Boilerplate and the Impact of Disclosure in Securities Dealmaking

Jeremy McClane*

Capital markets dealmaking, like many kinds of business transactions, is built on a foundation of copied and recycled language—what many call boilerplate. Regulators and the bar periodically call for less reliance on boilerplate, but despite these pressures, boilerplate remains a fixture of ever-growing securities disclosures. This Article explores why boilerplate persists and how it affects investors, showing that boilerplate may have a more complex role than commonly recognized. This Article does so by developing a theory on the effect of boilerplate in securities disclosure—a context that is little studied despite a wealth of literature on boilerplate in other settings—and analyzes disclosure empirically using language processing techniques on a dataset of initial public offering disclosure spanning twenty years, from 1996 to 2015. The data shows that in the aggregate, the use of boilerplate is associated with some efficiency gains. For example, 10% more boilerplate in IPO disclosure is associated with a savings of $65,000 in legal fees, on average, controlling for other relevant factors. But the measurable gains are generally outweighed by boilerplate's information-related costs: greater use of boilerplate is associated with several indicia of information asymmetry that see issuing firms give up as much as $5 to $6 million in the market on average for each additional 10% of their disclosure that consists of rote recitations. Greater use of generic boilerplate language is also related to greater incidence of securities litigation and is associated with lower readability of already complex registration statements. The evidence points to the conclusion

that, whether through its content or its signaling effect, boilerplate disclosure in the aggregate represents greater costs for IPO issuers and does little to advance the goal of better informing the investing public.

    In addition to discussing implications for law and policy, this Article addresses a puzzle raised by the data: Why do securities issuers continue to use boilerplate when it has the potential to lose them money, draw litigation, and buck regulatory pressure? Theory developed in legal scholarship provides a number of possible answers. The explanation most consistent with these findings is that boilerplate serves as a substitute for information production, meaning that issuers can obscure sensitive information or shortcut due diligence if they are willing to pay the price for doing so.

INTRODUCTION

      Lawyers entering corporate securities practice are sometimes surprised to discover that, of all the things they have learned through years of education and legal training, the skill they employ most is the ability to cut and paste. This may have caused some to wonder how much value this seemingly commodified work adds for clients and whether the lawyers' specialized skills are being put to their best use. This Article begins to answer those questions by providing a theoretical and empirical analysis of boilerplate in securities disclosure.

      The theory and evidence point to answers that are more nuanced than intuition would suggest. As this Article shows, there is evidence that incorporating boilerplate into securities disclosure does add some value by reducing transaction costs. But there is also evidence that, in the aggregate, using boilerplate comes with its own costs that are often greater than scholars have previously recognized. Nonetheless, the analysis suggests that some types of boilerplate language may enhance

communication by its presence or signaling effect, at least with respect to sophisticated investors.

Although legal scholars have written dozens of articles about boilerplate in contracts,[1] as well as in corporate and sovereign bonds[2] (which in many ways function like contracts), to date there has been little attention given to boilerplate in securities disclosure.[3] This is a significant omission because securities disclosure drives the multitrillion-dollar securities market in the United States,[4] and boilerplate in that market plays an important but distinct role than the role it plays in contracts or bond covenants. Whereas contract terms and bond covenants generally serve to define the rights and duties of different parties involved in the course of a transaction, securities disclosure is a legally mandated mechanism for providing information about companies raising capital—an objective that reduces information asymmetries and facilitates the creation of reliable securities markets. In the former context, boilerplate may provide an efficient stand-in for terms that have been negotiated countless times in the past. In the securities context, boilerplate represents essentially identical

---

1.   *See, e.g.*, Douglas G. Baird, *The Boilerplate Puzzle*, 104 MICH. L. REV. 933, 949–50 (2006) (discussing the application of contract law to boilerplate in commercial transactions); Omri Ben-Shahar, *Regulation Through Boilerplate: An Apologia*, 112 MICH. L. REV. 883, 895–96 (2014) (reviewing BOILERPLATE: MARGARET JANE RADIN, THE FINE PRINT, VANISHING RIGHTS, AND THE RULE OF LAW (2013)) (arguing that consumers benefit from boilerplate terms through better price or quality and providing examples of such tradeoffs); Margaret Jane Radin, *Response: Boilerplate in Theory and Practice*, 54 CAN. BUS. L.J. 292, 298–99 (2013) (arguing that many boilerplate contract provisions should be subject to greater judicial scrutiny).

2.   *See, e.g.*, Stephen J. Choi & G. Mitu Gulati, *Innovation in Boilerplate Contracts: An Empirical Examination of Sovereign Bonds*, 53 EMORY L.J. 929 (2004) (analyzing boilerplate language in sovereign bond terms); Marcel Kahan & Michael Klausner, *Standardization and Innovation in Corporate Contracting*, 83 VA. L. REV. 713 (1997) (discussing boilerplate's role in corporate bond terms).

3.   One notable article that touches on disclosure boilerplate is Karen K. Nelson & A.C. Pritchard, *Carrot or Stick? The Shift from Voluntary to Mandatory Disclosure of Risk Factors*, 13 J. EMPIRICAL LEGAL STUD. 266 (2016). Although their interesting analysis does not itself focus on boilerplate, one aspect of it looks at how securities issuers copy their own past disclosures in periodic reports. *See id.* at 273 (hypothesizing that firms at greater risk of litigation provide less boilerplate risk-factor disclosure). This Article looks at boilerplate in a different sense. It focuses broadly on boilerplate as generic language that is copied in different deals among different issuers, which is a perennial concern among regulators, courts, and practitioners.

4.   *See* SEC. INDUS. & FIN. MKTS. ASS'N, CAPITAL MARKETS REPORT—MODERNIZING AND RATIONALIZING REGULATION OF THE U.S. CAPITAL MARKETS 4 (2017), https://www.sifma.org/wp-content/uploads/2017/08/Capital-Markets-Report-%E2%80%93-Modernizing-and-Rationalizing-Regulation-of-the-U.S.-Capital-Markets.pdf [https://perma.cc/C3EQ-SA4M] ("As of December 31, 2016, market capitalization—the total value of all publicly traded domestic companies—of the U.S. stock market was $27.4 trillion . . . ."); *see also Report for Selected Countries and Subjects*, INT'L MONETARY FUND, http://www.imf.org/external/pubs/ft/weo/2017/01/weodata/weorept.aspx?pr.x= 57&pr.y=7&sy=2016&ey=2020&scsm=1&ssd=1&sort=country&ds=.&br=1&c=111&s=NGDPD% 2CNGDPDPC%2CPPPGDP%2CPPPPC&grp=0&a= (last visited Sept. 15, 2018) [https://perma.cc/ N3VR-FX8Q] (reporting that the United States' GDP in 2016 was only $18.5 trillion).

disclosures about inherently different companies that may or may not be accurately described in identical terms.

Whether securities boilerplate is problematic has no obvious a priori theoretical answer, but the issue is not merely academic. The Securities Exchange Commission ("SEC") has tried a number of times to limit boilerplate recitations in securities disclosures and, in December 2015, Congress took aim at boilerplate by mandating that the SEC revise its disclosure regulations to eliminate such language as much as possible.[5] The stated purpose for targeting boilerplate language has been to reduce costs and increase efficiency both for the companies that must produce disclosure and for the investors who consume it.[6] The premise is that boilerplate recitations provide little useful information and are difficult for investors to wade through. That rationale, however, conflicts with what some scholars have argued about boilerplate in other contexts, namely that such language has the power to efficiently convey information because over time it becomes a standardized language that is readily understood by market initiates.[7] The idea underlying these arguments is that boilerplate is a type of modular language—a settled formulation of a set of ideas or information—that can be easily recognized and understood by its intended audience.

One reason for these divergent views is that the word "boilerplate" itself, while referring generally to standardized, recyclable language, can actually encompass different kinds of language with different uses depending on its purpose. Boilerplate is sometimes used to refer to standardized legal language, like disclaimers or choice of law

---

5. The mandate was buried in a bill primarily aimed at fixing the United State's highway infrastructure. *See* Fixing America's Surface Transportation Act of 2015 (FAST Act), Pub. L. No. 114-94, § 72002, 129 Stat. 1312, 1784–85 (2015) ("[T]he Securities and Exchange Commission shall take all such actions to revise regulation S–K . . . to eliminate provisions of regulation S–K, required for all issuers, that are duplicative, overlapping, outdated, or unnecessary . . . ."); *see also* Business and Financial Disclosure Required by Regulation S-K, 81 Fed. Reg. 23,915 (Apr. 22, 2016) (setting out the SEC's analysis of the disclosure issues in the FAST Act and the process for gathering comments).

6. *See* Business and Financial Disclosure Required by Regulation S-K, 81 Fed. Reg. at 23,917 ("We are specifically seeking comment on . . . whether, and if so how, we could revise our requirements to . . . promote efficiency, competition, and capital formation . . . ."). To be sure, such regulation encompasses more than just copied language. However, as discussed below, copied language is a major component of what is typically defined as boilerplate, and what regulators target.

7. *See, e.g.*, Baird, *supra* note 1, at 936 (arguing that boilerplate in consumer contracts provides information when sophisticated consumers select products on the basis of boilerplate terms); Kahan & Klausner, *supra* note 2, at 718 (discussing how common use of terms in corporate bond contracts create learning and networking externalities); Henry E. Smith, *Modularity in Contracts: Boilerplate and Information Flow*, 104 MICH. L REV. 1175, 1176–77 (2006) (arguing that boilerplate gives rise to modular contractual terms that can be readily understood and transplanted to a wide variety of contract contexts, albeit at the cost of customization).

provisions, that serve a similar purpose and might be included in a document to fulfill a legal mandate, out of an abundance of caution, or because there are only so many ways to say certain things. Such language might be full of legal jargon or other terms that are poorly understood even by their drafters but are included out of habit or caution, and no one has bothered to change them or think them through in the course of many transactions. Boilerplate can also refer to clauses that, while not legally mandated or necessary, have been proven through experience to convey certain ideas or terms in a way that is predictable and reliable. Boilerplate might also be a means to be intentionally generic or vague, either because specifics are unknown, because specifics are better left for a later time, or because an issuer desires to be vague and blend in with other disclosing entities.

This Article analyzes these possibilities, providing a systematic, theoretical, and empirical analyses of boilerplate in securities disclosure. In doing so, it contributes to the policy and scholarly debates. To undertake this analysis, I draw insights from the contract and bond boilerplate literature and show how these literatures help frame the analysis of securities boilerplate, even though they are different in important ways. I then explain the results of an empirical examination of boilerplate in Initial Public Offering ("IPO") documents, using computerized natural language processing methods common in social science and studies of literature to measure boilerplate and assess its effects on issuers and investors. I use these methods to analyze an original dataset of 2,751 IPOs for operating companies taking place between 1996 and 2015.[8] The IPOs are confined to those for private companies going public for the first time, as opposed to spinoffs, securitization vehicles, investment companies, or real estate investment trusts. This provides a good setting for the analysis because little company information is typically available to the public for such companies prior to the filing of the registration statement, and so a company's legally mandated disclosure document—the locus of the deal's boilerplate—is often a critical means by which issuers shape investors' perceptions.

The results of the analysis shed light on boilerplate's potential value as well as its risks. A securities class action filed against Wayfair.com provides an illustration.[9] The lawsuit alleged that when

---

8.   The study period begins in 1996, in part because that is the first year for which disclosures are available on the SEC's electronic filing system, EDGAR. The study covers the adoption of the USA JOBS Act in 2012, which introduced confidential filing provisions for certain IPO issuers. *See* Jumpstart Our Business Startups Act § 106(a), 15 U.S.C. § 77f(e) (2012).

9.   *See* Dingee v. Wayfair Inc., No. 15cv6941(DLC), 2016 WL 3017401 (S.D.N.Y. May 24, 2016) (granting the defendants' motion to dismiss).

Wayfair, an online retailer specializing in home goods and furniture, went public, its prospectus misled investors by forgoing a specific disclosure in favor of a boilerplate risk factor about its competitors: "Our business is rapidly evolving and intensely competitive, and we have many competitors in different industries. Our competition includes: furniture stores, big box retailers, department stores, specialty retailers, and online home goods retailers and marketplaces . . . ."[10]

Absent from this disclosure was any mention of a specific competitor, Overstock.com, whose similar business model posed a serious competitive threat to Wayfair.[11] When securities analysts finally noticed the omission several months later, Wayfair stock fell and the investors lost money.[12] This risk factor was not the only generic disclosure Wayfair used: fifty percent of the risk factors in its IPO prospectus included language copied from other companies' recent deals, according to the measure used in this study. By comparison, the average amount of risk factor boilerplate across all prospectuses in the dataset is thirty-two percent.

This example highlights the problems raised by securities boilerplate. On the one hand, it is easy to see why efficiency might have prompted Wayfair's counsel—a busy, national law firm—to use ready-made language if experience told them that it was sufficient to convey the risk of competition to investors. On the other hand, one can imagine how a prospectus that is nearly eighty percent boilerplate might gloss over important issues faced by a relatively unknown company. The data shows that both views may have merit given that boilerplate is associated with lower legal costs on average, but is also associated with higher average losses to issuers from mispricing and more securities fraud litigation. The analysis in this Article explores the data to help understand how the effects of boilerplate balance out and how issuers and lawyers might use it.

In order to study boilerplate, I first develop a way of defining it and measuring it. Using that measure, I then examine observational data for evidence about the relative effects of more and less boilerplate. First, I find that more boilerplate is associated with lower legal costs, but find no evidence that it is associated with lower auditing fees or underwriting fees. Moreover, I find no significant association between

---

10. *Id.* at *2 (emphasis omitted).

11. *Id.* ("The [First Amended Complaint] alleges that the Registration Statement deliberately omits naming Overstock as Wayfair's 'prime competitor.' ").

12. *Id.* ("On the day that the [analyst firm's] Report was published, shares of Wayfair fell . . . over [eleven percent] . . . .").

boilerplate and faster deal-completion times, the average amount a prospectus is amended, or the scrutiny it receives from the SEC.

However, I also find evidence that large quantities of boilerplate are associated with more information asymmetry, which costs issuers amounts far outweighing any savings in fees, on average. Specifically, a 10% increase in boilerplate in certain important sections of a registration statement is associated with as much as a 5.1% to 6.2% increase in deal underpricing—a phenomenon by which IPO's are sold at prices below what the market will bear, and which is thought to be in part a product of information asymmetry.[13] This translates, on average, to as much as $6 million that an issuer leaves on the table for each 10% increase in the use of boilerplate in its disclosure. Higher levels of boilerplate are also associated with higher risk of prospectus-related litigation: a 10% increase in the amount of boilerplate in the some sections of the prospectus is associated with a 1.5% to 4% increase in the probability of being sued for securities fraud related to the offering.[14] The analyses also show a relationship between boilerplate and three other indicia of information asymmetry: greater probability of pre-IPO price revision, wider first-day bid-ask trading spreads, and greater divergence of recommendations among analysts following the issuer. To conduct this analysis, I rely primarily on regression analysis to assess the basic relationship between boilerplate and the various outcomes I study and on propensity score matching to attempt to rule out the possibility that the informational outcomes may be driven by underlying features of each issuer or by the transactions that initially prompt the use of boilerplate. Although causality cannot be definitively inferred from the analysis in this Article, the results taken together provide strong evidence of the consequences of including too much boilerplate in securities disclosure.

In sum, the data shows that boilerplate has some value, but in the aggregate it is associated with deal outcomes, indicating that neither the issuer nor the investing public are well served. The Wayfair example above bears this out. If it is true that neither issuers nor investors benefit from boilerplate, one might wonder why sophisticated law firms and well-counseled issuers continue to use so much of it in

---

13.   For an explanation of the underpricing phenomenon and the theories about its causes, see Randolph P. Beatty & Jay R. Ritter, *Investment Banking, Reputation, and The Underpricing of Initial Public Offerings*, 15 J. FIN. ECON. 213 (1986); Tim Loughran & Jay Ritter, *Why Has IPO Underpricing Changed over Time?*, 33 FIN. MGMT. 5 (2004); and Jay R. Ritter & Ivo Welch, *A Review of IPO Activity, Pricing, and Allocations*, 57 J. FIN. 1795 (2002).

14.   In addition, boilerplate is associated with indicia of greater investor uncertainty about an issuer, as well as lower readability scores, measured by the Gunning Fog readability index—in fact, prospectuses containing large amounts of boilerplate have scores indicating that they are essentially unreadable by most human beings.

securities disclosure. One possible explanation is that issuers benefit from boilerplate disclosure and information asymmetry in some way that is hard to measure. After all, in the Wayfair example, the company's prospectus had a much higher level of boilerplate than most of the other deals done by their law firm: the firm's average risk factor boilerplate across all deals since 2010 is approximately thirty-five percent—slightly higher than average, but still much lower than the amount used for Wayfair. Moreover, the law firm's fees showed no evidence that Wayfair saved money relative to similar deals, since the fees were in line with the average amount the firm received for similar deals in the dataset. Perhaps the deal was sui generis for the firm, or perhaps the issuer derived a benefit from using a vague disclosure that is not readily observable in the data. But the most consistent story is that the market ended up with less information than it should have had, and Wayfair ended up paying a price for it.

The high-level empirical conclusion, however, cannot answer whether the overall societal effects of securities boilerplate are negative, or whether boilerplate simply shifts value from one party to another. The theory on boilerplate developed in other contexts provides some possibilities. One possibility consistent with the data is that boilerplate provides a signal to investors that the burden of due diligence will be placed largely on them. Thus, the costs of research are transferred from the banking syndicate and the issuer to the investors. As in any bargain, those investors will expect something in return, and that reality will be reflected in the price at which an issuer goes public. In other words, issuers will leave more money on the table in exchange for leaving the transaction costs of mitigating information asymmetry to investors.

Another possibility is that the inclusion of boilerplate is a result of a network externality. In that case, the loss to the issuer (and possibly to investors as well) is the result of a human tendency to favor what has been done before and to exhibit reluctance to expend resources on creating a new template when the benefit of doing so will largely be captured by others. The data are least consistent with the theory that boilerplate is used primarily for its efficiency, at least in the aggregate sense.

Notwithstanding the analysis with respect to aggregate boilerplate, it is still possible that certain individual boilerplate phrases could provide value by conveying information and meeting regulatory requirements more easily. There is some evidence in the data to support this. Nonetheless, even if that is the case, it appears that boilerplate phrases that are individually efficient may still cloud information transmission if too much other boilerplate is used. Some simple changes

to regulation could help address both types of boilerplate and help declutter securities disclosures without losing whatever benefits such boilerplate might have.

The remainder of this Article proceeds as follows. Part I provides definitional background, as well as a general description of IPO deals and how boilerplate is produced for them. It also describes the SEC's disclosure regime and its efforts to regulate boilerplate, as well as the academic discussion of boilerplate to date. Part II explains the methods used in the empirical portion of the paper and discusses the various means of analyzing boilerplate. Part III discusses the results of the analysis, and Part IV discusses the implications of the analysis for the law, legal scholarship, and the SEC's reform efforts.

I. BACKGROUND

Some background on the process by which securities disclosure is produced is important both for formulating an accurate definition of boilerplate and for appreciating how the use of boilerplate might affect a transaction through pricing and other measurable outcomes. In brief, when a company decides to conduct an IPO, the issuing company, often unfamiliar with the norms and practices of securities markets, gathers a group of advisors, including its law firm, its auditor, the investment banks that will underwrite the deal, and the investment banks' lawyers.[15] The issuer's counsel usually takes the lead drafting the disclosure document with help from the bankers and their lawyers. As the lawyers and bankers conduct due diligence on the issuer, the parties meet in drafting sessions to refine the disclosure.[16] The issuer's lawyers rarely start from scratch to draft the disclosure; as in many business transactions, they pull text from the disclosures issued in precedent deals, usually those that were completed recently for other companies in the issuer's industry group, those with which they are familiar, or those suggested by the underwriting banks.[17] The process, as detailed below, is intended to comply with the law as well as market practice.

---

15. *See* Carl W. Schneider, Joseph M. Manko & Robert S. Kant, *Going Public: Practice, Procedure, and Consequences*, 27 VILL. L. REV. 1 (1981) (describing the tasks performed by bankers, lawyers, and issuing-company management when conducting an IPO).

16. *See id.* at 17–19 (discussing the role of company counsel in collecting due diligence and drafting the registration statement).

17. *See id.*

A. Drafting IPO Disclosures

The types of disclosure a company must make when seeking to issue securities are set out in the Securities Act of 1933 (the "Securities Act") and the regulations promulgated by the SEC.[18] The primary goal of the disclosure rules is to reduce information asymmetry that exists between a company entering the public securities market and potential investors.[19] The information asymmetry exists because the people who manage issuing companies have better information about the company's past and future performance than investors are presumed to be able to obtain on their own.[20] Company insiders—such as its officer and directors—may have incentives to reinforce this asymmetry by selectively disclosing certain information about their company to entice investors, highlighting positive information and obscuring negative information; they may also be reluctant to reveal sensitive information that might benefit their competitors.[21] The law thus requires companies to disclose certain kinds of information, and creates incentives for the investment banks and law firms advising the issuer to conduct due diligence, gather information, ensure its accuracy, and adequately communicate it in the prospectus accompanying the registration statement, the primary disclosure document required by the Securities Act.[22]

---

18. Securities Act of 1933, 15 U.S.C. §§ 77a–77aa (2012). Regulation S-K is the set of rules that primarily implements the disclosure requirements of the Securities Act of 1933. 17 C.F.R. §§ 229.1–229.1200 (2018). The Securities Act of 1933 adopts a regime of full disclosure to protect investors, as opposed to a regime regulating the merits of any particular investment (as the Food and Drug Administration ("FDA") does with new medicines, for example), on the theory that "[p]ublicity is justly commended as a remedy for social and industrial diseases. Sunlight is said to be the best of disinfectants; electric light the most efficient policeman." LOUIS D. BRANDEIS, OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT 92 (1914); *see also* Hersh Shefrin & Meir Statman, 49 FIN. ANALYSTS J. 21, 25 (1993) ("[M]andatory disclosure regulations seek to protect investors from insufficient and misleading information, rather than to protect or prevent them from choosing securities lacking merit.").

19. *See, e.g.*, *In re* Initial Pub. Offering Sec. Litig., 358 F. Supp. 2d 189, 205 (S.D.N.Y. 2004) ("The Securities Act of 1933 . . . was designed to provide investors with full disclosure of material information concerning public offerings of securities . . . ." (first alteration in original) (quoting Ernst & Ernst v. Hochfelder, 425 U.S. 185, 195 (1976))).

20. *See* Robert A. Prentice, *The Inevitability of a Strong SEC*, 91 CORNELL L. REV. 775, 804 (2006) ("Even sophisticated investors are not skilled at obtaining private information possessed by firms."). However, for an argument that the law should regulate on the basis of such bargaining between issuers and investors, see Stephen Choi, *Regulating Investors Not Issuers: A Market-Based Proposal*, 88 CALIF. L. REV. 279, 283 (2000).

21. *See* Merritt B. Fox, *Shelf Registration, Integrated Disclosure, and Underwriter Due Diligence: An Economic Analysis*, 70 VA. L. REV. 1005, 1019 (1984) ("Management['s] best interests . . . are sometimes different from those of the shareholders.").

22. *Cf. id.* (proposing additional due diligence requirements for Exchange Act filings to increase efficiency). The terms "registration statement" and "prospectus" have different legal meanings under the Securities Act. The prospectus forms part of the registration statement,

The prospectus has two main purposes that are sometimes in tension. On the one hand, the prospectus serves to comply with the disclosure requirements mandated by the law and the SEC[23] and limit the company's liability for material misrepresentations.[24] Achieving this purpose often involves a balancing act between providing specific and detailed information about a company, while cautiously incorporating caveats, and using language that has been tested by the market, the SEC, and the courts.[25] On the other hand, the prospectus forms the basis of the marketing effort that the company and the investment bankers undertake to sell the stock.[26] For that purpose, the company and the underwriting banks want to highlight the company's potential and avoid negative language that would undermine management's confident narrative regarding the company's future performance.[27] Lawyers on both sides of the deal—those representing the issuer and those representing the group of investment banks underwriting the transaction—perform most of the work of drafting the prospectus.[28] And although the content of the disclosure is mandated by the SEC, the issuer has the final say on what goes into the document,

although industry participants often simply refer to the prospectus. For ease of reference, I adopt this convention and refer to the disclosure as the prospectus here.

23. *See* 17 C.F.R. §§ 229.10–229.915 (2018) (providing disclosure requirements). Required disclosure includes (1) information about the company's business, *see* §§ 229.101–229.103; (2) the management's discussion and analysis of financial condition of the company, including future projections if desired, *see* § 229.303; (3) financial statements and an auditor's opinion covering them, *see* § 210; (4) a description of material contracts, *see* §§ 229.10–229.915; (5) information about legal and regulatory problems facing the company, *see* § 229.103; (6) information about the officers and directors of the company and their compensation, *see* §§ 229.403–229.405; and (7) certain industry-specific information, *see Industry Guides*, U.S. SEC. & EXCH. COMM'N, https://www.sec.gov/about/forms/industryguides.pdf (last visited Sept. 15, 2018) [https://perma.cc/YN29-Y9R8].

24. The company's advisors can shield themselves from liability under Section 11 of the Securities Act of 1933, as long as they have conducted adequate due diligence. *See* Securities Act of 1933, § 11(b)(3), 15 U.S.C. § 77k(b)(3) (2012); *see also* Escott v. BarChris Constr. Corp., 283 F. Supp. 643, 688–89 (S.D.N.Y. 1968) (establishing the terms of the due diligence defense for nonissuer defendants in prospectus-related litigation if the defendant can show reasonable grounds for that belief after a reasonable investigation into the truth of the alleged misstatements).

25. *See* SEC. INDUS. ASS'N, CAPITAL MARKETS HANDBOOK 272 (John C. Burch, Jr. & Bruce S. Foerster eds., 2006) (reviewing the IPO process); Schneider et al., *supra* note 15, at 10 ("[T]he prospectus . . . is prepared as a brochure describing the company and the securities to be offered.").

26. *See* SEC. INDUS. ASS'N, *supra* note 25, at 272.

27. *See* Schneider et al., *supra* note 15, at 14 (describing the tensions that sometimes arise between the desires of underwriters and counsel and the desires of issuers); *see also* STEVEN E. BOCHNER, JON C. AVINA & CALISE Y. CHENG, WILSON SONSINI GOODRICH & ROSATI, GUIDE TO THE INITIAL PUBLIC OFFERING 29 (8th ed. 2016), https://www.wsgr.com/publications/PDFSearch/IPOGuide2016.pdf [https://perma.cc/K9KN-TVPD] (stating that the issuer's counsel and underwriters must seek to serve the marketing and disclosure functions of the prospectus).

28. *See* Schneider et al., *supra* note 15, at 14–17 (discussing the role of company and underwriters' counsel in the drafting of disclosure).

and thus exercises discretion over what gets disclosed and in what level of detail.[29] The issuer's management typically relies on its law firms and bankers to advise it on the right level of disclosure, although it can decide how much deference to give these advisors.

In addition to working for the issuer, the lawyers and investment bankers are presumed to perform a gatekeeping function on behalf of investors.[30] In theory, since no one is at the table to represent investors when the disclosure is being drafted, the bankers and lawyers consider what information the investors need and how best to communicate it, which might be at odds with what the issuers themselves want.[31] Thus, the lawyers and bankers must negotiate with the issuer's management over what to disclose, how it will be worded (vaguely or in detail), and how to manage ambiguities in the regulations.[32] Boilerplate language serves as an opening salvo in this negotiation—a basis for filling out the draft and possibly helping the lawyers and bankers anchor the prospectus around their preferred wording. The prospectus, which begins as language copied from the text of prior transactions, is thus revised iteratively in meetings involving both sets of counsel, the underwriters, and representatives from the company. Moreover, as more due diligence is done, more information comes to light and, assuming the issuer agrees, the text becomes more specific.[33]

Once the prospectus is drafted, the issuer files the preliminary version of it (referred to as the "Red Herring") with the SEC as part of

---

29.   *See id.* at 18 ("In the last analysis, the company and its management must assume the final responsibility to determine that the information in the registration statement is accurate and complete."); *see also* Tom Arnold, Raymond P.H. Fishe & David North, *The Effects of Ambiguous Information on Initial and Subsequent IPO Returns*, 39 FIN. MGMT. 1497, 1500 (2010) (describing the issuing management's control over the message conveyed in the disclosure).

30.   *See* Schneider et al., *supra* note 15, at 14 (noting the potential conflict that may arise between the desires of underwriters' and issuer's counsel and the desires of issuers).

31.   *Id.*

32.   *See* Arnold et al., *supra* note 29, at 1500 (discussing how the Risk Factors section of Regulation S-K forces companies to make judgments about what to disclose and the wording used when disclosing); *see also* Kathleen Weiss Hanley & Gerard Hoberg, *The Information Content of IPO Prospectuses*, 23 REV. FIN. STUD. 2821, 2825 (2010) (discussing the issuer's role in producing information for the prospectus).

33.   Schneider et al., *supra* note 15, at 23 ("This drafting is an iterative process, as knowledge gained in due diligence informs what needs to be said about the issuer."); *see also* Royce R. Barondes & Gary C. Sanger, Lawyer Experience and IPO Pricing 8 (May 4, 2000) (unpublished manuscript),   https://papers.ssrn.com/sol3/papers.cfm?abstract_id=227729   [https://perma.cc/ 2YSK-JBX3]:

> Lawyers can cause disclosure to be more negative in a number of ways. They can uncover negative information in the due diligence process that everyone agrees needs to be disclosed. Additionally, they can take a more hard line in the drafting of the prospectus, compelling disclosure of information the business thought could remain undisclosed, or requiring language that casts information in a more negative light.

the registration statement on Form S-1.[34] The preliminary prospectus includes a price range at which the issuer and its underwriters expect to market the stock based on information learned in the due diligence review.[35] The SEC reviews the preliminary prospectus, and gives comments and requests clarifications or additions that the issuer must address.[36]

B. The Prospectus and IPO Pricing

As previously mentioned, one of the functions of a prospectus is to provide a basis of information for marketing the company, and the success of the marketing effort helps set the price at which the company's stock will debut.[37] While the SEC is reviewing the preliminary prospectus, the underwriter and issuer's management begin marketing the stock primarily, and in some cases exclusively, to large institutional investors—such as pension funds, mutual funds, and hedge funds—that will be the initial investors in the securities.[38] The lead underwriter and the issuer's management pitch the securities to institutional investors in various cities by presenting the company's story and prospects consistent with the information in the preliminary prospectus.[39] During this process, known as a road show, the initial investors are able to gather additional information not necessarily reflected in the prospectus by speaking to the company's management and the bankers more or less privately.[40] After these presentations, the initial investors express their level of interest in investing by making indicative orders, which are nonbinding but which the investors and bankers are usually expected to honor.[41] In this way, these initial investors play an indirect role in determining the final price for the transaction, because the underwriters use the indicative orders to

---

34.   The most common form of registration statement used for IPOs is Form S-1, although Form SB-1 was available for smaller issuers until the repeal of the small business disclosure rules in 2005. For simplicity, I will refer to all of these documents as S-1 or Form S-1.

35.   *See* Schneider et al., *supra* note 15, at 22–23 (describing the distribution of "red herring" prospectuses).

36.   *See id.* at 19–22 (discussing the SEC comment and review process); *see also* William W. Barker, *SEC Registration of Public Offerings Under the Securities Act of 1933*, 52 BUS. LAW. 65, 70–72 (1996) (describing the SEC staff's role in the registration and disclosure process).

37.   *See* Manuel A. Utset, *Producing Information: Initial Public Offerings, Production Costs, and the Producing Lawyer*, 74 OR. L. REV. 275, 284 (1995) (describing the lawyers' job in an IPO to be the production of an "information bundle").

38.   *See* Schneider et al., *supra* note 16, at 22–24 (discussing "red herring[s]" and "dog and pony show[s]").

39.   *See id.*

40.   *See id.*

41.   *See id.*

assess investor demand.[42] The lead underwriter and the issuing company's management negotiate the final offering price based largely on the investor demand ascertained during the road show.[43] Once the final price is negotiated, a final version of the prospectus and final pricing information are filed with the SEC and, once deemed effective,[44] the shares can be sold to the wider market and the company officially goes public.[45]

C. Disclosure Regulation and the Boilerplate Problem

As the discussion above highlights, language borrowed from prior transactions is a ubiquitous feature of IPO prospectuses (and securities disclosure more broadly), and the SEC has taken note. For many years, the regulator has taken steps to discourage too much copying and reliance on generic language in its attempt to promote clearer securities disclosure more broadly.

For instance, the SEC created a Task Force on Disclosure Simplification in 1995 and charged it with improving and streamlining disclosure.[46] The stated rationale for creating the task force was the perception that IPO prospectuses had become overly dense due to the

---

42. *See id.*; *see also* JOHN C. COFFEE, JR. & HILLARY A. SALE, SECURITIES REGULATION 115, 122–25 (11th ed. 2009) (discussing the book-building process); Barondes & Sanger, *supra* note 34, at 7 (describing how the offering process is decided and the use of the initial filing price range as a proxy for the estimate developed during the "beauty pageant").

43. *See* Barondes & Sanger, *supra* note 33, at 8:

> In a customary IPO, there is not a definitive agreement on the price at which the underwriters will purchase and resell the stock in the IPO until after the preliminary marketing process is complete, some time after a preliminary prospectus has been circulated. SEC rules, however, require that a preliminary prospectus for an IPO circulated prior to the pricing include a bona fide estimate of the price, frequently stated as a range, at which the stock will be sold . . . . This price estimate may change in subsequent preliminary prospectuses, as the lead managerd acquires information during the marketing process.

44. *See* 17 C.F.R. §§ 230.424(b), 230.430A (2018) (regulating the filing of prospectuses and the prospectus in a registration statement at the time of effectiveness); *see also* COFFEE & SALE, *supra* note 42, at 128–29 (discussing Rule 430A). Prior to the adoption of Rule 430A, underwriters had to file pricing information in the form of an amendment to the registration statement before the SEC would declare the registration statement effective. *See* COFFEE & SALE, *supra* note 42, at 128–29. Rule 430A allows the registration statement to be declared effective before pricing-related information is filed as long as a complete final prospectus is filed shortly thereafter. *See* 17 C.F.R. § 230.430A (regulating the prospectus in a registration statement at the time of effectiveness).

45. *See* COFFEE & SALE, *supra* note 42, at 128–9 (discussing the procedure surrounding effectiveness).

46. *See Report of the Task Force on Disclosure Simplification*, U.S. SEC. & EXCH. COMM'N (Mar. 5, 1996), https://www.sec.gov/news/studies/smpl.htm [https://perma.cc/D539-3M66] ("The Task Force was asked to review rules and forms affecting capital formation, with a view toward streamlining, simplifying, and modernizing the overall regulatory scheme . . . .").

incorporation of large amounts of nonspecific boilerplate language.[47] The SEC believed that some securities lawyers had come to view prospectuses less as sources of information for issuers and more as forms of insurance against liability for both issuers and underwriters.[48] This trend was propelled by the fact that it is more difficult to find issuers liable for events or circumstances that are adequately disclosed; therefore, including greater quantities of disclosure that had already been vetted by the SEC and tested in the market became a rational approach to drafting.[49] The SEC attempted to curb generic boilerplate and make prospectuses clearer through rules designed to reduce duplicative and generic information.[50]

    In 1998, the SEC also advanced its so-called Plain English Initiative, which went into force in October of that year. As part of the initiative, the SEC created guidelines for making prospectuses more readable,[51] including a "plain English rule" mandating that certain parts of the prospectus—in particular the Summary and Risk Factors sections—be written more clearly.[52] The rule specifically provided that the entire prospectus should avoid boilerplate.[53] Other guidance issued

---

47.   *See id.* ("Dense writing, with legal boilerplate and repetitive descriptions of the company, has become the standard convention."); *see also* BOCHNER ET AL., *supra* note 27, at 34 ("The Risk Factors section is commonly perceived simply as lawyers' boilerplate . . . .").

48.   *See* Regulation of Securities Offerings, 63 Fed. Reg. 67,174, 67,176 (proposed Dec. 4, 1998) (to be codified at 17 C.F.R. pts. 200, 202, 210, 228, 229, 230, 232, 239, 240, 249) ("[W]e seek to discourage drafters from just routinely providing the boilerplate transactional disclosure that some have suggested the standardized disclosure items have evoked. This alternative would re-focus drafters on analyzing and including the information particular to that deal that is material to investors. More focused disclosure could result."); *see also* BOCHNER ET AL., *supra* note 27, at 34 (describing historical trends in risk-factor drafting).

49.   *See* BOCHNER ET AL., *supra* note 27, at 34.

50.   *See* Regulation of Securities Offerings, 63 Fed. Reg. at 67,176. In particular, the SEC streamlined disclosure requirements relating to the description of the registrant's business by eliminating duplication of quantitative information provided in the financial statements; revised the description of property to elicit more meaningful and material disclosure; limited the scope of Item 507, relating to securities offered for the account of a company's individual security holders, so that a company only would have to disclose information regarding certain of its selling affiliates and significant beneficial owners rather than all of its selling security holders; and modernized the existing guides for industry-specific disclosure. *See id.*

51.   *See id.* at 67,222 ("In recognition of the importance of the prospectus to investors, we recently adopted rules that require the use of plain English in the prospectus.").

52.   Plain English Disclosure, 63 Fed. Reg. 6,370, 6,370 (Dec. 4, 1998) (codified at 17 C.F.R. pts. 228, 229, 230, 239, 274) ("The rule requires issuers to write the cover page, summary, and risk factors section of prospectuses in plain English."). The Rule went into effect and required compliance as of October 1, 1998. *Id.* Securities Act Rule 421(d), within Regulation C, is the plain English rule. *See* 17 C.F.R. § 230.421 (2018).

53.   *See* Plain English Disclosure, 63 Fed. Reg. at 6,371 ("[A]void . . . [v]ague boilerplate explanations that are readily subject to differing interpretations . . . ."); *see also A Plain English Handbook: How to Create Clear SEC Disclosure Documents*, U.S. SEC. & EXCH. COMM'N 65 (Aug. 1998), https://www.sec.gov/pdf/handbook.pdf [https://perma.cc/4MSB-S43J] (summarizing the plain English rule).

around the same time addressed boilerplate in the Risk Factors section, which the SEC was concerned had become overly filled with generic risks to help issuers escape liability.[54] The SEC issued guidance admonishing issuers to remove boilerplate risk factors completely from prospectuses.[55] Such risk factors, the SEC feared, overwhelmed investors with risks that were generic and highly improbable and caused investors to discount or miss risks that might actually be important.[56] Similarly, the SEC targeted boilerplate in the Management's Discussion and Analysis section of SEC filings between 2000 and 2003. Finding that too much of the section had become rote copying and pasting, the Commission issued guidance warning issuers not to rely on standardized, untailored language in that section.[57]

In addition to the SEC's explicit targeting of boilerplate language, Congress's passage of the Private Securities Litigation Reform Act ("PSLRA") in 1995 indirectly affected the use of boilerplate. The law afforded protection from lawsuits for issuers that made forward-looking statements—statements about their future prospects—as long as the statements were accompanied by meaningful cautionary language.[58] The legislative history of the PSLRA indicates

---

54. *See* Plain English Disclosure, 62 Fed. Reg. 3,152, 3,163 (proposed Jan. 21, 1997) (to be codified at 17 C.F.R. pts. 228, 229, 230, 239) ("Often the risk factor disclosure in a prospectus is boilerplate, listing risks that could apply to any offering or that are not likely to occur.").

55. *Id.*

56. *See id.* (emphasizing "the problem of listing many risk factors that are so general that they are not meaningful and add to the length of the document making the document difficult to read").

57. *See* U.S. SEC. & EXCH. COMM'N, SUMMARY BY THE DIVISION OF CORPORATION FINANCE OF SIGNIFICANT ISSUES ADDRESSED IN THE REVIEW OF THE PERIODIC REPORTS OF THE FORTUNE 500 COMPANIES (Feb. 27, 2003), https://www.sec.gov/divisions/corpfin/fortune500rep.htm [https://perma.cc/X3TQ-QTLW] ("Our comments addressed situations where companies simply recited financial statement information without analysis or presented boilerplate analyses that did not provide any insight into the companies' past performance or business prospects as understood by management."); Interpretation: Commission Guidance Regarding Management's Discussion and Analysis of Financial Condition and Results of Operations, Securities Act Release No. 33-8350, Exchange Act Release No. 34-48,960, 68 Fed. Reg. 75,056, 75,063 (Dec. 29, 2003) ("Any such discussion should be specific to the circumstances and informative, and companies should avoid generic or boilerplate disclosure."). In addition, an earlier release admonished against the use of boilerplate language in the MD&A section when discussing the impact of the technological issues associated with the approach of the year 2000. *See* Interpretation: Statement of the Commission Regarding Disclosure of Year 2000 Issues and Consequences by Public Companies, Investment Advisers, Investment Companies, and Municipal Securities Issuers, Securities Act Release No. 33-7558, Exchange Act Release No. 34-40,277, Investment Company Act Release No. 23,366, 63 Fed. Reg. 41,394, 41,398 (Aug. 4, 1998) ("This reflects our view that a flexible approach best elicits meaningful disclosure and avoids boilerplate discussions."); *see also* Notice: Disclosure in Management's Discussion and Analysis About the Application of Critical Account Policies, Securities Act Release No. 33-8098, Exchange Act Release No. 34-45,907, 67 Fed. Reg. 35,620, 35,622 (proposed May 20, 2002).

58. *See* 15 U.S.C. § 78u-5(c)(1)(A) (2012) (noting that a person "shall not be liable with respect to any forward-looking statement" so long as the forward-looking statement "is accompanied by

that such cautionary language could not be mere boilerplate, but that specific disclosures would instead be needed for issuers to have protection from suit.[59] While this provision is not available to IPO issuers, the law nonetheless reflects congressional concern with rote, meaningless cautionary statements and presumably had an impact on how practitioners thought about the cautionary language they used.

Most recently, in December 2015, Congress passed the Fixing America's Surface Transportation ("FAST") Act, a bill aimed at providing infrastructure repair to the nation's highway system, but that also included several provisions related to offerings under the Securities Act.[60] In particular, the FAST Act required the SEC to review its disclosure regulations and look for ways to eliminate redundancies and the use of needless boilerplate language.[61] As a result, the SEC proposed a set of relatively modest rules that have not been adopted as of the time of this writing, but has not otherwise taken steps to reduce or eliminate boilerplate.[62]

D. Defining Boilerplate

Before analyzing the theory and evidence on securities boilerplate, I attempt to define the term in a way that can be analyzed theoretically and empirically. In light of the forgoing background discussion, I use two related definitions of boilerplate in this Article and employ text-analysis techniques to measure boilerplate and assess its impact. The first definition of boilerplate is simply the amount of overlap between documents—or to put it another way, the amount of common language that is reproduced in multiple prospectuses. A second and related definition of boilerplate is copied language that conveys only generic information about a company, as opposed to mere framing or structural language (e.g., "Our results from operations last year were . . . .").

I use these definitions because they best align with traditional legal views of boilerplate as well as the motivations behind the SEC's

---

meaningful cautionary statements identifying important factors that could cause actual results to differ materially from those in the forward-looking statement").

59.   H.R. REP. NO. 104-369, at 43 (1995) (Conf. Rep.).

60.   Fixing America's Surface Transportation Act of 2015, Pub. L. No. 114-94, §§ 72001–03, 129 Stat. 1312, 1784–85.

61.   *Id.* § 72002.

62.   *See* FAST Act Modernization and Simplification of Regulation S-K, 82 Fed. Reg. 50,988, 51,024 (proposed Nov. 2, 2017) (to be codified at 17 C.F.R. pts. 229, 230, 232, 239, 240, 249, 270, 274, 275) (proposing amendments "to modernize and simplify certain disclosure requirements in Regulation S-K . . . in a manner that reduces the costs and burdens on registrants while continuing to provide all material information to investors").

policies. The SEC's pronouncements on boilerplate are surprisingly thin as to what the agency actually means by the term.[63] A basic definition drawn from *Black's Law Dictionary* is "content that is made to fit many uses."[64] The SEC's regulatory discussions of boilerplate refer to it as "imprecise" language "subject to differing interpretations."[65] In addition, the SEC in its release and disclosure rules makes reference to language that can be used for any issuer[66] and that is copied from the disclosures of other issuers. This definition makes sense given that IPO prospectuses are written using language borrowed from the disclosure of other, similar companies. This definition of boilerplate shares similarities with that adopted by other scholars who have examined the issue empirically. Professors Karen Nelson and Adam Pritchard examine standardized cautionary language in the Risk Factors section of company annual reports on 10-K forms using a similar definition, and measuring identical text by comparing the overlap in the number of bigrams and trigrams in those sections.[67] Professor Kathleen Hanley and Gerard Hoburg study standardized language in IPO prospectuses, adopting a similar definition encompassing overlapping text.[68] In their

---

63. *See e.g.*, Plain English Disclosure, 62 Fed. Reg. 3,152, 3,163 (proposed Jan. 21, 1997) (to be codified at 17 C.F.R. pts. 228, 229, 230, 239) (admonishing issuers against the use of repetitive boilerplate language, but not defining boilerplate language). To be sure, the SEC's regulations speak to more than boilerplate alone. They reference legalese and overly dense repetitive text, among other drafting issues. Given the traditional definition of boilerplate, and the central role of standardized disclosure copied from other deals, I use the definition adopted here.

64. *Boilerplate,* BLACK'S LAW DICTIONARY (9th ed. 2009).

65. Plain English Disclosure, 62 Fed. Reg. at 3,155.

66. For example, Item 503(c) of Regulation S-K states that issuers should not "present risks that could apply to any issuer or to any offering." *See* 17 C.F.R. § 229.503. In describing "generic" disclosure, the SEC explains:

> [A]lthough Item 503(c) instructs registrants not to present risks that could apply to any registrant, risk factor disclosure typically includes generic risk factors. Registrants often use risk factors that are similar to those used by others in their industry or circumstances as the starting point for risk disclosure, and the disclosure is not always tailored to each registrant's particular risk profile.

Business and Financial Disclosure Required by Regulation S-K, 81 Fed. Reg. 23,916, 23,955 (Apr. 11, 2016). In addition, it quotes the FAST Act's admonition to get company-by-company "information . . . disseminated to investors without boilerplate language or static requirements"; such boilerplate and static requirements imply language that remains unchanged from company to company. *Id.* at 23,921.

67. Karen K. Nelson & A.C. Pritchard, *Litigation Risk and Voluntary Disclosure: The Use of Meaningful Cautionary Language* (2nd Annual Conference on Empirical Legal Studies Paper) (Aug. 2007), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=998590## [https://perma.cc/6ENG-7KGX] (comparing the number of identical trigrams in one annual disclosure to the next year's annual disclosure to measure the level of boilerplate language). As further explained below, I used this method in untabulated alternative specifications as a robustness check. The results were consistent with the method employed here.

68. Hanley & Hoburg, *supra* note 32, at 2831 (explaining how to use root words to identify standard language in prospectuses). As explained below, I employed this method in untabulated

study, Hanley and Hoburg created wordcount vectors from the text of all IPO prospectuses filed in the ninety days before a given offering and all industry prospectuses filed between ninety-one days and one year before a given offering; they then regressed the word content of each prospectus on the word content of the precedent deals to develop a measure of standardized text.[69]

Although recycling text from other issuers is likely to generate generic disclosure, it is not necessarily the kind of boilerplate that investors or the SEC might worry about. The generality or specificity of the borrowed language, as well as the importance of what it is describing, are also relevant. For example, an issuer might use a copied risk factor stating that "the company faces risks if its supply chain is disrupted." This vague statement might apply to any company, but a more specifically tailored version of the disclosure indicating the kind of events that are likely to affect the supply chain would be more valuable. On the other hand, issuers might use generic framing language that introduces other disclosures—such as "Our net operating losses last year were . . . ."—without reducing informativeness.[70]

Thus, the definition I employ for the empirical analysis takes into account both overlap and topic. In order to assess which boilerplate is truly generic and which is either informative or at least innocuous, I construct a topic model that groups the repeated sentences thematically. This facilitates separating generic from potentially informative boilerplate and allows me to assess the impact of these topics on the outcomes used in the cosine similarity analysis. Together, these techniques shed light on the boilerplate issue further discussed below. In addition, I employ robustness checks of this methodology. As explained below, I use a second algorithmic similarity measure, known as Word2vec, that takes into account the context, order, and meaning

---

alternative specifications as a robustness check. The results were consistent with the method employed here, albeit more complicated to execute.

69. *See id*. at 2841.

70. Courts have also addressed the issue in a related context. Some have defined boilerplate in securities disclosure as overly generic language when construing the Private Securities Litigation Reform Act ("PSLRA") forward-looking statements safe harbor, which requires meaningful cautionary language. *See, e.g.*, Slayton v. Am. Express Co., 604 F.3d 758, 772 (2d Cir. 2010) ("A vague or blanket (boilerplate) disclaimer which merely warns the reader that the investment has risks will ordinarily be inadequate to prevent misinformation. To suffice, the cautionary statements must be substantive and tailored to the specific future projections, estimates or opinions in the prospectus which the plaintiffs challenge." (quoting Inst. Inv'rs Grp. v. Avaya, Inc., 564 F.3d 242, 256 (3d Cir. 2009))). In addition, Congress has expressed a view on the issue in the PSLRA context. The Conference Report for the PSLRA states: "Under this first prong of the safe harbor, boilerplate warnings will not suffice . . . . The cautionary statements must convey substantive information about factors that realistically could cause results to differ materially from those projected in the forward-looking statement, such as, for example, information about the issuer's business." H.R. REP. NO. 104-369, at 43 (1995) (Conf. Rep.).

of words when calculating similarity of texts. While this method is computationally expensive to perform, it provides assurance that the basic cosine similarity measure captures boilerplate well.

E. Legal Theory and Boilerplate

To develop hypotheses about the role of boilerplate in securities transactions and to provide context for the empirical analysis, I survey the legal literature on boilerplate more generally. Legal scholars have written numerous articles about boilerplate in a variety of legal transactions, including consumer contracts, sophisticated merger deals, corporate debt instruments, and sovereign bonds.[71] Despite all of this, securities disclosure boilerplate has received relatively little attention until recently. Theory that has been developed in contracts and consumer disclosure literatures, however, is valuable for understanding what role boilerplate might play in securities disclosure.

To draw from these other literatures, it is important to be clear about the similarities and differences between securities disclosure and other settings in which boilerplate is used. Like many other types of drafting that lawyers do, securities disclosure lends itself to commodification because the kinds of information requiring disclosure for many deals are similar, at least in a very general sense. As in other contexts, securities boilerplate may be language that has proven its usefulness before, either with investors or with the SEC. However, disclosure exists to correct information asymmetries between issuers and investors in securities markets, which are uniquely prone to fraud and manipulation. Investors purchasing company shares invest in the company's current assets as well as the expectation of its future prospects.[72] The law seeks to ensure that the company discloses enough information to allow investors to make an informed decision about the value of those assets and future prospects, which are inherently difficult to value without detailed information generally only possessed by company insiders.[73] Without such regulation, company insiders might make overly positive claims or withhold negative information, thereby skewing investors' ability to assess the company's true value.[74] Thus the

---

71. *See supra* notes 1–2 and accompanying text.

72. *See, e.g.*, Merritt, *supra* note 21, at 1015–22 (noting that security value is based on asset value and expectation of future dividends).

73. *See* Prentice, *supra* note 20, at 778 (highlighting the empirical link between mandatory disclosure requirements and efficient capital markets). For an argument that the law regulates on the basis of such bargaining, see Stephen Choi, *Regulating Investors Not Issuers: A Market-Based Proposal*, 88 CALIF. L. REV. 279, 283 (2000).

74. *See* Prentice, *supra* note 20, at 812. ("A central premise of disclosure theory is that any entity contemplating making a disclosure will disclose information that is favorable to the entity,

question that arises with regard to boilerplate in securities disclosures is whether it is an effective way to inform investors and fulfill the SEC's regulatory mandate.

By contrast, much of the prior scholarly work on boilerplate concerns documents that serve a similar function to contracts or notice provisions: their purpose is to set out the legal relationships between various parties or put parties (such as consumers) on notice of their rights (or lack thereof). Corporate and sovereign bond documents set out, for example, the mechanisms by which bondholders get paid and specify what will happen upon the occurrence of a default. An example of such language is the *pari passu* clause in sovereign bond covenants, about which a number of important articles have been written.[75] Unlike many such boilerplate provisions which are easily understood by contracting parties, Professors Mark Weidemaier, Robert Scott, and Mitu Gulati's research has shown that among lawyers who regularly employed such boilerplate, "there was no widely held understanding of what the clause actually meant."[76] The confusion became notorious after it resulted in problems for sovereign issuers, such as the Republic of Argentina, which was unable to restructure its sovereign debt due in part to the clause.[77]

While these bonds are also securities, scholars studying boilerplate in bond markets have focused less on disclosure about issuers and more on the language that governs cashflows and bondholder rights. Likewise, in consumer transactions, scholars have written primarily about how boilerplate is deployed to define parties' rights or limit consumer remedies.[78] In many corporate and commercial situations the boilerplate that lawyers produce is intended to facilitate contracting rather than fulfill a complex regulatory regime or provide information about an intangible and speculative financial product. Moreover, in many contracting situations the onus is on the parties to

---

and will not disclose information unfavorable to the entity.'" (quoting Ronald A. Dye, *An Evaluation of "Essays on Disclosure" and the Disclosure Literature in Accounting*, 32 J. ACCT. & ECON. 181, 184 (2001))).

75. *See* Mark Weidemaier, Robert Scott & Mitu Gulati, *Origin Myths, Contracts, and the Hunt for Pari Passu*, 38 L. & SOC. INQUIRY 72, 96 (2013) (discussing how the *pari passu* clause became an embedded ritual within sovereign bond boilerplate). A typical *pari passu* clause states: "These Notes rank, and will rank, equally (or Pari Passu) in right of payment with all other present and future unsecured and unsubordinated External Indebtedness of the Issuer." Choi & Gulati, *supra* note 2, at 990.

76. Weidemaier et al., *supra* note 75, at 74.

77. *See* NML Capital, Ltd. v. Republic of Argentina, 699 F.3d 246, 259 (2d Cir. 2012) (holding that the *pari passu* clause prohibits Argentina from formally subordinating the bonds by issuing superior debt and from paying on other bonds without paying on the FAA Bonds).

78. *See, e.g.*, Kearney & Trecker Corp. v. Master Engraving Co., 527 A.2d 429, 436 (N.J. 1987) (reviewing boilerplate language that limited a buyer's remedy).

negotiate for the terms, rights, or information that they need. In these contexts, the important questions are whether boilerplate provides an effective substitute or starting point for negotiated provisions, whether boilerplate terms are salient and legitimate to contracting parties, and whether parties to the deal understand their implications.

Despite the differences between securities disclosure and other kinds of corporate and commercial documents, scholarship across a range of areas is informative for thinking about securities boilerplate and assessing the validity of the SEC's rationale for targeting it. Synthesis of this scholarship helps bring to light some possible roles that securities boilerplate might play and explain why so much boilerplate persists. For ease of exposition, I group these reasons into three categories: efficiency (meaning savings in time and monetary costs), market forces (using, and indeed fearing to deviate from, language that has proven its value with the regulators and the market), and strategic vagueness (employing language that fulfills a regulatory requirement without expressing anything material). These categories overlap in some ways, but the thread of each is analytically distinct. I will explain each of these, and the literatures that describe them, in more detail in the following sections.

1. Efficiency and Bargaining

Much of the robust literature on boilerplate espouses the idea that boilerplate enhances efficiency, reducing transaction costs to the contracting parties. It does so by providing a starting point for the drafting process and an analogue to default terms that parties can choose to rely on without negotiation or tailoring. Where parties explicitly bargain over terms, such as in a merger transaction, boilerplate is efficient because it provides tested and readily understood clauses or modules that can be employed more quickly than if terms were developed from scratch. Where a transaction is more one-sided, such as in the sale of consumer products for which bargaining with individual purchasers over terms is not feasible, boilerplate terms have been described as features of products that can be accepted or rejected wholesale by the consumer and which, in theory, might be subject to market forces. This theory justifies the use of boilerplate on the grounds that it reflects an implicit bargain between contracting parties or, at the very least, it is what the parties would have wanted had explicit bargaining been feasible.

Whether the efficiency gains from boilerplate outweigh its

potential drawbacks is highly disputed.[79] For example, there is a forceful debate about boilerplate terms in consumer contracts that force consumers to arbitrate claims and limit their ability to aggregate claims.[80] On one side of this debate, scholars theorize that boilerplate reduces costs for consumers and producers of goods alike, and therefore reflects an optimal outcome.[81] According to this argument, contracts containing boilerplate are the result of an implicit bargain between consumers and producers in which consumers receive a discount on the goods in exchange for agreeing to the boilerplate terms forcing them to take any disputes to arbitration.[82] Although boilerplate terms are usually presented in a take-it-or-leave-it manner, consumers are free to reject undesirable terms by taking their business elsewhere.[83] Those who critique this line of reasoning argue that the boilerplate limitations on rights are not salient to consumers[84] and, in any event, it is questionable whether consumer markets are sophisticated enough to

---

79. *See* Radin, *supra* note 1, at 298 (asserting that judges should more carefully scrutinize many boilerplate contract provisions).

80. *See id.*; *see also* Jean R. Sternlight & Elizabeth J. Jensen, *Using Arbitration to Eliminate Consumer Class Actions: Efficient Business Practice or Unconscionable Abuse?*, 67 L. & CONTEMP. PROBS. 75 (2004) (arguing that boilerplate arbitration clauses unfairly deprive consumers of class action litigation rights); Jean R. Sternlight, *Creeping Mandatory Arbitration: Is It Just?*, 57 STAN. L. REV. 1631 (2005) (making a similar case with regard to the growing use of boilerplate mandatory arbitration clauses).

81. *See* Ben-Shahar, *supra* note 1, 895–96 (arguing that consumers actually benefit from boilerplate terms through a reduction in prices).

82. *See* Radin, *supra* note 1, at 294–95:

> The business will save money by deleting its customers' legal rights; the business will pass on these savings to the consumer; the consumer will value her legal rights less than the amount of the price reduction; so therefore, the consumer is choosing . . . to sell off her individual rights for the price reduction.

83. As one scholar put it, "If . . . a particular contract is a mass-produced inalterable thing, then the words that make it up are just elements of the thing, like wheels and carburetors." Arthur Allen Leff, *Contract as Thing*, 19 AM. U. L. REV. 131, 138 (1970); s*ee also* RICHARD POSNER, ECONOMIC ANALYSIS OF LAW 144 (8th ed. 2011) ("[W]hat is important is not whether there is haggling in every transaction but whether competition forces sellers to incorporate in their standard contracts terms that protect purchasers."); Robert A. Hillman & Jeffrey J. Rachlinksi, *Standard-Form Contracting in the Electronic Age*, 77 N.Y.U. L. REV. 429, 442 (2002) ("[T]he aggregate decisions of many consumer can pressure businesses into providing an efficient set of contract terms in their standard forms."); Russell Korobkin, *Bounded Rationality, Standard Form Contracts, and Unconscionability*, 70 U. CHI. L. REV. 1203, 1209 (2003) (arguing that there is a "market discipline established by the ability of buyers to shop among sellers for the most desirable package of product attributes, including contract terms").

84. *See, e.g.*, Melvin Aron Eisenberg, *The Limits of Cognition and the Limits of Contract*, 47 STAN. L. REV. 211, 241 (1995) (exploring the relationship between cognitive limitations and the ability of consumers to rationally evaluate preprinted terms); Robert A. Hillman, *Online Boilerplate: Would Mandatory Website Disclosure of E-Standard Terms Backfire?*, 104 MICH. L. REV. 837, 850 (2006) (discussing the problem of information overload in boilerplate terms).

price the legal terms into the price of goods.[85]

Research on boilerplate in the realm of negotiated agreements among relatively sophisticated market actors embraces similar efficiency arguments. This strand of research examines the use of boilerplate terms in areas such as secured transactions,[86] bonds,[87] and corporate contracts.[88] Much of the research focuses on the cost effectiveness of the use of boilerplate terms for agreements that often have complex but highly standardized features.[89] Efficiency and consistency are assumed to be the purposes of boilerplate language in many such situations.[90] However, although parties are frequently sophisticated enough to demand favorable terms, this research suggests that needless or even harmful terms are sometimes reproduced through the unthinking use of boilerplate.[91] Thus the drawbacks of boilerplate may sometimes outweigh its efficiency benefits.

The general argument that boilerplate promotes efficiency is plausible in securities disclosure as well: It should be quicker to take language from existing disclosure than it would be to draft everything from scratch. Therefore, one hypothesis is that boilerplate enhances efficiency by lowering transaction costs. Testing that hypothesis requires defining precisely what efficiency would mean in a securities deal. Cost and time are obvious candidates, as they are issues that could hinder completion of a successful deal. However, the boilerplate literature suggests that whether boilerplate is efficient depends in part on its other costs or benefits, including whether it accomplishes the intended purposes of disclosure (to convey useful information) or

---

85. Eisenberg, *supra* note 84, at 241 (asserting that the baseline rights of the consumers often diverge from the contract terms contained in a form contract, thus leading to the consumer's confusion over her basic legal position when accepting the contract).

86. *See* Baird, *supra* note 1, at 949–50 (discussing the inappropriate application of contract law to boilerplate in secured transactions where property is exempt from execution by operation of law).

87. *See* Mitu Gulati & Robert E Scott, *The Three and a Half Minute Transaction: Boilerplate and the Limits of Contract Design*, 40 HOFSTRA L. REV. 1, 2–5 (2012) (discussing the reasons for the persistence of *pari passu* clauses in sovereign debt covenants and the problems with contract terms that are included for the sake of efficiency only).

88. *See* Smith, *supra* note 7, at 1176 (describing modular contracting as a way to divide complex negotiations involving multiple issues into smaller parts that are easier to understand).

89. *See* Baird, *supra* note 1, at 935.

90. *See* Gulati & Scott, *supra* note 87, at 4 (describing the conventional wisdom that boilerplate language allows lawyers to be quick and efficient in serving their clients' needs).

91. *See, e.g.*, *id.* at 5 (providing an illustration of how the ubiquitous repetition of the *pari passu* clause in sovereign debt contracts led to a harmful result for a client); *cf.* Bernhard Ganglmair & Malcolm Wardlaw, Measuring Contract Completeness: A Text Based Analysis of Loan Agreements (Mar. 31, 2015) (unpublished manuscript), http://www.econ.jku.at/papers/2015/ressem/FoSe_Contract_Completeness_ganglmair.pdf [https://perma.cc/BHJ2-D8V4] (finding that complexity in contracts increases growth opportunities for clients).

whether it represents the result of an actual or implicit bargain between its producer (the issuer) and its consumers (the investors). Evidence of tradeoffs between boilerplate and other costs or benefits also needs to be explored to determine whether efficiency primarily explains boilerplate's use. Measuring these costs and benefits is one goal of the empirical analysis below.

2. Market Forces and Network Externalities

A related area of legal theory posits that standard language emerges when text comes to be accepted over time by market participants through many iterations of a type of transaction. Once a standard becomes accepted in the market, there is little incentive to change it because change is costly and the benefit of doing so will not accrue to the individual who implements the change, but instead to all future users of the new language. Boilerplate has thus been described as the product of a network externality: it becomes more valuable as more parties use it, but that value accrues to parties outside of the deal in which the standard was created (i.e., parties to future deals).[92] Experienced deal participants might seek to use terms that are "market," meaning that they are widely used and therefore widely recognized and understood by market participants, such as investors.[93] Boilerplate terms might also be the product of learning externalities if they have been formed through the experience of those who used them in the past. The new corporate law associate given the task of drafting a prospectus can produce a quality product by relying on years of others' experience. Such disclosures have three attractive features. First, they might reduce risk and uncertainty because they have already been tested by the markets and regulators. Second, they might be more easily understood by market participants who read such language frequently and immediately recognize the basic message. Third, they might signal quality by indicating that the company and its advisors understand the market and are familiar with the latest deals and trends.[94] In a sense,

---

92. *See* Kahan & Klausner, *supra* note 2, at 730 (noting that learning externalities allow early adopters of a term to "confer positive externalities on later adopters"). Reuse of boilerplate disclosures may also represent the less rational phenomenon of status quo bias. Given that the actors in IPO deals (apart from the issuer) are repeat players, one would expect the costs and benefits of boilerplate to play a greater role than a cognitive bias in determining boilerplate use over time.

93. *See* Roberta S. Karmel, *Disclosure Reform—The SEC is Riding Off in Two Directions at Once*, 71 BUS. LAW. 781, 818–22 (2016) (describing a situation in which market standards have developed for disclosure of emerging growth companies).

94. Any of these might be described as a learning externality. *See* Kahan & Klausner, *supra* note 2, at 730.

market forces might then act to sort disclosure language that is effective from language that is not, and thus perpetuate the effective language.[95] There are potential downsides, however, as parties may be reluctant to abandon a widely used standard for one that would be more value creating due to the collective action problem inherent in creating the new standard. Thus, boilerplate might represent a network or learning externality, which in turn might have positive implications if it is the equivalent of easily recognizable language that makes disclosure more efficient for those who are familiar with it, but may have negative implications if reliance on it undermines innovating more effective language.[96]

A hypothesis that emerges from this literature is that lawyers, underwriters, and issuers might find it valuable to employ language that has already passed SEC scrutiny and been tested on the market. The SEC reviews every IPO prospectus and issues comment letters seeking more information for parts of the document that are unclear or misleading.[97] Comment letters and the revisions they require take time to process and can delay a deal, and therefore the lawyers on a deal might use boilerplate to minimize the number of comments they draw.[98] Several sources of evidence might be examined to empirically test this hypothesis. If boilerplate is the result of a positive learning externality, one would expect to see benefits in terms of the outcomes that would drive the terms' use; transaction speed, the number of SEC comments disclosures draw, and indicia of investor uptake of information are all candidates, given that costs or benefits with regard to these outcomes would create incentives for the use or abandonment of standardized disclosures. If boilerplate represents language that is market standard and readily understood by investors and analysts, one would expect to

---

95.   For a parallel point made in the context of fine print in consumer contracts of adhesion, see Baird, *supra* note 1, at 949:

> By enforcing fine print that most never read, we may be enabling sellers to customize terms and offer a package that is far better than one that imposed only a general obligation to conform to generally recognized norms. If there are enough sophisticated buyers in the marketplace and it is easy enough for them to understand what is in the fine print, the forces of competition will drive sellers toward efficient terms.

96.   For a similar point with respect to contracts, see Smith, *supra* note 7, at 1176 (describing contract boilerplate as a modular language through which large, complex deals are broken into more easily understandable pieces). *See also* Erik F. Gerding, *Contract as Pattern Language*, 88 WASH. L. REV. 1323, 1326 (2013) (describing contract boilerplate as a pattern language through which large transactions are pieced together using various smaller, standardized transactions).

97.   Schneider et al., *supra* note 15, at 23.

98.   *Id.* (discussing the SEC comment process); *see also* CHARLES J. JOHNSON, JR., JOSEPH MCLAUGHLIN & ERIC S. HAUETER, CORPORATE FINANCE AND THE SECURITIES LAWS § 3.06 [F] (5th ed. 2016). From 2013 to 2015, the average number of comments the SEC made in connection with IPOs ranged from thirty to forty-two per deal. *See id.* at 3-94 (referencing a 2016 study by Proskauer Rose LLP examining market practices and trends for U.S.-listed IPOs).

see indications that it corresponds to greater certainty about an issuer. As a corollary, the empirical results would show that boilerplate is associated with less uncertainty with regard to an issuer. However, a contrary result might indicate a negative network externality.

3. Strategic Use of Boilerplate

Another related group of theories suggests that generic boilerplate helps issuers to be strategically vague. Strategic vagueness has been posited in contract design as a way for parties to defer issues that would be easier to negotiate at a later time or to save resources when contract issues are unlikely to be relevant.[99] In the context of disclosure, overuse of boilerplate could serve the additional purpose of making disclosure intentionally vague, which might be desirable for some companies.[100] A company may prefer vagueness because it does not want to disclose specific, sensitive information for fear that investors will punish it because of that information or that competitors will use that information against it.[101] There may also be more benign, strategic reasons to disclose in only vague terms. If there is only a very small chance that a certain kind of information will be relevant to investors or for the company's prospects, but it is mandated by the SEC's regulation nonetheless, it may be worthwhile to use an easier, more standard set of disclosures and worry less about whether language is rote or generic.[102]

In either of these cases, boilerplate makes sense as a means of complying with SEC disclosure regulations while disclosing only vague information. To the extent companies are strategically vague (either on their own or at their advisors' urging), one would expect the benefits of making such disclosures to be at least equal to the costs of negative investor reactions or a possible sanction for noncompliance with the

---

99. *See* Albert Choi & George Triantis, *Strategic Vagueness in Contract Design: The Case of Corporate Acquisitions*, 119 YALE L.J. 848, 852 (2010) (discussing the strategic use of vague language in M&A contracts as a means of saving resources against the remote possibility that the language will be important in the future).

100. *Cf.* Baird, *supra* note 1, at 949 ("Sellers that want to send signals have to devise ways of assuring buyers that the promise is not being undercut by what is in fine print.").

101. *See* Merritt B. Fox, *Retaining Mandatory Securities Disclosure: Why Issuer Choice Is Not Investor Empowerment*, 85 VA. L. REV. 1335, 1340–62 (1999) (making a similar argument for investor decisions to disclose generally).

102. *See, e.g.*, Choi & Triantis, *supra* note 99, at 852 ("If a provision matters only in remote contingencies, for instance, then the back-end costs should be discounted by that remote probability, and it may be correspondingly efficient to save front-end costs by using a standard (or a vague term) rather than a rule."); Eric Talley, *Disclosure Norms*, 149 U. PA. L. REV. 1955, 1956 (2001) (observing that, with respect to disclosure in situations of asymmetric information, "informed parties have an incentive to capitalize on their advantage by devising strategies to exploit their less knowledgeable counterparts").

disclosure requirements. These possible costs and benefits are further assessed below.

4. Tying Theory to Reality: How Disclosure Impacts Investors

One point that should be addressed before proceeding with the rest of the analysis is the widespread doubt that investors actually read much, if any, of the disclosure documents produced in connection with securities offerings.[103] Although the SEC's regulations are heavily concerned with the ability of individual investors to obtain and comprehend disclosure,[104] many (including the author) doubt that investors actually read the IPO disclosure carefully. If that is true, how can boilerplate make any difference?

There are several plausible mechanisms through which disclosure can influence investors and markets even if the language is not completely read, and these possible mechanisms motivate the empirical approach. First, most of the initial investors in IPOs are large, institutional money managers such as mutual funds and pension systems.[105] These entities employ teams of people who can process and analyze the disclosure and who possess automated means of parsing it as well. These institutions are key actors in the IPO process for setting prices by setting demand and by prompting market reaction through the signaling power of their participation.[106] In fact, these entities' role is so prominent that some scholars have argued that securities regulation ought not to consider individual investors at all, and should instead focus on institutional investors, since they are in the best position to understand and transmit disclosure.[107] Regardless of

---

103. *See, e.g.*, Ben-Shahar, *supra* note 1, at 891 ("The regulatory agenda that requires the sophisticated party to provide comprehensive information to its clients so as to help the clients reach autonomous, educated choices has never worked . . . ."); *see also* Tom C.W. Lin, *A Behavioral Framework for Securities Risk*, 34 SEATTLE U. L. REV. 325 (2011) (discussing the divergence between rational actor models of investor behavior and more realistic behavioral models).

104. For a recent illustration of this, see, for example, Mary Jo White, *A Conversation with Chair Mary Jo White*, U.S. SEC. & EXCH. COMM'N (Jan. 26, 2016), https://www.sec.gov/news/ speech/securities-regulation-institute-keynote-white.html [https://perma.cc/GC6C-CAKS] (noting that individual investors, when it comes to IPOs, "may get very excited from an article or a blog and invest their money, and so you worry about them not getting sufficient or accurate information").

105. Sean J. Griffith, *Spinning and Underpricing*, 69 BROOK. L. REV. 583, 585–90 (2003); *cf.* Kasim Alli, Jot Yau & Kenneth Yung, *The Underpricing of IPOs of Financial Institutions*, 21 J. BUS. FIN. & ACCT. 1013, 1014–16 (1994) (assuming financial institutions act as initial investors in IPOs).

106. *See* Alli et al., *supra* note 105, at 1014 (noting that financial institutions are not as underpriced as other firms for larger insurance against legal liability for misrepresenting value).

107. *See, e.g.*, Zohar Goshen & Gideon Parchomovsky, *The Essential Role of Securities Regulation*, 55 DUKE L.J. 711 (2006) (arguing that securities regulation would maximize efficiency

whether one agrees with that argument, it is rooted in the reality that disclosure is likely (and perhaps most effectively) consumed primarily by sophisticated entities.

Investment analysts also influence this process. Although many firms do not have any analyst following at the IPO stage, for those that do, analysts read, parse, and synthesize disclosure about an issuer to be used by other investors. Accordingly, analysts' recommendations may influence those who do not read the documents themselves.[108]

Second, the issuer's road show involves presentations that track the disclosure in the prospectus.[109] Boilerplate does not bear directly on such presentations, but it may have indirect effects, because the investors reading through a prospectus may need to process it quickly to determine what kind of further information is needed. Perhaps more importantly, boilerplate may signal to investors that the company has not conducted thorough due diligence or is intentionally avoiding disclosing something negative.

Third, and relatedly, the mere act of drafting specific disclosure prompts the lawyers and bankers to ask questions, conduct research, and uncover information. In this way, boilerplate may reflect that the company and its advisors have not asked hard questions or done thorough research, which may itself result in poor information dissemination. These possible mechanisms are each relevant to the analysis below.

II. EMPIRICAL ANALYSIS OF BOILERPLATE IN SECURITIES DOCUMENTS

Natural language processing ("NLP") tools provide methods for assessing the impact of boilerplate in ways that would have been difficult a decade ago. In recent years, language processing has entered wider use in social science and legal scholarship. Legal scholars have

---

if it were designed for sophisticated institutional investors); *see also* Luigi Zingales, *The Future of Securities Regulation*, 47 J. ACCT. RES. 391, 417 (2009) (suggesting that securities laws should be designed for efficiency rather than investor protection, and proposing that individual investors impose unjustified regulation costs on the market).

108. *See* Jill E. Fisch, *Does Analyst Independence Sell Investors Short?*, 55 UCLA L. REV. 39, 47–48 (2007) (noting that research sold by analysts is rapidly dissipated and "others cannot readily be excluded from using the information."); *see also* Shefrin & Statman, *supra* note 18, at 28 (noting that "[i]nsiders create new information when they trade").

109. Liability may attach for false statements made during the road show, *see* 15 U.S.C. § 77l (2012), and any statements made during the road show that are material but not included in the prospectus can form the basis for liability for material omissions under the Securities Act of 1933, *see* 15 U.S.C. § 77l, or Rule 10b-5, *see* 17 C.F.R. § 240.10b-5 (2018). Although some new information may be offered during the road show, if provided in written form, it must be accompanied by a preliminary or final prospectus. *See* 17 C.F.R. § 230.433 (2018).

used NLP methods to analyze various types of legal documents[110] and finance scholars have used these methods to examine prospectuses[111] and periodic filings with the SEC.[112] In this Part, I describe the methods used for measuring and analyzing boilerplate in the empirical portion of the study.

A. Data

The data for the empirical analysis is taken from a variety of sources. First, a set of IPOs was constructed for the years 1996 through 2015. IPOs for all issuers other than operating companies going public for the first time are removed, meaning that IPOs of companies spun off from larger companies, real estate investment trusts, securitization vehicles, and investment companies were kept out of the dataset. The reason for doing so is that the goal of this Article is to assess the information effects of boilerplate and there may be other sources of information for such companies apart from the IPO prospectus. The result is a dataset of 2,751 IPOs.

I downloaded the prospectus for each issuer from the SEC's online database, EDGAR, in either a text or HTML format. For purposes of the analysis, I used the initial S-1 filing containing a preliminary prospectus since this version of the prospectus is the one most likely to be seen by initial investors. I also gathered each final

---

110. A growing number of legal academics have used language processing and machine learning to study the law. *See, e.g.*, Jonathan Macey & Joshua Mitts, *Finding Order in the Morass: The Three Real Justifications for Piercing the Corporate Veil*, 100 CORNELL L. REV. 99 (2014) (using a machine-learning algorithm and text processing to analyze the most common conditions for the successful employment of corporate veil piercing doctrine); Nelson & Pritchard, *supra* note 67, at 1–2 (using computational linguistics to examine cautionary language in annual reports filed with the SEC); Gabriel Rauterberg & Eric Talley, *Contracting Out of the Fiduciary Duty of Loyalty: An Empirical Analysis of Corporate Opportunity Waivers*, 117 COLUM. L. REV. 1075, 1078 (2017) (using machine learning to quantify the occurrence of fiduciary duty corporate opportunity waivers in public companies); Eric Talley & Drew O'Kane, *The Measure of a MAC: A Machine-Learning Protocol for Analyzing Force Majeure Clauses in M&A Agreements*, 168 J. INST. & THEORETICAL ECON. 181, 183 (2012) (using machine-learning classifiers to identify and analyze material adverse event clauses).

111. *See, e.g.*, Hanley & Hoberg, *supra* note 32, at 2821–25 (describing a project in which natural language processing is used in prospectuses).

112. *See* Andriy Bodnaruk, Tim Loughran & Bill McDonald, *Using 10-K Text to Gauge Financial Constraints*, 50 J. FIN. QUANTITATIVE ANALYSIS 623, 623–30 (2015) (using an automated parsing algorithm to measure financial constraints); Tim Loughran & Bill McDonald, *Measuring Readability in Financial Disclosures*, 69 J. FIN. 1643, 1643–46 (2014) (noting a recent trend of text-based analysis in accounting and finance literature); Tim Loughran & Bill McDonald, Plain English, Readability, and 10-K Filings 2–4 (Aug. 4, 2009) (unpublished manuscript), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.531.6671&rep=rep1&type=pdf [https://perma.cc/HXN2-J5QD] [hereinafter Loughran & McDonald, 10-K Filings] (evaluating readability of 10-K filings and linking readability with investor response).

prospectus filed pursuant to Rule 424(b),[113] as well as each amended version in between.[114] I then cleaned the documents in a manner that is standard for text processing.[115] This involves the elimination of graphic content, punctuation, and stop words—words such as articles, personal pronouns, and conjunctions that appear often but provide little information relevant to meaning.[116] Removal of stop words is standard in language processing, as these words have been found not to contribute to the analysis of texts.[117] For all HTML documents, I removed tags and other HTML code. For all documents, tables containing numeric information were also removed because such tables cannot be easily compared and, in any event, are not the source of boilerplate that the SEC has expressed concern about. Finally, each word was stemmed—that is, any endings were removed and the word was reverted to its root, as is standard practice in text processing.[118]

In addition to the text data, I gathered other relevant data for each deal. I gathered financial information on issuers from Compustat. I obtained information on the market performance of issuers' securities from the Center for Research on Securities Prices ("CRSP") and the New York Stock Exchange Trade and Quote database ("TAQ"). From the Thomson One database, I obtained the dealsheet for each IPO and extracted information such as the identity and roles of the underwriters, the identity of counsel for issuers and underwriters, the auditors, auditor fees, and the size and timing of the deals.

B. Text Similarity

Once the text was processed and cleaned, I used a number of methods to assess the amount of boilerplate in each document. The main method was cosine similarity—a simple but powerful method for measuring the amount of resemblance in the language of different documents, which in turn provided the starting point for measuring boilerplate in the IPO documents. Cosine similarity has been used in

---

113. 17 C.F.R. § 240.424(b) (2018).

114. Amendments appear on Form S-1/A. *See* U.S. SEC. & EXCH. COMM'N, SEC 870 (09-18), FORM S-1 (Sept. 2018), https://www.sec.gov/files/forms-1.pdf [https://perma.cc/QYW7-NTSL].

115. *See* Matthew Gentzkow, Bryan T. Kelly & Matt Taddy, *Text As Data* 6–7 (Nat'l Bureau of Econ. Res., Working Paper, Mar. 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2941254 [https://perma.cc/FE5C-AYF9] ("[C]leaning steps reduce the number of unique language elements we must consider and thus the dimensionality of the data.").

116. *Id.*

117. *Id.*

118. *Id.*

information retrieval systems.[119] While this methodology is widely used in finance and other fields, it is still relatively new in legal research and therefore I provide a detailed description of it below. As an additional validation step, I used another measure of similarity to check the performance of the cosine similarity method. The method, called Word2vec, accounts for the location of groups of words together to assess word context. This method yielded results consistent with cosine similarity, but required far more computing resources. Therefore, I used cosine similarity for the Article's primary analysis.

1. General Methodology Description for Measuring Similarity

Cosine similarity is one means of measuring the amount by which two selections of text overlap. Measuring cosine similarity entails converting a selection of text to a numerical vector that represents the number of times each word in the text appears. The appearances of these words are weighted based on how frequently they appear across all documents using the common term frequency-inverse document frequency ("tf-idf") method. The method reduces the importance of words that appear very frequently across all documents so that the ultimate comparisons more strongly characterize the similarities and differences in less common language. The similarity score is calculated as the dot product, or what can be conceptualized as the cosine of the angle between the vectors, which can then be interpreted as measures of similarity. Since the cosine yields a number between zero and one, the measure can be interpreted as a percentage of similarity or overlap between the compared texts.[120]

---

119. Cosine similarity is widely used in studies of information processing. *See* Rada Mihalcea, Courtney Corley & Carlo Strapparava, *Corpus-based and Knowledge-based Measures of Text Semantic Similarity*, AM. ASS'N FOR ARTIFICIAL INTELLIGENCE 776–77 (2006), http://www.aaai.org/Papers/AAAI/2006/AAAI06-123.pdf [https://perma.cc/2VBL-NUG6] (using cosine similarity to measure the semantic similarity between text segments); *see also* Dekang Lin, An Information-Theoretic Definition of Similarity 1 (1998) (unpublished manuscript), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.55.1832&rep=rep1&type=pdf [https://perma.cc/BG5E-CHT6] (noting that cosine coefficients are utilized for measuring similarity when "objects are represented as numerical feature vectors").

120. *See* Mihalcea et al., *supra* note 119, at 776 ("This similarity score has a value between 0 and 1, with a score of 1 indicating identical text segments, and a score of 0 indicating no semantic overlap between the two segments.").

As an illustration, consider an example involving two sentences from the dataset, which I label Text 1 and Text 2:

Example 1:

Text 1: "Our growth depends on our ability to retain existing sellers and attract new sellers"

Text 2: "New and existing sellers are key essential factors for our growth"

These two sentences are first converted into numeric vectors based on the words they contain. If one imagines these vectors as though plotted on a plane, the angle between the two vectors shows the degree at which they are inclined with one another. If they are exactly the same, then the vectors will be parallel to each other, resulting in zero degrees of inclination (cosine of one), and if they are completely opposite, then they will be at ninety degrees of inclination (cosine of zero).[121]

The inclination of the vectors thus provides a useful means of measuring the similarity (or difference) between the two. In the example above, the cosine value equals 0.70, which allows for an interpretation that the sentences are 70% similar to one another.

For purposes of illustration, some other examples from the dataset are:

Example 2:

Text 1: "Investors should also understand that holding a portfolio of stocks even for an extended period of time can result in negative returns"

Text 2: "Investors holding individual stocks for an extended period of time also face the risk that the company they are invested in could enter a state of permanent decline or go bankrupt"

The cosine similarity value for the above example is 0.55204, or approximately 55%.

---

121. The angle of inclination between these two vectors is calculated as their dot product. *See* Lin, *supra* note 119.

Example 3:

Text 1: "The company produces automobiles"

Text 2: "The company produces aircraft and it produces satellites"

This gives cosine similarity of 0.575, or approximately 57%.

2. Measuring the Copied Language Between Transactions

I performed comparisons to determine the cosine similarity, or the effective overlap of the language, that each prospectus shares with every other prospectus. In order to examine the patterns with respect to various types of disclosure, I also excerpted several sections of each prospectus that are considered important for investors: the Risk Factors, the Use of Proceeds, the Management's Discussion and Analysis ("MD&A"), and the Business Description sections.[122]

I compared the text of every document, and each of the four individual sections, to the corresponding text from every other document in the dataset to obtain the cosine similarity for each pair. The result was a matrix containing 3,784,000 observations.

Using this matrix, I created a measure of the average amount of overlap in every selection of text to every other corresponding selection in the same industry, within the preceding year.[123] This approach best matched what the literature and anecdotal reports describe regarding the way in which lawyers and bankers draft IPO disclosure—by starting with recent deals from within the same industry.[124] This measure served as a basis for the analysis below. [125]

---

122. *See* Hanley & Hoberg, *supra* note 32, at 2823 (describing the prospectus's most important sections as "the Prospectus Summary, the Risk Factors section, the Use of Proceeds and Management's Discussion and Analysis").

123. This corresponds to the definition that the SEC suggests in some of its releases as well. *See, e.g.*, Plain English Disclosure, 62 Fed. Reg. 3,152, 3,152 (proposed Jan. 21, 1997) (to be codified at 17 C.F.R. pts. 228, 229, 230, 239) (proposing rule to require plain English "in writing the front and back cover pages, summary and risk factor sections of prospectuses; revise current requirements for highly technical information in the front of prospectuses").

124. *Cf.* Schneider et al., *supra* note 15, at 12 (noting that choosing the correct form is merely the beginning of prospectus preparation as they are applicable to nearly all industries).

125. This method is different from the one employed by Professors Hanley and Hoberg, who construct a "standard content" measure using a statistical regression to determine the relationship between words in a prospectus and words from other company prospectuses occurring within an industry within the preceding ninety days, and prospectuses between ninety-one days and one year before each IPO, and outside of the industry in the preceding ninety days. *See* Hanley & Hoberg, *supra* note 32, at 1643. As a robustness check, I employed the method used by Hanley and Hoberg and obtained similar results.

C. Phrase Analysis

A second component of the definition of boilerplate supplements the simple "copied language" definition. The phrase analysis described here attempts to separate copied language that vaguely and generically describes substantive issues (which might be problematic for investors) from copied language that serves a more innocuous framing or other function. The goal is to address the problem that cosine similarity may gloss over important differences in seemingly similar language. To illustrate, if one considers Examples 2 and 3 from the preceding Section, each sentence pair has a similar value for cosine similarity, but a human reader would likely detect more important thematic differences between the two sentences in Example 3 than in the two sentences in Example 2. These thematic differences are not completely captured by similarity measurements alone. In addition, although cosine similarity is a useful means for determining the extent to which two documents or sections are similar, it does not help to identify precisely which text within those sections or documents is similar and which is not.

1. Sentence-Level Analysis

In order to address these issues, I analyzed each sentence in each document and created groupings based on the topics they covered or the functions they performed. This in turn allowed me to distinguish generic disclosure on important issues, repetitive but possibly informative language, and framing language that should have no import to investors.

In order to do this, I first extracted every individual sentence from every prospectus and from each of the individual sections studied—Risk Factors, Use of Proceeds, MD&A, and Business Description. I then compared each phrase in each of these sections to every other phrase in the corresponding sections for all other deals to determine the degree of similarity between all phrases. I then identified identical or nearly identical sentences. For this purpose I considered sentences to be substantially identical if they had a cosign similarity score of at least 0.7 or greater (and thus could be considered 70% or more similar).[126] A visual examination of the sentences revealed that

---

126. The seventy percent cutoff was determined after experimenting with cutoffs of fifty percent, seventy percent, and ninety percent. There are relatively few sentences that are completely identical. Lawyers drafting deal documents typically alter the language of even very similar documents, if even in relatively minor ways, such that the new document is not an exact replica of its precedent. It is not surprising, therefore, that when I attempt to identify only

the 70% cutoff was appropriate—it was wide enough so that two sentences with the same meaning but a few words changed were counted as being the same (as in Example 1 above), but narrow enough that it excluded sentences having different meanings but sharing similar words. The cutoff also effectively separated sentences that shared most of the same words but differ by one or two words in a way that would change the meaning. I preserved sentences that appeared in at least twenty deals.

2. Creating a Topic Model

The process just described yielded the most frequently occurring sentences across all the documents and their corresponding appearance throughout all the years and different industry sectors in the dataset. It was difficult to discern any clear patterns from reading the sentences. I therefore created topic clusters using principal component analysis, a statistical technique that is useful for finding latent patterns and rooting out redundancies in large datasets.[127]

For each section studied above, and for the document as a whole, groupings (or principal components) were determined for the boilerplate sentences throughout all of the prospectuses. Each principal component consisted of a weighted average for each sentence, with weightings assigned such that the components describe the maximum amount of variation in the data. The weightings of each component indicated the importance of each sentence to its respective topic or component. This allowed me to group the boilerplate sentences together by both their meaning and the importance of their recurrence. The components could also be used as variables that describe the prevalence of topics in each document.[128] The clusters revealed which repeated sentences are merely framing language or rote disclaimers that appear in every prospectus and which pertain to information that investors might reasonably want specific information about. The framing sentences and rote disclaimers were then removed from the corpus, and the cosine similarity measures were recomputed. In removing these sentences, I used a conservative approach because even seemingly

---

sentences that are ninety percent similar or more, I obtain only a relatively small number of results.

127. The principal component analysis reduces the dimensionality of a large data construct by calculating a number of vectors equal to the number of components in the construct, each of which is orthogonal (or nearly so) to every other, thus each conveying the maximum possible information. Hervé Abdi & Lynne J. Williams, *Principal Component Analysis*, 2 WILEY INTERDISC. REVS.: COMPUTATIONAL STAT. 433, 433–40 (2010).

128. The twelve most prominent boilerplate topics in each section are set out in Appendix Table D.

boilerplate phrases might give information by their presence. For example, a relatively common phrase states that "summary pro forma financial data have been derived from the Company's consolidated financial statements." This seemingly banal statement could provide information because pro forma financial data is required of companies that have acquired or plan to acquire other companies, and thus the reference could itself be an informative signal in such a situation. A conservative approach is justified. A list of the most common boilerplate phrases removed is included in Appendix Table H.

D. Robustness Checks

I employed additional tests to check the robustness of the methods just described. First, I compared all of the documents (as well as the individual documents' sections) using the Word2vec algorithm. The Word2vec algorithm is implemented in such a way that it compares not only the order of all of the words in the prospectuses but also takes into account the context with which words are used, drawing on a corpus of word embeddings—i.e., instances in which words occur near or adjacent to each other in sentences.[129] This method yielded results highly consistent with the more efficient cosine similarity method.[130] In addition, I did the analysis using groups of three words (trigrams) instead of individual words as the basic unit for measurement. The results from this were similar to those obtained using unigrams, and for the sake of efficiency I use the original measure for the analysis. I note that, although different approaches were tested to ensure that the measure I use is appropriate, no quantitative model of language that currently exists is perfect, and there will inevitably be some shortcoming. The model and others like it are useful nonetheless, and insights can be drawn from it despite the fact that there may be specific circumstances that the model does not capture.

III. RESULTS OF ANALYSIS

This Part explains the results of the analysis. In doing so, it describes how the data relates to the theoretical arguments about boilerplate that can be gleaned from the literature.

---

129. *See* Gentzkow et al., *supra* note 115, at 23 ("[Word2vec] trains the vector representations for each word to be highly probable given the vector representations of the surrounding context."). For space economy, these results are not reported here.

130. *See infra* Appendix Figure 2.

A. Summary Trends in the Use of Boilerplate

As described above, I calculated a single boilerplate measure for each deal in a manner that was consistent with practitioner accounts of how boilerplate is incorporated into prospectuses.[131] The measure used was the average similarity between a given deal document and all other deal documents in the same industry group within the preceding year.[132] This was done by comparing each relevant section of text to its corresponding section in every other document (or comparing whole documents for the portion of the graph relating to the whole document) in the dataset that came from the same industry and was filed in the preceding year. The averages for each of these are shown in Figure 1, which graphs the use of boilerplate in the entire prospectus document and the Business Description, MD&A, Risk Factors, and Use of Proceeds sections.

Figure 1 provides average similarity percentages for all documents (whether in the same industry or not, denoted by the dark gray bar), as well as averages for comparisons between documents within the same industry (denoted by the light gray bar) in the past year.

Some interesting patterns are evident from Figure 1. The overall level of boilerplate in the typical registration statement is considerable. On average, the prospectuses across the entire time period share approximately 47% identical content to other recent industry deals. The averages for the individual sections studied are as follows: 32% for Risk Factors, 34% for MD&A, 23% for Use of Proceeds, and 15% for the Business Description. Of course, these averages vary over time, likely driven by trends in the market as well as regulation, as discussed below.

B. Boilerplate Trends Over Time

Of course, Figure 1 represents only averages, and the actual levels of similarity for individual deals vary widely by deal and by time. It is worth examining what might account for these differences and what impact the SEC's policies have had.

---

131. *See* BOCHNER ET AL., *supra* note 27, at 34 (providing an overview of a prospectus summary's typical requirements).

132. For industry groups, I used the SEC's Standard Industry Classification ("SIC") categories. As a robustness check, I also performed the analysis using Fama French 48 industries. The results were consistent. As a further robustness check, I calculated average similarity for deals outside the same industry.

FIGURE 1: AVERAGE AMOUNT OF IDENTICAL TEXT
(ALL DEALS AND DEALS IN THE SAME INDUSTRY IN THE PAST YEAR)
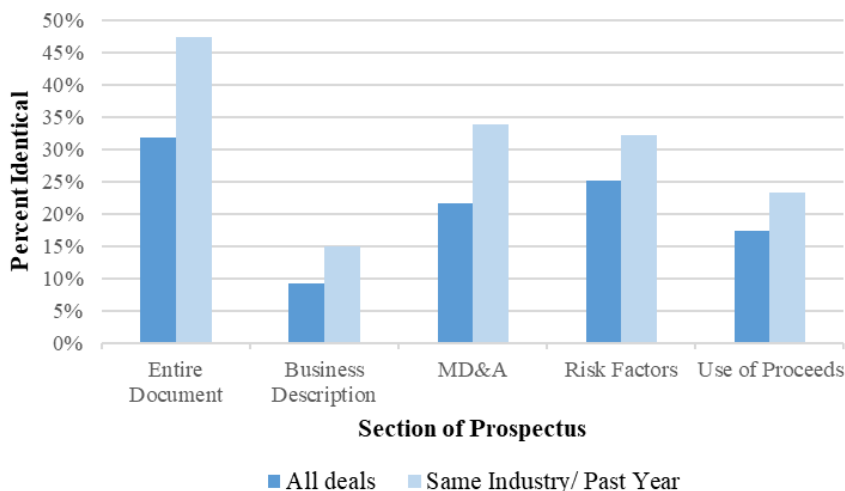
First, a pattern that is clear from the data, regardless of when a deal is completed, is that the size of the transaction (measured by the gross proceeds) bears a strong negative relationship to the amount of boilerplate that is used in the disclosure, as illustrated in Appendix Figure 1. The size of a transaction is related to its importance and the amount of attention it is likely to attract from investors, market observers, and regulators. It follows that size would be related to the amount of attention and tailoring that go into the disclosure, and therefore the size of the offering must be taken into account for any analysis of boilerplate to be meaningful beyond merely capturing the effect of deal size.

Second, the similarity measures described above allow for analysis of how patterns of boilerplate use have changed with time and evolving regulation. Indeed, an analysis of the trends over time shows that the level of boilerplate has grown steadily, despite some pullback based on SEC guidance. Figure 2 depicts the use of boilerplate over the time period covered in the data.

FIGURE 2: AVERAGE SIMILARITY BETWEEN DEALS IN THE SAME
INDUSTRY AND IN THE PRECEDING YEAR
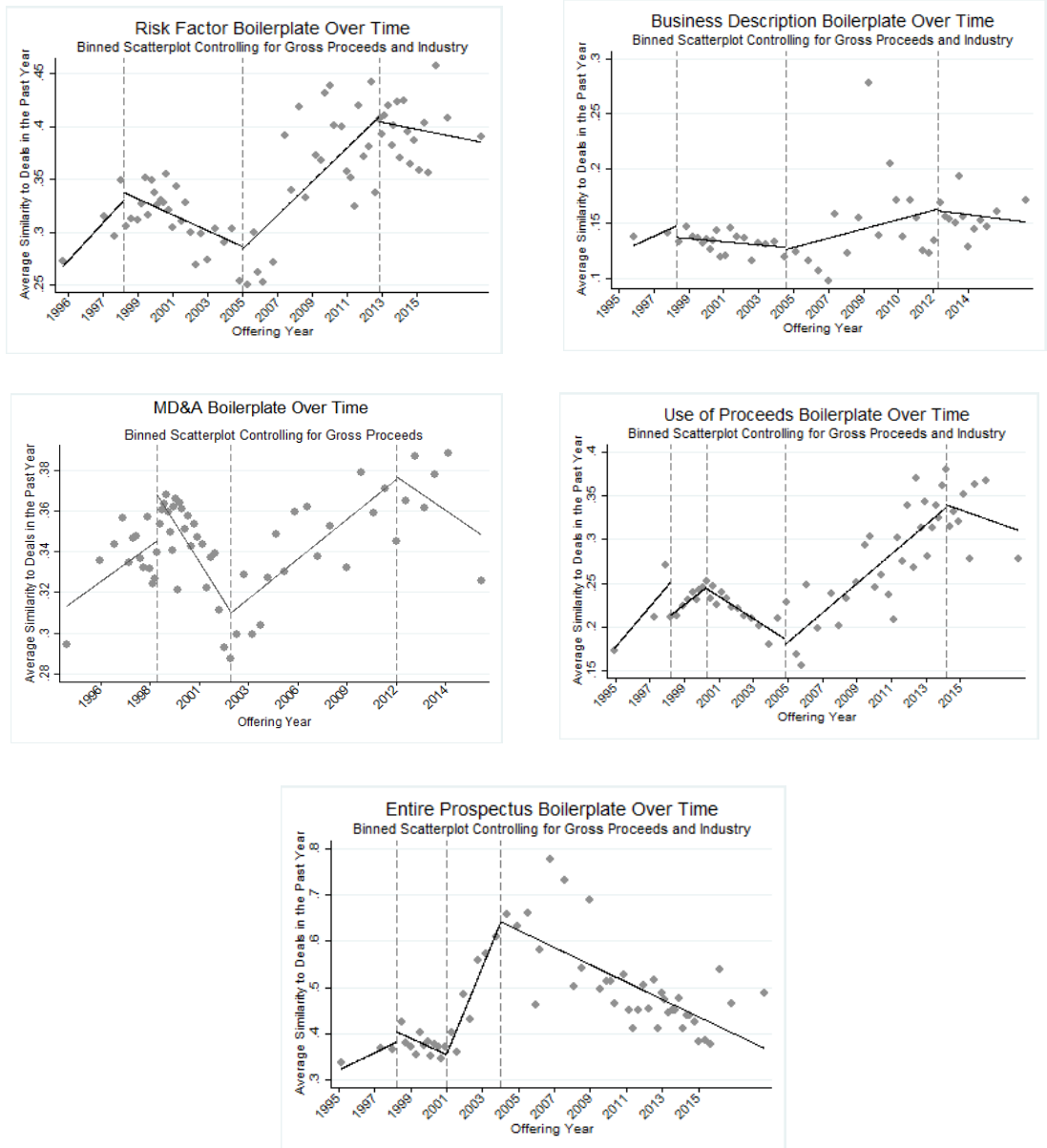
Controls are included for gross proceeds and industry group (using two-digit Standard Industry Classification ("SIC") codes) to show that the trend is independent of size, industry, or importance of the deal. The first discontinuity line corresponds to October 1998, when the

SEC's plain English rule governing boilerplate in disclosure went into effect.[133] Subsequent lines are placed to illustrate points at which the trends in boilerplate use changed over time. For several sections, the average boilerplate use shows an upward trend prior to October 1998, and the trend turns immediately downward thereafter before rising again in subsequent years.

It is plausible that the SEC's plain English rules spurred the downward trend in boilerplate after 1998. It is not clear, however, what caused boilerplate use to rise again a few years later in most cases. The reversal of the downward trend in boilerplate occurred in 2003. There are a number of possible explanations for this. One possibility is that it is due to the enactment of the Sarbanes-Oxley Act of 2002 ("Sarbanes-Oxley") and the entry into force of many of the regulations mandated by it in 2003. Although the law did not target boilerplate specifically, it is plausible that it might have led to more boilerplate because Sarbanes-Oxley introduced numerous new disclosure requirements that had not previously been tested by the markets or reviewed by the SEC.[134] In particular, it introduced new requirements for the management of public companies to either assess their internal controls and financial reporting processes or face liability for failure to do so.[135] As explained in Part I, if boilerplate represents a network externality, then boilerplate might be a means for parties to comply with disclosure requirements through the use of precedent documents, because such language has passed the scrutiny of the markets and regulators.[136] It

---

133. The Plain English Rule was proposed in early 1998 and formally went into effect in October 1998. *See* Plain English Disclosure, 63 Fed. Reg. 6,370, 6,370 (Dec. 4, 1998) (to be codified at 17 C.F.R. pts. 228, 229, 230, 239, 274) (adopting a rule that "requires issuers to write the cover page, summary, and risk factors section of prospectuses in plain English"); *see also* 17 C.F.R. § 230.421(d) (2018).

134. *See generally* Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745 (2002).

135. *Id.* §§ 301–03, 401–09, 116 Stat. at 775–78, 785–91.

136. Sarbanes-Oxley spawned many new disclosure requirements. *See, e.g.*, Disclosure Required by Sections 406 and 407 of the Sarbanes-Oxley Act of 2002; Correction, 68 Fed. Reg. 15,353 (Mar. 31, 2003) (to be codified at 17 C.F.R. pts. 228, 229); Disclosure in Management's Discussion and Analysis About Off-Balance Sheet Arrangements and Aggregate Contractual Obligations, 68 Fed. Reg. 5,982 (Jan. 28, 2003) (to be codified at 17 C.F.R. pts. 228, 229, 249). Other new requirements were enacted gradually over time, either incidentally to Sarbanes-Oxley or as new initiatives. *See, e.g.*, Securities Offering Reform, 70 Fed. Reg. 44,722 (Dec. 1, 2005) (to be codified at 17 C.F.R. pts. 200, 228, 229, 230, 239, 240, 243, 249, 274) (allowing more flexible incorporation of information by reference for certain companies); Additional Form 8-K Disclosure Requirements and Acceleration of Filing Date, 69 Fed. Reg. 15,594 (Aug. 23, 2004) (to be codified at 17 C.F.R. pts. 228, 229, 230, 239, 240, 249) (expanding the number of reportable events for public firms and shortening the filing deadline for many events to four business days); U.S. Sec. & Exch. Comm'n, Advisory Comm. on Improvements to Fin. Reporting, Final Report of the Advisory Committee on Improvements to Financial Reporting to the United States Securities and Exchange Commission (Aug. 1, 2008),

stands to reason that company officers and their counsel might be cautious in complying with the new disclosure requirements and might thus seek to copy language that has already been tested by the market and regulators. However, whether this explanation accurately captures the reasons for the upward trend in boilerplate after 2003 is not clear from the data. Nonetheless, the rise suggests that if dealmakers are rational, they derive some value from boilerplate that outweighs any negative consequences from using it.

C. Efficiency Explanations for Boilerplate

As discussed above, theory suggests that using cut-and-paste prospectuses might be more efficient, and if the language is tried and tested, it might lead to less regulatory scrutiny. Indicia of efficiency that can be observed in the data are the cost of a transaction to the issuer in terms of both fees and time, as well as the extent to which the deal is subject to regulatory delay. In this Section, I test the hypothesis that boilerplate is beneficial to issuers in terms of costs, time, and regulatory scrutiny. I do so by examining advisory costs to the issuer (legal fees, the underwriters' discount, and auditing fees), timing of deal completion, and the number of SEC comments that are associated with increased or decreased levels of boilerplate.

1. Advisors' Fees

The data provides some support for the hypothesis that boilerplate enhances efficiency, at least as measured by issuer costs, in that higher levels of boilerplate are associated with lower legal fees. However, there is no statistically significant association between boilerplate and auditor fees, or the underwriters' discount, even though auditors and underwriters might also contribute to boilerplate use either directly—through participation in drafting—or indirectly—by engaging in less due diligence and necessitating less alteration of a boilerplate precedent. Legal fees in the dataset range from approximately $20,000 to over $5 million, with average fees equal to approximately $850,000 and median fees of about $600,000. Linear regression analysis shows a negative and statistically significant relationship between boilerplate and legal fees (as the natural log of the fees). In other words, the analysis bears out the intuition that more boilerplate is related to lower legal fees, even when controlling for

http://www.sec.gov/about/offices/oca/acifr/acifr-finalreport.pdf [https://perma.cc/U2ST-4W94] (recommending measures to streamline financial reporting and eliminate redundancies).

factors that might otherwise affect each. The factors that I control for in the analysis include the IPO year, the issuer industry (using two-digit SIC codes), the gross proceeds from the offering, the age of the company, the size of the company (by assets), and the presence of venture capital backing. I also use fixed effects for each underwriting bank and each issuers' law firm. According to the analysis, each additional 10% of boilerplate in a document is associated with legal fees that are approximately $46,000 to $84,000 lower (or $65,000 lower on average).[137] While this is a significant amount, it is relatively modest given the average fee total.

One might suspect that the relationship between boilerplate and fees is driven by the quality of the law firms doing the deal or, relatedly, by how busy the law firm is. The quality of law firm in a practice area can be measured by both the market share (in dollars) of the deals the firm does in a given year in a given practice area and the raw number of deals.[138] Although there are obviously other features that define the quality of a law firm, market share and deal numbers are a reasonable approximation of how the market regards the firm's work. These statistics are also related to the firm's level of busyness. Very busy law firms may try to be more efficient by using more boilerplate; on the other hand, they may try to limit boilerplate if it is thought to produce poor quality disclosures. Neither supposition is supported by the data, however.

While certain law firms tend to use less boilerplate than others, the quality or busyness of the firms is not associated with the amount of boilerplate used. Appendix Figure 2 illustrates that the amount of boilerplate does not vary significantly from the average regardless of the quality of the firm. When the relationship between the quality of a law firm (as proxied by both its experience in raw number of deals and its deal market share) and the amount of boilerplate is examined using linear regression analysis, no statistically significant relationship is apparent for most specifications, and where the result is significant, the magnitude of the effect is nearly zero, as indicated in Appendix Table B. Similarly, law firm experience and deal flow turn out not to bear a statistically significant relationship to the amount of boilerplate used, once relevant factors are controlled for. One might assume that firms with greater deal flow would use more boilerplate because they have

---

137. *See infra* Appendix Table A.

138. *See, e.g.*, BLOOMBERG, GLOBAL EQUITY CAPITAL MARKET LEAGUE TABLES FY 2017, at 5 (2017), https://data.bloomberglp.com/professional/sites/10/UPDATE-2-Bloombergs-FY-2017-Global-Equity-League-tables.pdf [https://perma.cc/N8NE-LQHX] (ranking underwriters by market share in the U.S. equity market).

more precedents to draw upon. However, the data does not reveal any such relationship, as indicated in Appendix Table B.

In contrast to legal fees, accounting fees (and the underwriting discount) in the dataset bear no statistically significant relationship to the boilerplate measure. Auditing fees in the dataset range from $35,000 to $8 million, with an average of approximately $590,000 and a median of $400,000. The raw data show a general relationship in which deals with higher amounts of boilerplate tend to have lower accounting fees; however, this pattern does not withstand the addition of standard controls in regression analysis, as reported in Appendix Table A. This suggests that, all else equal, accounting fees do not appear to be a major factor that should drive reliance on boilerplate and there is no good evidence that boilerplate leads to greater efficiency with respect to the accountants' fees.

2. Transaction Speed

There is no statistically significant association between the use of boilerplate and the speed at which a deal in the dataset is completed.[139] I measure deal speed as the length of time that elapses between the time that the issuer files the registration statement and the date of the offering. This time-period average is 101 days for deals in the dataset, and the median is seventy-seven days.

In theory, using boilerplate might lead to faster deal completion because cutting and pasting is faster than drafting and editing nonstandard disclosure language. On the other hand, boilerplate might slow deals down if it draws SEC comments which must then be dealt with through revisions.

The data does not reveal a pattern in the raw data or in a regression analysis either way. Although this does not definitively indicate the absence of a relationship, any such relationship is not strong enough to be apparent in the raw data or when using relevant controls. Nonetheless, given that the timing of an IPO depends on a number of factors in addition to disclosure issues, it is possible that any effect is too attenuated to be perceptible. In any event, the evidence does not suggest that transaction speed is a compelling reason for the use of boilerplate.

---

139. *See infra* Appendix Table A.

3. Number and Extent of S-1 Amendments

In order to further assess whether boilerplate impacts the time efficiency of a deal, I analyze the amount of work that goes into amending the preliminary prospectus after it is filed. I do so by looking at the number of times an amendment to each prospectus is filed before the prospectus is finalized. In addition, I examine the proportion of the initial prospectus that is amended between the preliminary and the final versions. If boilerplate is more efficient because it is tested by the market, there should be less need to amend the prospectus as the deal is marketed and due diligence proceeds. Therefore, there should be fewer amendments, and there should be fewer changes between the preliminary prospectus and the final prospectus.

The data reveals no statistically significant relationship between boilerplate and either the number or extent of amendments to the prospectus, as reported in Appendix Table C. The average prospectus in the data is amended five times in the course of an IPO transaction. While some deals in the dataset are not amended at all before the final version is filed, some are amended as many as twenty-three times. However, no relationship is apparent between the number of amendments (or the log of the number of amendments) and the amount of boilerplate used.

Similarly, there is no statistically significant relationship between the use of boilerplate in Form S-1 and the amount of the prospectus that is amended before the final prospectus is filed, as reported in Appendix Table C. To calculate the amount by which a prospectus is amended, I calculate the "edit distance" between the preliminary prospectus filed with the registration statement on Form S-1 and the final prospectus. Edit distance is a similarity measure that calculates the number of insertions, deletions, and changes from one document to another. The data shows that on average, around 16% of the text of a preliminary prospectus is changed before the final version is filed. The amount of change for the deals in the data ranges from 2% of the text to 32% of the text. The lack of any discernable relationship between the number of amendments and the amount by which the disclosure is amended weighs against the efficiency theory of boilerplate.

4. SEC Comments

In theory, boilerplate might represent a positive learning externality if it is language that has passed muster with the SEC, and thus can be used efficiently to help navigate the review process. If that

is the case, then, on average, one might expect that the amount of boilerplate used in a preliminary prospectus would be associated with fewer SEC comments asking for clarification. Alternatively, if boilerplate creates noise, it might lead to a greater number of SEC comments. In order to assess whether either of these possibilities is true, I gathered the comment letters for the deals in the dataset that are available on EDGAR. One complication to this step is that comment letters were only made publicly available on EDGAR beginning in 2005 and are not available for prior years. Nonetheless, I collect the letters for the deals that occur during or after 2005. I then parse the letters to determine the number of comments in each letter.

IPOs in the dataset occurring after 2005 go through an average of four rounds of comment letters. These deals receive an average of thirty-four comments between the letters, with a minimum of two and a maximum of one hundred comments. The data reveals no relationship between the amount of language copied from other deals and the number of letters or comments, as set out in Appendix Table C. While this does not prove the absence of a relationship, it does cast doubt on the theory that boilerplate borrowed deal to deal enhances efficiency or represents a learning externality with respect to regulatory review of the registration statement. It could be the case that learning externalities exist with respect to specific boilerplate passages, but that in the aggregate there is no evidence that boilerplate facilitates the SEC review process. However, it is noteworthy that, despite the SEC's interest in boilerplate and despite the fact that the SEC flags boilerplate language for comments, the overall amount of boilerplate in a document bears no statistically significant relationship to the number of comments the prospectuses in the dataset drew.

D. Information Asymmetry

Regardless of boilerplate's effect on the cost or speed of transactions, an important question for lawyers as well as the SEC is whether boilerplate has any real impact on a prospectus's ability to convey information. The question is difficult to answer because investor reactions are hard to measure directly. Nonetheless, a number of indirect proxies for investor reactions can, when taken together, provide evidence as to whether any effect is present. But investors are not homogenous, and they range in sophistication and ability to acquire their own information. To tease out the possible impact on different types of investors, I employ several proxies that are discussed in the financial economic literature: document readability, offering price accuracy and the level of underpricing, and the propensity for the

offering price to be revised or corrected as new information is developed during road shows.[140]

I assess three possible hypotheses regarding the impact boilerplate might have on investors in the market that tie back into the literature on boilerplate in other contexts.[141] The first possibility is that boilerplate use has no discernable impact whatsoever. The second possibility is that boilerplate tends to obscure information in the aggregate, either out of intentional strategic vagueness or a negative network externality that leads to overinclusion of needless verbiage. If that is the case, then, on average, greater amounts of boilerplate would make it harder for investors or the market to determine the "true" value of the company's shares. A third possibility is that boilerplate increases information flow, either as a modular language or due to a network externality, or both. If that is the case, it could happen in one of two ways—either through its content, as a form of modular language that seasoned investors understand, or because its mere presence has a signaling effect.

1. Readability

One way to assess boilerplate's effect on comprehensibility is to examine readability. Whether an investor is scrutinizing a prospectus or flipping through it quickly, readability plays a role in her ability to absorb information. This is true for investors of all types, but especially for retail investors. A number of automated readability scoring systems exist to assess the readability of a text. All of these measures have their limitations, especially with regard to highly technical documents such as securities disclosures. However, they have been shown to provide a reasonable, if only approximate, measure of text readability and, therefore, in conjunction with the other analysis in this Article, add to the understanding of boilerplate's role in disclosure.

The empirical analysis shows that boilerplate bears a strong relationship to the readability of disclosure. I assess readability by means of the Gunning Fog readability index, a widely used, simple, and consistent method of determining the readability of texts.[142] The index

---

140. *See* B. Espen Eckbo et al., *Security Offerings*, *in* HANDBOOK OF CORPORATE FINANCE: EMPIRICAL CORPORATE FINANCE 233, 236–38 (B. Espen Eckbo ed., 2007) (reviewing the method used for empirical finance literature on IPOs).

141. *See supra* Section I.E.

142. For one of the first explanations and reviews of the Gunning Fog index, see George R. Klare, *Assessing Readability*, 10 READING RES. Q. 62, 73 (1974). The Gunning Fog index has been used widely in financial research. *See, e.g.*, Aymen Ajina, Mhamed Laouiti & Badreddine Msolli, *Guiding Through the Fog: Does Annual Report Readability Reveal Earnings Management?*, 38 RES. INT'L BUS. & FIN. 509, 510 (2016).

level of a particular document is calculated using a formula that accounts for the average number of words per sentence and the ratio of complex words (defined as words with three or more syllables) to total words.[143] Although the measure is simple, it is considered a reasonable indicator of the reading difficulty of a text.[144]

As demonstrated by Figure 3, the changes in boilerplate by regulation closely track similar changes in the readability of the various sections of prospectuses.

---

143. Klare, *supra* note 142, at 73.

144. As a robustness check, I performed the analysis using several other readability measures: the Flesch Reading Ease Index, the Flesch-Kincaid Index, the Smog Index, the Coleman-Liau Index, and the Automated Readability Index. All produced consistent results. Although the change in readability varied depending on the test used, the overall pattern was the same.

FIGURE 3: READING DIFFICULTY OF PROSPECTUSES OVER TIME, MEASURED BY THE GUNNING FOG INDEX
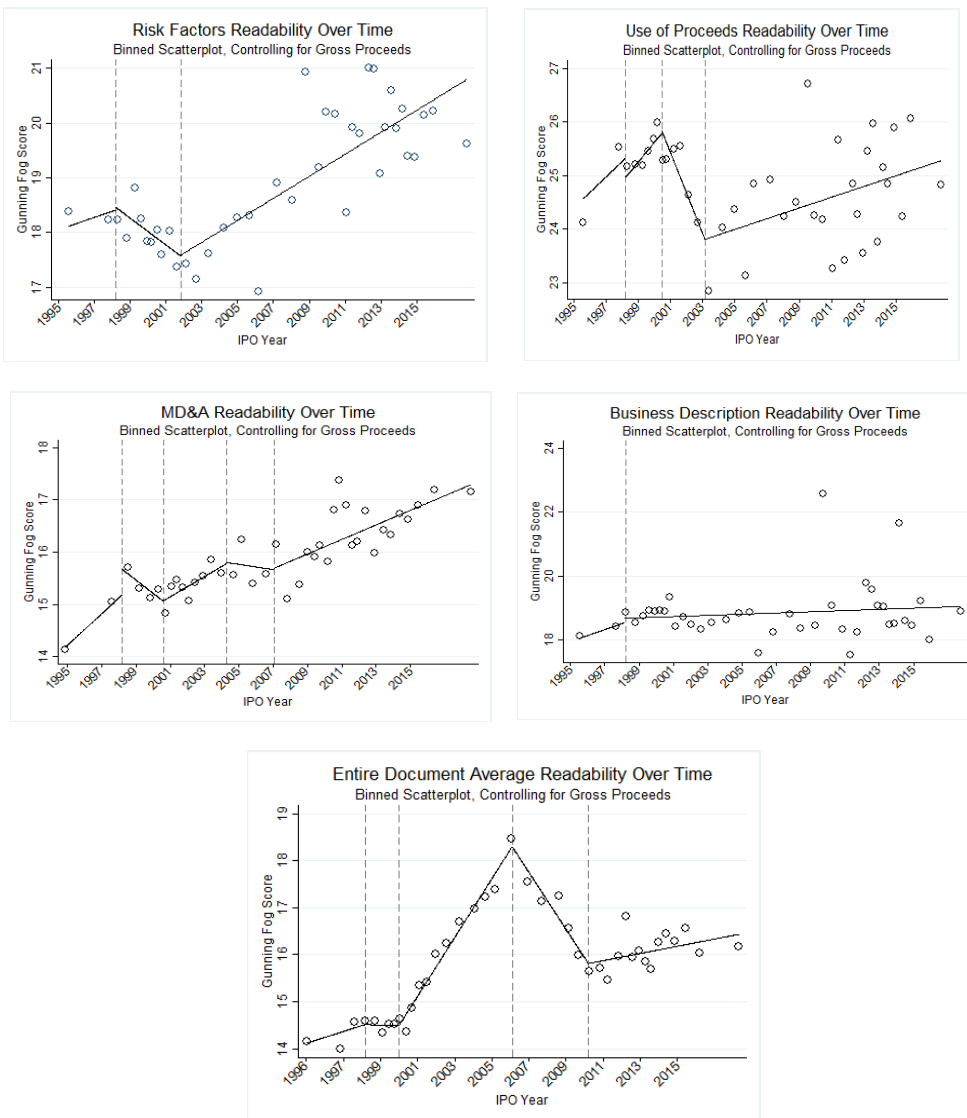
The scatterplots are binned into fifty quantiles that average the observations in the data. The graphs demonstrate the trend with respect to readability of the prospectuses over time. The lower numbers indicate documents that are easier to read, while the higher numbers indicate a trend toward more difficult documents.

In general, a Gunning Fog score of eight is considered appropriate for most audiences, whereas a score above eighteen is

considered unreadable by most audiences.[145] The average Gunning Fog score for entire prospectuses in the dataset is 17.9. The easiest prospectus to read has a total score of 10.8, and the most difficult has a score of 27.3. To put the scores in context, consider that *Green Eggs and Ham* by Dr. Seuss has a score of 2.5 and three recent issues of the *Yale Law Journal* have a collective score of 14.3.[146]

As in Figure 2 above, the graphs in Figure 3 are demarcated to note the points at which the plain English rule was enacted in late 1998 and any points at which the trend changed in subsequent year. Further analysis confirms a statistically significant change in the relationship between boilerplate and readability, as set out in Appendix Table G.[147]

Although a link between readability in general and overuse of boilerplate might seem intuitive, it is not clear from the data precisely why there would be a link between boilerplate and numerical readability scores such as Gunning Fog that are calculated based on the number of words per sentence and syllables per word, which are not necessarily related to boilerplate. One plausible explanation is that boilerplate develops by accretion—words and phrases are added over time and not eliminated. Moreover, when language is not tailored, it means that no one is making any attempt to streamline it and make it more readable by removing complex words and phrases. Moreover, although other components of the rule may have also affected readability when the rule was adopted, the relationship between boilerplate and readability remains even when a control is used for the other components of the rule. This suggests that boilerplate itself, or at the very least, the effort to reduce the use of boilerplate, has a large role to play in readability.

It has been pointed out that common measures of readability may not be appropriate for highly technical documents that are consumed primarily by experts, since these measures were created for general-use texts. This argument has traction when considering that experts consume prospectus language and standard readability scores may not capture the impact of language on these individuals. However, the readability scores are still relevant to retail investors who are not necessarily experts in reading such disclosures. To the extent that these scores reveal relative changes in reading ease, they convey something

---

145. *See* Loughran & McDonald, 10-K Filings, *supra* note 112, at 12 (suggesting that 10-K filings are unreadable by most audiences because the filings average a Gunning Fog score above eighteen).

146. Credit goes to Tim Loughran and Bill McDonald for the idea for these examples. For further analysis of the Gunning Fog Index in relation to various texts, see Loughran & McDonald, 10-K Filings, *supra* note 112, at 10, 12–13.

147. *See supra* Appendix Table G; *see also supra* Figure 3.

about how boilerplate might affect the accessibility of the disclosure to the average retail investor.

2. Analyst Assessments

The recommendations of securities analysts also serve as an indirect proxy for information asymmetry, and thus a gauge for whether boilerplate contributes to or reduces information asymmetry. Securities analysts, especially those on the sell side—i.e., those that work for investment banks and brokerage houses—issue reports and recommendations about whether to buy, sell, or hold a particular security. These analysts are sophisticated consumers of securities disclosures and important intermediaries who digest company information for other investors who may lack time or expertise to assess issuances themselves.[148] One way of measuring information asymmetry for a particular issuance is to look at the divergence of securities analysts' recommendations regarding the IPO.[149] If analysts' recommendations are uniform or nearly so (for example, if they all recommend buy, or give a similar earnings forecast), then it indicates that, whether the analyst community is right or wrong, they are getting strong signals about a stock's potential performance.[150] If the analysts' recommendations are all different (for example, some say strong buy, some say hold, others say sell) it indicates that the information available about an issuer is less certain.[151] With respect to IPOs, since the company usually has not issued any public information prior to the offering, the analysts must rely more on the disclosure than they might for companies that have been public for some time.

To investigate the analysts' perceptions, I obtained data on analyst recommendations from the Institutional Brokers Estimate System ("IBES"). Although not all companies that go public are followed by analysts initially, the dataset includes a sufficient number of IPO firms with at least two analysts covering them (1,859) within the first sixty days of going public to provide a useful sample. Since these companies are newly public, many analyst opinions do not give earnings forecasts, but they do give recommendations ranging from strong buy, buy, hold, sell, and strong sell. To measure the dispersion, I assign a

---

148. *See* COFFEE & SALE, *supra* note 42, at 122 (describing the role of securities analysts in equity markets).

149. *See, e.g.*, Jon A. Garfinkel, *Measuring Investors' Opinion Divergence*, 47 J. ACCT. RES. 1317, 1344 (2009) (linking the level of analyst coverage to analyst opinion divergence).

150. *Cf. id*. at 1317–18 (suggesting that the degree of heterogeneity among analyst opinions greatly affects the price of a company's stock).

151. *Id.*

numeric variable of one to five to each type of recommendation and then use a probit model to assess the likelihood that the analysts' opinions will diverge by more than one category of recommendation. Of course, many factors besides disclosure affect analysts' forecasts, such as general market conditions and trends in a certain industry at a given time. Therefore, I use fixed effects to control for the year, the industry category, and the underwriter of each transaction. Controls were also included for the log of the company's age, the log of its total assets, the log of the number of analysts covering the company, the presence or absence of venture capital, and the log of the gross proceeds of the deal. The results in Table 1 demonstrate that boilerplate bears a positive relationship to divergence of analyst opinions when controlling for those factors.

TABLE 1: DIVERGENCE OF ANALYST OPINIONS AS A FUNCTION OF
BOILERPLATE

Dependent Variable: Probability of Divergence in Analyst Opinions			Probability Change per 10% Change in Similarity
	(1)	(2)	(3)
Overall Similarity	-0.10	-0.09	**-1.00 to -0.90%**
	(0.15)	(0.27)	
Pseudo R^2	0.20	0.25	
Number of Observations	1,389	1,111	
Risk Factors Similarity	0.51**	0.16	**1.60 to 5.10%**
	(0.23)	(0.14)	
Pseudo R^2	0.20	0.25	
Number of Observations	1,299	1,285	
Use of Proceeds Similarity	0.70**	0.21	**2.10 to 7.00%**
	(0.29)	(0.32)	
Pseudo R^2	0.20	0.25	
Number of Observations	1,299	1,285	
MD&A Similarity	0.34*	0.48**	**3.40 to 4.80%**
	(0.17)	(0.22)	
Pseudo R^2	0.23	0.25	
Number of Observations	1,227	1,229	
Business Similarity	-0.03	-0.08	**-0.03 to -0.08%**
	(0.08)	(0.08)	
Pseudo R^2	0.19	0.24	
Number of Observations	1,194	1,282	
Prior to 2002 FE	X	X	
Industry FE	X	X	
Year FE	X	X	
IPO * Year FE	X	X	
Bank FE		X	

Probit model with marginal effects reported. Controls include fixed effects for each lead underwriting bank, the IPO year, the industry, and the interaction of these two sets. An indicator is also included for deals done prior to 2002. Additional controls for deal gross proceeds (log), the number of analysts making recommendations (log), issuer age (log), issuer total assets (log), and prospectus wordcount (log) are included for all specifications but not tabulated. Robust standard errors are in parentheses. Estimates marked with *, ** and *** are statistically significant at the 10%, 5%, and 1% level respectively.

Table 1 shows that boilerplate in the MD&A, Risk Factors, and Use of Proceeds sections is associated with an increased probability of disagreement among the analysts' predictions, even after underwriter fixed effects are included. It is plausible that boilerplate in these

sections in particular would affect analyst recommendations, given that they would be scrutinized by anyone looking to assess an issuer.

Two caveats are in order with respect to analyst coverage. First, prior to 2002, many securities analysts were found to have significant conflicts of interest because those that worked for investment banks (which are many of them) were compensated directly by the bankers and sales and trading operations.[152] Therefore, many analysts had an incentive to give favorable recommendations to stocks that their banks were underwriting.[153] This behavior included accompanying bankers and issuer's management on their road shows and helping to convince investors of the worth of the stock.[154] The conflict of interest led to the "global settlement" among the buy-side analysts and to regulations mandating that analysts be separated from commercial operations by a "Chinese wall."[155] To account for the significant change in analyst regulation, a control is used for all deals during or prior to 2002.

Second, as in other parts of this article, a caveat is in order, because other explanations are possible. For example, it could be the case that boilerplate is used by companies that are high risk or whose future performance is more uncertain, and that analysts' opinions on such companies diverge due to that risk or uncertainty. The regression analysis includes controls for indicia of risk and uncertainty, but it is possible these controls do not completely capture the risk and uncertainty. As a further test to ascertain whether boilerplate has a direct effect on analysts, I conduct an analysis using matched samples of issuers, as explained in the explanatory note for Appendix Tables F1 and F2. The results are consistent with those above and provide further evidence that boilerplate has an effect on analyst recommendations. This, in turn, bolsters the conclusion that boilerplate matters for information asymmetry.

---

152. *See* James C. Spindler, *Conflict or Credibility: Research Analyst Conflicts of Interest and the Market for Underwriting Business*, 35 J. LEGAL STUD. 303, 303–05 (2006) (noting that investment banks would offer positive analyst research in quid pro quos for underwriting businesses).

153. *See id.*

154. *See id.* ("Analysts were, essentially, part of the investment banking team–pitching deals to issuers, marketing offerings in roadshow presentations to investors . . . ."); *see also* Daniel J. Bradley, Bradford D. Jordan & Jay R. Ritter, *Analyst Behavior Following IPOs: The "Bubble Period" Evidence*, 21 REV. FIN. STUD. 101 (2008); Michael T. Cliff & David J. Denis, *Do Initial Public Offering Firms Purchase Analyst Coverage with Underpricing?*, 59 J. FIN. 2871 (2004); Raghuram Rajan & Henri Servaes, *Analyst Following of Initial Public Offerings*, 52 J. FIN. 507 (1997); Steven X. Zheng & David A. Stangeland, *IPO Underpricing, Firm Quality, and Analyst Forecasts*, 36 FIN. MGMT. 45 (2007).

155. *See* Spindler, *supra* note 152, at 304–05 ("With Sarbanes-Oxley's mandate implemented by the exchanges and NASD, the market for analyst research is effectively outlawed.").

3. Pricing and Market Reaction

Underpricing, price revision, and the changes in the bid-ask spread with respect to a newly public company's trading price are additional proxies of information asymmetry with regard to a company's securities. As discussed below, underpricing, price revision, and bid-ask spread have variously been used as proxies of information asymmetry in a large swath of the financial economics literature. I use these proxies to assess how boilerplate might affect securities disclosures' ability to reduce information asymmetry.

These signals are particularly relevant with respect to institutional investors, because they are the earliest consumers of the disclosure and the level of institutional investor interest in a deal helps determine the level at which the initial and final prices of the stock are set. These investors' level of interest is affected by information asymmetry about the issuer, which should, in theory, be mitigated by the disclosure. If the disclosure does a poor job of mitigating information asymmetry, then the initial investors are more likely to be uncertain about investing in an issuer. This means that they would require more inducement before giving the issuer serious consideration, given the uncertainty involved and the fact that they would have to expend their own resources, to the extent they could, to obtain the necessary information. The most salient inducement that issuers and their bankers have available is the price of an offering. If there is more information asymmetry with respect to a given issuer, the initial price is set lower to attract more interest, which in turn will affect the level of underpricing—the phenomenon of pricing the IPO below the level that the market will ultimately bear. The lower price also induces investors to do their own research, including by speaking to the issuer's management at the roadshow. The result of this process is that these investors may signal demand that deviates from what the banks anticipated based on how they thought the deal (and issuer information) would be received. In that event, the issuers and bankers might ultimately price the deal above the initial price range. The interaction between information and pricing makes underpricing and price correction reasonable proxies for information asymmetry vis-à-vis the institutional investors.

The bid-ask spread—that is, the difference between the price a market maker will buy a stock for and the price it will sell for—has also been widely used to measure information asymmetry. Market makers keep the spread as compensation for providing liquidity in a security. A wider spread is thought to represent, in part, higher compensation for

the additional risk of trading in securities for which information asymmetry is high and where market makers may be at a disadvantage to informed traders. Although information asymmetry is not the only determinant of the spread, it nonetheless serves as a useful proxy for means of assessing a firm's information environment and its relationship to disclosure.

a. Underpricing and Bid-Ask Spread as Proxies for Information Asymmetry

Vast literatures exist on the relationship between information asymmetry and both underpricing and the bid-ask spread. In this Section, I provide a brief explanation of the salient points from those literatures needed to provide context for the analysis. .

Underpricing is a common feature of U.S. IPOs, and the extent to which it occurs provides a proxy for the asymmetric information with respect to a transaction. Underpricing refers to the propensity for an IPO stock's price to rise on the first day of trading in the market and come to an equilibrium above the offering price. One might assume that a large price increase is a positive outcome for an IPO, and indeed some level of price increase (the first day bounce or "pop") has historically been a mark of a successful offering. For this reason, underpricing is often intentional to a certain extent; investment banks routinely and transparently pitch their IPO clients on an offering price set at a fifteen percent discount below what they think the market will bear in order to ensure a pop and attract interest from investors.[156] However, as the name implies, underpricing indicates that the initial offering price was set lower than what the market would bear and the issuer thus gave up proceeds it could have otherwise collected.[157] Thus, whether underpricing is problematic is a matter of degree. The more a stock is priced below the level needed to ensure a successful deal, the more underpricing represents an unnecessary loss for the issuer.[158] Scholars have puzzled over why issuers would tolerate excessive underpricing, which has often far exceeded fifteen percent, even surpassing one

---

156. *See id.* at 599–612 (describing the potential benefits of underpricing to issuers, including positive signaling effects and rewarding investor disclosure of demand); Goldman Sachs Grp., Presentation to eToys Regarding Initial Public Offering 4 (Jan. 18, 1999), https://archive.nytimes.com/www.nytimes.com/interactive/2013/03/10/opinion/sunday/nocera-goldman-sachs-etoys.html?ref=Sunday [https://perma.cc/3G4V-XMEP] (describing a "discount," usually ten to fifteen percent, necessary to ensure adequate post-offering appetite for stock).

157. *See* Griffith, *supra* note 105, at 583–90 (noting that underpricing allows individuals to quickly resell the shares they were allocated in an IPO for a profit).

158. *See id.* at 599–602 ("Issuers lose $0.93 per dollar of underpricing. Underpricing is thus much more expensive to issuers than it is to underwriters.").

hundred percent in many deals over the past few decades. A number of
theoretical explanations have been advanced,[159] the bulk of which focus
on the role underpricing might play in mitigating the effects of
uncertainty and risk related to the issuer or reducing litigation risk to
the issuer and the underwriter. These explanations in turn also relate
to problems of information asymmetry. According to theory, stock
issuances for companies about which investors have less information
exhibit more underpricing due to greater variance in predictions about
the company's performance and the institutional investors' need for
greater compensation for the risk of investing in an opaque company.[160]
Since the share prices rise quickly to the equilibrium level,
underpricing compensates those investors with immediate returns.

  The uncertainty can stem from inherent risk related to the
business, from lack of information about the business, or both. Investors
commonly deal with the inherent risk of particular businesses through
diversification. When lack of information is a problem, underpricing can
provide investors with an inducement to expend resources to gather
their own information about a company. Thus, it is plausible that more
underpricing would be present where a company and its underwriters
underinvest in producing information, which would mean more

---

  159. Most of the theoretical explanations come from finance literature. *See, e.g.*, James R. Booth & Richard L. Smith II, *Capital Raising, Underwriting and the Certification Hypothesis*, 15 J. FIN. ECON. 261, 261 (1986) (hypothesizing that an "underwriter can be employed to 'certify' that the issue price is consistent with inside information about future earnings prospects of the firm"); Kevin Rock, *Why New Issues Are Underpriced*, 15 J. FIN. ECON. 187, 188 (1986) (arguing that underpricing follows from the fact that underwriters discount the price to reach the "uninformed investor" and counteract adverse selection in the IPO market); Seha M. Tiniç, *Anatomy of Initial Public Offerings of Common Stock*, 43 J. FIN. 789, 790 (1988) (demonstrating that "gross underpricing serves as an efficient form of protection against legal liabilities and the associated damages to the reputations of both the investment bankers and the issuers"). However, the legal literature has addressed the issue as well. *See, e.g.*, Janet Cooper Alexander, *The Lawsuit Avoidance Theory of Why Initial Public Offerings Are Underpriced*, 41 UCLA L. REV. 17, 17–22 (1993) ("The 'lawsuit avoidance' theory posits that underpricing IPOs can avoid some suits altogether and reduce the potential damages in others, thereby serving as a form of insurance against legal liability."); Barondes & Sanger , *supra* note 34, at 169 (reviewing theories for underpricing); Yoram Barzel et al., *Prevention Is Better Than Cure: The Role of IPO Syndicates in Precluding Information Acquisition*, 79 J. BUS. 2911, 2911–13 (2006) (arguing "that information preclusion explains why managing underwriters require members of the underwriting syndicate to behave passively with respect to valuing and pricing an issue, engage in 'reciprocal participations' with syndicate members across successive IPOs, and maintain stable syndicate membership over time"); Richard A. Booth, *Going Public, Selling Stock, and Buying Liquidity*, 2 ENTREPRENEURIAL BUS. L.J. 649, 654 (2007) (offering a solution to underpricing by "conduct[ing] IPOs by means of a modified Dutch Auction"); James C. Spindler, *IPO Underpricing, Disclosure, and Litigation Risk* 15–16 (Univ. S. Cal. L. Sch., L. & Econ. Working Paper Series, Paper No. 94, 2009), http://law.bepress.com/cgi/viewcontent.cgi?article=1105&context=usclwps-lewps [https://perma.cc/4BMU-Z6JS] (providing evidence consistent with theories that underpricing is caused by information asymmetry).

  160. Barondes & Sanger, *supra* note 34, at 18.

boilerplate if boilerplate tends to be uninformative and less boilerplate if it tends to be more informative.[161] Although this is one of several explanations that has been posited for underpricing, it is especially plausible where there is high variance in first day returns because it indicates uncertainty that is at least partly determined by information asymmetry.[162] Consequently, if boilerplate has an impact on information asymmetry, a relationship should be present between boilerplate and underpricing.[163]

The bid-ask spread of a company's stock is another indicator of information asymmetry used in the finance literature.[164] Information asymmetry with regard to an issuer affects the bid-ask spread because the spread represents, in part, compensation to market makers for the risk involved in providing liquidity—i.e., standing ready to buy and sell a given security. If there is more information asymmetry with regard to a security, market makers require a larger spread in anticipation of increased risk posed by other, more informed traders in the market.[165] If boilerplate is associated with information asymmetry, one should expect to see a relationship between it and larger spreads, controlling for relevant factors (including other determinants of the spread that are not related to information, discussed below). As time goes on, informed traders trading in the company's stock provide information and thus reduce the information asymmetry. Thus if disclosure quality impacts information asymmetry, one should expect to see more boilerplate

161. *See id.* (positing that underpricing may occur to compensate uninformed buyers for the buyers' lack of knowledge about the company, which often leads to the uninformed buyers purchasing a great portion of the "bad (overpriced) IPOs").

162. This is a variation of Akerlof's lemons problem. For a more in-depth discussion of this problem as it relates to underpricing in IPOs, see Rock, *supra* note 159, at 187. It should be noted that other theories have been advanced to explain underpricing. *See* Ritter & Welch, *supra* note 13, at 1795. In any event, information, or lack thereof, is likely to play into underpricing, particularly when viewed in light of the propensity for price correction (which would counteract underpricing).

163. *See* Arnold et al., *supra* note 29, at 1516 ("Companies with more ambiguity in their offering prospectuses experience higher underpricing at the IPO."); Hanley & Hoberg, *supra* note 32, at 2860 ("Greater informative (standard) content decreases (increases) both the price change from the filing midpoint to the IPO price and underpricing."); Tim Loughran & Bill McDonald, *IPO First-day Returns, Offer Price Revisions, Volatility, and Form S-1 Language*, 109 J. FIN. ECON. 307, 324 (2013) ("In sum, one of the more appealing conceptual frameworks for IPO underpricing emphasizes the role of ex ante uncertainty.").

164. *See* Dan Amiram, Edward Owens & Oded Rozenbaum, *Do Information Releases Increase or Decrease Information Asymmetry? New Evidence From Analyst Forecast Announcements*, 62 J. ACC. & ECON. 121, 125 (2016) (employing bid-ask spreads as a measure of information asymmetry and noting that it is well established that spreads reflect information asymmetry, among other factors).

165. *See id.* at 124–125 ("One way to conceptualize this . . . is that as a specialist observes order flow, she increases [the] bid-ask spread to protect herself when it is more likely that order flow is coming from sophisticated investors that have superior [information] processing abilities.").

associated with wider spreads that narrow as information enters the market in other ways, such as informed trading.

b. Analysis of the Data

The data support the hypothesis that boilerplate is related to asymmetric information and that more boilerplate (in the aggregate at least) in fact increases investor uncertainty. This is apparent first in the underpricing data. Figures 4A and 4B illustrate the relationship evident in the raw data with respect to some of the key sections of the prospectus. The figures show percentage boilerplate in the Risk Factors, Use of Proceeds, and MD&A sections by quartile in relation to underpricing (Figure 4A) and variance of first-day returns (Figure 4B). The dominant (although not universal) pattern in the raw data is that underpricing and the variance of first-day returns both increase as boilerplate increases.

FIGURE 4A: AVERAGE UNDERPRICING IN RELATION TO BOILERPLATE
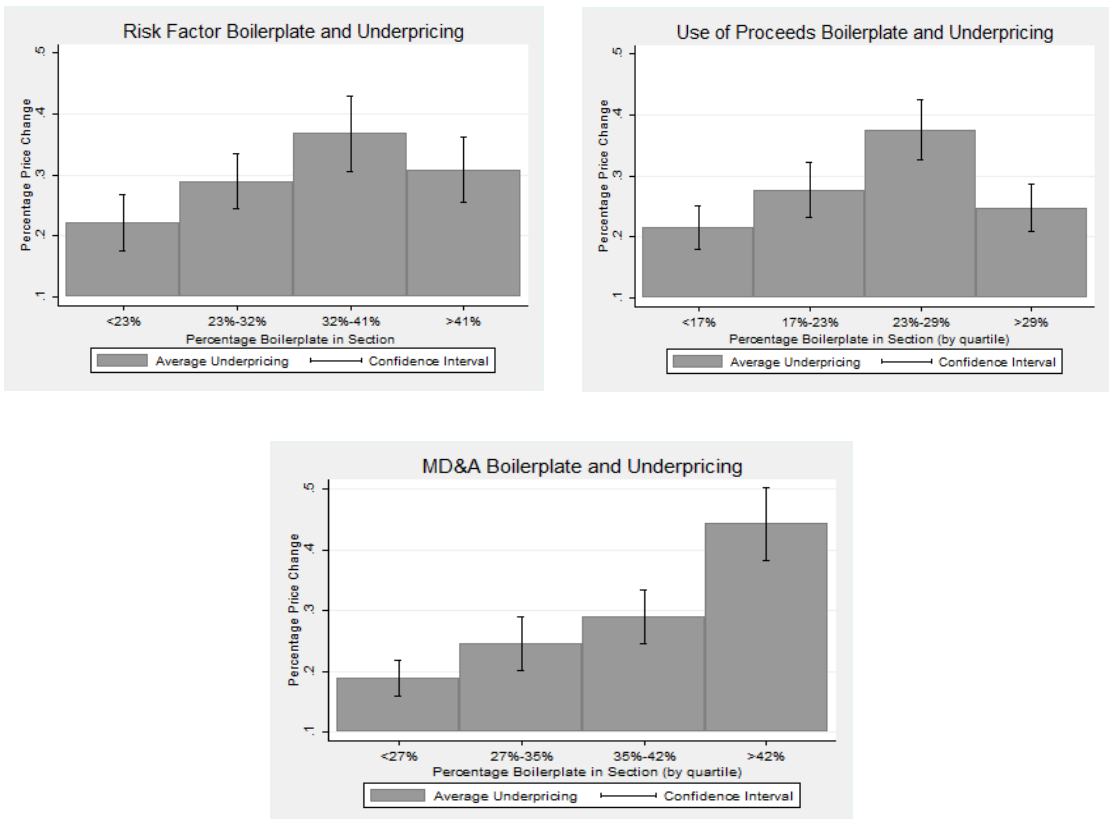
FIGURE 4B: VARIANCE OF FIRST-DAY RETURNS IN RELATION TO
BOILERPLATE
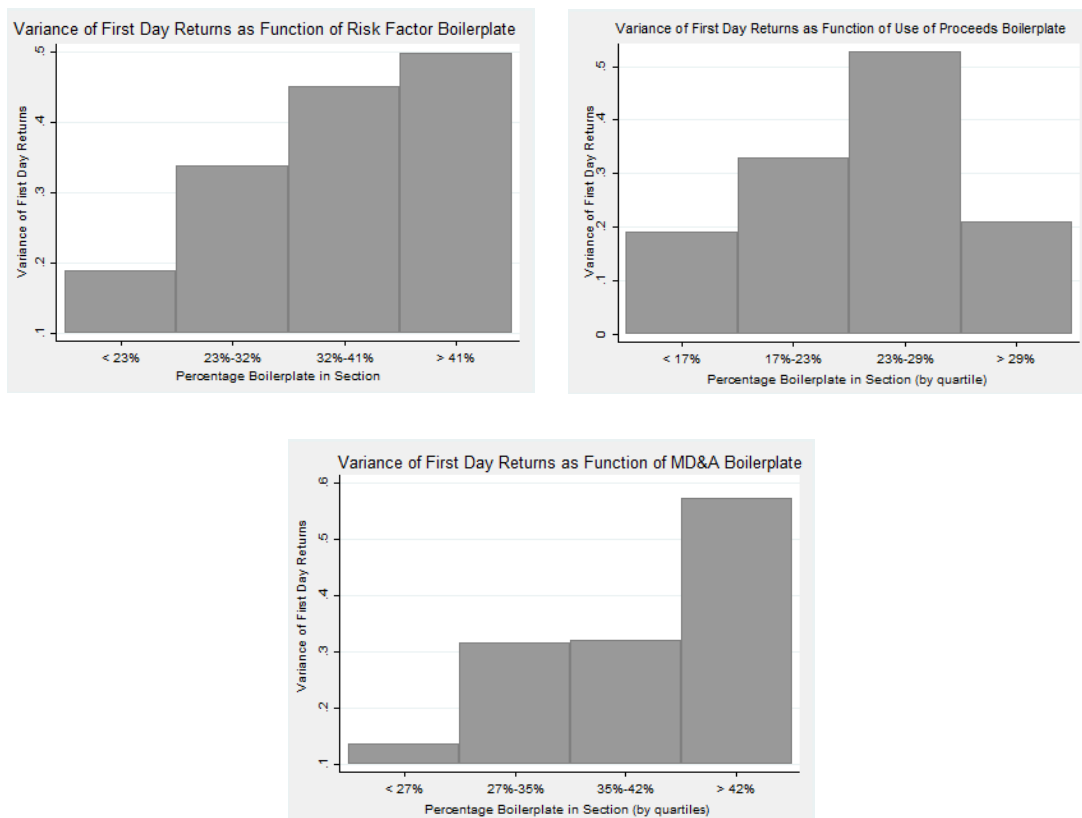
As Table 2 below shows, the relationship between prospectus language similarity and first-day price returns remains in ordinary least squares ("OLS") regressions that include controls for numerous factors that would have an effect on underpricing.[166] The dependent variable is the level of underpricing, defined as the percentage price increase on the first day of trading. The main independent variable is

---

166. OLS is a statistical method that attempts to determine the relationship between a set of explanatory variables and an outcome variable of interest by finding a function that approximately fits a set of data, holding a number of other factors (controls) constant. *See* JEFFREY M. WOOLDRIDGE, ECONOMETRIC ANALYSIS OF CROSS SECTION AND PANEL DATA 49 (2001).

I note that for the Use of Proceeds section, both underpricing and variance appear to drop off above 29% copied language. It is not clear why that is the case, although it should be noted the section is typically very short and certain standardized language about the proceeds may have relatively little impact on underpricing, relative to other parts of the disclosure.

level of boilerplate, defined as the average similarity between IPO prospectuses as a whole, and broken down by section.

TABLE 2: UNDERPRICING AS A FUNCTION OF BOILERPLATE

First-Day Returns			**Change in Returns per 10% Increase in Similarity**
	(1)	(2)	(3)
Overall Similarity	**0.30****	**0.34****	**3.00-3.40%**
	(0.12)	**(0.15)**	
Gross Proceeds (log)	0.09	0.13*	
	(0.05)	(0.07)	
Adj. R^2	0.37	0.34	
Number of Observations	2,050	2,164	
Risk Factor Similarity	**0.22*****	**0.23****	**1.70-2.80%**
	(0.13)	**(0.12)**	
Gross Proceeds (log)	0.09	0.12	
	(0.05)	(0.00)	
Adj. R^2	0.36	0.39	
Number of Observations	2,023	2,162	
Use of Proceeds Similarity	**0.60*****	**0.55****	**5.55-6.00%**
	(0.17)	**(0.14)**	
Gross Proceeds (log)	0.10	0.13	
	(0.05)	(0.19)	
Adj. R^2	0.36	0.39	
Number of Observations	2,023	2,160	
MD&A Similarity	**0.62*****	**0.51****	**5.10-6.20%**
	(0.18)	**(0.19)**	
Gross Proceeds (log)	0.10	0.12	
	(0.05)	(0.06)	
Adj. R^2	0.32	0.34	
Number of Observations	2,050	2,162	
Business Similarity	**0.00**	**0.07**	**0.00-1.00%**
	(0.14)	**(0.16)**	
Gross Proceeds (log)	0.08	0.11	
	(0.05)	(0.06)	
Adj. R^2	0.31	0.33	
Number of Observations	2,020	2,162	
Industry FE	X	X	
IPO Year FE	X	X	
Industry * Year FE	X	X	
Law Firm FE		X	
Bank FE	X		

Controls include deal proceeds (log) and fixed effects for each lead underwriting bank, the issuer's law firm, the IPO year, the industry, and the interaction of these two sets. Additional controls for issuer age (log), issuer total assets (log), venture capital involvement, volatility, syndicate size, whether the firm is a technology firm, debt to asset ratio, and prospectus wordcount (log) are included for all specifications but not tabulated. Robust standard errors in parentheses are clustered at the industry and year levels. Alternative untabulated specifications clustering standard errors at the bank and year levels for the first specification and the law firm and year levels in the second specification were significant at or below the 10% level for all specifications. Estimates marked with *, **, and *** are statistically significant at the 10%, 5%, and 1% level respectively.

The regression analysis controls for the log of the gross proceeds for each deal, serving as a control for the size and quality of the deal, as well as the log of the issuer's assets to control for issuer size.[167] The analysis also uses fixed effects for each lead underwriter, each issuer law firm, each industry category,[168] the presence of venture capital backing in the deal, each year, and the interaction of year and industry group.[169] Additional controls are used for the company's age, the market's volatility (measured by the CBOE Volatility Index), and an indicator for whether the company is a technology firm, according to SDC categories.

The table shows that for several of the sections, as well as for the prospectus as a whole, the average underpricing increases as the level of boilerplate language increases. That in turn suggests greater information asymmetry in the offering process as boilerplate increases.[170]

A similar effect is evident with respect to bid-ask spreads. The spread is determined by information asymmetry as well as market makers' order-processing costs and inventory carrying costs.[171] Accounting for those other factors, wider spreads indicate more information asymmetry and narrower spreads indicate less. Although

---

167. In line with the financial economic literature on IPOs, the regressions described in this Article use the natural log of the gross proceeds of each IPO in order to mitigate skewness in the distribution of dollar amounts. Eckbo et al., *supra* note 140, at 275–79. The gross proceeds are highly correlated with the size of the issuer and are frequently used as a measure of the issuer's quality. *Id.*

168. SIC codes are used to categorize the industry of issuing companies and are assigned for each securities issuer. *See* Eckbo et al., *supra* note 140, at 425. The analyses in this Article use the first two digits of the SIC code, which represents broad enough category to create groupings of similar deals, but specific enough to ensure that deals in the same category are in related industries. As a robustness check, I also tested each specification with Fama French industry classifications.

169. Fixed effects provide a method of controlling for variation within certain categories of variables by removing the mean of the observations for the dependent variable of interest. *See* WOOLDRIDGE, *supra* note 166, at 265–66 (illustrating a fixed-effect method). For example, if 2007 was a year that saw a particularly large amount of underpricing, the fixed effect would remove the year-specific average of the underpricing and leave only the variation attributable to other factors. The same is done for each IPO quarter, each lead underwriter, each industry, and the interaction of each industry and year.

170. The argument that unclear disclosure generally affects underpricing has been advanced elsewhere and supported by other empirical studies. *See* Spindler, *supra* note 159, at 30 (providing an empirical study to illustrate that disclosure affects uncertainty and underpricing); *see also* John L. Campbell et al., *The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings*, 19 REV. ACCT. STUD. 396, 405–06 (on file with author) (discussing market uptake of risk information); Todd D. Kravet & Volkan Muslu, *Textual Risk Disclosures and Investors' Risk Perceptions*, 18 REV. ACCT. STUD. 1088 (2013) (on file with author) (analyzing market absorption of risk information from periodic filings on Form 10-K).

171. *See* Amiram, et al., *supra* note 172, at 125 (using controls for various non-information asymmetry components of the bid-ask spread).

the pattern with respect to spread and disclosure is more difficult to visualize in the raw data, it can be analyzed in a regression model. Appendix Table E provides results of a regression model of bid-ask spreads on the first day of trading and at thirty days and sixty days after the IPO date in relation to the level of boilerplate. In addition to the controls included in the underpricing analysis, the model accounts for noninformation-related determinants of the spread by including controls for daily turnover, return volatility, and firm size, following the finance literature. The Table shows that greater levels of boilerplate are associated with wider bid-ask spreads on the first day of trading, indicating greater information asymmetry. The Table also shows that greater levels of boilerplate are more associated with spreads that become narrower over time compared to issues with less boilerplate disclosure. This is consistent with the hypothesis that boilerplate in a registration statement either contributes to, or indicates, information asymmetry at the time of an IPO and that this asymmetry is reduced over time as information enters the market by other means.

I note that the analysis with respect to both underpricing and spreads does not demonstrate that these relationships are causal. It could be the case that the boilerplate obscures information, or it could be the case that low-quality issuers use more boilerplate, but would have experienced more underpricing and wider spreads in any event. I address this problem in three ways. First, I employ controls for company age, venture capital involvement, amount of proceeds, managing underwriters, and law firms since these variables are correlated with the quality of the issuer.[172] Older companies have a longer track record from which to draw information. Those companies also tend to have long-standing relationships with their bankers and lawyers, and might therefore expect more effort from those parties in drafting disclosure.

Deal size is also important for disclosure and the outcome I analyze here. Larger deal proceeds are correlated with stronger and larger companies, since it easier for those companies to raise capital and to attract high quality law firms and underwriters to do so. Fixed effects for managing underwriters control for some aspects of issuer quality because underwriters provide a sorting function, in which high-profile underwriters generally take on the highest quality clients. Venture capital involvement can also indicate quality, given the resources that such firms can offer and given the fact that such firms have invested based on information not known the rest of the market.

---

172. *See* Eckbo et al., *supra* note 140, at 276–79 (providing a table of multiple different studies where a variety of variables are used to determine underpricing).

The second approach to addressing the causation problem is to perform a test using propensity score matching. The test involves performing similar regression analysis on samples of issuers matched on their predicted propensity to use boilerplate based on criteria that would otherwise indicate their quality, industry, deal timing, and other factors, as further explained in the text accompanying Appendix Tables F1 and F2. The method simulates an experiment in which companies in a control (low boilerplate) group are compared with similar companies in a treatment (high boilerplate) group. The results are consistent with the OLS analysis above.[173]

c. Price Revision

As previously noted, an additional proxy for the information asymmetry between investors and the issuer at the time they review the disclosure is the amount of offering price revision that takes place during the road show.[174] In particular, upward price revision, when viewed in conjunction with the other indicia analyzed here, provides a means to assess the relationship between boilerplate and information asymmetry.

When the price is revised upward during the road show, it indicates that the initial offering price range, set based on the information available to the underwriter after the preliminary prospectus is drafted but before due diligence is complete, was set far lower than the already underpriced level that the underwriter anticipated would be necessary to attract sufficient investor attention. A systematic pattern of such revisions linked to boilerplate would be most consistent with situations in which there is information asymmetry at the start of the marketing process that is alleviated by information that became available to the investors through channels other than the disclosure. The most likely such channels would be the road show or the investors' own research.

In either case, a positive relationship between boilerplate disclosure in the preliminary prospectus and upward price revision would imply that the boilerplate represents less information regarding an issuer, making the initial pricing less accurate and leaving investors to obtain information in other ways. Moreover, scholars have pointed out that pricing can be done in two ways: through preoffering

---

173. *See infra* Appendix Tables F1 and F2 and accompanying text.

174. Issuers and their bankers set an indicative price range in the initial Form S-1 based on estimates of what the ultimate price might be. The final price will end up either within that range, or it might be revised up or down from the initial range based on investor interest after the road show. *See* BOCHNER, *supra* note 27, at 47 (describing the road show process).

information discovery (via due diligence) or through book building, during which investors express their level of interest (i.e., demand) for the stock after being allowed to meet the company's management, read the disclosure, and ask questions.[175] These scholars posit that there is a tradeoff between due diligence and book building. An issuer (and perhaps more importantly, its underwriters) can invest in conducting due diligence and set the price accordingly, or it can simply set a low price and effectively pay off the initial investors to do their own research on the company. Less due diligence would result in less specific disclosure and more boilerplate borrowed from other deals, while reliance on book building would be marked by more price revision as the issuer and its investment banks set the initial price low to induce initial investors to do their own research.[176] Accordingly, a relationship between boilerplate and price revision would also indicate a relationship between boilerplate and information asymmetry.

d. Analysis of Price Revision

The pattern in the price revision data is consistent with the conclusion that boilerplate is related to information asymmetry. Looking at the raw data first, Figure 5 illustrates a generally positive relationship between the amount of copied language and a greater likelihood of upward price revision. The Figure shows the percentage of the relationship between boilerplate and price revision in the Risk Factors, Use of Proceeds, and MD&A sections (demarcated at the 25th and 50th percentiles). The mean occurrence of price revision trends upward for all three sections, although I note that the confidence intervals overlap significantly.

---

175. *See* Loughran & McDonald, *supra* note 163, at 315 ("The higher the uncertainty surrounding the IPO's valuation, the more likely new information (positive or negative) revealed during the bookbuilding process will have a significant impact in offer price revisions.").

176. *See id.* at 318 ("We could expect the IPOs with substantial uncertain/ negative language to have, on average, low preliminary offer prices, large upward price revisions, and high first-day returns due to the need of bankers to compensate investors for their information production.").

FIGURE 5: RISK FACTOR BOILERPLATE AND PRICE REVISION
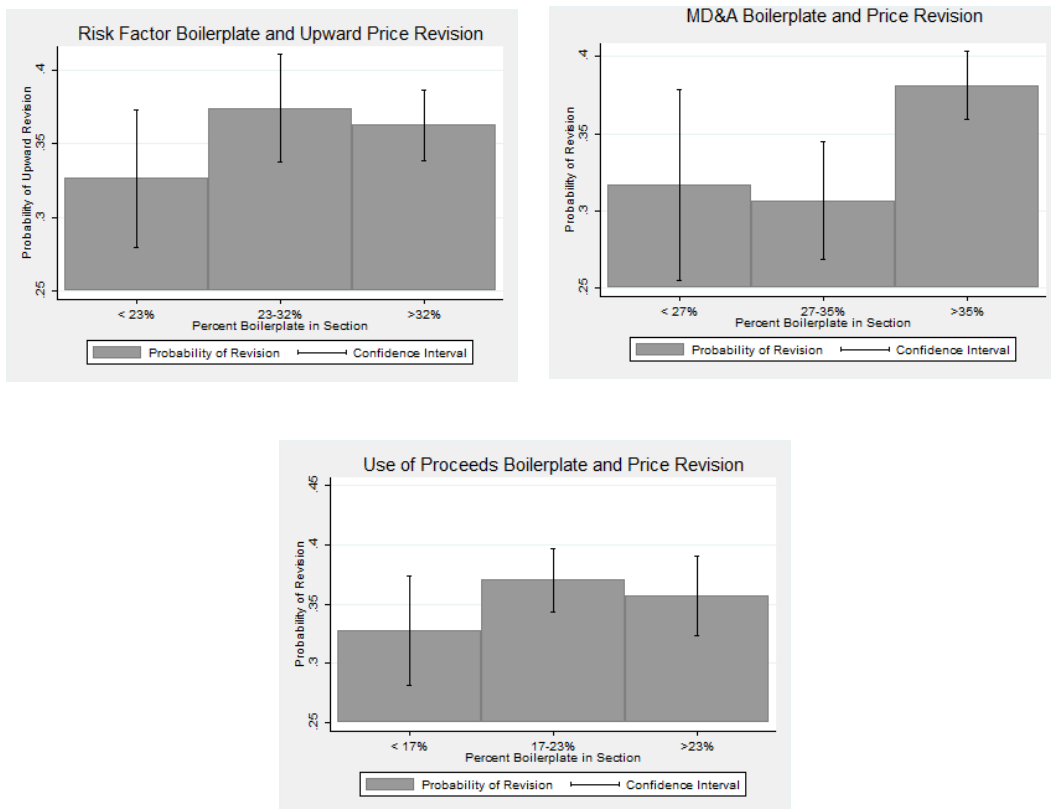
The relationship remains in a probit regression analysis of the probability that a deal's price will be revised outside the initial range, given the amount of boilerplate in the disclosure,[177] as shown in Table 3.

___

177. A probit model is a statistical model in which the outcome variable can take on only one of two values; it is useful for estimating the probability of an event occurring, versus the probability of the event not occurring. *See* WOOLDRIDGE, *supra* note 166, at 457–58 (illustrating the probit model).

TABLE 3: PRICE REVISION AS A FUNCTION OF BOILERPLATE

Price Revision			Probability Change per 10% Change in Similarity
	(1)	(2)	(3)
Overall Similarity	0.26**	0.24*	**2.40 to 2.60%**
	(0.11)	(0.13)	
Pseudo R^2	0.20	0.23	
Number of Observations	1,805	1,589	
Risk Factors Similarity	0.29***	0.33**	**3.00 to 3.30%**
	(0.09)	(0.16)	
Pseudo R^2	0.22	0.24	
Number of Observations	1,873	1,658	
Use of Proceeds Similarity	0.12	0.04	**0.04 to 1.20%**
	(0.15)	(0.20)	
Pseudo R^2	0.23	0.24	
Number of Observations	1,873	1,658	
MD&A Similarity	0.19*	0.29**	**1.19 to 2.90%**
	(0.11)	(0.14)	
Pseudo R^2	0.23	0.24	
Number of Observations	1,867	1,652	
Business Similarity	0.04	0.26	**0.40 to 2.60%**
	(0.21)	(0.27)	
Pseudo R^2	0.22	0.23	
Number of Observations	2,372	2,157	
Industry FE	X	X	
IPO Year FE	X	X	
Industry * Year FE	X	X	
Bank FE	X	X	
Law Firm FE		X	

Marginal effects reported. Controls include fixed effects for each lead underwriting bank, the issuer's law firm, the IPO year, the issuer's industry category, and the interaction of these two sets. Additional controls for issuer age (log), issuer total assets (log), deal proceeds (log), venture capital involvement, volatility (as VIX level), syndicate size, whether the firm is a technology firm, debt to asset ratio, and prospectus wordcount (log) are included for all specifications but not tabulated. Robust standard errors in parentheses are clustered at the bank and year levels for the first specification and the law firm and year levels in the second specification. An alternative specification clustering standard errors at the year and industry levels was also performed but not tabulated. Estimates marked with *, **, and *** are statistically significant at the 10%, 5% and 1% level respectively.

Table 3 indicates that for several of the sections, an increase in boilerplate is associated with an increased likelihood of upward price revision. In particular, boilerplate in the Risk Factors and MD&A sections leads to a higher probability of upward price revision, indicating a relationship between boilerplate in those sections and information asymmetry.

Once again, this analysis does not demonstrate that these relationships are causal. The analysis includes controls for company

age, venture capital involvement and amount of proceeds, since these variables are highly correlated with the quality of the issuer.[178] Fixed effects are used for each lead underwriter and the issuer's law firm, which are related to the quality of the issuer and characteristics of the deal, including disclosure. Fixed effects are also used for the issuer's industry, the IPO year, and the interaction of the two, to account for trends in each. In addition, as with underpricing and spread, I perform an analysis using matched samples of issuers, as further explained in Appendix Table G and the accompanying text.[179]

4. Litigation

The probability of drawing prospectus-related securities litigation provides a final metric of the effectiveness of boilerplate in creating information asymmetry. Issuers, and in some cases their underwriter and auditors, can be held liable for material misstatements or omissions in their disclosures.[180] Litigation might be related to boilerplate in those disclosures in a number of ways. First, boilerplate language can act as a type of catch-all disclosure that could potentially act as cheap insurance to protect issuers from litigation arising out of alleged omissions or misrepresentations in the offering document.[181] If, for example, certain risk factors have been found to sufficiently warn investors of certain risks and resulted in dismissal of securities lawsuits in the past, other firms would adopt the same disclosures to be afforded the same protection. If these disclosures offer such protection more often than not, then one would expect boilerplate to be associated with lower litigation risk on average.

On the other hand, generic boilerplate disclosures may fail to protect issuers from liability if they are too generic to be considered meaningful, as described anecdotally in the Wayfair case discussed in

---

178. *See* Eckbo et al., *supra* note 140, at 276–79 (providing a table of multiple different studies where a variety of variables are used to determine underpricing).

179. For a fuller account of this analysis, see *infra* explanatory note for Appendix Tables F1 and F2 .

180. Securities Act of 1933, § 11, 15 U.S.C. § 77k (2012).

181. *See* BOCHNER, *supra* note 27, at 34 ("The Risk Factors section is commonly perceived simply as lawyers' boilerplate, but a properly prepared risk factors discussion is carefully tailored to the company and can provide substantial protection from liability."); *US IPO Guide*, LATHAM & WATKINS LLP 58 (May 31, 2018), https://m.lw.com/thoughtLeadership/lw-us-ipo-guide [https://perma.cc/4TLD-CSFU]; Patrick J. Schiltheis et al., *The Initial Public Offering: A Guidebook For Executives And Directors*, WILSON SONSINI GOODRICH & ROSATI 140 (3rd ed. 2008), https://www.wsgr.com/publications/PDFSearch/IPO-guidebook-3.pdf [https://perma.cc/TAN8-G6AV] ("[G]ood risk factor disclosure may be the company's 'cheapest form of insurance.' There are certainly plenty of examples in securities litigation where risk factor disclosure led to the dismissal of multi-million dollar securities class action lawsuits.").

the Introduction.[182] Such disclosure might fail to convey specific information about a company that turns out to be critical for assessing a company's risks and prospects. If a company's share price does poorly and the performance is related to matters covered by generic boilerplate disclosure, then litigation is more likely to ensue based on the theory that more specific information should have been given. Indeed, some scholars have theorized that generic risk disclosures might even attract litigation by obscuring problems with the company, which could deceive investors or provide a roadmap for plaintiffs' attorneys seeking language to ground a claim upon.[183] Moreover, if expansive use of boilerplate results from a lack of careful due diligence of the company by its lawyers and underwriters, or reticence to disclose harmful information, then it is more likely that important information was not disclosed to investors, and possibly not even discovered by its advisors. It might thus call into question the reliability of the disclosure as a whole and result in more situations ripe for litigation.

The raw data indicates that more boilerplate in certain sections of the prospectus is generally related to more litigation. Figure 6 illustrates this relationship, indicating a higher average share of litigation in those deals with boilerplate higher than the 25th percentile for Risk Factors, Use of Proceeds, and MD&A sections.

---

182. *See* Dingee v. Wayfair Inc., No. 15cv6941(DLC), 2016 U.S. Dist. LEXIS 68322, at *10–11 (S.D.N.Y. May 24, 2016) (reviewing vaguely worded risk factor disclosure in an IPO). As noted above in Part I, the Private Securities Litigation Reform Act of 1995 created a safe harbor, protecting issuers from litigation based on forward-looking statements, as long as they provide "meaningful cautionary language" in the prospectus. *See* 15 U.S.C. § 78u-5(c)(1)(A) (2012) (stating that a person should not be liable for a forward-looking statement that is accompanied by meaningful cautionary statements). However, the safe harbor explicitly does not apply to IPOs. *See id.* ("[T]his section shall not apply to a forward-looking statement . . . that is . . . made in connection with an initial public offering . . . .").

183. *See* Spindler, *supra* note 159, at 33 ("[C]ontrolling for risk of litigation, more disclosure makes a firm more likely to be sued for in relation to its IPO."); *see also* Saumya Mohan, Disclosure Quality and Its Effect on Litigation Risk 38 (Sept. 1, 2006) (unpublished dissertation, McCombs School of Business, University of Texas at Austin), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=956499 [https://perma.cc/53P9-MY9E] ("My findings are that filings which are long but contain a smaller proportion of informative words such as numbers, past and future related words are more likely to precede shareholder lawsuits."); Nelson & Pritchard, *supra* note 67, at 28 ("High risk firms also revise their cautionary language more from year-to-year, suggesting they avoid boilerplate warnings that are unlikely to garner legal protection under the statute.").

FIGURE 6: RISK FACTOR BOILERPLATE AND CLASS ACTION FILINGS
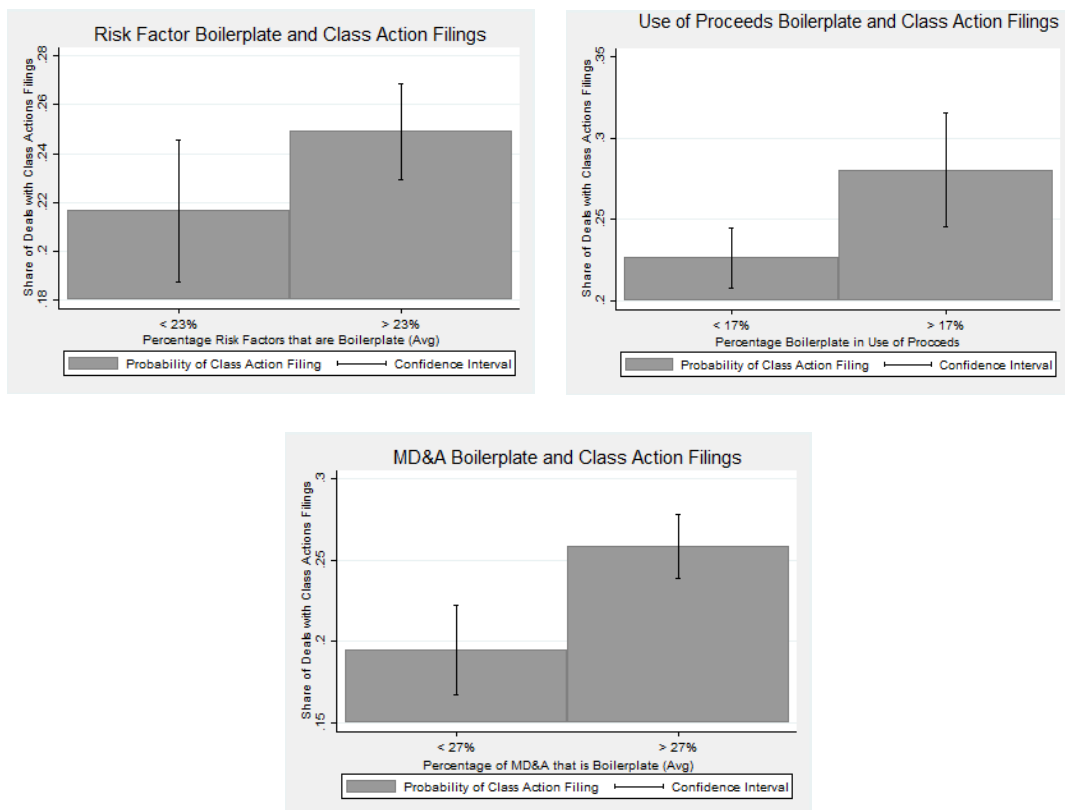
As in other parts of this Article, Table 4 shows the results of a linear regression of boilerplate on IPO-related class actions filed in federal courts within three years after a company's IPO.[184] The analysis shows a statistically significant relationship with respect to prospectus boilerplate and litigation: a 10% greater level of boilerplate is associated with between a 1.5% and 4% increase in the likelihood that an issuer will be subject to securities litigation.

---

184. Three years is the statute of limitations period for lawsuits brought under Sections 11 and 12 of the Securities Act. Securities Act of 1933, 15 U.S.C. §§ 77m (2012). It is the shortest statute of limitations period for the legal provisions that form the bases of most IPO litigation.

TABLE 4: LITIGATION AS A FUNCTION OF BOILERPLATE

Occurrence of Federal Securities Litigation Within Three Years of IPO	(1)	(2)	(3)	(4)	Probability Change per 10% Change in Similarity (5)
Overall Similarity	0.37***	0.10	0.28*	0.10	**1.00 to 3.70%**
	(0.11)	(0.07)	(0.16)	(0.07)	
Adj. R^2	0.46	0.28	0.33	0.59	
Number of Observations	1,751	2,317	1,971	2,317	
Rf Similarity	0.25**	0.16**	0.23**	0.24**	**1.60 to 2.50%**
	(0.11)	(0.05)	(0.09)	(0.10)	
Adj. R^2	0.50	0.36	0.38	0.60	
Number of Observations	1,749	2,317	2,329	2,317	
UP Similarity	0.02	0.18**	0.19**	0.26***	**2.60 to 4.30%**
	(0.20)	(0.09)	(0.08)	(0.05)	
Adj. R^2	0.51	0.60	0.59	0.59	
Number of Observations	2,331	2,317	2,332	2,317	
MD&A Similarity	0.24**	0.30***	0.30***	0.34***	**2.40 to 3.40%**
	(0.12)	(0.08)	(0.08)	(0.09)	
Adj. R^2	0.70	0.59	0.60	0.61	
Number of Observations	2,223	2,305	2,257	2,305	
Business Similarity	0.22	0.11	0.11	0.11	**1.90 to 3.90%**
	(0.30)	(0.32)	(0.32)	(0.35)	
Adj. R^2	0.50	0.48	0.61	0.40	
Number of Observations	2,327	2,314	2,327	2,314	
Industry FE	X	X	X	X	
IPO Year FE	X	X	X	X	
Industry * Year FE	X	X	X	X	
Litigation Target FE	X	X	X	X	
Law Firm FE	X				
Law Firm Market Share		X			
Bank FE			X		
Bank Market Share				X	

Marginal effects reported. Controls include fixed effects for the issuer's law firm (in the first specification), fixed effects for each lead underwriter (in the third specification), an indicator for whether the firm is a likely litigation target, the IPO year, the issuer's industry category and the interaction of industry and year. Issuer's law firm market share (in the second specification) and lead underwriter market share (in the fourth specification) are also included. Additional controls for issuer age (log), issuer total assets (log), deal proceeds (log), issuer's level of debt at the time of the IPO (log), venture capital involvement, debt to asset ratio, turnover (log), and prospectus wordcount (log) are included for all specifications but not tabulated. Robust standard errors reported in parentheses. Estimates marked with *, ** and *** are statistically significant at the 10%, 5% and 1% level respectively.

As in other parts of this Article, caveats are in order. The analysis does not demonstrate that these relationships are causal. For example, it could be the case that issuers who are already more likely

to face litigation rely on boilerplate to mitigate their risk. If that is the case, it is also possible that such issuers would have experienced more litigation had they not used the boilerplate that they did. I use a number of approaches to address these possibilities. I include a variable to indicate firms that are especially likely to be targets of securities litigation to control for firms' ex ante litigation risk, borrowing from other literature.[185] I also include controls for other indicia of firm quality and transaction quality, both of which affect the probability of litigation. This includes controls for company age, venture capital involvement, amount of proceeds, the size of the issuer (by assets), the issuer's amount of sales, and the amount of debt the issuer holds. I also control for the quality of the issuer's law firm and issuer's underwriter, as represented by their dollar market shares in IPO transactions in the preceding year, in addition to including underwriter and issuer law firm fixed effects in some specifications. These controls are related to the quality of the issuer as well as the quality of the disclosure. Fixed effects are also used for the issuer's industry, the IPO year, and the interaction of the two to account for trends in each, especially given that certain industries draw more litigation than others at different times. In addition, I perform an analysis using propensity score matching on samples of issuers, as discussed in the explanatory text preceding Appendix Tables F1 and F2. The analysis yielded statistically significant results (at or below the 10% level) consistent with those in Table 4 with respect to the Risk Factors, MD&A, and Use of Proceeds sections. Results for the entire document and for the Business Description were not statistically significant, which means they do not bolster, but do not necessarily refute, the results of the analysis above.

E. Discussion of the Results

The analyses lead to a number of conclusions but also raises more questions. First, the analyses show that, in the aggregate, boilerplate is associated with more indicia of information asymmetry, suggesting either that boilerplate tends to hinder investors from becoming informed about issuers or that its mere presence signals lower quality information about an issuer. Moreover, boilerplate is associated with higher costs to issuers: underpricing forces issuers to leave money on the table when deals are priced lower than what the market would

---

185. *See* Nelson & Pritchard, *supra* note 67, at 2, 7–10 (explaining the use of a variable indicating firms with a high risk of litigation and firms with a low risk of litigation). For purposes of this analysis, specific industry groups identified as being high risk for securities litigation are those with SIC codes 2833–2836, 2911, 3571–3577, 3612-3679, 4925, 4931, 4911, 4812–4813, 6162–6163, 6211-6289, 7370–7379, 8721, and 8731–8734.

have born, and higher litigation risk means higher legal fees and potential settlement costs. It is plausible that the boilerplate language itself creates these adverse consequences; however, even if that is not the case, at the very least the presence of a large amount of boilerplate provides a signal about the information asymmetry the market perceives regarding a company.

The results do not mean that all boilerplate in IPO disclosures is linked to higher costs or less efficient outcomes. Indeed, every prospectus in the dataset contains language that appears nearly the same form in other prospectuses. The important point is the general trend: the greater the level of copied language, the less information conveyed on average and the higher the costs for the issuer.

These conclusions have implications for boilerplate theory as it applies to securities disclosure. Although the results do not establish that any single theory explains the boilerplate phenomenon completely, they indicate strengths and weaknesses of various theories' explanatory power. Intuitively, one would imagine that boilerplate is efficient, perhaps the product of a learning externality that helps new lawyers draft quality documents and more seasoned lawyers negotiate them. If boilerplate is used to make deals more efficient, however, it does not seem to be a good value proposition from the perspective of the issuing companies. For instance, although boilerplate may be an efficient (and perhaps strategically vague) means by which to make disclosure, that efficiency (or strategy) comes at a price. If an issuer from the dataset pays, on average, $65,000 less in legal fees for each additional 10% of boilerplate disclosure in a whole prospectus (taking the average of the range in fee reduction from Appendix Table A), that savings will be overwhelmed by the loss an average issuer is likely to experience due to underpricing. For instance if the additional 10% boilerplate were found in the MD&A section, the transaction would be expected to lose, on average, between $5 to $6.2 million to underpricing.[186] In addition, the company would face extra litigation risk, increasing the expected settlement amount of a class action claim by approximately $600,000 to $850,000 on average, and possibly much more if the litigation went to trial or failed to settle within the typical range.[187] That amount includes

---

186. The average size of deals in the dataset is $106,000,000. This number would be multiplied by the increased level of underpricing that corresponds to risk factor boilerplate.

187. The average payment for settlement of securities class actions during the period of the study is approximately $25 million, while the median settlement amount is approximately $6 million. *See Securities Class Action Settlements 2015 Review and Analysis*, CORNERSTONE RES. 8–10 (2016), http://securities.stanford.edu/research-reports/1996-2015/Settlements-Through-12-2015-Review.pdf [https://perma.cc/B4NA-DMKS] (presenting data on all securities class action settlements from 1996 through 2015). A simple formulation of the average expected loss for a class action settlement would be the increase in probability (approximately 1.6% to 3.1%) multiplied by

only the actual settlement; one would still have to account for attorneys' fees, time costs, reputation costs, and the loss of value of the company's stock to assess the full impact of such litigation. It is possible that boilerplate merely signals low quality or high variance regarding an issuer that cannot be changed or mitigated, in which case perhaps such issuers would rather save money on legal fees if no amount of tailored disclosure would help. However, the analyses, all taken together, supports the conclusion that boilerplate has a role in affecting information asymmetry and, resultantly, costs to issuers, either through its information effects or through its signaling effects. Tailoring disclosure could therefore result in less information asymmetry and lower costs. Even assuming that some of the costs are justified and there are time savings from using boilerplate that are not captured through reduced advisory fees, it seems that boilerplate is not always worth its costs.

A question remains as to why high levels of boilerplate continue to appear in some deals, despite the SEC's efforts to reduce it and the costs that it carries with it. It is possible that at least some individual boilerplate provisions do facilitate communication or signal conformity with market norms. An analysis of the topics taken from the principal component analysis ("PCA") shows that some individual boilerplate topics are associated with indicia of less information asymmetry (i.e., more information flow). These results are not presented here, but they raise questions for future work. Although the topic model used here must be interpreted carefully given that PCA has no causal interpretation, it nonetheless raises the possibility that individual modules of boilerplate can convey information. But the clearest pattern in the data analyzed in this Article suggests that, in the aggregate, boilerplate is associated with negative information effects. If it has no use at all, one might conclude that market forces would reduce it, at least to the point that its benefits equal or outweigh its costs. But that does not seem to be the case.

One explanation for why large quantities of boilerplate are used despite potential drawbacks is that there are benefits from using it that are difficult to observe and measure. It is possible that boilerplate is used strategically by issuers, perhaps on the advice of underwriters or counsel; or issuers might use generic language to vary their chosen level of disclosure—in essence choosing to reveal less information while still complying with the letter of the law. The benefits of doing so might be to delay giving detail about negative information until the company is

---

the expected average loss ($25 million). Disclosure dollar loss—the amount of stock market value lost after a company discloses a securities class action settlement—would be even larger. *Id*. at 11.

in a better position, or to prevent competitors from obtaining sensitive information about the company's workings. If issuers are rational, then they might be willing to pay the price for these benefits. That explanation would be consistent with theories of strategic vagueness. Institutional investors may be indifferent if they are appropriately compensated for doing their own research through underpricing, which allows them to realize large short-term returns. For such investors, the disclosure is just the starting point of the process through which they will gather information from the issuer and elsewhere. It is difficult to see what benefit retail investors derive, however, since they pay full price in the market. Thus, if the law is concerned with protecting small retail investors, the equilibrium that allows so much boilerplate to persist is more problematic.

Another possibility is that very large amounts of boilerplate in a securities disclosure represent the capture of a deal's value by parties other than the issuer. Strategic vagueness could be driven by a firm's advisors—in particular, the underwriters—who lose little but stand to gain a lot if an IPO is underpriced. In particular, high levels of underpricing create substantial benefits for underwriters that easily outweigh any losses they suffer from forgone commissions.[188] The underwriters lose out on seven percent (the typical underwriting commission) of the underpriced amount. However, underwriters gain substantial benefits by allocating the underpriced stock to favored investors, who return that favor through future business and trading commissions.[189] Indeed, the underwriter might have an incentive to underprice the issue to a much higher degree than what the issuer should want because the underwriter captures the benefits from underpricing. Consequently, the underwriter can save time and effort

---

188. *See* Griffith, *supra* note 105, at 590–99 (evaluating methods by which underwriters can minimize risk and maximize reward, including underpricing).

189. *See* Ritter & Welch, *supra* note 13, at 1810 ("There is some evidence that underpriced share allocations have been used by underwriters to enrich buy-side clients in return for quid pro quos . . . ."); *see also* Griffith, *supra* note 105, at 593–94:

   [U]nderwriters may be able to increase profits above their base compensation by engaging in underpricing. This may seem contradictory since, as noted above, underwriter compensation is a percentage of aggregate offering proceeds, which are maximized by raising, not lowering, the offering price. However, underpricing creates an additional profit opportunity for underwriters by enabling the practice of spinning.;

Loughran & Ritter, *supra* note 13, at 8–9 (analyzing the factors that incentivize underwriters to underprice IPOs). The profitability of allocating underpriced stock to favored investors was a significant issue in the litigation surrounding the eToys IPO. *See* Joe Nocera, *eToys v. Goldamn Sachs: The Documents*, N.Y. TIMES 1–8 (Mar. 9, 2013), https://archive.nytimes.com/www.nytimes.com/interactive/2013/03/10/opinion/sunday/nocera-goldman-sachs-etoys.html?ref=Sunday [https://perma.cc/3G4V-XMEP] (Goldman Sachs internal documents reflecting compensation "owed" to the bank from investors receiving underpriced IPO allocations).

on due diligence by impelling the use of boilerplate while losing little (and perhaps gaining). Litigation is also less of a concern for underwriters because they have more defenses available to them than issuers do in suits based on misleading registration statements. The issuer's management, which relies on the underwriter and counsel for advice, may not realize the tradeoff being made. If that is true, then the intuitive and theoretical premise that boilerplate is efficient is indeed true—just not for the issuer. Rather, boilerplate's efficiency gains accrue to the underwriters, who can substitute due diligence for generic disclosure and pay few if any of the costs.

Finally, boilerplate may be the product of a network externality, which could be helpful or harmful. Lawyers or bankers might borrow from others within their network, free-riding on the drafting that others have done and avoiding having to reinvent the wheel. There is some evidence in the data that different boilerplate topics revealed by the topic model have diffused differently in different geographic regions, lending credence to the network externality explanation. The externalities from network effects may be positive or negative. For certain discrete types of disclosure, a standard form might provide value in deals and provide a positive externality. However, some boilerplate disclosures may be passed on without much critical revision. Since no one party internalizes the full benefit of refining general disclosures, there is no incentive to try to make them better or to think about them much at all. This, coupled with human tendency to gravitate toward the status quo and precedent, might explain the perpetuation of boilerplate even when too much of it can have harmful consequences.[190] If that is the case, then boilerplate represents another type of value transfer from issuers to their advisors. These consequences may be unrecognized and unintended, stemming from habitual practices of dealmakers who fear straying from the template provided by other successful deals.[191] Moreover, these precedents provide a way to lessen the cognitive cost of learning and complying with the SEC's complex requirements. Since the law firms drafting the

---

190. This would be the textual version of the cognitive bias known as the anchoring effect—the formation of a belief about the value of something, based on a specified initial value, regardless of whether that initial value is salient or entirely irrelevant. *See* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 14 (Daniel Kahneman et al. eds., 1982) ("In many situations, people make estimates by starting from an initial value that is adjusted to yield a final answer. . . . [D]ifferent starting points yield different estimates, which are biased toward the initial values. We call the phenomenon anchoring.").

191. *Cf.* Mark Weidemair, Robert Scott & Mitu Gulati, *Origin Myths, Contracts, and the Hunt for* Pari Passu, 38 L. & SOC. INQUIRY 72, 96 (describing lawyers using boilerplate language acting as "custodians of some ancient and sacred document—one whose inscrutable text they would not dare to alter").

disclosure do not internalize the cost of using it, they may not always have incentives to refine it. The SEC's complex regulations may, ironically, be partly responsible for all the boilerplate; at the same time, it is an area where energetic SEC action would help. Addressing boilerplate may be both simpler and harder than it would be if it were primarily used strategically. It would be simpler because there is no party with a particular vested interest in using it; it would be more difficult because human habits can be hard to change, even when the humans know they should change them. Moreover, given the SEC's other priorities, boilerplate may end up low on its list of enforcement or rulemaking activities.

In sum, the data points to a deleterious role for boilerplate in IPO disclosure, at least in the aggregate. However, the analysis in this study does not definitively determine why boilerplate persists or how it fits with theory, but it does shed light on some of the main possibilities, leaving questions for further inquiry. A better understanding of the role that boilerplate plays in the interaction between issuers, investors, bankers, and lawyers might help to guide dealmakers and policymakers in future transactions.

IV. FURTHER ANALYSIS

For any lawyer who has spent late nights crafting a well-tailored narrative from seemingly lifeless boilerplate, wondering if the exercise was worthwhile, the evidence from the past few decades of IPOs indicates that it often is. The results also indicate that the SEC is justified in trying to limit boilerplate, and courts are right to be skeptical of boilerplate disclaimers, because high levels of such generic language are associated with deals in which information is not conveyed to investors very well. But despite the best efforts of the profession and the SEC to regulate boilerplate, it continues to be an ever-growing staple of securities disclosure. This Part describes the implications that this study supports and makes suggestions for the SEC's reform efforts.

A. Disclosure Tradeoffs

One possibility that these findings raise is that, to some degree, good disclosure is traded off against IPO price to the initial investors. That tradeoff represents a process with some characteristics of a tacit bargain between investors (at least some of the institutional ones) and securities issuers (or their advisors). Boilerplate is important in such a process because it serves as a substitute for specific disclosure and costly information gathering. Whether it is true that issuers make a

strategic choice to withhold information or willingly pay a price, or whether investment bankers and lawyers make that choice for them, the tradeoff between boilerplate and value lends support to the theory developed in the contract boilerplate literature that cut-and-paste language is the product of an implicit bargain and might therefore be efficient in ways that are not salient in the data.

That possibility, if true, has implications for securities regulation. The idea that this type of implicit bargain affects securities disclosure is related to a broader debate in securities law about whether disclosure should be mandated by the government or whether investors and issuers should have more autonomy to decide what is disclosed.[192] To briefly outline this debate, those in favor of mandatory disclosure posit that issuers are unlikely to disclose sufficient information without being forced to do so.[193] This is true because the substantial internal costs of producing disclosure will almost always be greater than the amount of benefit the issuer will capture from such efforts.[194] The costs to the issuer arise not only as internal operational costs, such as the fees to advisers and management's time costs, but also by divulging sensitive information—for example by disclosing information that allows a supplier to negotiate for higher prices or by revealing to

---

192. *See* Stephen J. Choi & Andrew T. Guzman, *National Laws, International Money: Regulation in a Global Capital Market*, 65 FORDHAM L. REV. 1855, 1878 (1997) (claiming that differing national securities regimes across many countries give investors the best information about issuers); Fox, *supra* note 101, at 1340–62 (arguing that government-mandated disclosure is necessary); Jonathan R. Macey, *Administrative Agency Obsolescence and Interest Group Formation: A Case Study of the SEC at Sixty*, 15 CARDOZO L. REV. 909, 928 (1994) ("As markets have become more efficient, society's need to devote resources to support a statutory regime of mandatory disclosure designed and enforced by the SEC has disappeared. Any information that was supplied by the force of law now is supplied by the marketplace."); Roberta Romano, *Empowering Investors: A Market Approach to Securities Regulation*, 107 YALE L.J. 2359, 2374 (1998) (arguing that the relationship between disclosure and share price creates powerful incentives for issuers seeking new funds to disclose). Although many of the central works in this debate come from the era before Dodd-Frank, the debate continues in a variety of contexts. *See, e.g.*, Kevin S. Haeberle & M. Todd Henderson, *Making a Market for Corporate Disclosure*, 35 YALE J. ON REG. 383 (2018) (arguing that market forces, rather than government regulation, should be allowed to determine what information issuers produce).

193. *See* John C. Coffee, Jr., *Market Failure and the Economic Case for a Mandatory Disclosure System*, 70 VA. L. REV. 717, 745 (1984) (illustrating that in the municipal bond market, which is exempt from SEC disclosure, critical information is not being disclosed to investors); Frank H. Easterbrook & Daniel R. Fischel, *Mandatory Disclosure and the Protection of Investors*, 70 VA. L. REV. 669, 672–73 (1984) (arguing that mandatory disclosure might be the best of any alternative); Fox, *supra* note 101, at 1361 (arguing that in a world where issuers choose their disclosure tactics, investors "will not be as well informed as [they] would be if all issuers were compelled to disclose at the higher level that some issuers choose voluntarily"); Joel Seligman, *The Historical Need for a Mandatory Corporate Disclosure System*, 9 J. CORP. L. 1, 6–8 (1983) (claiming that critics of mandatory disclosure fail to take into account evidence that persuaded Congress in 1933 and 1934 concerning securities fraud and excessive underwriter compensation).

194. *See* Fox, *supra* note 101, at 1344–45 (illustrating via a graph that the cost to managers of disclosing is higher than the benefits the company receives).

competitors that barriers to entry in an industry are low.[195] These costs will be greater to an issuing firm than any benefit the issuer captures. However, such disclosures are socially optimal because society as a whole does capture some benefits, making the overall cost-benefit analysis favor disclosure that would not happen without regulation.[196]

Proponents of allowing issuers freedom to disclose as they please argue that markets are capable of forcing issuers to make adequate disclosure more efficiently than government regulators.[197] The rationale is that efficient markets readily reward issuers who disclose information the market finds valuable and punish firms that do not.[198] Another variation on their argument is that a firm's decision to be fully transparent is itself a signal of quality that will be rewarded in the markets, while firms that do not disclose signal low quality and drive investors away.[199] Thus, under most circumstances, firms will have strong incentives to be transparent. This view of capital markets regulation gives investors an important role: market-driven regulation would only work if investors are savvy enough to reward firms that disclose and punish firms that do not.

The analysis of boilerplate language supports both accounts but also shows how they are incomplete. On the one hand, even mandatory disclosure rules cannot ensure that issuers disclose information completely. The use of boilerplate is evidence of the way in which issuers can make disclosures that technically comply with regulations, but fail to provide specific or probing information. On the other hand, this means that issuers already have some choice about how much disclosure to make under the mandatory regime (even if that choice is limited).

Viewed through the lens of the securities disclosure debate, boilerplate demonstrates that even mandating disclosure is not always a complete solution—there are ways to comply without being informative. But it also demonstrates that investors are capable of

---

195. *Id.* at 1345.

196. *See id.* at 1346 (arguing that as a result of interfirm costs, a manager will never choose the socially optimal level of disclosure).

197. *See* Romano, *supra* note 192, at 2374–75 (arguing that information production in less-regulated European markets is no less efficient than the U.S. mandatory-disclosure regime).

198. *See id.* at 2374 ("Because firms need capital and investors need information, firms have powerful incentives to disclose information if they are to compete successfully for funds against alternative investment opportunities.").

199. *See* Stephen A. Ross, *Disclosure Regulation in Financial Markets: Implications of Modern Finance Theory and Signaling Theory*, *in* ISSUES IN FINANCIAL REGULATION 177, 183–93 (Franklin R. Edwards ed., 1979) (describing signaling theory as the proposition that issuers who openly choose to disclose news to the market will signal the company's quality, while investors will infer from those that do not disclose that the company has lower worth).

providing incentives for companies to disclose and garnering compensation when disclosures are incomplete. It is less clear, however, how well market mechanisms price vague disclosures. It is not possible from the analysis above to disaggregate overall boilerplate levels from specific boilerplate disclosures to know if there is a differential impact. Doing so is an area ripe for further research.

B. Regulating Boilerplate

As the SEC tries to streamline disclosure, it walks a fine line between regulating in a way that simplifies disclosure while also preserving the value of disclosure's content. Members of the Commission have expressed a willingness to decrease disclosure requirements, emphasizing that more disclosure is not necessarily better,[200] but investor groups are divided over which disclosures are useful and which can be eliminated.[201] And despite the call for streamlined disclosure rules, the SEC has repeatedly espoused the view that "better disclosure is not at all synonymous with less disclosure."[202] This leaves the SEC in a difficult position as it attempts to regulate disclosure.[203] Volume of disclosure is a problem, but investors want more disclosure anyway.

The solution that the SEC proposes involves two basic approaches: tailoring disclosures to the needs of different investor groups[204] and giving issuers more flexibility in choosing what to disclose.[205] The SEC has also suggested that issuers should be allowed

---

200. *See* Keith F. Higgins, Dir., Div. of Corp. Fin., Sec. & Exch. Comm'n, Disclosure Effectiveness: Remarks Before the American Bar Association Business Law Section Spring Meeting (Apr. 11, 2014), http://www.sec.gov/News/Speech/Detail/Speech/1370541479332#. VPUKkeF0ecE [https://perma.cc/UC3R-7G22]; *see also* Henry T.C. Hu, *Too Complex to Depict? Innovation, "Pure Information," and the SEC Disclosure Paradigm*, 90 TEX. L. REV. 1601, 1652 (2012) (claiming that the volume of disclosures mandated by the SEC make them "very hard to evaluate" given the large number of pages per filing).

201. *See, e.g.*, Higgins, *supra* note 200 ("Investors in different securities also might have different needs.").

202. Manuel F. Cohen, Chairman, Sec. & Exch. Comm'n, Remarks Before the Nineteenth Annual Conference of the Financial Analysts Federation (May 24, 1966) (emphasis omitted), https://www.sec.gov/news/speech/1966/052466cohen.pdf [https://perma.cc/8QME-9Y9D].

203. The paradox has been explored at length by Professor Roberta Karmel. *See* Karmel, *supra* note 93, at 828 ("What may be considered 'disclosure overload' for one investor group could simultaneously be regarded by another as insufficiently informative.").

204. *See* Kara M. Stein, Comm'r, Sec. & Exch. Comm'n, Remarks to the Council of Institutional Investors (May 8, 2014), https://www.sec.gov/news/speech/2014-spch050814kms [https://perma.cc/ DV57-BYP6] ("But, with rapid shifts in technology, the rise of increasingly large and complex businesses, and a growing understanding of our connection to each other and the planet, investors' needs and expectations have changed.").

205. *See* Higgins, *supra* note 200 (positing whether "companies [should] have flexibility to determine how they can convey information more effectively").

to experiment with different manners of disclosure.[206] With respect to both approaches, companies would be required to assess the materiality of the information, since that is the standard by which liability for securities fraud is assessed.[207] The challenge with either approach is that materiality is typically assessed from the point of view of the "reasonable" investors,[208] which makes the standard amorphous and sometimes difficult to apply ex ante. The amorphousness of the standard is at least partly responsible for the problem of too much disclosure—especially boilerplate disclosure—in the first place.

The market's reaction to past disclosures potentially provides the SEC with an additional tool for navigating this thicket. The topics uncovered by the topic model described in Part II provide a potential guidepost for giving investors flexibility and allowing for tailored disclosure, at least with regard to much of the boilerplate or pro forma disclosure. This is because it reveals patterns in the boilerplate disclosure that allow one to assess how useful or superfluous it really is to market participants, assuming that discrete selections of boilerplate can be useful even when an avalanche of such language is not.

Less useful types of disclosure that are frequently made using rote boilerplate could be standardized and incorporated by reference to a central website of generally applicable disclosures. It might also be possible for the SEC to exempt such topics from the disclosure requirements, although the Commission may be reluctant to exclude them altogether given the cautious approach it usually takes with respect to major changes. In any event, removing such language from the main document would allow prospectus drafters to focus disclosure on information that is more pertinent to the offering without losing any information that they feel the generic text might provide, while also allowing those who consume the text to focus on what is unique about a company. The length of prospectuses could be reduced, and investors who want to read the general disclosures would still have the ability to do so.

To illustrate, I note that the most common substantive topics in Appendix Table D fall into three broad categories. The first category consists of topics that would seem to a human reader to be obvious, generic, and convey little or no useful information. Even still, topics from the first category persist despite regulation. For example, Risk

---

206. *See id.*

207. *See* 15 U.S.C. § 77q (2012) (stating that an untrue statement of a material fact is considered securities fraud).

208. *See* TSC Indus., Inc. v. Northway, Inc., 426 U.S. 438, 449 (1976) (stating that a statement or omission is material "if there is a substantial likelihood that a reasonable shareholder would consider it important").

Factor Topic 1 (warnings that the issuer's new status as a public company will entail new compliance and other costs) would apply to the vast majority of IPO companies and does not offer useful information relative to other similar investments. Nonetheless, most issuers include this language, presumably out of caution. This conclusion is supported by the fact that the prevalence of this topic increases linearly throughout the time period covered in the data, despite the SEC's regulations, whereas the appearances of many others ebb and flow in sync with regulation. However, if such statements make the disclosure documents less useful for conveying more pertinent information, then it would be better for investors as well as issuers if the language could be safely left out.

The same is true for other types of disclosures, including the risk that a company may issue preferred stock in the future (Risk Factor Topic 3), the possibility that an active trading market for the company stock may not develop (Risk Factor Topic 8), and the fact that markets are volatile and that volatility increases litigation risk (Risk Factor Topic 12). These warnings are true for almost every company that goes public, and would be obvious to most investors. Yet, their persistence in the face of regulation suggests that issuers derive some value from these obvious statements or are too risk averse to shed them.

Such language could easily be standardized and provided on a central website. Issuers could incorporate these statements by reference, or perhaps more simply, these statements could become default disclosures unless issuers elected to vary them. In either case, the issuers would be deemed to adopt the generic statements unless they explicitly varied or disclaimed them.

The second category of topics laid out in Appendix Table D are broad statements that may only apply to some companies. Examples include statements about the risk that trade secrets might be obtained by competitors (Risk Factor Topic 5 and Business Description Topic 1), the risk that the company may have difficulty hiring and retaining highly skilled employees (Risk Factor Topic 6), or risks related to ownership concentration (Risk Factor Topic 7). This second category of disclosures could also be standardized, and issuers could incorporate them by reference. In addition, issuers would be encouraged to provide any specific information beyond the standard disclosure if there is anything material to add.

The third category of boilerplate contains general language relating to certain types of business models (such as companies that have large internet operations) or certain types of heavily regulated industries (such as biotechnology or pharmaceuticals). Examples from this category include issues related to government regulation by the

FDA (Business Description Topic 2), foreign regulators (Risk Factor Topic 4), or the level of competitiveness in the industry (Risk Factor Topic 2 and Business Description Topic 11). This third category deals with topics for which generic information is unlikely to apply broadly to most companies, but is very important for certain sets of companies. It may help for the SEC to know that such disclosures are routinely copied when more thorough information might be warranted. It is these types of disclosures that the SEC can focus on in reviewing prospectuses, for example, or in formulating mandates for more thorough company disclosure.

Perhaps as important, changes in boilerplate use provides information about what the market perceives as important to the reasonable investors, and how that changes over time. This would guide regulators when trying to prune disclosure regulations that may no longer be warranted but that have grown overly burdensome through years of accretion. In addition, to the extent the SEC wants to eliminate needless disclosures, analyzing the market impact (or lack thereof) of certain boilerplate disclosures can guide the SEC on what to eliminate, or at least where to begin asking market participants for input.[209]

C. The Future of Boilerplate

Language processing techniques like those used in the analysis above are becoming increasingly available in user-friendly formats. Given that fact, concerns about boilerplate language may soon become less important, because its presence will do less to hinder extraction of information from disclosure. Whatever its limitations, the analysis shows that boilerplate is likely persist in securities disclosure, and this may have drawbacks for issuers and investors. However, the emergence of computerized language processing as well as algorithmic trading and advisory services provide an opportunity to leverage the positive aspects of uniformity in disclosure.[210] The standardized nature of boilerplate makes it especially well suited to detecting latent patterns and signals that the presence or absence of tailoring provide. As text analysis and machine-learning modalities become more commonplace, user-friendly, and commercialized, boilerplate may in fact become extremely useful for sophisticated and ordinary investors alike. Investors who would ordinarily recoil at even the most detailed and

---

209. This analysis is beyond the scope of this Article, but is the subject of future work.

210. *See* Tom C.W. Lin, *Reasonable Investor(s)*, 95 B.U. L. REV. 461, 508–09 (2015) (arguing that the reasonable-investor standard is antiquated in the age of algorithmic trading).

thorough disclosure would have access to a digest of the information in even the driest boilerplate-filled prospectus.

To illustrate, it may be possible to learn specific information about companies from the boilerplate they use by comparing the company's boilerplate to that of other companies using the same language. For example, it may be the case that a risk factor about the inability of a company to deal with cyber-security issues conveys information that makes pricing more difficult because such boilerplate language provides no information regarding the specific problems a company faces. However, it may be the case that the presence of the risk factor by itself provides a signal about the company's challenges which, when compared with other companies using the same language, could provide meaningful information.

The precise way in which this might work is an area for further research, but in any event, boilerplate may be a less significant problem if an algorithm is distilling the text to its most important elements. These tools have the potential to help even relatively unsophisticated investors. Even in the absence of boilerplate, most securities disclosure is too complex for an investor with no finance or accounting training to digest meaningfully. Even the SEC has recognized that creating disclosure that is universally digestible by all investors is not realistic and should not be a goal of disclosure reform.[211] However, mechanical text processing tools can help to distill these complex documents and easily compare them to other similar disclosures in the market. Boilerplate would facilitate this process, and thus may ironically help average investors more than it harms them. Although few investors have access to complex machine-learning tools, that is likely to change.[212]

CONCLUSION

Boilerplate is ubiquitous in securities disclosure, and the evidence from this Article shows that its value is different from what might be assumed. Although it may not be efficient in the aggregate, it

---

211. *See* Higgins, *supra* note 200 ("While an individual investor may feel overloaded—and a bit overwhelmed—with information in a periodic report, other investors have said there is not a 'part of the disclosure pie that goes uneaten.' ").

212. For instance, products are widely marketed to help lawyers assess the ambiguity and readability of complex documents. *See, e.g.*, INTELLIGIZE, www.intelligize.com/products/transactions (last visited Sept. 18, 2018) [https://perma.cc/UV8E-DBJK] (describing a product that compares SEC filings to look for significant deviations from the norm); LEXCHECK, http://www.lexcheck.com/#precise_documents (last visited Sept. 18, 2018) [https://perma.cc/V5WX-EMKM] (describing a product for the legal market that uses natural language processing and machine learning to look for ambiguous language and poor drafting in contracts).

might perform a useful function by allowing issuing companies (and their advisers) to choose the specificity of the disclosures they make, while remaining in compliance with the SEC's regulations. Issuers may pay a price for using too much boilerplate disclosure, but this tradeoff can be seen as the result of a larger bargain with investors, in which issuers may find the price for vague disclosure worth paying. It remains an open question whether issuers are making this tradeoff wittingly or not. Nonetheless, boilerplate's role helps to explain its persistence. But the fact that boilerplate is here to stay may not be such a bad thing in the context of securities disclosure. Given that automated analysis of documents is becoming more commonplace and accessible, boilerplate might prove especially useful. Its standardized nature might ironically make it easier to digest than other kinds of disclosure, facilitating comparisons among deals and providing signals about what information is vague, what information is specific, and what information is important to investors. Thus, boilerplate can fit well into the SEC's disclosure reforms if it is able to leverage investors' preferences to better tailor disclosure mandates.

The results in this Article provide support for the conclusion that boilerplate—defined as language that is copied from one deal to the next—is related to information asymmetry and may obscure information in the aggregate. This reality appears to have become part of the fabric of the securities marketplace. The fact that the presence or absence of standardized language can be predictive of informativeness is useful nonetheless. If such signals can be readily interpreted by mechanical processes, boilerplate can communicate something by its presence, if not by its content. Thus, despite its potential to obscure, it is a source of information that should not be overlooked when regulators reconsider disclosure laws.

APPENDIX

SUMMARY STATISTICS

	Mean	Median	25th percentile	75th percentile	N
	(1)	(2)	(3)	(4)	
Deal Proceeds ($)	107,000,000	60,000,000	34,100,000	102,000,000	2,751
Company Assets (S)	539,758,000	134,398,000	56,936,000	358,865,000	2,751
Company Age (years)	12	7	<1	157	2,751
Legal Fees ($)	850,604	596,098	350,000	1,200,000	2,743
Prospectus Length (words)	101,725	71,266	45,766	100,473	2,751
S-1 Amendments (number)	7.6	7	5	10	2,751
Amount of Prospectus Amended (%)	84%	84%	76%	91%	2,751
Overall Prospectus Similarity (%)	47%	42%	35%	57%	2,617
Risk Factors Similarity (%)	32%	32%	23%	41%	2,617
Use of Proceeds Similarity (%)	23%	23%	17%	29%	2,617
MD&A Similarity (%)	34%	35%	27%	42%	2,617
Business Description Similarity (%)	15%	13%	9%	17%	2,617
Overall Gunning Fog Score	17.9	16.8	13.6	19.9	2,751
Risk Factors Gunning Fog Score	18.1	17.3	16.3	18.6	2,751
Use of Proceeds Gunning Fog Score	15	14.7	13.2	18.6	2,751
MD&A Gunning Fog Score	15.9	15.0	13.6	16.7	2,751
Business Description Gunning Fog Score	17.1	16.5	14.7	19.4	2,751

APPENDIX TABLE A: LEGAL FEES, ACCOUNTING FEES,
UNDERWRITER DISCOUNT, TRANSACTION TIMING,
AND BOILERPLATE

	(1)	(2)	Change per 10% Change in Similarity (Antilog * 0.1 * Average Fees) (3)
Panel A:Total Legal Fees for All Counsel (natural log)			
Overall Similarity	-0.33**	-0.35**	**$42,000 to $84,000 lower fees**
	(0.15)	(0.16)	
Gross Proceeds (log)	0.22***	0.23***	
	(0.03)	(0.03)	
Adj. R²	0.66	0.67	
Number of Observations	2,617	2,617	
Panel B: Total Accounting Fees (natural log)			
Overall Similarity	-0.19	-0.10	
	(0.19)	(0.21)	
Gross Proceeds (log)	0.16***	0.17***	
	(0.03)	(0.04)	
Adj. R²	0.58	0.60	
Number of Observations	2,612	2,600	
Panel C: Total Underwriting Spread (percent)			
Overall Similarity	0.08	0.20	
	(0.18)	(0.15)	
Gross Proceeds (log)	0.19***	0.26***	
	(0.04)	(0.03)	
Adj. R²	0.35	0.36	
Number of Observations	2,601	2,601	
Panel D: Time to Completion (number of days)			
Overall Similarity	-1.39	-0.65	
	(22.80)	(25.35)	
Gross Proceeds (log)	-10.72***	-8.62**	
	(2.79)	(4.04)	
Adj. R²	0.11	0.22	
Number of Observations	2,616	2,616	
Industry * Year FE	X	X	
Bank FE	X	X	
Auditor FE		X	
Law Firm Market Share		X	

Controls include deal proceeds (log) and fixed effects for each lead
underwriting bank, the IPO year and industry, and the interaction of these

two sets. The second specification for each analysis includes a fixed effect for the issuer's auditor, as well a control for the dollar market share of the issuer's law firm in the IPO market. Additional controls for the issuer age (log), issuer total assets (log), venture capital involvement, and prospectus wordcount (log) are included for all specifications but not tabulated. Robust standard errors reported in parentheses. Estimates marked with *, **, and *** are statistically significant at the 10%, 5%, and 1% level respectively.

APPENDIX TABLE B: LAWYER DEALS AND BOILERPLATE

Dependent Variable: Change in Boilerplate for Each Additional Deal by Law Firm in Relevant Time Period			
	(1)	(2)	(3)
Number of Deals in the Past Year	0.002** (0.001)		
Number of Deals in the Past Two Years		0.001 (0.001)	
Number of Deals in the Past Three Years			0.000 (0.001)
Industry FE	X	X	X
IPO Year FE	X	X	X
Industry * Year FE	X	X	X
Adj. R^2	0.35	0.35	0.34
Number of Observations	2,605	2,605	2,605

Controls include fixed effects for the IPO year and industry and the interaction of these two sets. Additional controls for the deal proceeds (log), the issuer age (log), venture capital involvement, syndicate size, and prospectus wordcount (log) are included for all specifications but not tabulated. Robust standard errors reported in parentheses. Estimates marked with *, **, and *** are statistically significant at the 10%, 5%, and 1% level respectively.

APPENDIX TABLE C: AMENDMENTS, AMENDED TEXT, AND SEC
COMMENTS

Panel A: Number of Amendments (log)		
	(1)	(2)
Overall Similarity	-0.47	-0.45
	(0.67)	(0.68)
Gross Proceeds (log)	0.39***	0.35**
	(0.11)	(0.13)
Adj. R^2	0.20	0.25
Number of Observations	2,549	2,549
Panel B: Amount of Text Amended (percent)		
Overall Similarity	-0.004	-0.02
	(0.02)	(0.02)
Gross Proceeds (log)	0.004*	0.01
	(0.00)	(0.00)
Adj. R^2	0.12	0.11
Number of Observations	2,551	2,551
Panel C: Quantity of SEC Comments (log)		
Overall Similarity	-0.745	0.03
	(0.91)	(0.09)
Gross Proceeds (log)	0.02	-0.70
	(0.95)	(1.25)
Adj. R^2	0.15	0.30
Number of Observations	804	804
Industry FE	X	X
IPO Year FE	X	X
Industry * Year FE	X	X
Bank FE	X	
Law Firm FE		X

Controls include deal gross proceeds (log), fixed effects for each lead
underwriting bank, the IPO year and industry, and the interaction of these
two sets. Fixed effects for each issuer law firm are included in the second
specification. Additional controls for issuer age (log), issuer total assets (log),
and prospectus wordcount (log) are included for all specifications but not
tabulated. Untabulated analysis for individual Risk Factors, MD&A, Use of
Proceeds, and Business sections were not statistically significant. Robust
standard errors are reported in parentheses Estimates marked with *, **,
and *** are statistically significant at the 10%, 5% and 1% level respectively.

APPENDIX FIGURE 1: RELATIONSHIP BETWEEN DEAL SIZE (AS GROSS PROCEEDS) AND BOILERPLATE (AS SIMILARITY SCORE)

Boilerplate and Deal Size
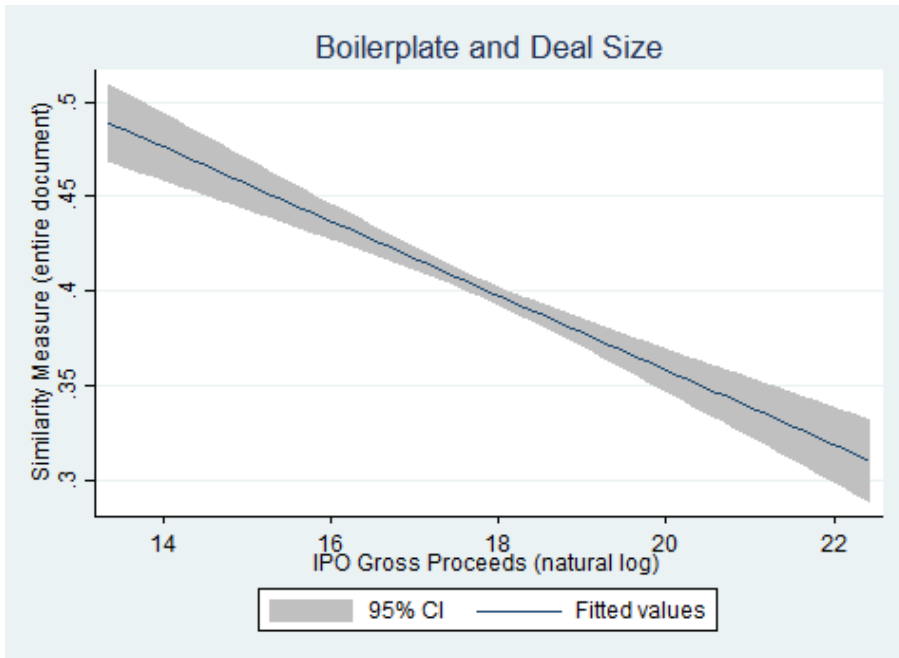
APPENDIX FIGURE 2: AVERAGE LAW FIRM DEALS PER YEAR AND
BOILERPLATE (TOP) AND AVERAGE LAW FIRM EQUITY CAPITAL
MARKETS MARKET SHARE BY YEAR (BOTTOM)
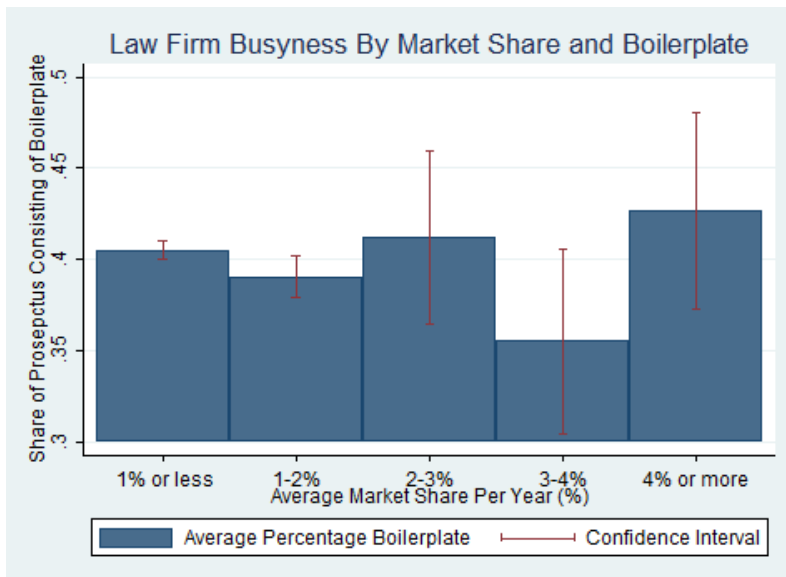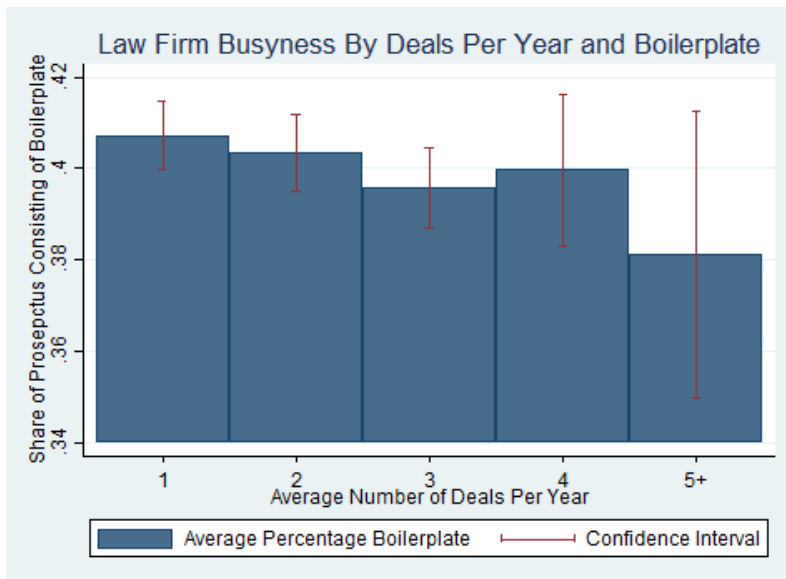
		Appendix Table D: Boilerplate Topics for Each Prospectus Section		
Topic Number	**Use of Proceeds**	**Risk Factors**	**MD&A**	**Business Description**
1	Purpose of the offering is to increase available working capital	Being a public company will incur significantly greater legal/accounting/ compliance expenses	Statements that the company's financials must rely on estimates/ assumptions/ historical experience	The company relies on proprietary rights, which unauthorized parties may obtain
2	Purpose of the offering is to create a public market in the company's stock/increase its visibility	It may be difficult to compete with more established competitors/ competitors have longer histories and more resources	Statements that the company's systems are being updated for cybersecurity/ new technology needs	The company's products require FDA or foreign regulatory approval and/or clinical trials
3	Amounts and timing of actual uses depend on numerous factors such as the company's marketing/ research/revenues	The company may issue preferred stock in the future that will impact the rights of common stock	Recitations regarding financial instruments and guarantees	None of the company's employees are covered by collective bargaining agreements
4	Framing language: the company estimates an amount of net proceeds based on assumed IPO price	Government regulation may impair the company's revenues/ability to raise new money/ ability to obtain licenses/limit expansion	Statements that financial data in the section is derived from audited/ unaudited financial statements	The company is involved in legal proceedings from time to time in the ordinary course of business
5	Management/the Board of Directors retains broad discretion to spend proceeds	Possible loss of proprietary rights/trade secrets	Recitations regarding derivative instruments and hedging activities	No current legal proceedings but one may arise in the ordinary course of business
6	Pending other uses, the proceeds will be invested in short term interest-bearing securities	Difficulty finding and retaining skilled employees	Recitations regarding variable interest entities/exit disposal activities: Fin 46; SFAS 146	The company has never experienced work stoppages/ employee relations are good
7	Proceeds may be invested in complementary business or technologies	Concentration of ownership/owner lockup agreements will expire and sales will cause price to decline	Statements regarding software capitalization costs	Statement that historical results are not necessarily indicative of future periods
8	Purpose of the offering is to	An active market for the shares may	Framing language regarding	Litigation may be necessary to

	attract employees and facilitate future access to the capital markets	not develop, making the shares less valuable and hard to trade	income/revenue and expenses from administration, depreciation, amortization, and operations	enforce the company's intellectual property rights/ trade secrets
9	The Board has discretion applying proceeds	Management may issue more shares in the future, causing dilution	Recitations regarding income tax accounting uncertainty	Litigation could result in costs/take management resources/adversely affect operations
10	Proceeds will be invested in income-producing investments	The company has never, and may never, declare dividends	Language regarding year-on-year comparisons, interest income, and expense	Litigation could be expensive and might not succeed
11	Proceeds will be used to redeem preferred stock/repay promissory notes issued by the founders	Antitakeover provisions/ ownership structure may deter changes in control	Recitations regarding accounting impairment for disposal of long lived assets	General statement that there is significant competition in the industry
12	Pending other uses, proceeds will be invested in U.S. government securities	Market price could decline/the company could be subject to class action if prices are volatile	Statement that accounting impairment will have no material impact on the company	The company relies on trademarks, copyright laws, trade secrets, and uses contractual restrictions and nondisclosure agreements

APPENDIX TABLE E: BID-ASK SPREADS AND BOILERPLATE

	Bid-Ask Spread on the First Day of Trading	**Change in Bid-Ask Spread After Thirty Trading Days**	**Change in Bid-Ask Spread After Sixty Trading Days**
	(1)	(2)	(3)
Overall Similarity	0.04	-0.07	-0.02
	(0.02)	(0.05)	(0.04)
Adj. R^2	0.15	0.12	0.33
Number of Observations	2,380	2,263	1,971
Rf Similarity	0.15***	-0.11**	-0.10**
	(0.03)	(0.04)	(0.04)
Adj. R^2	0.16	0.12	0.15
Number of Observations	2,464	2,342	2,305
UP Similarity	0.15***	- 0.81*	- 0.06
	(0.03)	(0.04)	(0.05)
Adj. R^2	0.15	0.12	0.15
Number of Observations	2,464	2,342	2,305
MD&A Similarity	0.10***	- 0.25	-0.003
	(0.03)	(0.04)	(0.04)
Adj. R^2	0.50	0.11	0.15
Number of Observations	2,455	2,322	2,296
Business Similarity	0.10***	- 0.79*	-0.07*
	(0.03)	(0.04)	(0.04)
Adj. R^2	0.15	0.12	0.15
Number of Observations	2,459	2,337	2,300
Industry FE	X	X	X
Year FE	X	X	X
Industry * Year FE	X	X	X
Bank FE	X	X	X

The Table shows results of an OLS regression in which the average bid-ask spread on the first day of trading (column 1), after thirty days of trading (column 2) and after 60 days of trading (column 3) is the dependent variable and boilerplate is the independent variable. Controls include fixed effects for each lead underwriting bank, the IPO year and industry, and the interaction of these two sets. Additional controls for the issuer age (log), issuer size total assets (log), venture capital involvement, volatility and turnover (as turnover on the first day of trading for the specification in column one, and average daily turnover for the 30 and 60 days post offering date for the specifications in columns two and three) are included for all specifications but not tabulated. Robust standard errors are reported in parentheses. Estimates marked with *, ** and *** are statistically significant at the 10%, 5% and 1% level respectively.

EXPLANATORY NOTE FOR APPENDIX TABLES F1 AND F2

The following analysis uses propensity score matching to address the possibility that firms that self-selected to provide more boilerplate and less specific disclosure differed systematically from a control sample of firms that used less boilerplate and more specific disclosure in a way that drove the results in the regression analysis. Propensity score matching attempts to simulate an experiment by using characteristics to match issuers in a "treatment" sample to the most similarly situated issuers in a "control" sample in the data. The match is performed using issuers' observable characteristics that should in theory be similar with respect to unobservable qualities.

To create a propensity score matching sample, the first step is to determine which variables predict treatment. Treatment here is defined as greater use of boilerplate. Although boilerplate in this Article is a continuous variable, I construct an indicator treatment variable using the 25th percentile measure for the level of the boilerplate for each section of the disclosure to separate "high" boilerplate from "low" boilerplate issuers. For example, 23% boilerplate in a disclosure is the 25th percentile for boilerplate in the Risk Factors section, meaning that this level and above is used as the treatment group in Appendix Table E. The first step determines which variables predict that a firm will use more boilerplate. The second step compares the treatment firms with the matched sample. The analysis matches each treatment observation to one or more control observations that are similar along a number of covariates that are likely proxies for unobserved qualities of an issuer that would give rise to more or less boilerplate, all else equal. I generate a propensity score, and thus match observations, by estimating a logistic regression on the following covariates: (1) an indicator for industry category, based on two digit SIC industry categories; (2) the log of deal gross proceeds; (3) log of total assets; (4) book-to-market ratio; (5) log of the company age; (6) an indicator for the involvement of venture capital investors pre-IPO; (7) the dollar market share of each lead underwriter; (8) the return on assets ("ROA") for the year preceding the IPO; (9) the log of research and development expenditures for the year preceding the IPO; (10) earnings before interest, tax, depreciation, and amortization ("EBITDA") for the year preceding the IPO (normalized);[213] (11) total revenues for the year preceding the IPO (normalized); (12) an indicator for whether the

---

213. For financial data that cannot be normalized using a log transformation (i.e., because it takes both positive and negative values), the variable is normalized by taking its cube root.

company is a technology company; (13) an indicator for whether the company is in an industry that historically draws greater than average litigation; and (14) the offering date of the deal.

APPENDIX TABLE F1: COVARIATE MEANS AND VARIANCES FOR RAW AND MATCHED SAMPLES

Panel A: Treatment = Risk Factors Boilerplate > 0.23				**Panel B: Treatment = MD&A Boilerplate > 0.27**			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standardized differences		Variance Ratio		Standardized differences		Variance Ratio	
Raw	Matched	Raw	Matched	Raw	Matched	Raw	Matched
Gross Proceeds (log)							
-0.4280	-0.1033	0.5776	0.6895	-0.3844	-0.0166	0.4099	0.5746
Venture Capital Involvement							
0.0102	-0.0212	0.9847	1.0359	0.2452	-0.0253	0.6828	1.0385
Total Assets (log)							
-0.3739	-0.0822	0.6294	0.6927	-0.4514	-0.0434	0.4669	0.6039
Book to Market							
0.1957	-0.0658	0.2676	0.8072	-0.0256	0.0414	0.7399	1.2389
Offering Date							
-0.3700	-0.0683	1.7190	1.6739	-0.2787	0.1044	1.1988	1.4201
Industry Category							
-0.0798	0.0358	0.7822	0.8555	-0.0058	-0.0596	0.6649	0.7180
Company Age (log)							
0.0412	-0.0079	0.8113	0.8920	0.0260	0.0373	0.6569	0.7957
Underwriter Market Share							
0.0235	-0.1589	1.3440	0.60724	-0.0299	0.0597	0.7695	0.8852
Return on Assets							
0.0072	0.0071	0.6810	1.1034	0.0254	0.0223	0.7002	0.8157
EBITDA (standardized)							
-0.2696	-0.0255	0.6267	0.7680	-0.4150	0.0015	0.6682	0.8042
Tech Company Indicator							
0.3837	0.0511	1.0741	1.0108	0.6358	0.0591	1.0385	1.0093
R&D Expenses (log)							
-0.1790	-0.0456	1.8058	1.0500	-0.1466	-0.0493	1.3189	0.9747
Total Revenues (standardized)							
-0.3503	-0.0680	0.4012	0.5541	-0.4017	0.0275	0.4718	0.7763

Panel C: Treatment = Use of Proceeds Boilerplate > 0.17				**Panel D: Treatment = Business Description Boilerplate > 0.09**			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standardized differences		Variance Ratio		Standardized differences		Variance Ratio	
Raw	Matched	Raw	Matched	Raw	Matched	Raw	Matched
Gross Proceeds (log)							
-0.7583	-0.0376	0.4109	0.4564	-0.0482	0.0339	0.9075	0.9251
Venture Capital Involvement							
0.3414	0.0691	0.6209	0.9189	-0.1960	0.0488	1.2979	0.9373
Total Assets (log)							
-0.9162	-0.0402	0.5703	0.5379	-0.0220	0.0179	0.8439	0.8882
Book to Market							
-0.1528	0.0341	0.5267	0.7004	-0.1057	-0.0548	1.1470	1.1146
Offering Date							
-0.4989	-0.0316	1.8866	1.4025	-0.4210	-0.0139	1.1125	1.0992

Industry Category	-0.2260	0.0816	0.8483	1.2716	0.2361	-0.0010	0.9345	1.0014
Company Age (log)	-0.2868	-0.1022	0.5160	0.7814	-0.0447	-0.0845	1.2581	1.3091
Underwriter Market Share	-0.0166	-0.0573	0.6799	0.3463	-0.1060	0.0507	0.5058	0.8779
Return on Assets	-0.1016	0.0123	9.0650	3.8129	0.1104	0.0467	0.1782	0.2969
EBITDA (standardized)	-0.9846	-0.0774	0.7934	0.8093	-0.0924	0.0674	0.9646	1.0020
Tech Company Indicator	0.7753	0.0163	1.3085	1.0031	0.1323	-0.0295	1.0174	0.9954
R&D Expenses (log)	-0.1373	-0.0999	2.1645	1.9391	0.0149	0.0709	0.8555	0.6701
Total Revenues (standardized)	-0.9036	-0.0777	0.3508	0.6149	-0.1293	0.0096	0.6824	0.8614

Panel E: Treatment = Entire Document Boilerplate > 0.33

	(1)	(2)	(3)	(4)
	Standardized differences		Variance Ratio	
	Raw	Matched	Raw	Matched
Gross Proceeds (log)	0.3118	0.1979	1.3071	0.8976
Venture Capital Involvement	0.1348	0.2866	0.8245	0.7476
Total Assets (log)	0.3848	0.2305	0.9619	0.5903
Book to Market	0.0103	0.1249	2.1275	2.0749
Offering Date	1.5408	-0.1981	0.7968	0.7291
Industry Category	-0.0674	0.0561	1.1065	1.2199
Company Age (log)	0.3357	0.0647	1.0276	1.6458
Underwriter Market Share	0.0696	0.0217	1.3591	1.7715
Return on Assets	0.1079	0.1759	0.0930	0.2271
EBITDA (standardized)	0.5795	0.3710	1.0344	0.6458
Tech Company Indicator	-0.0300	0.1017	0.9948	0.9991
R&D Expenses (log)	0.2905	-0.0619	0.3361	1.1395
Total Revenues (standardized)	0.4115	0.0889	1.2373	0.6209

In an alternative, unreported specification, I use two-digit North American Industry Classification System codes to match industries and obtain consistent results. For all specifications, I force an exact match for industry category and use a running day-count variable to match on offering date. In a further alternative but unreported specification, I

force exact matches for industry and IPO year, instead of matching on the continuous offering date variable. This specification results in consistent estimates as well, although it is less theoretically justifiable since there is no reason to assume that an IPO at the end of one year is meaningfully different than one at the beginning of the following year.

The soundness of a propensity score matching method to estimate treatment effects depends upon the extent to which the control and treatment groups are balanced with respect to covariates. Balance indicates that the control and treatment groups would likely experience the same outcomes in the absence of treatment. To ensure validity of matching estimator, I analyze the covariate balance in two ways: by examining the matched sample means and by plotting the sample propensity score densities.

Appendix Table F1 compares covariate means between the raw and matched samples for the five different text selections from which treatment and control groups are constructed. Columns one, two, five, and six of the Table show that the specification leads to good balance on the covariates, given that the differences in standardized means of the matched sample is small to negligible; the differences in means of the raw samples are larger for most covariates. It also shows, in columns four and eight, that the variance ratios of the matched treatment and control sample are relatively close to one.

In addition, I visually compare the distribution of the propensity score for the given covariates for each of the five different types of treatment (i.e., greater than median boilerplate in the five text selections studied) before and after matching on the covariates. Appendix Figure 3 shows this distribution before and after matching. The similarity in the density of the two propensity scores after matching suggests that the two groups are balanced with respect to the propensity score.

APPENDIX FIGURE 3: BALANCE PLOTS OF DISTRIBUTION OF
COVARIATES BEFORE AND AFTER MATCHING
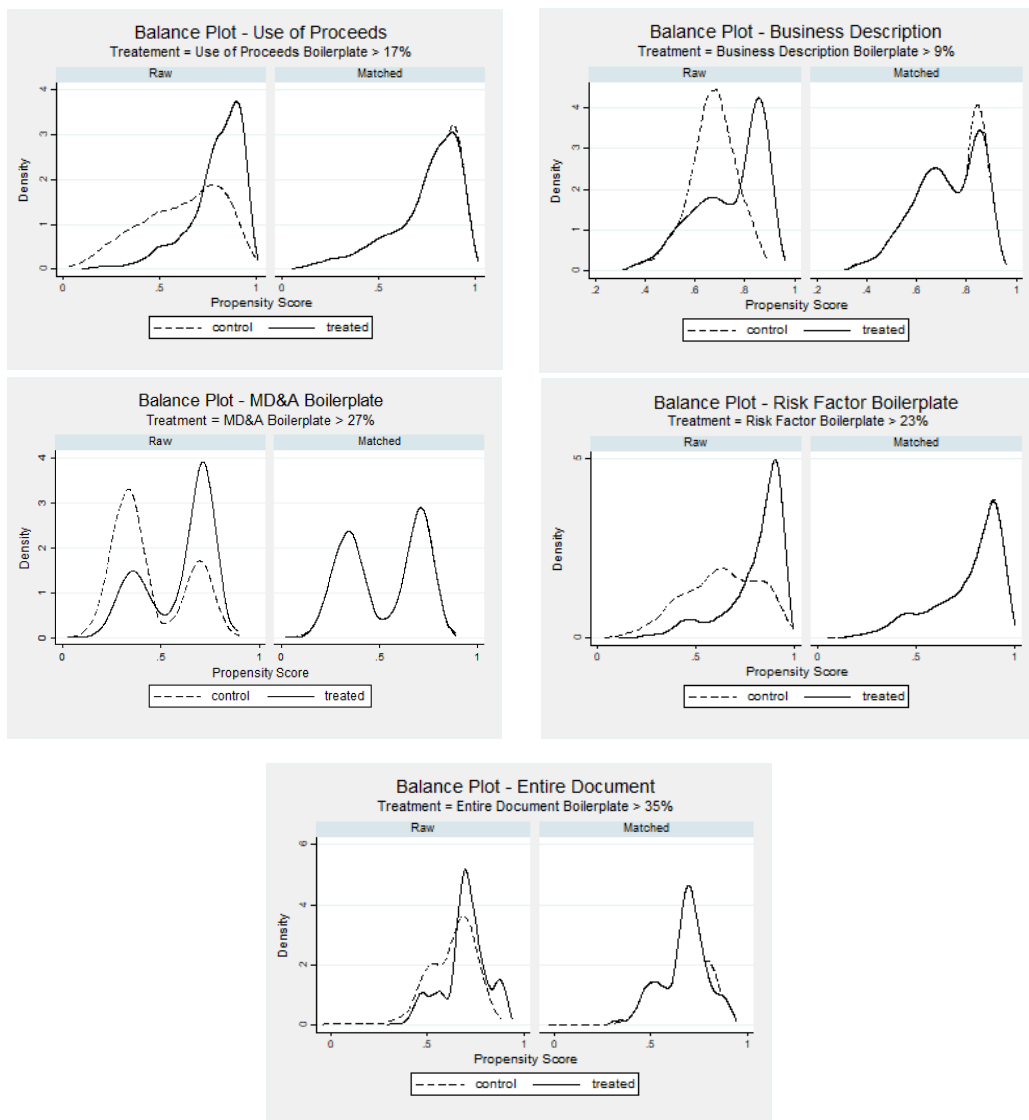
Propensity Score Balance Tests. In both left and right panels, the density of propensity scores is plotted for treatment groups (solid lines) and control groups (dashed lines), comparing the raw controls with the propensity-score matched observations.

As shown below, the consistency of the coefficient estimates in significance and magnitude strongly suggests that the results in the text are not driven by spurious variation in covariate balance.

Appendix Table E2 shows that the treatment effects on the outcomes of interest remain consistent with the regression analysis described in detail in the text. The analysis here employs propensity-score matching, and estimates the average treatment effect for the issuers with high boilerplate disclosure (defined as above median boilerplate for each section) by propensity score matching them with issuers with relatively lower levels of boilerplate that are comparable with regard to important characteristics. Z-scores based on robust standard errors are reported in parentheses.

APPENDIX TABLE F2: AVERAGE TREATMENT EFFECTS

	(1)	(2)	(3)	(4)	(5)
Average Treatment Effect for:	Underpricing	Price Revision	Litigation	Analyst Dispersion	Bid-Ask Spread
Overall Similarity	0.013 (0.57)	0.041* (1.86)	0.046 (0.75)	-0.019 (-0.94)	-0.005 (-0.79)
Number of Observations	1,847	2,317	1,923	1,907	2,271
Rf Similarity	0.106** (2.81)	0.060** (2.40)	0.076** (2.34)	0.173*** (4.22)	0.017* (1.79)
Number of Observations	2,277	2,305	2,314	1,904	2,252
UP Similarity	0.092* (2.34)	0.110*** (3.54)	0.027* (1.92)	0.042 (1.22)	0.002 (0.31)
Number of Observations	2,300	2,282	2,291	1,890	2,229
MD&A Similarity	0.161*** (8.47)	0.075** (2.86)	0.12*** (3.58)	0.080** (2.06)	0.015** (2.33)
Number of Observations	2,277	2,305	2,126	1,890	2,229
Business Similarity	0.007 (0.19)	-0.052** (-1.99)	-0.004 (-0.16)	0.011 (0.37)	-0.002 (-0.12)
Number of Observations	2,109	2,118	2,124	1,826	2,065

Z-scores based on robust standard errors are reported in parentheses. Estimates marked with *, ** and *** are statistically significant at the 10%, 5%, and 1% level respectively.

The results with respect to Risk Factors and MD&A sections are consistent with the OLS analysis and significant at the 10% level or better, with the exception of the relationship between MD&A boilerplate and class action litigation, which does not give a statistically significant result. Boilerplate in the Use of Proceeds section also yields a positive and statistically significant result (at the 10% level or better) for the underpricing, price revision and litigation dependent variables.

With respect to the Business Description section, boilerplate continues to bear a statistically significant relationship only to upward price revision, and the relationship is negative. With respect to the entire document, boilerplate bears a significant and positive relationship to upward price revision. For Risk Factors, MD&A, and, to a large degree, Use of Proceeds, the results here support those obtained in the regression analysis in the main text, while the results for the entire document and the Business Description section are inconclusive.

APPENDIX TABLE G: GUNNING FOG INDEX AND BOILERPLATE

Gunning Fog Score	(1)	(2)
Overall Gunning Boilerplate	**6.90***	**6.57***
	(3.00)	(3.02)
Gross Proceeds (log)	0.12***	0.08
	(0.04)	(0.05)
Adj. R²	0.21	0.30
Risk Factor Boilerplate	**1.32****	**1.00***
	(0.43)	**(0.50)**
Gross Proceeds (log)	0.23***	-0.02
	(0.04)	(0.04)
Adj. R²	0.471	0.57
Use of Proceeds Boilerplate	**-1.08**	**-1.55†**
	(0.72)	(0.82)
Gross Proceeds (log)	-0.08***	-0.23***
	(0.04)	(0.05)
Adj. R²	0.10	0.20
MD&A Boilerplate	**6.52****	**7.95****
	(2.93)	**(3.22)**
Gross Proceeds (log)	-0.28***	-0.13***
	(0.03)	(0.03)
Adj. R²	0.22	0.41
Business Boilerplate	**0.84***	**0.99***
	(0.37)	**(0.41)**
Gross Proceeds (log)	-0.01	-0.02
	(0.18)	(0.03)
Adj. R²	0.09	0.20
Industry FE	X	X
Year FE	X	X
Industry * Year FE	X	X
Law Firm FE	X	X
Bank FE		X
Number of Observations	2,451	2,451

Controls include deal proceeds (log) and fixed effects for each lead underwriting bank, the issuer's law firm, the IPO year and industry , and the interaction of these two sets. Additional controls for issuer age (log), issuer total assets (log), and prospectus wordcount (log) are included for all specifications but not tabulated. Robust standard errors are in parentheses. Estimates marked with †, *, ** and *** are statistically significant at the 10%, 5%, 1%, and 0.1% level respectively.

APPENDIX TABLE H: REPRESENTATIVE SAMPLE OF BOILERPLATE PHRASES

[] If Form filed register additional securities offering pursuant Rule b Securities Act please check following box list Securities Act registration statement number earlier effective registration statement offering

If Form posteffective amendment filed pursuant Rule c Securities Act check following box list Securities Act registration statement number earlier effective registration statement offering

That purpose determining liability Securities Act posteffective amendment shall deemed new registration statement relating securities offered therein offering securities time shall deemed initial bona fide offering thereof

The registrant hereby amends Registration Statement date dates may necessary delay effective date registrant shall file amendment specifically states Registration Statement shall thereafter become effective accordance Section a Securities Act Registration Statement shall become effective date Commission acting pursuant said Section a may determine

The underwriters initially propose offer shares common stock part directly public initial public offering price set forth cover page prospectus part certain dealers including underwriters price less concession excess per share

This table read conjunction financial statements notes thereto included elsewhere Prospectus Managements Discussion Analysis Financial Condition Results Operations"

SELECTED CONSOLIDATED FINANCIAL DATA The following selected consolidated financial data read conjunction Company\s Consolidated Financial Statements Notes thereto "Management\s Discussion Analysis Financial Condition Results Operations" included elsewhere herein

This Prospectus forms part Registration Statement contain information set forth Registration Statement exhibits schedules thereto

THIS PROSPECTUS DOES NOT CONSTITUTE AN OFFER TO SELL OR SOLICITATION OF AN OFFER TO BUY ANY SECURITY OTHER THAN THE SHARES OF COMMON STOCK OFFERED HEREBY NOR DOES IT CONSTITUTE AN OFFER TO SELL OR SOLICITATION OF ANY OFFER TO BUY ANY OF THE SECURITIES OFFERED HEREBY TO ANY PERSON IN ANY JURISDICTION IN WHICH IT IS UNLAWFUL TO MAKE SUCH AN OFFER OR SOLICITATION

This Prospectus contains forwardlooking statements involve risks uncertainties

The Company intends furnish stockholders annual reports containing audited financial statements reported independent auditors quarterly reports first three quarters fiscal year containing unaudited interim financial statements

The following summary information qualified entirety detailed information including "Risk Factors" Company\s Consolidated Financial Statements Notes thereto appearing elsewhere Prospectus

[_] CALCULATION OF REGISTRATION FEE PROPOSED PROPOSED TITLE OF EACH CLASS OF MAXIMUM MAXIMUM SECURITIES AMOUNT TO OFFERING PRICE AGGREGATE AMOUNT OF TO BE REGISTERED BE REGISTERED PER SECURITY OFFERING PRICE REGISTRATION FEE Common Stock $ par value

For six months ended June net cash used operating activities $ million

It currently estimated initial public offering price Common Stock $ $ per share

As of June the Company had cash and cash equivalents of $ million and working capital of $ million

In opinion financial statement schedule considered relation basic financial statements taken whole presents fairly material respects information set forth therein

This represents immediate increase net tangible book value $ per share Common Stock current holders Common Stock immediate dilution approximately $ per share new investors purchasing shares Offering

The results operations interim periods necessarily indicative results expected future periods

Net Loss Per Common Share Net loss per common share computed using weighted average number common shares outstanding year