ARTICLE

REGULATING SUBSTANCE THROUGH FORM: LESSONS FROM THE SEC'S PLAIN ENGLISH INITIATIVE

JEREMY R. MCCLANE*

TABLE OF CONTENTS

* Associate Professor, University of Connecticut School of Law. I would like to thank Amitai Aviram, Bernard Black, James Kwak, Robert Lawless, Kate Litvak, Jeffrey Mullins, Gillian Popodak, Peter Siegelman, Michael Sinkinson, Connor Raso, Verity Winship, participants at the 2016 Conference on Empirical Legal Studies, the University of Illinois Faculty Workshop, and reviewers at the Securities and Exchange Commission for valuable comments, criticisms and insights on this project. All errors are my own.

Mandatory disclosure is an important piece of government regulation in many spheres, including securities offerings. Regulators recognize that the form of disclosure matters significantly for how well the substance is conveyed. Thus, regulators like the Securities Exchange Commission ("SEC") that focus on disclosure spend time and energy thinking about form as well as substance. As I explain and support empirically in this Article, at least some of these "stylistic regulations"—regulations that govern the form of disclosure—can provide useful tools for improving how well substance is communicated, but at the same time, long-term compliance with these rules can be difficult to ensure.

I support this analysis with three contributions. First, I employ natural language processing methods to measure compliance with the SEC's "plain English" regulations for new issuances, and investigate how well these regulations have worked over time. I do so by creating a measure of plain English based on the SEC's rules. This measure of plain English goes beyond prior research, providing a more robust analysis of compliance with these rules and their potential effects Using a dataset of 2,255 initial public offering ("IPO") prospectuses and related documents, I find that the SEC's regulations increased the degree to which disclosure was written in "plain English," although the impact of the regulations has faded over time.

Second, I provide evidence that the style corrections mandated by the SEC corresponded to more readable disclosures, and that these disclosures were associated with proxies of better information supplied to the market regarding securities issuers. Moreover, I document that using plain English can be cost effective for issuers who must comply with the regulations.

Third, I demonstrate how language processing methods might impact stylistic regulation in the future. These methods can help the SEC to streamline its disclosure mandates, but they also demonstrate why disclosure style may become trivial in the future, because machines can process text and condense disclosure into easily comprehensible forms regardless of style. Nonetheless, until that technology becomes widely democratized, stylistic regulation will remain an important tool in securities law.

I. Introduction

Mandatory disclosure is a cornerstone of governmental intervention in numerous domains,[1] and this is especially true in securities regulation.[2]

[1] A necessarily small selection of timely examples includes: mandatory disclosures by pharmacy benefit managers required by the Affordable Care Act, *see* Joanna Shepherd, *Is More Information Always Better? Mandatory Disclosure Regulations in the Prescription Drug Market*, 99 Cornell L. Rev. Online 14 (2013), campaign finance disclosures, *see* Katherine Shaw, *Taking Disclosure Seriously*, Yale L. & Pol'y Rev. Inter Alia 18, 21 (2016), and disclosure by prosecutors of exculpatory evidence in criminal trials, *see* Emmet G. Sullivan, *Enforcing Compliance with Constitutionally-Required Disclosures: A Proposed Rule,* Cardozo L. Rev. de novo 138, 139 (2016).

However, mandatory disclosure regimes face an inherent tension between producing adequate information on the one hand, and avoiding information overload on the other.[3]

In the context of securities disclosure, the tension is especially acute, as every year mandated disclosures grow. Between 2000 and 2015 for example, the average prospectus for a new issue of stock grew from just under 43,000 words to over 115,000.[4] Amidst the growth in prospectus length, there is widespread debate over disclosure reform.[5] On one side of this debate, advocates argue for curtailing disclosure regulations, noting that they impose high costs on the companies that must comply with them,[6] while doing little for investors who do not read them[7] or who may be overwhelmed by them.[8] Others argue that mandated disclosures should be expanded even further, and point to crises and corporate scandals that have resulted when wrongdoers were able to evade transparency requirements.[9]

---

[2] For example, firms going public for the first time must file a registration statement including numerous disclosures with the SEC pursuant to the Securities Act of 1933, §§ 5 and 10, 15 U.S.C. §§ 77f and 77g (2012). Pursuant to the Exchange Act of 1934, §§ 12–13, 15 U.S.C. §§ 78m–n and 78o (2012), public firms must disclose in their annual report detailed information on such items as the firm's financial results, properties, legal proceedings against the firm, information on the firm's officers and directors, and a discussion by management of the firm's "financial condition, changes in financial condition and results of operations." *See* Item 303, Regulation S-K, 17 C.F.R. § 249.308a (2012).

[3] For an early analysis of both sides of this tension and the usefulness of disclosure regulations for protecting investors, see Frank H. Easterbrook & Daniel R. Fischel, *Mandatory Disclosure and the Protection of Investors*, 70 VA. L. REV. 669, 669–71 (1984). For a more contemporary take, see Steven M. Davidoff Solomon & Claire A. Hill, *Limits of Disclosure*, 36 SEATTLE U. L. REV. 599, 600 (2013).

[4] These numbers were derived from the data used in this Article, as well as more recent prospectuses taken from the SEC's EDGAR system. *See* EDGAR, https://www.sec.gov/edgar (last visited Feb. 20, 2018) [https://perma.cc/RG6D-FY95].

[5] For examples of this debate, critiques of the system and various proposed solutions see, e.g., Geoffrey Manne, *The Hydraulic Theory of Disclosure Regulation and Other Costs of Disclosure*, 58 ALA. L. REV. 473, 473–77 (2007), Davidoff Solomon & Hill, *supra* note 3, at 669–72.

[6] *See, e.g.*, Manne, *supra* note 5, at 474–75; *see also* Easterbrook & Fischel, *supra* note 3, at 671.

[7] *See generally* Omri Ben-Shahar & Carl Schneider, *The Failure of Mandated Disclosure*, 159 U. PA. L. REV. 647 (2011) (arguing that investors turn away from disclosures that are too long and detailed to be useful); Stephen Bainbridge, *Mandatory Disclosure: A Behavioral Analysis*, 68 U. CIN. L. REV. 1023 (2000) (arguing that behavioral biases cause investors to ignore mandatory disclosure or act in ways that make it irrelevant).

[8] *See generally* Troy A. Paredes, *Blinded by the Light: Information Overload and Its Consequences for Securities Regulation*, 81 WASH. U. L.Q. 417 (2003) (arguing that overly detailed disclosure can be counterproductive to a system whose goal is increasing transparency); Steven L. Schwarcz, *Rethinking the Disclosure Paradigm in a World of Complexity*, 2004 U. ILL. L. REV. 1 (suggesting that the mandatory disclosure regime conceived in the 1930s is no longer appropriate for modern complex securities); Steven L. Schwarcz, *Disclosure's Failure in the Subprime Mortgage Crisis*, 2008 UTAH L. REV. 1109, 1110 ("Most, if not all, of the risks giving rise to the collapse of the market for securities backed by subprime mortgages were disclosed, yet the disclosure was insufficient, in part because complexity made the risks very difficult to understand.").

[9] *See* Davidoff Solomon & Hill *supra* note 3, at 600 (observing that "disclosure is too often a convenient path for policymakers and many others looking to take action and hold onto

Often forgotten in this debate is the importance of "stylistic regulation," meaning regulation that governs the form of disclosure rather than the content or subject matter itself. Style offers a distinct policy lever that can appease critics as well as supporters of mandatory disclosure regimes.[10] That is because concerns that disclosure is too costly to produce, or too overwhelming to consume have as much to do with the way information is presented as they do with the amount of disclosure required. Likewise, calls for more disclosure could often be answered through clearer compliance with existing requirements, rather than the creation of new ones.[11] Regulation of form provides a theoretically powerful tool for improving substance without imposing new substantive disclosure requirements.

However, it is unclear whether stylistic regulation is effective as an empirical matter. Stylistic regulation is tricky to enforce, in part because style is relatively subjective, and in part because it is difficult to impose harsh sanctions for non-compliance. Moreover, it has been difficult in the past to measure style in a way that permits investigation into whether the regulations are followed.[12]

Thus, this Article seeks to leverage relatively new analytical methods to assess the effectiveness of the SEC's stylistic regulations, first by investigating whether market participants obey them, and second by determining the impact of stylistic issues on issuers and investors. It does so by examining a particular instance of stylistic regulation: the SEC's attempt to regulate plain English.[13] The SEC's so-called plain English rule encompasses a number of stylistic concerns, including grammatical suggestions—e.g., use of active voice, avoidance of superfluous language, use of short sentences and avoid-

---

comforting beliefs in the face of a bad outcome."). *See generally* Henry Hu, *Too Complex to Depict? Innovation, "Pure Information," and the SEC Disclosure Paradigm*, 90 TEX. L. REV. 1601 (2012) (arguing that modern financial innovation demands expanding disclosure to include pure data on complex financial products); Richard E. Mendales, *Collateralized Explosive Devices: Why Securities Regulation Failed to Prevent the CDO Meltdown, and How to Fix It*, 2009 U. ILL. L. REV. 1362 (arguing that the 2008 financial crisis was "a failure in transparency"); Charles W. Murdock, *Why Not Tell the Truth?: Deceptive Practices and the Economic Meltdown,* 41 LOY. U. CHI. L.J. 801 (2010) (noting the role of deceptive and incomplete disclosure in the economic crisis); Robert B. Thompson, *Market Makers and Vampire Squid: Regulating Securities Markets after the Financial Meltdown*, 89 WASH. U. L. REV. 323 (2011) (arguing that "[i]n modern securities markets, a focus on disclosure by issuers alone has come up short").

[10] One of few examples I found dealing directly with the topic is an excellent student note, *see generally* Kenneth Firtel, Note, *Plain English: A Reappraisal of the Intended Audience of Disclosure under the Securities Act of 1933*, 72 S. CAL. L. REV. 851 (1999). Other articles deal with disclosure format indirectly but do not address the SECs efforts at regulating style directly. *See generally* Hu, *supra* note 9; Schwarcz, *Disclosure's Failure*, *supra* note 8.

[11] *See* Mendales, *supra* note 9, at 1062 (noting that existing regulation should have forced issuers to disclose more risks leading up the financial crisis).

[12] The methods herein are relatively new to law, and although they have been used in finance, I have found no research investigating stylistic regulation per se using the methods that are used herein.

[13] *See* Firtel*, supra*, note 10, at 851–52. *See generally* Plain English Disclosure, Security Act Release No. 33-7497, Exchange Act Release No. 34-39,593, Investment Company Act Release No. 23011, 63 Fed. Reg. 6370 (Feb. 6, 1998).

ance of legal jargon—that companies issuing securities should take into account when writing disclosures.[14] The rules also discourage "boilerplate" language, or generic disclosure copied from other documents that have little specific bearing on a particular securities issuer.[15]

I analyze plain English using IPOs as a case study, because prior to a company's IPO, little company information is typically available to the public, so the prospectus is often the most important source for forming investors' perceptions.[16] The analysis is conducted on a dataset of IPO prospectuses for operating companies going public for the first time. I analyze each deal in terms of the entire IPO prospectus, as well as a subset of the data that includes the sections of each prospectus dealing with risk factors, the use of the deal proceeds, the management's discussion and analysis of company operations, and the description of the business.

The analysis measures plain English usage, as defined in the SEC's rules and style guide. The empirical analysis in the Article does three things: First, it documents and analyzes the mixed results of the SEC's past efforts to force the use of so-called plain English writing in securities disclosures. Second, it provides preliminary evidence that compliance with plain English rules is related to the outcomes of transactions. I hypothesize that this occurs either because the regulations promote clearer disclosure and thus facilitate communication with investors, or because they force issuers to examine their disclosures more critically, and thus produce more information. Third, I find evidence that compliance with plain English rules does not result in significantly higher costs for issuing firms in terms of legal fees and time taken to complete a deal.

To briefly preview the analysis and the results, I assess the impact of the plain English rules on investors by analyzing the relationship between boilerplate and several indicia of uncertainty, including the variance of first day returns for the issuer's stock, the probability of price revision prior to the offering date, and the incidence of litigation.[17]

In the first set of results, I find that some of the SEC's previous attempts to regulate style have been successful, but some have had no effect and even backfired, leading to worse plain English compliance in the section of the prospectus targeted by regulation. Compliance with plain English rules has no discernable effect on the speed with which deals are completed, and does not significantly save issuers legal fees. I find evidence that lower compliance with plain English rules is associated with higher levels of underpric-

---

[14] *See* Plain English Disclosure, 63 Fed. Reg. at 6384; *see also* 17 C.F.R. § 230.421 (2014).

[15] *See* Plain English Disclosure, 63 Fed. Reg. at 6371, 6373, 6384; *see also* Section II.B, *infra*.

[16] *See* JOHN C. COFFEE, JR. & HILLARY A. SALE, SECURITIES REGULATION 115, 122–25 (11th ed. 2009) (describing the importance of disclosure in IPOs since comparatively little information is available about new issuers).

[17] *See also* George R. Klare, *Assessing Readability*, 10 READING RES. Q. 62 (1974). *See generally* Section III.C., *infra*.

ing—a phenomenon by which IPOs are sold at prices below what the market will bear, and often an indicator of investor uncertainty regarding an issuing firm.[18] Meanwhile, better plain English drafting scores overall bear a negative association with underpricing, although the results are not statistically significant.

The positive association between non-compliance with plain English rules and underpricing lends support to the conclusion that poor attention to clarity in prospectus drafting signals lack of transparency about an issuer. These findings support a story that drafting style has an impact on the outcome of the deal, and thus matters to investors, either because it affects the intelligibility of disclosure, or because better drafting forces more information to be disclosed in the first place.

The more troubling lesson of the SEC's experience with the plain English rules is that even if stylistic regulations matter, the effect of such regulations is ephemeral and compliance drops off over time. Moreover, issuers are apt to disregard plain English drafting concerns, even though the costs of doing so are far greater than the potential savings on average.

Nonetheless, another conclusion to this Article is suggested not by the content of the analysis but by the methods: style and stylistic disclosure may soon become obsolete despite their current usefulness. This Article begins to explore how plain English drafting could eventually become a relic of the past. Language processing and machine learning have the potential to mitigate and even eliminate information overload by condensing and analyzing language in ways that even the most sophisticated humans cannot. In fact, "poor" drafting techniques like overuse of boilerplate is well suited to machine processing precisely because it is so standardized. Standardization allows an algorithm to compare disclosure between deals easily, and glean information through the presence, absence, and placement of different kinds of disclosure.

The remainder of this Article will proceed as follows. Part II will give background information on IPO deals, as well as the SEC's disclosure regime and its efforts to restrict the use of boilerplate. This Part will also survey the legal scholarship on the stylistic issues discussed in this Article, as well as the empirical literature that bears on the analysis in this Article. Part III will introduce the methods used in the empirical portion of the Article, and discuss the various means of defining and analyzing boilerplate and plain English. Part IV discusses the results of the quantitative analysis. Part V discusses the implications of the analysis for the law and the SEC's reform efforts.

---

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

II. Background

An IPO is a company's introduction to the public markets, and therefore gathering and disseminating information about the issuing company and its prospects is one of the most important components of the transaction. The information a company must divulge is dictated by the regulatory system set out in the Securities Act of 1933 (the "Securities Act").[19] The Securities Act and the rules promulgated under it exist primarily to ensure that investors have adequate information about an issuing company, rather than be left to the mercies of caveat emptor as they would in other kinds of markets.[20] The theory behind this is that issuing companies have access to information about their operations that ordinary investors cannot learn on their own, and cannot negotiate for in most circumstances.[21] The Securities Act's mandated disclosure regime is intended to reduce information asymmetries that exist between sellers of securities and buyers in their markets that would otherwise go uncorrected.[22] Investors purchasing company shares are buying into the current assets of the company, as well as the company's prospects for future performance.[23] Thus the process of taking a company public is, in large part, a process of gathering information and deciding how to present it. The primary participants in these deals—the lawyers, investment bankers, accountants and the issuing company's management—are responsible for gathering the required information, ensuring its accuracy and attempting to adequately communicate it in the registration statement and accompanying prospectus, the main disclosure document required by the Securities Act.[24] In order to provide context for the remaining analysis, this Section will set out in more detail the process by which IPO disclosure is created, and then briefly describe the recent history of the SEC's efforts to regulate drafting style.

---

[19] 15 U.S.C. §§ 77a–77aa (2012).

[20] *See, e.g.*, *In re* Rediff.com India Ltd., 358 F. Supp. 2d 189, 205 (S.D.N.Y. 2004).

[21] *See id.*

[22] The primary regulation governing the SEC's unified disclosure regime is Regulation S-K, 17 C.F.R. §§ 229.10–229.915 (2014). The Securities Act 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 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, *Ethics, Fairness and Efficiency in Financial Markets*, 49 Fin. Anal. J. 21, 21–29 (1993).

[23] *See generally* Merritt B. Fox, *Shelf Registration, Integrated Disclosure, and Underwriter Due Diligence: An Economic Analysis*, 70 Va. L. Rev. 1005 (1984) (noting that security value is based on asset value and expectation of future dividends).

[24] *See* Fox, *supra* note 23, at 1010. The prospectus is both a legal term and a term of art that refers to the disclosure document that is filed as part of the registration statement in an IPO. For ease of reference, I will refer to both the registration statement and prospectus as "prospectus."

A. Background on IPO Disclosure

While the issuer's management usually has the final say on what goes into the prospectus, lawyers for both the company and the group of investment banks underwriting the deal are heavily involved in the discussions and drafting of the prospectus.[25] Typically, the issuer's counsel bears primary responsibility for drafting the document.[26] And although the content of the disclosure is mandated by the SEC and the Securities Act, the issuer has significant leeway both in terms of what is disclosed and how it is described.[27] Consequently, the issuing company's management, their lawyers, and the investment banks often engage in lengthy negotiations over issues such as the specific wording of certain disclosures (whether vague or detailed), and decisions to exclude important information about which the disclosure regulations are ambiguous.[28]

The prospectus is important both for introducing a previously private company to the investors, and for helping set the price at which the company's stock will debut.[29] Creating the prospectus involves a thorough due diligence review, in which the lawyers and the underwriter investigate the issuing company's business.[30] During the due diligence review, the attorneys gather and verify information for the prospectus, which in turn helps to manage liability risk from material omissions and misstatements.[31] Typical legal due diligence often includes a review of material contracts, related party transactions, cross default provisions, negative pledge agreements, and rights of third parties to terminate contracts.[32] Due diligence requires the issuing company to gather the relevant information for the lawyers to review. It also necessarily involves the underwriters, who may raise questions or ask for verification on particular matters during the process.[33]

---

[25] *See* Carl W. Schneider et al., *Going Public: Practice, Procedure, and Consequences*, 27 VILL. L. REV. 1 *passim* (1981) (discussing the role of underwriters in the drafting of disclosure).

[26] *See id.*

[27] *See id.*; *see also* Tom Arnold, Raymond Fishe & David North, *The Effects of Ambiguous Information on Initial and Subsequent IPO Returns*, 39 FIN. MGMT. 1497, 1500 (2010) (describing the issuer's management's control over the message conveyed in the disclosure).

[28] *See* Arnold et al., *supra* note 27 at 1500; *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).

[29] *See* Manuel Utset, *Producing Information: Initial Public Offerings, Production Costs, and the Producing Lawyer*, 74 OR. L. REV. 275, 277 (1995) (describing the lawyers' job in an IPO to be the production of an information bundle).

[30] *See* Royce de Rohan Barondes et al., *Underwriters' Counsel as Gatekeeper or Turnstile: An Empirical Analysis of Law Firm Prestige and Performance in IPOs*, 2 CAP. MKTS. L.J. 164, 167 (2007).

[31] *See id.*; *see also* Schneider et al., *supra* note 25, at 4–5.

[32] *See* Barondes et al., *supra* note 30, at 167 (explaining the details of the lawyers' involvement in the IPO process).

[33] *See id.*

The prospectus is usually drafted while due diligence proceeds. The lawyers for the issuer typically take the lead, basing the draft off precedent documents taken from past deals and tailoring the disclosure as the issuer and the due diligence process furnish information.[34] The prospectus is revised iteratively in meetings involving both sets of counsel, the underwriters, and representatives from the company.[35]

The prospectus serves as both a marketing document as well as a regulatory one, and the underwriter typically has requirements of its own that make the prospectus a more effective tool for marketing purposes.[36] Additionally, the issuing company usually takes a strong interest in how its story is told.[37] Thus, the drafting process requires collaboration between all the parties and requires counsel to work closely with each other and with both sets of clients.

Once the prospectus is drafted, the issuer files it with the SEC on Form S-1.[38] The SEC's review of the preliminary prospectus typically involves several rounds of comments and requests for clarifications, additions, or alterations to the disclosure.[39] As the SEC reviews the preliminary prospectus, the underwriter and issuer's management begin the marketing effort by visiting institutional investors in various cities and presenting the company's story as set out in the prospectus.[40] Through this process, known as a road show, the underwriter assesses the demand for the stock by building a book of orders from interested investors.[41] The final offering price is ultimately negotiated between the lead underwriter and the issuing company's management based largely on the investor demand ascertained during the road show.[42] A final version of the prospectus and pricing information are filed,

---

[34] *See id.*

[35] *See id.* ("This drafting is an iterative process, as knowledge gained in due diligence informs what needs to be said about the issuer.").

[36] *See id.*; *see also* Schneider et al., *supra* note 25, at 14.

[37] *See* Schneider et al., *supra* note 25, at 14, 18.

[38] The most common form is S-1, although S-3 is sometimes used, as are SB-1 and SB-2, until the repeal of the small business disclosure rules. For simplicity, I will refer to all of these documents as S-1.

[39] *See* Schneider et al., *supra* note 25, at 45–50 (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).

[40] *See* Schneider et al., *supra* note 25, at 22 (noting that the "red herrings" are distributed after filing and while the SEC reviews the filing).

[41] *See id.* at 22–23; *see also* COFFEE & SALE, *supra* note 16, at 122–25 (describing the road show and bookbuilding process); Barondes et al., *supra* note 30, at 168–69 (describing the development of an offering price and using the initial filing range as a proxy for the estimate developed during the "beauty contest").

[42] *See* Barondes et al., *supra* note 32, at 168 ("In a customary IPO, there is not a definitive agreement on the price at which the underwriters will resell the stock to the public 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

and once deemed effective by the SEC,[43] the shares are sold to the investors and the company officially goes public.[44]

B. Disclosure Reform and the Plain English Problem

Given the way in which IPO prospectuses (and securities disclosures in general) are drafted, it is not hard to see why comprehensibility sometimes suffers, and why stylistic regulation has periodically become a priority. Vestigial language carried over from precedent documents is ubiquitous, and the presence of many drafting parties under time pressure make good writing difficult. Moreover, the presence of numerous lawyers poring over each word to ensure legal accuracy can lead to highly technical, jargon-filled prose. For years, regulators have taken steps to fight this problem by urging issuers to use less boilerplate and encouraging them to use clearer language. Most recently, in December 2015, Congress passed the FAST Act,[45] which requires the SEC to review its disclosure regulations, eliminate redundancies and look for ways to limit the use of needless boilerplate language.[46] The SEC has subsequently solicited feedback on how disclosure might be streamlined to reduce or eliminate boilerplate.[47] The SEC has proposed some reforms in the wake of the FAST Act, although they are relatively modest ones.[48]

Congress and the SEC have tried to improve drafting quality and reduce repetitive, cut-and-paste language in securities disclosures for many years. The most forceful efforts to streamline disclosure date back to the 1980's when the SEC created a unified disclosure system with the promulgation of Regulation S-K, which was an attempt to consolidate and streamline what

---

prospectuses, as the managing underwriter acquires information during the marketing process.").

[43] *See* 17 C.F.R. §§ 230.424(b), 230.430A (2014); *see also* COFFEE & SALE, *supra* note 16, at 128–29. Before the promulgation of Rule 430A, the underwriters were required to file pricing information in the form of an amendment to the registration statement before the SEC declared the registration statement effective. *See* COFFEE & SALE, *supra* note 16, at 128–29. Under Rule 430A, the registration statement can be declared effective before the filing of pricing-related information as long as a complete final prospectus is filed shortly thereafter. *See* 17 C.F.R. § 230.430A (2014).

[44] *See* COFFEE & SALE, *supra* note 16, at 129.

[45] *See generally* Fixing America's Surface Transportation Act of 2015 (FAST Act), Pub. L. No. 114-94, 129 Stat. 1312 (2015).

[46] *See id.* § 72002.

[47] *See* FAST Act Modernization and Simplification of Regulation S-K, Securities Act Release No. 33-10425, Exchange Act Release No. 34-81851, Investment Company Act Release No. 32,858, 82 Fed. Reg. 50,988 (proposed Nov. 2, 2017). *See generally* SEC, REPORT ON MODERNIZATION AND SIMPLIFICATION OF REGULATION S-K (2016), https://www.sec.gov/reportspubs/sec-fast-act-report-2016.pdf [https://perma.cc/3LAQ-XHH4]; Business and Financial Disclosure Required by Regulation S-K, Securities Act Release No. 10064, Exchange Act Release No. 77599, 81 Fed. Reg. 23,916 (proposed Apr. 22, 2016).

[48] *See* FAST Act Modernization and Simplification of Regulation S-K, *supra* note 47. The notice and comment period ended January 2, 2018. No action has been taken at the time of this writing.

had previously been a dispersed array of disclosure provisions.[49] In 1992, the SEC sought to simplify disclosure requirements for small business issuers (at the time, those with less than $25 million in annual revenues).[50] In 1995, Congress passed the Private Securities Litigation Reform Act ("PSLRA"),[51] and while it did not directly address the streamlining of disclosure requirements, it did provide additional protection from lawsuits for companies that issued forward-looking statements, as long as the statements were accompanied with cautionary language.[52] The legislative history of the PSLRA makes clear that boilerplate cautionary language would not be enough to protect companies' disclosures from allegations of fraud—specific disclosures would be needed.[53] That same year, the SEC created a Task Force on Disclosure Simplification that was charged with reviewing disclosure trends and requirements.[54] By this time, many practitioners recognized that IPO prospectuses had become dense and repetitive documents that incorporated a large amount of non-specific language.[55] The SEC decided to address these trends through rules designed to reduce duplicative information from a number of prospectus sections.[56]

In 1998, the SEC also advanced its Plain English Initiative that resulted in guidelines for making prospectuses more readable.[57] The "plain English rule" coming out of the initiative mandated that certain parts of the prospectus—in particular the summary and risk factors—employ short sentences, concrete language, and active voice, and avoid jargon and legalese.[58] In addi-

---

[49] *See id.*; *see also* 17 C.F.R. §§ 229.10–229.915 (2014).

[50] *See* Small Business Initiatives, Securities Act Release No. 33-6949, Exchange Act Release No. 34-30968, Investment Company Act Release No. 39-2287, 57 Fed. Reg. 36442 (Aug. 13, 1992).

[51] *See* Private Securities Litigation Reform Act of 1995, Pub. L. No. 104-67, 109 Stat. 737.

[52] 15 U.S.C. § 78u-5(c)(1)(A) (2012).

[53] *See* H.R. Rep. No. 104-369, at 1890 (1995).

[54] *See* SEC, Report of the Task Force on Disclosure Simplification (1996), www.sec.gov/news/studies/smpl.htm [https://perma.cc/HYN5-SRZK].

[55] *See id.*; *see also* Wilson Sonsini Goodrich & Rosati, Initial Public Offerings 151–52 (3d ed. 2008), https://www.wsgr.com/publications/PDFSearch/IPO-guidebook-3.pdf [https://perma.cc/N3EV-3NZU].

[56] *See* SEC, *supra* note 54. 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.*

[57] *See* The Regulation of Securities Offerings, Securities Act Release No. 33-7606A, Exchange Act Release No. 34-40632A, Investment Company Act Release No. 23519A, 63 Fed. Reg. 67,174 (proposed Dec. 4, 1998).

[58] *See* Plain English Disclosure, Security Act Release No. 33-7497, Exchange Act Release No. 34-39,593, Investment Company Act Release No. 23011, 63 Fed. Reg. 6370, 6370 (Feb. 6, 1998). The Rule went into effect and required compliance as of Oct. 1, 1998. *Id.*

tion, the rule provided that the entire prospectus should avoid boilerplate.[59] The SEC reasoned that, the more information an investor received, the less likely he or she would be to absorb it, and the presence of boilerplate risk language would cause investors to ignore the language altogether.[60]

The SEC made another push to simplify disclosure language in 2003. A series of releases that year highlighted the overuse of boilerplate language in the Management's Discussion and Analysis ("MD&A") section of SEC filings, and admonished against the use of standardized, untailored language.[61] The SEC further streamlined disclosure for smaller companies in 2007, creating a new category of smaller reporting issuers with less than $75 million in public float.[62] Those efforts clashed, however, with a raft of new requirements for public company disclosure in the wake of the Sarbanes-Oxley Act of 2002,[63] a sweeping corporate governance reform statute that was enacted as a reaction to several corporate scandals at the beginning of the century.

C. Empirical Literature on Text Analysis

Before continuing the analysis, I undertake a review of the relevant literature on natural language processing and analysis of text-based regulation. Natural language processing has quickly become a tool for use in aca-

---

[59] *See id. Cf. generally* Office of Inv'r Educ. & Assistance, SEC, A Plain English Handbook: How to Create Clear SEC Disclosure Documents (1999), https://www.sec.gov/pdf/handbook.pdf [https://perma.cc/95RW-J948].

[60] *See* SEC, *supra* note 59.

[61] *See* SEC, Summary by the Division of Corporation Finance of Significant Issues Addressed in the Review of the Periodic Reports of the Fortune 500 Companies (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 MD&As 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. 2366, 63 Fed. Reg. 41,394, 41,398 (August 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. 33-809, Exchange Act Release No. 34-45,907, 67 Fed. Reg. 35,620, 35,622 (proposed May 20, 2002).

[62] *See generally* Smaller Reporting Company Regulatory Relief and Simplification, Securities Act Release No. 33-8876, Exchange Act Release No. 34-56,994, Investment Company Act Release No. 39-2451, 73 Fed. Reg. 934 (proposed January 4, 2008).

[63] Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745.

demic analysis of contracts,[64] prospectuses,[65] periodic filings with the SEC,[66] and corporate documents,[67] among other things. Language processing encompasses a variety of different techniques that can be used to parse and analyze large amounts of text using computer algorithms.[68]

One study relevant to boilerplate in IPOs reviews prospectuses to investigate informative versus non-informative text.[69] This paper parsed a large sample of IPO prospectuses, created a list of the most standard words, and analyzed the overlap between each prospectus and the standard word list.[70] The study assumed common language to be uninformative, while all uncommon language was assumed to be informative.[71] Using these measures of informative and uninformative text, the authors concluded that informative text is associated with more accurate pricing.[72]

Another study that is relevant to the research presented here involves an analysis of all company annual reports filed with the SEC on form 10-K between 1996 and 2013.[73] These documents were similarly parsed and analyzed for comprehensibility using readability indexes.[74] The study found that

---

[64] This Article builds upon the work of a growing number of legal academics who use language processing and machine learning to study the law. *See, e.g.*, Pamela Corley, Paul Hamer & Jesse Collins, *The Influence of Amicus Curiae Briefs on U.S. Supreme Court Opinion Content*, 49 L. & SOC'Y REV. 917, 917–20 (2015) (using plagiarism detection software to identify the importance of parties' briefs in Supreme Court opinions); Jonathan Macey & Joshua Mitts, *Finding Order in the Morass: The Three Real Justifications for Piercing the Corporate Veil*, 100 CORNELL L. REV. 99, 135 (2014) (using a machine learning algorithm and text processing to analyze the most common conditions for the successful employment of corporate veil piercing doctrine); Eric Talley & Drew 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, 181–85 (2012) (using machine learning classifier to identify and analyze material adverse event clauses); Eric Talley & Gabriel Rauterberg, *A Machine Learning Classifier for Corporate Opportunity Waivers*, 117 COLUM. L. REV. 1075, 1119–23 (2017) (using machine learning to quantify the occurrence of fiduciary duty corporate opportunity waivers in public companies).

[65] *See, e.g.*, Hanley & Hoberg, *supra* note 28, at 2821–25 (describing a project in which natural language processing is used in prospectuses).

[66] *See* Tim Loughran & Bill McDonald, *Measuring Readability in Financial Disclosures*, 69 J. FIN. 1643, 1643–46 (2014).

[67] *See* Andriy Bodnaruk, Tim Loughran & Bill McDonald, *Using 10-K Text to Gauge Financial Constraints*, 50 J. FIN. QUANTITATIVE ANAL. 623, 623–30 (2015); Tim Loughran & Bill McDonald, *Plain English, Readability, and 10-K Filings* (2009) [hereinafter *10-K Filings*] (unpublished manuscript), https://www3.nd.edu/~tloughra/Plain_English.pdf [https://perma.cc/3KBA-FRW9].

[68] For a general survey of techniques and their uses, see Tim Loughran & Bill McDonald, *Textual Analysis in Accounting and Finance: A Survey*, 54 J. ACCT. RES. 1187 (2015); *see also* Scott E. Masten & Stephane Saussier, *Econometrics of Contracts: An Assessment of Developments in the Empirical Literature on Contracting*, 92 REV. D'ÉCONOMIE INDUS. 215 (2000).

[69] *See* Hanley & Hoberg, *supra* note 28, at 2821. Some of the methodology for the first part of this analysis is adapted from the work done there. In addition, several robustness checks were performed using the techniques in Hanley & Hoberg's paper.

[70] *See id.*

[71] *See id.*

[72] *See id.*

[73] *See 10-K Filings*, *supra* note 67, at 3.

[74] *See id.*

companies with more readable (as measured using the Gunning-Fog index, a standard readability index) 10-Ks have more participation from smaller investors (as measured by small lot purchases and sales).[75] While this study does not directly bear on boilerplate or plain English, readability and comprehensibility are relevant to whether or not language—whether standard or not—is informative.[76] Therefore, I use several methods of analyzing comprehensibility taken from the methodology in this study to determine if boilerplate and plain English impact disclosures' informativeness.

III. Methodology: Measuring Disclosure Style

In this Part, I describe the methods I use to measure plain English, in order to explore compliance with the regulations. Before delving into the details of the empirical analysis, I first set out the elements of the regulation that can be measured through text analysis. I then discuss ways to measure the main components of the rule in light of the foregoing background discussion.[77]

I define "plain English" to be compliance with the specifics set out in the SEC's plain English rule, other than the part of the rule that deals with boilerplate language which, for reasons discussed below, is measured according to a different method. The SEC has provided specific guidance on what counts as plain English for its purposes. Specifically, those things are the use of short sentences, short words, concrete language, active voice, tables, and bullet points, wherever possible.[78] It also requires that issuers avoid double negatives, legal jargon, and overly technical language.[79] In addition, the SEC has published a *Plain English Handbook*, which gives additional guidance, including lists of specific words and phrases that are compliant and certain others to avoid.[80] The manual also included a number of so-called superfluous words to avoid.[81] Given the specificity of the terms of the

---

[75] *See id.* Small lot sales are defined by the authors as sales of lots of 100 shares. Such denominations are typically traded by smaller, less sophisticated investors. *See id.*

[76] *See id.*

[77] I realize that separating boilerplate and plain English for analysis may seem somewhat artificial since boilerplate may affect the extent to which drafting is clear; nonetheless, it is useful to separate them because the SEC defines them distinctly and emphasizes them differently. In particular, it admonishes issuers against the use of boilerplate, but defines the term only vaguely. *See, e.g.*, Plain English Disclosure, Securities Act Release No. 33-7380, Exchange Act Release No. 34-38,164, Investment Company Act Release No. 22,464, 62 Fed. Reg. 3152, 3159 (proposed Jan. 21, 1997). By contrast, the SEC sets out fairly well-defined criteria for what it considers to be plain English, even going as far as providing lists of words to avoid. *See, e.g.*, 17 C.F.R. § 240.421(d) (2014).

[78] *See* 17 C.F.R. § 240.421(d).

[79] *See id.*

[80] *See* Office of Inv'r Educ. & Assistance, *supra* note 59; *see also* Plain English Disclosure, Securities Act Release No. 33-7497, Exchange Act Release No. 34-39,593, Investment Company Act Release No. 23,011, 63 Fed. Reg. 6370 (Feb. 5, 1998).

[81] *See* Office of Inv'r Educ. & Assistance, *supra* note 59, at 22.

plain English rule and *Handbook*, I use the items specifically listed as a basic definition of plain English.

The plain English rule also admonishes issuers to minimize the use of boilerplate language, a simple command with large-scale implications given the copy-and-paste way that prospectuses are often drafted. The SEC's regulations governing boilerplate suggest that one of their main objectives is to minimize "language copied from the prospectuses of similar companies or companies in the same industry."[82] Given the way that most prospectuses are written, it makes sense that boilerplate would often appear in the form of language that is simply borrowed from other companies' documents, and that conveys little about the specific issuer using it.

However, while copying language wholesale from similar issuers is likely to lead to generic disclosure, it is not necessarily a bad thing. Of equal relevance is the level of generality of a given sentence or passage, as well as the importance of its subject matter, and therefore a definition of boilerplate for regulatory purposes should presumably be more nuanced. For instance, a copied sentence vaguely suggesting that "the company is subject to risks from its competitors" is both obvious and not very informative. However, a phrase stating "Our net operating profits last year were . . ." may be copied and generic, but innocuous, providing a basic frame for the required disclosure.[83]

Given these considerations, I leverage two different text analysis techniques to assess the use and impact of boilerplate. In the first method, I quantify the language that is copied from one prospectus to another. As discussed in more detail below, I perform comparisons of the text from each prospectus, in a manner that provides the amount of similarity between each

---

[82] *See generally* Plain English Disclosure, 62 Fed. Reg. at 3152 (admonishing issuers against the use of repetitive boilerplate language, but not defining exactly what such language consists of).

[83] Courts have implicitly expressed the view that boilerplate in securities disclosure is overly generic language, albeit in the specific context of construing the 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."); Asher v. Baxter Int'l Inc. 377 F.3d 727, 732 (7th Cir. 2004); Southland Sec. Corp. v. INSpire Ins. Sols. Inc., 365 F.3d 353, 372 (5th Cir. 2004) (holding "a boilerplate litany of generally applicable risk factors" inadequate to meet the requirement for meaningful cautions); Sawant v. Ramsey, Civil Action No. 3:07-CV-980 (DLVLB), 2010 WL 3937403, at *15 (D. Conn. Sept. 28, 2010).

In addition, the Conference Report for the PSLRA states that "[u]nder 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." SECURITIES LITIGATION REFORM, H.R. REP. NO. 104-369, at 43 (1995).

one, to formulate a measure of copied text.[84] Because this method cannot capture the nuanced differences between generic and specific copied language, in the second method I also construct a topic model of the repeated sentences using an algorithm. This allows for differentiating between generic and potentially informative boilerplate. I can then use the prevalence of generic versus non-generic topics to create a measure of copied text that includes generic topics and excludes framing language or plausibly specific disclosure. These techniques are further explained below.

A. Data Sourcing, Cleaning, and Processing

Before compliance with plain English rules could be adequately measured, the prospectuses and other relevant data had to be gathered and cleaned. Each prospectus was downloaded from the SEC's EDGAR website in either a text or HTML format. For purposes of the initial analysis, the S-1 filing containing the first draft of the prospectus (often called the "red herring") was used, since this version of the prospectus is the one most likely to be seen by initial investors. I also gathered the final version of each prospectus, filed under rule 424(b),[85] as well as each amendment submitted in between.[86] The documents were then cleaned. This process involved eliminating numeric digits, graphic content, punctuation, and stop words—words such as articles, personal pronouns, conjunctions that appear frequently but do not contribute significantly to an understanding of the text, or more relevantly, how repetitive language is from one document to the next. For all HTML documents, tags and other HTML code were removed. For all documents, tables containing numeric information were also removed because such tables cannot be easily compared and are not usually the source of stylistic concerns that the SEC is concerned about. Each word was then stemmed: any endings were removed and the word was reverted to its root.

B. Measuring Plain English

As previously explained in Section II, the plain English rule includes a number of stylistic requirements beyond the elimination of boilerplate language.[87] In order to assess compliance with the rule, I create a measure of the

---

[84] In a further robustness test, I perform the analysis using a difflib similarity. Difflib is a means of measuring similarity that accounts for exact word order, which cosine similarity does not. The results are consistent using both methods.

[85] 17 C.F.R. § 230.424(b) (2014).

[86] Amendments appear on Form S-1/A.

[87] The plain English rule became effective October 1, 1998. Then-SEC Chairman Arthur Levitt described the need for the new rules to improve disclosure documents:

> Investors need to read and understand disclosure documents to benefit fully from the protections offered by our federal securities laws. Because many investors are neither lawyers, accountants, nor investment bankers, we need to start writing disclosure documents in a language investors can understand: plain English.

extent to which a prospectus follows the requirements laid out by the SEC in the regulation, and in the *Plain English Handbook* that the SEC published shortly after the regulations became effective.[88]

The plain English rule and the SEC's *Plain English Handbook* set out several stylistic requirements. Specifically, the rule requires the use of short sentences and words, the use of concrete language, the use of active voice, the use of tables and bullet points, and the avoidance of double negatives and legal jargon.[89] The *Plain English Handbook* further expands on the points set out in the rule, indicating that issuers should use personal pronouns and includes a list of superfluous words that issuers should avoid.[90]

I use the rule's criteria and wordlists to construct a measure of plain English. I do this by first tallying the occurrence of each of the items listed in the rule and in the *Handbook* for each prospectus. Of course, some of the items are more difficult to count than others. For instance, whether a word is concrete or vague is difficult to assess mechanically, and depends on context and other things that cannot be easily quantified. However, most of the items in the SEC's rule and lists can be retrieved mechanically. I used an algorithm to identify instances in which a relevant style rule was breached or complied with, and aggregated those instances for each prospectus and prospectus section. Since each rule relates to categories of "error" that may occur with more or less frequency in any document (for example, long sentences and complex words occur more frequently than double negatives or superfluous words), the total for each category of style error was converted to a standard range having a mean of zero and standard deviation of one. The standardized scores were then compiled to create a composite score representing a prospectus's compliance with the plain English rule and style guide.[91]

C. Measuring Copied Language

A definition of boilerplate that I adopted for this analysis is language that can be used with little modification across many different companies' documents. That definition implies that language that has been copied with minimal modification from one document to the next is essentially boilerplate. Simple measures of similarity between the text contained in different prospectuses provide a metric of the amount of boilerplate that is present in the documents. One such measure of similarity that is widely used in infor-

---

OFFICE OF INV'R EDUC. & ASSISTANCE, *supra* note 59, at 3; *see also* Firtel, *supra* note 10, at 897.

[88] *See* OFFICE OF INV'R EDUC. & ASSISTANCE, *supra* note 59, at 1.

[89] *See* 17 C.F.R. § 230.421(d).

[90] *See id.*

[91] This methodology has been used in studies of plain English in 10-K documents. *See* Loughran & McDonald, *supra* note 66, at 1647–52; *10-K Filings*, *supra* note 67, at 2.

mation retrieval systems is the *cosine similarity* of the samples of text, which I employ to measure boilerplate.[92]

In the most general terms, cosine similarity is a measure of the amount of overlap that occurs between two samples of text. Finding cosine similarity entails converting a selection of text to a numerical vector representing the number of occurrences of each word in a text. Selections of text can be compared using the cosine of the angle between the vectors representing each selection. Cosine similarity is a useful way to compare texts because it is relatively simple computationally and yields a readily interpretable number between zero and one that corresponds to a percentage of similarity between the two bodies of text being compared.[93]

Using this method, each document was converted into numeric vectors.[94] The cosine of the angle between the two vectors shows the degree at which they are inclined with one another.[95] The precise inclination of the vectors provides a useful means of measuring the amount of overlap (or difference) between the two selections of text. For example, if the cosine value equals 0.70, one can interpret this to mean that the sentences are 70% similar to one another.

Once the text was processed and cleaned, I performed comparisons to determine the cosine similarity of each prospectus to every other prospectus. Since using the full prospectus for each IPO had the potential to yield noisy results, I also excerpted several sections of the prospectus that the literature describe as important for investors considering IPO stock: the Risk Factors, the Use of Proceeds, the MD&A, and the Business Description.[96]

---

[92] Cosine similarity is widely used in studies of information processing. *See generally* Rada Mihalcea, Courtney Corley & Carlo Strapparava, *Corpus-Based and Knowledge-Based Measures of Text Semantic Similarity*, *in* Proceedings of the 21st National Conference on Artificial Intelligence 775 (2006); *see also* Dekang Lin, *An Information-Theoretic Definition of Similarity*, *in* Proceedings of the Fifteenth International Conference on Machine Learning 296 (1998).

[93] *See id.*

[94] The numeric version of these sentences in vector form can be represented as shown below, where each vector object represents a word separately and the scalar quantity associated represents the number of times the word has appeared in this text. Thus, in the example below, each letter represents a unique word, and the number next to it represents number of times it appears in the sentence:

Numeric version of text1 = Vec1 = $\{2\vec{a} + 1\vec{b} + 1\vec{c} + 1\vec{d} + 1\vec{e} + 1\vec{f} + 2\vec{g} + 1\vec{h} + 1\vec{i} + 1\vec{j}\}$

Numeric version of text2 = Vec2 = $\{1\vec{i} + 1\vec{h} + 1\vec{f} + 1\vec{g} + 1\vec{k} + 1\vec{l} + 1\vec{m} + 1\vec{a} + 1\vec{b}\}$

[95] The angle of inclination between these two vectors can be mathematically calculated by using the standard approach, i.e.,

Cosine $= \frac{|vec1| * |vec2|}{|vec1| . |vec2|}$ and logically that can be represented as

Cosine $= \frac{sum([vec1[x] * vec2[x]\ for\ each\ word\ in\ intersection])}{\sqrt{\text{sum([vec1[x]**2 for each word in vec1])} * \text{sum([vec2[x]**2 for each word in vec2])}}}$

Where intersection = vec1[x] $\cap$ vec2[x].

See Lin, *supra* note 92, at 394.

[96] *See* Hanley & Hoberg, *supra* note 28, at 2823.

For each of these five selections of text, I compared the text for each IPO to the text from every other IPO in the dataset using the method described above. Thus, for example, to get the average amount of similarity for the entire prospectus for a particular deal, the prospectus for that deal was compared to each of the 2,255 other prospectuses in the dataset. The result was a matrix containing 2,542,512 observations (2,255 x 2,255, divided by 2 since one half are duplicate comparisons). I repeated that step for each of the four excerpted sections, resulting in 12,712,560 observations that represented all the possible combinations of comparisons between deal documents and their excerpts. These similarity measures were then used as an initial means to assess the use of boilerplate, as further discussed below.

Using this matrix, boilerplate could be measured in a number of ways. One way of measuring it was as the average similarity of each document to every other document. However, this method ignores the most probable mechanism by which standard language is copied from one document to another, namely, by companies within a particular industry and over a relatively short time period. Another way of measuring boilerplate was to use the average similarity of a text to the text from other deals in the same industry within the preceding year. This approach comported to what the literature and anecdotal reports revealed about how lawyers and bankers draft IPO disclosure—by starting with recent deals from within the same industry. Therefore, as a starting point I constructed a boilerplate measure using each deal's similarity to all recent industry deals.[97]

D. Topic Analysis

The kind of copied language one might worry about in securities disclosure is vague and generic about substantive issues, while other types of copied language may be simply framing language and therefore less important. The phrase analysis described here separates the two types of disclosure.

In order to address these issues, I performed an analysis at the level of each sentence in each document to determine the topics of the copied "boilerplate" language. This in turn allowed me to distinguish generic disclosure on important issues from framing language that should have no import to investors. This required using an algorithm to identify and extract every individual sentence from every prospectus. Each sentence from a deal's Risk Factors, Use of Proceeds, MD&A, and Business Description sections was

[97] This method is different from the one employed by 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 recent IPOs within industry, and outside of the industry in the preceding ninety days. *See* Hanley & Hoberg, *supra* note 28, at 2830. As a robustness check, I also calculated boilerplate using the same methodology and obtained results consistent with the method used here. I use the method herein since it is more straightforward, and since I am interested in duplication from deal to deal, in addition to "standard" content.

compared to every other phrase for all other deals to determine the degree of similarity between all phrases.

I then identified sentences that were identical or nearly identical to one another. For this purpose, I considered sentences to be nearly identical if they were 70% or more similar to each other (meaning they share a cosine similarity of at least 0.7). The 70% cutoff was determined after attempting the analysis using a cutoff of 50%, 70%, or 90%.[98] Using 70% as a cutoff was broad enough to allow two sentences with the same meaning but a few words changed to be counted as the same, but narrow enough to exclude sentences that had different meanings while sharing similar words. The 70% cutoff was also broad enough to capture sentences that shared most of the same words but differed by one or two words in a way that would change the meaning.

The end result was a large, sparse matrix with the most frequently occurring sentences across all the documents and their corresponding appearance throughout all the years and different industry sectors in the dataset. This formed the basis of a topic model.

The large matrix of sentences revealed no obvious pattern (and in fact was rather boring to read); however principal components analysis ("PCA"), a statistical technique that is usually used for eliminating redundancies in large datasets and preparing text for machine learning operations, provided a fitting tool for boiling down the boilerplate to reveal latent patterns in it.

PCA eliminated redundancies in the boilerplate sentences by identifying groupings of them (the principal components) that each individually conveyed the most information.[99]

For each section studied above, and for the document as a whole, principal components were determined for the collection of repeated sentences throughout all of the prospectuses. Each principal component was a weighted average of each sentence, with weightings that ensured that the components together describe the maximum amount of variation in the data. The principal component loadings resulted in lists of sentences and weights to indicate their respective importance to each component. This meant that a human could examine the most heavily weighted sentences for each princi-

---

[98] In the corpus it is relatively uncommon to encounter sentences that are completely identical, even in similar deal documents that are clearly drawn from the same precedent. Lawyers drafting deal documents may begin with boilerplate language from a prior deal, but they typically alter the language, 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 sentences that are 90% similar or more, I obtain only a relatively small number. Conversely, a 50% cutoff yields far too many sentences, many of which are not similar enough to have been copied or to be plausibly considered boilerplate.

[99] More technically speaking, the PCA 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. *See* Hervé Abdi & Lynne J. Williams, *Principal Component Analysis*, 2 WILEY INTERDISC. REVS.: COMPUTATIONAL STAT. 433, 433–40 (2010).

pal component and determine the basic pattern or topic. Thus, the boilerplate sentences could be grouped together in terms of meaning as well as the importance of their recurrence. Instead of considering 400 sentences spread across 2255 documents, I could examine a small set of principal components that captured the common meaning among the repeated sentences. The variables that described framing language, or topics that were arguably nongeneric, were then used to create a residual of the cosine similarity measure of copied language for each document, to create a final measure of boilerplate that described only the copied language that is generic or vague in nature.

E. Results of Analysis

This section sets out the basic results of the different sets of analysis described in the preceding section. At the outset, I provide some basic data to give context to the results.

1. Overall Use of Plain English Features

To analyze compliance with the plain English regulations, I examined boilerplate and the directives of the SEC's style handbook separately, given the different ways each was measured. Between 1996 and 2010, the average size of IPO prospectuses rose from just under 40,000 words to just over 100,000. Unsurprisingly, the methods described herein revealed that much of the content of these prospectuses was generic language copied from past transactions.

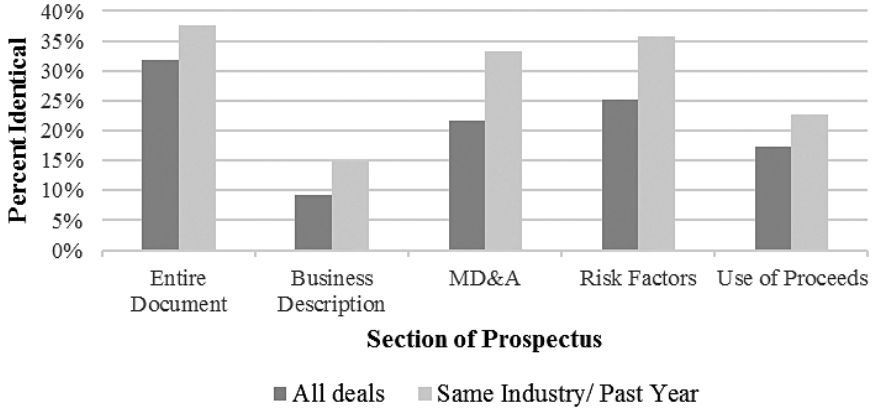
The first task in analyzing boilerplate was to calculate a single similarity score for each deal that would represent the amount of boilerplate contained in that deal. To do so consistent with practitioner accounts of how boilerplate is incorporated into prospectuses,[100] I calculated the average cosine similarity between each deal and all other deals in the same industry group, within the preceding 365 days.[101] As an alternative measure, I also calculated the average similarity between each deal and every other deal that preceded it (whether in the same industry or not) using the same method. The averages for each of these are shown in Figure 1 for each of: the entire prospectus document, the Risk Factors section, the Use of Proceeds section, the Business Description, and the MD&A.

---

[100] *See* WILSON SONSINI GOODRICH & ROSATI, *supra* note 55, at 151.

[101] As a robustness check, I also calculated average similarity for deals outside the same industry. I also used the method employed by Hanley & Hoberg, *supra* note 28, at 2829–34, and obtained consistent results, albeit in a more complicated way.

FIGURE 1: AVERAGE AMOUNT OF IDENTICAL TEXT



Averages are provided for comparisons between all documents, as well as averages for comparisons between documents within the same industry in the past year. As would be expected, average similarity is greater within industries in the recent past than on average overall. In addition, the individual sections that contain the most copied language in the graph are consistent with the literature and anecdotal evidence on where lawyers tend to use boilerplate.[102] The average similarity of each deal to each other deal provides a measure of the amount of standard content present in a given prospectus.[103]

The amount of boilerplate varies widely depending on the section of the document. For the entire document, similarity to other documents within the industry and in the past year ranges from a minimum of 10% to a maximum of 80%. On average, the prospectuses all share approximately 37% identical content to other recent industry deals. Sections that stand out for containing a relatively large proportion of boilerplate are the Risk Factors section (36% identical language on average), and the MD&A (33% identical language on average). The relatively high average levels of similarity in those sections are consistent with the fact that the MD&A and Risk Factors are perennial targets of attempted regulation for containing too much boilerplate.

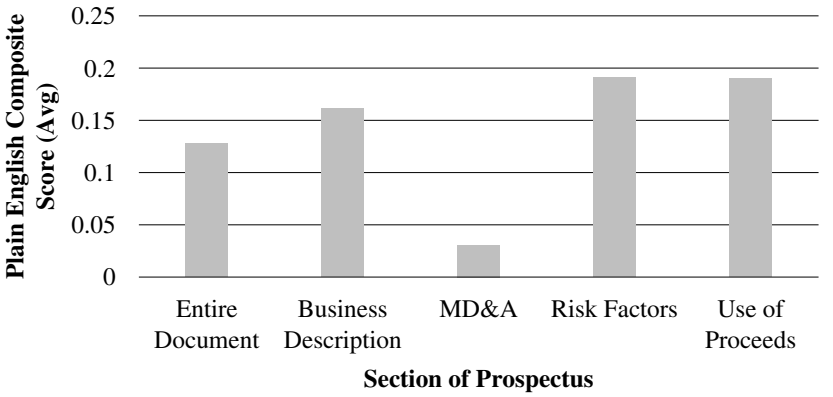
More modestly, the Use of Proceeds section contains on average 23% boilerplate, while the Business Description contains on average 15% boilerplate, which makes sense given that each business is different, and its narrative description ought not to contain a large amount of standard language. The analysis also reveals that plain English scores vary by section of the document, as demonstrated in Figure 2.

---

[102] *Cf.* WILSON SONSINI GOODRICH & ROSATI, *supra* note 55, at 151.

[103] This is similar to, but distinct from the technique used by Hanley and Hoberg, who used a statistical regression to estimate standard content. *See* Hanley & Hoberg, *supra* note 28, at 2835.

FIGURE 2: AVERAGE PLAIN ENGLISH SCORES



The plain English composite scores represent a standardized score for the presence or absence of various elements that the SEC lays out in the rules and its *Plain English Handbook*. Lower scores indicate less compliance, while higher scores indicate more compliance. Once again, the scores vary based on section. The MD&A section, often one of the most technical, has the lowest average scores. Risk Factors and Use of Proceeds, two sections that often have high amounts of narrative and descriptive content, have some of the highest scores on average over the course of the period studied. These scores do not correspond directly with the amount of copied text in the documents shown in Figure 1. This may be due to the fact that the decision to use copied or boilerplate language takes additional considerations into account beyond compliance with the SEC's directive. An exhaustive exploration of those considerations is beyond the scope of this Article, but is the basis for future work.

2. *The Impact of SEC Stylistic Disclosure Policies*

Have the SEC's stylistic disclosure policies affected the use of boilerplate or plain English writing? The similarity measures and plain English score described above allow for analysis of whether these policies have made an impact on prospectus drafters. I begin with an analysis of boilerplate, and then move to an analysis of plain English use.

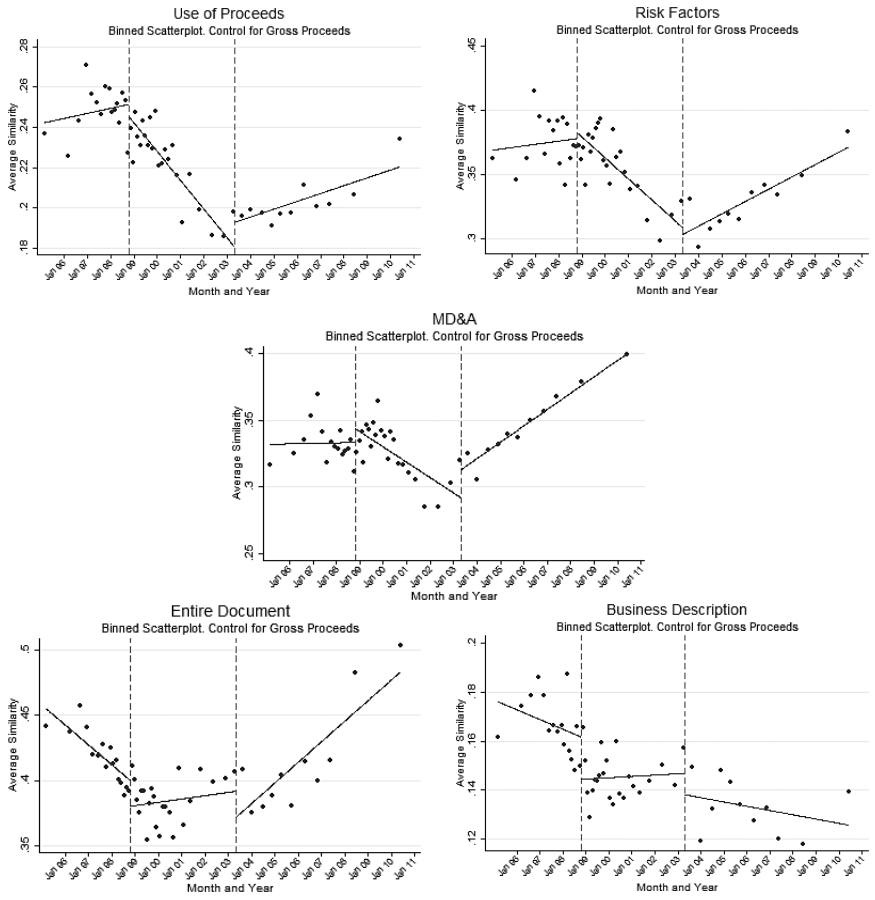
The discussion above in Section II describes the SEC's attempts to target unclear writing in the past few decades. The SEC's plain English rules became effective on October 1, 1998.[104] To assess compliance with the rule, I examine the trends of over time with respect to the use of copied language as well as compliance with the SEC's style guidance, and look for a break in the trend at the time that the new guidance was issued. A look at the data for

[104] Plain English Disclosure, 63 Fed. Reg. 6370, 6370 (Feb. 6, 1998).

those time periods reveals that such a break exists: with respect to the Risk Factors as well as several other sections, the data show a downward trend in plain English scores and an upward trend in the use of boilerplate before the effectiveness of the plain English rule. After the enactment of the new policies, these trends reverse sharply. A further statistical test similarly confirms a statistically significant break in the trends at this time, when controlling for variables that might otherwise predict a change in trends with respect to text features.[105]

The data thus provides evidence that the SEC's attempts to limit boilerplate in the late 1990's were effective, at least for a time. Figure 3a below depicts the use of boilerplate over the time period covered in the data.

FIGURE 3A: AVERAGE SIMILARITY OF DEALS TO DEALS IN THE SAME INDUSTRY AND IN THE PRECEDING YEAR



[105] *See infra* Appendix Table A.

The binned scatterplots in Figure 3a, in which each dot represents one of 50 quantiles, show the trend in the use of boilerplate (defined as average cosine similarity to deals in the same industry, in the preceding year) over the fifteen years in the dataset. The first discontinuity line is placed at the time that the plain English regulation became effective (October 1998), and the second is placed at the time when the downward trend ends in 2003.[106] For several sections, the use of boilerplate shows an upward trend prior to October 1998, and the trend turns immediately downward afterward. For Risk Factors, the downward trend begins slightly earlier in 1997, corresponding to the SEC's earlier targeting of boilerplate risk factors.

As previously noted, the downward trend in boilerplate reverses sharply in 2003, and the rate at which language is copied climbs for the entire document and the individual sections except for the Business Description. It is not clear why this may have happened, although one possibility is that the enactment of Sarbanes-Oxley and its implementing regulations at that time led risk averse issuers to rely on precedent documents to comply with new disclosure requirements. If company officers and their counsel were risk averse in complying with new disclosure requirements, then they might have been more likely to borrow precise prospectus language that had already been vetted by other law firms, the SEC staff, and market participants.

The upward trend in the use of boilerplate continues through the early part of the century. This may occur for the same reason, as the SEC continued to issue a steady flow of highly technical disclosure recommendations and regulations.[107] However, whether this occurred due to new regulations or not, the upward trend indicates that the SEC's boilerplate regulations from the late 1990's may have lost effectiveness (or perhaps were overwhelmed by other regulations) over time.

3. Targeted Regulation of Boilerplate in the MD&A

Although this analysis does not provide a precise measure of the impact of the SEC's regulations or Sarbanes-Oxley in affecting the use of boilerplate language, a rule change from that period creates a quasi-experiment that can be used to assess the impact of the SEC's stylistic regulations. In 2003, the SEC attempted to limit the use of boilerplate disclosure by issuing multiple releases specifically admonishing against repetitive, cut-and-paste

---

[106] The plain English Rule was proposed in early 1998, and formally went into effect in October 1998. *See* Plain English Disclosure, Security Act Release No. 33-7497, Exchange Act Release No. 34-39,593, Investment Company Act Release No. 23011, 63 Fed. Reg. 6370, 6370 (Feb. 6, 1998); *see also* 17 C.F.R. §230.421(d) (2011).

[107] Numerous new requirements grew out of Sarbanes Oxley itself, as the SEC refined its implementation of the law. *See generally* SEC 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 (2008), https://www.sec.gov/about/offices/oca/acifr/acifr-finalreport.pdf [perma.cc/HJL4-J835] (recommending measures to streamline financial reporting and eliminate redundancies).

language in the MD&A.[108] The SEC releases were merely guidance, but such guidance is usually followed by practitioners who wish to avoid upsetting the regulators or drawing comments on their S-1 filings that could cause costly delays. The specific targeting of the MD&A in SEC releases provides an opportunity to assess the SEC's ability to rein in duplicative disclosure, at least through tools that come short of official regulation.

As indicated in Figure 3a above, an upward trend in boilerplate is evident for every section except the Business Description following 2003. Appendix Table B below sets out estimates from a difference-in-difference analysis of the change in duplicative language (within year and industry) in the MD&A section following the SEC's pronouncements.

Table B shows difference-in-difference estimates for similarity (within year and industry). Column 1 shows the analysis with respect to deals done by underwriters' counsel, and column 2 shows the analysis with respect to deals done by underwriting banks. The "After" variable is defined as 1 after the second of the SEC's pronouncements restricting MD&A boilerplate disclosure.[109] The "Treatment" variable is defined as 1 for the section affected, the MD&A. The interaction term is the difference-in-difference estimate. Both specifications include fixed effects for industry, year, each section of the document, and each law firm or investment bank. The analyses were conducted as a time series for which each unit of observation was defined as the deal-to-deal change in boilerplate (for each section and the entire document). The analysis was done three times, to assess the impact of the new guidance with respect to the three parties in the best position to interpret and act upon the guidance: the underwriters, the underwriters' counsel and the issuers' counsel. Thus the analysis was done using: the deal-to-deal change with respect to each underwriter, the deal-to-deal change with respect to law firms that represented underwriters in the dataset, and the deal-to-deal change for issuers' lawyers. The estimates imply that the SEC's attempt to reduce the use of boilerplate led, if anything, to an increase in the magnitude of duplication and boilerplate in the MD&A section, with respect to deals done by underwriters and their counsel. No effect is seen for deals done by issuers' law firms after the SEC pronouncements.

A few caveats are in order with respect to this analysis. The validity of the results depends on the assumption that trends in the treatment group (the

---

[108] *See supra* note 61 and accompanying text.

[109] This method follows the generally accepted technique for differences-in-differences, clustering standard errors at the level of the relevant actor (bank or law firm) as the entity applying the policy change. *See generally* Marianne Bertrand, Esther Duflo & Senhil Mullainathan, *How Much Should We Trust Differences-in-Differences Estimates?*, 119 QUAR. J. ECON. 249 (2004). For a good general discussion of differences-in-differences techniques and potential pitfalls, see Jonah B. Gelbach & Jonathan Klick, *Empirical Law and Economics*, *in* 1 THE OXFORD HANDBOOK OF LAW AND ECONOMICS 29, 31–33 (2017). *See generally* Vladimir Atanasov & Bernard Black, *Shock-Based Causal Inference in Corporate Finance Research*, 5 CRITICAL FIN. REV. 207 (2016) (discussing best practices for using causal inference techniques).

MD&A) and the control groups (the other sections) would have remained the same if not for the policy intervention, once relevant controls are used.[110] Figure 3a above shows that the trends for all groups, apart from the Business Description, were the same around the time of the intervention and afterward. One might suppose that the trend in the MD&A would have been different in any event because of the raft of new regulations following Sarbanes-Oxley. However, those regulations did not speak to boilerplate, unlike the very targeted guidance the SEC issued. A plausible mechanism for the observed effect that would comport with theory is that in targeting the MD&A, the SEC signaled that it would be paying closer attention to that section, and ironically, this caused issuers and their lawyers to behave more cautiously, and employ more boilerplate language which they thought would be safe. Another plausible explanation is that there were strategic reasons to use vague boilerplate language in that section given that the SEC would be giving it more scrutiny. Another possibility is that short of an official rulemaking, the guidance was ignored, although that would not explain why the use of boilerplate increased more in the MD&A than it did in other sections.

The overall conclusion to be drawn is that the SEC's attempt to limit boilerplate in the MD&A in 2003 largely resulted, if anything, in greater amounts of copied language. This highlights a potential challenge for the SEC as it attempts to reform disclosure. Any reform efforts face countervailing forces from market participants who, caught between uncertain new regulation and the SEC's efforts to make disclosure comprehensible to investors, would rather conservatively use standard language than listen to the SEC and take chances on newly drafted disclosure text. As the next section elaborates, this is more than just an issue of style. Although boilerplate may save issuers money, it also obscures information and leads to undesirable results for issuers and investors alike. The analysis in the following sections explores boilerplate's potential to obscure or convey information.

4. Regulation of Plain English

The SEC's plain English policy can similarly be analyzed with respect to its effect on issuer's drafting practices. As with boilerplate, I look for a break in the trend with respect to plain English usage at the time the plain English policy was enacted. Once again, a look at the data for that time period reveals that such a break exists. Plain English usage (as measured using the plain English scoring system described above) demonstrates a relatively stable, and in some sections a gentle upward, trend until January 1998 (when the plain English rule was announced, although before it became effective in October). After this time, plain English scores exhibit a steep upward trend, peaking in mid-2000. This trend can be seen in all sections and

[110] *See* Gelbach & Klick, *supra* note 109 at 31.

the document as a whole, and is confirmed in statistical tests.[111] The fact that all sections are affected is noteworthy given that the plain English rule itself originally applied only to the Risk Factors and summary information, and did not apply explicitly to other sections.[112] The SEC encouraged issuers to use plain English everywhere, but did not mandate it via a rule.[113] The data show evidence that the SEC's rule impacted not only its target disclosure, but that it created spillover effects to other parts of the prospectus. The trend can be seen in Figure 3b below.

[111] *See infra* Appendix Table B.

[112] *See* Plain English Disclosure, Security Act Release No. 33-7497, Exchange Act Release No. 34-39,593, Investment Company Act Release No. 23,011, 63 Fed. Reg. 6370, 6370 (Feb. 6, 1998); *see also* 17 C.F.R. § 230.421(d)(1) (2014); OFFICE OF INV'R EDUC., SEC, *supra* note 61, at 3, 65.

[113] *See* Plain English Disclosure Plain English Disclosure, 63 Fed. Reg. at 6370; *cf.* Mark Malyszko, *Compliance Clarified: Plain English*, COMPLIANCE REP. Aug. 18, 1997, at 12. The SEC Chairman at the time the rules were passed, Arthur Levitt, wrote a forward to the Plain English Handbook that ended with: "I urge you—in long and short documents, in prospectuses and shareholder reports—to speak to investors in words they can understand." OFFICE OF INV'R EDUC., SEC, *supra* note 61, at 4. Similarly, the SEC in its proposing release for Rule 421 stated: "Our ultimate goal is to have all disclosure documents written in plain English . . . . We also encourage you to use these techniques for drafting your other disclosure documents." *See* Plain English Disclosure, Securities Act Release No. 33-7380, Exchange Act Release No. 34-38,164, Investment Company Act Release No. 22,464, 62 Fed. Reg. 3152, 3154, 3159 (proposed Jan. 21, 1997).

FIGURE 3B: PLAIN ENGLISH SCORES OVER TIME

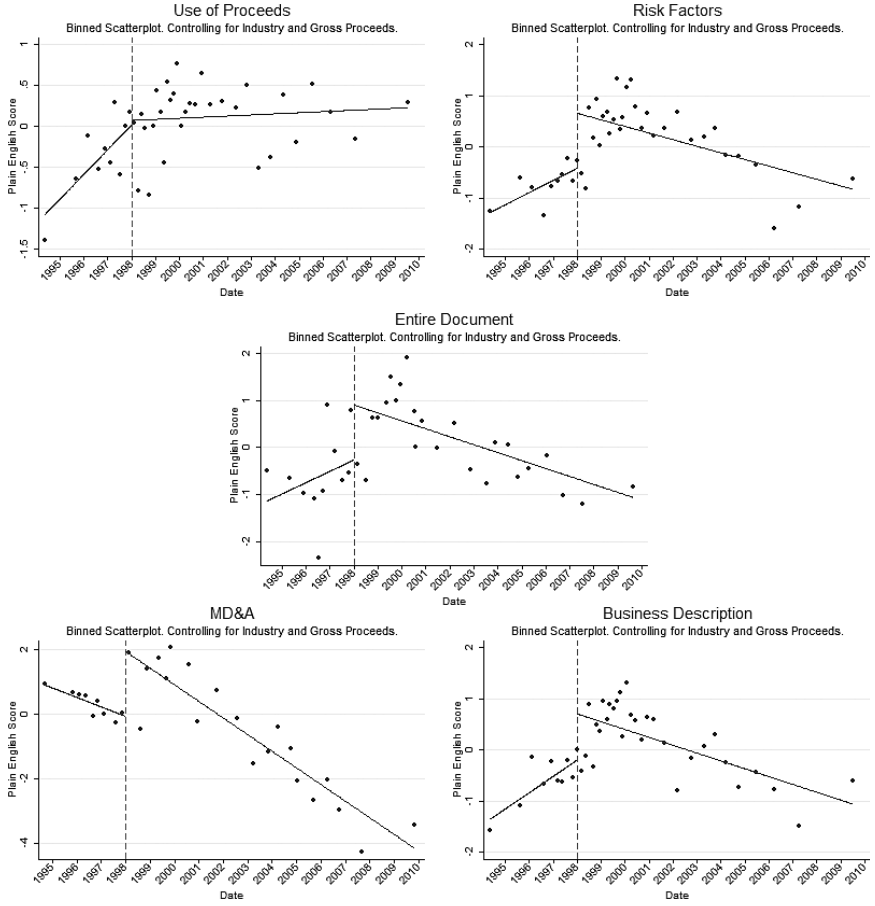


Figure 3b shows binned scatterplots (in which each dot represents one of 40 quantiles) of the average plain English scores over time, with controls for gross proceeds and industry category so that the trend can clearly be seen, with other relevant factors being equal. Higher scores mean greater compliance with the stylistic points that the SEC laid out specifically in the rule and in its *Handbook*.

As the figure shows, the adoption of the rule created a sudden spike in better plain English drafting (at least, as the SEC defines it). The trend begins to taper off shortly after that, with a downward drift in scores beginning in mid-2000. Notably, plain English scores in the highly technical MD&A drop to well below their pre-plain English averages, perhaps due to new regulations that mandated greater technical disclosure. The Use of Proceeds is the only section for which the plain English average does not drop off after the initial spike.

It is not clear why the plain English scores taper off over time, although it is possible that for both plain English and boilerplate, compliance with the regulation slipped once the SEC or the bar stopped focusing on it. Given the subjective nature of the regulation, it could also be the case that the SEC was unwilling or unable to enforce the policy, and allowed compliance to slacken over time. In either case, theory would predict that issuers would try to evade the plain English rules if the cost of complying with them were greater than the cost of any potential penalty.[114] Given that the rule does not impose harsh penalties other than the possibility of drawing SEC comments, one might conclude that it would be natural for issuers to ignore the rule if they could. However, as the discussion and analysis below indicate, complying with plain English reveals no association with higher costs or slower transactions, and so the data undercut this theoretical explanation for the rule's waning effect after 2000.

F. Information Effects of Boilerplate and Plain English

The discussion so far has focused on the ability (or lack thereof) of the SEC to regulate disclosure through regulating its style. But perhaps a more important question is: what do these stylistic issues mean for the ability of a prospectus to convey information to investors? Investor reaction to the disclosure provides one means of assessing if any effect is present. However, investor reactions are difficult to measure directly. To tease out the possible impact on investors, I use a number of proxies that are employed in the financial economic literature as indirect indicators of impact on investors. These proxies are: the readability of the documents, the accuracy of the offering price that is ultimately set for the IPO, and the propensity for the price to be revised or corrected as the issuer and its bankers solicit interest from investors. I explain each of these measures further below.

Before continuing, I describe three possible hypotheses about what impact disclosure style might have on investors. The first possibility is that plain English and boilerplate use have no impact whatsoever. The second possibility is that the overuse of boilerplate and poor drafting obscure information, making it harder for investors or the market to determine the appropriate price for the shares. The third possibility is that robust use of boilerplate and poor drafting practices could actually improve information flow. Boilerplate in particular could convey information as a form of standardized content, or as a modular language which seasoned investors know how to parse. Alternatively, boilerplate or drafting style could convey information through their mere presence, as a signal that affects investor's impressions of an issuer's quality or the quality of the information in the disclosure.

---

[114] For a more detailed discussion of this theory, *see* Section II, *supra*.

1. Possible Mechanisms for Information Effects

One caveat must be addressed before proceeding with this analysis. There is widespread doubt that investors actually read much, if any, of the disclosure documents produced in connection with a securities offering.[115] If it is true that investors do not read the disclosure, then stylistic issues cannot plausibly influence the market's reception of the issuer, and any indications to the contrary would be coincidental.

I am sympathetic to the argument that most investors do not read much of the disclosure, particularly individual retail investors. However, there are good reasons to think that the disclosure matters by way of several plausible mechanisms. First, with respect to most IPOs, few of the initial investors are retail investors, but rather large institutional money managers such as mutual funds (e.g., Vanguard, Prudential or MFS), pension systems (e.g., CalPERS), hedge funds and some wealthy individuals.[116] These investors are the key actors for indicating demand for the IPO security that in turn determines the price, and to some degree the market reaction.[117] These institutions employ individuals who can process and analyze the disclosure, and pay attention to at least some parts of it.[118] Investment analysts also play a part in this process. Although most firms have relatively few investment analysts following until after the IPO stage, these analysts play an important role in communicating with investors.[119] These individuals are likely to read at least some of the disclosure.

Second, the issuer's road show involves presentations that must be based on the disclosure in the prospectus. Stylistic concerns do not bear directly on such presentations, but they may have indirect effects, because the investors quickly reading through a prospectus at a road show may need to process it quickly to determine what kind of further information is needed.

---

[115] *See, e.g.*, Ben-Shahar & Schneider, *supra* note 7, at 647–50; *see also* Feit v. Leasco Data Processing Equip. Corp., 332 F. Supp. 544, 565 (E.D.N.Y. 1971), in which Judge Weinstein quoted:

> There are also the perennial questions of whether prospectuses, once delivered to the intended reader, are readable, and whether they are read. The cynic's answer to both questions is 'No'; the true believer's is 'Yes'; probably a more accurate answer than either would be: 'Yes'—by a relatively small number of professionals or highly sophisticated non-professionals; 'No'—by the great majority of those investors who are not sophisticated and . . . are not 'able to fend for themselves' and most 'need the protection of the ['33] Act.'

Id. at 565 (quoting Milton H. Cohen, *"Truth in Securities" Revisited*, 79 HARV. L. REV. 1340, 1351–52 (1966)).

[116] *See* Kasim Alli, Jot Yau & Kenneth Yung, *The Underpricing of IPOs of Financial Institutions*, 21 J. BUS. FIN. ACCT. 1013, 1014–16 (1994); Sean J. Griffith, *Spinning and Underpricing: A Legal and Economic Analysis of the Preferential Allocation of Shares in Initial Public Offerings*, 69 BROOK. L. REV. 583, 583–90 (2003).

[117] *See* Alli, Yau & Yung, *supra* note 116, at 1014.

[118] *See id.*

[119] *See* Jill Fisch, *Does Analyst Independence Sell Investors Short?* 55 UCLA L. REV. 39, 40–43 (2007); *see also* Shefrin & Statman, *supra* note 22, at 21–29.

Third, the process of drafting better disclosure itself might prompt information discovery. For example, if an issuer wants to edit a boilerplate risk factor to say something more specific about the company, its lawyers may have to investigate just how likely the risk is, and what mitigating measures the issuer has taken. Thus, attention to style in the drafting effort can plausibly help to produce substantive information that is valuable to investors. Looking for indications of how investors might receive the information is a reasonable approach to assessing the impact of stylistic regulations.

2. First Day Returns

The performance of the IPO stock on the first day it is offered provides information about the views of the professional investors who are first offered the IPO stock. In order to better understand how to interpret first day price increase it is necessary to explain briefly the underpricing phenomenon that occurs in virtually every IPO.

a. Underpricing as a Proxy of Investor Uncertainty

The underpricing phenomenon is the propensity for the stock price of many IPOs to rise by a large amount on the first day of trading in the market. The phenomenon is called underpricing because a large increase in the market price implies that the initial offering price was set much lower than what the market would bear.[120] Numerous theoretical explanations for underpricing have been advanced.[121] Some element of underpricing is intentional; investment banks routinely and transparently aim to offer IPOs at a 15% discount to what they think the market will bear.[122] A number of explanations have been offered for why underpricing often exceeds the relatively modest intentional amount. One such explanation is that the underpricing serves as a compensation mechanism for investment banks' favored institutional clients, who often bear risk by agreeing to purchase shares in IPO

[120] *See* Griffith, *supra* note 116, 583–90 (2003).

[121] Most of the theoretical explanations come from finance literature. *See generally* James R. Booth & Richard L. Smith III, *Capital Raising, Underwriting and the Certification Hypothesis*, 15 J. FIN. ECON. 261 (1986); Kevin Rock, *Why New Issues Are Underpriced*, 15 J. FIN. ECON. 187 (1986); Seha M. Tiniç, *Anatomy of Initial Public Offerings of Common Stock*, 43 J. FIN. 789 (1988). 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); Barondes et al., *supra* note 30, at 16–21; 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); Richard A. Booth, *Going Public, Selling Stock, and Buying Liquidity*, 2 ENTREPRENEURIAL BUS. L.J. 649, 649–51 (2007); James Spindler, *IPO Underpricing, Disclosure, and Litigation Risk* 15–16 (Univ. S. Cal. Law Econ. Working Paper Series, Paper No. 94, 2009), http://law.bepress.com/usclwps-lewps/art94 [https://perma.cc/ZWY2-2X9F].

[122] *See* Rock, *supra* note 121, at 188.

issuers.[123] The banks need these investors to ensure adequate demand for stock in certain offerings, including offerings in which the risk of return is uncertain. According to this theory, stock issuances for companies about which there is less information will exhibit more underpricing, related to the greater variance in predictions of the company's performance, and the institutional investors need a greater compensation for the risk of investing.[124]

A related explanation for underpricing holds that the discount provides an incentive for investors to produce their own information about the company, instead of forcing the company and the underwriters to invest in producing information.[125] According to this theory, the large discount from what the market will bear provides initial investors with the incentive to investigate the company and buy into the offering, even though the prospectus produced by the issuer and the banks is not particularly informative.[126] The level of first day returns (or the first day bounce), represents, among other things, an indication of the level of information gathering that the parties to the deal have prior to the offering. The theory behind this conclusion is that less information about the company in the prospectus means that the initial investors, whose level of demand set the price in the bookbuilding phase, will demand a greater discount to purchase shares in the offering.[127]

These explanations share the underlying idea that underpricing is related to a lack of information about the company and the perceived uncertainty about its future performance. Consequently, if plain English usage increases or decreases the information quality of securities disclosure, this should have an impact on the level of underpricing seen in the IPO as well.[128]

---

[123] *See generally* Murat M. Binay et al., *The Role of Underwriter-Investor Relationships in the IPO Process*, 42 J. FIN. QUANTITATIVE ANAL. 785 (2007).

[124] *See* Hanley & Hoberg, *supra* note 28, at 2823.

[125] *See id.*

[126] *See id.*

[127] 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 121, at 187–212. It should be noted that other theories have been advanced to explain underpricing. *See* Ritter & Welch, *supra* note 18, at 1795–1828. The most prolific scholar of underpricing has stated that the true explanation most likely combines a number of theories previously advanced. *See id.* 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). Thus, the first day bounce can be used as a proxy for information completeness within the market. This is consistent with recent findings that small companies going public under the emerging growth company on-ramp under the USA JOBS Act have seen greater levels of underpricing: since there is less information available about these companies than there is even with respect to typically IPO firms, investors demand a steeper initial discount. *See* Mary E. Barth, Wayne R. Landsman & Daniel J. Taylor, *The JOBS Act and Information Uncertainty in IPO Firms*, ACCT. REV., Nov. 2017, at 25.

[128] *See generally* Arnold, Fishe & North, *supra* note 27; Hanley & Hoberg, *supra* note 28.

b. Analysis of Underpricing Data

The raw data support the hypothesis that boilerplate is related to investor uncertainty; however, the data show less support for concluding the same about plain English drafting. Appendix Table C demonstrates the relationship between prospectus language similarity, plain English, and first day price returns. The dependent variable is the percentage price increase on the first day of trading and the main independent variable is the average similarity between IPO prospectuses (as a whole, and broken down by section).

The regression uses the log of the gross proceeds for each deal as a control for the quality and size of the issuer.[129] The measure is used in numerous finance studies, and corresponds well to issuer quality and assets. The regression also uses fixed effects for each underwriter, each law firm, each industry, the presence of venture capital backing in the deal, each year, and the interaction of year and industry group. In addition, in alternative specifications, controls are included for the company age, the log of the issuer's assets, the market share of the underwriting banks and law firms in the preceding year.

As the table shows, for several of the sections as well as for the prospectus as a whole, the average first day bounce increases as the level of boilerplate language increases. The coefficients in the table can be interpreted as a change in underpricing for each level of cosine similarity equal to one. Since each standard deviation is close to 0.10, a simpler way to interpret the coefficients is to calculate change in returns for each 10% increase in similarity, as per column three of Appendix Table D.[130]

The table indicates in particular that increased use of boilerplate in Risk Factors, MD&A, and Use of Proceeds is associated with underpricing, which can be interpreted as greater noise in the offering process and less consensus among investors, related to more boilerplate in the offering document.[131] The coefficient on the plain English score is not significant for any section of the document, and therefore one cannot draw any definitive conclusions about its importance for purposes of first day returns.

I note that the analysis does not demonstrate that these relationships are causal. It could be the case that the boilerplate obscures information, or it

---

[129] The gross proceeds are highly correlated with the size of the issuer, and are frequently used as a measure of the issuer's quality. *See* B. Espen Eckbo et al., 1 HANDBOOK OF CORPORATE FINANCE: EMPIRICAL CORPORATE FINANCE 287–88 (B. Espen Eckbo ed., 2007).

[130] These results are consistent with other research on standard language in IPOs and first day price returns. *See* Hanley & Hoberg, *supra* note 28, at 2841–42.

[131] The argument that unclear disclosure generally affects underpricing has been advanced elsewhere and supported by other empirical studies. *See* Spindler, *supra* note 122, at 30. *See also* John L. Campbell et al., *The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings*, 19 REV. ACCT. STUD. 396, 396 (discussing market uptake of risk information); Todd D. Kravet & Volkan Muslu, *Textual Risk Disclosures and Investors' Risk Perceptions*, 18 REV. ACCT. STUD. 1088, 1113 (2013) (discussing market absorption of risk information from periodic filings on Form 10-K).

could be the case that more boilerplate simply goes hand in hand with a deal for which the issuer is lower quality, which leads to more underpricing. One way in which I try to rule out the second possibility is by employing controls for company age, venture capital, and amount of proceeds, since these variables act as proxies for the quality of the issuer which in turn affects the effort spent on the deal. Older companies with more assets tend to have standing relationships with their bankers and lawyers, and would expect more effort (indeed, greater deal size is strongly correlated with lower amounts of boilerplate). In addition, I undertake a number of analyses using a subset of the data involving only the busiest and most prestigious underwriters, since these banks carefully screen deals and act as lead underwriters for companies whose quality is relatively good. These analyses produce results that are consistent with those above.[132]

3. Price Revision

Another proxy for the information conveyed to investors in a deal is the propensity with which the security's offering price is revised from the initial filing range set out in the Form S-1.[133] In connection with underpricing, price revision thus provides another means to assess the importance of boilerplate and plain language in a prospectus.

a. Price Revision as a Proxy of Uncertainty

When the price is revised before the offering date, it indicates a disconnect between the initial price range set based on the initial disclosure and due diligence, and the impressions of the initial investors upon reviewing the disclosure and questioning the issuer's management during the roadshow. If investor demand for the offering is either high or low in relation to the initial price determined by the underwriters and the issuer, the price will be revised. Price revision is therefore an indirect signal about the information contained in disclosure, or the ability of investors to absorb the information, or both.

Scholars posit that there is a tradeoff between due diligence-related ex-ante pricing and ex-post price discovery through bookbuilding.[134] The due diligence process and the drafting process work in concert and affect each other—the more that parties attempt to craft specific disclosure, the more information they will need to gather, and the more likely it is that the disclo-

---

[132] *See infra* Appendix Table C.

[133] Recall that issuers set an indicative price range in the initial Form S-1, based on estimates of what the ultimate price might be. The ultimate price is either within that range, or it can be revised up or down from the initial range based on whether investor interest after the road show is stronger or weaker than expected. *See generally* Tim Loughran & Bill McDonald, *IPO First-Day Returns, Offer Price Revisions, Volatility, and Form S-1 Language*, 109 J. FIN. ECON. 307 (2013).

[134] *See id.* at 312–13.

sure will be specific and that attention will be paid to the language and drafting used.[135] The issuer and its bankers may attempt to set the "right" price range based on what they assume the demand will be, and how they believe the institutional investors will perceive the information.[136]

If there is price revision, then it suggests that either 1) the underwriters did not produce enough information to price the stock in the correct range, or that 2) information was revealed during the road show that affected initial investors' perceptions of the issuer in a way that the underwriters did not anticipate based on the information and disclosure they had, or both. In either case, if disclosure is more informative, then there will be a lower incidence of price revision during bookbuilding.[137] Likewise, if less diligence is conducted, then disclosure will reflect more generic language that has not been revised extensively, and thus is likely to comply less with plain English directives. That, in turn, results in the need to rely on the bookbuilding process for price discovery and leading to more price revision.[138]

b. Analysis of Price Revision Data

The data tell a story consistent with the explanation above. With regard to both boilerplate and compliance with the SEC's plain English guide, the raw data reveal a relationship between a document's score and price revision. However, in statistical tests, only the boilerplate measure remains significant once relevant controls are used.

Appendix Table E below provides the results of a probit regression to determine the change in probability that the offer price of a deal was revised from the initial offer range, in relation to the change in the amount of boilerplate (represented by copied language) in the prospectus and its plain English score.

The table 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, MD&A, and Use of Proceeds sections leads to a higher probability of upward price revision, indicating that the initial price was not accurately set, and in fact was likely set low intentionally to induce institutional investors to invest in their own research instead of relying on the disclosure. The component of plain English compliance dealing with factors other than boilerplate does not bear a significant relationship to price revision. This could be because no such relationship exists, or it could be the case that any such relationship is difficult to discern through the noise of the data.

---

[135] *See id.* at 312–13.
[136] *See id.* at 313.
[137] *See id.* at 313.
[138] *See id.* at 313.

4. Prospectus-related Litigation

The frequency of securities litigation, based on allegations of material misrepresentations in the prospectus, provides a final metric of the effectiveness of disclosure style. In theory, prospectus-related litigation implicates drafting in two ways. First, attorneys are thought to use generic boilerplate language as catch-all disclosure—a means to protect issuers from litigation arising out of omissions or misrepresentation in the offering document. Issuers might also try to use what the SEC's plain English Handbook describes as "poor" drafting techniques like long sentences, legalese or double negatives to disclose risks in ways that are not readily apparent to investors. Ironically, such disclosures often fail to protect issuers from liability,[139] and some theorize that they attract litigation.[140]

Second, even if disclosure is not generic or overly complex, unclear language may obscure problems with the issuer or mislead investors in ways that are colorably actionable.[141] For example, despite disclosure's role in protecting issuers from lawsuits, unclear or boilerplate language in sections such as the Risk Factors has been held by courts not to protect issuers, because the language is too vague to be truly informative.[142] Moreover, expansive use of boilerplate and poor drafting may signal to investors a lack of careful investigation of the company, calling into question the disclosure as a whole. Thus, boilerplate and plain English writing may have a relationship to litigation if they obscure or clarify important information.

The data reveal a relationship between the amount of boilerplate in some sections of the disclosure and litigation risk; however, no such relationship exists with respect to plain English use. Figure 4 below depicts the relationship in the raw data between risk factor boilerplate and litigation. The graph demonstrates a positive relationship between risk of litigation and boilerplate, indicating an increased probability of litigation as boilerplate increases past 30%.

---

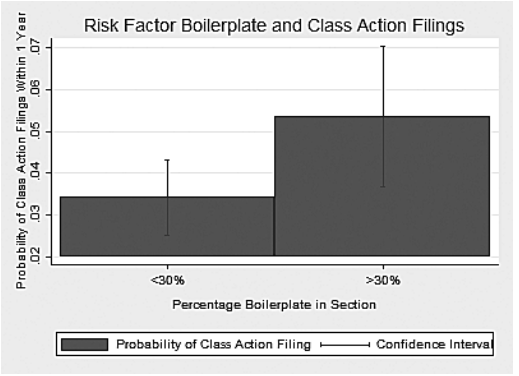
[139] *See e.g.*, Slayton v. Am. Expres. Co., 604 F.3d 758, 771 (2d Cir. 2010); *see also* Erin M. Hardtke, *What's Wrong with the Safe Harbor for Forward-Looking Statements? A Call to the Securities and Exchange Commission to Reconsider Codification of the Bespeaks Caution Doctrine*, 81 MARQUETTE L. REV. 133, 140–41 (1997). As noted above in Section 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). However, the safe harbor explicitly does not apply to IPOs. *See id.* § 78u-5(b)(2)(D).

[140] *See* Spindler, *supra* note 121, at 4.

[141] For further discussion of this theory, *see* Alexander, *supra* note 121; *see also* Saumya Mohan, *Disclosure Quality and Its Effect on Litigation Risk* (Jan. 13, 2007) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=956499.

[142] *See Slayton*, 604 F.3d at 758–59.

FIGURE 4: RISK FACTOR BOILERPLATE AND CLASS ACTION FILINGS



The relationship holds for several sections of the prospectus in a regression analysis. A probit regression of the occurrence of boilerplate on class action litigation filed within the first year of an IPO reveals a significant association between the Risk Factors and the Business Description sections. Risk factors show a 2-3% greater probability of litigation for every 10% increase in risk factor boilerplate, but a 2-3% decrease for a 10% increase in boilerplate in the business description. None of the coefficients for plain English are significant, although for most sections they are negative, which would be expected, since plain English corresponds to less misleading information in the prospectus. The results are set out in Appendix Table F.

These associations must be taken cautiously, and they do not prove a causal relationship. It could be the case that riskier companies include more boilerplate risk factors, and their higher litigation risk is related to their underlying riskiness. I attempt to rule this out by controlling for a number of indicators of company quality, as well as other factors known to be associated with litigation risk that are not reported in the table: gross proceeds (because deep pockets attract litigation), company age (younger companies are often riskier), investment bank and law firm identity (indicating quality of the company and the due diligence), leverage (companies with more debt are more likely to sued), and sales (known to be associated with litigation risk). The results are robust to these controls, but nonetheless cannot confirm a causal story. However, it is clear that boilerplate is related to litigation, and even if it is not itself a cause of litigation, it may provide a signal of greater litigation risk to investors.

G. Effect of Stylistic Regulations on the Costs and Speed of Transactions

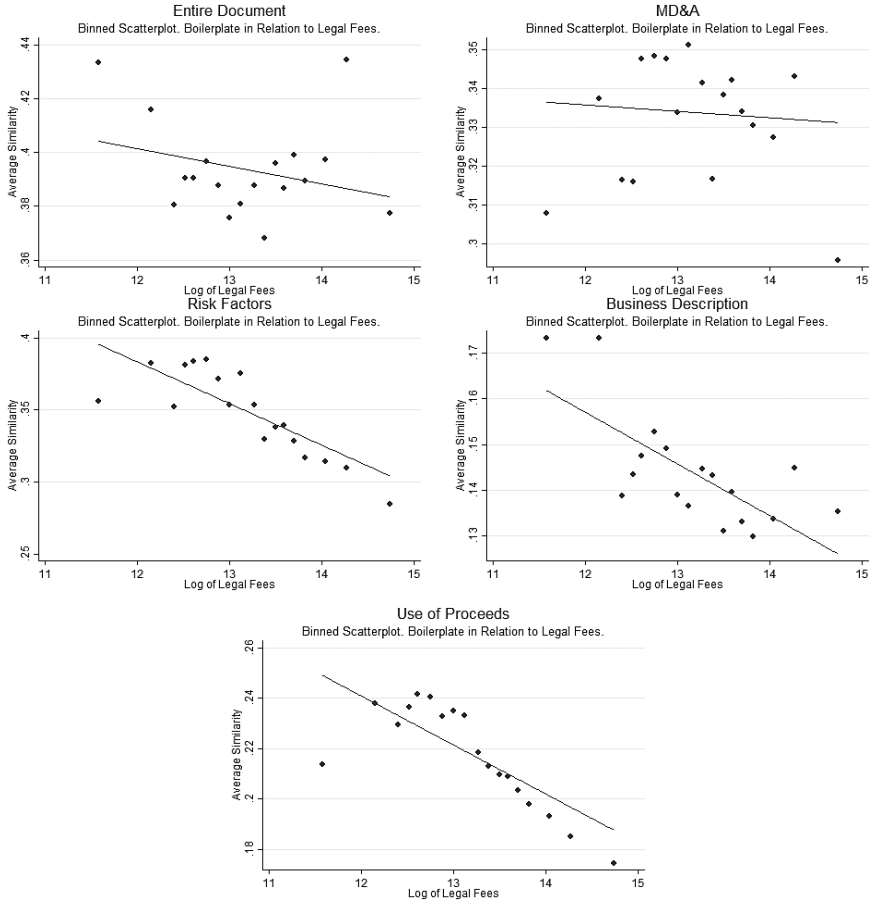
The analysis above demonstrates some problems associated with poor drafting, especially the overuse of boilerplate. But what are the benefits of using cut-and-paste prospectuses and ignoring plain English? Two potential

benefits are lower fees to counsel, and faster deal completion. I analyze each of these here.

1. Legal Fees

Analysis of the data supports the hypothesis that lower legal fees are associated with more boilerplate and vice versa. Figure 5 depicts the relationship between aggregate legal fees and average boilerplate compared to recent industry deals.[143]

FIGURE 5: BINNED SCATTERPLOT OF COSINE SIMILARITY VERSUS LEGAL FEES (AS THE NATURAL LOG OF LEGAL FEES)



[143] *Cf.* Eckbo et al., *supra* note 159, at 275–355 (reviewing the voluminous empirical finance literature on IPOs).

The graphs above show scatterplots (with individual observations averaged and plotted in bins) of the relationship between the use of boilerplate within industry in the recent past, and legal fees. Legal fees in the dataset range from $30,000 to over $5 million. The graphs above are presented in terms of the log of legal fees, instead of raw numeric values, in order to show the relationship between a percentage increase in legal fees and the change in the use of boilerplate. The graphs above represent only the raw data; however the relationships in the graphs also bear out in linear regressions with controls for year, industry, gross proceeds, age of the company, the presence of venture capital backing, and fixed effects for underwriting banks, as set out in Appendix Table G. The analysis reveals that each additional 10% of similarity between documents is associated with a decrease in legal fees equaling between $39,600 and $52,000 on average.[144]

As would be expected, for many of the individual prospectus sections, higher legal fees correspond to less identical language, indicating that there is more tailoring of the documents occurring. Notably, however, the trend is weaker for the document as a whole, and for the MD&A section. The document as a whole may have a weak relationship because of the occurrence of language that issuers must include, but that does not bear in any significant way on the deal. The lack of a significant relationship with respect to the MD&A is also interesting, as the MD&A has been the focus of SEC regulation. Nonetheless, the relationship above demonstrates that, on average, the greater use of boilerplate is less costly to issuing firms.

Surprisingly, greater compliance with the stylistic points in the SEC's plain English handbook is associated with lower legal fees, as set out in Appendix Table G, although the statistical significance is weak for most sections. The negative relationship is surprising because one would expect compliance with the SEC's various stylistic directives to require more lawyer hours, and therefore entail greater fees. However, the average reduction in fees related to a greater plain English score is very small compared to that associated with greater boilerplate. As with other analyses, it is difficult to measure exactly how much the fees fluctuate in relation to specific components of plain English usage since it is measured using a scaled score, however, the direction of the trend is informative.

2. Time to Transaction Completion

By contrast, compliance with plain English rules have no discernable association with the speed at which a deal is completed, as set out in Appendix Table G. Timing is measured as the length of time that elapses between the filing of the S-1 with the SEC and the date of the offering. One might hypothesize that the time to transaction completion would be strongly affected by the SEC's stylistic regulations for two reasons. First, one would

[144] *See infra* Appendix Table D.

assume that drafting and editing non-standard disclosure would take longer than using boilerplate; the same would be true for lawyers paying attention to the SEC's grammar and style directives. Second, non-compliance with the rules might draw SEC comments, forcing time-consuming revisions and re-drafting of the prospectus.

The analysis does not reveal any statistically significant relationship between compliance with the style rules and timing of the offering. While the lack of result does not definitively indicate that there is no relationship, it is noteworthy that any such relationship is not strong enough to be revealed either in the raw data or in a regression analysis. Transaction speed, as analyzed here, is measured from the time that the S-1 is filed, and so it is possible that any decrease in transaction speed from following plain English rules occurs before the filing, while the document is being drafted. In addition, given that the timing of an IPO depends on numerous factors including the conditions prevailing in the market, if any timing benefit exists from ignoring the rules, it may be the case that the benefit is attenuated by other circumstances. Nonetheless, the absence of any measurable increase in speed garnered by disregarding the concerns that are central to the SEC's plain English rules indicates that after the S-1 is filed, neither SEC review nor investor reaction to the purportedly bad drafting style have an appreciable effect on the speed with which the deal is closed, on average.

H. Summary and Synthesis of Results

A number of conclusions can be drawn from the analysis above. Even though the style regulations were adopted for "ordinary" investors, the analysis above provides evidence that disclosure style might matter for sophisticated investors and issuers as well. The overuse of boilerplate language and lack of attention to clarity can mean that issuing companies leave money on the table and increase their litigation risk.

The graphs below help to visually summarize the results while also representing a paradox that the SEC faces in regulating disclosure.

FIGURE 6A: UNDERPRICING AS A FUNCTION OF PROSPECTUS LENGTH (AS TOTAL NUMBER OF WORDS)

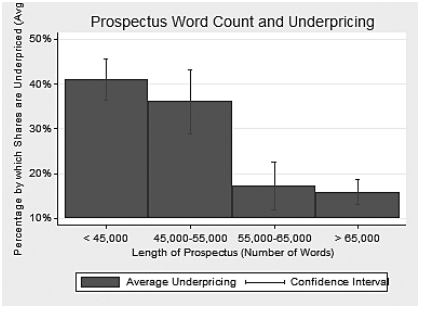
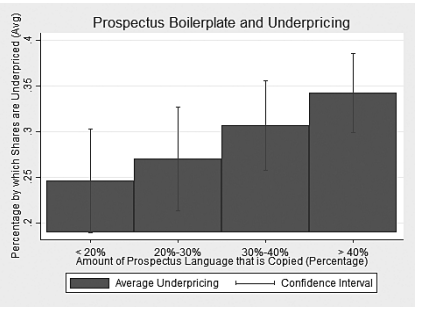


FIGURE 6B: UNDERPRICING AS A FUNCTION OF BOILERPLATE (AS PERCENTAGE OF TEXT COPIED FROM RECENT SIMILAR DEALS)



The graph on the left shows the relationship between prospectus length and underpricing in the raw data. On the right, the graph shows the relationship between underpricing and boilerplate. On the one hand, longer prospectuses may be harder for investors to read and digest, but they are nonetheless associated with less investor uncertainty, on average, as proxied by underpricing. However, greater levels of boilerplate, while providing investors with a better basis for comparison to recent transactions, result in greater investor uncertainty. The results support the conclusion that more information is not necessarily bad, and that the form of information is important for communicating to investors. At the same time, the analysis indicates that stylistic directives may have only temporary effect, and that over time other considerations weigh more heavily in issuers' decisions about how to tailor their disclosure.

If stylistic disclosure regulations are worthwhile, then the important question then is why compliance with the regulations drops off, and what the SEC can do to ensure future compliance. One reason that compliance might drop off is that the cost of ignoring the plain English rule is less than the costs of following it. That would be true if any potential sanction for ignoring the rule were outweighed by the costs of compliance. That possibility is not supported by the data, however. Although issuers do save legal costs from using boilerplate, they do not appear to save much if anything from ignoring the stylistic points in the plain English Handbook. And even though the sanction for ignoring the rule is probably relatively small in most cases (SEC comments and delay of the deal) there are unofficial sanctions that the market appears to impose for violating at least some of the plain English directives. For instance, although boilerplate may save legal fees, those savings must be weighed against the costs imposed by greater underpricing, and the greater potential for litigation that are associated with such drafting practices. . To illustrate the point, consider an issuer from the dataset which pays, on average, $52,000 less in legal fees for each additional 10% of boilerplate disclosure in a whole prospectus. If the additional 10% were found in risk factors, however, the transaction would lose, on average, approximately $5.1 to $6.1 million to underpricing.[145] In addition, the company would face extra litigation risk, increasing the expected settlement amount of a class action claim by approximately $500,000 on average, and possibly much more if the litigation went to trial or failed to settle within the typical range.[146] That

[145] The average size of deals in the dataset is $93,200,000. This number would be multiplied by the increased probability of litigation.

[146] The average probability of class action litigation within the first year for all deals in the dataset is 4 percent. 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* CORNERSTONE RESEARCH, SECURITIES CLASS ACTION SETTLEMENTS, 2015 REVIEW AND ANALYSIS 8–10 (2016) (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 2%) multiplied by the expected average loss ($25 million). Disclosure dollar loss – the amount of stock market

amount includes only the actual settlement; one would still have to add attorneys' fees, time costs, reputation costs and the loss of stock market value of the company's stock to get the full impact of such litigation. Thus, even assuming that some of the costs are justified and there are time savings from using boilerplate or ignoring plain English that are not captured by legal fees, it is hard to see how poor drafting practices outweighs their costs. There may be other benefits from ignoring plain English, but it is not clear what they might be. For example, it is unclear why ignoring the plain English rules would convey uniformity, compliance with market norms, or anything positive about an issuer at all. It is possible that regulations like the plain English rule focus the attention of issuers and their advisors on stylistic concerns, but that their attention wanes as other issues overtake the SEC's regulatory agenda. The analysis in this Article does not give a definitive answer to the question, but it provides an area to be explored in future work.

IV. Stylistic Disclosure and the Securities Act Regime

The primary lesson of the SEC's experience with the plain English rules is that stylistic regulations work, and there is evidence that they matter. The SEC's efforts to regulate the form of disclosure can be effective, and there is evidence that stylistic choices have substantive consequences for issuers and investors. However, the results also show that the effect of such regulations is ephemeral and boilerplate and non-plain English drafting resurge when issuers have incentives to use them.

The ultimate implication of the results with respect to quantifiable costs is that through style regulation, government actors can ease the burdens of other regulations on both the regulated parties and the intended beneficiaries of the regulation. In other words, ironically, sometimes more regulation may help to ease the burdens of pre-existing regulation. These general conclusions have implications for future style regulations, as well as disclosure regulation generally. The discussion below describes the implications that this study supports, and makes suggestions for the SEC's reform efforts.

A. Implications for Future Disclosure Regulation

Securities disclosure reform faces an inherent tension: investors complain that volume of disclosure is a problem, but they want more anyway. Stylistic regulation is one way of dealing with the problem by mandating

---

value lost after a company discloses a securities class action settlement – would be even larger. The value of such losses ranged from an average of $89 million to $121 million in the period studied. *Id.* at 11; *see also* Michelle Lowry & Susan Shu, *Litigation Risk and IPO Underpricing*, 65 J. Fin. Econ., 309, 310–15 (2002) (noting average settlement payment); Qing Hao, *Securities Litigation, Withdrawal Risk and Initial Public Offerings*, 17 J. Corp. Fin. 438, 454 (2011) (reporting the results of a recent empirical analysis showing no reliable relation between underpricing and subsequent litigation risk for U.S. IPOs from 1996 to 2005).

disclosures in a format that is more easily digestible to investors. Indeed, when it comes to complex instruments like asset-backed securities or collateralized debt obligations, some scholars have proposed, and the SEC has permitted, disclosure of data about the securities in tabular formats that are easier for investors in those sophisticated markets to analyze. When it comes to more traditional securities, disclosure format can similarly help investors to digest information, and format rules give issuers incentives to conduct more extensive due diligence. Rules intended to do this already exist, and the analysis in the Article shows that the rules have the potential to help both issuers and investors. A modest proposal, then, would be for the SEC to do more to enforce these rules, or at the very least, to draw more attention to them.

I recognize that issuers may have other reasons for ignoring these rules than the threat (or lack thereof) of SEC enforcement. It may, therefore, also make sense for the SEC to revisit its stylistic regulations to determine whether they still serve the needs of the issuer and investors communities, and find out if there are other reasons that compliance has dropped off over time. That knowledge would provide a basis for any future action with respect to these rules.

In addition, when reexamining stylistic regulations, the SEC can use past experience with stylistic regulation as an indicator of what kinds of disclosures are important to the market, leading to more targeted regulation. In the past, members of the Commission have stated that they are willing to streamline required disclosures.[147] Moreover, the SEC has historically espoused the view that "better disclosure is not at all synonymous with less disclosure."[148] The Director of the SEC Division of Corporation Finance reports that investors share that sentiment, and in practice they continue to demand more disclosure, rather than less.[149] However, investor groups are divided over what requirements should be cut and what kinds of disclosures are useful.[150] This presents something of a paradox in the SEC's attempts to regulate disclosure. In addressing this puzzle, the SEC could make use of the same language processing tools employed here to assess exactly which kinds of stylistic regulations issuers follow and which they have chosen to ignore over time. For instance, the regulators can relatively easily find out which kinds of boilerplate language issuers have clung to over time, and which

---

[147] *See* Keith F. Higgins, Dir., Div. of Corp. Fin., SEC, 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/9R6W-TT47] (claiming that "with hundreds of pages per filing for some large issuers," the volume of disclosures mandated by the SEC make them "very hard to evaluate."); *see also* Hu, *supra* note 9, at 1652 (citing KPMG, FINANCIAL INSTITUTION RISK DISCLOSURE BEST PRACTICE SURVEY 2008, at 5 (2008)).

[148] Manny Cohen, Chairman, SEC, Remarks Before the 19th Annual Conference of the Financial Analysts Federation (May 24, 1966).

[149] *See* Higgins, *supra* note 147.

[150] *See* Higgins, *supra* note 147.

kinds seem responsive to regulation. This would reveal the kinds of disclosures that issuers or the market find useful, or at least provide a targeted starting point for asking the market for comment. This would in turn help the SEC to walk the fine line between simplifying and preserving the value of disclosure's content as it attempts future reform.

B. Language Processing and the Obsolescence of Style

An equally important conclusion of this Article comes not from the results themselves, but from what they demonstrate: that language can be analyzed and processed easily even when poorly written or riddled with boilerplate. Thus, for all of the SEC's efforts to improve clarity, concerns about stylistic regulation of disclosure may become obsolete in the relatively near future due to machine-driven language processing. In fact, boilerplate may ironically be even more useful in the future than tailored disclosure precisely because its standardized nature makes it easier to analyze mechanically, and its presence, as demonstrated above, conveys information about issuers even independently of its content.

Clarity of drafting and the use of boilerplate are likely to continue to play an important role in disclosure, at least in the near future. However, even if clarity begins to suffer and compliance with plain English regulation slips, negative consequences may not be inevitable because the emergence of computerized language processing may obviate the importance of disclosure's style and form.[151] Indeed, the uniformity of boilerplate language may even make it easier for computer algorithms to process, and identify deviations from the norm, which can themselves provide important signals about issuers. Unlike human readers, machines are unlikely to be overwhelmed by a large volume of such disclosure, as they can quickly assess and sort through what is standard and what is not. Similarly, grammatical plain English issues are thus less significant if an algorithm is distilling the text to its most important elements. Eventually (and perhaps relatively soon) issuers may be free to ignore drafting problems if they think it more efficient to do so, and investors may be no worse off for it. Although few retail investors have access to natural language processing tools now, that is likely to change as those tools become more widely available and usable.[152]

---

[151] For an interesting argument that the reasonable investor standard is antiquated in the age of algorithmic trading, *see* Tom C.W. Lin, *Reasonable Investor(s)*, 95 B.U. L. Rev. 461, 508–09 (2015).

[152] For example, user-friendly products are now available to help lawyers assess the ambiguity and readability of contracts and M&A documents. *See, e.g.*, Intelligize, www.intelligize.com/products/transactions [https://perma.cc/2QGG-S9MV] (describing a product that compares SEC filings to look for significant deviations from the norm); Litiq http://www.litiq.com/#precise_documents [https://perma.cc/P5F8-UW9P] (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).

V. Conclusion

Stylistic regulation is a valuable tool for walking the line between useful mandatory disclosure laws and the potential for information overload that they tend to create. This Article provides evidence that such regulations can work to help make disclosure more useful to investors, but that compliance with these regulations is likely to wane without either better enforcement or more compelling reasons for issuers to follow them. Nonetheless, in addressing disclosure reforms more generally, stylistic regulations might provide information that is useful for guiding substantive disclosure regulation.

Ironically, the fact that style regulations are effective may soon be irrelevant to the disclosure debate. In an era when machine-based analysis of documents is becoming more common, style matters less. Indeed, assumedly uninformative boilerplate may actually provide more easily digestible means of assimilating information, and picking out noteworthy differences among deals, than more standard disclosure can. A number of tools that are now widely used for information retrieval and machine learning can help regulators and investors to understand boilerplate in a new way, and separate that which might be informative from that which is not.

Appendix

Appendix Table A: Change in Boilerplate in Response to Policy Change

Boilerplate before and after October 1998					
	(1)	(2)	(3)	(4)	(5)
	Risk Factors	MD&A	Use of Proceeds	Business	Entire Document
Pre October 1998	0.775***	0.372***	0.575***	0.135*	0.876***
	(0.083)	(0.073)	(0.053)	(0.055)	(0.084)
Post October 1998	0.715***	0.361***	0.524***	0.103+	0.811***
	(0.086)	(0.075)	(0.055)	(0.056)	(0.087)
Log of Gross Proceeds	-0.0139***	-0.0142***	-0.032***	-0.002	-0.008
	(0.005)	(0.004)	(0.003)	(0.003)	(0.005)
Venture capital	X	X	X	X	X
Syndicate Size	X	X	X	X	X
Log of Company Age	X	X	X	X	X
Industry Dummies	X	X	X	X	X
Bank Dummies	X	X	X	X	X
Law Firm Dummies	X	X	X	X	X
Adj. R^2	0.945	0.944	0.948	0.907	0.959
Number of Observations	2,237	2,237	2,237	2,237	2,237

Robust standard errors are in parentheses. Estimates marked with +, *, ** and *** are statistically significant at the 10%, 5%, 1%, 0.1% level respectively.

APPENDIX TABLE B: DIFFERENCE-IN-DIFFERENCE ESTIMATES

	Cosine Similarity	
	(1)	(2)
After	-0.0360	0.0517***
	(0.0309)	(0.0136)
Treatment	-0.0407***	-0.0476***
	(0.0065)	(0.0079)
After*Treatment	0.0207**	0.0373***
	(0.0097)	(0.0070)
Year Fixed Effects	X	X
Industry Fixed Effects	X	X
Document Section Fixed Effects	X	X
Mgr Counsel Fixed Effects	X	
Bank Fixed Effects		X
Clustering:	Manager's Counsel	Bank
N	8,235	12,190
Adj. R^2	0.4467	0.431

Clustered standard errors at the level of manager's counsel (column 1) and underwriter (column 2) are in parentheses. Estimates marked with *, **, and *** are statistically significant at the 5%, 1%, and 0.1% level respectively.

Appendix Table C: Change in Plain English in Response to Policy Change

Plain English Scores before and after October 1998										
	Risk Factors		MD&A		Use of Proceeds		Business		Entire Document	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Pre October 1998	4.682+	4.574	8.697***	11.513***	1.390	-0.430	3.862	-2.098	5.829	5.114
	(2.480)	(3.895)	(2.170)	(3.529)	(2.146)	(3.598)	(3.039)	(3.624)	(3.952)	(4.456)
Post October 1998	7.485**	7.510+	9.557***	12.462***	2.544	0.412	6.160*	0.144	7.761+	6.917
	(2.549)	(3.971)	(2.234)	(3.626)	(2.243)	(3.629)	(3.105)	(3.703)	(4.006)	(4.530)
Log of Gross Proceeds	-0.558***	-0.558***	-0.598***	-0.492***	-0.108	0.043	-0.372**	-0.139	-0.491***	-0.294
	(0.123)	(0.123)	(0.124)	(0.144)	(0.128)	(0.156)	(0.136)	(0.163)	(0.162)	(0.209)
Venture capital	X	X	X	X	X	X	X	X	X	X
Syndicate Size	X	X	X	X	X	X	X	X	X	X
Log of Company Age	X	X	X	X	X	X	X	X	X	X
Industry Dummies	X	X	X	X	X	X	X	X	X	X
Bank Dummies	X	X	X	X	X	X	X	X	X	X
Law Firm Dummies		X		X		X		X		X
Adj. R^2	0.214	0.268	0.251	0.272	0.186	0.151	0.258	0.384	0.089	0.166
Number of Observations	2,164	2,162	2,164	2,162	2,164	2,162	2,164	2,162	2,164	2,162

Robust standard errors are in parentheses. Estimates marked with +, *, **, and *** are statistically significant at the 10%, 5%, 1%, 0.1% level respectively.

Appendix Table D: Boilerplate, Plain English, and First Day Returns

First Day Returns			*Change in returns per 10% increase in similarity*
	(1)	(2)	(3)
Overall Similarity	**0.243***	**0.164**	**1.6-2.4%**
	(0.095)	**(0.097)**	
Overall Plain English	**0.004**	**0.005**	
	(0.004)	**(0.005)**	
Log (Gross Proceeds)	0.095***	0.076***	
	(0.140)	(0.020)	
Adj. R^2	0.193	0.316	
Risk Factor Similarity	**0.237*** **	**0.271** **	**2.3-2.7%**
	(0.118)	**(0.143)**	
Risk Factor Plain English	**-0.001**	**0.001**	
	(0.004)	**(0.004)**	
Log (Gross Proceeds)	0.135***	0.130***	
	(0.025)	(0.035)	
Adj. R^2	0.193	0.314	
Use of Proceeds Similarity	**0.644*** **	**0.497*** **	**4.9-6.3%**
	(0.127)	**(0.131)**	
Use of Proceeds Plain English	**0.004+**	**0.005**	
	(0.002)	**(0.003)**	
Log (Gross Proceeds)	0.103***	0.083***	
	(0.014)	(0.028)	
Adj. R^2	0.200	0.320	
MD&A Similarity	**0.456*** **	**0.444*** **	**4.4-4.6%**
	(0.103)	**(0.115)**	
MD&A Plain English	**0.001**	**0.004**	
	(0.005)	**(0.006)**	
Log (Gross Proceeds)	0.094***	0.078***	
	(0.137)	(0.028)	
Adj. R^2	0.200	0.321	
Business Description Similarity	**0.083**	**0.060**	**1.0%**
	(0.107)	**(0.148)**	
Business Description Plain English	**0.007**	**0.006**	
	(0.003)	**(0.006)**	
Log (Gross Proceeds)	0.092***	0.076***	
	(0.137)	(0.028)	
Adj. R^2	0.194	0.318	
Industry Dummies	X	X	
IPO Year Dummies	X	X	
Industry*Year Dummies	X	X	
Law Firm Dummies	X	X	
Bank Dummies		X	
Number of Observations	2,232	2,232	

Additional controls for issuer age, venture capital involvement, syndicate size included for all specifications but not tabulated. Standard errors clustered at the industry level are in parentheses. Estimates marked with *, **, and *** are statistically significant at the 5%, 1%, and 0.1% level respectively.

APPENDIX TABLE E: PRICE REVISION AS A FUNCTION OF BOILERPLATE

Price revision			*Probability change per 10% change in similarity*
	(1)	(2)	(3)
Overall Similarity	0.146 (0.117)	0.040 (0.187)	**0.40 to 1.46%**
Overall Plain English	-0.0003 (0.147)	-0.001 (0.003)	
Pseudo R^2	0.164	0.160	
Number of Observations	2,113	1,678	
Risk Factors Similarity	0.367** (0.120)	0.389* (0.163)	**3.67 to 3.89%**
Risk Factor Plain English	-0.003 (0.003)	-0.004 (0.004)	
Pseudo R^2	0.168	0.163	
Number of Observations	2,114	1,679	
Use of Proceeds Similarity	0.464** (0.170)	0.558* (0.263)	**4.64 to 5.58%**
Use of Proceeds Plain English	0.005 (0.003)	0.006+ (0.003)	
Pseudo R^2	0.169	0.162	
Number of Observations	2,113	1,678	
MD&A similarity	0.570*** (0.130)	0.402* (0.089)	**4.02 to 5.70%**
MD&A Plain English	0.005 (0.003)	0.002 (0.005)	
Pseudo R^2	0.172	0.163	
Number of Observations	2,114	1,679	
Business Similarity	0.377* (0.185)	0.247 (0.204)	**1.06 to 1.63%**
Business Plain English	0.004 (0.004)	0.003 (0.004)	
Pseudo R^2	0.164	0.160	
Number of Observations	2,114	1,679	
Venture capital	X	X	
Industry Dummies	X	X	
IPO Year Dummies	X	X	
Industry*Year Dummies	X	X	
Bank Dummies		X	

Additional controls for issuer age, venture capital involvement, syndicate size included for all specifications but not tabulated. Robust standard errors are in parentheses. Estimates marked with *, **, and *** are statistically significant at the 5%, 1%, and 0.1% level respectively.

APPENDIX TABLE F: LITIGATION AS A FUNCTION OF BOILERPLATE

Probability of IPO Litigation			*Probability change per 10% change in similarity*
	(1)	(2)	(3)
Overall Similarity	0.001 (0.050)	-0.047 (0.122)	**-0.17 to 0.47%**
Overall Plain English	-0.001 (0.001)	-0.001 (0.002)	
Rf Similarity	0.188** (0.076)	0.319*** (0.081)	**1.88 to 3.19%**
Rf Plain English	-0.001 (0.001)	-0.001 (0.001)	
UP Similarity	0.022 (0.097)	-0.057 (0.143)	**-0.22 to -0.57%**
UP Plain English	-0.001 (0.001)	-0.001 (0.001)	
MD&A similarity	0.002 (0.087)	-0.056 (0.123)	**0.02 to -0.56%**
MD&A Plain English	-0.002 (0.087)	-0.002 (0.002)	
Business Similarity	-0.205* (0.086)	-0.336** (0.132)	**-2.05 to -3.36%**
Business Plain English	0.002 (0.001)	0.002 (0.002)	
Log (Gross Proceeds)	0.025*** (0.004)	0.035*** (0.006)	
Venture capital	X	X	
Industry Dummies	X	X	
IPO Year Dummies	X	X	
Company Age	X	X	
Industry*Year Dummies	X	X	
Bank Dummies		X	
Pseudo R^2	0.094	0.122	
Number of Observations	2,225	1,475	

Standard errors clustered at the industry level are reported in parentheses. Estimates marked with +, *, **, and *** are statistically significant at the 10%, 5%, 1%, and 0.1% level respectively.

Appendix Table G: Legal Fees and Transaction Timing

Total Legal Fees for all Counsel			*Change per 10% change in similarity*
	(1)	(2)	(3)
Overall Similarity	-518527.40** (196926.90)	-671559.70** (212945.80)	**$51,853 to $67,156 lower fees**
Overall Plain English	-13498.86** (5116.39)	-10515.19+ (5526.07)	
Log (Gross Proceeds)	265756.60*** (85925.85)	287298.4*** (078291.69)	
Adj. R^2	0.572	0.457	
Number of Observations	1,702	1,475	
Time to Completion			
Overall Similarity	-1.387 (22.801)	-0.649 (25.349)	
Overall Plain English	-0.375 (0.481)	-0.248 (0.521)	
Log (Gross Proceeds)	-10.723*** (2.788)	-8.619* (4.039)	
Adj. R^2	0.110	0.219	
Number of Observations	2,147	2,147	
IPO Year Dummies	X	X	
Industry*Year Dummies	X	X	
Bank Dummies	X	X	
Law Firm Dummies		X	

Robust standard errors are in parentheses. Estimates marked with *, **, and *** are statistically significant at the 5%, 1%, and 0.1% level respectively.

APPENDIX TABLE H: ROBUSTNESS CHECKS FOR FIRST DAY RETURNS ANALYSIS

First Day Returns		
	(1)	(2)
Overall Similarity	**0.411***	**0.358**
	(0.203)	**(0.247)**
Overall Plain English	**0.002**	**0.002**
	(0.005)	**(0.247)**
Log (Gross Proceeds)	0.095***	0.076***
	(0.140)	(0.020)
Adj. R^2	0.299	0.277
Risk Factor Similarity	**0.534****	**0.830*****
	(0.205)	**(0.260)**
Risk Factor Plain English	**-0.001**	**-0.004**
	(0.005)	**(0.007)**
Log (Gross Proceeds)	0.135***	0.130***
	(0.025)	(0.035)
Adj. R^2	0.299	0.290
Use of Proceeds Similarity	**1.160*****	**1.197*****
	(0.317)	**(0.433)**
Use of Proceeds Plain English	**0.008+**	**0.011+**
	(0.004)	**(0.006)**
Log (Gross Proceeds)	0.103***	0.083***
	(0.014)	(0.028)
Adj. R^2	0.300	0.281
MD&A Similarity	**0.977*****	**1.002*****
	(0.262)	**(0.311)**
MD&A Plain English	**-0.0002**	**0.005**
	(0.008)	**(0.010)**
Log (Gross Proceeds)	0.094***	0.078***
	(0.137)	(0.028)
Adj. R^2	0.310	0.289
Business Description Similarity	**0.138**	**0.730+**
	(0.297)	**(0.400)**
Business Plain English	**0.013***	**0.019***
	(0.006)	**(0.008)**
Log (Gross Proceeds)	0.092***	0.076***
	(0.137)	(0.028)
Adj. R^2	0.300	0.318
Industry Dummies	X	X
IPO Year Dummies	X	X
Industry*Year Dummies	X	X
Law Firm Dummies	X	X
Bank Dummies		X
Manager>=40	X	
Manager>=80		X
Number of Observations	1,135	666

Robust standard errors are in parentheses. Estimates marked with +, *, ** and *** are statistically significant at the 10%, 5%, 1%, 0.1% level respectively. Additional controls for issuer age, venture capital involvement, syndicate size included for all specifications but not tabulated. The table above reports the results of OLS regressions testing the possibility that selection is driving the observed increase in IPO market performance in the periods studied. The table reports tests using the main specification for the opening day price jump but limiting the sample to banks that manage at least 40 issues in the dataset (the top 18 banks), and to banks that manage at least 80 issues in the dataset (the top 7 banks).

Appendix Table I: Lawyer Deals and Boilerplate

Change in Boilerplate for each deal in relevant time period			
	(1)	(2)	(3)
Number of Deals in the past 1 year	0.0021* (0.0010)		
Number of Deals in the past 2 years		0.0008 (0.0008)	
Deals in the past 3 years			0.0004 (0.0007)
Venture capital	X	X	X
Industry Dummies	X	X	X
IPO Year Dummies	X	X	X
Industry*Year Dummies	X	X	X
R^2	0.305	0.304	0.304
Number of Observations	2,149	2,149	2,149

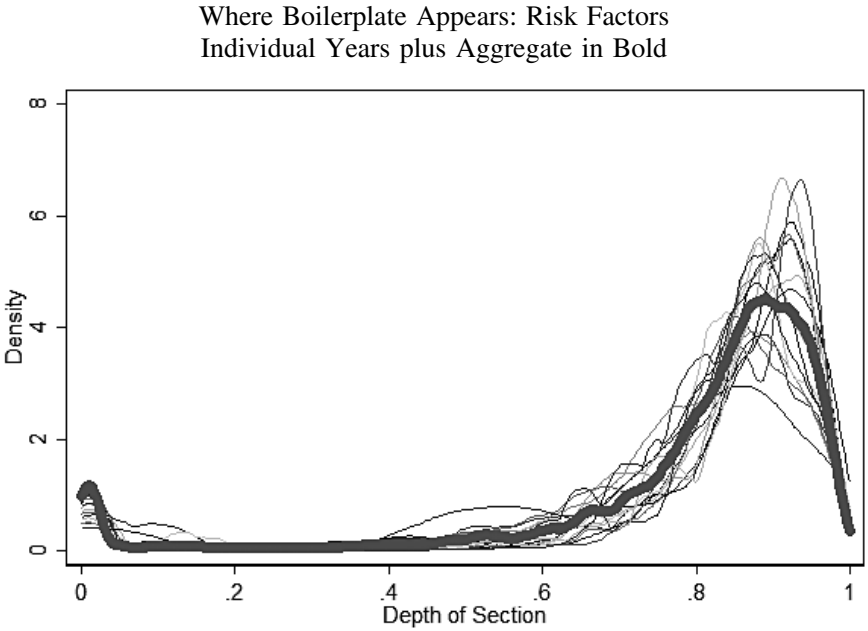
Robust standard errors are in parentheses. Estimates marked with *, **, and *** are statistically significant at the 5%, 1%, and 0.1% level respectively.

APPENDIX TABLE J: ROBUSTNESS CHECKS FOR PRICE REVISION ANALYSIS

Price revision		
	(1)	(2)
Overall Similarity	0.062	0.104
	(0.174)	(0.256)
Overall Plain English	0.002	-0.002
	(0.004)	(0.005)
Pseudo R^2	0.100	0.100
Number of Observations	1,135	666
Risk Factors Similarity	0.329*	0.120
	(0.157)	(0.203)
Risk Factor Plain English	-0.012*	-0.018*
	(0.005)	(0.007)
Pseudo R^2	0.091	0.093
Number of Observations	1,135	666
Use of Proceeds Similarity	0.355	0.250
	(0.248)	(0.326)
Use of Proceeds Plain English	0.012*	0.008
	(0.006)	(0.008)
Pseudo R^2	0.08	0.08
Number of Observations	1,135	666
MD&A similarity	0.222	0.215
	(0.180)	(0.234)
MD&A Plain English	0.005	0.004
	(0.006)	(0.008)
Pseudo R^2	0.08	0.08
Number of Observations	1,135	666
Business Similarity	0.544+	0.812+
	(0.308)	(0.426)
Business Plain English	0.007	-0.001
	(0.006)	(0.008)
Pseudo R^2	0.100	0.100
Number of Observations	1,135	666
Venture capital	X	X
Industry Dummies	X	X
IPO Year Dummies	X	X
Industry*Year Dummies	X	X
Law Firm Dummies	X	X
Manager>=40	X	
Manager>=80		X

Robust standard errors are in parentheses. Estimates marked with +, *, **, and *** are statistically significant at the 10%, 5%, 1%, and 0.1% level respectively. Additional controls for issuer age, venture capital involvement, syndicate size included for all specifications but not tabulated. The table above reports the results of probit regressions testing the possibility that selection is driving the observed increase in IPO market performance in the periods studied. The table reports tests limiting the sample to banks that manage at least 40 issues in the dataset (the top 18 banks), and to banks that manage at least 80 issues in the dataset (the top 7 banks).

APPENDIX FIGURE 1: WHERE BOILERPLATE APPEARS IN RISK FACTORS



Appendix Figure 1 depicts the location within the Risk Factors section that boilerplate sentences most frequently appear. Location is denoted as the percentage of the distance from the beginning to the end of the section, whereas the concentration of boilerplate is denoted by the density.

APPENDIX FIGURE 2: CHANGE IN PREVALENCE OF RISK FACTOR TOPIC 1 (GENERIC RISKS OF BEING A PUBLIC COMPANY) OVER TIME