Fact, Opinion, and the Function of Securities Law*

Adam B. Badawi, UC Berkeley Law

Andrew C. Baker, UC Berkeley Law

April 10, 2026

Abstract

This study examines the relationship between the threat of liability under our securities statutes and the phrasing used by firms in their disclosures. We fine-tune a BERT model to classify sentences in SEC filings as fact or opinion, and apply it to (i) 17,276 public company annual reports from 2007–2021 and (ii) 2,229 IPO prospectuses from 2011–2024. We show that firms that increase the amount of factual language in the Risk Factors section of their annual reports encounter a trade-off: they experience an increase in abnormal stock returns in the weeks following the change in content, but are also more likely to be sued for making more certain statements in the ensuing years. In the IPO sample, we find that the Supreme Court's expansion of liability for opinion-based statements was associated with an increase in the use of opinions and an increase in underpricing. We develop evidence that this effect is largely attributable to the way that the market prices such disclosures, rather than the composition of the disclosures themselves. Our results show that the threat of liability under the securities laws can improve price accuracy through its effect on the credibility of firm disclosures.

1 Introduction

American securities law creates a disclosure framework backed by liability for material misstatements and omissions. Disclosure allows issuers to inform investors about the value of assets and the threat of liability gives investors confidence in the content of those disclosures. This system requires drawing a line between statements that are verifiable enough to support liability and those that remain too speculative to warrant it. Although the precise location

*We thank Sabrina Arias, Elliot Ash, and participants at the 2025 Law and Data Science Conference for helpful comments.

of that line has shifted over time, the distinction between fact and opinion can determine whether an issuer is liable under the securities laws for the content of its disclosures.

The relationship between disclosures and asset prices is the subject of a long line of research in law and finance. Recent advances in digitization and computational power have allowed researchers to study that relationship at scale, albeit using relatively coarse measures. That research finds that period-to-period changes in word counts can have a significant effect on stock prices (Cohen et al., 2020), language with a specific tenor has a more positive effect on stock price than more general language (Hope et al., 2016), and adding more risk factors to disclosures produces stock price declines. There is little research, however, on the consequence of the phrasing in disclosures at the sentence level. This phrasing matters because, under the securities laws, a plaintiff must show that a specific misstatement or omission would have been important to a reasonable investor's decision-making.

We investigate the importance of this phrasing by training a BERT-based language model to classify sentences in firm disclosures as opinions or facts. We use the results of this exercise for two sets of analyses. In the first, we examine the effect of changes in the amount of factual statements in a firm's annual reports. These disclosures are subject to liability under Rule 10b-5 if found to be material misstatements or omissions. More certain statements in annual reports allow public investors to be more confident about the disclosed information and thus should increase the equity prices of the firms that make them (Heinle and Smith, 2017). But stating a risk with certainty should also increase the potential for legal liability if those factual statements turn out to be false.

We find evidence that is broadly consistent with these expectations. Firms that increase the factual content of their disclosures tend to experience an increase in their equity prices in the days and weeks following these disclosures. But these firms also increase the chance that they will be subject to a securities lawsuit in the years following the increase in factual content. As this implies, firms that increase the amount of opinion-based content in their disclosures experience a relative decrease in equity prices but also diminish the chance that they will be sued.

In the second analysis, we look at the relationship between the factual content of IPO offering documents and IPO underpricing and how that changed when the Supreme Court altered the legal standard for liability for factual statements and opinion statements in the *Omnicare* case. Statements in IPO offering documents are subject to liability under Section 11 of the Securities Act, which imposes a stricter standard on issuers that Rule 10b-5, and thus those documents should be especially sensitive to this shift.

We find a negative relationship between factual content and IPO underpricing. Offering documents that contain more facts are associated with smaller first-day price increases, a measure of placement underpricing, than those with more opinions. Examining the relationship around the *Omnicare* decision, which ostensibly increased liability for opinions in the federal circuits with the most public firms and decreased it in others, we find a slight increase in IPO underpricing. When we decompose the market's shift in the pricing of opinions after *Omnicare* and the general increase in the opinions in offering documents after *Omnicare*, we find that most of the change is due to the shift in pricing. This finding is consistent with market actors placing less weight on the fact/opinion distinction; in the circuits with the most public firms, those opinions were now more likely to result in liability.

This paper proceeds as follows. Part II provides the relevant background on securities law and reviews the literature on securities pricing and IPO underpricing. Part III describes the data collection process. Part IV presents our analysis of that data, beginning with an intrafirm analysis of the Management's Discussion and Analysis and Risk Factors sections of annual reports. This part then analyzes the amount of factual and opinion-based information in registration statements that firms file in advance of their IPOs before and after the *Omnicare* decision. Part V discusses these results and concludes.

2 Legal Background and Literature Review

This part provides some brief background on the securities laws that govern disclosure by public firms and the liability that can result from any deficiencies in that disclosure. We next review the relevant literature on securities disclosure and IPO underpricing.

2.1 Background on Relevant Securities Law

The Securities Acts of 1933 and 1934 provide much of the framework for disclosure requirements for firms listed in the United States. The Securities Acts and the regulations promulgated under them obligate firms to disclose information that investors would find material. These requirements extend to the information that firms include in a prospectus for an initial public offering and to annual reports (Form 10-K) and quarterly reports (Form 10-Q). Regulations promulgated by the SEC govern the specific content of these documents. For the purposes of this paper, the two most important required sections are Management's Discussion & Analysis (MD&A) and the Risk Factors.

The requirement to include MD&A in 10-Ks is longstanding and predates the adoption of the electronic filing of disclosures in the 1990s. The SEC requires registered companies to provide "material information relevant to an assessment of the financial condition and results of operations of the registrant." 17 CFR § 229.303. This regulation also requires

MD&A to appear in more periodic updates such as quarterly statements (10-Qs).

The obligation to provide a discussion of risk factors is a more recent development. That requirement began in 2005 with the adoption of rule 33-8591, which requires firms to disclose qualitative risk factors in their reports. The rule obligates firms to include risk factors in their annual reports and to provide any updates in quarterly filings. Firms comply with these mandates in a variety of ways. Some include the entirety of their risk factors in all filings, including 10-Qs, while others reference the risk factors listed in their 10-Ks in their quarterly filings and only disclose risk factors that are new or that have changed since the firm's most recent 10-K filing. Smaller reporting companies, which is a definition that has varied over time, are not required to disclose risk factors, although some choose to do so voluntarily.

Material misstatements and omissions that appear in 10-Ks and 10-Qs are usually governed by Rule 10b-5. To establish liability under Rule 10b-5 a plaintiff needs to show that a party made a material misstatement or omission with scienter (a state of mind embracing an intent to deceive), that the plaintiff relied on that misstatement or omission, and that this reliance proximately caused loss to the plaintiff.

In contrast, statements made in a registration statement or prospectus, such as those filed as part of an IPO, are subject to the easier standard provided by Section 11. A plaintiff proceeding under Section 11 need only show that there was a material misstatement or omission in the registration statement and that the plaintiff purchased a security traceable to that statement. There is no requirement to prove scienter, reliance, or loss causation for plaintiffs in a Section 11 case, and courts generally refer to this standard as one of "strict liability."

2.1.1 Omnicare

In *Omnicare, Inc. v. Laborers Dist. Council Const. Industry Pension Fund*, 575 U.S. 175 (2015), investors brought a Section 11 claim against the company for including certain opinions regarding compliance and contractual soundness in its registration statement. Before *Omnicare*, federal circuit courts had split over the requirements to make an opinion actionable under Section 11. The Sixth Circuit promulgated the most lenient and plaintiff-friendly standard. Leaning on the strict liability of Section 11, the Sixth Circuit held that as long as a statement is material and untrue, there is Section 11 liability regardless of whether the statement was phrased as an opinion or a fact. On the other hand, the Second and Ninth Circuits adopted a stricter standard; for an opinion to be actionable under Section 11, plaintiffs would have to prove that the opinion was both objectively false and that the speaker disbelieved the statement.

This split was resolved by the Supreme Court in *Omnicare*. The Court approved three separate paths to Section 11 liability for an opinion, lowering the guardrails in the Second and Ninth Circuits but curtailing the broadness of the Sixth Circuit's standard. First, the Court held that an opinion is actionable if the speaker does not actually hold the view espoused. Second, an opinion is actionable if it contains an embedded statement of fact that turns out to be untrue. Both are categorized as actionable misrepresentations. Third, under an omission theory of liability, the Court held that an opinion statement is actionable if it omits material facts about the issuer's investigation or knowledge about the subject of the opinion. Lower courts applied the holding in *Omnicare* to both dismiss and permit challenges against firms issuing opinions in financial filings which were later alleged to be untrue.[1] In addition, lower courts have broadly expanded the applicability of Omnicare past Section 11, including to Section 10(b) claims in the Second and Ninth Circuits.[2]

The reaction to *Omnicare* varied. Omnicare's lawyers, having defeated the expansive Sixth Circuit standard, heralded the Supreme Court decision as a "significant victory" for the defense bar.[3] To many, *Omnicare* provided welcome guidance from the Court, in the sense that it resolved the previous split between the courts of appeal on the standard for opinion statements in Section 11 claims.[4] Others, however, complained that "*Omnicare* raises more questions than it answers"[5] and that the court's holding was too confusing to be named a win for either plaintiffs or defendants.[6]

Moreover, the opinion inherently intersected with the existing procedural elements of securities litigation. The path to *Omnicare* liability, which require a showing of fraud, would trigger the stricter pleading standards of Federal Rule of Civil Procedure 9(b), making it "more difficult for a plaintiff to adequately plead a Section 11 claim based on a statement of opinion that turns out to have been objectively false."[7] While Rule 9(b)'s standards

[1] *See, e.g., In re Investment Technology Group, Inc. Securities Litigation*, 251 F.Supp.3d 596, 619 (S.D.N.Y. 2017) (dismissing a Section 11 challenge); *Securities and Exchange Commission v. Thompson*, 238 F.Supp.3d 575, 601 (finding a Section 11 claim actionable).

[2] *See Tongue v. Sanofi*, 816 F.3d 199, 209-10 (2d Cir. 2016); *City of Dearborn Heights Act 345 Police & Fire Retirement System v. Align Technology, Inc.*, 856 F.3d 605, 611 (9th Cir. 2017).

[3] Robert A. Van Kirk and John S. Williams, The Supreme Court's Decision in Omnicare: The View From Two Years Out, Bloomberg Law (Aug. 7, 2017).

[4] Darren Robbins, The Supreme Court Rules on Securities Issuers' Liability for Misleading Statements of Opinion, Robbins Geller Rudman & Dowd LLP (Mar. 24, 2015).

[5] Kelly Margolis Dagger, The Supreme Court Addresses Scope of Section 11 Liability for Statements of Opinion, Ellis Winters (Jul. 16, 2015)

[6] Jon Eisenberg, Supreme Court's Omnicare Decision Muddies Section 11 Opinion Liability Standards, K&L Gates LLP (Mar. 31, 2015).

[7] Jordan D. Hershman, Omnicare Decision Clarifies Grounds for Section 11 Liability, Morgan Lewis (Mar. 26, 2015).

would not apply to omission-based allegations, the PSLRA's safe harbor for forward-looking statements offered protection for defendants facing such *Omnicare* challenges. Still, the omission theory was blamed for increasing litigation over opinion statements, as it opened a new path for liability previously unavailable in the Second and Ninth Circuits.[8]

2.2 Literature Review

Our paper analyzes intra-firm changes in annual reports and the amount factual statements in IPO registration statements. This involves research on the relationship between disclosure and asset pricing and the large literature on IPO underpricing.

2.2.1 Disclosure Under the Securities Laws

Early theoretical work on disclosure argued that precise and verifiable disclosure improves asset pricing and lowers a firm's cost of capital by reducing information asymmetry and encouraging liquidity (Diamond and Verrecchia, 1991; Verrecchia, 2001; Healy and Palepu, 2001). Subsequent research recognizes that firms can disclose information with varying degrees of certainty, and finds that the potential for liability or falsification can lead firms to release vaguer information that is more difficult to disprove. Although the net effect of more precise information will depend on the underlying content of that information, theory suggests that the market will pay a premium for more certain information, which also brings increased legal liability if that information turns out to be false.

Despite claims from practitioners and some academics that risk disclosures are boilerplate (Kravet and Muslu, 2013), there is evidence that confirms the theoretical predictions. Campbell et al. (2014) find that the introduction of the obligation to disclose risk in annual reports produced varying amounts of disclosure by firm and that market actors respond to the variations. Such evidence suggests that investors place weight on the content of risk disclosures and that they are not merely boilerplate. Hope et al. (2016) distinguish between general risk disclosures and more specific information, finding that more precise risk information is associated with positive abnormal returns. These empirical findings support the expectation that backing more certain statements with a more certain prospect of liability should reward companies that provide more factual information about risk with a certainty premium while also increasing the risk that they will be sued under the securities laws.

[8] Cleary Gottlieb Alert Memorandum, In Omnicare, Supreme Court Clarifies the Scope of Liability for Statements of Opinion Under Section 11 of the Securities Act of 1933, Cleary Gottlieb (Apr. 3, 2015).

2.2.2 IPO Underpricing

IPO underpricing is commonly interpreted as a market response to information asymmetry between issuers, underwriters, and heterogeneous investors. In the Rock (1986) "winner's curse" model, uninformed investors face adverse selection because better-informed investors only participate when an issue is attractive. To keep the uninformed investors in the game, issuers have to leave "money on the table," i.e., underprice. Benveniste and Spindt (1989) extend this logic to bookbuilding: underwriters elicit private signals from informed institutions by promising them preferential allocations and partial price adjustments, which again implies systematic underpricing as compensation for revealed information.

Empirically, studies have shown that proxies for ex ante uncertainty—a stand-in for asymmetric information—re strongly related to the magnitude of underpricing. Beatty and Ritter (1986) documents that IPOs with greater uncertainty (younger firms, smaller sales, riskier industries) exhibit higher initial returns. Hanley (1993) finds that when the final offer price moves above the preliminary filing range—evidence that valuable information surfaced during marketing—underpricing is especially pronounced, consistent with only "partial" incorporation of new information into the offer price. Loughran and Ritter (2004) further show that while the average level of underpricing varies dramatically across eras, shifts in how information is produced and allocated in bookbuilt offerings help explain these swings.

Survey pieces identify information asymmetry as being among the core explanations for underpricing. Ritter and Welch (2002), reviewing the theory and data, highlight that asymmetric information accounts for many—but not all—IPO regularities, while Ljungqvist (2007) catalogs the principal asymmetric-information models (winner's curse, signaling, bookbuilding) alongside institutional and behavioral alternatives. Together, they reinforce that underpricing is both predicted by, and used to infer the severity of, informational frictions at the IPO stage.

Because first-day returns are observable and tightly linked to these mechanisms, researchers routinely employ underpricing as an empirical proxy for the informational gap surrounding an IPO. The practice rests on the theoretical prediction that greater private information (and costlier information revelation) requires larger inducements, and on the empirical regularity that higher measured uncertainty correlates with larger initial returns.

3 Data

We explain the development of our data in two steps. We begin with the process to extract information from securities disclosures and to classify the content as either fact or opinion at the sentence level. We then discuss the additional data used, including stock price information, details regarding securities lawsuits, and the data about IPOs and underpricing.

3.1 Training a Large Language Model to Classify Opinion and Fact

To discern statements of fact and statements of opinion, we begin with an existing model, FinBERT (Araci, 2019), and train it for this purpose. FinBERT is a variant of the BERT model[9] that has been trained on a corpus of 46,000 Reuters articles that contain certain financial keywords. An advantage of a BERT-based model is that its ability to understand context means that less training data should be necessary to adapt the model for our purpose. Training the model requires ground truth information, which we supply by randomly selecting several hundred sentences from 10-K disclosures and hand coding them as fact, opinion, or neither.

We use the traditional definitions of facts, which are verifiable statements, and opinions, which are statements about states of mind. For example, the statement "some open source software mimics the features and functionality of our products" is a verifiable statement and is classified as a fact. The statement "[w]e believe the closing of the acquisition will likely occur in the first quarter of calendar year 2022" describes a state of mind, and is thus labeled an opinion. Consistent with prevailing securities law, we classify contingent statements using "may," "could," or other speculative language as opinions because this phrasing is generally treated as a non-actionable forward-looking statement. As an example, we label the statement, "[t]hese risks may increase as our products are introduced into new devices, markets, technologies and applications or as new versions are released" as an opinion.

We use this training data to build and test a large language model that classifies a sentence as fact or opinion. To do so, we split the hand coded data into a training set, a tuning set, and a test set. We use the tuning set to assess performance of the model with different parameters and select the model with the best performance. The model produces a probability for each category: opinion, fact, or neither, which represents the strength of its classification into that category. These parameters sum to one. For example, the model might produce an estimate of (.55, .34, .01), in which case the sentence would be classified as an opinion. We then apply that model to the test set of hand-coded data to determine the accuracy of that model against known cases. This approach returns an f1 score for opinions

[9]BERT stands for Bidirectional Encoder Representations from Transformers. It was developed by Google researchers (Devlin et al., 2019) and produced what were, at the time, state-of-the-art performance gains. Unlike earlier context-free models, such as word2vec, BERT is able to understand multiple meanings for the same word.

of .91 and an f1 score for facts of .90.[10]

We next download every 10-K available on the EDGAR platform through 2021—that is every one of the disclosures issued by public firms since the inception of electronic filing—and use an algorithm to extract the MD&A and risk factors text from those filings. We screen the text in each of these sections to take out sentences that are mostly numbers. We then apply the language model to each remaining sentence. If the model predicts that the sentence expresses an opinion or a fact with greater than .80 confidence, we categorize the sentence as an opinion or a fact accordingly. The resulting panel data set produces a number that represents the proportion of sentences that we categorize as facts for each section—MD&A and risk factors–for each filing by each firm.

We filter the data on risk factors to reflect the regulations that govern the filing of that information. The obligation for public issuers to include a risk factors section began in fiscal year 2006. We begin our analysis in 2007 because at that point nearly every firm had entered or completed fiscal year 2006. As discussed above, SEC regulations require complete disclosures about risk factors in a public firm's annual report, but only obligate those firms to provide updates to those risk factors in the interim periods.

Because firms vary in how they respond to this regulation, we choose to focus only on the risk factors in the annual report. That is, for each firm subject to this disclosure obligation, we have an annual observation of the factual content in the risk factors section of that firm's annual report. The requirement to include risk factors excludes so-called "smaller reporting companies." The definition of a "smaller reporting company" has shifted over time, but the relevant threshold depends on a firm's "public float," which is the value of its trading equity. We obtain data on each firm's public float from Ewens et al. (2024) from 2007 to 2019. We require all firms in the risk factors data set to have a public float above the relevant threshold for the year of its filing through 2019. For the same reason, we require that each firm also have at least three years of continuous data in the panel.

In addition to 10-Ks, we also download all S-1 registration statements filed between 2010 and 2024. We extract the Risk Factors section from these statements and apply our fact/opinion algorithm to them. For each S-1, this produces the proportion of sentences in the Risk Factors section that the algorithm classifies as a factual statement.

To provide a sense of how the model operates, we provide several examples of classified sentences from the dataset.

- **Sandisk, 2010 10-K, Risk Factors**, .898 fact: "Our new products have, from time-to-time, been introduced with design and production errors at a rate higher than the

[10] The f1 score for the neutral category performs rather poorly at .52, but this category is rare as it makes up about 6% of the test sample.

error rate in our established products."

- **Impact Mortgage, 2010 10-K, Risk Factors**, .934 opinion: "We believe that our future results will also depend in part upon our attracting and retaining highly skilled and qualified management."

- **Schweitzer Mauduit, 2010 10-K, Risk Factors**, .601 fact: "We expect approval and, if successful, this and other actions should allow our Brazilian operation to utilize more credits than it generates on an annual basis."

3.2 Additional Data

We supplement our data on the factual content of disclosures with additional data from other sources. We obtain information on firm financial reporting information and characteristics from Compustat. For stock price analysis, we obtain abnormal returns from the Eventus platform, which uses security price information from the Center for Research on Securities Prices (CRSP). We obtain information on securities class action lawsuits and settlements from Stanford Securities Litigation Analytics. We use Professor Jay Ritter's website for the IPO data, after removing direct offerings, ADRs, and offerings under $5 per share. Offering price and proceeds are retrieved from the London Stock Exchange Group (LSEG).

4 Analysis

Our analysis proceeds in two parts. We first examine within-firm variation in the amount of factual content in the MD&A and Risk Factor sections of annual reports. We analyze the effect of shifts in that content on abnormal returns and the incidence of securities class action lawsuits. The second section examines the factual content in the Risk Factors section of the S-1 prospectuses for IPOs. We test how the factual content of these filings, and the market response, changed after the United States Supreme Court handed down the *Omnicare* decision.

4.1 Fact and Opinion in Annual Reports

We begin with an overview of the summary statistics from our annual report dataset. The primary variables of interest are "MDA Fact" and "RF Fact", which measure the proportion of sentences that we classify as factual in each of these respective sections. Table 1 presents a yearly summary of the firms in the sample, the variables of interest, the percent of firms incorporated in Delaware, and the average word counts for each section. Several trends are evident from these summary statistics. First, the number of firms that meet the criteria for reporting risk factors has grown over time, roughly doubling from the beginning of the sample to the end. When it comes to the factual content of the MD&A and risk factor sections, we observe different patterns over time.

Table 1: **Yearly Summary of Firms and Financial Metrics**

Year	Firms	$\overline{MDA_{fact}}$	$\overline{RF_{fact}}$	% DE_{incorp}	$\overline{MDA_{words}}$	$\overline{RF_{words}}$
2007	841	0.924	0.559	69.6	8532	4518
2008	826	0.922	0.547	68.0	9034	5033
2009	904	0.918	0.535	69.1	9848	5467
2010	933	0.915	0.531	69.9	9918	5840
2011	969	0.913	0.526	70.3	9918	6330
2012	994	0.913	0.523	71.1	9977	6821
2013	1010	0.913	0.518	72.2	10369	7462
2014	1069	0.913	0.511	74.1	10586	8203
2015	1146	0.912	0.506	75.0	10540	8916
2016	1209	0.913	0.501	76.1	10620	9573
2017	1245	0.912	0.497	76.6	10703	10150
2018	1301	0.911	0.496	77.5	10664	10972
2019	1444	0.910	0.492	78.3	10017	12461
2020	1592	0.905	0.489	78.0	8973	14034
2021	1834	0.901	0.483	80.0	8658	16142

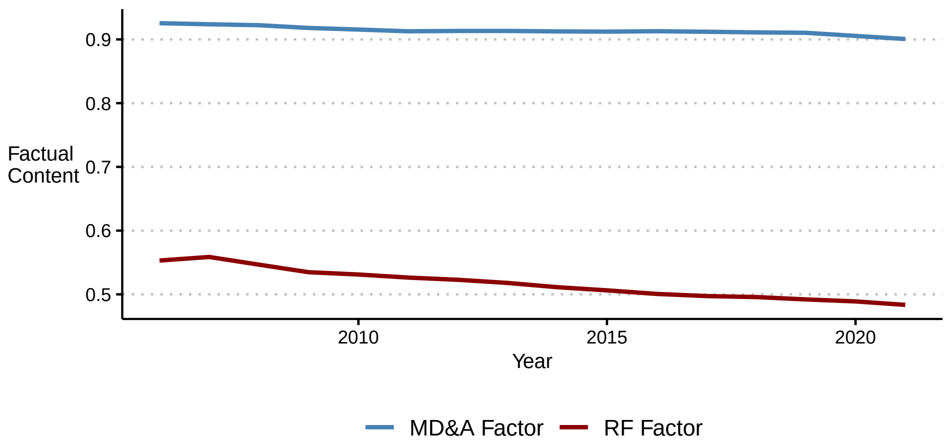
This table reports yearly summary statistics f or firms in our sample. MD&A Fact and RF Fact measure the proportion of sentences classified as factual (with >0.80 confidence) in the Management's Discussion & Analysis and Risk Factors sections of 10-K filings, respectively. The sample includes firms with public float above the smaller reporting company threshold and at least three years of continuous data. Word counts exclude sentences that are predominantly numerical

Figure 1 displays the trends for both the factual content and word counts of the MD&A and risk factor sections. The high amount of factual content in MD&A sections has stayed relatively stable, varying from 92% to 90% over the fifteen years in the sample. In contrast, the decline in factual content for risk factors has been more dramatic. At the beginning of the sample, nearly 56 percent of the classified sentences in the average firm's risk factors section contained factual information. By the end of the sample in 2021, that figure decreases to about 48 percent. The length of these two sections also shows different trends. The MD&A section increases in length at the beginning of the sample, but then decreases at the end. The final average of 8714 words in 2021 is quite close to the average of 8591 in the first year of the sample. The length of the risk factors section has increased dramatically over this period, roughly quadrupling from around 4500 to over 16000 words.

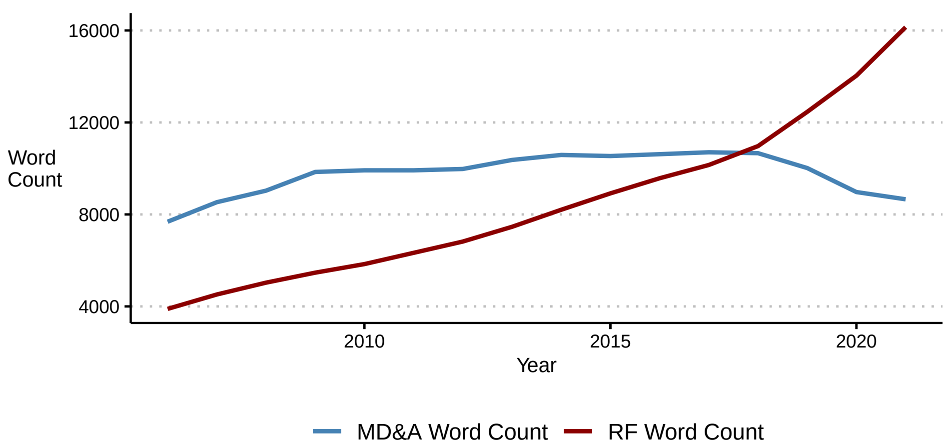
Table 2 provides the yearly levels of factual content in the MD&A and RF sections by

Figure 1: **Time trends in factual content and length.**

(a) Average Factual Content of MD&A and RF Sections



(b) Average Word Count for MD&A and RF Sections



This figure displays yearly averages of factual content and word counts for the Management's Discussion & Analysis (MDA) and Risk Factors (RF) sections of 10-K filings from 2007–2021. Panel A shows the proportion of sentences classified as factual with greater than 0.80 confidence by our fine-tuned FinBERT model. Panel B shows the average word count for each section, excluding sentences that are predominantly numerical. The sample includes firms with public float above the smaller reporting company threshold and at least three years of continuous data.

the eight highest-levels of the Standard Industrial Classification (SIC). The trends for most industries mirror the overall averages, but it is notable that there is substantial variation in the starting and ending points for certain industries. Mining, for example, has roughly 60 percent factual content at the beginning of the sample and remains over 50 percent by the end. Finance, in contrast, starts at about 54 percent factual content and declines to less than 47 percent by 2021.

Table 2: **Average Factors by Industry and Year**

Year	Mining	Cons.	Manuf.	Utilities	Wholesale	Finance	Services	Pub. Admin.
Panel A: MD&A Factors								
2007	0.930	0.923	0.922	0.930	0.924	0.922	0.927	0.898
2008	0.930	0.923	0.919	0.933	0.921	0.918	0.930	0.931
2009	0.923	0.921	0.914	0.927	0.912	0.917	0.929	0.940
2010	0.919	0.917	0.911	0.924	0.911	0.918	0.923	0.929
2011	0.920	0.911	0.908	0.926	0.908	0.915	0.921	0.937
2012	0.919	0.913	0.908	0.926	0.910	0.913	0.929	0.933
2013	0.918	0.912	0.909	0.926	0.910	0.914	0.927	0.926
2014	0.917	0.911	0.908	0.924	0.911	0.913	0.927	0.942
2015	0.916	0.911	0.908	0.923	0.909	0.912	0.926	0.943
2016	0.914	0.910	0.909	0.927	0.910	0.915	0.926	0.944
2017	0.918	0.909	0.907	0.929	0.908	0.913	0.927	0.947
2018	0.915	0.909	0.906	0.925	0.908	0.912	0.924	0.946
2019	0.916	0.903	0.908	0.926	0.908	0.913	0.931	0.907
2020	0.912	0.896	0.907	0.920	0.904	0.906	0.924	0.906
2021	0.909	0.886	0.904	0.918	0.903	0.905	0.913	0.907
Panel B: Risk Factors								
2007	0.603	0.558	0.557	0.571	0.553	0.539	0.562	0.556
2008	0.591	0.546	0.544	0.554	0.547	0.526	0.553	0.530
2009	0.570	0.536	0.533	0.548	0.530	0.514	0.532	0.567
2010	0.571	0.531	0.526	0.547	0.532	0.508	0.534	0.552
2011	0.570	0.530	0.519	0.536	0.530	0.498	0.542	0.538
2012	0.565	0.531	0.515	0.535	0.519	0.496	0.543	0.517
2013	0.555	0.525	0.513	0.535	0.512	0.490	0.539	0.517
2014	0.541	0.516	0.510	0.523	0.508	0.483	0.530	0.533
2015	0.537	0.512	0.505	0.520	0.503	0.480	0.519	0.529
2016	0.533	0.502	0.501	0.515	0.498	0.475	0.515	0.532
2017	0.526	0.501	0.496	0.513	0.494	0.473	0.511	0.515
2018	0.522	0.500	0.493	0.512	0.492	0.474	0.517	0.503
2019	0.515	0.495	0.489	0.510	0.490	0.473	0.520	0.486
2020	0.512	0.489	0.486	0.514	0.485	0.476	0.506	0.481
2021	0.505	0.481	0.484	0.511	0.481	0.469	0.494	0.511

This table reports the yearly average of our factual intensity measure by year and industry. Panel A reports the average values for the MD&A section of the annual report, while Panel B reports the corresponding value for the Risk Factor section.

4.1.1 Returns

Previous research on the content of disclosures provides evidence that period-to-period changes in firm disclosures predict significant changes to the security prices of those firms. These papers use changes in raw word counts or the appearance of new items in these disclosures to measure periodic shifts. Our approach instead captures the legal significance of changes to the language that a firm uses in its disclosures. A firm may add significant content to its risk factors, but if that language is phrased as opinion rather than fact, that language is unlikely to be the basis for a plausible securities class action. Investors are therefore less likely to make certain inferences from that language. In contrast, a disclosure may make very few changes in terms of overall words or the items discussed, while making significant changes to the factual content of a disclosure. At the sentence level, a firm only needs to alter a handful of words for that sentence to have significant consequences for potential legal liability.

As explained above, our focus is on year-to-year changes in the factual content of the MD&A and RF sections of each firm's annual report. Our variables of interest are the lagged difference between the current value and the previous value for each section:

$$\Delta MDA_{i,t} = MDA_{i,t} - MDA_{i,t-1}$$

$$\Delta RF_{i,t} = RF_{i,t} - RF_{i,t-1}$$

For ease of interpretation, all independent variables are standardized to have mean zero and standard deviation one, so that each coefficient reflects the effect of a one standard deviation change in the relevant variable. In the analysis that follows, we use cumulative abnormal stock returns and estimate regressions of the following form:

$$\widehat{CAR_{i,t}} = \beta_1 \Delta Section_{i,t} + \delta_i + \phi_t + \epsilon_{i,t}$$

Where $\widehat{CAR_{i,t}}$ is the cumulative abnormal return in excess of value-weighted market returns for the relevant window for firm i in time period t, $\Delta Section_{i,t}$ is the change in factual content for the relevant disclosure section for firm i in period t, and X_i are a firm fixed effect and a year fixed effect. The regressions reported in Table 4 report the CARs for the [-1,+10] and [-1,+20] windows, which covers the cumulative abnormal stock returns for the periods running from the day before the release of the annual report to ten and twenty trading days after the release.

Table 3 shows a large positive effect for increases in factual content in the risk factors section of a firm's annual report. This effect is statistically significant at the one-percent

Table 3: **Cumulative Abnormal Returns and Disclosure Changes**

Panel A: Baseline

	CAR[-1,+10]				CAR[-1,+20]			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta_{RF_{fact}}$	0.220**				0.372***			
	(0.086)				(0.109)			
$\Delta_{RF_{words}}$		-0.139				-0.224***		
		(0.100)				(0.085)		
$\Delta_{MDA_{fact}}$			-0.033				-0.118	
			(0.118)				(0.133)	
$\Delta_{MDA_{words}}$				-0.163				-0.400
				(0.453)				(0.350)
Observations	12,996	12,996	12,996	12,996	12,996	12,996	12,996	12,996
R^2	0.154	0.154	0.154	0.154	0.166	0.166	0.166	0.166
Adjusted R^2	0.048	0.048	0.048	0.048	0.061	0.061	0.061	0.061
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Controlling for $\Delta_{LM_{unc}}$

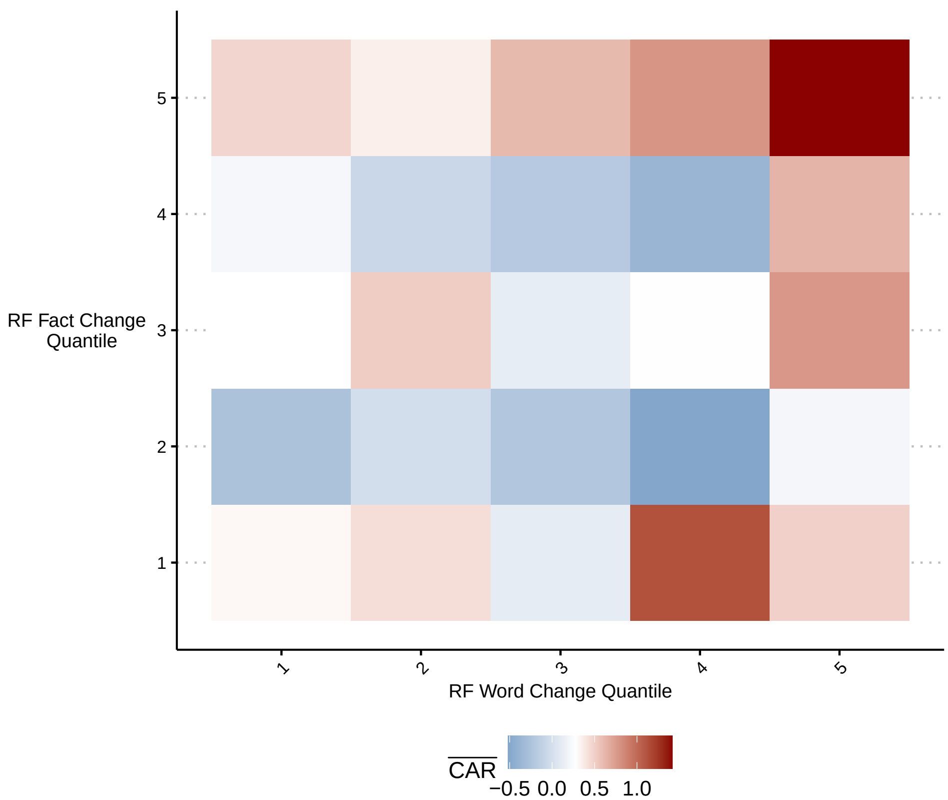
	CAR[-1,+10]				CAR[-1,+20]			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta_{RF_{fact}}$	0.216**				0.375***			
	(0.086)				(0.109)			
$\Delta_{RF_{words}}$		-0.139				-0.224***		
		(0.100)				(0.085)		
$\Delta_{MDA_{fact}}$			-0.038				-0.116	
			(0.117)				(0.131)	
$\Delta_{MDA_{words}}$				-0.162				-0.400
				(0.453)				(0.350)
$\Delta_{LM_{unc}}$	-0.070	-0.079	-0.081	-0.079	0.051	0.036	0.030	0.036
	(0.091)	(0.091)	(0.091)	(0.091)	(0.116)	(0.116)	(0.115)	(0.116)
Observations	12,996	12,996	12,996	12,996	12,996	12,996	12,996	12,996
R^2	0.154	0.154	0.154	0.154	0.166	0.166	0.166	0.166
Adjusted R^2	0.048	0.048	0.048	0.048	0.061	0.061	0.061	0.061
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The dependent variable is the cumulative abnormal return in excess of the value-weighted market return for each firm in the relevant window. The independent variables measure the year-to-year changes in the percent of sentences classified as facts and the word counts in the RF and MD&A sections of the firm's annual report. Panel B additionally controls for $\Delta_{LM_{unc}}$, the year-to-year change in the Loughran-McDonald uncertainty word proportion for the full 10-K filing. All independent variables are standardized to mean zero and standard deviation one. Robust standard errors, which are reported in parentheses, are clustered by industry. *$p< 0.1$; **$p< 0.05$; ***$p< 0.01$.

level for both the [-1,+10] and [-1,+20] day windows. This result suggests that increasing the amount of factual content in this section relative to the previous year increases equity returns for the firms that do so. This result is consistent with there being a certainty premium that comes with providing information to investors with more confidence. We also find no significant effect associated with changes to the factual content or word counts in the MD&A sections. This is, perhaps, not surprising because the factual content of this information tends to be high in almost all cases.

Consistent with prior work (Cohen et al., 2020), we also find a negative effect of an increase in word counts in the RF section on abnormal returns. Our results suggest that it is not just the addition of words that matters. Rather, the content and phrasing of those words is also important. In Panel B, we show that these results are robust to controlling for the year-over-year change in the Loughran and McDonald (2013) textual uncertainty measure, defined as the fraction of "uncertainty" words in the 10-K filing. Figure 2 provides a heat map of the average cumulative abnormal return by decile of lagged changes in RF fact and word counts. The figure shows that, where firms change a large amount of factual content, they can experience large positive abnormal returns, even if the overall word count in those sections does not change very much.

Figure 2: **Heatmap of CAR and Factual Intensity**



This figure displays average cumulative abnormal returns (CARs) for the [-1,+20] day window around 10-K filing dates, binned by deciles of year-over-year changes in Risk Factors word count (x-axis) and factual content (y-axis). CARs are calculated in excess of value-weighted market returns. Red cells indicate positive abnormal returns; blue cells indicate negative abnormal returns. The figure shows that firms with large increases in factual content (top rows) tend to experience positive abnormal returns regardless of changes in word count.

As discussed earlier, an increase in factual content may also bring the added risk of liability under the securities laws. Many securities lawsuits allege that a firm has misstated the risk that it faces. Framing a statement about risk in a factual way is substantially more likely to result in liability than if that statement were phrased as an opinion. To explore that possibility, we run regressions similar to those in Table 3, but instead of using cumulative abnormal returns as the dependent variable, we use whether the firm is subject to a securities class action in the period following the release of the annual report as the dependent variable.

Table 4 reports the results of these regressions. In the year following the release of the annual report, the coefficient for changes to the factual content of risk factors section is positive, but not statistically significant. If we instead use an indicator for whether a firm gets sued in the two years following the annual report, the coefficient is positive and is statistically significant at the one-percent level. This is consistent with securities violations taking time to materialize, which will typically not be discovered until the firm makes a corrective disclosure. Our finding is consistent with the tradeoff suggested above: increasing factual content provides investors with more confidence about the risk the firms disclose, but can also subject the firm to a greater risk of being sued under the securities laws.

Table 4: **Future Lawsuits and Disclosure Changes**

	Sued $t+1$				Sued $t+1$ or $t+2$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta_{RF_{fact}}$	0.003*		0.003**		0.005***		0.005***	
	(0.002)		(0.002)		(0.002)		(0.002)	
$\Delta_{MDA_{fact}}$		-0.001		-0.001		-0.001		0.000
		(0.002)		(0.002)		(0.002)		(0.002)
$\Delta_{LM_{unc}}$			0.002	0.002			0.002*	0.002
			(0.001)	(0.001)			(0.001)	(0.001)
Observations	14,690	14,690	14,690	14,690	14,690	14,690	14,690	14,690
R^2	0.152	0.152	0.152	0.152	0.253	0.253	0.254	0.253
Adjusted R^2	0.050	0.050	0.050	0.050	0.164	0.164	0.164	0.164
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The dependent variable is whether the firm is the subject of securities class action in the relevant time period. The independent variables measure the year-to-year changes in the percent of sentences classified as facts in the RF and MDA sections of the firm's annual report. Columns (3), (4), (7), and (8) additionally control for $\Delta_{LM_{unc}}$, the year-to-year change in the Loughran-McDonald uncertainty word proportion for the full 10-K filing. Robust standard errors, which are reported in parentheses, are clustered by industry. *p< 0.1; **p< 0.05; ***p< 0.01.

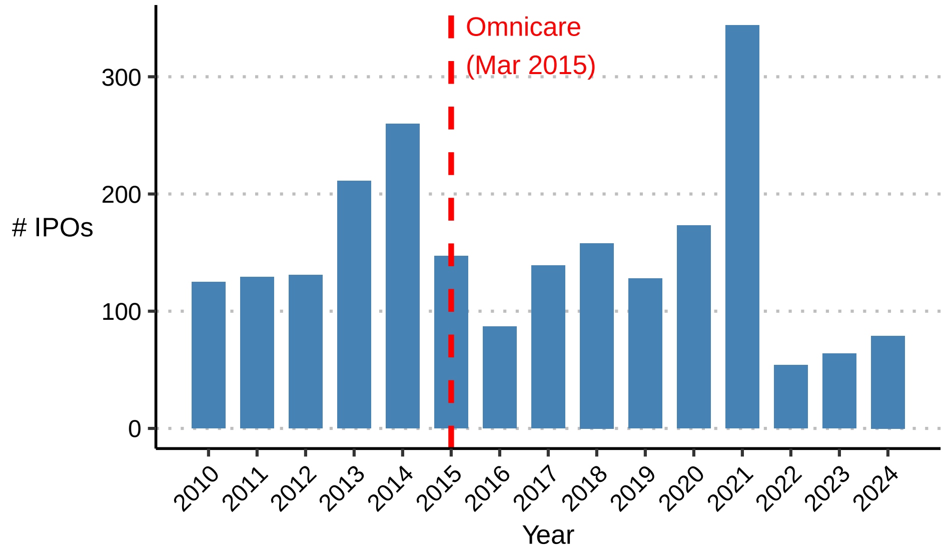
4.2 *Omnicare* and IPO Underpricing

When the stock price of a newly listed company jumps on the first day of trading, it suggests that the company left money on the table during the course of its offering. Information asymmetry between the issuer and the investing public is the traditional explanation for why this underpricing happens, and empirical evidence supports the argument that greater asymmetry is associated with more underpricing. Litigation under our securities laws plays a role in this information asymmetry; where a new issuer makes a statement that can result in liability, investors should be more likely to trust those statements. When managers choose to frame disclosures as statements of opinion, making them more difficult to recover than factual statements, rational investors will place less weight on them. The litigation backstop means that the mixture of fact and opinion in the firm's discussion of the risks it faces should provide a measure of information asymmetry between the firm and potential investors.

To analyze the relationship between legal enforceability and IPO underpricing, we construct a dataset containing our fact and opinion measure for the risk factor section from firms' Form S-1 or similar filings in advance of the offering. Our sample is constructed from the set of all IPOs listed on Professor Jay Ritter's IPO database that had an offer date between 2010 and 2024, from which we remove all American depository receipts (ADRs), closed-end funds, domestic or international equity offerings, and municipal and taxable bond offerings.[11] We match the observations in the Ritter dataset to the text of the corresponding SEC registration statement in the SEC Analytics Suite from Wharton Research Data Services (WRDS). If the filing is amended, we use the most recent filing before the offering date. We extract the text of the Risk Factor section from the filing, and calculate the RF fact measure each IPO in the same manner as described above.

[11] Formally, we require that the ADR column is equal to 1.

Figure 3: **IPO Volume by Year**



This figure shows the annual number of IPOs in our sample that are not American depository receipts (ADRs), closed-end funds, domestic or international equity offerings, and municipal and taxable bond offerings, that have an offering price above $5, and for which we can calculate our measures of fact-intensity, underpricing, and covariates.

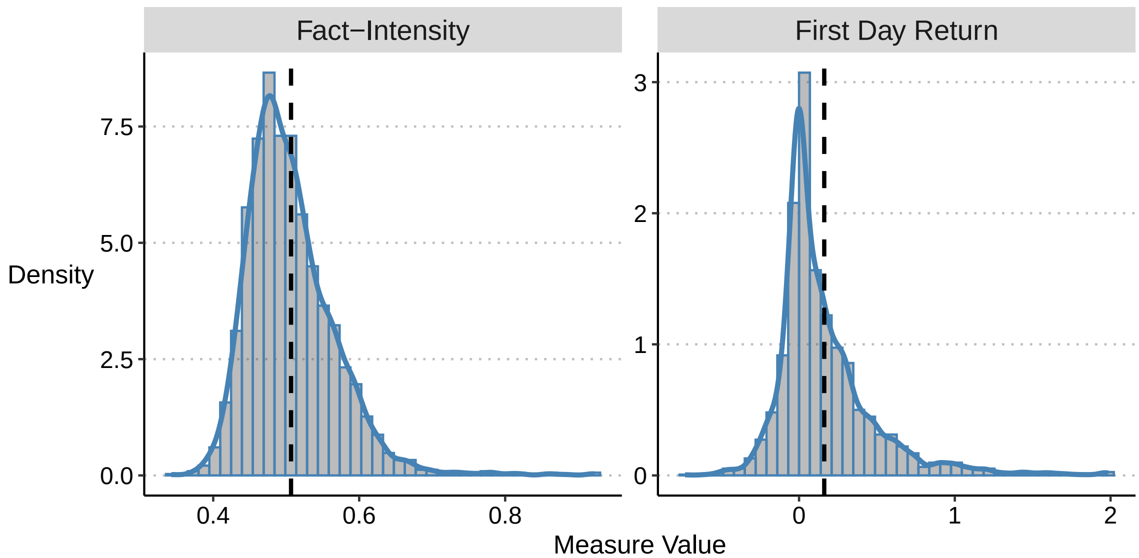
Figure 3 shows a histogram of the yearly initial public offering count that satisfy our inclusion criteria. For these IPOs, we retrieve the offering price from the London Stock Exchange Group (LSEG) database and publicly reported trading data from the Center for Research in Security Prices (CRSP). Our primary underpricing measure is defined as:

$$U_i = \frac{p_{i,1} - p_{i,O}}{p_{i,O}}$$

where U_i is the underpricing value of the IPO of firm i, $p_{i,1}$ is the closing price of the stock of firm i on the first trading date after the offer, and $p_{i,0}$ is the final offering price. Following prior work in this literature, we remove observations with an offering price below $5 per share to focus on economically relevant equity issuances.

Finally, we also generate a set of control variables that have been identified as determinants of underpricing in prior work, including the size of the firm (as measured by the natural logarithm of firm assets before going public), the size of the offering (the logarithm of the offering proceeds), underlying market fundamentals (the rolling 250-day cumulative return of the CRSP value-weighted market return index for the offering date), the firm's industry profile (the Fama-French 12 industry designation as mapped from their 4-digit SIC code), and the proportion of words in the risk factor section that appear in the Loughran and McDonald (2013) uncertainty dictionary. We require there to be no missing values for the underpricing measure or covariates for an IPO to enter our sample. The resulting dataset contains 2,229 IPOs between 2010 to 2024.

Figure 4: **Fact Intensity and IPO Underpricing**



This figure displays the distribution of first-day returns for U.S. IPOs and our fact-intensity measure. First-day return is calculated as the percentage change from the offer price to the closing price on the first trading day. Histograms show the density of observations in each period, with kernel density estimates overlaid. Vertical dashed lines indicate sample means. Fact-intensity is measured using a BERT-based classifier that scores sentences on a fact-opinion continuum, with higher values indicating more factual language.

4.2.1 Fact-Intensity and Underpricing

Figure 4 reports histograms (with overlaid kernal density estimates) for our fact-intensity and underpricing measures. For the fact-intensity variable, higher values represent risk factor sections with more factual statements (and lower values represent sections with more opinions). We see that over the entire sample the mean of our measure is around 0.5, with some mass skewed towards offerings in the right tail of factualness. For our underpricing measure—the first day return from the offering—the median return is 8%, and the mean is 19%, showing the right-tail skewness in the measure. The third quartile of the distribution has a first-day return of 28% and the 90% quintile has a return of 57%. Although there is some mass below zero, the corresponding first quartile of the distribution is a return of only -1.3%, consistent with popular media and academic commentary on *under*pricing in the IPO market.

Figure 5 presents two ways of visualizing the relationship between our fact-intensity measure and IPO underpricing. In the left panel we report a simple binscatter of the relationship, both with (red) and without (blue) the inclusion of covariates. Binscatter is a nonparametric visualization method that shows the conditional expectation $E[Y|X]$ by first grouping the independent variable (fact-intensity) into bins and then taking the average of the outcome variable (underpricing) within each bin. It is essentially a scatter plot of conditional or unconditional means when you have too many datapoints for a raw scatter plot to be non-informative. We follow Cattaneo et al. (2024) and construct confidence intervals that are validly uniform across bins, while correctly partialing out covariates from both the outcome and independent variable. We see that there is a negative relationship between fact-intensity and underpricing, which is somewhat obscured by the most factually-intense offering documents. This relationship is slightly attenuated when the covariates are included, but the general pattern is consistent.

The right panel displays marginal effects from a distribution regression specification (Chernozhukov et al., 2013). Distribution regression models the entire conditional distribution of Y given X by estimating a series of binary regressions:

$$P(Y_i > \tau \mid X_i) = \Lambda(X_i'\beta(\tau))$$

for a grid of thresholds τ across the support of Y and a given link function Λ. Each threshold gives you a binary outcome model, and stacking them traces out the full conditional CDF.

Here, for the quantiles $q \in \{0.5, 0.95\}$ of the first-day return distribution, we estimate:

$$P(\mathrm{U}_i > q \mid \mathrm{RF}_i, X) = \Lambda(\beta_0(q) + \beta_1(q) \cdot \mathrm{RF}_i + X'\gamma(q))$$

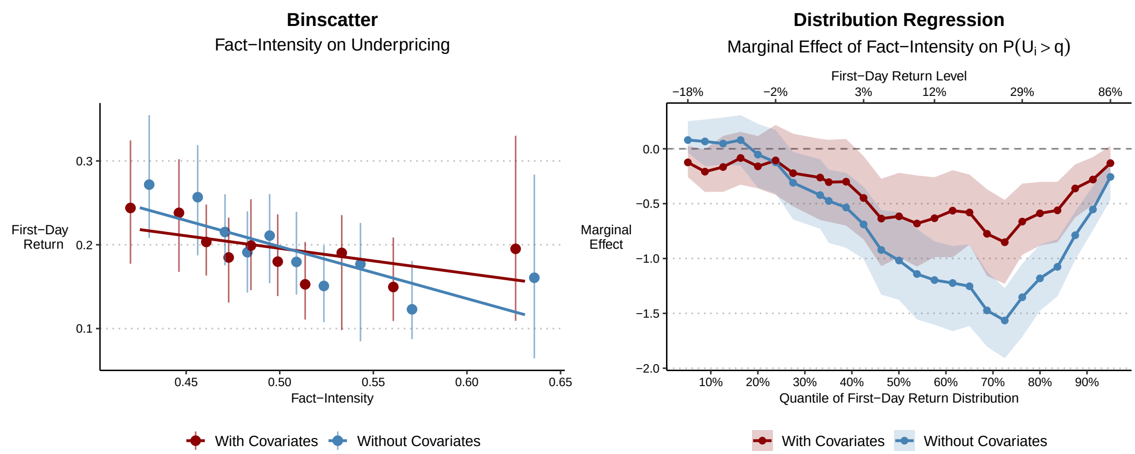
where U_i is the first-day return, RF_i is our fact-intensity measure, and X is a matrix of covariates including firm size (log assets), deal size (log proceeds), market conditions (30-day rolling CRSP value-weighted return), and Fama-French 12 industry fixed effects. The plotted line in Figure 5 is $\hat{\beta}_1(q)$ — the marginal effect of fact-intensity on the probability of exceeding quantile q — across the distribution. In Figure 5, for each quantile q of the first-day return distribution, we estimate $P(U_i > q|X)$ and plot the marginal effect of the fact-intensity measure throughout the distribution.[12] The top axis shows the first-day return levels U_i corresponding to the quantile thresholds. Negative values indicate that higher fact-intensity reduces the probability of exceeding that return threshold. Shaded regions denote 95% pointwise confidence intervals.

We see that there are negative effects throughout the distribution, with higher fact-intensity being associated with a lower probability that returns exceed the threshold value across the distribution. The strongest effects are at the 50th-70th percentiles, suggesting that

[12] We follow Delgado et al. (2022) and use a cloglog link function.

fact-intensity has the largest impact on moderate-to-high underpricing, not the extreme tails. The blue line reports the marginal effects without covariates, and the red line shows the associated values once controlling for the other confounding variables. The inclusion of controls reduces the magnitude here somewhat, but the pattern persists. Whereas a simple ordinary least squares regression gives you one conditional effect number, $E[Y|X]$, distribution regression shows where in the distribution the effect operates. Here, fact-intensity compresses the distribution by reducing the probability of moderate-to-high underpricing, rather than shifting the entire distribution uniformly.

Figure 5: **Relationship Between Fact-Intensity and IPO Underpricing**

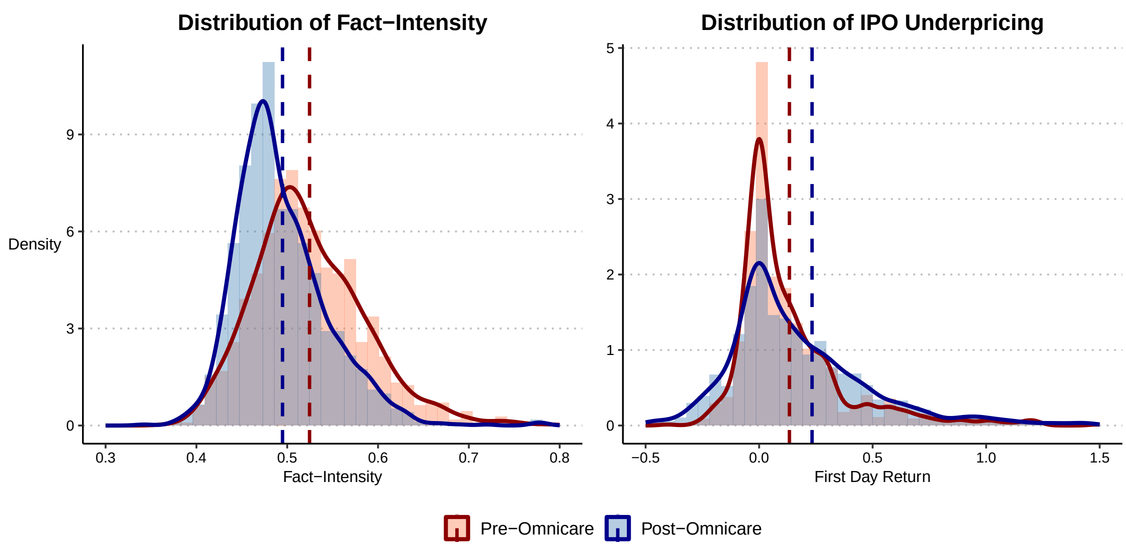


This figure examines the relationship between prospectus fact-intensity and first-day returns. The left panel presents binscatter plots using the binning procedure of Cattaneo et al. (2024), showing conditional means of first-day returns within each bin of fact-intensity. Panel B presents distribution regression estimates following Chernozhukov et al. (2013), plotting the marginal effect of fact-intensity on the probability that first-day returns exceed each threshold. In both panels, blue series show baseline estimates without controls; the red series include controls for firm size (log assets), deal size (log proceeds), market conditions (30-day rolling CRSP value-weighted return), and Fama-French 12 industry fixed effects. Shaded areas represent 95% confidence intervals (bootstrap, 200 replications for distribution regression). The sample includes U.S. IPOs from 2010–2024 with offer prices of at least \$5 (N = 2,229).

4.2.2 Change in IPO Underpricing and Fact-Intensity Following *Omnicare*

In general, for both binscatter and distribution regression in Figure 5, we see a negative relationship between the amount of facts used in the risk factor sections of the offering document and the amount of proceeds required to be left on the table in an IPO. As discussed in Section 2.1.1, the *Omnicare* decision plausibly shifted the liability standard for statements of opinions in the registration statement. While previously the 2nd and 9th circuits (which contain a disproportionate number of publicly traded firms) followed a rule where there could be no Securities Act liability for statements phrased as opinions, the *Omnicare* decision opened multiple (if difficult) avenues for remedy.[13] Figure 6 reports the distributions of the fact-intensity and underpricing measures, split by the time period relative to the decision. In the post-*Omnicare* period we see that firms issue risk factor sections with a higher proportion of opinions than before the decision. In addition, we see that the first day returns shift outward, particularly in the right tail, showing a higher percentage of underpricing after the decision.

Figure 6: **Distributional Shifts in Fact-Intensity and Underpricing: Pre vs. Post *Omnicare***



This figure displays the distributions of prospectus fact-intensity and IPO first-day returns before and after Omnicare v. Laborers District Council. Panel A shows the distribution of fact-intensity, measured using a BERT-based classifier that scores sentences on a fact-opinion continuum. Panel B shows the distribution of first-day returns, calculated as the percentage change from offer price to first-day closing price. Histograms display the density of observations with kernel density estimates overlaid. Vertical dashed lines indicate period means. The sample includes U.S. IPOs from 2010–2024 with offer prices of at least $5 (N = 882 pre-Omnicare; N = 1,347 post-Omnicare).

Figure 7 examines how the relationship between prospectus fact-intensity and IPO underpricing changed after *Omnicare*. The top row presents binscatter plots and the bottom row presents distribution regression marginal effects, while the left column shows baseline estimates, and the right column includes controls. We also split the estimates by period; the relationship before *Omnicare* is reported in red, and after *Omnicare* is reported in blue.

There is a strong negative relationship between fact-intensity and first-day returns before *Omnicare*. In the binscatter, the slope is steeply negative, reflecting how IPOs with higher

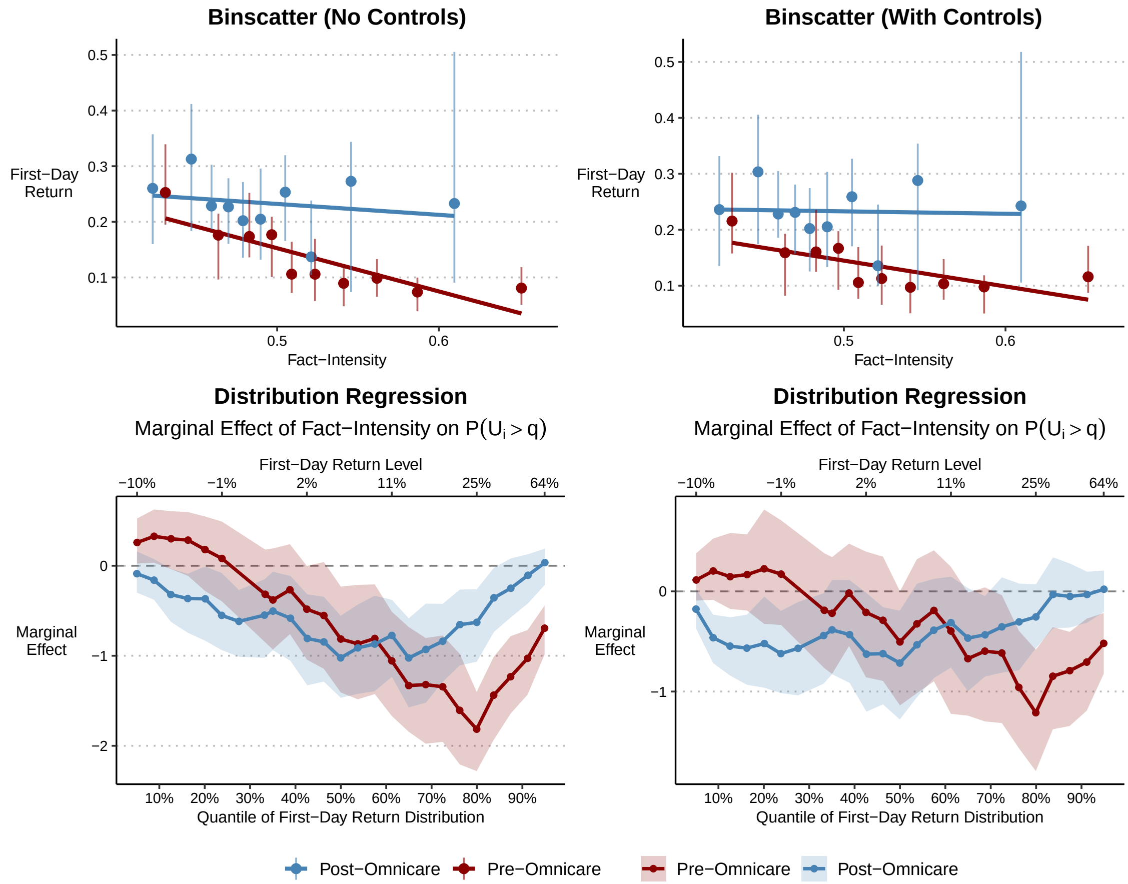
[13]The prevailing standard shifted in the other direction for firms headquartered in the 6th circuit; however, these firms make up a very small portion of the sample in both the pre- and post-decision period, and the results are substantively similar if we exclude these firms from the analysis.

fact-intensity in their risk factor disclosures experienced substantially lower underpricing. The distribution regression confirms this pattern is not confined to the conditional mean; marginal effects are negative and statistically significant across nearly the entire distribution, with the strongest effects concentrated between the 40th and 80th percentiles. The pattern persists in both columns, although the inclusion of controls modestly attenuates the pre-*Omnicare* relationship.[14] This suggests that factual risk disclosure reduced the probability of both moderate and high underpricing during this period.

Following *Omnicare*, the relationship becomes less sensitive. The binscatter slope flattens considerably, indicating that fact-intensity no longer correlates strongly with underpricing. Average first-day returns are on the order of 25% across quantile bins of factual intensity. The distribution regression reveals a similar pattern, with the marginal effect curve shifting upwards across most quantiles, and confidence intervals frequently including zero. The negative relationship, although not entirely eliminated, is substantially weaker, which is consistent with *Omnicare* altering the informativeness of prospectus disclosures. Before the decision, fact-intensive risk factor sections appeared to resolve investor uncertainty, reducing information asymmetry, and thus lowering underpricing. After *Omnicare* potentially increased issuer liability exposure for opinion statements, the signaling value of factual disclosure decreased relative to opinions, consistent with investors acknowledging a smaller difference in liability standards across disclosure type.

[14] Moreover, the smaller sample in the pre-*Omnicare* period, coupled with the inclusion of the control variables, increases the 95% simultaneous confidence in intervals in the plot, particularly for the middle part of the distribution.

Figure 7: **Fact-Intensity and Underpricing by Period: Pre vs. Post Omnicare**

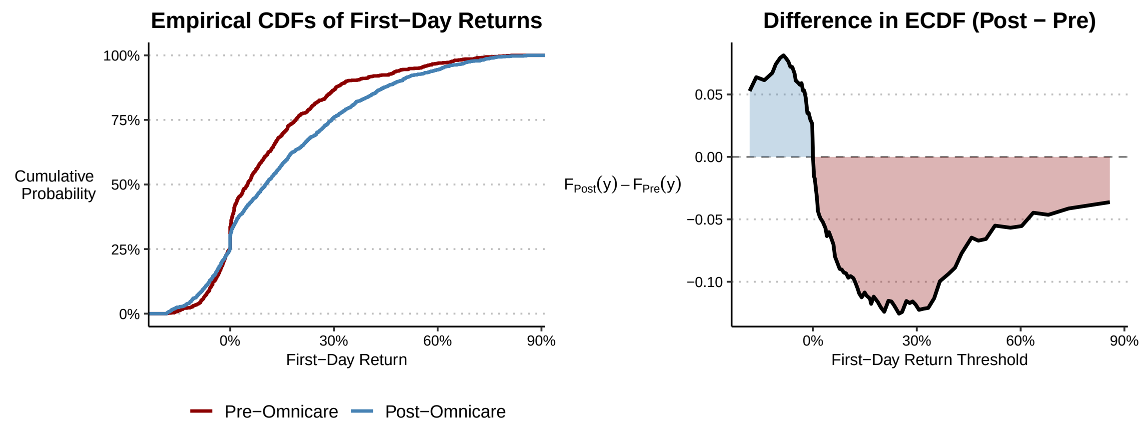


This figure examines how the relationship between prospectus fact-intensity and IPO first-day returns differs before and after *Omnicare*. The top row presents binscatter plots using the binning procedure of Cattaneo et al. (2024), showing conditional means of first-day returns within each bin of fact-intensity. The bottom row presents distribution regression estimates (Chernozhukov et al., 2013), plotting the marginal effect of fact-intensity on the probability that first-day returns exceed each threshold τ. The left column shows baseline estimates; the right column includes controls for firm size (log assets), deal size (log proceeds), market conditions (30-day rolling CRSP value-weighted return), and Fama-French 12 industry fixed effects. Red series denote the pre-*Omnicare* period; blue series denote post-*Omnicare*. Shaded areas and vertical bars represent 95% confidence intervals. The sample includes U.S. IPOs from 2010–2024 with offer prices of at least \$5 (N = 882 pre-Omnicare; N = 1,347 post-Omnicare).

Figure 6 showed the shift the probability distribution of IPO underpricing using kernel density estimation. Another way to visualize a distributional shift in first day returns is to integrate over the PDF to uncover the empirical cumulative distribution function (ECDF). Figure 8 presents empirical ECDFs of first-day returns for the pre- and post-*Omnicare* periods. The left panel overlays the two ECDFs, while the right panel reports the difference in ECDFs over the return distribution, $F_{Post}(y) - F_{Pre}(y)$, for different values of first-day returns y.

Again, we see how IPOs generally exhibit higher first-day returns (greater underpricing) after *Omnicare*, as the cumulative probability for the pre-*Omnicare* period (in red) lies above the post-*Omnicare* line (in blue). This is not true for negative returns—in fact we see a higher mass of IPOs that *lose* money on the first day after trading. However, for offerings that earn a positive return on the first day, the right panel shows that $F_{Post}(y) - F_{Pre}(y)$ is negative, indicating that the post-Omnicare return distribution has shifted to the right.

Figure 8: **Change in Empirical Cumulative Distribution Functions After *Omnicare***



This figure compares the cumulative distribution of IPO first-day returns before and after *Omnicare*. The left panel displays empirical cumulative distribution functions (ECDFs) for each period. The right panel plots the difference $F_{Post}(y) - F_{Pre}(y)$ at each return threshold. Negative values (red shading) indicate the post-*Omnicare* CDF lies below the pre-Omnicare CDF, implying a rightward shift toward higher returns. The sample includes U.S. IPOs from 2010–2024 with offer prices of at least \$5 (N = 882 pre-Omnicare; N = 1,347 post-Omnicare).

This observed distributional shift could arise from different sources. It could be that the composition of the offerings has changed (i.e. firms are issuing more factual statements or of different size). In addition, it could be that the market prices these factors differently. As mentioned earlier, in the post-*Omnicare* world, firms have more avenues to be sued for false opinions in their registration statement, and rational investors should price the IPOs accordingly. To address this question, we employ the counterfactual decomposition framework of Chernozhukov et al. (2013), which allows us to decompose the total distributional change into a composition effect—the change attributable to shifts in the distribution of the offering characteristics—and a structure effect that reflects how the market prices those characteristics. This is the distributional analog to the commonly-used decomposition of mean effects from Oaxaca (1973) and Blinder (1973).

Let U_j represent underpricing and RF_j again represent our fact-intensity variable for offerings in period j, with $j = 0$ denoting offerings that occur before the decision and $j = 1$ afterwards. The conditional distribution functions $F_{U_0|RF_0}(U|RF)$ and $F_{U_1|RF_1}(U|RF)$ describe the stochastic assignment of underpricing to firms with fact-intenstiy RF, both before and after *Omnicare* respectively. Following Chernozhukov et al. (2013), we simplify notation with $F_{Y\langle 0|0\rangle}$ and $F_{Y\langle 1|1\rangle}$ representing the observed distribution function of underpricing for firms that IPO before and after respectively. Here we can specify a distribution $F_{Y\langle 0|1\rangle}$ as the counterfactual distribution function of underpricing that would have prevailed for offerings after *Omnicare* had they faced the same underpricing schedule as firms that did offerings before *Omnicare* $\left(F_{Y\langle 0|0\rangle}\right)$:

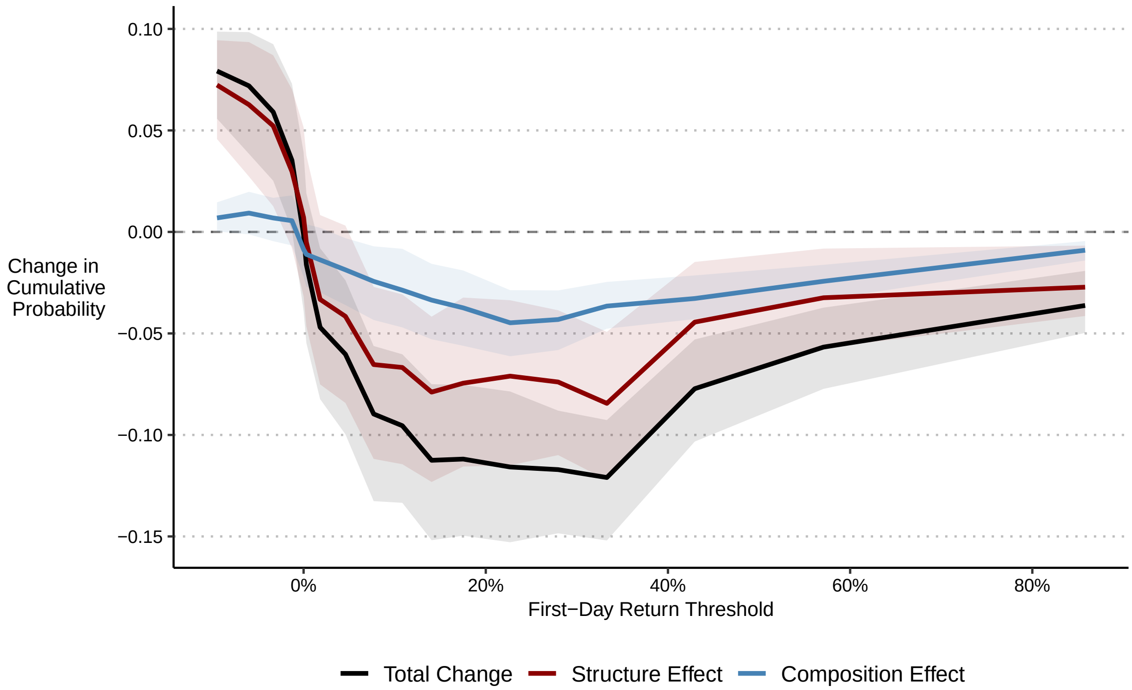
$$F_{Y\langle 0|1\rangle}(U) = \int_{RF_1} F_{\langle U_0|RF_0\rangle}(U \mid RF)\,\mathrm{d}F_{RF_1}(RF).$$

We do not observe this distribution. Instead, it is constructed by integrating the conditional distribution of underpricing for firms before the IPO with respect to the distribution of our fact-intensity variable for firms that IPO afterward. The difference in the distributions of underpricing before and after the decision from Figure 8 can be decomposed as:

$$F_{Y\langle 1|1\rangle} - F_{Y\langle 0|0\rangle} = \left[F_{Y\langle 1|1\rangle} - F_{Y\langle 0|1\rangle}\right] + \left[F_{Y\langle 0|1\rangle} - F_{Y\langle 0|0\rangle}\right]$$

The first term in brackets is due to difference in the structure effect and the second term is the composition effect due to differences in the underlying mix of fact and opinion. That is, the first term shows how the market reactions changed to a given level of facts in the risk factors, while the second term represents the difference in how frequently firms use facts versus opinions.

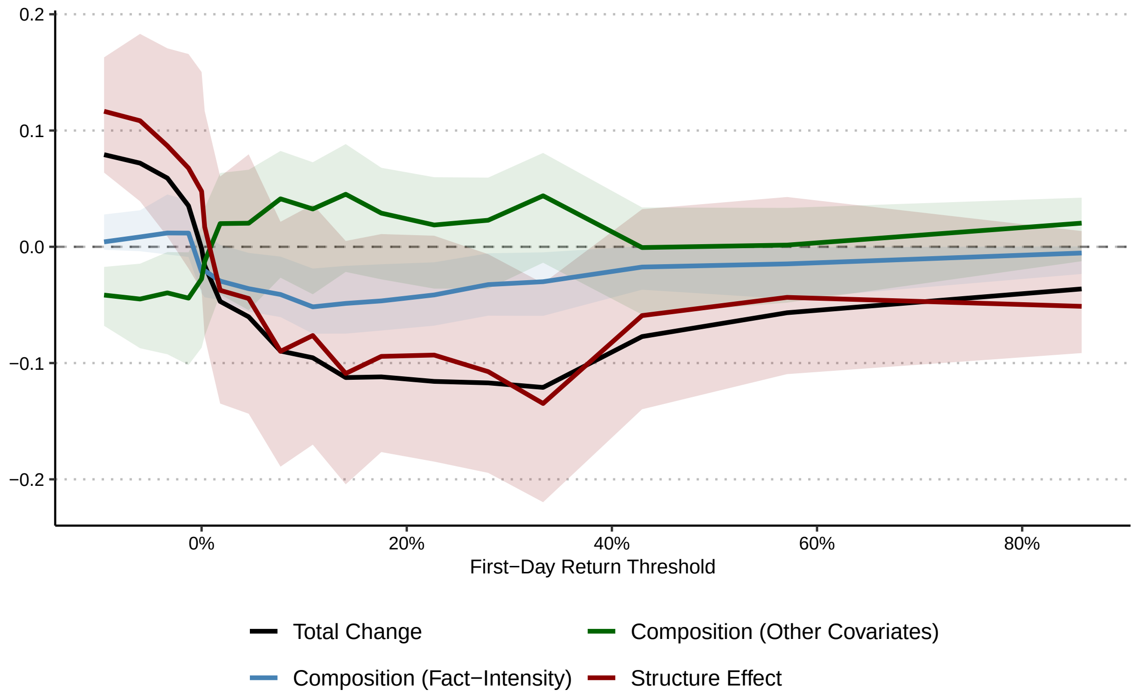
Figure 9: **Counterfactual Decomposition of Changes in IPO Underpricing: Pre vs. Post Omnicare Without Covariates**



This figure decomposes the change in the distribution of IPO first-day returns between the pre- and post-*Omnicare* periods using the counterfactual decomposition framework of Chernozhukov et al. (2013). The y-axis plots the change in cumulative probability at each first-day return threshold; negative values indicate reduced probability mass at or below that threshold (i.e., a rightward shift toward higher returns). The total change (black) is decomposed into a structure effect (red)—reflecting changes in how the market prices given characteristics—and a composition effect (blue)—reflecting changes in the distribution of characteristics themselves. Shaded areas represent 95% bootstrap confidence intervals (300 replications). The sample includes U.S. IPOs from 2010–2024 with offer prices of at least \$5.

We present this decomposition in Figure 9. The black line shows the change in the cumulative density of IPO underpricing after the *Omnicare* decision, which is the same line shown in the right panel of Figure 8. The blue line shows the composition effect, or the expected change in IPO underpricing given the change in fact-intensity of the offering documents. Although negative, it does not do much to explain the observed increase in underpricing (or equivalently, the decrease in the cumulative density function of the first-day returns). The decomposition reveals that virtually all of this shift is attributable to the structure effect (red line), which tracks the total change closely throughout the distribution. The structure effect captures changes in how the market prices a given level of fact-intensity—that is, the relationship between disclosure quality and underpricing. The magnitude of this component suggests that the conditional distribution of returns given fact-intensity fundamentally changed after *Omnicare*. The increase in underpricing does not appear to be driven largely by issuers providing less factual disclosure; rather, it reflects a change in how investors value that disclosure.

Figure 10: **Counterfactual Decomposition of Changes in IPO Underpricing: Pre vs. Post Omnicare With Covariates**



This figure extends the CFM decomposition to a three-way attribution that controls for firm size (log assets), deal size (log proceeds), and market conditions (30-day rolling CRSP value-weighted return). The total change (black) is decomposed into: a structure effect (red), reflecting changes in how the market prices given characteristics; a composition effect from fact-intensity (blue); and a composition effect from other covariates (green). Shaded areas represent 95% bootstrap confidence intervals (300 replications). The sample includes U.S. IPOs from 2010–2024 with offer prices of at least $5.

Figure 10 reports the corresponding decomposition when controlling for other determinants of IPO underpricing identified in previous work, including firm size (log assets), deal size (log proceeds), market conditions (30-day rolling CRSP value-weighted return), and industry fixed effects. Here we see that the shift towards more opinion statements in the post-*Omnicare* period identified in Figure 6 plays a negligible role in explaining the shift in underpricing after the opinion. As in the baseline specification, the composition effect from fact-intensity (blue line) remains negligible throughout the distribution. Similarly, the composition effect from the other covariates (green line) is, if anything, positive, indicating that the characteristics of post-*Omnicare* IPOs—larger firms, different deal sizes, varying market conditions—would have reduced underpricing under pre-*Omnicare* pricing.

In other words, the composition of the IPO market shifted toward firm types that historically experienced less underpricing. However, this favorable compositional shift is more than offset by the negative structure effect. The market's repricing of IPO characteristics after *Omnicare* swamps any compositional advantage, resulting in a net increase in underpricing. Changes in disclosure practices contribute little to the observed shift; the story is one of changed pricing, not changed disclosure.

5 Discussion and Conclusion

We study the relationship between legal liability and the phrasing firms use in their annual reports and IPO registration statements. Stating risks with more certainty increases the likelihood that those statements will form the basis of securities class actions. That stronger phrasing also makes those statements more believable. Using a large language model to classify sentences in firm disclosures as either fact or opinion, we measure how much litigation risk those disclosures pose.

For statements in firms' annual reports, we show that increasing the factual content of Risk Factors has countervailing effects. That increase produces positive abnormal returns in the following days and weeks, but it also heightens the risk that the firm will be sued in a securities class action in the next two years. Our analysis of IPO registration statements shows that adjusting the liability threshold for opinions and facts affects how markets respond to those statements. When the Supreme Court increased the potential for opinion-based liability in the circuits that produce the most IPOs, we find that market actors distinguished less between statements of opinion and fact when pricing IPOs. Together, these results show that securities law is playing its intended role. By imposing liability for more certain statements, those laws influence the credibility of firms' disclosures.

References

Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

Beatty, R. P. and Ritter, J. R. (1986). Investment banking, reputation, and the underpricing of initial public offerings. *Journal of Financial Economics*, 15(1–2):213–232.

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

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4):436–455.

Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H.-m., and Steele, L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, 19(1):396–455.

Cattaneo, M. D., Crump, R. K., Farrell, M. H., and Feng, Y. (2024). On binscatter. *American Economic Review*, 114(5):1488–1514.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268.

Cohen, L., Malloy, C., and Nguyen, Q. (2020). Lazy prices. *Journal of Finance*, 75(3):1371–1415.

Delgado, M. A., García-Suaza, A., and Sant'Anna, P. H. C. (2022). Distribution regression in duration analysis: An application to unemployment spells. *The Econometrics Journal*, 25(3):675–698.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 4171–4186.

Diamond, D. W. and Verrecchia, R. E. (1991). Disclosure, liquidity, and the cost of capital. *Journal of Finance*, 46(4):1325–1359.

Ewens, M., Xiao, K., and Xu, T. (2024). Regulatory costs of being public: Evidence from bunching estimation. *Journal of Financial Economics*, 153:103775.

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

Healy, P. M. and Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1–3):405–440.

Heinle, M. S. and Smith, K. C. (2017). A theory of risk disclosure. *Review of Accounting Studies*, 22(4):1459–1491.

Hope, O.-K., Hu, D., and Lu, H. (2016). The benefits of specific risk-factor disclosures. *Review of Accounting Studies*, 21(4):1005–1045.

Kravet, T. D. and Muslu, V. (2013). Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies*, 18(4):1088–1122.

Ljungqvist, A. (2007). IPO underpricing. In Eckbo, B. E., editor, *Handbook of Corporate Finance: Empirical Corporate Finance*, pages 375–422. Elsevier/North-Holland, Amsterdam.

Loughran, T. and McDonald, B. (2013). Ipo first-day returns, offer price revisions, volatility, and form s-1 language. *Journal of Financial Economics*, 109(2):307–326.

Loughran, T. and Ritter, J. R. (2004). Why has IPO underpricing changed over time? *Financial Management*, 33(4):5–37.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

Ritter, J. R. (2025). IPO data. University of Florida, `https://site.warrington.ufl.edu/ritter/ipo-data/`. Accessed 2025.

Ritter, J. R. and Welch, I. (2002). A review of IPO activity, pricing, and allocations. *Journal of Finance*, 57(4):1795–1828.

Rock, K. (1986). Why new issues are underpriced. *Journal of Financial Economics*, 15(1–2):187–212.

Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics*, 32(1–3):97–180.