeHaan et al. (2019) estimate that 69% of all S&P500 ticker searches go to websites that do not contain investing information, emphasizing that SVI is a noisy proxy for investor attention. The ambiguous mapping between company and attention is one disadvantage media coverage calculation is also prone to. We discuss further limitations of SVI and explain why SVI is not effectively capturing attention induced effects in the cross-section of IPOs.

While Google SVI is likely to capture mostly retail investor attention, Ben-Rephael et al. (2017) use Bloomberg terminal searches to proxy for institutional attention. Similar to SVI, the Bloomberg measure is formulated as an indicator for abnormal

Figure 2.2: Temporal Development of Different Sophistication Quantiles



Notes: Illustration of the temporal development of sophistication quantiles q_t^α for different quantile levels $\alpha \in \{10\%, 25\%, 50\%, 75\%, 90\%\}$. The sophistication quantiles belong to the sophistication distribution, which builds on $SEB_{*,t}$. $SEB_{i,t} = SEB_{i,t}^{365}(T_\Omega)$ corresponds to the sum of daily unique EDGAR filings accessed by IP i in the year before t.

attention as it builds upon aggregated and obfuscated underlying data.

We are convinced that the EDGAR log files are the best available source for revealed investor attention due to the unique level of detailedness of the EDGAR database that allows a direct, granular mapping of attention from an IP address to a specific piece of information of a company, for example its preliminary IPO prospectus (Form S-1). Our view is supported by numerous studies mapping a specific group of entities, e.g. mutual funds (Chen et al., 2020), hedge funds (Crane et al., 2018), or analysts (Gibbons et al., 2020), to the IP addresses. However, these studies focus on answering research questions related to the respective entity group and do not intend to measure different sophistication levels.

To do so, a recent paper by Drake et al. (2020) maps all available IP addresses to their underlying entities and separates the entities into a sophisticated and unsophisticated category based on reasonable assumptions about the entities. We show that our approach strongly relates to Drake et al. (2020) and extends their measurement. Our approach does not require a cumbersome construction of a large IP database through which researchers face significant challenges arising from combining several databases, assessing historic IP-to-entity mappings, and string-matching tasks. These difficulties do not only complicate a correct mapping between IP address and entity but also impede replication of the results.

The approach proposed in this paper is solely built upon information readily available in the EDGAR log files. It is computationally inexpensive and robust, e.g. we find that bypassing the filtering steps for both robot searches and dynamic IPs, and the construction of daily one-year sophistication distributions, by a smart choice of upper and lower thresholds for the sophistication groups, yields to qualitatively similar sophistication classifications.

However, there are some issues that may disturb the sophistication measurement. We acknowledge that information acquisition, in general, is nearly always the result of some economic stimulus that induces the acquirer to seek information and is hence non-random. Further, as also discussed in Drake et al. (2015), investors can obtain the same information from alternative sources, or gather some of the information from websites that summarize the information contained in regulatory filings. A single IP can represent the information demand of many users. Among the possible reasons are that the IP represents a Virtual Private Network (VPN), that a computer is used by more than a single user, or that a single user executes a search request for other users.[9] Last, even if the IP is in principle only used by a single user, the user can change in arbitrary intervals for instance due to technical changes or new staff.

The latter two mentioned issues are more likely to occur for company users, but likely to be rather rare. In addition, we calculate our sophistication measure using a rolling window scheme with a daily shift and a window size of one year. This reduces the impact of all temporally changes to users.

2.3 Price Pressure: Unsophisticated Attention to Initial Public Offerings

Starting with Ibbotson (1975), IPO underpricing has been well documented in the literature and various economic explanations have been discussed. The most recognized models propose compensation mechanisms. Rock (1986) suggests compensation for uninformed investors increasing in the riskiness of the offer. Instead, Benveniste and Spindt (1989) suggest compensation for (institutional) investors as a reward for a truthful revelation of valuations during the IPO process. Somewhat later, the joint role of long-term performance and underpricing has been investigated (Loughran and Ritter, 1995) partly questioning true “underpricing” (Purnanandam and Swaminathan, 2004) since IPOs underperform on average. With a special focus on hot markets Ljungqvist et al. (2006) propose a model to explain empirically observed IPO patterns for first-day return and long-term performance jointly. In their model irrationally exuberant investors combined with short-sale constraints lead to long-run underperformance.[10]

In general, sentiment investors are likely to be related to the (retail) attention via Google searches to IPOs as investigated in Da et al. (2011) who test the attention and price pressure theory from Barber and Odean (2008) by focusing on short-term and

[9] Regarding the first, we refer to Drake et al. (2015) who also state that VPNs are sometimes used in companies. Regarding the latter, we find specific IPs in the EDGAR logs who access EDGAR quite steadily in a typical nine-to-five time frame.

[10] Researchers have often assumed short-sale constraints in the early IPO aftermarket. However, Edwards and Hanley (2010) report that for a sample of IPOs in 2005 and 2006 all but two IPOs had short sales on the first trading day, and short sales accounted for 12 % of the trading volume. Hence, as opposed to earlier beliefs, short-sale constraints might not play a major role in explaining initial returns. Nevertheless, increased short-sale difficulty still might play a role for IPOs, which makes them better suited than seasoned stocks to test the price pressure hypothesis.

long-term pricing of 185 IPOs from 2004 to 2007. According to Barber and Odean (2008), the impact of attention towards buy decisions is more pronounced for retail investors since they have not the resources to screen all potential investment options but reduce their set to those stocks grabbing their attention, producing increased demand for such stocks. We test this theory and expect IPOs with larger attention to have higher initial returns in general but that this effect is more pronounced for attention from unsophisticated investors. Further, we follow Da et al. (2011) and identify IPOs subject to price pressure with both high unsophisticated attention and high first-day returns and expect that these IPOs underperform to a higher extent.

2.3.1 Sample Construction

Attention Variables

We use the experience-based sophistication measure from Section 2.2 to define EDGAR-attention from a sophisticated and unsophisticated group of EDGAR-users. We use each IPO's first trading day t_j on CRSP as the reference date and focus on the seven previous days. This time span corresponds with the literature, e.g. Da et al. (2011), and guarantees that a sufficient amount of attention can be measured, that the attention is contemporary and hence relevant to the issue, and reduces dependencies on day-of-the-week effects.

Denote with $\mathbf{A}_j = \{i : |\bigcup_{d=t_j-7}^{t_j-1} F_{i,d}(T_j)| > 0\}$ all IP addresses with at least one access to one of the IPO filings T_j of firm j within the period $t_j - 7, \ldots, t_j - 1$. Then we define the overall attention to these filings as $N_j = |\mathbf{A}_j|$ and sophistication-constrained attention as $N_j(s_j) = |\{i \in \mathbf{A}_j : SEB_{i,t_j-7} \in s_j\}|$.

For our main specification we choose the median $q_{t_j-7}^{0.5}$ of the active user sophistication distribution as illustrated in Figure 2.1 of Section 2.2 to distinguish sophistication groups. These medians are between 8 and 14. More detailed, we use $s_j = [0, q_{t_j-7}^{0.5}]$ to measure unsophisticated attention and $s_j = (q_{t_j-7}^{0.5}, \infty)$ for sophisticated attention. T_j are all filings from firm j of form type S-1 (initial prospectus) or S-1/A (amendments).

These raw user counts cover a wide range of our sample IPOs. Quite a few IPOs generate barely any attention on EDGAR while others instead create immense user counts. We conjecture that the size of the offering explains much of the observed attention since underwriters need to attract investors' attention in a magnitude related to an IPO's size (Bauguess et al., 2018) and scale the raw user counts by proceeds. In further tests, we employ two other ways to investigate differences between attention from both sophistication groups. First, we use the *proportion of unsophisticated attention* $U_j^{\%} = N_j[0, q_{t_j-7}^{0.5}]/N_j$ of each IPO firm j. Second, we test also residuals U_j^{Abn} from a regression of N_j on $N_j[0, q_{t_j-7}^{0.5}]$ as a measure of *abnormal unsophisticated attention.*

IPO Sample

Our IPO list is extracted from Thomson Financial's SDC New Issues database with additional items and corrections supplied by Professor Jay Ritter. Although the EDGAR log file data is available from 2003 onwards, data errors from September 25th, 2005, to May 11th, 2006, and our requirement of at least one year of EDGAR activity for sophistication measurement, restrict us to a sample covering roughly ten years from May 11th, 2007, to June 30th, 2017.[11] We exclude offerings that are associated with limited partnerships, closed-end funds, units, financial companies, real estate investment trusts, and dual-class capital structures or have an offer price less than \$5, all of which are typically excluded from IPO studies.

We merge the SDC IPOs to stock data from CRSP, to accounting data from Compustat, and to EDGAR via the EDGAR master index file.[12] See Section A.1 of the Appendix for details on the sample construction process and particularly Table A.1 where we give a detailed overview of all variables' definitions and sources used throughout this study. We obtain a sample size of up to 794 IPOs.

Our dependent variables are first-day returns, calculated as the percentage change from offer to the first closing price, and benchmark adjusted post-IPO buy-and-hold returns from three respectively six to twelve months. Intervals closer to the IPO offering likely comprise lasting price pressure effects, and market-making and price stabilization efforts by lead underwriters, e.g. documented in Ellis et al. (2000) for several months after the offering. As benchmarks, we use the corresponding Fama-French 48 industry value-weighted portfolio, as well as the corresponding Fama-French 25 value-weighted

[11] See Bauguess et al. (2018) for similar exclusions. By omitting the period from January 1st, 2003, to September 25th, 2005, we miss 173 observations. These observations are dominated by IPOs with no unsophisticated attention due to low EDGAR usage volume before the data errors. As a robustness check, we include these observations and find our results to be robust.

[12] Corrections to SDC are available on `https://site.warrington.ufl.edu/ritter/ipo-data/`.

portfolio formed on size and book-to-market, and the CRSP value-weighted market portfolio. We acknowledge that the choice of an appropriate benchmark is nontrivial and follow recent literature on this choice (Liu and Wu, 2020). See Lowry et al. (2017) for a discussion on measurement on long-run performance of IPOs. For convenience, we present only results for the industry adjusted post-IPO returns from three to twelve months. The results for the other abnormal post-IPO returns are available upon request from the authors. Table 2.1 shows the descriptive statistics of the sample.

Table 2.1: Summary Statistics

Variable	Mean	Std. dev.	Min	Median	Max	Obs.
Attention Variables:						
$U^{\$}$	0.05	0.08	0	0.03	1.27	794
$S^{\$}$	1.6	1.38	0.02	1.37	14.78	794
$A^{\$}$	1.6469	1.4233	0.0674	1.4282	16.0476	656
$U^{\%}$	0.0267	0.0205	0	0.0225	0.15	656
U^{Abn}	-0.01	0.52	-1.52	0.01	1.25	656
Dependent Variables:						
First-Day Return	16.3%	28.4%	-56%	8.3%	206.7%	794
Abnormal post-IPO Return	-7.2%	52.8%	-116.5%	-12.8%	348.4%	734
Abnormal post-IPO Return$_{\text{Size, B/M}}$	-4.5%	52.4%	-106.5%	-10.5%	341.5%	724
Abnormal post-IPO Return$_{\text{Market}}$	-6%	53.6%	-107.3%	-12.1%	348.1%	793
Raw post-IPO Return	1.1%	55.8%	-96.5%	-4.8%	368.1%	793
Controls:						
log(Sales + 1)	3.99	2.44	0	4.35	11.56	656
Up revision	4%	6.8%	0%	0%	45.5%	656
log(Filing Range)	4.4	0.81	3.22	4.43	7.41	656
log(Proceeds)	4.69	0.96	1.39	4.57	9.68	656
VC dummy	0.56	0.5	0	1	1	656
Share overhang	2.91	1.91	0	2.52	15.81	656
Bookrunner Market Share	0.32	0.25	0	0.32	0.75	656
Debt over Assets	0.97	1.31	0.03	0.73	15.2	656
Positive EPS dummy	0.37	0.48	0	0	1	656
Pre-IPO $\bar{r}_{\text{Market}}$	23.53%	41.65%	-83.66%	18.38%	496.32%	656
Pre-IPO σ_{Market}	13.23%	5.77%	5.94%	11.88%	75.76%	656

Notes: This table presents summary statistics for our sample of 794 IPOs from 1st August, 2007, to 28th June, 2017. The sample includes only offerings satisfying the usual IPO sample selection criteria. See Section A.1 in the Appendix for definitions and sources of the variables as well as a description of the sample construction process.

2.3.2 Empirical Results

Evidence From Portfolio Sorts

We start with an analysis of portfolios formed on sample sorts for unsophisticated $U^{\$}$ and sophisticated attention per Dollar $S^{\$}$ and report mean first-day returns for the four portfolios in Figure 2.3 (a) to study the impact of sophistication-constrained attention on first-day returns. For both variables, we identify a significant increase in average first-day returns, which confirms the generally presumed relation between attention and first-day returns. Consistent with the attention-induced price pressure hypothesis, we find the increase in first-day returns to be significantly higher for IPOs with higher unsophisticated attention. The difference between the two portfolios is about twice as high as the difference of portfolio sorts on sophisticated attention (21.99%−10.66%=11.33% respectively 19.18%−13.47% = 5.71%).

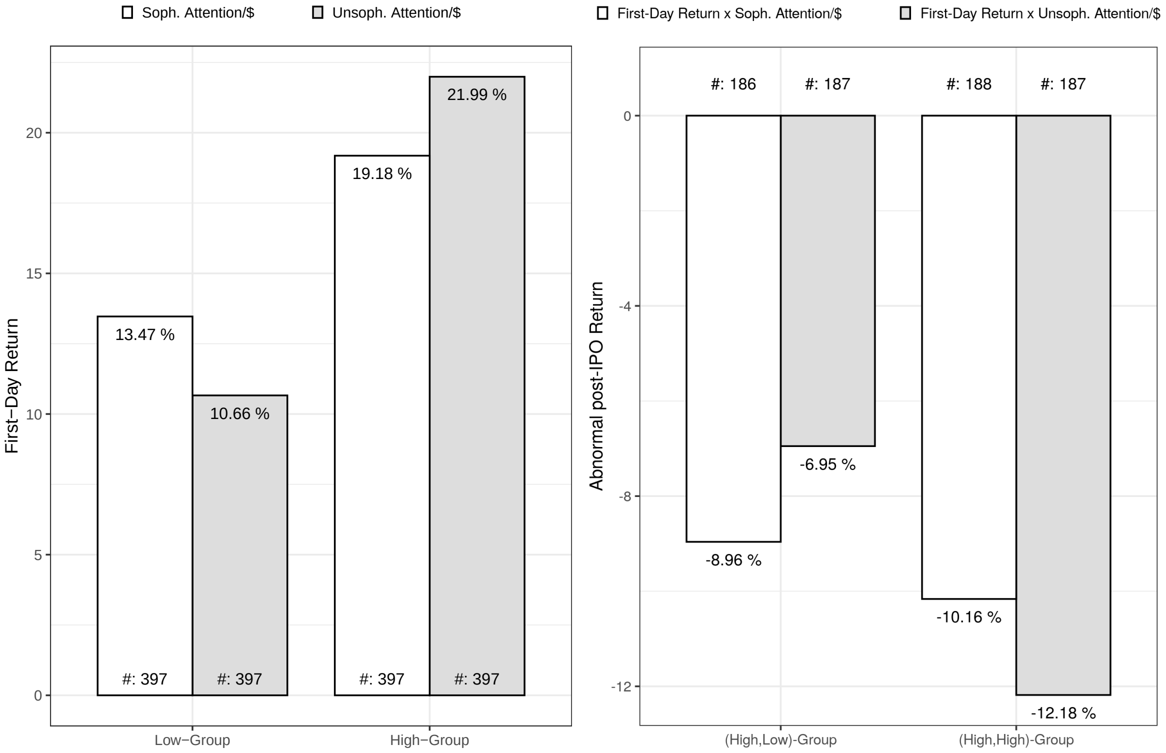
Studying the post-IPO returns we first sort on first-day returns and then on the attention measure resulting in eight portfolios from which we report the four belonging to high first-day returns in Figure 2.3 (b). We focus on high first-day returns since these ought to indicate increased price pressure combined with high unsophisticated attention. Consistent with the literature reporting long-run underperformance of IPOs (Loughran and Ritter, 1995; Brav et al., 2000) and the negative correlation between first-day and long-run returns also present in our sample (-8%), we find a significant price reverse in the long run for all portfolios. However, the portfolios sorted additionally on sophisticated attention exhibit almost identical post-IPO returns (-8.96% to -10.16%), while the portfolios sorted additionally on unsophisticated attention reveal a significant difference (-6.95% to -12.18%). These bivariate findings suggest the existence of price reversals induced by attention from unsophisticated investors.

Finally, we double-sort the sample on both sophistication-constrained attention variables in both possible orders and again report average first-day returns. The results in Figure 2.4 (a) show that after sorting for sophisticated attention, sorting for unsophisticated attention still leads to a significant increase in mean first-day returns, namely from 10.31% to 16.64% and from 12.11% to 26.3%. Both differences are significant at the 1% level. Sorting for unsophisticated attention first and then for sophisticated attention, presented in Figure 2.4 (b), we do not find significant differences in the average portfolio returns. Taking both double-sort orders into account suggests that

Figure 2.3: IPO Attention and Average IPO Returns

(a) First-Day Return

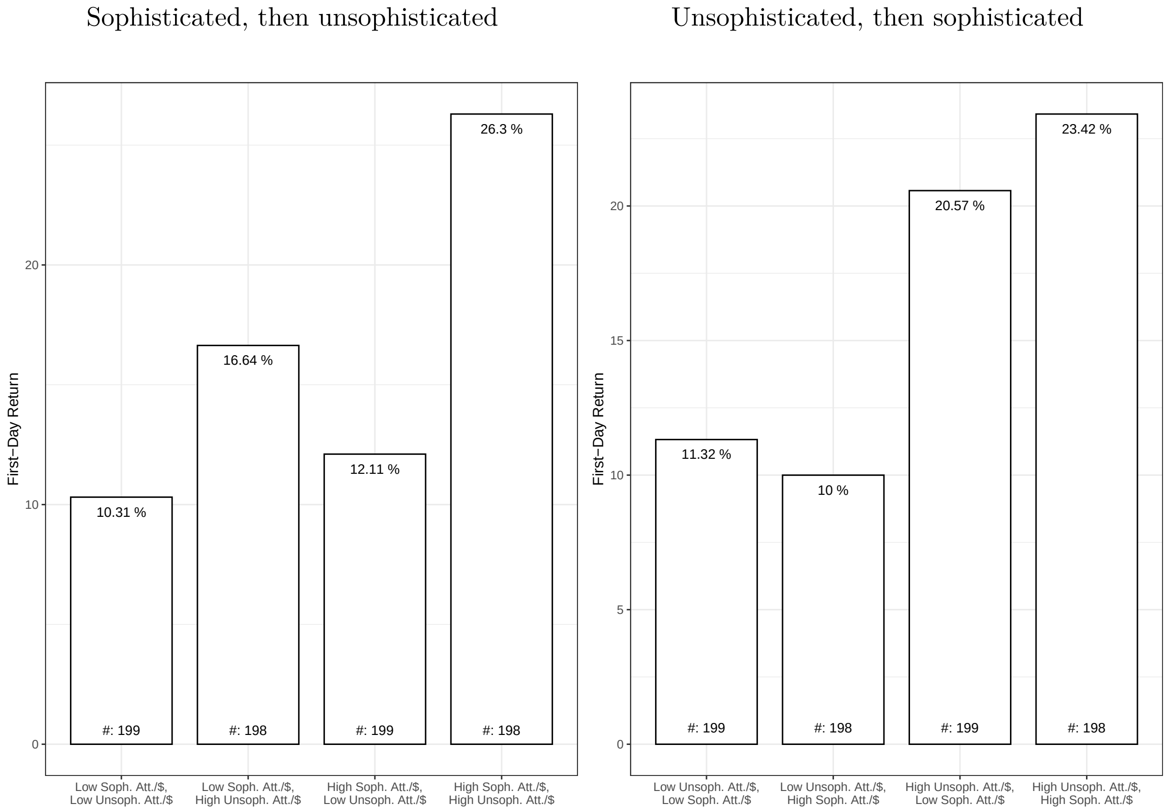
(b) Abnormal post-IPO Return



Notes: This figure presents first-day return averages of portfolios based on univariate sorts on attention variables in (a) and mean post-IPO return averages of portfolios based on double-sorts on first-day returns and attention variables in (b). The sample consists of 794 IPOs.

average first-day returns are boosted by unsophisticated attention, consistent with the price pressure hypothesis.

Figure 2.4: Average First-Day Returns and (Un-)sophisticated IPO Attention


Sophisticated, then unsophisticated

Unsophisticated, then sophisticated

Notes: This figure presents first-day return means of portfolios based on double-sorts on attention variables. In (a) IPOs are initially sorted on sophisticated attention $S^{\$}$ and then on $U^{\$}$ while the order is reversed in (b). The sample consists of 794 IPOs.

Multivariate Evidence

We proceed by formalizing regression models to provide evidence that the measured bivariate effects identified in the previous section are not driven by other IPO characteristics. The main regression model takes the form

$$FDR = \beta_0 + \beta_1\, U^{\$} + \beta_2\, S^{\$} + X\delta + \epsilon \tag{2.3.1}$$

where X denotes the design matrix whose rows correspond to the control variables, including Fama-French 48 industry and year fixed effects, and ϵ the error terms. The results are summarized in Table 2.2.

The first column shows the baseline regression, which exhibits typical associations found in the IPO literature. In line with almost any study explaining IPO characteristics, we find a strong association between higher first-day returns and upwards revised issues (Hanley, 1993). IPOs from venture capital backed companies (Lee and Wahal, 2004) or companies with positive trailing earnings are found to be associated with higher first-day returns as well. Both the logarithm of trailing sales and leverage, defined as total debt scaled by total assets, negatively relate to initial returns (Butler et al., 2014).

From columns 2 to 5, we include the attention measures in different model specifications. While all signs of the control variables remain the same, the logarithm of proceeds and the venture capital dummy turn significant respectively insignificant in some specifications. Consistent with the bivariate findings, we find attention from the group of EDGAR-users below the median user to be a strong determinant of first-day returns. A one-standard-deviation (0.075) increase in attention from less active EDGAR-users leads to a 5.6% higher first-day return. In a regression together with attention from more active EDGAR-users, the latter even turns insignificant, suggesting that price pressure effects in new issues are driven by less sophisticated investors. The last column highlights the positive relation between attention in general and first-day returns.

To assess the existence of a price reversal among IPOs with high first-day returns and high unsophisticated attention, we follow Da et al. (2011) and include an interaction term between our attention measures and first-day returns in the regression model of abnormal post-IPO buy-and-hold returns and additionally control for the magnitude of the first-day returns:

$$APR = \beta_0 + \beta_1 U^{\$} + \beta_2 S^{\$} + \beta_3 U^{\$} \times FDR + \beta_4 S^{\$} \times FDR + \beta_5 FDR + X\delta + \epsilon \tag{2.3.2}$$

where X again denotes the design matrix featuring the control variables from the previous regression and ϵ the error terms. We summarize the results of eight different specifications in Table 2.3.

Table 2.2: (Un-)sophisticated pre-IPO Attention and IPO First-Day Returns

	Dependent variable: First-Day Return				
	(1)	(2)	(3)	(4)	(5)
$U^{\$}$		72.651***	74.764***		
		(2.917)	(4.902)		
$S^{\$}$		0.203		3.686***	
		(0.169)		(3.670)	
$A^{\$}$					3.663***
					(4.145)
Up revision	1.730***	1.613***	1.614***	1.659***	1.654***
	(5.281)	(5.770)	(5.816)	(5.531)	(5.555)
log(Filing Range)	−1.029	−0.249	−0.256	−0.505	−0.470
	(−0.960)	(−0.232)	(−0.242)	(−0.453)	(−0.423)
log(Proceeds)	2.021	3.726*	3.579	5.498**	5.552**
	(0.993)	(1.871)	(1.573)	(2.438)	(2.412)
VC dummy	5.472*	3.279	3.244	4.965**	4.859*
	(1.951)	(1.181)	(1.207)	(1.994)	(1.954)
Share overhang	0.658	0.475	0.487	0.355	0.349
	(0.968)	(0.702)	(0.756)	(0.540)	(0.536)
Bookrunner Market Share	2.624	1.313	1.335	1.562	1.505
	(0.682)	(0.357)	(0.357)	(0.438)	(0.422)
log(Sales)	−1.690***	−1.682***	−1.678***	−1.741***	−1.740***
	(−2.876)	(−2.964)	(−2.952)	(−2.734)	(−2.735)
Debt over Assets	−1.489***	−1.347***	−1.338***	−1.569***	−1.561***
	(−5.058)	(−4.640)	(−4.678)	(−4.925)	(−4.933)
Pos. EPS dummy	3.264**	2.398*	2.402*	2.743**	2.704**
	(2.480)	(1.930)	(1.957)	(1.995)	(1.971)
Pre-IPO $\bar{r}_{\text{Market}}$	4.442**	4.160**	4.157**	4.348**	4.334**
	(2.422)	(2.238)	(2.255)	(2.429)	(2.419)
Pre-IPO σ_{Market}	−11.879	−13.803	−13.730	−14.160	−14.236
	(−0.573)	(−0.686)	(−0.673)	(−0.706)	(−0.709)
Constant	19.394*	5.255	5.992	−0.860	−1.385
	(1.706)	(0.453)	(0.451)	(−0.084)	(−0.132)
Fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	656	656	656	656	656
Adjusted R^2	0.267	0.296	0.297	0.282	0.284
F Statistic	5.048***	5.508***	5.609***	5.293***	5.323***

Notes: This table shows regression results for the three attention measures unsophisticated attention $U^{\$}$, sophisticated attention $S^{\$}$ and overall attention $A^{\$}$ - each scaled by proceeds - to IPO filings on EDGAR in the pre-IPO week on first-day returns. Detailed variable definitions can be found in Section A.1 of the Appendix. The numbers in brackets below the coefficient estimates show t-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

Typically for studies drawing inference on long-term returns, we find relatively small adjusted R^2 ranging from 4.9% to 6.2% throughout all specifications. Summing up the effects of the control variables, we find new issues of companies with positive trailing earnings to have an on average 12% higher abnormal returns from three months to one year after the offering (Field and Lowry, 2009). We further document a positive (but insignificant slightly below the 10% significance level) effect of reputable underwriters on post-IPO performance as initially suggested by Carter and Manaster (1990). Higher levels of debt relative to assets before an offering is associated with significantly lower post-issue returns. In line with the predictions, the negative linkage between first-day returns and long-term underperformance dissipates in specifications where we included the interaction terms with our attention variables.

Column 4 of Table 2.3 shows the regression as formulated in Equation 2.3.2 for $\beta_4 = 0$. While controlling for attention from sophisticated users, we find a significant, negative coefficient for the interaction term between first-day returns and attention from unsophisticated users. This effect holds when incorporating the interaction term between first-day returns and attention from more active users as shown in column 5. However, this model is subject to high variance inflation factors due to including first-day returns and two interactions with them, which likely explains the positive coefficient of the interaction between first-day returns and attention from more active users. Excluding $U^{\$}$ and its interaction $U^{\$}$ × *First-Day Return* leaves the interaction term $S^{\$}$ × *First-Day Return* insignificant, as shown in column 8. Recent studies suggest a positive association between the participation of sophisticated investors and stock price performance (Field and Lowry, 2009; Drake et al., 2020). Note that we do not find a significantly positive effect of attention by more active users on abnormal post-IPO returns up to one year.

Robustness

By now, we have classified EDGAR-users as (un-)sophisticated based on their trailing one-year continuity in accessing EDGAR-filings compared to the median $q^{0.5}_{t_j-7}$ of the distribution of all active EDGAR-users. Choosing a quantile, e.g. the median, has the appealing property that for each IPO the threshold is relative to the overall information acquisition level at that time. However, all choices in this respect are rather ad hoc

Table 2.3: (Un-)sophisticated pre-IPO Attention and post-IPO Returns

	Dependent variable: Abnormal post-IPO Return							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$U^{\$}$	11.024	80.431	−8.501	60.523	87.701			−17.403
	(0.320)	(1.318)	(−0.129)	(0.674)	(1.136)			(−0.199)
$U^{\$}$ × FDR		−0.920**		−0.927**	−3.254***			
		(−2.288)		(−2.386)	(−3.164)			
$S^{\$}$			1.877	1.964	−0.459	1.478	1.249	1.716
			(0.391)	(0.404)	(−0.186)	(0.552)	(0.495)	(0.399)
$S^{\$}$ × FDR					0.219**		0.005	0.013
					(2.377)		(0.126)	(0.228)
FDR	−0.085**	−0.020	−0.085**	−0.020	−0.277**	−0.088*	−0.099	−0.111
	(−2.331)	(−0.406)	(−2.258)	(−0.398)	(−2.378)	(−1.777)	(−1.355)	(−0.993)
Up revision	−0.088	−0.134	−0.093	−0.140	−0.057	−0.094	−0.090	−0.081
	(−0.234)	(−0.358)	(−0.256)	(−0.387)	(−0.161)	(−0.264)	(−0.245)	(−0.202)
log(Filing Range)	2.812	3.246	2.877	3.317	3.685	2.905	2.899	2.832
	(0.543)	(0.633)	(0.571)	(0.667)	(0.721)	(0.586)	(0.583)	(0.555)
log(Proceeds)	−1.421	−0.446	−0.057	0.989	0.819	−0.251	−0.399	−0.227
	(−0.274)	(−0.081)	(−0.018)	(0.285)	(0.222)	(−0.061)	(−0.105)	(−0.069)
VC dummy	1.349	1.203	1.674	1.542	1.960	1.489	1.430	1.719
	(0.231)	(0.203)	(0.315)	(0.285)	(0.411)	(0.245)	(0.237)	(0.331)
Share overhang	−1.089*	−1.423**	−1.198*	−1.540**	−1.202**	−1.184*	−1.149*	−1.126*
	(−1.738)	(−2.305)	(−1.804)	(−2.341)	(−2.044)	(−1.769)	(−1.922)	(−1.927)
Bookrunner Market Share	11.715	13.049	11.512	12.846	12.292	11.486	11.381	11.274
	(1.515)	(1.642)	(1.520)	(1.647)	(1.589)	(1.522)	(1.461)	(1.453)
log(Sales)	0.898	0.936	0.868	0.905	0.861	0.871	0.869	0.860
	(0.377)	(0.387)	(0.370)	(0.379)	(0.364)	(0.370)	(0.368)	(0.369)
Debt over Assets	−3.422***	−3.246***	−3.502***	−3.329***	−3.020***	−3.480***	−3.474***	−3.510***
	(−4.142)	(−4.003)	(−4.515)	(−4.414)	(−3.705)	(−4.206)	(−4.257)	(−4.494)
Pos. EPS dummy	12.957***	12.777***	12.918***	12.735***	12.573***	12.884***	12.877***	12.936***
	(2.868)	(2.823)	(2.918)	(2.872)	(3.052)	(2.928)	(2.936)	(2.914)
Pre-IPO $\bar{r}_{\text{Market}}$	−13.153**	−13.240**	−13.125**	−13.211**	−13.248**	−13.137**	−13.138**	−13.115**
	(−2.038)	(−2.080)	(−2.069)	(−2.112)	(−2.306)	(−2.056)	(−2.058)	(−2.075)
Pre-IPO σ_{Market}	60.724**	61.742**	60.041**	61.036**	60.328**	60.049**	59.975**	59.846**
	(2.244)	(2.191)	(2.253)	(2.203)	(2.095)	(2.236)	(2.218)	(2.245)
Constant	−34.784	−44.330	−41.591	−51.523	−48.922	−40.878	−39.917	−39.914
	(−0.853)	(−1.061)	(−1.247)	(−1.536)	(−1.332)	(−1.144)	(−1.168)	(−1.170)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	656	656	656	656	656	656	656	656
Adjusted R^2	0.051	0.053	0.050	0.052	0.062	0.052	0.050	0.049
F Statistic	1.578***	1.594***	1.557***	1.573***	1.681***	1.584***	1.557***	1.531***

Notes: This table shows regression results for unsophisticated attention $U^{\$}$ and sophisticated attention $S^{\$}$ - each scaled by proceeds - in the pre-IPO week on abnormal post-IPO returns from three months to twelve months after the offering. Detailed variable definitions can be found in Section A.1 of the Appendix. The numbers in brackets below the coefficient estimates show *t*-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

and not compelling. Hence, we test four alternative ways of forming sophistication groups, which are subsumed in Table 2.4, and analyze how the changes impact the results related to price pressure.

Table 2.4: Alternative Sophistication Classifications

Case	Unsophisticated s_j	Sophisticated s_j	Attention Vars.
Base	$[0, q^{0.5}_{t_j-7}]$	$(q^{0.5}_{t_j-7}, \infty)$	$U^\$, S^\$$
A	$[0, 25]$	$(25, \infty)$	$U^\$[0, 25], S^\$(25, \infty)$
B	$[0, 25]$	$(200, \infty)$	$U^\$[0, 25], S^\$(200, \infty)$
C	$(5, 25]$	$(200, 1000]$	$U^\$(5, 25], S^\$(200, 1000]$
D	$(25, 200]$	$(200, 1000]$	$U^\$(25, 200], S^\$(200, 1000]$

Notes: This table presents four alternative sophistication classifications A-D to our base specification. The set s_j represents all admissible sophistication levels for IPO j. Note that our baseline classification relies on a time-variant threshold (median of the sophistication distribution, cf. Figure 2.1 and 2.2) while cases A-D are constant over time.

Case A is relatively similar to our base case since the chosen fixed threshold of 25 is somewhat higher than the average median of all sophistication distributions, see Figure 2.2. One benefit of this choice is its concrete nature as it is constant for all IPOs and hence also easier to implement. Case B introduces a considerable gap between the most sophisticated user possible in the unsophisticated group and the most unsophisticated user possible in the sophisticated group by shifting the lower threshold for the sophisticated group to 200. It may be meaningful to assume that users in the middle of the distribution are better to be omitted instead of classified into one of the groups. Case C is a truncated version of Case B in order to check the heuristic filtering procedures applied to the log file data. It may be the case that some IP addresses are still dynamic, which would likely be those with the lowest sophistication. Hence, we require a sophistication level of at least 5 for a user to be included. Additionally, despite the robot cleaning procedure applied, it may happen that some of the high sophistication levels are due to robots or VPNs. Hence, we exclude all too sophisticated IP addresses above 1,000. Finally, Case D checks the importance of a relatively low upper threshold for the unsophisticated group by allowing sophistication levels of up to 200 for unsophisticated users.

Using these alternative attention measures we repeat the main analysis. Results for each case are presented in Table 2.5 for regressions on first-day returns and in Table 2.6 for regressions on post-IPO returns.

For Case A and Case B both tables reveal no material differences to our base regressions. In other words, a constant threshold for discrimination between sophistication groups, which is also somewhat larger than the average median, yields qualitatively similar effects on first-day and post-IPO returns. The same is true for omitting users from the middle of the distribution, where the classification is more ambiguous. Case C shows an improvement in terms of both magnitude of estimates and significance. This suggests that omitting users in both tails of the sophistication distribution removes noise from the data, which may come from not flawless heuristic filtering approaches. While in Case D the relation between unsophisticated attention and first-day returns gets more significant, the effect on post-IPO returns vanishes. This shows that the upper threshold cannot be shifted arbitrarily high due to more and more inclusion of in fact rather experienced users, which suggests that a meaningful upper threshold for unsophisticated users is somewhere below 200.

We perform three further robustness tests. First, we do not filter out IP addresses that are likely dynamic as we discuss in Section 2.2. The described IP filtering process is necessary since dynamic IP addresses change over time impeding to obtain a reliable snapshot of its historical EDGAR activity. Second, we use an expanded sample starting in January, 2003, and third, we base our sophistication proxy on a limited set of admissible filings containing “fundamental” information only. The results of these analyses are qualitatively similar to those of our main analysis and are available upon request from the authors.

Table 2.5: (Un-)sophisticated pre-IPO Attention Based on Alternative Sophistication Classifications and IPO First-Day Returns

	Dependent variable: First-Day Return		
	(1)	(2)	(3)
Case A			
$U^{\$}[0, 25]$	37.178*** (3.005)	33.215*** (4.377)	
$S^{\$}(25, \infty)$	−0.945 (−0.670)		3.697*** (3.101)
Adjusted R^2	0.299	0.300	0.280
Case B			
$U^{\$}[0, 25]$	38.009*** (3.217)	33.215*** (4.377)	
$S^{\$}(200, \infty)$	−1.592 (−0.944)		3.979** (2.336)
Adjusted R^2	0.300	0.300	0.276
Case C			
$U^{\$}[5, 25]$	37.376*** (2.679)	41.072*** (4.596)	
$S^{\$}(200, 1000]$	2.857 (0.541)		18.351*** (4.805)
Adjusted R^2	0.297	0.298	0.286
Case D			
$U^{\$}[25, 200]$	20.003*** (4.398)	20.443*** (8.128)	
$S^{\$}(200, 1000]$	0.531 (0.106)		18.351*** (4.805)
Adjusted R^2	0.291	0.293	0.286

Notes: This table presents the impacts of different attention measures on first-day returns of IPOs. Unsophisticated attention per Dollar $U^{\$}(*)$ and sophisticated attention per Dollar $S^{\$}(*)$ towards EDGAR filings of IPOs in the pre-IPO week are used in four different variants (Case A - D) of defining sophistication groups. While control variables are included in the regressions as in Table 2.2, we omit their results for brevity. The numbers in brackets below the coefficient estimates show t-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

Table 2.6: (Un-)sophisticated pre-IPO Attention Based on Alternative Sophistication Classifications and post-IPO Returns

	Dependent variable: Abnormal post-IPO Return							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Case A								
$U^{\$}[0,25]$	−0.611	21.975	−16.795	4.876	17.982			−32.306
	(−0.055)	(1.084)	(−0.736)	(0.155)	(0.736)			(−0.989)
$U^{\$}[0,25]$ × FDR		−0.284**		−0.248*	−1.227***			
		(−2.096)		(−1.775)	(−3.194)			
$S^{\$}(25,\infty)$			3.839	3.379	0.646	1.789	1.300	3.644
			(0.755)	(0.646)	(0.297)	(0.604)	(0.499)	(0.835)
$S^{\$}(25,\infty)$ × FDR					0.251***		0.013	0.054
					(2.705)		(0.250)	(0.809)
FDR	−0.078**	−0.028	−0.075*	−0.032	−0.305***	−0.088*	−0.113	−0.170
	(−1.975)	(−0.609)	(−1.841)	(−0.667)	(−2.797)	(−1.756)	(−1.354)	(−1.493)
Adjusted R^2	0.051	0.051	0.051	0.051	0.063	0.052	0.050	0.051
Case B								
$U^{\$}[0,25]$	−0.611	21.975	−14.068	7.231	19.848			−33.614
	(−0.055)	(1.084)	(−0.708)	(0.244)	(0.843)			(−1.153)
$U^{\$}[0,25]$ × FDR		−0.284**		−0.243*	−1.095***			
		(−2.096)		(−1.694)	(−3.325)			
$S^{\$}(200,\infty)$			4.430	3.803	0.375	2.418	1.191	4.039
			(0.747)	(0.619)	(0.155)	(0.626)	(0.399)	(0.861)
$S^{\$}(200,\infty)$ × FDR					0.311***		0.038	0.098
					(2.773)		(0.487)	(1.070)
FDR	−0.078**	−0.028	−0.074*	−0.032	−0.328***	−0.087*	−0.148	−0.215*
	(−1.975)	(−0.609)	(−1.830)	(−0.664)	(−2.934)	(−1.707)	(−1.380)	(−1.674)
Adjusted R^2	0.051	0.051	0.051	0.051	0.064	0.052	0.051	0.053

(Continued on next page.)

Notes: This table presents the impacts of different attention measures on abnormal post-IPO returns from three to twelve months. Unsophisticated attention per Dollar $U^{\$}(*)$ and sophisticated attention per Dollar $S^{\$}(*)$ towards EDGAR filings of IPOs in the pre-IPO week are used in four different variants (Case A - D) of defining sophistication groups. While control variables are included in the regressions as in Table 2.3, we omit their results for brevity. The numbers in brackets below the coefficient estimates show t-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

Table 2.6: (Un-)sophisticated pre-IPO Attention Based on Alternative Sophistication Classifications and post-IPO Returns (Continued)

	Dependent variable: Abnormal post-IPO Return							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(Continued.)								
Case C								
$U^{\$}[5,25]$	−2.248	20.279	−1.918	21.027	29.010			−9.065
	(−0.189)	(1.011)	(−0.060)	(0.605)	(0.999)			(−0.267)
$U^{\$}[5,25]$ × FDR		−0.303**		−0.303**	−1.451**			
		(−2.203)		(−2.238)	(−2.320)			
$S^{\$}(200,1000]$			−0.256	−0.564	−7.370	−1.035	−2.964	−0.768
			(−0.011)	(−0.024)	(−0.444)	(−0.085)	(−0.183)	(−0.033)
$S^{\$}(200,1000]$ × FDR					0.924*		0.047	0.084
					(1.944)		(0.246)	(0.436)
FDR	−0.076*	−0.031	−0.076*	−0.031	−0.238**	−0.077	−0.099	−0.111
	(−1.836)	(−0.713)	(−1.783)	(−0.691)	(−2.220)	(−1.464)	(−1.610)	(−1.461)
Adjusted R^2	0.051	0.051	0.049	0.049	0.056	0.051	0.049	0.048
Case D								
$U^{\$}[25,200]$	4.959	11.472	21.635	29.245	44.405**			23.065
	(0.492)	(0.865)	(0.824)	(1.437)	(2.468)			(1.004)
$U^{\$}[25,200]$ × FDR		−0.117		−0.126	−1.316*			
		(−0.853)		(−0.969)	(−1.700)			
$S^{\$}(200,1000]$			−20.124	−20.846	−35.299*	−1.035	−2.964	−19.932
			(−0.679)	(−0.718)	(−1.677)	(−0.085)	(−0.183)	(−0.660)
$S^{\$}(200,1000]$ × FDR					1.418*		0.047	−0.036
					(1.723)		(0.246)	(−0.229)
FDR	−0.088*	−0.044	−0.087*	−0.041	−0.210***	−0.077	−0.099	−0.072
	(−1.931)	(−1.066)	(−1.943)	(−1.014)	(−2.597)	(−1.464)	(−1.610)	(−1.265)
Adjusted R^2	0.051	0.051	0.051	0.051	0.055	0.051	0.049	0.050

Notes: This table presents the impacts of different attention measures on abnormal post-IPO returns from three to twelve months. Unsophisticated attention per Dollar $U^{\$}(*)$ and sophisticated attention per Dollar $S^{\$}(*)$ towards EDGAR filings of IPOs in the pre-IPO week are used in four different variants (Case A - D) of defining sophistication groups. While control variables are included in the regressions as in Table 2.3, we omit their results for brevity. The numbers in brackets below the coefficient estimates show t-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

2.3.3 Empirical Results for Alternative Attention Measures

Based on our financial statement experience measurement proposed in Section 2.2.1, we discuss two alternative ways to construct variables for attention generated by rather unsophisticated EDGAR-users. First, we focus on the *relative proportion* of unsophisticated attention relative to overall attention. Second, we consider a measure of *abnormal* unsophisticated attention. Information on the descriptives can be found in Table 2.1.

Relative Proportion of Unsophisticated Attention

We calculate the proportion of unsophisticated attention as the number of EDGAR-users left of the median of the sophistication distribution divided by the number of all users who accessed a Form S-1 and S-1/A filing within one week prior to the IPO date. $U^{\%}$ is by definition a number between zero and one, easy to interpret, and captures disparities in attention by less and more active users in one measure. Bivariate correlations between this measure and the other variables are highest for first-day returns, upwards revisions, and proceeds. In a regression model analogous to Equation 2.3.1, we find a one-standard-deviation (0.0205) increase in this ratio to be associated with 3.76% higher first-day returns. Again, these issues experience a significant price reverse from three to twelve months after the offering as tabulated in Panel A of Table 2.7.

Abnormal Attention by Unsophisticated Users

Given a general attention level for an IPO, we expect a proportion of this attention to be driven by unsophisticated investors. Great deviations from this expectation define offerings with considerably low or high attention from unsophisticated investors. Following, we regress the number of all users on the number of unsophisticated users who accessed an IPO's Form S-1 or S-1/A within one week before the offering, both in their logarithm and including time and industry fixed effects:

$$\log(N[0, q^{0.5}_{t-7}]) = \beta_0 + \beta_1 \log(N) + \alpha + \gamma + \epsilon \tag{2.3.3}$$

where α and γ denote Fama-French 48 industry respectively time fixed effects and ϵ the residuals. The residuals of this regression comprise the number of unsophisticated

EDGAR-users that cannot be explained by the overall attention level, neither by industry nor time. We denote these as U^{Abn} and include them in regression models analogous to Equations 2.3.1 and 2.3.2. U^{Abn} is positively correlated with first-day returns, upwards revisions, and share overhang. Panel B of Table 2.7 confirms that abnormal unsophisticated attention is associated with a short-run price pressure effect as measured by higher first-day returns, followed by a price reversal after the offering.

Table 2.7: Alternative pre-IPO Attention Measures and IPO Returns

Dependent variable:	*FDR*	*Abnormal post-IPO Return*	
	(1)	(2)	(3)
Panel A: Relative proportion of unsophisticated attention			
$U^{\%}$	183.502***	−93.560	9.440
	(2.618)	(−1.344)	(0.120)
$U^{\%}$ × FDR			−4.806***
			(−4.248)
FDR		−0.069	0.107*
		(−1.534)	(1.870)
Controls	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes
Observations	656	656	656
Adjusted R^2	0.280	0.052	0.054
Panel B: Abnormal attention by unsophisticated EDGAR-users			
U^{Abn}	4.355***	−7.111*	−4.368
	(3.927)	(−1.763)	(−1.148)
U^{Abn} × FDR			−0.167***
			(−3.864)
FDR		−0.065	−0.040
		(−1.424)	(−0.962)
Controls	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes
Observations	656	656	656
Adjusted R^2	0.272	0.055	0.056

Notes: This table presents regression results for the proportion of unsophisticated attention $U^{\%}$ on first-day and abnormal post-IPO returns in Panel A. Results for abnormal unsophisticated attention U^{Abn} are shown in Panel B. While control variables are included in the regressions as in Table 2.2 and Table 2.3, respectively, we omit their results for brevity. The numbers in brackets below the coefficient estimates show *t*-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

2.4 Empirical Relation to Existing Sophistication (Attention) Measures

At least since Barber and Odean (2008), the list of suggested proxies for investor attention has rapidly been growing. Since some of these measures are quite close to ours in certain aspects, we perform comparisons to two alternatives in this section. First, we focus on the EDGAR-logs-based attention measure by Drake et al. (2020), which differentiates sophistication levels via IP matching to specific entities. Second, we turn towards the widely used Google-based attention measure initially proposed by Da et al. (2011).

2.4.1 The Experience vs. Entity-Matching Approach

Recently, Drake et al. (2020) propose (un-)sophisticated attention measures from EDGAR that are similar in spirit to us but differ in how sophistication is measured. They create a large IP database that matches IPs to entities such as internet service providers, banks, or funds. Then they separate the entities into a sophisticated and unsophisticated category based on reasonable assumptions. For instance, an EDGAR hit from Goldman Sachs reflects likely the attention of an investment professional whereas a hit from the internet service provider AT&T reflects likely attention from a non-professional retail investor. We denote this idea *entity-matching* sophistication subsequently. Using hits of 10-K and 10-Q filings Drake et al. (2020), find that sophisticated attention is predictive for variables associated with future firm performance as for instance future abnormal returns or future earnings announcement news whereas unsophisticated attention is not.

Both the approach proposed in Drake et al. (2020) as well as our have pros and cons. While entity-matching in general and also specifically for the purpose of judging sophistication may be useful it comes also with some difficulties. First, the construction of a large and reliable IP-entity database is quite time-consuming and not always unambiguous. The approach requires an extensive number of Whois requests and the mapping between a Whois report and an entity is often equivocal. Further, the owner of an IP can change over time, which is why services like WhoWas exist, which makes the entity-matching task more complicated. Second, in addition to these rather technical aspects, it is not clear which entities are sophisticated and which are not (especially for companies the human classifier is unfamiliar with). Further, and rather tautological, one could think of sophisticated individuals whose main internet connection is via an internet service provider. Some professionals might access EDGAR not exclusively from their offices but instead also from home.

The approach presented in this paper is self-contained, hence needs no cumbersome matching and allows differentiating further between entities in a data-driven manner. However, while we argue that the approach in Drake et al. (2020) likely over-classifies unsophisticated EDGAR-users, we agree that in some situations our approach might over-classify sophisticated users. For example, empirical researchers who regularly access EDGAR to pull disclosure data would be classified as sophisticated. Yet, it is unlikely to find such observations in samples that measure prompt attention towards an event.

We proceed by comparing *sophistication by experience* with *sophistication by entity* on 1,227,745 pre-issue EDGAR-hits for 656 IPOs.[13] In order to get an impression of the relation between the two approaches, we build an IP database containing IP blocks belonging to 20 important real-world entities as AT&T and Wells Fargo. We use both the Top-10 most frequent sophisticated and unsophisticated entities according to Drake et al. (2020).

For each EDGAR hit, we match the corresponding IP to our IP database and calculate the proportion of hits from Top-10 internet service providers (ISPs) and Top-10 institutions. We perform this again for the subset of hits from less and more EDGAR-experienced IPs. The results can be found in Table 2.8.

We are able to identify the origin of approximately 34.6 % of all pre-IPO-issue hits with only 20 entities.[14] With 32.9 % of all hits internet service providers make up most of it while only 1.68 % come from one of the institutions.

Subsetting the hits to less experienced IPs reveals that almost no institutional IP is “unsophisticated-by-experience”, which supports the view that institutions are often “sophisticated-by-entity”. In addition, relatively more ISP-IPs are in this group compared to all hits, which is in line with the notion that retail investors tend to be unsophisticated.

[13] See Section 2.3.1 for sample details.

[14] For two entities we do not find any match.

Table 2.8: Comparison of Sophistication Classifications

	Hits from Top-10 ISPs	Hits from Top-10 Inst.
All Hits	0.32934	0.01695
Hits from rather ...		
unexperienced IPs	0.35612	0.00058
experienced IPs	0.25342	0.03665

Notes: This table shows proportions of the four different subgroups "Top-10 ISPs", "Top-10 Institutions", "unexperienced IPs", "experienced IPs" and combinations thereof for 1,227,745 EDGAR hits from a sample of 654 IPOs.

Probably rather surprising, we find that among the experienced hits a substantial portion still comes from ISPs. Classifying them as unsophisticated may be misleading since their EDGAR hits show that they do considerable research using EDGAR. As expected, the proportion of hits from Top-10 institutions is substantially larger in this case.

All in all these results suggest that our financial statement experience measure is related to the underlying entity in an expectable manner but enables to differentiate the IPs in a less burdensome manner. A combination of the two approaches suggests promising opportunities for future research.

2.4.2 Unsophisticated EDGAR-Users vs. Google Searches

Before Da et al. (2011), attention measures were typically not directly linked to a specific investor group. However, since search engines in general and Google specifically are easy to use by anyone and since the plain number of retail investors should exceed the number of institutional investors it seems obvious to interpret Google searches as retail attention. We compare this commonly used direct proxy for retail attention to the unsophisticated *experience-based* measures derived from the EDGAR logs.

The Google search volume index (SVI) tracks the number of Google searches on a specific keyword, e.g. a company name, over time. Well known, Da et al. (2011) use this measure to provide empirical evidence for the Barber and Odean (2008) price pressure hypothesis in the U.S. stock market from 2004 to 2007. They further provide evidence in a sample of 185 IPOs documenting significantly positive effects of unsophisticated attention on initial returns and a subsequent long-term reverse within the same year.

There are some important aspects to get a better understanding of SVI (Stephens-Davidowitz and Varian, 2015): i) Google obfuscates the absolute amount of searches for a keyword by scaling the time-series by its maximum, such that it takes a value between 0 and 100 at any point of time. Consequently, SVI represents *within-search-term* variation[15] and can vary for the same week over different time periods. ii) If total searches are below Google's unreported privacy threshold, a zero will be reported. iii) SVI is based on subsamples and thus might differ slightly when downloaded on different days. iv) SVI is not cleaned from robot searches.[16]

For each IPO we obtain the daily SVI for 8 weeks prior to one day before the first trading day on CRSP and aggregate them to weekly values by summing up each week and scale it by its maximum. To account for the base search level, the final IPO attention measure based on Google SVI is defined as:

$$\mathrm{ASVI}_{t-1} = \frac{\mathrm{SVI}_{t-1} - \frac{1}{7}\sum_{i=2}^{n=8} \mathrm{SVI}_{t-i}}{\sigma\left(\mathrm{SVI}_{t-2}, ..., \mathrm{SVI}_{t-8}\right)} \tag{2.4.1}$$

where σ is the standard deviation. Da et al. (2011) use a similar definition using the logarithm of the median SVI value within the trailing seven weeks. We prefer our definition since it drops fewer observations resulting from zeros in trailing SVI.

We question the use of ASVI as an attention measure to explain the cross-section of IPO characteristics. We argue that firms with high base levels of search volume will on average have systematically smaller peaks around the IPO date than firms with low base levels.[17] Prominent examples are Groupon and Twitter, which have negative ASVI in the week before their IPOs despite their huge popularity. We trace this back to the enormous use of the provided online services of these companies.[18] For example, Twitter had more than 200 billion monthly active users and over 500 billion tweets

[15] Google allows a maximum of five separate search words at once to be cross-compared.

[16] On `https://support.google.com/trends/answer/4365533` some information on "automated searches" is provided by Google.

[17] We do not allude to companies with ambiguous firm names here, for example Box Inc. or Cyan Inc, which is another concern to deal with.

[18] Since we are not aware of a direct proxy variable to identify these firms, we cannot provide empirical evidence for this hypothesis.

per day before going public according to their initial prospectus.[19] Our observations relate to DeHaan et al. (2019) who provide a rigorous analysis on measurement errors in Google search volume.

To weaken this issue, we use a second specification of ASVI in which we append the term " ipo" after the company name, thus measuring the abnormal search volume for the IPO itself. Panel A of Figure 2.5 shows the effect of this modification on the obtained SVIs of Twitter. Anyway, firms whose IPOs have been intensively searched earlier than one week before the issuing date will still have systematically downwards-biased abnormal attention values. Besides this, we observe increased volatility in the base search levels induced by zeros for firms with search volume slightly above Google's privacy threshold. See Figure Panel B of Figure 2.5.

These weaknesses are especially pronounced for the cross-section of firms with heterogeneous base search levels, as present in our study. Logically, longitudinal – within-firm – regressions are not as much affected by these issues.

We repeat our main analysis using ASVI instead of the EDGAR-based attention measures. As a consequence of the discussion above, we also include two other modifications of ASVI, namely dummy variables indicating positive ASVI, to analyze the direction of the attention rather than its levels. We view these versions of ASVI as least cross-sectional biased.

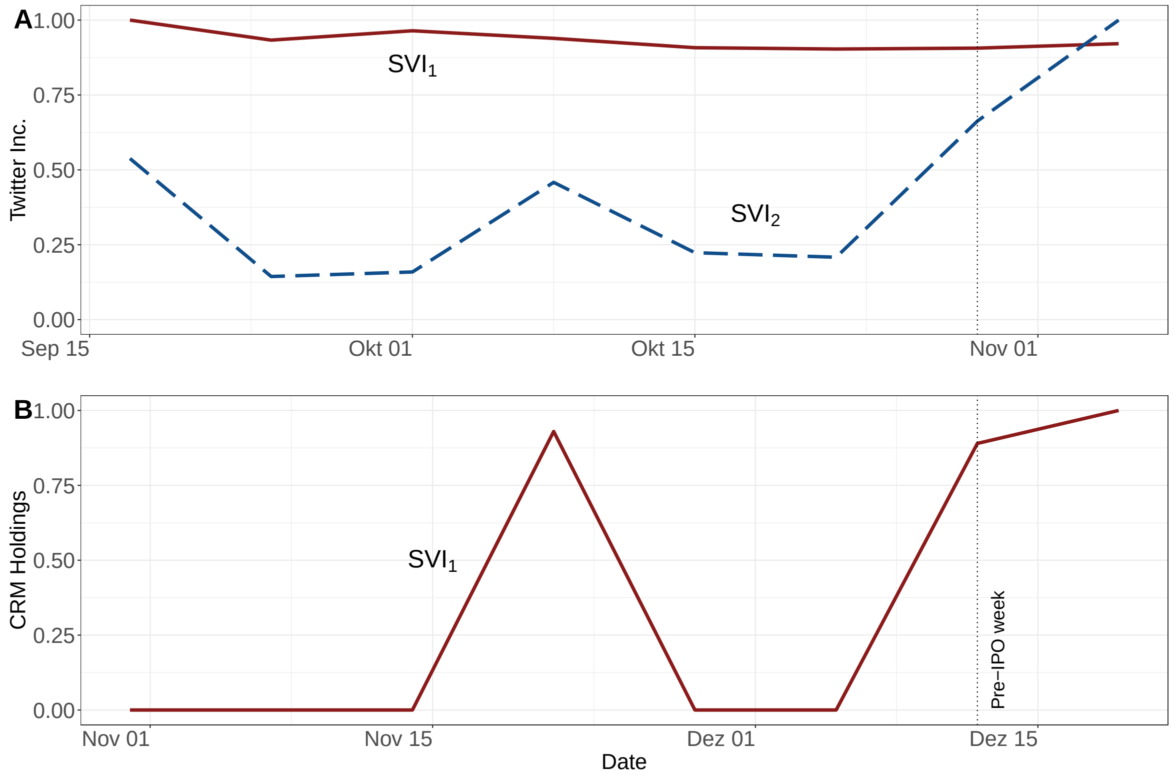
In the price pressure setting, we find only some, if any, evidence for the existence of price pressure induced price reversals using Google SVI. However, we find the correlations of the ASVI measures with the EDGAR-based measures to be consistently higher for attention from the unsophisticated half (8% vs 0%) of the EDGAR universe, which we reinforce in a multivariate setting in the following.

2.4.3 Attention Determinants

In this section, we investigate the determinants of three IPO attention variables. First, we focus on logarithms of the raw user counts $N_j[0, q^{0.5}_{t_j-7}]$ and $N_j(q^{0.5}_{t_j-7}, \infty)$ of viewers of IPO-related EDGAR filings in the pre-IPO week where the first represents unsophisticated and the latter sophisticated attention. Second, we analyze the proportion of

[19] `https://www.sec.gov/Archives/edgar/data/1418091/000119312513390321/d564001ds1.htm#toc564001_13`

Figure 2.5: Exemplary pre-IPO Google Search Volume Index (SVI)



Notes: Panel A (top): SVI of the search terms "twitter" and "twitter ipo" denoted as SVI_1 and SVI_2 respectively. Due to the high base search levels of the term "twitter", SVI_1 is relatively constant over time, while SVI_2 rises to its peak in the week before Twitter's IPO.

Panel B (bottom): SVI for the search term "crm holdings". Due to Google's unreported privacy threshold, the majority of values is set to zero leading to increased volatility in the base search level.

unsophisticated attention $U^{\%}$.

Due to many systematic differences between sophisticated and unsophisticated investors documented in the literature (Barber and Odean, 2013; Miller, 2010; Field and Lowry, 2009; Barber and Odean, 2008) we expect to observe also heterogeneous IPO attention determinants for both groups. We employ a rich set of explanatory variables, of which some have already been used before in this study. First, we describe the newly added variables.

We conjecture that some aspects explaining attention to the IPO of a given firm are related to unique traits of firm or IPO, which are often constant over time but also hard to measure in general. Consider the Facebook IPO as an example that attracted the greatest attention in our sample by raw user counts. Apart from the obvious feature of being one of the largest IPOs ever (Krigman and Jeffus, 2016), anecdotal factors contributing to the immense attention include Facebooks' strong public awareness through their frequently used social media platform, that their founder and CEO Mark Zuckerberg was only 27 years old at the time when the S-1 was filed, and a controversial founding history, which even was the base of a Hollywood movie less than two years before the IPO.[20] We control for such cross-sectional variations by including a measure of initial IPO attention analogously to our base attention measure where we replace the pre-IPO week with the first week after the S-1 filing. We expect initial attention to be a strong determinant for pre-issue attention.

In order to calculate initial attention, we replace our definition of the viewer set A_j for IPO j with $A_j = \{i : |\bigcup_{d=t_{\text{S-1}_j}}^{t_{\text{S-1}_j}+7} F_{i,d}(T_{\text{S-1}_j)}| > 0\}$ containing all IPs with at least one access on the Form S-1 of firm j within 7 days after its filing date. Again, we then define overall attention to the initial prospectus as the cardinality of this set and classify attention into attention generated by unsophisticated respectively sophisticated EDGAR-users based on their SEB-values within the previous year. See Section 2.3.1 for a formal presentation of our definitions.

Da et al. (2011) show that abnormal Google search volume peaks $ASVI$ capture mainly retail investor attention. Hence, we expect Google search volume to be stronger related to our unsophisticated attention. However, as discussed in the previous sub-

[20] "The Social Network" was released to US theatres on October 1st, 2010, grossed $224.9 million at the worldwide box office, won three Academy and four Golden Globe awards, but is also known for its historical inaccuracy.

section, we prefer to use a dummy variable in our regressions, indicating an attention peak in the week prior to the IPO date.

Behavioral studies show that cognitive abilities are limited. This carries over to a limited ability to pay attention, which has been proposed as a possible explanation for many market anomalies related to delayed reactions. For instance, Hirshleifer et al. (2009) find weaker market reactions to earnings announcements in the presence of more same-day earnings announcements and stronger post-announcement drifts. Similarly, we expect distracted investors to more likely skip a specific IPO more, which should reduce the IPOs' received attention. We control for two factors associated with distraction related to IPO timing. First, we calculate the number of all newly filed fundamental EDGAR filings within the pre-IPO week. Second, we calculate the number of all IPOs within the previous 90 days before the IPO. While the first measure captures distraction by any kind of financial information the second one is more related to distraction within the IPO industry and to a potential IPO fatigue.

Table 2.9 summarizes the results. In line with our expectations, we find initial attention to be a strong predictor of pre-issue attention. While significant at the 1 % level for all three attention variables, the magnitude of sophisticated users' initial attention is highest emphasizing their attention persistence during IPO processes.

For abnormal Google Search Volume, we find a strong link to both raw user counts. However, consistent with the notion of Google representing retail attention proposed by Da et al. (2011), the association is stronger for unsophisticated attention. More specifically, we find an attention peak in Google searches to be associated with a 20 % increase in attention from unsophisticated users, significant at the 1 % level, almost twice as high as the 10 % increase in attention from sophisticated users, which is significant slightly above the 5% level. This stronger association between Google and our unsophisticated EDGAR attention is supported via the proportion of unsophisticated attention $U^{\%}$, which is positive but slightly insignificant below the 10 % level.

Our results provide evidence in favor of distracted investors. First, we find the number of newly filed EDGAR documents to withdraw attention from IPOs. While this effect is present for both sophistication groups, our second distraction measure, the number of previously completed IPOs, is only associated with reduced sophisticated attention. A potential explanation may be found in Khanna et al. (2008) who suggest increased information acquisition costs for sophisticated investors in hot markets reducing IPO

Table 2.9: Determinants of (Un-)sophisticated pre-IPO Attention

Dependent variable:	log $N_j[0, q^{0.5}_{t_j-7}]$	log $N_j(q^{0.5}_{t_j-7}, \infty)$	$U^{\%}$
	(1)	(2)	(3)
Initial Attention	0.393***	0.432***	0.225***
	(4.892)	(13.528)	(4.741)
Google ASVI dummy	0.200***	0.105**	0.003
	(2.740)	(2.139)	(1.563)
# Pre-IPO EDGAR filings	−0.046**	−0.048***	−0.000 3
	(−2.241)	(−4.790)	(−0.502)
# Pre-IPO IPOs	−0.001	−0.003***	0.000 1
	(−0.311)	(−3.242)	(1.193)
Pre-IPO $\bar{r}_{\text{Market}}$	0.043	0.024	0.001
	(0.746)	(0.521)	(0.409)
Pre-IPO σ_{Market}	0.438	0.814***	−0.004
	(0.742)	(2.590)	(−0.205)
Revision	0.143	0.251*	−0.001
	(0.501)	(1.931)	(−0.195)
log(Filing Range)	−0.131***	−0.122***	−0.001
	(−3.321)	(−8.092)	(−1.372)
log(Proceeds)	0.234***	0.212***	0.003**
	(4.426)	(8.810)	(2.431)
VC dummy	0.103	0.058	0.003
	(1.425)	(1.635)	(1.615)
Bookrunner Market Share	0.069	0.170**	−0.002
	(0.388)	(2.287)	(−0.578)
log(Sales)	−0.024	−0.005	−0.001
	(−0.852)	(−0.427)	(−1.196)
Debt over Assets	−0.014	−0.004	−0.001**
	(−1.614)	(−0.543)	(−2.155)
Positive EPS dummy	0.092**	0.098***	0.001
	(2.166)	(2.600)	(1.321)
Constant	0.415	1.499***	0.031***
	(1.448)	(10.502)	(3.197)
Fixed effects	Yes	Yes	Yes
Observations	587	587	587
Adjusted R^2	0.591	0.764	0.354
F Statistic	14.882***	32.015***	6.259***

Notes: This table presents results for determinants of three IPO attention variables. We analyze raw (un-)sophisticated user counts $N_j[0, q^{0.5}_{t_j-7}]$ and $N_j(q^{0.5}_{t_j-7}, \infty)$ as well as the proportion of unsophisticated users $U^{\%}$. Detailed variable definitions can be found in Section A.1 of the Appendix. The numbers in brackets below the coefficient estimates show t-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

screening, which is likely to not occur to a comparable extent for unsophisticated investors.

With the pre-IPO 30-day CRSP value-weighted index return and volatility we included two variables related to sentiment. While pleasant pre-IPO market returns are not significantly related to our attention measures, we find that increased pre-IPO market risk is significantly related to increased attention by sophisticated EDGAR-users. This may suggest that sophisticated investors become increasingly careful when markets are in turmoil and are more likely to take a deeper look at relevant primary data.

Generally, with filing ranges from the initial prospectus (Form S-1) to the first public trading IPOs between 25 days and several years, IPOs are typically time-consuming. For instance, Lowry et al. (2016) find larger filing ranges for firms with more SEC interaction and SEC concerns to the firms' disclosures. Further, Dunbar and Foerster (2008) argue that delayed IPOs may distract the management from the actual business and increase the likelihood of weakened market conditions leading to larger withdrawal risk. We find that lengthy IPOs have less pre-IPO attention, which is consistent with investors facing more potential distraction increasing the probability of losing interest.

Field and Lowry (2009) study differences in individual and institutional IPO investment with a focus on public data usage and voluntarily institutional holdings sometime after the issue. While institutions are commonly thought to have systematic advantages over private investors due to private information or monitoring activities Field and Lowry (2009) find that much of institutions' superior IPO selection can be attributed to better use of public data.

Similar to Field and Lowry (2009), we find sophisticated IPO attention to be associated with larger proceeds, with certification via venture capital firms and market-leading underwriters, and with profitable firms. However, we find that unsophisticated attention is related to larger proceeds and profitable firms as well. Moreover, our regression (3) using the unsophisticated attention proportion $U^{\%}$ reveals a significantly stronger relation between unsophisticated attention and proceeds. While not being significant in one of the raw attention regressions we observe that significantly less attention is paid to more leveraged firms by unsophisticated attention. If we interpret our attention measures, which are calculated in a close-to-issue phase, as a revelation of investors' willingness to invest then our results suggest that unsophisticated investors do not necessarily disregard relatively simple accounting information such as profitability as previously thought.

Finally, Table 2.9 shows that the revision from the midpoint of the filed price range to the offer price is significantly positively related to sophisticated attention. Revision is typically interpreted as being related to positive and negative information revelation from institutional investors to underwriters during the bookbuilding process (Benveniste and Spindt, 1989; Hanley, 1993). Hence, since we do not find this relation for unsophisticated attention, we find support for information revelation theories and also for the notion that EDGAR attention is a positive demand indicator.[21]

Overall, with adjusted R^2 values above approximately 60 % we are able to explain a substantial portion of the variation in both raw attention measures. Furthermore, the adjusted R^2 for sophisticated attention exceeds the unsophisticated R^2 by about 17 percentage points. This underlines that sophisticated users are more systematic in how they retrieve information from disclosures.

2.5 Conclusion

In this paper, we propose a direct measure of timely financial disclosure experience, capturing how frequent and continuous individuals assess relevant firm disclosures on the information retrieval system EDGAR. The measurement adds to the ongoing debate on the role of acquiring and processing disclosure information and its accompanying costs, which have recently been stressed to have implications for a broad range of research and phenomena (Blankespoor et al., 2020). Due to the flexibility and simplicity of our approach, we believe in its potential to be useful in many other contexts.

Building upon our experience measurement, we propose and test a new measure of investor sophistication that avoids challenging IP address matching, ambiguous mappings between company and attention, and reliance on obfuscated data that impedes cross-sectional comparisons, among other things. While our baseline approach makes use of time-dependent sophistication thresholds for classifications, our robustness tests show that appropriate choices of constant thresholds allow a further simplification.

[21] In contrast, Field and Lowry (2009) find a negative association between revision and voluntarily institutional holdings. This tendency is interpreted as an institutional flipping activity. That we find a contrary relation of revision and attention highlights differences between pre-issue attention and voluntarily holdings sometime after the issue as used in Field and Lowry (2009).

Using several versions of the proposed sophistication measurement, we provide unique empirical evidence that pre-IPO-week attention from less sophisticated investors is associated with higher underpricing and subsequent price depreciations within one year after the offering, consistent with the price pressure hypothesis proposed in Barber and Odean (2008). Due to the lack of measures adequately capturing unsophisticated investor attention in the cross-section of IPOs, the recent empirical literature on this issue is fragmentary.

3 SEC Workload, IPO Filing Reviews, and IPO Pricing

The following is based on Köchling et al. (2020b).

3.1 Introduction

The U.S. Securities and Exchange Commission (SEC) Division of Corporation Finance (CF) is one of five divisions within the SEC. Its goal is to ensure the completeness and quality of the information provided by firms enabling investors to make informed decisions based on reliable information (SEC, 2019a).[1] By means of their filing review process, the offices of the CF examine corporate filings and issue comments if needed. For instance, in 2019, the CF performed overall 4,090 reviews, including 590 reviews for new issues (SEC, 2020). Almost all IPOs are getting reviewed, often resulting in a considerable number of comments, which makes the SEC an important stimulator of information production.

For IPOs, information production is a process traditionally associated with large, institutional investors attempting to value the offering. Their privately produced information plays a crucial role in models of underpricing where underwriters compensate investors for truthfully revealing their positive information by adjusting the price of the offering only partially (Benveniste and Spindt, 1989; Hanley, 1993). This leads to the well-known positive relation between price revision in the primary market and underpricing.

The role of the issuer as an information producer has recently gained increased academic attention. Lowry et al. (2020) focus on how the SEC induces issuers to

[1] For periodic filings such as quarterly and annual reports, the literature reports beneficial effects associated with the SEC filing review. For instance, Cunningham et al. (2020) find fewer earnings management, Bozanic et al. (2017) find fewer information asymmetry, and Kubick et al. (2016) find fewer tax avoidance. A natural prerequisite for such effects is a sufficient review quality.

disclose information. Hanley and Hoberg (2010) study the extent to which issuers produce information via due diligence prior to the filing of a preliminary prospectus. They develop a measure of prospectus informativeness and find that prospectuses with more informative, non-standard content result in more accurate prices. This supports the view that more initial information production by the issuer, including the help of advisers such as underwriters, represents an alternative or additive to information production via bookbuilding.

In this paper, we examine how high workload from time-varying filing activity impacts the SEC filing review process for IPOs and the SEC's ability to prompt information production. Considering the unique role of IPOs in the history of a firm as well as the substantial uncertainty and information asymmetry accompanying these events, the role of regulatory authorities and potential deficiencies are of great importance.

We construct a daily workload measure to proxy the number of filings in urgent review each day for each industry office in the Division of Corporation Finance. The workload measure passes three initial tests where we explain organizational changes between SEC offices that are likely to be related to workload as well as self-reported SEC workload data. The workload measure used in our study is inspired by the one proposed in Ege et al. (2020) but differs in several details.

Next, we build a comment letter database from the publicly available EDGAR data and match SEC comment letters to IPO filings, namely preliminary prospectuses as well as their amendments. Building on this, we investigate the relationship between high workload and comment letter quality, remediation costs via response times, and implications for IPO pricing.

As the starting point of our empirical analysis, we focus on quantitative quality measures of the SEC comment letters, such as the number of comments for each IPO. On average, the first letter in our sample contains already 74% of all comments issued during the IPO and hence is most important. However, using negative binomial models, we find no compelling evidence in favor of decreases in quantitative quality in the first letter when the workload is high. This is consistent with the presumably high priority of these reviews but contrary to what has been documented for annual reports (Ege et al., 2020; Gunny and Hermis, 2020). Turning to the subsequent letters after the first one (2.6 on average), we find that a high workload on the filing date of the corresponding IPO filing is associated with a significant 11% decrease in the number of comments.

A comment-similarity clustering reveals that a considerable portion of the comments, between 5% and 21% depending on parameters, are similar across different IPOs. We employ this procedure to approach a more content-related measure of quality. For each initial comment for each IPO, we determine the most similar comment from a set of recent IPOs based on cosine similarity. Then we classify all comments having a cosine similarity larger than 80% to their most similar comment as being standard. We find that both the number of standard comments and the proportion of standard comments are more extensive for high workload IPOs.

We then turn to the response times by the SEC staff, which are particularly important for IPOs since any exogenously prolonged registration time can be regarded as costs due to a distraction of the management (Falato et al., 2014), forfeiting of favourable market conditions (e.g., Pástor and Veronesi (2005)), or an increased risk of IPO withdrawal (Busaba et al., 2001), among other things. We study aggregated and letter-level SEC response times using Cox (1972) proportional-hazard models. Across different specifications, we find that high workload is associated with significantly quicker responses.[2] Regarding solely the time in active SEC review proxied by the sum over all letter-level response times, we find the IPO review process to be completed about 29% earlier.

At first glance, quicker responses appear to be counterintuitive since high workload could also be associated with a delay in order to guarantee a certain level of quality. For instance, the SEC staff conducting the reviews states in some letters that reviews of the one letter might yield a delay for other letters.[3] Taken together, quicker responses can be interpreted as a sign of either lower quality or increased efficiency. Psychological theories such as the job demands-resources model (Bakker and Demerouti, 2014) and the challenge-hindrance framework (Crawford et al., 2010) as described by Tadić et al. (2015) show that "challenge job demands" (as opposed to hindrance job demands) can have a positive relationship with work engagement.[4]

Due to the evidence regarding high workload consequences for IPO reviews, we explore how filing reviews and workload relate to IPO pricing. We begin by revisiting existing findings regarding filing review outcomes and IPO price revisions from the

[2] These analyzes exclude the first letter due to the considerable clustering of first-letter response times around 27 days with only little variation.

[3] See, e.g., `https://www.sec.gov/Archives/edgar/data/1533932/000000000011067372/filename1.pdf`.

[4] Often, workload and time urgency are regarded as a challenge demand.

midpoint of the first price range to the offer price. In addition to the overall number of comment letters (Li and Liu, 2017), the number of comment letters prior to the first price range (Lowry et al., 2020), we use a measure of SEC concerns based on various comment counts - e.g., all comments in the first letter - as measures of SEC prompted information production and find consistent results that SEC concerns are related to absolute price revisions and down revisions.

Building on this, we examine the interaction of raised SEC concerns and high workload and find that the relation between SEC concerns and (absolute) revision becomes smaller under high workload. The statistically significant effect of SEC concerns on price revision doubles when controlling for the interaction with high workload. However, the estimate of the interaction term is almost diametrically to the effect of the SEC concerns. Similar results hold for absolute revision. Hence, for all IPOs subject to high workload, we find no relation between SEC concerns and price revisions.

The disappearance of the association between SEC concerns and price revision under high workload suggests that not all expressed SEC concerns are similarly informative for price changes. This receives support when we calculate SEC concerns conditional on standard and non-standard comments. We find that non-standard SEC concerns are significantly related to price revisions while standard concerns are not. Moreover, non-standard concerns are associated with more information production and standard concerns with less. A potential explanation of these results is a lack of quality under high workload, which, however, does not affect the overall number of comments but is potentially reflected in a tendency to more standard content in the letters.

If high workload is associated with less SEC induced information production, we expect that more information needs to be produced by institutional investors via bookbuilding. In turn, this should be compensated via underpricing by underwriters and issuers (Hanley and Hoberg, 2010). Examining the relation between underpricing and high workload, we find 2% higher first-day returns under high workload, which is significant at the 1% level and consistent with this hypothesis.

Our primary variable of interest is the high workload dummy. Its assignment to IPO filings is non-random since two firms matched to the same SEC office filing sufficiently close will have the same treatment. This complicates the estimation of a high workload effect. We address this by applying entropy-balancing to our sample where adequate (Hainmueller, 2012).[5] Generally, we include a variety of standard IPO control variables, which, however, are not necessarily sufficiently rich. For instance, the central determinant of initial comments is undoubtedly the true number of issues within the issuing firm, which we cannot control for since its revelation is one of the goals of the SEC review process. Interestingly, while we do not find an effect regarding the first letter, we find fewer subsequent letter comments under high workload. This is robust to the inclusion of issuer fixed effects, which should largely control for issues associated with the IPO firm.[6]

Our study contributes to the literature in the following four directions. First, we contribute to the IPO literature by shedding light on the role of regulatory reviews and information production for IPOs (Benveniste and Spindt, 1989; Hanley, 1993; Hanley and Hoberg, 2010). Second, our paper is related to the distraction literature where the focus was traditionally on investor distraction, reactions to information, and implications for asset prices (Hirshleifer et al., 2009; Dellavigna and Pollet, 2009). We widen the horizon of this strand by examining regulator distraction in the IPO process. Third, the present study adds to the literature on SEC filing reviews (see Cunningham and Leidner (2019) for a summary), particularly to the scant evidence for IPO filing reviews (Agarwal et al., 2017; Li and Liu, 2017; Lowry et al., 2020). We expand the former literature strand by focusing not only on the first letter. Due to our focus on potentially varying review quality, we advance also the IPO filing review strand. Fourth, we expand the textual analysis literature in finance and accounting by clustering similar SEC comments (see Loughran and McDonald (2016) for a survey).

Our results should be of interest to the regulatory authorities. First, we believe that additional resources can help to ensure that all IPOs experience regulatory information production of the same high quality. Our results can be interpreted in a way that this was not always the case in the past. Second, even without additional funding, a reconsideration of the internal structure of the CF might also mitigate the consequences of high workload. Since workload originates at the SEC office level, a higher number of offices combined with a rather rigid mapping between firms and offices can result in some offices being under high workload even when the overall resources are not fully

[5] Entropy-balancing calculates sample weights to achieve moment conditions for the covariates in both the treatment (high workload) and control group. This method was similarly applied by Ege et al. (2020).

[6] Further concerns for other regressions are discussed in the respective sections.

used.[7] Interestingly, recent changes to the internal structure have led to a reduction to only seven offices. We believe that this change can help to avoid potential problems arising from high workload.

Our results can also be of interest to all those involved with IPOs. For instance, for issuers, we provide insights into the nature of comments issued by the SEC by quantifying their similarity and we provide evidence regarding help from high-quality companions when going public such as a Big 4 auditor. Together with considerations regarding SEC busyness, such aspects can inform decision-makers.

The remainder of this article is organized as follows. Section 3.2 describes how we build our IPO sample with a strong focus on the matching between IPO filings and SEC comment letters. This section also contains summary information about the IPO filing review over the years and the comment similarity clustering. Section 3.3 defines our workload measure, details regarding its implementation, including inherent limitations, as well as initial evidence that it is able to capture stressed periods. In Section 3.4 and 3.5, we focus on the relationship between the quality of comment letters issued by the SEC, respectively their response times, and high workload. Section 3.6 studies the relation between the filing review, IPO pricing, and high workload. Section 3.7 concludes.

3.2 IPO Sample, IPO Filings, and Comment Letters

In this section, we describe our IPO sample selection process (Subsection 3.2.1), how we match IPO filings and SEC comment letters (Subsection 3.2.2), and give overview figures on the SEC filing review (Subsection 3.2.3).

3.2.1 IPO Sample

Our IPO list is extracted from Thomson Financial's SDC New Issues database with additional items and corrections supplied by Professor Jay Ritter.[8] Since SEC comment letters are available on EDGAR since 2004, we restrict the sample to August 2004 till

[7] Essentially, the *industry* offices are organized to map industries. However, some offices process filings of quite different firms such as the *Office of Beverages, Apparel, and Mining.*

[8] SDC Corrections and founding dates are taken from `https://site.warrington.ufl.edu/ritter/ipo-data/`. We thank Professor Ritter for making this data publicly available.

December 2018 covering slightly more than 14 years. We follow Lowry et al. (2017) and perform typical exclusions. We exclude offerings that are associated with limited partnerships, closed-end funds, units, financial companies, real estate investment trusts, and dual-class capital structures or have an offer price less than USD $5.

We merge the SDC list to stock data from CRSP, to annual accounting data from Compustat, to the founding dates provided by Professor Jay Ritter, and to EDGAR via the EDGAR master index file and the SEC file number available in SDC. For all IPOs, we determine relevant IPO filings (including Draft Registration Statements) and SEC Letters (using a self-created comment letter database) and match the letters to the filings via one of three methods (by order, by date, or by Amendment Number).[9] Similar to Lowry et al. (2020), we keep only IPOs with at least one comment letter and omit also IPOs where we could not match all letters. Additionally, we exclude IPOs where we detect one of the following conditions: indication of a material fail or of a limited review in the first SEC letter, multiple Draft Registration Statements prior to the first public filing, a 10-12G filing prior to the first IPO filing, mismatch between first EDGAR SIC Code and SIC Code of the final prospectus, or existence of last reported sale price on an exchange.

After all exclusions, we obtain 922 IPOs where all standard IPO control variables are available. Table 3.1, Panel A, shows the descriptive statistics of the final sample. Variable definitions can be found in Table B.1 of the appendix.

3.2.2 Matching IPO Filings and Comment Letters

The public part of the IPO process in the U.S. starts with the filing of a preliminary prospectus. With this prospectus, the issuer presents itself and the offering to the general public for the first time. Common parts of the prospectus are describing the business model, risk factors, and the financial situation. Hence, the prospectus is a primary information source when evaluating the issuer. For the majority of firms, the prospectus is subject to a detailed review by staff from the SECs' Division of Corporation Finance. In order to ensure the quality of the disclosure, the SEC typically replies with a list of comments demanding amendment or further explanations. Since

[9] Details regarding the matching can be found in Subsection 3.2.2. However, those who are not interested in the details may want to skip to the overview in Subsection 3.2.3.

Table 3.1: Descriptive Statistics

Panel A: IPO-level Summary

	Mean	Std. dev.	perc(0.1)	Median	perc(0.9)
Workload Variables:					
Workload	0.65	0.28	0.21	0.72	0.97
High Workload (D)	0.40	0.49	0.00	0.00	1.00
Filing Review Variables:					
#Letters	3.60	1.54	2.00	3.00	5.00
#Letters$_{\text{Before PR}}$	3.02	1.32	2.00	3.00	5.00
#Comments$_{\text{First Letter}}$	39.17	19.94	16.00	36.00	66.00
#Comments$_{\text{Before PR}}$	56.11	36.23	20.00	50.00	100.00
#Stand. Comments	2.44	2.21	0.00	2.00	5.00
#Non-Stand. Comments	36.50	19.45	14.00	34.00	62.00
Proportion(Stand. Com.)	0.08	0.08	0.00	0.05	0.19
SEC Concerns	−0.01	0.39	−0.45	−0.06	0.50
Stand. SEC Concerns	0.02	0.87	−1.00	−0.19	1.19
Non-Stand. SEC Concerns	−0.01	0.42	−0.48	−0.07	0.55
Dependent IPO Variables:					
First-Day Return (%)	17.41	26.91	−6.25	11.08	51.51
Revision (%)	−4.04	20.38	−30.95	0.00	18.75
Abs. Revision (%)	15.33	14.02	0.00	12.50	33.88
Controls:					
ln(Age)	2.54	0.81	1.61	2.40	3.71
ln(Sales)	3.91	2.44	0.00	4.30	7.03
Leverage	0.90	1.16	0.18	0.70	1.52
Pos. EPS (D)	0.35	0.48	0.00	0.00	1.00
VC (D)	0.57	0.49	0.00	1.00	1.00
Bookrunner Market Share	0.30	0.24	0.00	0.28	0.64
Lawyer Market Share	0.03	0.04	0.00	0.01	0.07
Big 4 (D)	0.83	0.38	0.00	1.00	1.00
ln(Review Size)	15.10	0.53	14.43	15.08	15.76
Market Return$_{\text{30 Days}}$	0.18	0.35	−0.23	0.16	0.60
Market Vola$_{\text{30 Days}}$	0.13	0.06	0.08	0.11	0.20

Panel B: Letter-level Averages

	Letter 1	Letter 2	Letter 3	Letter 4
#IPO (abs.)	922.00	882.00	711.00	435.00
#IPO (%)	100.00	96.00	77.00	47.00
#Comments	39.17	10.84	5.18	4.77
#Words	2 174.71	718.32	354.89	298.03
Response Time (Days)	26.93	14.76	11.53	9.10
Response Time (Workdays)	18.44	10.20	7.87	6.35
Workload	0.65	0.65	0.65	0.63
High Workload (D)	0.40	0.38	0.40	0.38
Review Size (MB)	4.18	1.33	1.19	1.27
Market Return$_{\text{30 Days}}$	0.18	0.17	0.19	0.21
Market Vola$_{\text{30 Days}}$	0.13	0.13	0.13	0.13

Notes: This table presents descriptive statistics of the dataset. Panel A shows a summary of the variables on the IPO-level. Panel B presents averages of variables that relate to a specific letter of the review process. See Table B.1 in the Appendix for detailed definitions and sources of the variables.

2004, these comment letters are filed publicly with some delay via EDGAR. In the following, we describe how we construct a sample of IPO filings and corresponding comment letters.

Identifying IPO Filings We match the IPO list to the EDGAR index file by identifying the (public) preliminary prospectus and the final prospectus. During this matching we allow the filing date (for the preliminary prospectus) and the issue date (for the final prospectus) from SDC to differ up to three days from the filing dates in the EDGAR index. Admissible form types for the preliminary prospectus are S-1, F-1, and SB-2. For IPOs without a match by this method, we use the SEC file number provided by SDC. For all IPOs after 2012, we search additionally for Draft Registration Statements (form type: DRS) in the EDGAR index prior to the public preliminary prospectus. These drafts were introduced with the JOBS Act in 2012 and are initially confidential and only made public with some delay. For each IPO, we denote all preliminary registration statements (including drafts if available) and their amendments as IPO filings. From the EDGAR index, we extract a list of these filings between the first and final prospectus.

Identifying SEC Comment Letters For each IPO, we reduce the set of all UPLOAD filings to the comment letters relevant to the IPO. In this process, we make use of a self-created comment letter database. This database covers 153,105 parsed UPLOAD filings representing 98,6% of all available UPLOADs on EDGAR until December 2019. Details of the database construction are described in Appendix B.1. We consider all UPLOADs up to two years after the issue. That is, we also examine UPLOADs prior to the first IPO filing. This is necessary since the Draft Registration Statements of a few IPOs are not contained in the EDGAR index. In these cases, we supplement the IPO filings with information from the letters. With the choice of a two-year-post-IPO window, we follow Lowry et al. (2020). For all required UPLOADs with parsing errors, we collect the data manually.[10] We omit all UPLOADs whose date of dispatch is not within the IPO registration range and that do not reference

[10] This applies to 19 cases in our sample. A common reason for a failure is that the UPLOAD is a scan or does not represent a comment letter.

an IPO form type.[11] Furthermore, we omit all IPOs where at least one UPLOAD references both an IPO filing and a non-IPO related filing since we cannot automatically distinguish between comments related to the IPO and potential other comments.

Matching IPO filings and Comment Letters For all IPOs with a non-empty set of comment letters, we match the letters to the IPO filings via the three following approaches, which are ordered by precedence:

1. Matching by Order:
   - Iterate over all letters starting with the earliest:
     - Determine all unmatched IPO filings prior to the letter.
     - If there is only one such filing, then match it to the letter.
     - If not, end the matching attempt unsuccessfully.
2. Matching by Date:
   - Determine all filing dates referenced in all letters.
   - If all letters reference at least one date, then match by date.
   - If not, end the matching attempt unsuccessfully.
3. Matching by Amendment Number:
   - Determine the referenced amendment numbers in all letters.
   - If all letters reference at least one amendment number, then match by Amendment Number.
   - If not, end the matching attempt unsuccessfully.

Which approach is suitable depends on the data contained in the letters and the type of mapping between IPO filings and letters. For instance, matching by order works only for a simple mapping structure where all IPO filings up to a certain one receive a

[11] Currently, we do not make use of the file number to identify relevant UPLOADs. A file number captures related filings on EDGAR. This alternative was used in Lowry et al. (2020) but is usually not applicable for draft comment letters since these often lack file numbers. The resulting summary statistics for both approaches are close, which gives trust to both approaches. See Table 3.1 of this paper and Table 1 of Lowry et al. (2020).

letter. Regarding the precedence, we use matching by order first, since it requires the least amount of parsed information from the letters. Then, we try matching by date due to its obvious accuracy.

Generally, we consider a match to be successful if all of the several conditions are satisfied. First, all letters should be matched to at least one IPO filing.[12] In contrast, not all IPO filings need to be matched to a letter. Second, we require that one IPO filing is matched to one letter at most.[13] Third, we require that the first IPO filing needs a matching comment letter.

Comments, Response Times, and Shifting From our comment letter database, we merge the number of comments to each letter. For all pairs of matched IPO filings and letters, we calculate the *Response Time of the SEC* as the number of days (and workdays) between the date of the IPO filing and the reply date contained in the SEC letter. Some of these response times are zero. Such an immediate response is rather unsuspicious for all later letters where the number of comments is typically low. However, for early letters, especially letters issuing quite a few comments, manual checking of these cases suggests that it can be more sensible to shift the matched IPO filing to its predecessor if the predecessor is an unmatched draft statement. In these cases, it appears that the issuer files a public version of an originally confidential draft filing under review and the SEC references the public filing instead of the original one, which explains seemingly quick responses. Hence, we conduct such a shift when the corresponding response time is below four.

3.2.3 Summary Statistics of the IPO Filing Review

Before omitting IPOs due to missing variables, our IPO sample from 3rd August 2004 to 30th August 2018 includes 1,339 IPOs.[14] For 1,206 IPOs we attempt to match IPO filings and letters and in 1,086 cases we obtain a complete match (592 matched by order, 447 matched by date, 47 matched by Amendment Number).

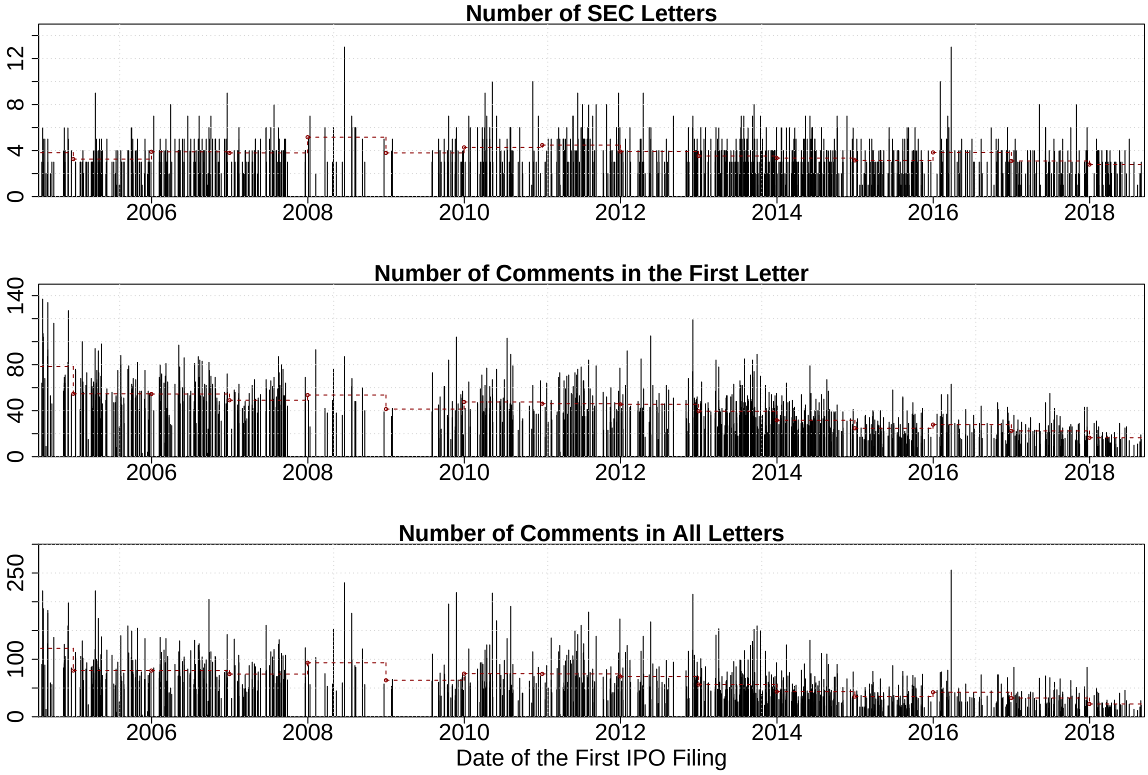
[12] Sometimes a single letter references more than one IPO filing.

[13] While more than one letter per IPO filing can occur in practice, for instance, when a few additional comments are submitted via a separate letter, we use this requirement to omit cases where erroneous matching would occur due to unclear referenced data in the letters.

[14] Filing date of the first IPO receiving a comment letter and filing date of the last IPO in our sample.

Table 3.1, Panel B, presents summary statistics of the described matching process for all 922 IPOs obtained after dropping all IPOs with missing control variables. With 39 comments on average, the first letter contains the most comments. This number decreases sharply for the following letters. Similar observations can be made for the SEC response time.

Figure 3.1: Key Measures of the IPO Filing Review over Time

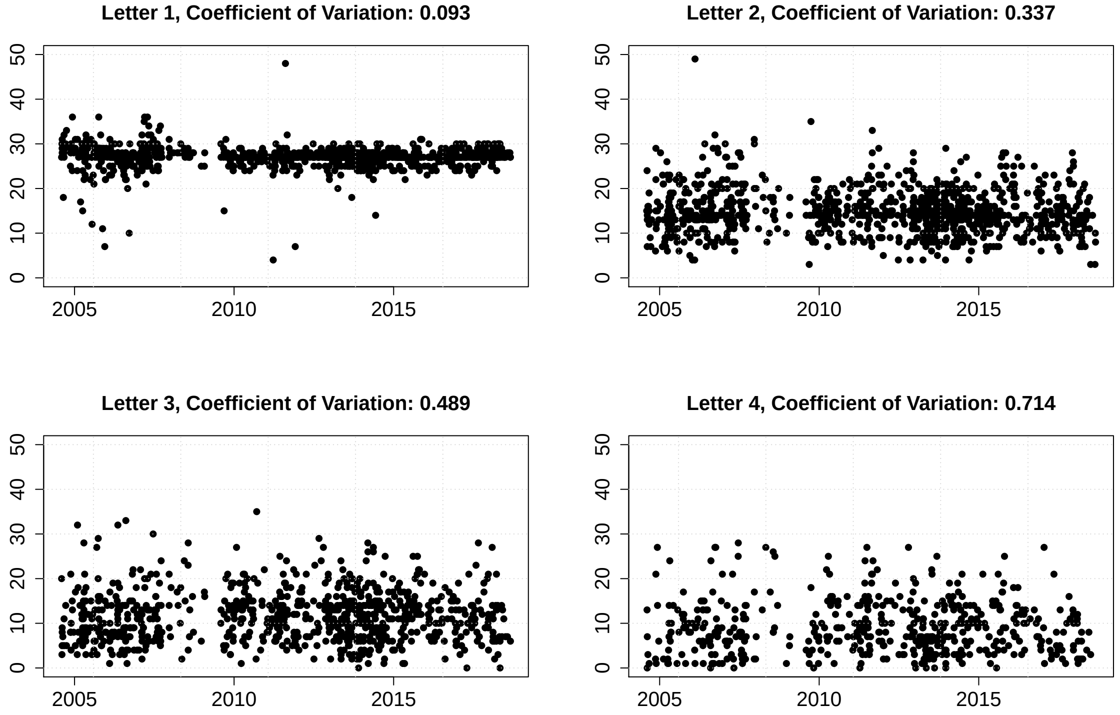


Notes: This figure shows the number of SEC Letters, the number of comments in the initial SEC letter as well as the overall number of comments issued by the SEC for all 1,046 IPOs between 2004 and 2018 where we obtained a complete match between IPO filings and letters. The red, dotted lines indicate yearly averages. While the number of comment letters decreased only slightly, if any, the comment counts decreased substantially.

Figure 3.1 shows several statistics of the IPO filing review process against time. The number of letters is relatively constant with a slight tendency to fewer letters. In contrast, the number of comments decreased considerably over time. From 2005 with

55 comments to 2011 with 46 comments, we observe already a decrease, which became even more pronounced thereafter and culminates in 22 comments in 2017. On the one hand, the publication of the SEC letters after 2004 is likely to help avoid standard SEC comments. On the other hand, the introduction of reduced disclosure requirements for emerging growth companies with the JOBS Act in 2012 contributes also to this trend. To account for the fact that the number of comments is not comparable over time and to avoid spurious regressions, we regress the number of comments on year dummies. We use the resulting residuals as a measure of *SEC concerns* in Section 3.6.

Figure 3.2: Response Times of SEC Comment Letters by Letter Number



Notes: This figure shows plots of SEC response times (in days) for SEC comment letter 1 till 4 for all 1,046 IPOs between 2004 and 2018 where we obtained a complete match between IPO filings and letters. They illustrate considerable increases in dispersion from letter number to letter number as documented by the rising coefficient of variation. Concurrently, the mean response time tends to decrease for higher letters numbers.

Figure 3.2 reveals considerable response time variations depending on the review round, that is SEC Letter number. While the plot of letter 1 resembles a horizontal line around 27 days with only a few outliers, mainly downwards, the response times become more and more dispersed during the review process, which is also emphasized by the increasing coefficients of variation.[15]

3.2.4 Standard and Non-standard Comments

When browsing SEC comments, one notices similar, rather boilerplate comments for different IPOs. In this section, we quantify the magnitude of this observation in our IPO sample. We transform each individual comment into a word root vector, cluster the data into subsets of similar comments, and compare the comments via cosine similarity.[16]

Clustering. We place relatively high demands on the similarity of two comments to be clustered. As a result, we aggregate only comments that are almost identical. That distinguishes our approach from the one pursued in Lowry et al. (2020) who perform a latent Dirichlet allocation (LDA) for comment letters. LDA models that documents (the comments for each IPO as a whole in Lowry et al. (2020)) are composed of a fixed set of relatively few topics. Instead, we exploit the (more or less) natural structure of the comment letters by clustering at the comment level and demanding a high degree of similarity. While being related in terms of the goal, our approach is also different from the procedure used by Hanley and Hoberg (2010) who measure informative and standard content of IPO prospectuses. That approach regresses the word root counts of the current document on word root counts from a set of past documents. Since the lengths of the SEC comment letters vary substantially, word root counts of shorter letters will have a tendency to be more "informative" and longer letters will be less "informative".[17] Hence, we do not use this approach and prefer direct comment comparisons, which are also more illustrative. However, we follow most

[15] The clustering around 27 days for the first letter seems to reflect internal SEC deadlines (Johnson et al., 2019). SEC, 2019b reports a target of "30 days or less" with actual values between 25.4 and 26.0 for the period between 2013 and 2018.

[16] Cosine similarity measures the similarity between two non-zero vectors based on the angle α between them as follows: $\text{sim} = \cos\alpha = (v_1 \cdot v_2)/(|v_1|\ |v_2|)$ where $\cdot$ is the dot product.

[17] In this framework, informative content is defined as the sum of the absolute residuals from the word root regression. Obviously, shorter documents, e.g. a single comment letter, tend to lack many of the roots contained in larger documents, e.g. the combined comments of a few past IPOs. Hence, absence of words can be classified informative.

of the text preprocessing steps used in Hanley and Hoberg (2010). Each comment is processed as follows:[18]

1. Initially, we parse all text between the beginnings of two consecutive comments. In many cases, this text still contains subheadings introducing the next set of comments at the end. We drop these subheadings.[19]
2. We convert the comment to lower case.
3. We tokenize the comment and keep only tokens contained in the Loughran-McDonald master dictionary. We drop stopwords and all tokens associated with articles, conjunctions, and personal or possessive pronouns.[20]
4. We stem the remaining words to word roots and drop all roots that occur fewer than five times in all comments of all initial letters combined.[21]
5. We apply a term frequency–inverse document frequency (tf-idf) weighting to the roots.

The text preprocessing steps are applied to 49,404 initial comments for all IPOs where we either obtained a full match between IPO filings and SEC letters or a partial match for the first letter. We then run the clustering algorithm DBSCAN on the transformed comments (Ester et al., 1996). DBSCAN is suited for large sample sizes, can handle quite many clusters, and is able to detect asymmetric cluster sizes. Not all data gets necessarily clustered. Instead, the data is classified into clusters and noise. In our application, noise comments are those that are more or less unique to an IPO, at least in terms of the word root vector. To control how the data gets clustered, DBSCAN requires two parameters: ε relates to the (euclidean) distance that determines the neighbors of a vector and m controls roughly the minimal cluster size. Exemplary baseline results for the case $\varepsilon = 0.5$ and $m = 5$ are illustrated in Figure 3.3.

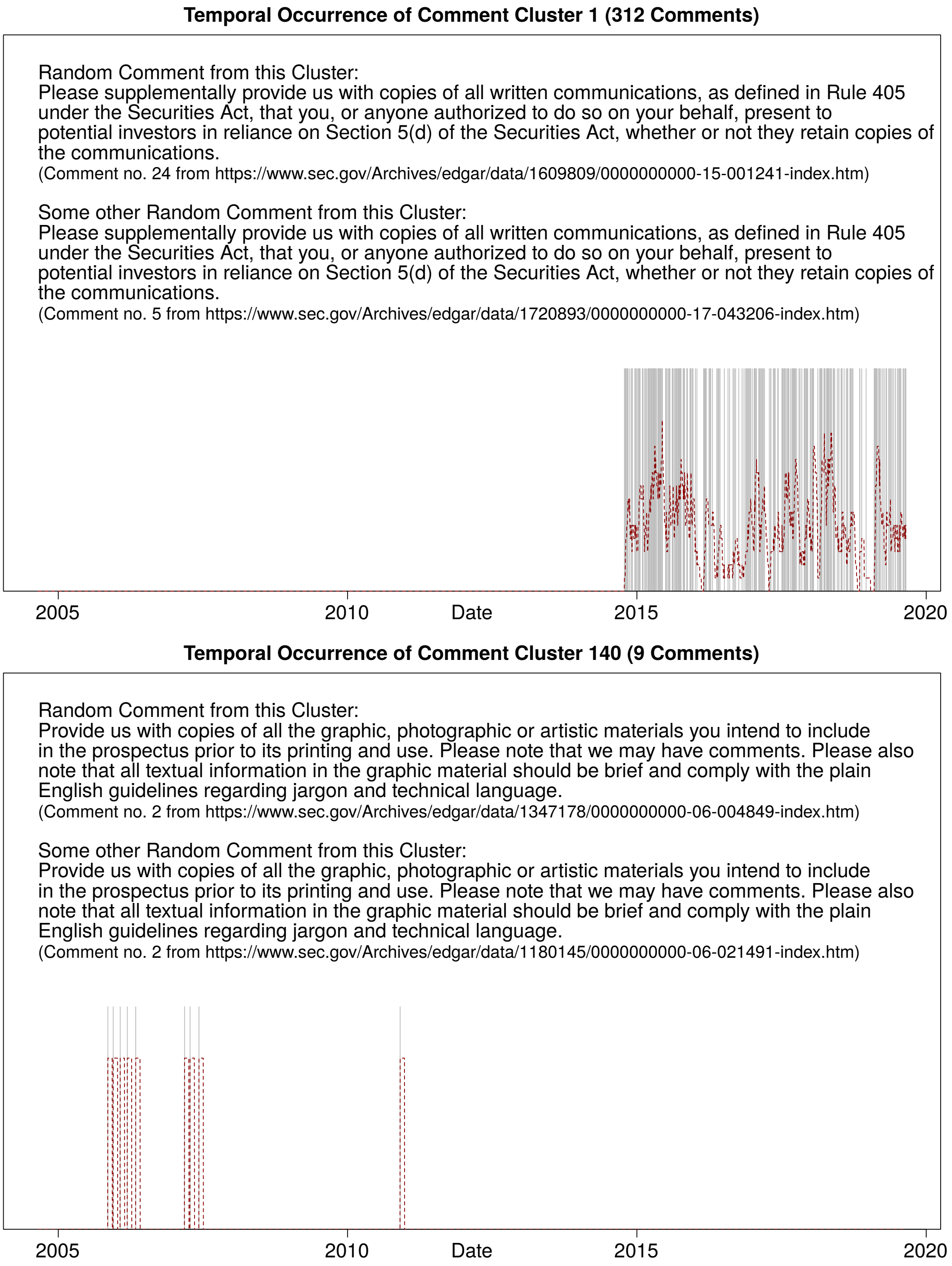
[18] We use the Python packages NLTK (Bird et al., 2009) and scikit-learn (Pedregosa et al., 2011).

[19] We use the PunktSentenceTokenizer from NLTK supplemented with specific common sentence endings occurring in the SEC comments to detect the subheadings.

[20] The master dictionary can be downloaded from `https://sraf.nd.edu/textual-analysis/resources/`. The stopwords to drop are from NLTK. Then, we drop also all words tagged with 'CC' (coordinating conjunction), 'DT' (determiner), 'PRP' (personal pronoun), or 'PRP$' (possessive pronoun) via NLTK.

[21] We use "PorterStemmer" from NLTK.

Figure 3.3: Clusters of Initial SEC Comments with DBSCAN



Temporal Occurrence of Comment Cluster 1 (312 Comments)

Random Comment from this Cluster:
Please supplementally provide us with copies of all written communications, as defined in Rule 405 under the Securities Act, that you, or anyone authorized to do so on your behalf, present to potential investors in reliance on Section 5(d) of the Securities Act, whether or not they retain copies of the communications.
(Comment no. 24 from https://www.sec.gov/Archives/edgar/data/1609809/0000000000-15-001241-index.htm)

Some other Random Comment from this Cluster:
Please supplementally provide us with copies of all written communications, as defined in Rule 405 under the Securities Act, that you, or anyone authorized to do so on your behalf, present to potential investors in reliance on Section 5(d) of the Securities Act, whether or not they retain copies of the communications.
(Comment no. 5 from https://www.sec.gov/Archives/edgar/data/1720893/0000000000-17-043206-index.htm)

Temporal Occurrence of Comment Cluster 140 (9 Comments)

Random Comment from this Cluster:
Provide us with copies of all the graphic, photographic or artistic materials you intend to include in the prospectus prior to its printing and use. Please note that we may have comments. Please also note that all textual information in the graphic material should be brief and comply with the plain English guidelines regarding jargon and technical language.
(Comment no. 2 from https://www.sec.gov/Archives/edgar/data/1347178/0000000000-06-004849-index.htm)

Some other Random Comment from this Cluster:
Provide us with copies of all the graphic, photographic or artistic materials you intend to include in the prospectus prior to its printing and use. Please note that we may have comments. Please also note that all textual information in the graphic material should be brief and comply with the plain English guidelines regarding jargon and technical language.
(Comment no. 2 from https://www.sec.gov/Archives/edgar/data/1180145/0000000000-06-021491-index.htm)

Notes: This figure shows occurrences of comments as well as examples from two comment clusters obtained by applying the DBSCAN clustering algorithm to a set of 49,404 initial SEC comments relating to IPOs. Each vertical, gray line represents a comment letter where a comment from the cluster was issued. The red line indicates the frequency of comments from the cluster issued within a 27-day window.

Figure 3.3 shows occurrences of comments from two clusters over time. The top plot shows the largest cluster identified by us containing 312 comments while the bottom plot shows a smaller cluster with only nine comments. The respective exemplary comments illustrate the similarity of the clustered comments. The baseline parameters yield about 10% clustered comments, 294 clusters, and an average cluster size of 16.4.[22] Note that the identified clusters do not necessarily represent distinct content, i.e. two different clusters can still be quite close.

Recent Standard and Non-Standard Comments. We use the presented evidence on the existence of similar comments and define a *number of recent standard (and non-standard) comments* for each IPO. For all initial comments of a given IPO, we determine the closest comment from a set of recently issued comments for other IPOs. If the matched comment has a cosine similarity in excess of 0.8, we classify the comment to be standard and else to be non-standard in terms of these recently issued comments.

With this approach, we account for the possibility that not all clustered comments are always standard. For instance, see the bottom plot in Figure 3.3, where a few large gaps between the dates are visible. The last comment in this plot is standalone and hence not standard relative to its last issuance date. Moreover, we omit concerns regarding a potential forward-looking bias when determining standard and non-standard comments. For instance, presumably, even the "earliest" comment of a large cluster was likely not standard at the time of its first issuance. With this approach we follow Hanley and Hoberg (2010) who also use past IPOs when calculating standard and informative content. In order to use only recent comments, we compare with the ten most recent IPOs.

There are at least two other ways of defining "recent". First, we could also compare to all comments issued within a constant time window, e.g. the past 90 days. However, by this method, we would have very large variation of the effective number of comments to compare with since IPO filing volumes vary. In doing so, we would mechanically find more similar comments when many IPOs are filed because we also compare to more comments. However, we want to assure that we compare to a broadly constant number

[22] Changing the parameters can also alter these numbers. For instance, a larger ε as well as a smaller m yields more clustered comments. For the values of the parameters we have tested, the percentage of clustered comments varies from about 5% to about 21%.

of past comments. Second, we could only consider past IPOs of the same industry or industry office. However, since there are sometimes only a few IPOs per industry, this would require us to include too old IPOs. Instead, the ten most recent IPOs are typically within 21 (1st quartile) and 45 (3rd quartile) days before the IPO, which appears to be sufficiently recent.

3.3 Measuring Filing Review Workload for SEC Offices

In this section, we describe how we construct our workload measure. Details can be found in Subsection 3.3.1 and initial tests for the measure in Subsection 3.3.2.

Generally speaking, the time required to accomplish any task should depend on its extent, the processing quality, and the resources allocated to its realization. Hence, the work of a SEC team entrusted with a specific filing review may be influenced by the amount of concurrent work at that time since it reduces available resources. Intuitively, one would suspect that especially (too) high workload affects the outcome of a review negatively, for instance with respect to quality or time. Such ideas have recently been tested. Ege et al. (2020) focus on unexpected workload from reviews of transactional filings, e.g. IPO and M&A filings, and consequences of high workload to reviews of periodic filings, e.g. 10-Ks and 10-Qs. Indeed, they find quality losses of periodic reviews measured by the number of comments, the involvement of a supervisor, and the tendency to induce disclosure changes. Instead, Gunny and Hermis (2020) analyze the impact of expectable high workload due to clustering of firms' fiscal year-ends at the calendar year-end. Together, both papers suggest that the SEC staff is influenced by high workload. Since reviews of periodic filings are affected by high transactional filing volume, they might buffer this workload already to an extent that the reviews of transactional filings themselves are not influenced. Whether or not there is a relation is an empirical question, which we examine in this study for the case of IPOs.

3.3.1 The Workload Measure

Our daily abnormal workload measure is constructed at the CF office level for each workday. The core of this measure is the estimation of the number of filings currently in urgent review for each office of the Division of Corporation Finance (CF). We perform a regression of today's raw workload numbers on past values to obtain an abnormal workload measure. Using the abnormal workload, we define a high workload dummy variable so that 20% of all workdays across all offices are classified as high workload.[23] For each filing, we measure high workload on its filing date. Since initial filings of IPOs create a large share of the workload, 40% (see Table 3.1) of them are classified as high workload, which is considerably more than the 20% threshold.

The workload measure is similar to and inspired by the one proposed by Ege et al. (2020). However, we differ in the following details: daily measurement instead of a monthly, slightly enlarged set of filings, and the introduction of hypothetical workload for calculations of abnormal workload to account for SIC Code swaps between offices.

During our sample period from 2004 - 2018 the Division of Corporation Finance consisted of eleven major offices (Office 1 - 11) and one to three rather minor offices.[24] Each office is managed by an Assistant Director[25] and historically endowed with 25 - 35 employees.[26] Filings to review are assigned to the offices by a time-changing industry mapping based on the Standard Industry Classification Codes (SICs).[27] The following paragraphs contain a detailed description of how we construct the workload measures.[28]

Step 1: EDGAR Index and Workdays Our approach is based on the estimation of the number of filings in urgent review for each office. We start after 14th January 2003 and estimate these numbers only for SEC workdays, which we determine from the EDGAR master index file. The focus on workdays simplifies a meaningful consideration of filings in review. An analysis of the EDGAR index reveals that the number of filings on weekends differs considerably from weekdays (2,082 filings on average on weekdays vs. less than one filing on weekend days on average). The maximal number of filings on

[23] By construction, the percentage of high workload days across offices can vary.

[24] From 14th January 2003 till 31st October 2019 we denote Office 12, the Office of International Corp Fin/99, and the Office of Structured Finance (OSF) as minor offices since they did not exist in all subperiods and have systematically lower filings counts, see the filing count plots in Figure B.1 in the Appendix. After 1st November 2019 a larger structural reform reduced the number of major offices to seven and the number of minor offices to two (pre-existing Offices of Structured Finance and International Corp Fin).

[25] Hence, the offices are sometimes called Assistant Director Offices (ADOs).

[26] See `https://web.archive.org/web/20150225012952if_/https://www.sec.gov/divisions/corpfin/cffilingreview.htm`.

[27] Hence, the offices are sometimes called industry offices although the pooled SIC Codes are not always very related.

[28] Those who are not interested in details can skip to the initial validity tests in Section 3.3.2

a weekend day is 76. Hence, we use a threshold of 100 filings to distinguish workdays from non-workdays in the EDGAR index.[29] The few filings filed on non-workdays are shifted to the next workday in order to count them properly.

Step 2: Form Types to Review The term *urgent* refers to the fact that not all eventually reviewed filings are time-sensitive, which is approximately the distinction between periodic and transactional filings in terms of urgency. Ege et al. (2020) provide a comprehensive overview of transactional form types, what they typically contain, and how certain their review is. Based on this discussion, Ege et al. (2020) use form types S-1, S-4, SC 13E3, and PREM14A (as well as their amendments) for their filing counts. We extend this list and use additionally the form types DRS, F-1, SB-2, and F-4 as well as their respective amendments. DRSs were introduced with the JOBS Act in 2012. In the cases where a firm files its prospectus confidentially via a Draft Registration Statement, the draft is subject to SEC review and replaces the first public registration statement regarding the review. Hence, DRS filings add to the workload. Furthermore, DRS filings do not only represent S-1s but also other registration statement form types, which are also part of our IPO sample. This is why we include also F-1 and SB-2.

Step 3: Matching Filings and Offices The CF assigns filings to industry offices by SIC Code. However, this mapping changes over time, which is why we reconstruct it historically via `archive.org`.[30]

The EDGAR index does not contain SIC information. Hence, for all filings having relevant transactional form types, we obtain historical SIC information from the respective EDGAR index-sites of the filings. However, not all index-sites contain SIC information. In these cases, we first try to assign a SIC Code via successor filings. If this also yields no SIC Code, we download the filings and extract the SIC Codes from the filings itself where possible. From all 149,975 relevant filings, we omit the 405 filings where we could not obtain a sufficiently timely SIC Code (0.27%).

The office assignments obtained by a combination of these two data sets are not always unequivocal. First, in some periods, there is no clear mapping between some

[29] This leads to 250 till 252 workdays per year with a median of 251 days.

[30] For instance, one historical snapshot is `https://web.archive.org/web/20140122054224/https://www.sec.gov/info/edgar/siccodes.htm` whose mapping was valid after 01/03/2011 (until next change).

SIC Codes and offices. For instance, SIC Code 7389 in 2011 is assigned to Office 2 and 3. Second, the EDGAR index contains multiple records for filings with several filing CIKs. In some of these cases, we obtain different SIC Codes and different offices for a single filing. We make use of all office possibilities and perform a step-wise weighting as follows: all filings are weighted with the reciprocal of the number of step-wise office possibilities. Step-wise refers to cases of the following kind: a filing is assigned to SIC Code 7389 in 2011 (Office 2 and 3) and to SIC Code 7385 (Office 11). In the first step, we weight both SIC possibilities, in the second step, we weight the office possibilities. This leads to the following weighting: Office 2 (25%), Office 3 (25%), and Office 11 (50%). However, such cases occur infrequently.

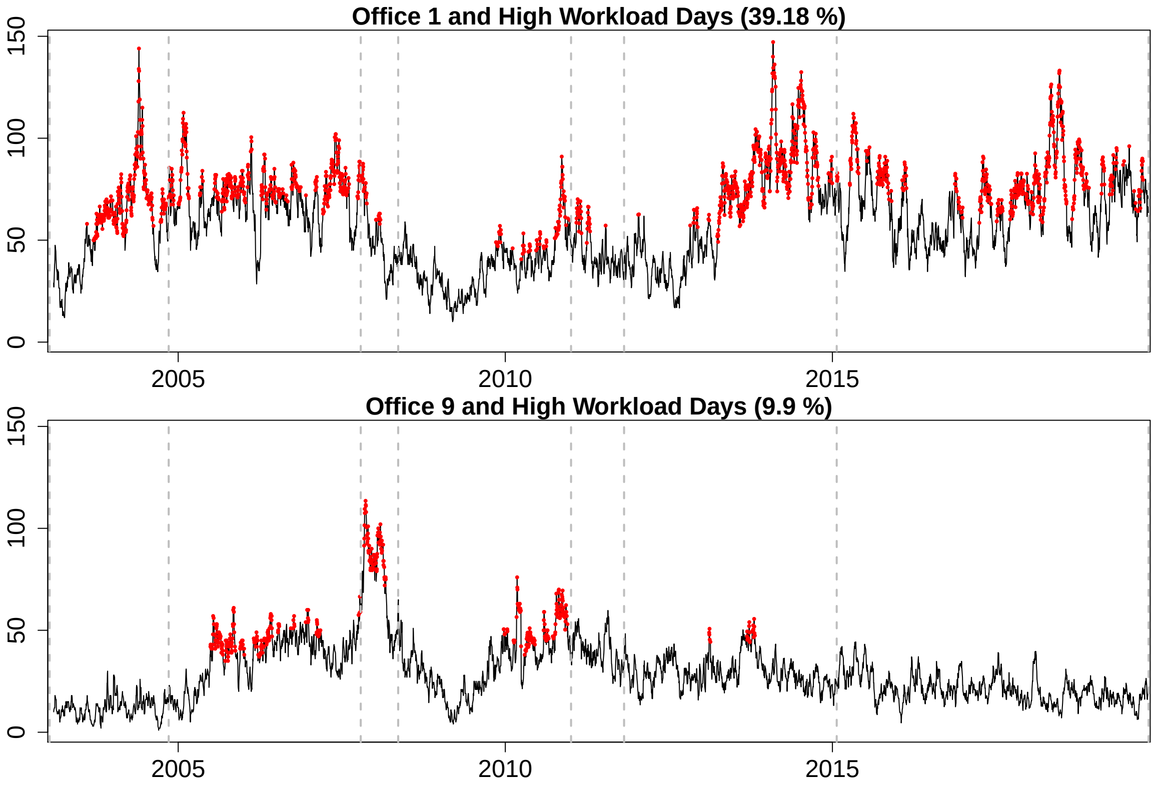
Step 4: Review Times and the Estimated Number of Filings in Review We assume that each filing of a specific form type is reviewed and that the review lasts a constant number of workdays, depending on the form type. Supported by the declining response times for later letters presented in Table 3.1, Panel B, we distinguish between initial and amended filings. We assume 17 workdays in review for all original filings and 5 workdays for all amended filings. Both choices are somewhat below their empirical means in Table 3.1, Panel B. This increases the fraction of filings that were indeed still under review at the time. Subject to these assumptions we calculate the estimated number of filings in review $w_{i,t}$ for office i and workday t as the sum over the weights mentioned in Step 3 for all relevant filings. Figure 3.4 presents $w_{i,t}$ time-series for Offices 1 and 9.

Step 5: Models for Abnormal Workload Based on the raw filing counts and following Ege et al. (2020), we calculate abnormal workload using a pooled regression. First, this is a convenient method to enhance the comparability of workloads across offices. Second, it allows incorporating both assumptions on how the SEC predicts future workloads and how flexible the SEC is regarding reducing potential workload consequences.

In our framework, the workload $w_{i,t}$ on day t for office i is explained by past (average) workloads $\bar{w}^c_{i,t,s,a}$, that is:

$$w_{i,t} = \beta_0 + \sum_{k=1}^{K} \beta_k \bar{w}^c_{i,t,s_k,a_k} + \epsilon_{i,t}, \tag{3.3.1}$$

Figure 3.4: Estimated Number of Filings in Review and High Workload Days



Notes: This figure shows time-series of workload as measured with the estimated number of filings in review for Offices 1 and 9. The gray, vertical, dashed lines indicate the dates where the SEC changed the SIC ranges for some of the offices. The red dots indicate high workload at the $c = 80\%$ level used throughout the paper.

for a specific period $t \in T$ and offices $i \in I$ where $\bar{w}^c_{i,t,s,a} := \sum_{j=t-s+1-a}^{t-s} w^c_{i,t}/a$. In this definition, c can represent actual workload, $w^{act}_{i,t} = w_{i,t}$ or hypothetical workload, which we motivate in the following.

Actual vs. Hypothetical Workload. We distinguish between actual and hypothetical workload to account for the changes in the SIC-office mappings over time. While "actual" refers to the historical, true workload an office was confronted with calculated with the valid SIC-office mapping on that day, "hypothetical" workload builds upon the current valid SIC-office mapping. We regard the latter option as more realistic in terms of resource allocation planning. The difference is illustrated in Figure 3.5.

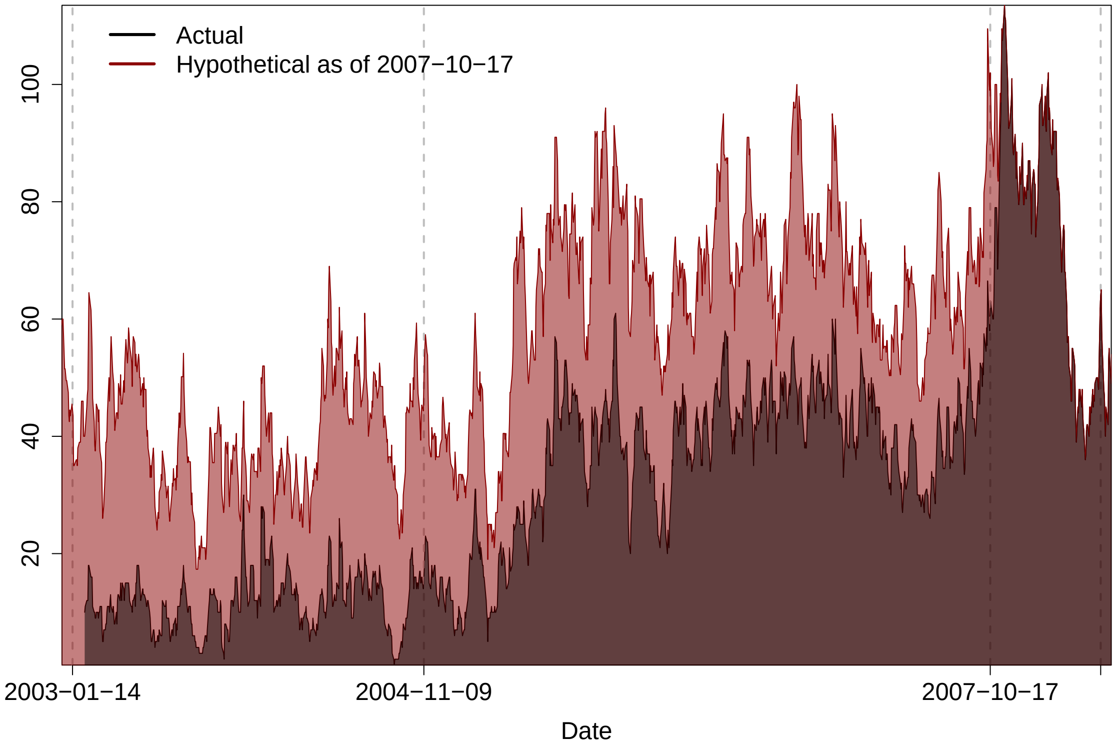
Figure 3.5 presents an extract of the actual workload for Office 9 (black), already contained in Figure 3.4. Additionally, the plot shows the hypothetical workload as of 17th October 2007 (red) where a considerable change to the SIC Code range of Office 9 was introduced.[31] Measured with the office-SIC mapping of that time, the past hypothetical workload is substantially larger than the actual one. We believe that it is more sensible to use hypothetical workloads to obtain abnormal workloads since it accounts for changes in the SIC Code range, which are most likely part of the SEC planning. Hence, the strong workload spike after 17th October 2007 can at least partly be attributed to the increase of the SIC Code range. Moreover, hypothetical workloads increase the number of days where an abnormal workload can be calculated since they are available for any date. This comes in handy for the SEC office structure change in November 2019 since it allows to calculate meaningful abnormal workload already for the first day of its effectivity.

Unexpected and Abnormal Workload. The choice of the parameters $s_1, s_2, \ldots$ and $a_1, a_2, \ldots$ is connected to an assumption of how the SEC plans workload and how the SEC is able to deal with expected workload. Eventually, we attempt to identify phases where the staff is most likely to face stress-inducing, abnormal workload since such workload could be associated with negative consequences. Obviously, the knowledge of upcoming high workload will not necessarily reduce the stress induced by the workload. How it is dealt with matters as well.

We choose $s_1 = 251$, $a_1 = 21$, $s_2 = 502$, and $a_2 = 21$, which is similar to Ege et al. (2020). This assumes that the SEC uses a planning horizon of two years and is able to react at the monthly frequency.

[31] The number of SIC Codes assigned to Office 9 increased from one to 39.

Figure 3.5: Actual vs. Hypothetical Estimated No. of Filings in Review of Office 9



Notes: This figure shows actual and hypothetical time-series of workload as measured with the estimated number of filings in review for Office 9. The black line indicates actual workload similar to Figure 3.4 while the red line indicates hypothetical workload as of 17th October 2007. Quickly after this date, actual and hypothetical workload coincide perfectly by definition. However, in the prior periods the hypothetical workload is substantially larger. Note that the time-series of actual workload starts only with some delay after the first date of the SIC-office matching used in this study (14th January 2003) while the time-series of hypothetical workload is calculated for each date.

Step 6: Estimation Techniques Using hypothetical workloads as regressors, we perform a full sample regression from 6th February 2003 to 31th October 2019 including all major offices, that is Offices 1 - 11. The residuals from these regressions $\hat{\epsilon}_{i,t}$ are transformed to empirical probability integral transforms $\hat{p}_{i,t} = \bar{F}(\hat{\epsilon}_{i,t})$ (PITs) where $\bar{F}$ is the empirical cdf of all residuals. We use these *Workload PITs* to define days with high workload $HW_{i,t,\alpha}$ via a threshold α as $HW_{i,t,\alpha} = \mathbb{1}_{\{\hat{p}_{i,t} \geq \alpha\}}$. Our high workload threshold is $\alpha = 20\%$ throughout the paper.[32]

Pitfalls of Workload Measurement There are some issues that may disturb the workload measurement. First, the EDGAR index misses a few filings (e.g. some confidentially filed Draft Registration Statements from 2012). Second, probably not all filings considered by us are getting reviewed. Third, the form type alone does not determine review workload. For instance, S-1 filings not associated with IPOs are sometimes only subject to a limited review. Another example would be that S-1 filings subsequent to a DRS should rather be interpreted as an amendment in terms of review effort. Fourth, the matching between filings and CF offices is not always unambiguous.

3.3.2 Initial Evidence: Does the Workload Measure Capture Stress?

Test 1: SIC Code Office and Signer Office In order to test the workload measure, we perform two tests. First, we match our IPO list to the SEC offices based on the first available SIC Code for the IPO from EDGAR. We call the resulting office *SIC Code office*. For each IPO, we expect that the SIC Code office coincides with the office associated with the signer (*signer office*) contained in the first letter. While this is usually the case, we identify 35 IPOs where we suspect that the SIC Code office did not actually perform the review. One potential explanation is that the SIC Code office was under too high workload and the signer office performing the review was not. We perform a logit analysis where we attempt to explain the detected office changes via

[32] Most of the results presented in this study are similar when we lower the threshold, e.g. to 70%, i.e. classify more IPOs as being under high workload. However, if we raise the threshold, e.g. to 90%, some results get weaker. This suggests that many of the IPOs above the 80% threshold (but below 90%) are indeed subject to high workload and should not be classified otherwise.

high workload in both office variants. Workload is measured on the date of the first IPO filing. Logit regression results are presented in Table 3.2.

Table 3.2: SIC Code Office, Signer Office, and High Workload

	Dependent variable: Signer does not belong to SIC Code Office (D)					
	(1)	(2)	(3)	(4)	(5)	(6)
High Workload$_{\text{SIC Code Office}}$ (D)	6.474***	6.671***	0.953***	0.983**		
	(4.082)	(4.008)	(2.756)	(1.987)		
High Workload$_{\text{Signer Office}}$ (D)	−6.247***	−6.405***			−1.497	−1.557
	(−3.020)	(−2.995)			(−1.031)	(−1.049)
ln(Age)		0.268		0.177		0.181
		(0.702)		(0.577)		(0.756)
ln(Sales)		0.417***		0.347***		0.357***
		(5.161)		(8.417)		(6.168)
Leverage		−0.029		0.164*		0.142
		(−0.263)		(1.877)		(1.598)
Pos. EPS (D)		−0.349		−0.163		−0.120
		(−0.722)		(−0.234)		(−0.226)
VC (D)		1.229		0.648		0.625
		(1.386)		(0.590)		(0.661)
Bookrunner Market Share		0.270		−0.280		−0.479
		(0.181)		(−0.223)		(−0.404)
Lawyer Market Share		−1.378		1.188		1.926
		(−0.309)		(0.229)		(0.385)
Big 4 (D)		−0.421		0.069		−0.183
		(−0.600)		(0.121)		(−0.432)
Prospectus Type (D)	Included	Included	Included	Included	Included	Included
Fixed Effects	SEC Office Year Month	SEC Office Year Month	SEC Office Year Month	SEC Office Year Month	SEC Office Year Month	SEC Office Year Month
Observations	922	922	922	922	922	922
Pseudo R^2	0.485	0.486	0.079	0.082	0.097	0.108

Notes: This table presents logit regression results for two different variants of matching SEC offices to IPOs. The dependent variable is a dummy indicating that the office matched via SIC Code does not coincide with the SEC office matched via the signer of the first SEC comment letter. These regressions provide a first test for the proposed workload measure. Main independent variables are high workload dummies for both office variants as measured on the filing date of the first IPO filing. Age is the age of the IPO firm, calculated with founding dates from Prof. Jay Ritter's website. Sales, Leverage, and Earnings per Share (EPS) are accounting variables from Compustat. VC is a dummy from SDC indicating Venture-Capital backed IPOs. Bookrunner (Lawyer) Market Share is the two-year trailing market share of the lead underwriter (law firm). Big 4 is a dummy variable indicating the auditor is a Big 4 audit firm. Prospectus Type (D) include dummies for the initial IPO prospectus type. See Table B.1 in the Appendix for detailed definitions and sources of the variables. The numbers in brackets below the coefficient estimates show *t*-statistics. Standard errors are clustered by SIC Code Offices. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

Table 3.2 shows that office changes are related to workload. We find that high workload of the SIC Code office is related to an increasing change likelihood and that high workload of the signer office impedes an office change. The effect is unchanged when we incorporate standard IPO control variables.

Test 2: SIC Code Swaps between Offices The second test focuses on the occasional SIC Code swaps between offices as mentioned in Step 3 of this Section. Again, a potential reason for such SIC Code swaps would be to balance workload across offices. In contrast to our first test for the workload measure, here, it is more sensible to consider the full range of workload and not only peaks.[33] For each SIC Code and swap date from 9th November 2004 to 25th January 2015, we predict changes in the mapped SEC offices using the average workload of the old and new office one year till one month prior to the swap date. Results of logit regressions can be found in Table 3.3.

Table 3.3: SIC Swaps between SEC Offices and Workload

	Dependent variable: SIC Swap occurred (D)		
	(1)	(2)	(3)
$\text{Workload}_{\text{Old Office}}$	45.415***	5.102***	
	(4.224)	(3.137)	
$\text{Workload}_{\text{New Office}}$	−52.746***		−13.209***
	(−3.817)		(−4.614)
Fixed effects	SIC, Date	SIC, Date	SIC, Date
Observations	2618	2626	2624
Pseudo R^2	0.796	0.399	0.48

Notes: This table presents logit regressions results for SIC Code swaps between SEC Offices on the six change-dates from 9th November 2004 to 25th January 2015. The dependent variable is a dummy indicating that the SIC Code was swapped to another office at the corresponding date. Independent variables are workload measures for the new and the old office calculated as the average of the daily workload from one year to one month prior to the corresponding swap date. The numbers in brackets below the coefficient estimates show *t*-statistics. Standard errors are clustered by SIC Code. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

Table 3.3 shows that swaps of SIC Codes are related to workload. Summarizing, the results of both tests support the workload measure and the idea that actions undertaken by the SEC are related to it.

[33] SIC Code rebalancing should not be restricted to high workload offices since, for instance, swaps between low and medium workload offices are also sensible.

Test 3: Self-reported SEC Workload In their annual performance reports, the SEC discloses actual workload data (as well as estimated and requested numbers) for several types of reviews (e.g., Reporting Company Reviews, New Issuer Reviews, . . .) at the annual level (see for instance SEC, 2019b). We extract the actual numbers for the years 2012-2019 from the reports.[34] Then we calculate a time-series of daily average workload PITs across offices 1 to 11. For this time-series, we calculate yearly workload PIT averages and regress the logarithmized workload data from the SEC reports on these. Results can be found in Table 3.4.

Table 3.4: Self-Reported SEC Workload and Estimated Workload

	New Issuer Reviews	Reporting Company Reviews	Total Reviews
	(1)	(2)	(3)
Workload	0.898*** (3.960)	0.050 (0.377)	0.154 (1.343)
Observations	8	8	8
R^2	0.649	0.006	0.063
F Statistic	11.071**	0.035	0.402

Notes: This table reports results for OLS regressions (with intercept) of logarithmized self-reported yearly SEC Workloads (number of reviews) on yearly estimated workload averages across offices 1 - 11 in a small sample. The numbers in brackets below the coefficient estimates show *t*-statistics based on robust standard errors with small sample size adjustment. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

Although there are only eight observations, we find that our workload measure is related to the self-reported SEC workload data for “New Issuer Reviews” but not for “Reporting Company Reviews” or “Total Reviews”.

3.4 Quality of Comment Letters and High Workload

The general quality of a comment letter is difficult to determine. Ultimately, this would require a content-based assessment of the comment letter (to analyze the comments that were issued) and the reviewed document (to detect potentially missed comments). Consequently, it is easier to fall back on relatively simple measures related to quantity such as the number of comments or the number of words, which we coin *quantitative quality.*[35]

[34] Older data seems not to be available.

[35] The number of comments was already used by Ege et al. (2020) as an output-based quality measure. Furthermore, they use the number of topics (from Audit Analytics) as an output-based measure,

In the textual comment analysis in Section 3.2.4, we found that a considerable amount of comments are similar to previously issued comments in antecedent letters for other IPOs. Based on this observation, we classify each comment of each SEC letter as a relatively standard or rather non-standard comment. It seems reasonable to expect that more standard comments or fewer non-standard comments are related to lower *content quality*. Readily available and somewhat generic comments may substitute unique, firm-specific comments that require more resources to produce.

There are several arguments for and against a relationship between high regulator workload and IPO reviews. In contrast to periodic filings, IPO filings are of a transactional type. Thus, there is a certain degree of time pressure associated with their assessment. Intuitively, time pressure and workload should add to stress and may lead to quality reductions. On the other hand, there are several reasons why high workload effects are not necessarily present. While Ege et al. (2020) and Gunny and Hermis (2020) document consequences for periodic reviews, their findings are consistent with the notion that these occasional, not time-sensitive reviews can be used as a buffer for time-varying workloads induced by transactional filings, including IPO filings. Furthermore, there might be several mechanisms to cope with high workloads, such as using efficiency leeways or activating additional workforce within the SEC.

3.4.1 Quantitative Quality

First, we focus on quantitative quality as measured by the number of comments. We diverge from the comment letter literature by focusing not only on the first letter (Cunningham and Leidner, 2019) but also on all *subsequent* letters after the first one by making use of the comment letter matching described in Section 3.2. Workload is always measured on the filing date of the corresponding IPO filing. Results of entropy-balanced negative binomial regressions can be found in Table 3.5.

Entropy-balancing is a data preprocessing method proposed by Hainmueller (2012) to balance a sample with respect to moment conditions of the covariates when estimating the effects of binary variables. We use it throughout this study to balance the covariate distributions across high workload and non-high workload observations. In

a supervisor's involvement as a measure of input-quality, and whether the firm states that it will amend or revise filings.

all regressions, we balance with regard to all standard IPO control variables using the high workload dummy as the treatment.[36]
First, we focus on the initial letter. Table 3.5 shows no detrimental effect of high workload on the number of comments in the first letter. Neither are there effects of review size, which is the size of the first prospectus (including exhibits but excluding images), and the two market variables. In contrast, older issuers tend to receive fewer comments, and firms with higher sales get more comments. Both results are consistent with the findings of Lowry et al. (2020) who analyze determinants of topics within the first letter. They find that age is negatively related to the extent of almost all topics and that the company size (most close variable to sales) is positively related to all topics, especially revenue recognition. Regarding IPO companions, we find several significant negative relations. Venture-capital backed IPOs, IPOs accompanied by large market share lawyers, and issuers audited by a Big 4 firm receive considerably fewer comments.

If we focus only on the subsequent letters, we find in the pooled specification (2) that high workload is associated with about 6% fewer comments, which is statistically significant. However, considering that the average number of comments for subsequent letters is about 4.4, this is effectively not a sizeable decrease, but it indicates existing workload effects. Furthermore, IPO letters with larger review sizes (defined as the size of all filed exhibits for the subsequent letters) receive more comments. Often, the control variables have qualitatively similar effects compared to specification (1). Lawyer Market Share and the Big 4 dummy approximately double their coefficients. Additionally, more indebted issuers are associated with more subsequent comments.

Most of the presented results are in line with expectations. For instance, an experienced lawyer can help avoid initial SEC concerns and produce better answers and amendments that satisfy the SEC. Similar thoughts apply to reputable audit firms as well as experienced shareholders.

Finally, we introduce an issuer dummy, which removes all IPO invariant covariates and examine the subsequent letters (specification 3) again. Qualitatively, the results for high workload and review size hold, and the high workload dummy coefficient almost

[36] Results for equally-weighted models are both qualitatively and quantitatively similar. Note that we discuss the coefficients of the covariates based on the tabulated results from the weighted models for convenience.

Table 3.5: Quantitative Comment Letter Quality and High Workload

	Dependent variable: #Comments per Letter		
	First Letter	Subsequent Letters	
	(1)	(2)	(3)
High Workload (D)	0.015	−0.061**	−0.110**
	(0.466)	(−2.088)	(−2.015)
ln(Review Size)	−0.006	0.010***	0.011***
	(−0.331)	(3.184)	(2.644)
Market Return$_{30\text{ Days}}$	−0.019	0.049	0.105***
	(−0.361)	(1.417)	(5.026)
Market Vola$_{30\text{ Days}}$	−0.221	−0.110	−0.106
	(−0.688)	(−0.301)	(−0.502)
ln(Age)	−0.076***	−0.074**	
	(−4.207)	(−2.274)	
ln(Sales)	0.064***	0.082***	
	(9.580)	(6.219)	
Leverage	0.003	0.032***	
	(0.627)	(3.222)	
Pos. EPS (D)	0.012	−0.018	
	(0.282)	(−0.770)	
VC (D)	−0.096***	−0.127**	
	(−3.052)	(−2.147)	
Bookrunner Market Share	0.015	−0.025	
	(0.254)	(−0.445)	
Lawyer Market Share	−0.352**	−0.865***	
	(−2.100)	(−7.516)	
Big 4 (D)	−0.121***	−0.244***	
	(−6.511)	(−10.233)	
Prospectus Type (D)	Included	Included	–
Fixed Effects	SEC Office	SEC Office	Issuer
	–	Letter	Letter
	Year, Month	Year, Month	Month
Observations	908	2359	2359
Pseudo R^2	0.530	0.368	0.604

Notes: This table presents results for weighted negative binomial regressions on the number of comments per SEC letter. The weights are estimated by entropy balancing using the presented set of control variables and High Workload as the treatment. High Workload is a dummy variable indicating abnormally high workload of the SEC office responsible for the IPO review process. Review Size is the combined file size of all new exhibits (+ prospectus for the first letter). Market Return (Vola) is the trailing annualized 30-day return (volatility) of the CRSP value-weighted market portfolio. Age is the age of the IPO firm, calculated with founding dates from Prof. Jay Ritter's website. Sales, Leverage, and Earnings per Share (EPS) are accounting variables from Compustat. VC is a dummy from SDC indicating Venture-Capital backed IPOs. Bookrunner (Lawyer) Market Share is the two-year trailing market share of the lead underwriter (law firm). Big 4 is a dummy variable indicating the auditor is a Big 4 audit firm. Prospectus Type (D) include dummies for the initial IPO prospectus type. See Table B.1 in the Appendix for detailed definitions and sources of the variables. The numbers in brackets below the coefficient estimates show t-statistics. Standard errors are clustered by SEC Offices respectively letter number for the panel regressions. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

doubles to 11% fewer subsequent comments. The market return prior to the review start shows a significant effect via doubling its estimate compared to specification (2).

Not all determinants regarding the number of comments for IPOs are similar to findings for periodic filings. For instance, the age coefficient is consistently positive for periodic filings while it is negatively associated here. Instead, the Big 4 dummy is negatively related to both types. Overall, these results provide some support for quantitative quality reductions under a high workload.

A likely important determinant for explaining the number of comments issued by the SEC would be a measure of the true extent of issues present within the IPO disclosure. Supposedly, the preciser the SEC performs its reviews, the more larger the correlation between revealed and true issues would be. While we cannot control for this in the cross-sectional model (1) of Table 3.5, we include an issuer dummy in specification (3) for subsequent comments, which controls for time-invariant general issuer problems. After considering these fixed effects, the high workload coefficient increases, which provides robust support for detrimental high workload effects.

Since the dependent variables in the regressions of Table 3.5 are count variables, we estimate negative binomial count variable models. Especially comment counts for subsequent letters can be small, which makes such models more appropriate. However, if we instead use OLS regressions with the logarithmized number of comments as the dependent variable like some papers of the filing review literature (e.g., Cassell et al. (2013) or Ege et al. (2020)), we typically obtain quantitatively unchanged results.

3.4.2 Content Quality

We now examine the more content-related quality measures based on the similarity of the comments in the first SEC letter to those issued in the ten most recent (first) letters for other IPOs. Based on the maximum similarity of one comment to another, we classify comments as standard and non-standard, see Section 3.2.4. Results of entropy-balanced negative binomial regressions are presented in Table 3.6.

Naturally, standard comments are relatively rare (about 10%). Hence, when we distinguish between standard and non-standard comments, regression results for non-standard comments (specification (2) of Table 3.6) largely resemble the previous results for all comments as presented in specification (1) of Table 3.5 but are often slightly

Table 3.6: Comment Letter Content and High Workload

	#Stand. Com.	#Non-Stand. Com.	Prop.(Stand. Com.)
	(1)	(2)	(3)
High Workload (D)	0.144*** (2.989)	0.008 (0.319)	0.134*** (4.153)
ln(Review Size)	−0.016 (−0.300)	−0.008 (−0.323)	0.056 (1.168)
Market Return$_{30\ \text{Days}}$	0.171 (1.444)	−0.029 (−0.519)	0.153 (1.541)
Market Vola$_{30\ \text{Days}}$	0.991 (0.934)	−0.523** (−2.004)	1.228 (1.591)
ln(Age)	0.005 (0.073)	−0.080*** (−5.734)	−0.002 (−0.038)
ln(Sales)	−0.043* (−1.736)	0.077*** (10.823)	−0.101*** (−5.311)
Leverage	0.012 (0.822)	0.009* (1.687)	−0.023* (−1.956)
Pos. EPS (D)	0.088 (1.357)	−0.002 (−0.052)	0.078*** (4.290)
VC (D)	0.009 (0.158)	−0.110*** (−2.928)	0.108* (1.740)
Bookrunner Market Share	−0.041 (−0.342)	−0.021 (−0.365)	−0.117 (−1.068)
Lawyer Market Share	0.010 (0.021)	−0.372* (−1.908)	−0.007 (−0.021)
Big 4 (D)	−0.004 (−0.055)	−0.128*** (−5.858)	0.088 (1.566)
Prospectus Type (D)	Included	Included	Included
Fixed Effects	SEC Office Year, Month	SEC Office Year, Month	SEC Office Year, Month
Observations	902	902	902
Pseudo R^2	0.266	0.510	0.439

Notes: This table presents results for weighted negative binomial regressions on the number of (standard, non-standard) comments as well as for a fractional regression on the proportion of standard comments in the first SEC letter. (Non-)Standard refers to the similarity between the comments of the corresponding SEC letter to the comments issued in antecedent letters. Proportion(Standard Comments) is the relative proportion of comments that are similar to comments issued in antecedent letters. The weights are estimated by entropy balancing using the presented set of control variables and High Workload as the treatment. High Workload is a dummy variable indicating abnormally high workload of the SEC office responsible for the IPO review process. Review Size is the combined file size of all new exhibits (+ prospectus for the first letter). Market Return (Vola) is the trailing annualized 30-day return (volatility) of the CRSP value-weighted market portfolio. Age is the age of the IPO firm, calculated with founding dates from Prof. Jay Ritter's website. Sales, Leverage, and Earnings per Share (EPS) are accounting variables from Compustat. VC is a dummy from SDC indicating Venture-Capital backed IPOs. Bookrunner (Lawyer) Market Share is the two-year trailing market share of the lead underwriter (law firm). Big 4 is a dummy variable indicating the auditor is a Big 4 audit firm. Prospectus Type (D) include dummies for the initial IPO prospectus type. See Table B.1 in the Appendix for detailed definitions and sources of the variables. The numbers in brackets below the coefficient estimates show *t*-statistics. Standard errors are clustered by SEC Offices respectively letter number for the panel regressions. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

stronger in terms of coefficients and significances. As before, the high workload dummy shows no effect. We find a difference for market volatility whose estimate approximately doubles and becomes significant. Research for periodic filings has shown that high firm volatility is associated with the receipt and extent of comments (Johnston and Petacchi, 2017; Cunningham and Leidner, 2019). Since higher market volatility is driven by an increase in firm volatility for many firms, our market volatility effect may be associated with an attention shift from transactional filings to periodic filings.

In contrast, the standard comments' coefficients and significances are often different from those estimated for all or non-standard comments. Most importantly, with 14.4%, the high workload dummy significantly associates with more standard comments supporting the notion of less quality. The variables Age, the VC dummy, Lawyer Market Share, as well as the Big 4 dummy are no longer significant. Interestingly, the sign of Sales flips. Moreover, note that non-standard comments are easier to explain ($R^2 = 0.510$) than standard comments ($R^2 = 0.266$).

Note that the regressions in this subsection are performed with six IPOs less than specification (1) of Table 3.5. The reason for this is due to the fact that we compute standard and non-standard comments based on the ten most recent IPOs. There are only six IPOs missing because the computation is conducted on all IPOs where we matched either all SEC letters or the first one successfully to the IPO filings and not only the regression IPOs.

The high workload results regarding the standard and non-standard content are similar when we increase the number of recent IPOs to moderately larger values, e.g., 20, 30, or 40, but slightly weaker. Such an increase is always accompanied by a sample size reduction and by comparisons to older comments. This suggests that the timeliness of this measure matters.

3.5 Cost of Remediation and High Workload

Initial Public Offerings (IPOs) often spend considerable time in registration. The time between the first prospectus and the first trading day on CRSP in our sample is 156 days on average. To complete the average review, the SEC needs 58 days (answers by the issuer excluded, otherwise 125 days), representing 37% (80%) of the overall registration length.

From an issuer perspective, an exogenously prolonged registration period should generally be avoided as they are associated with costs for several reasons. First, going public is a major step for a company and requires considerable attention from the issuer, especially at the management level. A delayed IPO may hence distract the company additionally from conducting and developing its actual business (Falato et al., 2014). Second, issuers tend to time their offerings to capture favorable conditions resulting in IPO waves (Benninga et al., 2005; Pástor and Veronesi, 2005; Ibbotson and Jaffe, 1975). Third, any additional day in registration adds to the risk of a deteriorating stock market, which increases the risk of withdrawal (Busaba et al., 2001), which would harm not only the issuer but also the reputation of the underwriters (Dunbar, 2000). Fourth, since IPOs are often a way of financing, the speed with which the proceeds become available should matter. Finally, Chaplinsky et al. (2017) note that the time in registration is also positively associated with the direct costs of an IPO, such as fees or gross spread. While these aspects are particularly important for IPOs, Cassell et al. (2013) provide a similar discussion for the review process of annual reports.

If the SEC is unable to compensate for abnormally high workload, there is no clear expectation on the relation between high workload and response times. On the one hand, the SEC could delay their review tasks in order to guarantee a certain level of quality.[37] On the other hand, the SEC may reply quicker for reasons such as increased efficiency or decreased quality. The idea of improved efficiency would go hand in hand with unexploited capacities in lower workload times, while a decreased quality would be accompanied by fewer average resources allocated to the reviews.

We model the response times of the SEC in two different dimensions. In the first dimension, we aggregate the response times (in days) for each IPO to proxy the full time in active SEC review. We regard this as modeling the remediation costs solely related to the SEC review[38]. Secondly, we analyze the letter-level response times to estimate the association between high workload and the number of workdays needed by the SEC to review a specific amendment of the IPO. In both variants, we include only IPOs in which the SEC letters are consecutive, i.e. where each consecutive amendment

[37] The SEC staff sometimes addresses this possibility in their review letters, see for example: `https://www.sec.gov/Archives/edgar/data/1533932/000000000011067372/filename1.pdf`.

[38] Cassell et al. (2013) define the time from the first letter until the last letter, including the response times by the issuer, as remediation costs. Since the matching between comment letters and IPO filings allows to decompose this period, we are able to focus solely on the SEC induced period.

of the IPO received a letter until the last issued letter by the SEC. Generally, the full time in active SEC review is not observable since presumably all IPO filings are getting reviewed but not necessarily receive comments. However, for IPOs with a clear, simple filing-letter structure, we can observe the time in active SEC review until all SEC concerns are resolved, which should be a good proxy for the time in SEC review. Further, we do this to focus on IPOs where timing is more likely to matter, as made evident by the fact that all essential material was filed early in the process, and to avoid measurement error.[39]

Our empirical approach consists of Cox (1972) proportional hazard models. Hazard models are regression models widely used for analyzing duration data, typically used in medical studies to model the effect of a medication on patients' survival times. In economics, hazard models have, for example, been employed to study the duration of venture capital investments (Gompers and Metrick, 2001), forecasting bankruptcy (Shumway, 2001), or CEO turnover (Hazarika et al., 2012), among other topics.[40]

Particularly suitable for our purpose, Cox proportional hazard models allow for time-varying covariates. This enables us to include our workload estimates at a granular resolution. More precisely, at the aggregated level, we employ a high workload dummy on the filing date of each filing and at the letter-level, we use a daily (each workday) high workload dummy time-series.

The Cox model is expressed by a hazard function h

$$h(t) = h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p) \tag{3.5.1}$$

that can be interpreted as the probability of SEC review completion at day t where $h(t)$ is the hazard function determined by a set of p covariates. The coefficients $\beta_1, \ldots, \beta_p$ measure the effect size of the covariates, similar to multivariate linear regressions.

The central assumption of the model is that each covariate has multiplicative and time constant effects. We test this assumption based on Grambsch and Therneau (1994) and find it to be violated when including the first SEC response letter. That is not

[39] For instance, two early IPO filings can receive comments, the third and the fourth one not, and then again the fifth one quite a time later. In such cases, it is probably not plausible to consider the time in active SEC review.

[40] We also formulate regression models analog to Table 3.5. Untabulated results are qualitatively similar and available upon request from the authors.

surprising as the response times for this letter seem to be a result of internal guidelines and clusters heavily around 27 days. In an untabulated regression, we find that the little response time variations (cf. Figure 3.2) are not explainable by our set of variables. Once we exclude the first letter, the model is well specified.[41]

Table 3.7 presents the results of the two estimated models. It is striking that high workload is in both models associated with a significant decrease in response times by the SEC reviewers. The results of the hazard models suggest a reduction of up to 26% ($\exp^{0.23} = 1.259$).[42] The letter-level hazard model confirms the estimated effect.

Comparing the results with those documented for the number of comments raised by the SEC in Section 3.4, we find the SEC to issue slightly fewer comments, more standard content, but also to respond faster after the first letter. Noteworthy, we find IPOs accompanied by a Big 4 auditor are not only associated with significantly fewer comments issued by the SEC (−12% for the first, −24% for the subsequent letters, cf. Table 3.5) but are also associated with significantly lower remediation costs in terms of response times by the SEC (−29%).

[41] The results, however, remain qualitatively similar when including the first letter.

[42] Note that this effect is not representative of the full registration length of an IPO but rather for the time the IPO is actively under review by SEC staff.

Table 3.7: IPO Remediation Costs and High Workload

	Dependent variable: Variants of Response Time	
	Hazard Models	
	(1)	(2)
High Workload (D)	0.230***	0.250***
	(2.896)	(5.120)
ln(Review Size)	0.011	0.003
	(0.968)	(0.250)
Market Return$_{30\ \text{Days}}$	0.099	−0.102
	(1.454)	(−1.284)
Market Vola$_{30\ \text{Days}}$	−1.870**	−0.567
	(−2.499)	(−0.959)
ln(Age)	0.057	−0.154***
	(0.606)	(−3.296)
ln(Sales)	−0.167***	−0.014
	(−6.262)	(−0.770)
Leverage	−0.004	0.034***
	(−0.206)	(2.676)
Pos. EPS (D)	0.115	0.042
	(0.960)	(0.509)
VC (D)	0.118	0.013
	(1.330)	(0.349)
Bookrunner Market Share	0.874***	0.343***
	(3.265)	(2.882)
Lawyer Market Share	−1.383	−1.190
	(−1.257)	(−0.780)
Big 4 (D)	0.256***	0.241***
	(3.826)	(3.509)
Prospectus Type (D)	Included	Included
Fixed Effects	SEC Office – Year, Month	SEC Office Letter Year, Month
Observations	1398	12969
Pseudo R^2	0.081	0.026

Notes: This table presents results for two weighted Cox proportional-hazard regressions on variants of SEC response time. The weights are estimated by entropy balancing using the presented set of control variables and High Workload as the treatment. The dependent variable in model (1) is the sum of all consecutive letter-level response times (in calendar days) for each IPO. In model (2), the dependent variable is the response time (in workdays) at the letter-level. Note that the signs of the coefficients in a Cox regression relate to hazard and hence need to be oppositely interpreted to OLS coefficient signs. High Workload is a dummy variable indicating abnormally high workload of the SEC office responsible for the IPO review process. Review Size is the combined file size of all new exhibits (+ prospectus for the first letter). Market Return (Vola) is the trailing annualized 30-day return (volatility) of the CRSP value-weighted market portfolio. Age is the age of the IPO firm, calculated with founding dates from Prof. Jay Ritter's website. Sales, Leverage, and Earnings per Share (EPS) are accounting variables from Compustat. VC is a dummy from SDC indicating Venture-Capital backed IPOs. Bookrunner (Lawyer) Market Share is the two-year trailing market share of the lead underwriter (law firm). Big 4 is a dummy variable indicating the auditor is a Big 4 audit firm. Prospectus Type (D) include dummies for the initial IPO prospectus type. See Table B.1 in the Appendix for detailed definitions and sources of the variables. The numbers in brackets below the coefficient estimates show *t*-statistics. Standard errors are clustered by SEC Offices respectively letter number for the panel regressions. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

3.6 IPO Pricing and High Workload

Primary Market Pricing The standard track of an IPO starts with the filing of a preliminary prospectus, which typically does not contain price ranges or shares offered. At some point, the issuer files an amendment containing an initial price range or an expected price as well as the number of shares offered. Together, they determine the expected offer size. If the issuer or the underwriter receives information during the bookbuilding, the price or the number of shares can be revised at any time.

Hanley and Hoberg (2010) find evidence suggesting a trade-off regarding the information production every issuer faces when conducting an IPO. On the one hand, issuers can decide to perform costly information production on their own via due diligence. That would allow the issuer to obtain a more substantiated value estimate, which will be believed by the market if it also yields more informative disclosure. Alternatively, if the aggregated costs (such as the use of advisors) or risks (such as disclosure of proprietary information) of this self-reliant information production are too high, issuers can also decide to produce less information on their own and instead rely on the information production of investors during the bookbuilding. However, information production by investors is also not cost-free. Empirical evidence suggests that investors get compensated via underpricing for their information production (Hanley, 1993; Benveniste and Spindt, 1989). Hanley and Hoberg (2010) use the extent of non-standard information in the initial IPO prospectus as a proxy for issuers' efforts regarding information production and find that IPOs with more informative content have more accurate initial price estimates.

Since the SEC performs an in-depth review of almost all IPO filings and raises comments that often yield to disclosure changes, the SEC review activities contribute to informative IPO disclosure. Both Lowry et al. (2020) and Li and Liu (2017) find that IPOs with prolonged SEC review activities tend to revise their initial price estimate downwards. While Li and Liu (2017) use the overall number of comment letters and responses between the issuer and the SEC, Lowry et al. (2020) employ the number of letters before the initial price range gets filed. They argue that SEC concerns expressed before the initial price range is determined are known to issuer and underwriters. Hence, they could already be incorporated into the initial price range. That seems not to be the case since IPOs with more SEC review tend to be down-revised.[43] Then, investors either use the updated information in the disclosure or discover similar concerns independently. In contrast to information production via bookbuilding, which is associated with costly underpricing, the SEC information production is likely to be not compensated via underpricing. However, the time increases related to the review are associated with costs (Cassell et al., 2013).

As is apparent from the workload time-series presented in Figure 3.4, workload can quickly change. Moreover, IPOs spent typically several months in registration. Hence, workload measuring at the IPO level is not unambiguous. Based on the fact that about 74% of all comments are already contained in the first SEC comment letter, we examine high workload at the review start, which is the filing date of the first IPO prospectus. Our initial SEC concerns measure is based on the number of comments in the first review round and defined in Section 3.2.3. It is similarly related to revision as the letter-count variables previously used. See Table B.2 for a baseline comparison. To study whether high workload is related to price changes and whether the relation of comments is influenced by high workload, we focus on revision and absolute revision. Entropy-balanced results are presented in Table 3.8.

Table 3.8 shows that both revision and absolute revision are significantly related to SEC concerns: more comments issued by the SEC associates with the production of negative information supported by the negative relation to revision. Besides, information production, in general, is positively related to the extent of comments. We find no evidence that high workload alone is related to the pricing variables. However, for revision and absolute revision, the interaction effect between high workload and comments is significant and almost diametrically to the effect of the comments variable. That is, the initial filing review outcome becomes less related to price changes under high workload. Compared to the regression without the interactions, the comment variable's coefficient doubles approximately from -3.458 (2.550) to -7.099 (4.950) for revision (absolute revision). This emphasizes that the association between the SEC concerns and price revision is stronger in the absence of but almost vanished under high workload.[44]

[43] This is in line with Lowry and Schwert (2004) who find that not all (public) information is priced by underwriters.

[44] We find similar results for down-revision, the absolute value of the negative part of revision, and no effects for up-revision, the positive part of revision, which can be found in Table B.3 of the Appendix.

Table 3.8: IPO Price Revisions and High Workload

	Abs. Revision			Revision		
	(1)	(2)	(3)	(4)	(5)	(6)
SEC Concerns	3.306***	5.519***	5.358***	−3.786*	−8.040***	−8.989***
	(3.180)	(3.850)	(4.030)	(−1.763)	(−2.931)	(−3.408)
High Workload (D)	−0.893	−1.053	−1.222*	0.620	0.928	1.312
	(−1.374)	(−1.548)	(−1.730)	(0.619)	(0.780)	(1.098)
SEC Concerns×		−4.811**	−5.588**		9.248*	10.255**
High Workload (D)		(−2.528)	(−2.570)		(1.938)	(2.264)
ln(Age)	0.233	0.102		−2.654***	−2.403***	
	(0.432)	(0.192)		(−4.456)	(−3.632)	
ln(Sales)	0.157	0.168		0.328	0.307	
	(0.551)	(0.583)		(0.676)	(0.609)	
Leverage	0.316*	0.317*		−0.316	−0.318	
	(1.824)	(1.836)		(−0.777)	(−0.796)	
Pos. EPS (D)	−2.781**	−2.644**		0.431	0.167	
	(−2.594)	(−2.540)		(0.178)	(0.068)	
VC (D)	3.433***	3.268***		2.597	2.914	
	(3.572)	(3.555)		(1.352)	(1.452)	
Bookrunner Market Share	−5.007	−4.972		17.001***	16.934***	
	(−1.546)	(−1.558)		(8.012)	(8.238)	
Lawyer Market Share	−2.208	−1.338		−7.618	−9.290	
	(−0.244)	(−0.148)		(−0.706)	(−0.848)	
Big 4 (D)	−0.324	−0.338		2.321	2.346	
	(−0.290)	(−0.315)		(1.390)	(1.507)	
Prospectus Type (D)	Included	Included	Included	Included	Included	Included
Fixed Effects	FF48	FF48	FF48	FF48	FF48	FF48
	Year	Year	Year	Year	Year	Year
	Month	Month	Month	Month	Month	Month
Observations	922	922	922	922	922	922
Adjusted R^2	0.098	0.100	0.096	0.203	0.208	0.164

Notes: This table presents weighted linear least squares results for regressions on IPO price revisions calculated as the percentage change from the midpoint of the first price range to the offer price. The weights are estimated by entropy balancing using the presented set of control variables and High Workload as the treatment. SEC Concerns are the time-adjusted number of comments raised in the first SEC Letter. High Workload is a dummy variable indicating abnormally high workload of the SEC office responsible for the IPO review process. Age is the age of the IPO firm, calculated with founding dates from Prof. Jay Ritter's website. Sales, Leverage, and Earnings per Share (EPS) are accounting variables from Compustat. VC is a dummy from SDC indicating Venture-Capital backed IPOs. Bookrunner (Lawyer) Market Share is the two-year trailing market share of the lead underwriter (law firm). Big 4 is a dummy variable indicating the auditor is a Big 4 audit firm. Prospectus Type (D) include dummies for the initial IPO prospectus type. See Table B.1 in the Appendix for detailed definitions and sources of the variables. The numbers in brackets below the coefficient estimates show t-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

To examine the relation between SEC comments and price revision in more depth, we regress IPO pricing variables on standard as well as non-standard SEC concerns. These variables are again detrended comment counts. The results are reported in Table 3.9. Table 3.9 reveals the different effects of both kinds of concerns. Regarding revision, non-standard concerns are significantly related to lower revision, while standard concerns are not. For absolute revision, non-standard concerns are positively related, while standard concerns are negatively associated. These results suggest that the average effect of SEC concerns on information production is driven by the non-standard comments.

Potentially, high SEC workload and hence SEC distraction is also associated with distraction of other parties, e.g., investors. Our findings indicate that the relations between the SEC review and price revisions are weaker under high workload. Alternatively, this might be driven by investors, whose information production capabilities are altered when distracted. The results in Table 3.8 and 3.9 provide little evidence in this regard since the high workload dummy is overall unrelated to price revisions.[45]. While institutional investors' resources are not unlimited (Khanna et al., 2008), they are typically thought to be quite large, which makes them less prone to distraction (Barber and Odean, 2008; Ben-Rephael et al., 2017), at least compared to retail investors. As opposed to the SEC who reviews almost all filings with a more or less fixed staff, the large set of institutional investors can act more selectively and react flexibly, making an overall distraction less likely. Moreover, the bookbuilding commonly starts several weeks to months after the filing of the first prospectus.

First-Day Pricing Summarizing, we find relations between workload and outcomes of the filing review, especially evidence for less informative comments, but no direct effect of high workload on revision. Hence, we conjecture that the information production inspired through SEC comment letters is not necessary for price revision but can improve the information environment, especially for the general public. Assuming that the information produced by the SEC is less informative under high workload, the information production role of institutional investors should become more important. Since these information production activities are commonly thought to be compensated via underpricing, we hypothesize that high workload should be associated with more

[45] Note, however, that there are negative coefficients regarding absolute revision in some specifications, indicating less information production.

Table 3.9: IPO Price Revisions and SEC Letter Content

	Abs. Revision			Revision		
	(1)	(2)	(3)	(4)	(5)	(6)
Stand. SEC	−0.502	−0.515		0.277	0.360	
Concerns	(−1.439)	(−1.520)		(0.322)	(0.454)	
Non-Stand. SEC	3.384***	5.808***		−3.889*	−8.063***	
Concerns	(3.351)	(4.462)		(−1.846)	(−3.292)	
High Workload (D)		−0.930			0.813	
		(−1.361)			(0.704)	
Non-Stand. Conc.×		−5.327***			9.222**	
High Workload (D)		(−3.104)			(2.270)	
Prop.(Standard)			−10.699**			12.404*
			(−2.534)			(1.782)
SEC Concerns			2.692**			−3.154
			(2.391)			(−1.511)
ln(Age)	0.344	0.212	0.290	−2.665***	−2.427***	−2.612***
	(0.660)	(0.417)	(0.565)	(−4.418)	(−3.726)	(−4.399)
ln(Sales)	0.018	0.053	0.025	0.397	0.364	0.409
	(0.063)	(0.178)	(0.089)	(0.817)	(0.688)	(0.829)
Leverage	0.296	0.293	0.268	−0.322	−0.318	−0.291
	(1.682)	(1.647)	(1.476)	(−0.800)	(−0.798)	(−0.741)
Pos. EPS (D)	−2.787**	−2.656**	−2.786**	0.369	0.147	0.358
	(−2.589)	(−2.541)	(−2.587)	(0.152)	(0.060)	(0.146)
VC (D)	3.432***	3.222***	3.408***	2.458	2.824	2.474
	(3.480)	(3.452)	(3.422)	(1.277)	(1.398)	(1.292)
Bookrunner Market	−4.881	−4.768	−4.923	17.081***	16.963***	17.108***
Share	(−1.463)	(−1.457)	(−1.467)	(7.878)	(8.172)	(7.767)
Lawyer Market	−1.661	−0.225	−1.734	−7.424	−9.803	−7.457
Share	(−0.187)	(−0.026)	(−0.196)	(−0.669)	(−0.890)	(−0.678)
Big 4 (D)	−0.257	−0.306	−0.253	2.143	2.175	2.111
	(−0.231)	(−0.287)	(−0.227)	(1.245)	(1.365)	(1.226)
Prospectus Type (D)	Included	Included	Included	Included	Included	Included
Fixed Effects	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month
Observations	916	916	916	916	916	916
Adjusted R^2	0.100	0.103	0.100	0.198	0.203	0.199

Notes: This table presents weighted linear least squares results for regressions on IPO price revisions calculated as the percentage change from the midpoint of the first price range to the offer price. The weights are estimated by entropy balancing using the presented set of control variables and High Workload as the treatment. SEC Concerns are the time-adjusted number of comments raised in the first SEC Letter. (Non-)Standard refers to the similarity between the comments of the corresponding SEC letter to the comments issued in antecedent letters. Proportion(Standard Comments) is the relative proportion of comments that are similar to comments issued in antecedent letters. Age is the age of the IPO firm, calculated with founding dates from Prof. Jay Ritter's website. Sales, Leverage, and Earnings per Share (EPS) are accounting variables from Compustat. VC is a dummy from SDC indicating Venture-Capital backed IPOs. Bookrunner (Lawyer) Market Share is the two-year trailing market share of the lead underwriter (law firm). Big 4 is a dummy variable indicating the auditor is a Big 4 audit firm. Prospectus Type (D) include dummies for the initial IPO prospectus type. See Table B.1 in the Appendix for detailed definitions and sources of the variables. The numbers in brackets below the coefficient estimates show *t*-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

underpricing. We test this hypothesis and present results in Table 3.10.

Table 3.10: IPO Underpricing and High Workload

	Dependent variable: First-Day Return			
	(1)	(2)	(3)	(4)
High Workload (D)	2.223**	2.241**	1.956***	2.139**
	(2.155)	(2.160)	(2.881)	(2.070)
Revision			0.581***	
			(14.090)	
SEC Concerns				3.012
				(1.414)
ln(Age)		−0.977	0.412	−0.768
		(−0.573)	(0.237)	(−0.452)
ln(Sales)		−0.073	−0.135	−0.248
		(−0.191)	(−0.338)	(−0.535)
Leverage		−1.339**	−1.136***	−1.365**
		(−2.168)	(−2.729)	(−2.153)
Pos. EPS (D)		0.422	0.206	0.376
		(0.167)	(0.123)	(0.143)
VC (D)		10.192***	8.525***	10.410***
		(3.550)	(3.787)	(3.605)
Bookrunner Market Share		10.409*	0.499	10.462*
		(1.711)	(0.083)	(1.715)
Lawyer Market Share		−20.251	−16.686	−19.075
		(−0.970)	(−1.052)	(−0.933)
Big 4 (D)		−0.397	−1.947	−0.119
		(−0.177)	(−0.887)	(−0.049)
Prospectus Type (D)	Included	Included	Included	Included
Fixed Effects	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month
Observations	922	922	922	922
Adjusted R^2	0.084	0.112	0.272	0.112

Notes: This table presents weighted linear least squares results for regressions on IPO first-day returns calculated as the percentage change from offer to the first closing price. The weights are estimated by entropy balancing using the presented set of control variables and High Workload as the treatment. High Workload is a dummy variable indicating abnormally high workload of the SEC office responsible for the IPO review process. Revision is the percentage change from the midpoint of the first price range to the offer price. SEC Concerns are the time-adjusted number of comments raised in the first SEC Letter. Age is the age of the IPO firm, calculated with founding dates from Prof. Jay Ritter's website. Sales, Leverage, and Earnings per Share (EPS) are accounting variables from Compustat. VC is a dummy from SDC indicating Venture-Capital backed IPOs. Bookrunner (Lawyer) Market Share is the two-year trailing market share of the lead underwriter (law firm). Big 4 is a dummy variable indicating the auditor is a Big 4 audit firm. Prospectus Type (D) include dummies for the initial IPO prospectus type. See Table B.1 in the Appendix for detailed definitions and sources of the variables. The numbers in brackets below the coefficient estimates show *t*-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

In Table 3.10 we find that IPOs have about 2% more underpricing when their initial filing was reviewed under high workload. This value can be interpreted as a cost related to additional information production by investors, which arises since the regulatory information production is less informative than usual. These costs are relativized by the lowered remediation costs due to lower times in review as reported in Section 3.5.

Since the employed workload measure is based on filing activity, which includes IPO filing activity, higher underpricing for high workload IPOs might also be driven by the "hot issue markets"-phenomenon (Ibbotson and Jaffe, 1975). This phenomenon is characterized by both high IPO volume and underpricing. However, the workload measure differs in several respects: it captures not exclusively IPOs, it is applied at the filing date of the first IPO prospectus, which often precedes the issue date by a large and heterogeneous number of days, it is SIC Code specific, and finally also regressed on past values. Indeed, we find that the high workload dummy is barely correlated with many variants of recent IPO activity.[46]

3.7 Conclusion

This study examines the role of high workload for the Division of Corporation Finance of the U.S. Securities and Exchange Commission and its implications for the process of going public. The office-specific workload measure we use in this paper can explain several organizational changes within the SEC and is correlated to self-reported SEC workload data.

Our results suggest that IPOs reviewed by SEC offices and exposed to high workload receive significantly fewer comments in later SEC comment letters. Despite no evidence for fewer comments in the first letter, our results indicate significantly more standard content. Further, the SEC tends to issue comment letters quicker while being busy (after the initial letter), which can be interpreted as a reduction of remediation costs from an issuing firm's perspective.

SEC concerns are associated with IPO price revisions, as empirically shown by Li and Liu (2017) and Lowry et al. (2020). We reinforce and extend this evidence in this paper by employing the extent of initial comments. Under high workload, however, we find this association to diminish, in some specifications even to vanish. We provide some evidence that relates this observation to the reduced contentwise quality of the comment letters. In line with a weaker information environment resulting from an

[46] Employed IPO activity variables are the number of completed IPOs within the n previous days and the average underpricing (and price revision) of the previous n IPOs, where $n \in \{30, 60, 90\}$. Inclusions of these variables in the regression leave the results qualitatively unchanged.

altered review process, we find IPOs under high workload to be associated with about 2% more underpricing. This is consistent with the view that additional information production in the bookbuilding via institutional investors is required and compensated through underpricing.

Our study emphasizes the need for a flexible balancing of workload across those responsible in regulatory authorities. Interestingly, the SEC recently reduced the number of Division of Corporation Finance offices to seven, which should ease workload disparities. Future research can show whether this change will affect the distribution of workload across the offices. For issuing companies, our paper provides several novel insights into the SEC filing review process. For instance, we find that a substantial number of comments are similar. Furthermore, the level of regulator business may be a part of future considerations when going public.

4 Using the Extremal Index for Value-at-Risk Backtesting

The following is based on Bücher et al. (2020):

Bücher, A., P. N. Posch, and P. Schmidtke (2020). Using the Extremal Index for Value-at-Risk Backtesting. *Journal of Financial Econometrics 18(3)*, 556–584. `https://doi.org/10.1093/jjfinec/nbaa011`

5 Volatility forecasting accuracy for Bitcoin

The following is based on Köchling et al. (2020c):

Köchling, G., P. Schmidtke, and P. N. Posch (2020c). Volatility forecasting accuracy for Bitcoin. *Economics Letters 191*, 108836. https://doi.org/10.1016/j.econlet.2019.108836

A Appendix for Chapter 2

A.1 Sample Construction and Variable Descriptions

While creating our sample we follow the Online Appendix of Lowry et al. (2017) where appropriate. We extract a list of IPOs from SDC Platinum between 1st January, 2003, and 30th June, 2017. The definitions and sources for our main variables and controls can be found in Table A.1.

Table A.1: Variable Definitions and Sources

Variable	Source	Description
Attention Variables:		
$U^{\$}, S^{\$}, A^{\$}$	EL, SDC	Sophisticated, unsophisticated, and full attention to an IPO in the week prior to the first trading date on CRSP relative to offer size, see Section 2.3 for details.
$U^{\%}$	EL	The proportion of attention from unsophisticated users relative to all users. See Section 2.3 for details.
U^{Abn}	EL	Residuals of a regression of the number of unsophisticated users paying attention to the IPO on the number of all users paying attention. See Section 2.3 for details.
Dependent Variables:		
First-Day Return	SDC, CRSP	$:= \frac{\text{First End-of-day price available from CRSP}}{\text{Offer Price from SDC}} - 1$ as a percentage
Abnormal post-IPO Return$_{\text{BM}}$ where BM ∈ {' ', 'Size, B/M', 'Market'}	CRSP, FF	$:= \frac{\text{250th (or DL) CRSP price}}{\text{60th CRSP price}} - 1 - r_{BM}^{60,250\text{ (or DL)}}$ where DL is the delisting date and $r_{BM}^{60,250\text{ (or DL)}}$ the return of matched benchmark portfolio (FF48 industry, FF25 size/book-to-market, CRSP value-weighted market) in the same period.
Raw post-IPO Return	CRSP	$:= \frac{\text{250th (or DL) CRSP price}}{\text{60th CRSP price}} - 1$ where DL is the delisting date.
(Continued on next page.)		

Notes: This table presents sources and definitions of the variables used throughout the paper. "EL" refers to the EDGAR log file data set available under https://www.sec.gov/dera/data/edgar-log-file-data-set.html. "CS" is short for the Compustat annual file from which all variables refer to the first value before the SDC Issue Date. "FF" refers to the Fama-French portfolios available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research. IDs refer to the variable identifiers in the corresponding databases.

Table A.1: Variable Definitions and Sources (Continued)

Variable	Source	Description
(Continued.)		
Controls:		
log(Sales + 1)	CS	Proxy for firm size, ID: "revt"
Up revision	SDC	$:= \frac{\text{Offer Price - Mid of Amended Price Range}}{\text{Mid of Amended Price Range}} \times \mathbb{1}(\frac{\text{Offer Price - Mid of Amended Price Range}}{\text{Mid of Amended Price Range}} > 0)$, IDs: "P", "MFILE2"
log(Filing Range)	EL, CRSP	Number of days between first trading day on CRSP and filing date of the initial prospectus (Form S-1)
log(Proceeds)	SDC	Proxy for offer size, ID: "PROCDS"
VC dummy	SDC	1 if issuer is backed by a venture capital firm, else 0, ID: "VE"
Share overhang	SDC, CRSP	$:= \frac{\text{Shares Outstanding from CRSP}}{\text{Shared Offered from SDC}}$, ID: "TOTSHSOVSLD"
Bookrunner Market Share	SDC	Two-year trailing market share of the lead underwriter, ID: "LEADMANAGERSLONG2"
Debt over Assets	CS	$:= \frac{\text{Debt}}{\text{Assets}}$, IDs: "at", "lt"
Positive EPS dummy	CS	$:= \mathbb{1}(EPS > 0)$, ID: "epspi"
Pre-IPO $\bar{r}_{\text{Market}}$	CRSP	Annualized return on the CRSP value-weighted market portfolio in the 30 days prior to first trading day on CRSP.
Pre-IPO σ_{Market}	CRSP	Annualized volatility of the CRSP value-weighted market portfolio in the 30 days prior to first trading day on CRSP.

Notes: This table presents sources and definitions of the variables used throughout the paper. "EL" refers to the EDGAR log file data set available under `https://www.sec.gov/dera/data/edgar-log-file-data-set.html`. "CS" is short for the Compustat annual file from which all variables refer to the first value before the SDC Issue Date. "FF" refers to the Fama-French portfolios available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research`. IDs refer to the variable identifiers in the corresponding databases.

B Appendix for Chapter 3

B.1 Comment Letter Data

We build a database of comment letters from the publicly available EDGAR data, which we use to match SEC comment letters to IPO filings and to calculate various letter-level variables such as the number of comments. We start by downloading all 155,320 unique "UPLOAD"-filings until 13th December 2019.[1] We apply a parsing script in order to extract all relevant data from these filings. With respect to identifying the date of the letter, the reference block, and the body of the letter we are successful for 153,105 filings (rate: 98.6%).[2] Concomitant, we extract 923,193 comments from 110,018 filings.

Where required, we supplement the automatically created data with hand-collected information from the UPLOADs. On the one hand, this is the case for UPLOADs relevant for our IPO sample where automatic parsing yielded no result. In this regard, we add 168 comments from 25 filings manually and further data for 34 filings. On the other hand, we correct information contained in the filings, mostly dates.

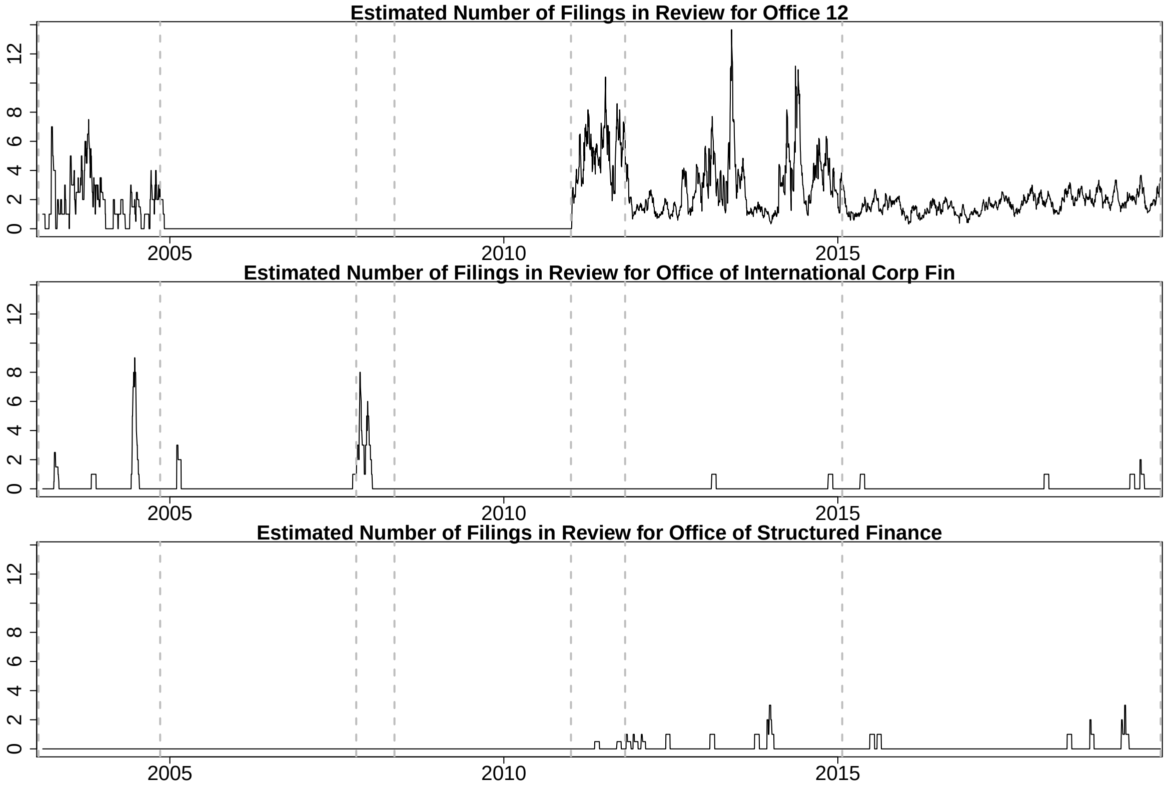
[1] These filings contain also many letters from the *Division of Investment Management*, which performs reviews under the Trust Indenture Act of 1939 and the Investment Company Act of 1940 (Cunningham and Leidner, 2019). Unique refers to the fact that some filings, UPLOADs too, are sometimes uploaded for several CIKs, which produces more than one entry in the EDGAR index file.

[2] The remaining filings typically represent scans or letters from a company to the SEC instead of a SEC response letter, which are not relevant for our purposes.

B.2 Supplementary Figures and Tables

This section contains additional figures and tables referenced throughout Chapter 3.

Figure B.1: Estimated Number of Filings in Review for “Minor” Offices



Notes: This figure shows time-series of workload for the three rather minor offices: Office 12, Office of International Corp Fin/99, and Office of Structured Finance (OSF). The one single phase for Office 12 without any filing is because no SIC was mapped to this office at that time. In the remaining periods, we observe always positive workloads for Office 12. However, these are relatively low compared to the major offices 1 - 11. The latter two minor offices show longer phases without any filing.

Table B.1: Variable Definitions and Sources

Variable	Source	Description
Workload Variables:		
Workload	SEC, EI, W	An empirical probability integral transform from abnormal workload regressions. Values are between zero and one. See Section 3.3.
High Workload	SEC, EI, W	A dummy indicating whether the workload is higher than a threshold. We use 0.8 throughout the paper, see Section 3.3.
Filing Review Variables:		
#Letters	E	The number of SEC letters issued during the review of an IPO. “Before PR” indicates that only the letters prior to the announcement of the first price range are counted. The date of the first price range is determined from EDGAR.
#Comments	E	The number of SEC comments contained in a specific letter. Standard (non-standard) refers to the comment classification performed using the ten most recent IPOs as described in Section 3.2.4.
SEC Concerns	E	Residuals of a regression of a comment count variable on calendar year dummies using a negative binomial count variable model as described in Section 3.2.3. Potential comment count variables are all initial comments, all standard comments, etc.
Response Time	E	The number of days between the filing of an IPO filing and the SEC answer, either measured in calendar or workdays.
Dependent IPO Variables:		
(Absolute) Revision	E, SDC	The (absolute) percentage change from the midpoint of the first filed price range (from EDGAR) and the final offer price (from SDC).
(Continued on next page.)		

Notes: This table presents sources and definitions of the variables used throughout the paper. “SDC” is the Securities Data Company (SDC) Platinum database. “CRSP” is data from The Center for Research in Security Prices. “SEC” refers to data from SEC websites, “E” refers EDGAR filings while “EI” refers to the EDGAR master index. “CS” is short for the Compustat annual file from which all variables refer to the first value before the SDC Issue Date. “R” is data from the website of Prof. J. Ritter. “W” refers to historical website data via `https://archive.org/`. IDs refer to the variable identifiers in the corresponding databases.

Table B.1: Variable Definitions and Sources (Continued)

Variable	Source	Description
(Continued.)		
Up/Down revision	E, SDC	The absolute value of the positive (negative) part of revision.
First-Day Return	SDC, CRSP	$:= \frac{\text{First End-of-day price available from CRSP}}{\text{Offer Price from SDC}} - 1$ as a percentage
Controls:		
log(Age)	R, SDC	Age is the difference between the issue year (SDC) and the founding year (R).
log(Sales)	CS	Sales is a proxy for firm size in million. We use log(Sales+1) since some firms have no revenues. ID: “revt”
Leverage	CS	$:= \frac{\text{Debt}}{\text{Assets}}$, IDs: “at”, “lt”
Positive EPS dummy	CS	$:= \mathbb{1}(EPS > 0)$, ID: “epspi”
VC dummy	SDC	1 if issuer is backed by a venture capital firm, else 0, ID: “VE”
Bookrunner Market Share	SDC	Two-year trailing market share of the lead underwriter, ID: “LEADMANAGERSLONG2”
Lawyer Market Share	SDC	Two-year trailing market share (based on IPO proceeds) of the lawyer, ID: “ILAW”
Big 4	CS	A dummy indicating whether the accounting firm is one of PwC, EY, KPMG, or Deloitte.
Review Size	E	The size of all exhibits contained in a filing in bytes plus the size of the main document if the filing is an initial IPO filing.
Market Return$_{\text{30 Days}}$ (Volatility)	CRSP	Market Return is the trailing annualized 30-day return while market volatility is the trailing annualized 30-day standard deviation based on daily data. The market portfolio is the CRSP value-weighted index.

Notes: This table presents sources and definitions of the variables used throughout the paper. “SDC” is the Securities Data Company (SDC) Platinum database. “CRSP” is data from The Center for Research in Security Prices. “SEC” refers to data from SEC websites, “E” refers EDGAR filings while “EI” refers to the EDGAR master index. “CS” is short for the Compustat annual file from which all variables refer to the first value before the SDC Issue Date. “R” is data from the website of Prof. J. Ritter. “W” refers to historical website data via `https://archive.org/`. IDs refer to the variable identifiers in the corresponding databases.

Table B.2: Outcomes of the SEC Filing Review Process and IPO Price Revisions

	Dependent variable: Revision			
	(1)	(2)	(3)	(4)
#Letters	−1.333			
	(−1.211)			
#Letters$_{\text{Before PR}}$		−2.202**		
		(−2.262)		
SEC Concerns			−3.634**	
			(−2.202)	
SEC Concerns$_{\text{Before PR}}$				−2.385***
				(−3.920)
ln(Age)	−2.591***	−2.423***	−2.783***	−2.616***
	(−4.020)	(−3.633)	(−4.413)	(−4.063)
ln(Sales)	0.005	−0.024	0.165	0.122
	(0.009)	(−0.045)	(0.357)	(0.239)
Leverage	−0.299	−0.338	−0.295	−0.282
	(−0.982)	(−1.250)	(−1.037)	(−1.007)
Pos. EPS (D)	1.093	1.297	1.110	1.270
	(0.501)	(0.562)	(0.529)	(0.565)
VC (D)	3.651**	3.793**	3.523*	3.751*
	(2.050)	(2.118)	(1.903)	(2.026)
Bookrunner Market Share	15.402***	15.841***	15.535***	15.875***
	(7.914)	(8.466)	(7.942)	(8.486)
Lawyer Market Share	−2.576	−3.885	−4.415	−5.540
	(−0.302)	(−0.448)	(−0.560)	(−0.661)
Big 4 (D)	3.216*	3.335**	2.916*	3.013*
	(1.968)	(2.031)	(1.746)	(1.820)
Prospectus Type (D)	Included	Included	Included	Included
Fixed Effects	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month
Observations	922	910	922	910
Adjusted R^2	0.194	0.197	0.197	0.200

Notes: This table presents OLS results for regressions on IPO price revisions calculated as the percentage change from the midpoint of the first price range to the offer price. #Letters is the number of SEC letters the IPO received. SEC Concerns are the time-adjusted number of comments raised in the first SEC Letter. “Before PR” means “before the first price range”. Age is the age of the IPO firm, calculated with founding dates from Prof. Jay Ritter’s website. Sales, Leverage, and Earnings per Share (EPS) are accounting variables from Compustat. VC is a dummy from SDC indicating Venture-Capital backed IPOs. Bookrunner (Lawyer) Market Share is the two-year trailing market share of the lead underwriter (law firm). Big 4 is a dummy variable indicating the auditor is a Big 4 audit firm. Prospectus Type (D) include dummies for the initial IPO prospectus type. The numbers in brackets below the coefficient estimates show t-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

Table B.3: Directional IPO Price Revisions and High Workload

	Neg. Revision			Pos. Revision		
	(1)	(2)	(3)	(4)	(5)	(6)
SEC Concerns	3.546***	6.779***	7.174***	−0.240	−1.260	−1.815*
	(3.025)	(3.634)	(3.812)	(−0.198)	(−1.101)	(−2.002)
High Workload (D)	−0.756	−0.991	−1.267	−0.137	−0.062	0.045
	(−1.010)	(−1.138)	(−1.433)	(−0.350)	(−0.146)	(0.106)
SEC Concerns×		−7.030**	−7.922**		2.218	2.333
High Workload (D)		(−2.378)	(−2.637)		(1.051)	(1.230)
ln(Age)	1.443***	1.252**		−1.211***	−1.150***	
	(3.021)	(2.504)		(−3.953)	(−3.476)	
ln(Sales)	−0.086	−0.070		0.242	0.237	
	(−0.238)	(−0.187)		(1.441)	(1.374)	
Leverage	0.316	0.317		0.000 03	−0.000 4	
	(1.248)	(1.278)		(0.000 2)	(−0.002)	
Pos. EPS (D)	−1.606	−1.406		−1.175	−1.239	
	(−1.186)	(−1.030)		(−0.909)	(−0.950)	
VC (D)	0.418	0.177		3.015***	3.091***	
	(0.390)	(0.164)		(2.802)	(2.735)	
Bookrunner Market Share	−11.004***	−10.953***		5.997***	5.981***	
	(−5.161)	(−5.316)		(3.494)	(3.478)	
Lawyer Market Share	2.705	3.976		−4.913*	−5.314*	
	(0.284)	(0.414)		(−1.699)	(−1.825)	
Big 4 (D)	−1.323	−1.342		0.998	1.004	
	(−1.081)	(−1.184)		(1.383)	(1.417)	
Prospectus Type (D)	Included	Included	Included	Included	Included	Included
Fixed Effects	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month
Observations	924	924	924	924	924	924
Adjusted R^2	0.167	0.173	0.146	0.180	0.180	0.132

Notes: This table presents weighted linear least squares results for regressions on directional IPO price revisions calculated as the positive respectively negative percentage change from the midpoint of the first price range to the offer price, both in absolute terms. The weights are estimated by entropy balancing using the presented set of control variables and High Workload as the treatment. SEC Concerns are the time-adjusted number of comments raised in the first SEC Letter. High Workload is a dummy variable indicating abnormally high workload of the SEC office responsible for the IPO review process. Age is the age of the IPO firm, calculated with founding dates from Prof. Jay Ritter's website. Sales, Leverage, and Earnings per Share (EPS) are accounting variables from Compustat. VC is a dummy from SDC indicating Venture-Capital backed IPOs. Bookrunner (Lawyer) Market Share is the two-year trailing market share of the lead underwriter (law firm). Big 4 is a dummy variable indicating the auditor is a Big 4 audit firm. Prospectus Type (D) include dummies for the initial IPO prospectus type. See Table B.1 in the Appendix for detailed definitions and sources of the variables. The numbers in brackets below the coefficient estimates show *t*-statistics. Standard errors are clustered by 48 Fama-French industries. Asterisks indicate levels of significance as follows: *** (1 %), ** (5 %), * (10 %).

Table B.4: Directional IPO Price Revisions and SEC Letter Standard Content

	Neg. Revision			Pos. Revision		
	(1)	(2)	(3)	(4)	(5)	(6)
Stand. SEC Concerns	−0.390 (−0.839)	−0.438 (−1.040)		−0.112 (−0.242)	−0.077 (−0.175)	
Non-Stand. SEC Concerns	3.637*** (3.134)	6.935*** (4.141)		−0.253 (−0.215)	−1.127 (−1.105)	
High Workload (D)		−0.872 (−0.994)			−0.059 (−0.162)	
Non-Stand. SEC Concerns× High Workload (D)		−7.275*** (−2.874)			1.947 (1.069)	
Prop.(Standard)			−11.551*** (−3.142)			0.853 (0.192)
SEC Concerns			2.923** (2.482)			−0.231 (−0.193)
ln(Age)	1.504*** (3.194)	1.320*** (2.741)	1.451*** (3.163)	−1.161*** (−3.752)	−1.108*** (−3.345)	−1.161*** (−3.718)
ln(Sales)	−0.190 (−0.539)	−0.155 (−0.404)	−0.192 (−0.535)	0.207 (1.135)	0.208 (1.109)	0.217 (1.196)
Leverage	0.309 (1.215)	0.305 (1.213)	0.280 (1.101)	−0.013 (−0.073)	−0.012 (−0.070)	−0.011 (−0.067)
Pos. EPS (D)	−1.578 (−1.163)	−1.402 (−1.040)	−1.572 (−1.146)	−1.209 (−0.932)	−1.255 (−0.961)	−1.214 (−0.931)
VC (D)	0.487 (0.456)	0.199 (0.186)	0.467 (0.434)	2.945** (2.690)	3.023** (2.620)	2.941*** (2.719)
Bookrunner Market Share	−10.981*** (−4.960)	−10.866*** (−5.160)	−11.016*** (−4.863)	6.100*** (3.516)	6.097*** (3.479)	6.093*** (3.562)
Lawyer Market Share	2.882 (0.300)	4.789 (0.507)	2.861 (0.301)	−4.542 (−1.515)	−5.014 (−1.672)	−4.595 (−1.540)
Big 4 (D)	−1.200 (−0.962)	−1.241 (−1.080)	−1.182 (−0.948)	0.943 (1.275)	0.935 (1.301)	0.929 (1.256)
Prospectus Type (D)	Included	Included	Included	Included	Included	Included
Fixed Effects	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month
Observations	916	916	916	916	916	916
Adjusted R^2	0.163	0.169	0.163	0.174	0.173	0.174

Notes: This table presents weighted linear least squares results for regressions on directional IPO price revisions calculated as the positive respectively negative percentage change from the midpoint of the first price range to the offer price, both in absolute terms. The weights are estimated by entropy balancing using the presented set of control variables and High Workload as the treatment. SEC Concerns are the time-adjusted number of comments raised in the first SEC Letter. (Non-)Standard refers to the similarity between the comments of the corresponding SEC letter to the comments issued in antecedent letters. Proportion(Standard Comments) is the relative proportion of comments that are similar to comments issued in antecedent letters. See the caption of Table B.3 for additional details that apply also to this table.

C Appendix for Chapter 4

Bibliography

Acharya, V. V., L. H. Pedersen, T. Philippon, and M. Richardson (2017). Measuring Systemic Risk. *Review of Financial Studies 30*(1), 2–47.

Adrian, T. and M. K. Brunnermeier (2016). CoVaR. *American Economic Review 106*(7), 1705–1741.

Agarwal, S., S. Gupta, and R. D. Israelsen (2017). Public and Private Information: Firm Disclosure, SEC Letters, and the JOBS Act. *SSRN Electronic Journal*.

Alexander, C., E. Lazar, and S. Stanescu (2013). Forecasting VaR using analytic higher moments for GARCH processes. *International Review of Financial Analysis 30*, 36–45.

Andersen, T. G. and T. Bollerslev (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review 39*(4), 885.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*(1), 43–76.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics 169*(1), 75–93.

Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol. 65*(2), 367–389.

Bakker, A. B. and E. Demerouti (2014). Job Demands-Resources Theory. In *Wellbeing*, pp. 1–28. Chichester, UK: John Wiley & Sons, Ltd.

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*(2), 785–818.

Barber, B. M. and T. Odean (2013). The Behavior of Individual Investors. pp. 1533–1570.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 64*(2), 253–280.

Bauguess, S. W., J. Cooney, and K. W. Hanley (2018). Investor Demand for Information in Newly Issued Securities.

BCBS (1996a). Overview of the Amendment to the Capital Accord to incorporate Market Risks. *Basel Committee on Banking Supervision.*

BCBS (1996b). Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements. *Basel Committee on Banking Supervision.*

BCBS (2016). Minimum capital requirements for market risk. *Basel Committee on Banking Supervision.*

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of extremes: Theory and Applications.* Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd.

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

Benninga, S., M. Helmantel, and O. Sarig (2005). The timing of initial public offerings. *Journal of Financial Economics 75*(1), 115–132.

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

Berghaus, B. and A. Bücher (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Ann. Statist. 46*(5), 2307–2335.

Berkowitz, J. (2001). Testing Density Forecasts, With Applications to Risk Management. *Journal of Business & Economic Statistics 19*(4), 465–474.

Berkowitz, J., P. Christoffersen, and D. Pelletier (2011). Evaluating Value-at-Risk Models with Desk-Level Data. *Management Science 57*(12), 2213–2227.

Bernard, D., T. Blackburne, and J. Thornock (2020). Information flows among rivals and corporate investment. *Journal of Financial Economics 136*(3), 760–779.

Bird, S., E. Klein, and E. Loper (2009). *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

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

Bollerslev, T. (1986a). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bollerslev, T. (1986b). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bonner, S. E. and B. L. Lewis (1990). Determinants of Auditor Expertise. *Journal of Accounting Research 28*, 1.

Bontemps, C. and N. Meddahi (2012). Testing distributional assumptions: A GMM aproach. *Journal of Applied Econometrics 27*(6), 978–1012.

Bozanic, Z., J. L. Hoopes, J. R. Thornock, and B. M. Williams (2017). IRS Attention. *Journal of Accounting Research 55*(1), 79–114.

Brav, A., C. Geczy, and P. A. Gompers (2000). Is the abnormal return following equity issuances anomalous? *Journal of Financial Economics 56*(2), 209–249.

Bücher, A., P. N. Posch, and P. Schmidtke (2020). Using the Extremal Index for Value-at-Risk Backtesting. *Journal of Financial Econometrics 18*(3), 556–584.

Busaba, W. Y., L. M. Benveniste, and R.-J. Guo (2001). The option to withdraw IPOs during the premarket: empirical analysis. *Journal of Financial Economics 60*(1), 73–102.

Butler, A. W., M. O. Keefe, and R. Kieschnick (2014). Robust determinants of IPO underpricing and their implications for IPO research. *Journal of Corporate Finance 27*, 367–383.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2009). Measuring the Financial Sophistication of Households. *American Economic Review 99*(2), 393–398.

Candelon, B., G. Colletaz, C. Hurlin, and S. Tokpavi (2011). Backtesting Value-at-Risk: A GMM Duration-Based Test. *Journal of Financial Econometrics 9*(2), 314–343.

Carter, R. and S. Manaster (1990). Initial Public Offerings and Underwriter Reputation. *The Journal of Finance 45*(4), 1045–1067.

Cassell, C. A., L. M. Dreher, and L. A. Myers (2013). Reviewing the SEC's Review Process: 10-K Comment Letters and the Cost of Remediation. *The Accounting Review 88*(6), 1875–1908.

Chaplinsky, S., K. W. Hanley, and S. K. Moon (2017). The JOBS Act and the Costs of Going Public. *Journal of Accounting Research 55*(4), 795–836.

Chen, H., L. Cohen, U. Gurun, D. Lou, and C. Malloy (2020). IQ from IP: Simplifying search in portfolio choice. *Journal of Financial Economics 138*(1), 118–137.

Christoffersen, P. and D. Pelletier (2004). Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics 2*(1), 84–108.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review 39*(4), 841.

Chu, J., S. Chan, S. Nadarajah, and J. Osterrieder (2017). GARCH Modeling of Cryptocurrencies. *Journal of Risk and Financial Management 10*(17).

Clement, M. B., L. Koonce, and T. J. Lopez (2007). The roles of task-specific forecasting experience and innate ability in understanding analyst forecasting performance. *Journal of Accounting and Economics 44*(3), 378–398.

Cook, D. O., R. Kieschnick, and R. A. Van Ness (2006). On the marketing of IPOs. *Journal of Financial Economics 82*(1), 35–61.

Costanzino, N. and M. Curran (2015). Backtesting general spectral risk measures with application to expected shortfall. *The Journal of Risk Model Validation 9*(1), 21–31.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological) 34*(2), 187–202.

Crane, A., K. Crotty, and T. Umar (2018). Do Hedge Funds Profit from Public Information?

Crawford, E. R., J. A. LePine, and B. L. Rich (2010). Linking job demands and resources to employee engagement and burnout: A theoretical extension and meta-analytic test. *Journal of Applied Psychology 95*(5), 834–848.

Cunningham, L. M., B. A. Johnson, E. S. Johnson, and L. L. Lisic (2020). The Switch-Up: An Examination of Changes in Earnings Management after Receiving SEC Comment Letters. *Contemporary Accounting Research 37*(2), 917–944.

Cunningham, L. M. and J. J. Leidner (2019). The SEC Filing Review Process: Insights from Accounting Research. *SSRN Electronic Journal*.

Da, Z., J. Engelberg, and P. Gao (2011). In Search of Attention. *The Journal of Finance 66*(5), 1461–1499.

DeHaan, E., A. Lawrence, and R. Litjens (2019). Measurement Error in Dependent Variables in Accounting: Illustrations Using Google Ticker Search and Simulations. *SSRN Electronic Journal*.

Dellavigna, S. and J. M. Pollet (2009). Investor Inattention and Friday Earnings Announcements. *The Journal of Finance 64*(2), 709–749.

Derrien, F. (2005). IPO Pricing in “Hot” Market Conditions: Who Leaves Money on the Table? *The Journal of Finance 60*(1), 487–521.

Dewey, J. (1938). *Experience and Education*. Kappa Delta Pi.

Diebold, F. X. and R. S. Mariano (1995). Comparing Predicitve accuracy. *Source: Journal of Business & Economic Statistics 13*(3), 253–263.

Dorn, D. (2009). Does Sentiment Drive the Retail Demand for IPOs? *Journal of Financial and Quantitative Analysis 44*(1), 85–108.

Drake, M. S., B. A. Johnson, D. T. Roulstone, and J. R. Thornock (2020). Is There Information Content in Information Acquisition? *The Accounting Review 95*(2), 113–139.

Drake, M. S., D. T. Roulstone, and J. R. Thornock (2012). Investor Information Demand: Evidence from Google Searches Around Earnings Announcements. *Journal of Accounting Research 50*(4), 1001–1040.

Drake, M. S., D. T. Roulstone, and J. R. Thornock (2015). The Determinants and Consequences of Information Acquisition via EDGAR. *Contemporary Accounting Research 32*(3), 1128–1161.

Du, Z. and J. C. Escanciano (2017). Backtesting Expected Shortfall: Accounting for Tail Risk. *Management Science 63*(4), 940–958.

Dufour, J.-M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *J. Econometrics 133*(2), 443–477.

Dunbar, C. (2000). Factors affecting investment bank initial public offering market share. *Journal of Financial Economics 55*(1), 3–41.

Dunbar, C. G. and S. R. Foerster (2008). Second time lucky? Withdrawn IPOs that return to the market. *Journal of Financial Economics 87*(3), 610–635.

Dyer, T., M. Lang, and L. Stice-Lawrence (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics 64*(2-3), 221–245.

Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar - A GARCH volatility analysis. *Finance Research Letters.*

Edwards, A. K. and K. W. Hanley (2010). Short selling in initial public offerings. *Journal of Financial Economics 98*(1), 21–39.

Ege, M., J. L. Glenn, and J. R. Robinson (2020). Unexpected SEC Resource Constraints and Comment Letter Quality. *Contemporary Accounting Research 37*(1), 33–67.

Eling, M. (2014). Fitting asset returns to skewed distributions: Are the skew-normal and skew-student good models? *Insurance Math. Econom. 59*, 45–56.

Elliott, W. B., S. M. Grant, and K. M. Rennekamp (2017). How Disclosure Features of Corporate Social Responsibility Reports Interact with Investor Numeracy to Influence Investor Judgments. *Contemporary Accounting Research 34*(3), 1596–1621.

Ellis, K., R. Michaely, and M. O'Hara (2000). When the Underwriter Is the Market Maker: An Examination of Trading in the IPO Aftermarket. *The Journal of Finance 55*(3), 1039–1074.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling extremal events for insurance and finance*, Volume 33 of *Applications of mathematics*. Berlin u.a.: Springer.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica 50*(4), 987.

Escanciano, J. C. and J. Olmo (2010). Backtesting Parametric Value-at-Risk With Estimation Risk. *Journal of Business & Economic Statistics 28*(1), 36–51.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, pp. 226–231. AAAI Press.

Falato, A., D. Kadyrzhanova, and U. Lel (2014). Distracted directors: Does board busyness hurt shareholder value? *Journal of Financial Economics 113*(3), 404–426.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance 25*(2), 383.

Fama, E. F. (1991). Efficient Capital Markets: II. *The Journal of Finance 46*(5), 1575–1617.

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

Ferro, C. A. T. and J. Segers (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol. 65*(2), 545–556.

Field, L. C. and M. Lowry (2009). Institutional versus Individual Investment in IPOs: The Importance of Firm Fundamentals. *Journal of Financial and Quantitative Analysis 44*(3), 489–516.

Ghalanos, A. (2017). Introduction to the rugarch package (version 1.3-1). Available at `http://cran.r-project.org/web/packages/rugarch`.

Gibbons, B., P. Iliev, and J. Kalodimos (2020). Analyst Information Acquisition via EDGAR. *Management Science.*

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance 48*(5), 1779.

Gompers, P. A. and A. Metrick (2001). Institutional Investors and Equity Prices. *The Quarterly Journal of Economics 116*(1), 229–259.

Grambsch, P. M. and T. M. Therneau (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika 81*(3), 515–526.

Gunny, K. A. and J. M. Hermis (2020). How Busyness Influences SEC Compliance Activities: Evidence from the Filing Review Process and Comment Letters. *Contemporary Accounting Research 37*(1), 7–32.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis 20*(1), 25–46.

Hanley, K. W. (1993). The underpricing of initial public offerings and the partial adjustment phenomenon. *Journal of Financial Economics 34*(2), 231–250.

Hanley, K. W. and G. Hoberg (2010). The Information Content of IPO Prospectuses. *Review of Financial Studies 23*(7), 2821–2864.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics 23*(4), 365–380.

Hansen, P. R. and A. Lunde (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics 20*(7), 873–889.

Hansen, P. R., Lunde Asgar, and Nason James M. (2011). The Model Confidence Set. *Econometrica 79*(2), 453–497.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the Cross-Section of Expected Returns. *Review of Financial Studies 29*(1), 5–68.

Hazarika, S., J. M. Karpoff, and R. Nahata (2012). Internal corporate governance, CEO turnover, and earnings management. *Journal of Financial Economics 104*(1), 44–69.

Hilary, G. and R. Shen (2013). The Role of Analysts in Intra-Industry Information Transfer. *The Accounting Review 88*(4), 1265–1287.

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009). Driven to Distraction: Extraneous Events and Underreaction to Earnings News. *The Journal of Finance 64*(5), 2289–2325.

Ibbotson, R. G. (1975). Price performance of common stock new issues. *Journal of Financial Economics 2*(3), 235–272.

Ibbotson, R. G. and J. F. Jaffe (1975). “Hot Issue” Markets. *The Journal of Finance 30*(4), 1027–1042.

Johnson, B. A., L. L. Lisic, J. S. Moon, and M. Wang (2019). SEC Comment Letters on Form S-4 and M&A Accounting Quality. *SSRN Electronic Journal*.

Johnston, R. and R. Petacchi (2017). Regulatory Oversight of Financial Reporting: Securities and Exchange Commission Comment Letters. *Contemporary Accounting Research 34*(2), 1128–1155.

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters 158*, 3–6.

Kerkhof, J. and B. Melenberg (2004). Backtesting for risk-based regulatory capital. *Journal of Banking & Finance 28*(8), 1845–1865.

Khanna, N., T. H. Noe, and R. Sonti (2008). Good IPOs draw in bad: Inelastic banking capacity and hot markets. *Review of Financial Studies 21*(5), 1873–1906.

Köchling, G., P. Schmidtke, and P. N. Posch (2020a). Information Acquisition Experience, Investor Sophistication, and IPO Price Pressure.

Köchling, G., P. Schmidtke, and P. N. Posch (2020b). SEC Workload, IPO Filing Reviews, and IPO Pricing.

Köchling, G., P. Schmidtke, and P. N. Posch (2020c). Volatility forecasting accuracy for Bitcoin. *Economics Letters 191*, 108836.

Kole, E., T. Markwat, A. Opschoor, and D. van Dijk (2017). Forecasting Value-at-Risk under Temporal and Portfolio Aggregation. *Journal of Financial Econometrics 15*(4), 649–677.

Kratz, M., Y. H. Lok, and A. J. McNeil (2018). Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. *Journal of Banking & Finance 88*, 393–407.

Krigman, L. and W. Jeffus (2016). IPO pricing as a function of your investment banks' past mistakes: The case of Facebook. *Journal of Corporate Finance 38*, 335–344.

Krische, S. D. (2019). Investment Experience, Financial Literacy, and Investment-Related Judgments. *Contemporary Accounting Research 36*(3), 1634–1668.

Kubick, T. R., D. P. Lynch, M. A. Mayberry, and T. C. Omer (2016). The Effects of Regulatory Scrutiny on Tax Avoidance: An Examination of SEC Comment Letters. *The Accounting Review 91*(6), 1751–1780.

Kupiec, P. H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives 3*(2), 73–84.

Laurent, S., J. V. Rombouts, and F. Violante (2012). On the forecasting accuracy of multivariate GARCH models. *Journal of Applied Econometrics 27*(6), 934–955.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete 65*(2), 291–306.

Lee, C. M., P. Ma, and C. C. Wang (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics 116*(2), 410–431.

Lee, P. M. and S. Wahal (2004). Grandstanding, certification and the underpricing of venture capital backed IPOs. *Journal of Financial Economics 73*(2), 375–407.

Li, B. and Z. Liu (2017). The oversight role of regulators: evidence from SEC comment letters in the IPO process. *Review of Accounting Studies 22*(3), 1229–1260.

Liu, L. X., A. E. Sherman, and Y. Zhang (2014). The Long-Run Role of the Media: Evidence from Initial Public Offerings. *Management Science 60*(8), 1945–1964.

Liu, X. K. and B. Wu (2020). Do IPO Firms Misclassify Expenses? Implications for IPO Price Formation and Post-IPO Stock Performance. *Management Science, Forthcoming.*

Ljungqvist, A., V. Nanda, and R. Singh (2006). Hot Markets, Investor Sentiment, and IPO Pricing. *The Journal of Business 79*(4), 1667–1702.

Longin, F. M. (2000). From value at risk to stress testing: The extreme value approach. *Journal of Banking & Finance 24*(7), 1097–1130.

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

Loughran, T. and J. R. Ritter (1995). The New Issues Puzzle. *The Journal of Finance 50*(1), 23–51.

Lowry, M., R. Michaely, and E. Volkova (2016). Information Revelation Through Regulatory Process: Interactions between the SEC and Companies Ahead of the IPO. *SSRN Electronic Journal*.

Lowry, M., R. Michaely, and E. Volkova (2017). Initial Public Offerings: A Synthesis of the Literature and Directions for Future Research. *Foundations and Trends® in Finance 11*(3-4), 154–320.

Lowry, M., R. Michaely, and E. Volkova (2020). Information Revealed through the Regulatory Process: Interactions between the SEC and Companies ahead of Their IPO. *The Review of Financial Studies*.

Lowry, M. and G. Schwert (2004). Is the IPO pricing process efficient? *Journal of Financial Economics 71*(1), 3–26.

McNeil, A. J. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance 7*(3-4), 271–300.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative risk management*. Princeton Series in Finance. Princeton, NJ: Princeton University Press.

Mikhail, M. B., B. R. Walther, and R. H. Willis (1997). Do Security Analysts Improve Their Performance with Experience? *Journal of Accounting Research 35*, 131.

Mikosch, T. and C. Starica (2000). Limit Theory for the Sample Autocorrelations and Extremes of a GARCH (1, 1) Process. *Ann. Statist. 28*(5), 1427–1451.

Miller, B. P. (2010). The Effects of Reporting Complexity on Small and Large Investor Trading. *The Accounting Review 85*(6), 2107–2143.

Northrop, P. J. (2015). An efficient semiparametric maxima estimator of the extremal index. *Extremes 18*(4), 585–603.

Pástor, L. and P. Veronesi (2005). Rational IPO Waves. *The Journal of Finance 60*(4), 1713–1757.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. In *Journal of Econometrics*, Volume 160, pp. 246–256.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research 12*, 2825–2830.

Pelletier, D. and W. Wei (2016). The Geometric-VaR Backtesting Method. *Journal of Financial Econometrics 14*(4), 725–745.

Pérignon, C., Z. Y. Deng, and Z. J. Wang (2008). Do banks overstate their Value-at-Risk? *Journal of Banking & Finance 32*(5), 783–794.

Pérignon, C. and D. R. Smith (2010). The level and quality of Value-at-Risk disclosure by commercial banks. *Journal of Banking & Finance 34*(2), 362–377.

Purnanandam, A. K. and B. Swaminathan (2004). Are IPOs Really Underpriced? *Review of Financial Studies 17*(3), 811–848.

Rocco, M. (2014). Extreme Value Theory in Finance: A Survey. *Journal of Economic Surveys 28*(1), 82–108.

Rock, K. (1986). Why new issues are underpriced. *Journal of Financial Economics 15*(1), 187–212.

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Ann. Math. Statist. 23*(3), 470–472.

Ryans, J. (2017). Using the EDGAR Log File Data Set.

Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business 74*(1), 101–124.

Smith, R. L. and I. Weissman (1994). Estimating the Extremal Index. *Journal of the Royal Statistical Society. Series B (Methodological) 56*(3), 515–528.

Stephens-Davidowitz, S. and H. Varian (2015). A Hands-on Guide to Google Data. *Google, Inc.*, 1–25.

Süveges, M. (2007). Likelihood estimation of the extremal index. *Extremes 10*(1-2), 41–55.

Süveges, M. and A. C. Davison (2010). Model misspecification in peaks over threshold analysis. *The Annals of Applied Statistics 4*(1), 203–221.

Tadić, M., A. B. Bakker, and W. G. M. Oerlemans (2015). Challenge versus hindrance job demands and well-being: A diary study on the moderating role of job resources. *Journal of Occupational and Organizational Psychology 88*(4), 702–725.

Tan, H.-T., E. Ying Wang, and B. Zhou (2014). When the Use of Positive Language Backfires: The Joint Effect of Tone, Readability, and Investor Sophistication on Earnings Judgments. *Journal of Accounting Research 52*(1), 273–302.

Tiniç, S. M. (1988). Anatomy of Initial Public Offerings of Common Stock. *The Journal of Finance 43*(4), 789–822.

Troster, V., A. K. Tiwari, M. Shahbaz, and D. N. Macedo (2018). Bitcoin returns and risk: A general GARCH and GAS analysis. *Finance Research Letters*.

U.S. Securities and Exchange Commission (2019a). Agency Financial Report (FY 2019). Technical report.

U.S. Securities and Exchange Commission (2019b). Congressional Budget Justification Annual Performance Plan (FY 2020) and Annual Performance Report (FY 2018). Technical report.

U.S. Securities and Exchange Commission (2020). Congressional Budget Justification Annual Performance Plan (FY 2021) and Annual Performance Report (FY 2019). Technical report.

West, K. D. (1996). Asymptotic Inference about Predictive Ability. *Econometrica 64*(5), 1067 – 1084.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica 68*(5), 1097–1126.

Wong, W. K. (2008). Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking & Finance 32*(7), 1404–1415.

Ziggel, D., T. Berens, G. N. Weiß, and D. Wied (2014). A new set of improved Value-at-Risk backtests. *Journal of Banking & Finance 48*, 29–41.