SEC Workload, IPO Filing Reviews, and IPO Pricing

Gerrit Köchling, Philipp Schmidtke, Peter N. Posch

January 27, 2021

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.

Keywords: Initial Public Offering; IPO; SEC; Workload; Stress; Comment Letter; Price Revision; Review Process; Information Production

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 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

[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.

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 midpoint of the first price range to the offer price. In addition to the overall

[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.

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

[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.

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 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 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 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 4 and 5, we focus on the relationship between the quality of comment letters issued by the SEC, respectively their response times, and high workload. Section 6 studies the relation between the filing review, IPO pricing, and high workload. Section 7 concludes.

2 IPO Sample, IPO Filings, and Comment Letters

In this section, we describe our IPO sample selection process (Subsection 2.1), how we match IPO filings and SEC comment letters (Subsection 2.2), and give

[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.*

overview figures on the SEC filing review (Subsection 2.3).

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 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 1, Panel A, shows the descriptive statistics of the final sample. Variable definitions can be found in Table A1 of the appendix.

[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.

[9]Details regarding the matching can be found in Subsection 2.2. However, those who are not interested in the details may want to skip to the overview in Subsection 2.3.

Table 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
$\#\text{Letters}_{\text{Before PR}}$	3.02	1.32	2.00	3.00	5.00
$\#\text{Comments}_{\text{First Letter}}$	39.17	19.94	16.00	36.00	66.00
$\#\text{Comments}_{\text{Before PR}}$	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
$\text{Market Return}_{\text{30 Days}}$	0.18	0.35	−0.23	0.16	0.60
$\text{Market Vola}_{\text{30 Days}}$	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
$\text{Market Return}_{\text{30 Days}}$	0.18	0.17	0.19	0.21
$\text{Market Vola}_{\text{30 Days}}$	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 A1 in the Appendix for detailed definitions and sources of the variables.

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 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. 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 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.

[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.

[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 1 of this paper and Table 1 of Lowry et al. (2020).

- 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 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

[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.

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.

2.3 Summary Statistics of the IPO Filing Review

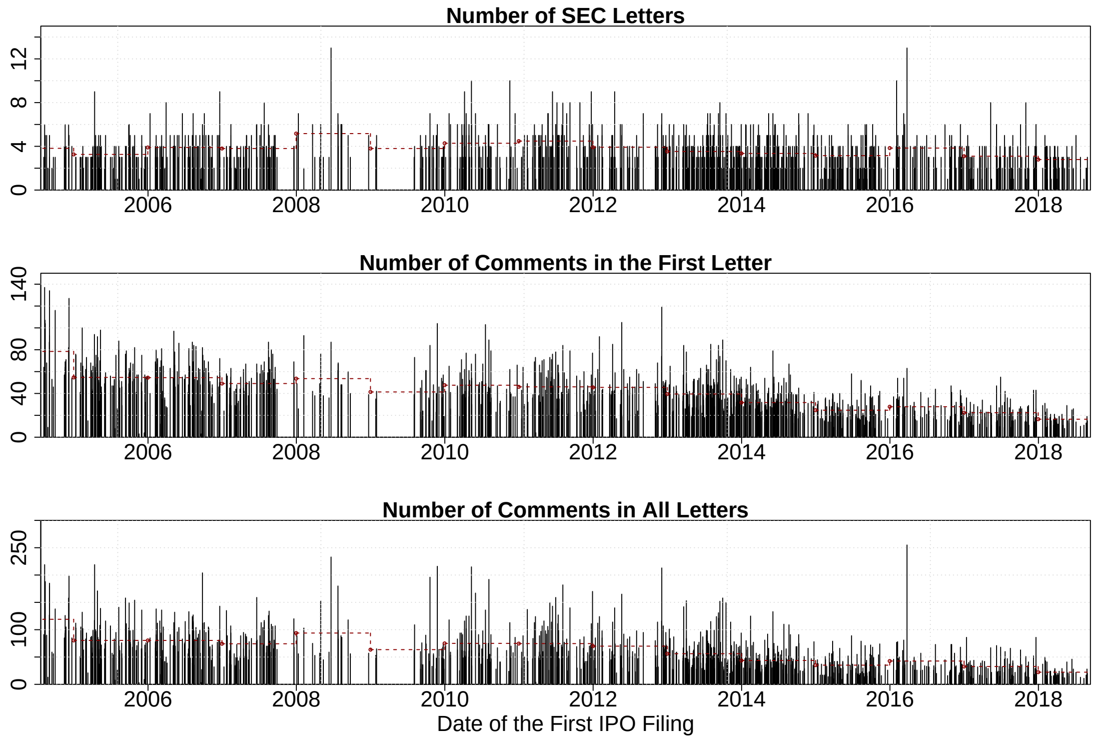
Before omitting IPOs due to missing variables, our IPO sample from 3rd August 2004 to 30th August 2018 includes 1,339 IPOs.[14] For 1,206 IPOs we attempt to match IPO filings and letters and in 1,086 cases we obtain a complete match (592 matched by order, 447 matched by date, 47 matched by Amendment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

Table 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 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 6.

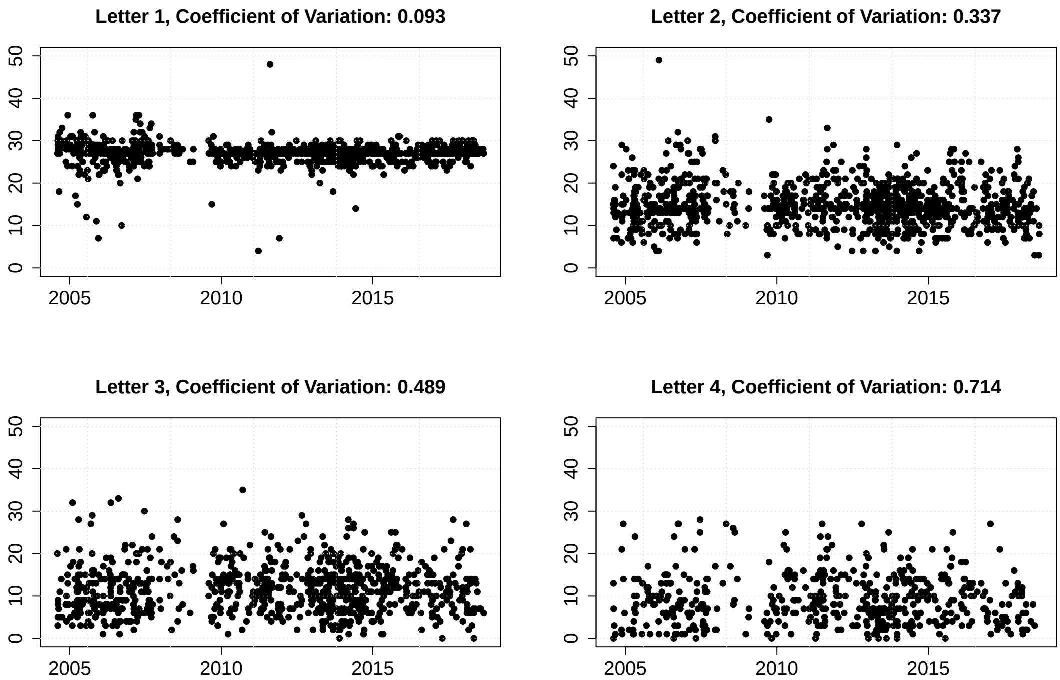
[14] Filing date of the first IPO receiving a comment letter and filing date of the last IPO in our sample.

Figure 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 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 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]

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]

[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: 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

Hence, we do not use this approach and prefer direct comment comparisons, which are also more illustrative. However, we follow most 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

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.

[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.

(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.
Figure 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.

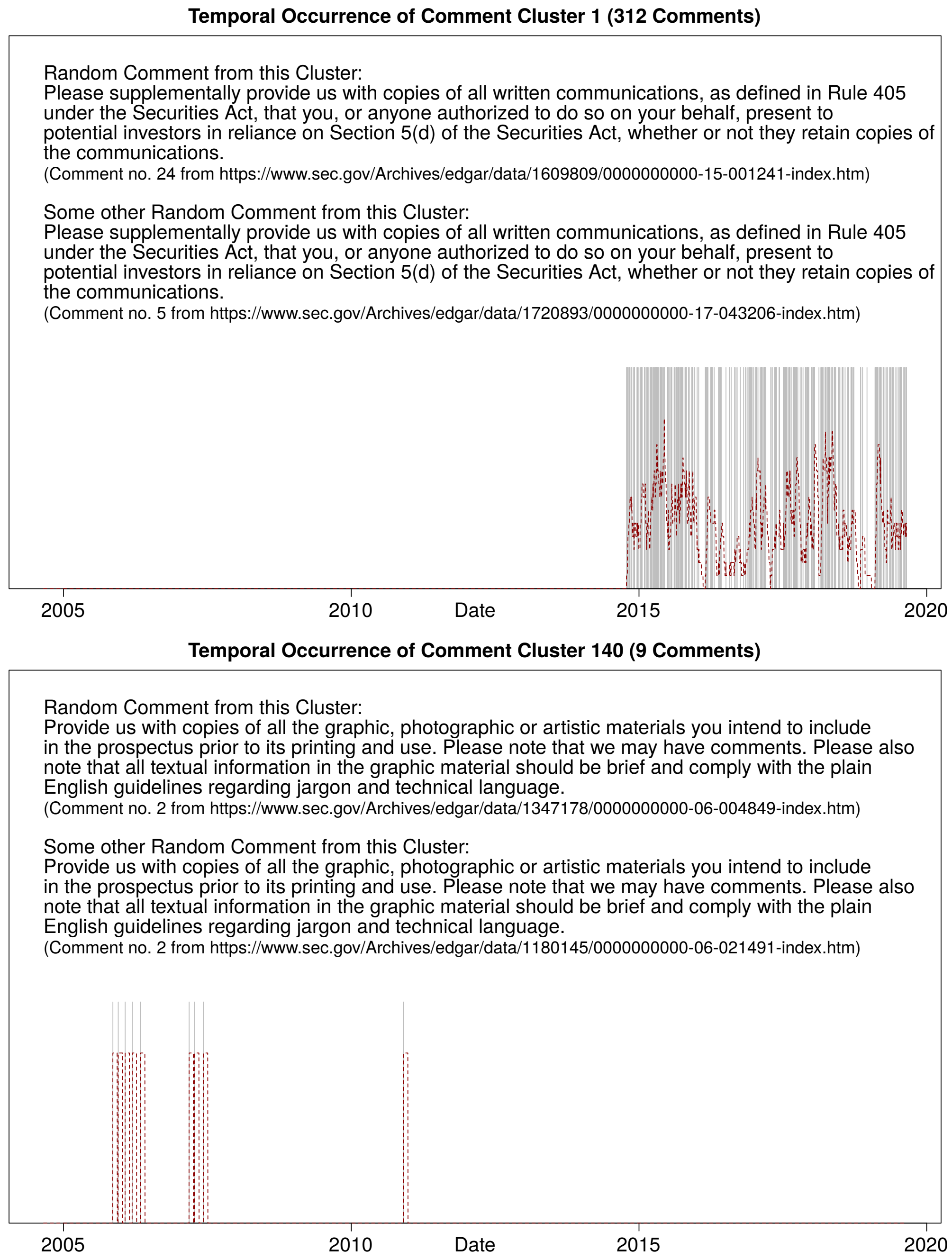
Recent Standard and Non-Standard Comments. We use the presented evidence on the existence of similar comments and define a *number of recent standard (and non-standard) comments* for each IPO. For all initial comments of a given IPO, we determine the closest comment from a set of recently issued comments 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, 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

[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%.

Figure 3: Clusters of Initial SEC Comments with DBSCAN



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.

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 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 Measuring Filing Review Workload for SEC Offices

In this section, we describe how we construct our workload measure. Details can be found in Subsection 3.1 and initial tests for the measure in Subsection 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.

3.1 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 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

[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 A1 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`.

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 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

[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 3.2
[29]This leads to 250 till 252 workdays per year with a median of 251 days.

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%).

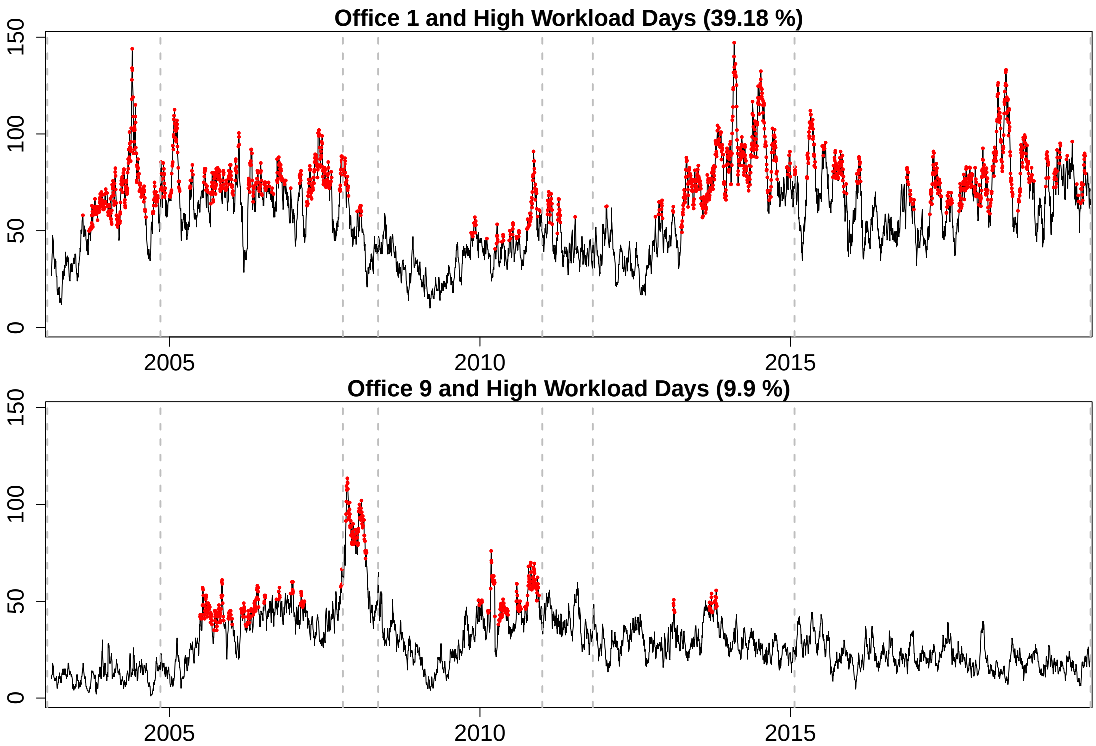
The office assignments obtained by a combination of these two data sets are not always unequivocal. First, in some periods, there is no clear mapping between some SIC Codes and offices. For instance, SIC Code 7389 in 2011 is assigned to Office 2 and 3. Second, the EDGAR index contains multiple records 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

[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).

lasts a constant number of workdays, depending on the form type. Supported by the declining response times for later letters presented in Table 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 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 4 presents $w_{i,t}$ time-series for Offices 1 and 9.

Figure 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.

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:

$$w_{i,t} = \beta_0 + \sum_{k=1}^{K} \beta_k \bar{w}^c_{i,t,s_k,a_k} + \epsilon_{i,t}, \tag{3.1}$$

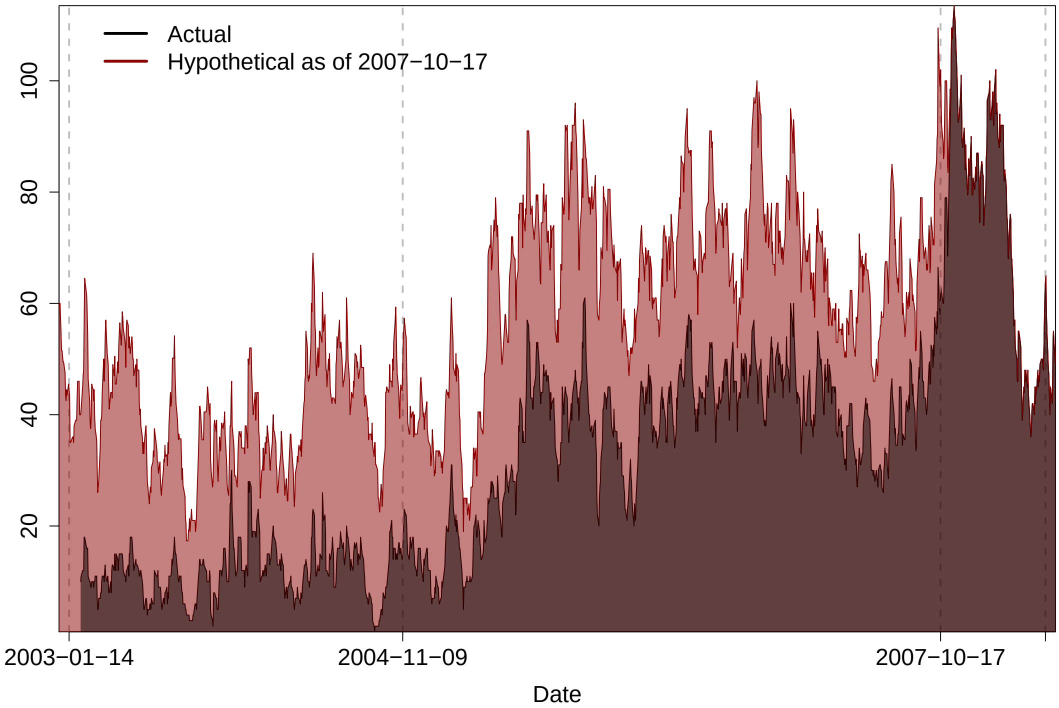
for a specific period $t \in T$ and offices $i \in I$ where $\bar{w}^c_{i,t,s,a} := \sum_{j=t-s+1-a}^{t-s} w^c_{i,t}/a$. In this definition, c can represent actual workload, $w^{act}_{i,t} = w_{i,t}$ or hypothetical workload, which we motivate in the following.

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 5.

Figure 5 presents an extract of the actual workload for Office 9 (black), already contained in Figure 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

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

Figure 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 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.

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.

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

[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.

amendment in terms of review effort. Fourth, the matching between filings and CF offices is not always unambiguous.

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 high workload in both office variants. Workload is measured on the date of the first IPO filing. Logit regression results are presented in Table 2.

Table 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 A1 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 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.

Table 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 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 4.

Table 4: Self-Reported SEC Workload and Estimated Workload

	New Issuer Reviews	Reporting Company Reviews	Total Reviews
	(1)	(2)	(3)
Workload	0.898***	0.050	0.154
	(3.960)	(0.377)	(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".

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

[34]Older data seems not to be available.

words, which we coin *quantitative quality*.[35]

In the textual comment analysis in Section 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.

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 2. Workload is always measured on the filing date of the corresponding IPO filing.

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

Results of entropy-balanced negative binomial regressions can be found in Table 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 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 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.

[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 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 – Year, Month	SEC Office Letter Year, Month	Issuer Letter 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 A1 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 %).

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 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 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 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.

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 2.4. Results of entropy-balanced negative binomial regressions are presented in Table 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 6) largely resemble the previous results for all comments as presented in specification (1) of Table 5 but are often slightly 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 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.

Table 6: Comment Letter Content and High Workload

	#Stand. Com.	#Non-Stand. Com.	Prop.(Stand. Com.)
	(1)	(2)	(3)
High Workload (D)	0.144***	0.008	0.134***
	(2.989)	(0.319)	(4.153)
ln(Review Size)	−0.016	−0.008	0.056
	(−0.300)	(−0.323)	(1.168)
Market Return$_{30\ \mathrm{Days}}$	0.171	−0.029	0.153
	(1.444)	(−0.519)	(1.541)
Market Vola$_{30\ \mathrm{Days}}$	0.991	−0.523**	1.228
	(0.934)	(−2.004)	(1.591)
ln(Age)	0.005	−0.080***	−0.002
	(0.073)	(−5.734)	(−0.038)
ln(Sales)	−0.043*	0.077***	−0.101***
	(−1.736)	(10.823)	(−5.311)
Leverage	0.012	0.009*	−0.023*
	(0.822)	(1.687)	(−1.956)
Pos. EPS (D)	0.088	−0.002	0.078***
	(1.357)	(−0.052)	(4.290)
VC (D)	0.009	−0.110***	0.108*
	(0.158)	(−2.928)	(1.740)
Bookrunner Market Share	−0.041	−0.021	−0.117
	(−0.342)	(−0.365)	(−1.068)
Lawyer Market Share	0.010	−0.372*	−0.007
	(0.021)	(−1.908)	(−0.021)
Big 4 (D)	−0.004	−0.128***	0.088
	(−0.055)	(−5.858)	(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 A1 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 %).

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.

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 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

[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.

[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.

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

$$h(t) = h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p) \tag{5.1}$$

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 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 2) are not explainable by our set of variables. Once we exclude the first letter, the model is well specified.[41]

Table 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

[40] We also formulate regression models analog to Table 5. Untabulated results are qualitatively similar and available upon request from the authors.

[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 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 A1 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 %).

effect.

Comparing the results with those documented for the number of comments raised by the SEC in Section 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 5) but are also associated with significantly lower remediation costs in terms of response times by the SEC (-29%).

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 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 2.3. It is similarly related to revision as the letter-count variables previously used. See Table A2 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

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

revision and absolute revision. Entropy-balanced results are presented in Table 8.

Table 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× High Workload (D)		−4.811**	−5.588**		9.248*	10.255**
		(−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 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year Month	FF48 Year 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 A1 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 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]

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 9.
Table 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 8 and 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

[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 A3 of the Appendix.

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

Table 9: IPO Price Revisions and SEC Letter Content

	Abs. Revision			Revision		
	(1)	(2)	(3)	(4)	(5)	(6)
Stand. SEC Concerns	−0.502	−0.515		0.277	0.360	
	(−1.439)	(−1.520)		(0.322)	(0.454)	
Non-Stand. SEC Concerns	3.384***	5.808***		−3.889*	−8.063***	
	(3.351)	(4.462)		(−1.846)	(−3.292)	
High Workload (D)		−0.930			0.813	
		(−1.361)			(0.704)	
Non-Stand. Conc.× High Workload (D)		−5.327***			9.222**	
		(−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 Share	−4.881	−4.768	−4.923	17.081***	16.963***	17.108***
	(−1.463)	(−1.457)	(−1.467)	(7.878)	(8.172)	(7.767)
Lawyer Market Share	−1.661	−0.225	−1.734	−7.424	−9.803	−7.457
	(−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 A1 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 %).

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 underpricing. We test this hypothesis and present results in Table 10.

In Table 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 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

Table 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 A1 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 %).

dummy is barely correlated with many variants of recent IPO activity.[46]

[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.

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 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.

A Appendix

B 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.[47] 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%).[48] 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.

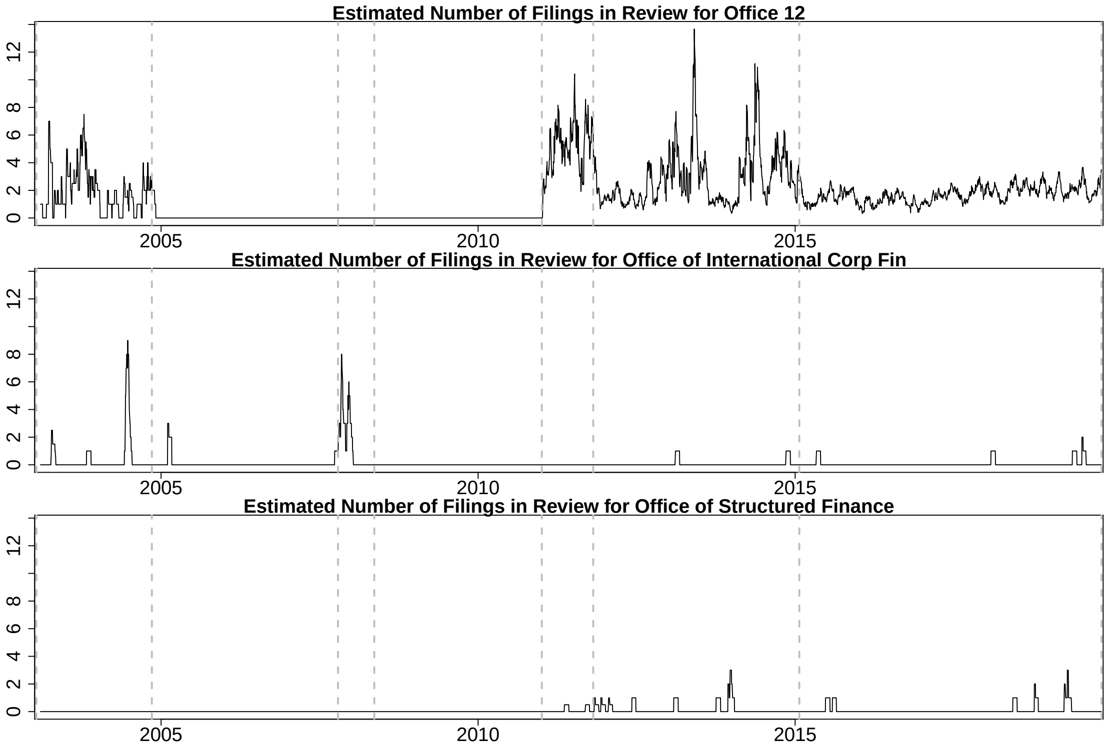
[47] These filings contain also many letters from the *Division of Investment Management*, which performs reviews under the Trust Indenture Act of 1939 and the Investment Company Act of 1940 (Cunningham and Leidner, 2019). Unique refers to the fact that some filings, UPLOADs too, are sometimes uploaded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

[48] 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.

C Supplementary Figures and Tables

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

Figure A1: 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 A1: 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.
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.
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 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 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 A1: 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 (based on IPO proceeds) of the (first) lead underwriter, ID: "LEADMANAGERS"
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$_{30\text{ Days}}$ (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 A2: 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*** (−4.020)	−2.423*** (−3.633)	−2.783*** (−4.413)	−2.616*** (−4.063)
ln(Sales)	0.005 (0.009)	−0.024 (−0.045)	0.165 (0.357)	0.122 (0.239)
Leverage	−0.299 (−0.982)	−0.338 (−1.250)	−0.295 (−1.037)	−0.282 (−1.007)
Pos. EPS (D)	1.093 (0.501)	1.297 (0.562)	1.110 (0.529)	1.270 (0.565)
VC (D)	3.651** (2.050)	3.793** (2.118)	3.523* (1.903)	3.751* (2.026)
Bookrunner Market Share	15.402*** (7.914)	15.841*** (8.466)	15.535*** (7.942)	15.875*** (8.486)
Lawyer Market Share	−2.576 (−0.302)	−3.885 (−0.448)	−4.415 (−0.560)	−5.540 (−0.661)
Big 4 (D)	3.216* (1.968)	3.335** (2.031)	2.916* (1.746)	3.013* (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 A3: 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× High Workload (D)		−7.030**	−7.922**		2.218	2.333
		(−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 A1 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 A4: Directional IPO Price Revisions and SEC Letter Standard Content

	Neg. Revision			Pos. Revision		
	(1)	(2)	(3)	(4)	(5)	(6)
Stand. SEC Concerns	−0.390	−0.438		−0.112	−0.077	
	(−0.839)	(−1.040)		(−0.242)	(−0.175)	
Non-Stand. SEC Concerns	3.637***	6.935***		−0.253	−1.127	
	(3.134)	(4.141)		(−0.215)	(−1.105)	
High Workload (D)		−0.872			−0.059	
		(−0.994)			(−0.162)	
Non-Stand. SEC Concerns× High Workload (D)		−7.275***			1.947	
		(−2.874)			(1.069)	
Prop.(Standard)			−11.551***			0.853
			(−3.142)			(0.192)
SEC Concerns			2.923**			−0.231
			(2.482)			(−0.193)
ln(Age)	1.504***	1.320***	1.451***	−1.161***	−1.108***	−1.161***
	(3.194)	(2.741)	(3.163)	(−3.752)	(−3.345)	(−3.718)
ln(Sales)	−0.190	−0.155	−0.192	0.207	0.208	0.217
	(−0.539)	(−0.404)	(−0.535)	(1.135)	(1.109)	(1.196)
Leverage	0.309	0.305	0.280	−0.013	−0.012	−0.011
	(1.215)	(1.213)	(1.101)	(−0.073)	(−0.070)	(−0.067)
Pos. EPS (D)	−1.578	−1.402	−1.572	−1.209	−1.255	−1.214
	(−1.163)	(−1.040)	(−1.146)	(−0.932)	(−0.961)	(−0.931)
VC (D)	0.487	0.199	0.467	2.945**	3.023**	2.941***
	(0.456)	(0.186)	(0.434)	(2.690)	(2.620)	(2.719)
Bookrunner Market Share	−10.981***	−10.866***	−11.016***	6.100***	6.097***	6.093***
	(−4.960)	(−5.160)	(−4.863)	(3.516)	(3.479)	(3.562)
Lawyer Market Share	2.882	4.789	2.861	−4.542	−5.014	−4.595
	(0.300)	(0.507)	(0.301)	(−1.515)	(−1.672)	(−1.540)
Big 4 (D)	−1.200	−1.241	−1.182	0.943	0.935	0.929
	(−0.962)	(−1.080)	(−0.948)	(1.275)	(1.301)	(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 A3 for additional details that apply also to this table.

References

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

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.

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.

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

Bozanic, Z., J. R. Dietrich, and B. A. Johnson (2017). SEC comment letters and firm disclosure. Journal of Accounting and Public Policy 36(5), 337–357.

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.

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.

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

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.

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

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

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.

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.

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.

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.

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. 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.

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.

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.

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.

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

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.

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

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.

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

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.

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.