Lawyers, Law Firms, and the Production of Legal Knowledge

Adam B. Badawi*

Abstract

Legal language is a prominent constraint on firm behavior. While some of this language is statutory or regulatory in origin, much of it is produced by the lawyers and law firms that draft the contracts that firms enter into and the public documents that they file. Despite the importance of this lever of corporate governance, there has been relatively little research on the role that lawyers and law firms play in the generation and evolution of this language. This paper addresses this gap in the literature by conducting a large scale word content analysis of fifteen years worth of offering prospectuses filed with the SEC by firms that seek to make an initial or secondary public offering. Extracting the names of the lawyers and law firms that produce these documents allows for an analysis of the similarity of documents produced by the same lawyers, the same law firms, the same industries, and by proximate law firms. By identifying lawyers that switch law firms, the analysis can assess the amount of similarity that is attributable to the lawyers themselves and how much is associated with the lawyers working at a specific law firms. The law firm effect is larger than the lawyer effect across multiple specifications of the model. This result suggests that if lawyers leave their law firms, the subsequent work product of those lawyers will look quite different. This effect implies that there are important organizational effects associated with the production of this sort of legal language. In addition to this result, the analysis also shows substantial similarity effects associated with issuer industry, law firm proximity, and documents drafted after a law firm has merged.

*School of Law, Washington University in St. Louis. I thank Curtis Anderson, Bernie Black, Caroline Gentile, and Kyle Rozema for helpful comments and discussion.

1 Introduction

Legal language is a significant source of control over firm behavior. While much of this language comes from statutes, regulations, and judicial opinions, a substantial amount of it is generated on behalf of the firm by the lawyers and law firms that they hire to draft public filings and contracts. Despite the importance of these documents, little is known about the role that individual lawyers and law firms play in the development of this language. This paper takes on this open question by performing content analysis on a unique set of registration statements filed with the Securities and Exchange Commission (SEC) in anticipation of an initial or secondary public offering. The dataset allows for an analysis of the lawyers and law firms that draft each of these documents. In so doing, this analysis relates to other studies that analyze the content of legal documents, such as the work of Hanley and Hoberg on the relationship between the content of IPO registration statements and IPO pricing (2010) and litigation risk (2012), the extensive work by Gulati and others on the evolution of sovereign bond contracts (Choi and Gulati, 2004; Choi et al., 2013, 2012; Gulati and Scott, 2012), and recent work by Coates (2016) and Anderson and Manns (2017) on the content of merger agreements.

Understanding the role that lawyers and law firms play in the generation and evolution of legal documents provides insight about the degree to which these documents are actually tailored to firm circumstances and, relatedly, whether firms are getting much value from legal service providers. The anecdotal take of lawyers on these documents is that they are boilerplate. But that blanket statement says little about the source of any borrowing (let alone whether they actually are boilerplate). The documents could be similar across the universe of all registration documents, across specific industries, across particular law firms, or across individual lawyers involved in the drafting.

The source of the variation matters for several reasons. If all registrations statements are highly similar, that fact would cast doubt on the effectiveness of disclosure because it would suggest that most firms are releasing the same type of information. If that were the case, market participants would learn little from these disclosures. This point also holds if similarity has a strong relationship with readily observable characteristics such as firm industry or issuer counsel. As long as market participants know that observable characteristic, there is little new information would be conveyed by the actual disclosure.

The issue is more subtle if there is only limited overlap across documents. If there are similarities across registration statements across industry that could be evidence of writing about a similar topic rather engaging in verbatim copying. To the degree there are similarities across statements prepared by the same lawyer or team of lawyers, that may be evidence of a particular linguistic style rather than repeated use of boilerplate. If documents prepared by the law firm are similar, even when prepared by different lawyers, that suggests that organizations exert influence over the language used in the registration statements.

The descriptive evidence developed in this paper provides some indications of answers to these questions. The degree of variation among the universe of all registration statements is large. Using a measure of similarity between each unique pair of documents shows that the average amount of similarity between any two documents is quite small. Within narrower categories, however, this measure is higher. There is, as one might expect, a notable industry effect. The average pairwise similarity measure for registration statements prepared for firms that are within the same Fama-French 48 industry classification is about four and a half times higher than the overall average similarity. Commonality of lawyers and law firms is also associated with substantial similarity of registration statements. When overlapping lawyers prepare a pair of documents, the average similarity score is even higher than when two documents are prepared for firms in the same industry. When the same law firm prepares two documents, the average similarity score is over two times higher than the average comparison score.

Regression analysis shows some more complex associations. The central question of interest is whether it is the lawyer or the law firm that is most strongly associated with document similarity. A potential answer to this question can be found in comparing the lawyer similarity variable with the interaction variable for the same lawyers preparing the same document at the same law firm. In every specification, the interaction effect is larger than the lawyer effect. These results suggest that if the lawyers who prepare registration documents move to another firm, those documents are likely to look substantially different than those prepared when the lawyers were at their previous firms. Or, to put this another way, law firm boundaries appear to matter when it comes to the content of attorney work product even when the same attorneys prepare that work product.

The analysis also examines the effect of law firm collapses on future document preparation. Law firms tend to collapse rather suddenly, which leaves the law firm's lawyers scrambling for a new position (Morley, 2015). This tendency allows for an assessment of whether the subsequent work product of lawyers who are forced to move differs from that of those who choose to do so. There is little evidence of such a difference, although the number of forced moves is low enough to warrant caution about reaching any conclusions.

Law firm mergers and acquisitions have the potential to affect the work product that the firm produces. To the degree that registration statements are a product of firm personnel and the internal databases that firms use to produce legal documents, both will be affected by a merger or an acquisition. Moreover, the most commonly cited reason for law firm mergers is ability to expand intrafirm referral networks (Aronson, 2007; Briscoe and Tsai, 2011). That capacity may increase the volume of registration statement work for the firm, which may affect their content through increased efficiencies. The analysis of law firm mergers shows that post-merger documents are more similar to each other than the pre-merge documents prepared by the previous firm. This result suggests that these merger-related channels may play a role in document content.

This paper proceeds as follows. Section II reviews the literature on the generation and evolution of boilerplate and on the structure of law firms before developing some expectations about associations between document features and document similarities. Section III discusses the data collection and provides some summary statistics. Section IV presents the core results and explores how law firm collapses and mergers are associated with registration statement content. Section V concludes and Appendix A provides variable definitions.

2 Legal Background, Related Literatures, and Theory Development

This section begins with a discussion of the legal and organizational context for the preparation of S-1 registration statements. It then discusses the existing finance, legal, and organizational research that relates to lawyers, law firms, and the content of the documents they produce. This literature review also develops some predictions about the factors that may contribute to similarity among registration statements.

2.1 Legal Background

Firms that wish to raise capital through initial or secondary public offerings must file appropriate disclosures with the SEC. In most cases, those firms will file an S-1 statement. The goal of the document is to provide information to prospective investors about the firm's business, the details of the offering, and the risks associated with the investment. The document has standardized sections such as a summary of the prospectus, a discussion of the risk factors, a statement of how the firm will use the proceeds, and the management's discussion and analysis.

Going public with insufficient disclosures in a registration statement has legal consequences for issuers and their directors and officers. The securities laws authorize any purchaser of the issuer's stock to sue for an untrue statement of a material fact or for an omission of a material fact in the initial registration statement. The SEC may also initiate civil or criminal actions against issuers who make false or misleading statements. This liability is strict for the firm, in the sense that good faith and due diligence are not viable defenses. Individuals, such officers and directors, do, however, have some limited recourse to due diligence defenses.

The documents themselves tend to be long and often stretch over several hundred pages. While many parties play a role in the preparation of an S-1—including the issuer, the underwriter, the underwriter's attorneys, and the issuer's auditors—there is evidence that the issuer's attorney has the most influence on these documents (Hanley and Hoberg, 2010). An S-1 typically has a fair amount of information that is tailored to each issuer, but there is also a fair amount of boilerplate in these documents. One might expect the public availability of these documents through the SEC's EDGAR system to facilitate the repurposing of boilerplate from earlier S-1s. Indeed, some practice manuals expressly recommend that parties begin drafting document by "selecting a comparable prospectus relating to a security already public" and relying on it to "take the drudgery out of the chore" (Bartlett, 1999, p. 137). There are other sources that the issuer's attorney may use to prepare the document. All, or nearly all, large law firms maintain a searchable electronic database of the documents prepared by the firm's lawyers. That source is a natural place to turn if a lawyer is working on a similar document. In the context of S-1s, that is likely to mean that lawyers will use previous S-1s prepared by that firm when drafting a new S-1.

2.2 Literature Review and Theory Development

This study of the influence that individuals and organizations have on the content of registration statements relates to three separate literatures in finance, law, and the theory of the firm. The first is the finance literature that uses similar textual analysis tools to shed light on topics such as IPO underpricing, interfirm competition, and firm financing decisions. The second is the legal literature on how contracts and other legal documents evolve over time. The third is the theory of the firm research on how and why professional service firms organize themselves in the way that they do.

The finance literature has embraced the use of text analysis quite broadly. The raw material for this analysis is the huge amount of text that firms publicly disclose through the SEC's EDGAR system. Two studies by Hanley and Hoberg use the same S-1 materials and cosine similarity measure that this paper uses. The first of those papers (Hanley and Hoberg, 2010) provides evidence of a relationship between the amount of individual tailoring in a registration statement and the amount of IPO underpricing. The authors find that the more informative a prospectus is–in the sense that it deviates from other prospectuses–the more accurate is the pricing. They conclude that investing in making a prospectus more informative is a substitute for performing more bookbuilding. A second paper (Hanley and Hoberg, 2012) examines the extent to which issuers update their registration statements in response to new information acquired during the offering period. They find a tradeoff between the use of underpricing and strategic disclosure as hedges against litigation risk.

Other studies use similar analytical techniques on different bodies of disclosures. Hoberg and Phillips (2010) use product descriptions in 10-Ks and find that mergers and acquisitions are more common among firms that use similar product language. In a separate paper, Hoberg and Phillips (2016) also use text analysis of product descriptions to reclassify industries based on competitors with similar products. A similar approach allows Hoberg et al. (2014)

to assess how competitive pressure–as measured by the textual similarity of product descriptions–affects firm payout policies.

There is a substantial empirical legal literature on the evolution and stickiness of boilerplate documents in the legal literature. Much of this analysis focuses on pari passu provisions in sovereign debt contracts. Gulati and Scott (2012)document that this provision–which determines whether the holders of sovereign debt can holdout during restructurings–has been stiffly resistant to change despite court decisions that have called previous understandings into question. The researchers who have analyzed this phenomenon have largely attributed it to the agency costs that are inherent in legal practice (Choi et al., 2016). These agency costs arise because lawyers do not fully internalize the costs associated with poor drafting and the ease of copying boilerplate means that they do not capture all of the gains associated with innovation. These dynamics can lead to lawyers capturing rents for the straightforward task of copying previously produced text.

Not everyone takes such a dim view of the lawyers task. While lawyers do sometimes draw on previously used boilerplate, it requires professional judgment to select the appropriate text and to tailor it to the new circumstances. If lawyers or law firms are drafting text for repeat clients, that dynamic should curtail some of the agency costs because any future liability that is attributable to the lawyer or law firm may result in the loss of future business (Kahan and Klausner, 1997). But what is not well known, and what this paper helps to answer, is the degree any use of previous materials is driven by individual lawyer effects or by law firm effects.

The final related literature is a strain of research on the theory of the firm that focuses on professional service firms. As usual, the question of interest is why activity takes place inside the firm instead through contracting outside of it. In the context of law firms, part of the reason for doing work entirely within lawyer-owned firms are undoubtedly related to legal and ethical restrictions on certain types of activity. For example, there are restrictions on non-lawyers having ownership interests in law firms (Adams and Matheson, 1998). These rules affect the way lawyers organize themselves and, presumably, how they conduct their work. Similarly, there are rules on potential and actual conflicts of interest between clients (Ribstein, 1998). These restrictions are likely to affect the size of law firms because, as law firms grow, conflicts become more and more inevitable. Such a limitation prevents some amount of information aggregation and sharing.

But beyond these legal restrictions, there must also be market forces that drive the organization of lawyers and other similar professionals. The early literature on the theory of the firm struggled to understand what these reasons might be. For example, (Alchian and Woodward, 1987, p. 126) suggest that in "a professional service firm, like law, architecture, medicine, engineering, or economic consulting, members may be so specific to certain customers that if they left the firm, remaining members and shareholders would hold an empty shell." This description poses the question of why these professionals choose to organize together at all if their value is unique to the individual relationships they have with their clients. There might be administrative economies of scale in working together and pooling income may diversify the risk associated with cyclical practice areas (Gilson and Mnookin, 1985), but this approach says little about how the nature of legal labor encourages organizing in firms.

Subsequent work has suggested what some of the benefits of organizing in a single entity might be. One theory is that it can be difficult to judge the quality of a professional service provider on the basis of a single interaction (Von Nordenflycht, 2010). Did, for example, a lawyer's involvement in a negotiation produce a better outcome for the client? Reputation may serve as a proxy for overall outcomes in repeated interactions. By aggregating together, lawyers (and other similar types of professionals) may be able to establish a reputation for effective outcomes that would be more difficult if they all worked individually. Another theory, which relates more directly to this paper, is that professionals such as lawyers can aggregate knowledge in a way that would be difficult individually (Grant, 1997). By using centralized databases, a law firm can generate a store of collective and historical knowledge that stays with the firm rather than with departing lawyers. Members of the firm can then draw on this collective knowledge for future work. To the degree that law firms do so, that should be borne out in a similarity across documents produced by the law firm even when different lawyers are working on those documents. If this is the case, document differences between firms may see some convergence if the two firms merge together.

An alternative explanation for the similarity of documents is the that they are authored by similar people or groups of people. There is a substantial literature that attempts authorship through the use of text analysis (Juola et al., 2008; Stamatatos, 2009). An underlying assumption in this approach is that individuals write in a similar way across documents. The legal research in this area largely focuses on judges and asks, for example, whether judges tend to write their own opinions or, instead, rely on a rotating group of law clerks (Rosenthal and Yoon, 2010, 2011). To the degree the assumption is valid, it allows us to examine the relative influence of law firms and individual lawyers. When a lawyer moves firms–either by choice or through necessity–we can examine how similar the S-1s prepared by that lawyer to previous S-1s prepared by that lawyer. A strong relationship of this sort would suggest that the ability of law firms to aggregate knowledge is not that important, at least in the context of S-1s.

There are other likely sources of similarity in these types of documents. One is existing documents from firms that are in related industries. It is certainly imaginable that lawyers and other involved parties would use these documents as templates to begin drafting an S-1 and previous work does show a relatively strong industry effect when comparing S-1 similarity (Hanley and Hoberg, 2010). That work has also shown a temporal effect, in the sense that documents filed within months of each other are more similar than those filed years apart. This effect could be due to parties using recent statements as templates and it could also reflect that changing regulatory landscapes produce different language over time.

A final potential source of similarity is geographic proximity to other law firms doing similar work. There are two likely channels for this influence. First, a lawyer is more likely to know other lawyers who are nearby and, if the lawyer thinks well of some of these lawyers, may seek out S-1 prepared by those lawyers to use as a template. Second, law firms will sometimes have internal opinion manuals that specify acceptable language that the firm may use in documents such as registration statements. Those law firms will sometimes be authorized to use the language of other firms, provided that those firms are known to high-level work. Geographic proximity may allow firms to learn enough about other law firms that they are comfortable using their language in an S-1.

3 Data and Summary Statistics

The data collection begins with the use of a PERL script to scrape every S-1 registration statement filed with the Securities and Exchange Commission's

EDGAR system between January 1, 2001 and December 31, 2015.[1] To reduce similarity that is a product of firms making multiple offerings, the dataset limits each individual firm to its earliest offering in the sample (i.e. there are no repeated firms in the sample).[2]. From each S-1, I extract the name of the company, EDGAR's Central Index Key (CIK), the initial filing date of the S-1, and the four-digit SIC code, if provided. To increase sample size, the dataset retains all S-1s filed, even if the issuer eventually withdrew the offering.

The processing of these S-1s requires several steps. Many of them are in HTML or XML format those tags must be stripped from the document. The lawyers and law firms are not listed in defined fields and so I extract the section of the S-1 where that information typically appears and then parse that language to look for terms that identify lawyers (such as "Esq." or "Esquire") and characters and terms that identify law firms (such as ampersands, "LLP", and "L.L.P.").[3] The dataset also provides the zip code of the issuer's law firm, which permits the identification of separate offices of law firms. Not all S-1 filers use outside counsel and some of the statements are not in a format that makes them scrapable. These documents are excluded from the sample. The final sample contains 2,721 registration statements.

The next step is to prepare the text of the S-1s for content analysis. The format of S-1s is standardized and I focus only on the content that appears between the table of contents and the section that details where investors can find more information. This approach eliminates the sometimes voluminous appendices that usually contain a substantial amount of information produced by the issuer's auditors. The process next converts all text to lower case and removes all numbers, punctuation, and symbols. As is standard in content analysis, words are stemmed to their roots and commons stop words in the English language are eliminated.[4]

I put the text of each S-1 into a vector of terms, called $terms_i$. I calculate a document frequency matrix, which tokenizes all unique terms across the

[1] The format of the electronic S-1s prior to 2001 rendered it difficult to reliably identify the law firm affiliations of lawyers.

[2] EDGAR's Central Index Key (CIK) is used to identify the each firm for this screening process

[3] The dataset only includes information for the first listed law firm (and its associated lawyers). In every handchecked S-1 that listed multiple law firms, the second law firm represented the underwriter.

[4] The processing uses the standard list of the top fifty stop words provided by R's quanteda package.

entire sample of S-1s and contains a vector for each individual S-1. That vector tallies the number of times each unique term appears in a given S-1. I then normalize the word vectors to give each term a weight, such that the sum of the weights for each document is equal to one. This process is the preferred method for comparing documents of differing lengths. I then calculate the cosine similarity between all unique pairs of documents. The cosine similarity is the dot product of the normalized word vectors of the two documents. Thus, for $terms_i$ and $terms_j$ the cosine similarity is $\frac{terms_i \cdot terms_j}{||terms_i|| \; ||terms_j||}$. The cosine similarity is bounded in $[0, 1]$.

To make the concept more concrete, imagine the two documents. The first, i states, "the company manufactures automobile parts" and the second, j states, "the company develops software for the manufacture of automobiles and the manufacture of automobile parts" After eliminating the common words ("the," "for," "of," and "and") and stemming the words, the total word vector has six words: company, manufacture, automobile, part, develop, and software. For each document the vectors are as follows: $terms_i = (1, 1, 1, 1, 0, 0)$ and $terms_j = (1, 2, 2, 1, 1, 1)$. The normalized vectors are then: $terms_i = (.25, .25, .25.25, 0, 0)$ and $terms_j = (.125, .25, .25, .125, .125, .125)$. Each normalized vector sums to one, which facilitates comparison across documents of different lengths. The cosine similarity of the two documents is about .094.

There are 2,721 registrations statements in the complete sample, which means that there are 3,700,560 unique pairs of documents.[5] Many of the registration statements do not include an SIC code and to remedy this issue, I match the data to Compustat using the EDGAR CIK. Doing so substantially increases the number of observations with information about industry.[6] This subsample contains 2,099 registration statements and that results in 2,201,851 unique pairs of documents. Many of the S-1s also provide a self-reported SIC code. This subsample includes 1,607 documents, which allows for 1,290,421 unique pairs of documents.

To provide some sense of how representative the sample is, Figure 1 depicts the annual distribution for the complete sample and completed IPOs during the time period. Although data for intended IPOs and secondary offerings are

[5] The formula for the number of unique pairs is $\frac{x^2 - x}{2} = uniquepairs$, where x is the number of documents.

[6] In a number of these cases, the information in Compustat is not contemporaneous with the S-1. This is likely because the filer initially withdrew the offering, but wound up going public later. For this reason, I refrain from obtaining other covariates from Compustat.

not readily available, the trend in Figure 1 mirrors the trend for completed IPOs over this period, with the exception of 2008.[7] The difference is almost certainly attributable to the high number of withdrawn IPOs in 2008 (the S-1s in the sample include all intended initial and secondary offerings).

[Insert Figure 1 about here]

The filing firms are quite geographically dispersed, as Figure 2 shows. And, as Table 1 demonstrates, the law firms that prepare the registration statements are not. New York City law firms dominate the landscape and account for over one-third of the market. Other important regions for this work include Boston, Washington, D.C., the San Francisco Bay Area, and Southern California.

[Insert Figure 2 about here]

[Insert Table 1 about here]

The market is also quite concentrated when it comes to the law firms that prepare registration statements. Table 2 lists the top 25 law firms that appear in the sample and the count of the number of times they appear. The list is quite top heavy with the number one firm, Latham & Watkins, preparing almost three times as many S-1s as the tenth ranked firm. In the list there is a mix of highly regarded law firms and lesser known firms. A review of the data suggests that the lesser known firms are doing commodity-like work for transactions, such as private investments in public equity (PIPEs), that require an S-1 filing.

[Insert Table 2 about here]

The S-1s provide the names of the lawyers involved in the preparation of the S-1.[8] In the complete sample, a total of 4,648 lawyers are listed, meaning that an S-1 lists an average of about 1.72 lawyers. Of the 4,648 lawyers listed in the sample, there are 2,068 unique lawyers. The lawyer who appears most is listed in 66 S-1s, the lawyer who appears in the tenth most S-1s appears 18 times, and the lawyer who appears in the hundredth most S-1s appears 7 times.

In an earlier age of law practice, a partner moving to another law firm was a rare event. Partners were loyal to law firms and the law firms, in turn, were loyal to them (Galanter and Palay, 1994). The past several decades have produced a much more competitive landscape for legal practice. Unproduc-

[7]The completed IPO data are from Ritter (2016).

[8]The S-1 states that copies of any correspondence should be sent to a listed group of lawyers. There is no standard for who gets listed. Sometimes it is one lawyer, presumably the lead partner, and in other instances there are up to seven lawyers listed.

tive partners are more likely to get pushed out the door and partners with a significant book of business can and do receive offers of higher compensation from other law firms. The dataset developed here reflects that change. A little over five percent of the lawyers, 105 of them, appear on an S-1 for more than one law firm. Of those 105, 96 appear with two different law firms, 8 appear with three different law firms, and one of them appears with four different law firms.

The pairwise analysis that follows uses two different measures of lawyer similarity when comparing two S-1s. The first is whether the first-listed lawyer in the first S-1 is the same as the first-listed lawyer in the second S-1. The second measure takes into account every lawyer listed on both S-1s and calculates the average proportional overlap in the lawyers listed on the first S-1 and the second S-1. For example, if the first S-1 lists two lawyers, each of those lawyers receives .5 credit for that S-1. If the second S-1 lists four lawyers, each one receives .25 credit for that S-1. If one of the lawyers from the first S-1 overlaps with one of the lawyers from the second S-1, the mean proportional overlap is $((.5+.25)/2)=.375$. For the full pairwise sample, the mean overlap is just .0016, which reflects the relative rarity of overlap. If the complete sample is limited to pairwise comparisons with some amount of positive overlap, the mean overlap is .692.

Table 3 provides the means for some of the binary variables in the for the pairwise dataset. In the full dataset, the first-listed lawyer is the same in the first and second S-1 for about .2 % of the observations, the law firm is the same in about 1.6% of the observations, and the preparers are from different offices of the same law firm in about .9% of the full sample. It is relatively rare for the first-listed lawyer to be the same, but for the law firms to be different. This occurs in .02% of the complete sample. The New York City dominance helps to account for the relatively high percentage of comparisons (1.5%) where the law firms are different, but are located in the same zip code. There is a high degree of industry overlap, with nearly 16% of the pairwise comparisons having the same Fama-French 12 industry code, while that number dips to about 8% when the categories are sliced into the Fama-French 48 industry codes.

[Insert Table 3 about here]

The metric for comparison between documents is the cosine similarity. As discussed above, this measure varies from zero to one with a higher number indicating more similarity between two documents. Table 4 displays descriptive statistics for the similarity scores of all unique pairs of documents and for documents that share common features, such as being prepared by the same lawyer or the same law firm. The overall amount of similarity across all pairwise comparisons is relatively low, coming in at .016 for all of the S-1s, and .018 for the S-1s that have been matched to Compustat and for the S-1s that provide an SIC code in the filing.

Table 4 suggests a relatively strong industry effect, with the similarity being stronger as the industry definition becomes narrower. The largest mean similarity is for S-1s prepared for firms that are in the same Fama-French 48 industry, with that mean (.081) being about four and a half times higher than the overall mean for Compustat-matched documents. The temporal aspects appears to be modest as S-1s filed within 90 days have an average textual similarity of .020, which is not much higher than the overall mean similarity of .016 across all observations in the complete dataset.

The mean document similarity suggests that there may be an effect associated with two documents being prepared by the same law firm. The overall mean similarity for documents prepared by the same law firm is .041 and, when limited to the same office of the same law firm, the mean is .050. For documents prepared by different offices of the same law firm, the mean falls to .034. Note, however, that some of this difference is likely attributable to lawyer and perhaps industry effects because lawyers rarely move from one office to another.

When there is any positive overlap in the lawyers preparing the the same document, the mean similarity is .087. That number is larger than any of the industry effects and it is over five times larger than the mean similarity across all document comparisons. That number edges up to .090 when there is positive overlap and the law firms preparing the two S-1s are the same. When there is lawyer overlap and the two S-1s have been drafted by different law firms, the mean similarity halves to .045. This fact provides an indication that some of amount of document similarity depends is a product of an organizational effect. This difference is even more pronounced when focusing on the first listed lawyer in the S-1. When the first lawyer is the same for two given S-1s, the mean similarity is .091. When that first listed lawyer has moved law firms, the mean comparison between the two documents more than halves to .043.

[Insert Table 4 about here]

3.1 Lawyers, Law Firms, and Prospectus Content

To assess the association between lawyers, law firms, other related variables on the content of registration statements, this subsection reports the results of a series of regressions where each observation is a pairwise comparison of a unique pair of documents in the relevant sample. The dependent variable in each regression is the textual similarity score between the unique pair of registration statements. The independent variables, most of which are binary, measure other similarities and differences between the pair of registration statements. The measures include whether the same law firm prepared the two documents, the commonality of the lawyers preparing the document, whether the documents were prepared by different law firm offices that are in the same zip code, and whether the filing firms are in the same industry. The regressions are ordinary least squares models that include S-1 fixed effects and the heteroscedasticity-robust standard errors are clustered by law firm.

Table 5 uses the mean proportional lawyer overlap measure discussed above to measure lawyer commonality between pairs of documents. This measure should capture authorship more accurately in comparison to a variable that indicates whether the first-listed lawyer in the two documents is the same. The first regression includes all of the S-1s in the sample. This sample does not include information on industry, which previous work shows has a substantial effect on prospectus similarity (Hanley and Hoberg, 2010). Nevertheless, this regression shows many of the associations that one would expect. There is a law firm effect even when including the individual, zip code, and office controls. This effect is, however, much smaller than the lawyer overlap measure (.013 vs. .032) and over three times smaller than the Lawyer Overlap X Same Law Firm interaction coefficient (.042). There is a geographic relationship that is not small relative to the mean of the textual similarity measure for that sample: the indicator variable for the two law firm offices having the same zip has a coefficient of .010 against a dependent variable mean of .016. The temporal relationship is quite small (.003), although it is statistically significant at the one-percent level. The last four specifications include a variable to indicate some degree of industry similarity. Doing so produces a sizable jump in the R^2 for these models relative to the first specification. In these four regressions the coefficients on the other independent variables are generally smaller than they are in the first regression. The statistical significance for each variable is generally similar across all five specifications.[9]

[Insert Table 5 about here]

The lawyer overlap variables show some interesting associations. In all specifications, the interaction of the lawyer overlap measure and the same law firm variable is larger than the coefficient on the lawyer overlap variable. Both the overlap and interaction coefficients are quite large, ranging from .015 to .042. The lawyer overlap result is not all that surprising. One would expect that documents prepared by the same lawyers would exhibit substantial similarity. But that effect is not as large as when the same lawyers prepare documents while working at the same law firm. To put this another way, if a lawyer goes to another law firm, a subsequent registration statement will look quite different compared to those that the lawyer prepared at the previous law firm.

There are several potential channels that could account for this association. One is that lawyers prefer to use language from S-1s they have previously prepared and that being at the same law firm provides them unique access to those documents. There may also be internal firm protocols that contribute to this effect. For example, some law firms have registration statement language that has been approved by auditors and malpractice insurers. If a different firm has a different set of approved language, that would help to account for the differences observed when a lawyer prepares a registration statement at a different law firm. There may also be personnel related effects. There are no precise requirements about which attorneys must be listed in an S-1. As the descriptive statistics show there is substantial variation in the number of listed attorneys. Cross referencing the listed attorneys to their LinkedIn profiles suggests that it is usually the law firm partners who get listed and not the law firm associates. When a partner changes firms, the associates may stay at the former law firm and it may be their missing contributions that account for the interaction effect.

The geographic and law firm office results are also noteworthy. The zip code result suggests that firms that are geographically proximate may borrow from one another. It could be, for example, that a lawyer who works at one

[9] The addition of a variable that measures the absolute value of the number of years between the two S-1s being filed does not change the results. In these unreported regressions the coefficients on the variables of interest are almost exactly equal to those in Table 5. The coefficient on the absolute year difference variable is exceptionally small and is only significant in the fifth specification.

firm knows a lawyer that works across the street at a different firm. If the lawyer thinks well of the other lawyer's work that lawyer may look up the language used by the other lawyer and use it as a template. But it appears that, once one accounts for lawyer overlap and geographic effects, there is nothing special about an individual office of a law firm. The coefficients for the same law firm, same office interaction are essentially zero and none of them are statistically significant.[10] This evidence is consistent with law firms being effective at centralizing some aspect of prospectus preparation. The internal databases and language guides may be being used in roughly equal measures by different offices of the law firm.

Table 6 repeats the regressions in Table 5 but substitutes the first-listed lawyer for the lawyer overlap measure. This measure is less precise than the lawyer overlap variable and its use helps determine the relative importance of the first listed lawyer and its use also serves as a validation exercise. As the regressions show, the coefficient for the indicator that the first-listed lawyer in the two documents is the same is lower than for the lawyer overlap variable in the previous table. This result is what one would expect if, for example, both documents list two lawyers, but only the first-listed lawyer is shared by the registration statements. The other results in Table 6 are largely consistent those in Table 5 with the exception of the lawyer similarity interaction term. In Table 6, the same first lawyer and same law firm interaction coefficient is only statistically significant in three of the five regressions and one of those is only at the ten-percent level. Moreover, the magnitude of the coefficient is lower relative to the regressions that uses the lawyer overlap measure.

[Insert Table 6 about here]

3.2 Law Firm Collapses

Law firms do not last forever and when the end comes it can happen very quickly. Accounts vary for why law firms collapse rather than enter bankruptcy. Morley (2015) argues that when the law firm is no longer economically viable all of its most valuable assets–its partners with big books of business–have already left the organization. Other accounts chalk law firm collapse up to empire building gone wrong–such as in the collapse of Brobeck, Pheleger Har-

[10] The construction of this variable assumes that law firms have only one office per zip code (i.e., it is equal to one when the two documents were prepared by the same law firm and the listed zip codes for the law firm are the same for both documents).

rison in the wake of the dot-com bust–or to taking on excessive debt–as in the relatively recent collapse of Dewey LeBoeuf. Whatever the reason for the quick end of these firms provides a potential way to assess the robustness of the results in the previous subsection.

It is possible that lawyers who voluntarily leave law firms differ in some important ways from the lawyers who do not leave law firms. It could be that rainmaking partners–who have large books of business, but leave the hands-on work to others–are more likely to move to another firm. These lawyers may provide less input to S-1s and so one might observe bigger difference in comparing S-1s relative to lawyer who do not switch law firms. When a law firm collapses–or is on its way to collapsing–the lawyers have little choice but to leave. This subsection attempts to assess whether the registration statements prepared by lawyers who left a law firm due to its collapse differ from previous statements prepared by that lawyer in a way that varies from other lawyers who move to another law firm.

Morley (2016) compiles a list of law firm collapses between 1988 and 2014 that received significant media attention. Of that list, five law firms produce registration statements in the complete sample. They are: Brobeck, Pheleger & Harrison (collapsed in 2003), Heller Ehrman (2008), Thelen, Reid, Brown, Raysman & Steiner (2008), Adorno & Yoss (2011), and Bingham McCutcheon (2014). The number of pairwise comparisons in the full sample where there is lawyer overlap greater than zero, the law firms are different, one of the documents was produced by a law firm that eventually collapsed, and the other document was filed during or after the year of collapse is quite small at only 41. The mean textual similarity measure for that group is .060, which is larger than for the 624 comparisons where there is some amount of lawyer overlap and the law firms are different, but the reason for the law firm switch was not a law firm collapse (.044). This difference would be consistent with the expectation that less involved lawyers are more likely to switch firms in the absence of a collapse, but in a two-sample t-test, this difference is not statistically significant (p-value=.196). An unreported triple difference regression has positive coefficients for the triple interaction variable, but none of those coefficients is statistically significant.[11]

[11]The triple interaction is the product of a variable for whether the firm is different, a variable that indicates whether any of the common lawyers in the comparison ever worked at a law firm that collapsed, and the lawyer overlap variable.

3.3 Law Firm Mergers & Acquisitions

Law firm mergers and acquisitions are not all that frequent, but they do occur. When these transactions happen, the primary assets of the two firms–its lawyers–are combined into a single entity. Law firms can do more or less to integrate the previously separate groups of personnel. If the law firms had offices in completely different cities, there may be little attempt to combine lawyers into new working groups. But if there are overlapping geographic and topical similarities, the merged law firm may try to put lawyers from the two former firms into new team combinations. A merged law firm may also integrate the document databases and practice protocols used by the formerly separate firms. These types of changes could produce differences in the subsequent work product formed by the new combination of lawyers.

The leading explanations for why law firms merge are that doing so facilitates intrafirm client sharing, diversifies income streams, and increases administrative economies of scale while doing so (Aronson, 2007; Briscoe and Tsai, 2011). If these explanations are correct, one should expect to see law firms merge with firms that are in different geographic areas and specialize in different areas of legal practice. Alternatively, law firms might combine because they seek to increase market share, which might help increase pricing power. In this case, one should observe merger and acquisition activity among firms that operate in the same geographic and topical areas.

The number of merged law firms in the sample that have prepared registration statements before and after the transaction is relatively modest. There are three large mergers that have before and after observations in the sample: that of Wilmer, Cutler & Pickering (about 560 lawyers) and Hale and Dorr (about 500 lawyers) in 2004 (now known as WilmerHale), the merger of Pillsbury Winthrop (about 600 lawyers) with Shaw Pittman (about 300 lawyers), the merger of Kirkpatrick & Lockhart Nicholson Graham (about 1000 lawyers) and Preston Gates & Ellis (about 400 lawyers) in 2007 (now known as K & L Gates), and the merger of Hogan & Hartson with Lovells to form Hogan Lovells (about 2500 lawyers) in 2011. Less sizable transactions that have before and after observations in the sample include Bryan Cave's (about 900 lawyers) 2011 acquisition of Holme Roberts & Owen (about 160 lawyers) and the 2012 merger of Faegre & Benson and Baker & Daniels to create the roughly

770 lawyer law firm of Faegre Baker Daniels.[12]

The descriptive evidence strongly suggests that, at least with respect to IPO and secondary offering work, the firms merged to diversify their practice areas rather than to enlarge them. Of the mergers identified above only one involves a case where both firms were doing prospectus work. Prior to the WilmerHale merger, Wilmer, Cutler & Pickering did not work on any registrations statements filed in the sample. Hale and Dorr, alternatively, had worked on ten S-1s. Kirkpatrick & Lockhart had a similarly sizable S-1 practice, participating in eleven filings, while Preston Gates & Ellis performed no pre-merger registration statement work. The only exception to this pattern is the Bryan Cave and Holme Roberts & Owen transaction where both firms had performed pre-merger S-1 work (ten and one, respectively).

The rarity of firms combining active S-1 practices might suggest that the content of subsequent registration statements is unlikely to change for reasons related to the merger. Why would it if, post-merger, the firm does not experience an increase in experienced personnel or additions to its database of language that it can use? But there may be some indirect effects on prospectus content. For example, a larger network of clients may lead to work that is in a different industry or in a different geographic area. The increased referrals may also increase overall firm volume, which might create efficiencies in the preparation these documents. These differences could come out in the language used in subsequent prospectuses.

To determine the associations that merging has on S-1 content, Table 7 runs the baseline specifications from Table 5 with the inclusion of independent variables associated with law firm mergers. The first variable is an indicator that is equal to one when the comparison is between documents prepared by a firm that eventually merges or acquires or has merged or acquired. The second variable is equal to one when the comparison is between a document prepared by firm that eventually merges to a document prepared by that same firm after the merger or acquisition. The last indicator is equal to one when the observation compares two documents prepared by the same firm after the merger or acquisition. The omitted category is the comparison of two documents that were both prepared by the same firm prior to the merger or acquisition.

[12] Altman Weil, Inc., a legal consulting firm, is the source for information on these mergers. http://altmanweil.com.

[Insert Table 7 about here]

The variables of interest in Table 7 are the comparison between the Pre and Post-Merger documents prepared by the same firm and the comparison of Post-Merger documents prepared by the same firm. The coefficients for the pre and post-comparison are near zero and none are statistically significant. This result suggests that there is little difference between a comparison of a pre-merger S-1 and a post-merger S-1 and a comparison of two pre-merger S-1s (the reference group). The result is quite different when the comparing two documents prepared by a firm after the merger to two documents prepared by the firm before the merger (the reference category). The coefficients for the post-merger comparison are positive, relatively large, and are statistically significant in all five specifications. This results provides evidence that the documents prepared after the merger are highly similar to each other, but are not similar to those prepared by the same firm prior to the merger. Mergers thus appear to be associated with some effect on the content of the documents.

It is difficult to tell what the reason is for the association with the data at hand. There are some clues, however, in the descriptive statistics that allow for some speculation. Prior to its merger with Seattle-based Preston, Gates & Ellis, the Pittsburgh based firm of Kirkpatrick & Lockhart prepared eleven S-1s and Preston prepared none. After the merger, K & L Gates assisted with 29 registration statements between 2008 and 2014. Most of the clients were in the technology business and were located on the West Coast or were international firms from the Pacific Rim. These facts suggest that the merged firm was able to develop Preston's clients to do work that Kirkpatrick & Lockhart had some expertise performing. The similarity in the post-merger documents could be driven by similarities in Preston's client base, the increased routinization that comes with higher volume, or some combination of these and other factors. But, whatever the reason, the regression results to suggest differences in the operation of pre and post-merger environments.

4 Concluding Remarks

Understanding how bringing together individuals with the boundary of a firm changes their behavior is a challenging task. This paper attempts to gain insight to this question by looking at how the movement of lawyers from one law firm to another affects the future work product of those lawyers. The evidence developed here suggests that there is sizable effect associated when lawyers switch firms, at least in the context of preparing registration statements for initial and secondary public offerings. Over half of the combined individual lawyer and lawyer and law firm interaction effect on document similarity is associated with the lawyer and law firm interaction. That implies that taking lawyers out of one firm and placing them into another one has a substantial effect on the way those lawyers prepare their documents. While the precise channel or channels for this effect is unknown, this evidence does suggest that organizations matter in ways that go beyond the identity of the individuals who work in those organizations.

The evidence developed here also suggests that there is a stable law firm effect on the content of documents. This effect is present even when controlling for individual lawyers, industry, and other likely influences on document similarity. This effect appears to be unaffected by the physical office of the law firm, which suggests that this organizational effect has a long reach. This paper also shows that law firm mergers have an effect on document production. Comparing post-merger work product to pre-merger work product shows that the post-merger documents are substantially more similar to each other than the pre-merger documents are to each other. While the precise channel is again unknown, large changes to personnel and available client networks appear to change how attorneys work.

References

Adams, E. S. and Matheson, J. H. (1998). Law firms on the big board?: A proposal for nonlawyer investment in law firms. *California Law Review*, pages 1–40.

Alchian, A. A. and Woodward, S. (1987). Reflections on the theory of the firm. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, 143(1):110–136.

Anderson, R. and Manns, J. (2017). The inefficient evolution of merger agreements. *Geo. Wash. L. Rev.*, 85:57.

Aronson, B. E. (2007). Elite law firm mergers and reputational competition: Is bigger really better-an international comparison. *Vand. J. Transnat'l L.*, 40:763.

Bartlett, J. W. (1999). *Fundamentals of venture capital*. Madison Books.

Briscoe, F. and Tsai, W. (2011). Overcoming relational inertia: How organizational members respond to acquisition events in a law firm. *Administrative Science Quarterly*, 56(3):408–440.

Choi, S. J. and Gulati, G. M. (2004). Innovation in boilerplate contracts: an empirical examination of sovereign bonds. *Emory LJ*, 53:929.

Choi, S. J., Gulati, G. M., and Scott, R. E. (2016). Contractual arbitrage. $https://papers.ssrn.com/sol3/papers.cfm?abstract_id = 2795264$.

Choi, S. J., Gulati, M., and Posner, E. A. (2012). The evolution of contractual terms in sovereign bonds. *Journal of Legal Analysis*, 4:131–179.

Choi, S. J., Gulati, M., and Posner, E. A. (2013). The dynamics of contract evolution. *NYUL Rev.*, 88:1.

Coates, J. C. (2016). Why have m & a contracts grown? evidence from twenty years of deals. $https://papers.ssrn.com/sol3/papers.cfm?abstract_id = 2862019$.

Galanter, M. and Palay, T. (1994). *Tournament of lawyers: The transformation of the big law firm*. University of Chicago Press.

Gilson, R. J. and Mnookin, R. H. (1985). Sharing among the human capitalists: An economic inquiry into the corporate law firm and how partners split profits. *Stanford law review*, pages 313–392.

Grant, R. M. (1997). The knowledge-based view of the firm: implications for management practice. *Long range planning*, 30(3):450–454.

Gulati, M. and Scott, R. E. (2012). *The three and a half minute transaction: Boilerplate and the limits of contract design*. University of Chicago Press.

Hanley, K. W. and Hoberg, G. (2010). The information content of ipo prospectuses. *Review of Financial Studies*, 23(7):2821–2864.

Hanley, K. W. and Hoberg, G. (2012). Litigation risk, strategic disclosure and the underpricing of initial public offerings. *Journal of Financial Economics*, 103(2):235–254.

Hoberg, G. and Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies*, 23(10):3773–3811.

Hoberg, G. and Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465.

Hoberg, G., Phillips, G., and Prabhala, N. (2014). Product market threats, payouts, and financial flexibility. *The Journal of Finance*, 69(1):293–324.

Juola, P. et al. (2008). Authorship attribution. *Foundations and Trends® in Information Retrieval*, 1(3):233–334.

Kahan, M. and Klausner, M. (1997). Standardization and innovation in corporate contracting (or" the economics of boilerplate"). *Virginia Law Review*, pages 713–770.

Morley, J. (2015). Why law firms collapse. $https://papers.ssrn.com/sol3/papers.cfm?abstract_id = 2580616$.

Ribstein, L. E. (1998). Ethical rules, agency costs, and law firm structure. *Virginia Law Review*, pages 1707–1759.

Ritter, J. R. (2016). Initial public offerings: Updated statistics.

Rosenthal, J. S. and Yoon, A. H. (2010). Judicial ghostwriting: authorship on the supreme court. *Cornell L. Rev.*, 96:1307.

Rosenthal, J. S. and Yoon, A. H. (2011). Detecting multiple authorship of united states supreme court legal decisions using function words. *The Annals of Applied Statistics*, pages 283–308.

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*, 60(3):538–556.

Von Nordenflycht, A. (2010). What is a professional service firm? toward a theory and taxonomy of knowledge-intensive firms. *Academy of management Review*, 35(1):155–174.

Appendix A: Variable Definitions

Variable	Definition
Textual Similarity Measure)	The cosine similarity of the normalized word vectors of two registration statements.
Same Law Firm)	The same law firm prepared the two documents that are being compared.
Lawyer Overlap	The average proportional overlap of the lawyers listed on each of the two registration statements being compared.
Same Zip Code	The zip code of the two law firms that have prepared the two registration statements is the same.
Same Law Firm, Same Office	The two registration statements being compared were prepared by the same office of the same law firm (i.e. the listed zip code of the office is the same).
Filed within 90 Days of Each Other	The two registration statements being compared were filed with 90 days of each other.
Same Fama-French 48 Industries	The Fama-French 48 industry code of the two firms filing the compared registration statements is the same.
Same Fama-French 30 Industries	The Fama-French 30 industry code of the two firms filing the compared registration statements is the same.
Same Fama-French 12 Industries	The Fama-French 12 industry code of the two firms filing the compared registration statements is the same.
Same SIC Two-Digit Code	The SIC code listed in the S-1s of the two firms filing the documents is the same.
Documents Prepared by Law Firm that Merges or Acquires	The two S-1s were prepared by the same law firm and that law firm had a merger or acquisition at some point in the sample.
Comparison of Pre and Post-Merger S-1 from Same Law Firm	The two S-1s were prepared by the same law firm and one of them was prepared prior to the law firm's merger or acquisition and the other was prepared after the law firm's merger or acquisition.
Comparison of Post-Merger S-1s from Same Law Firm	The two S-1s were prepared by the same law firm and both were prepared after the law firm's merger or acquisition.

Figure 1: Comparison of S-1s in the Complete Sample and Completed IPOs

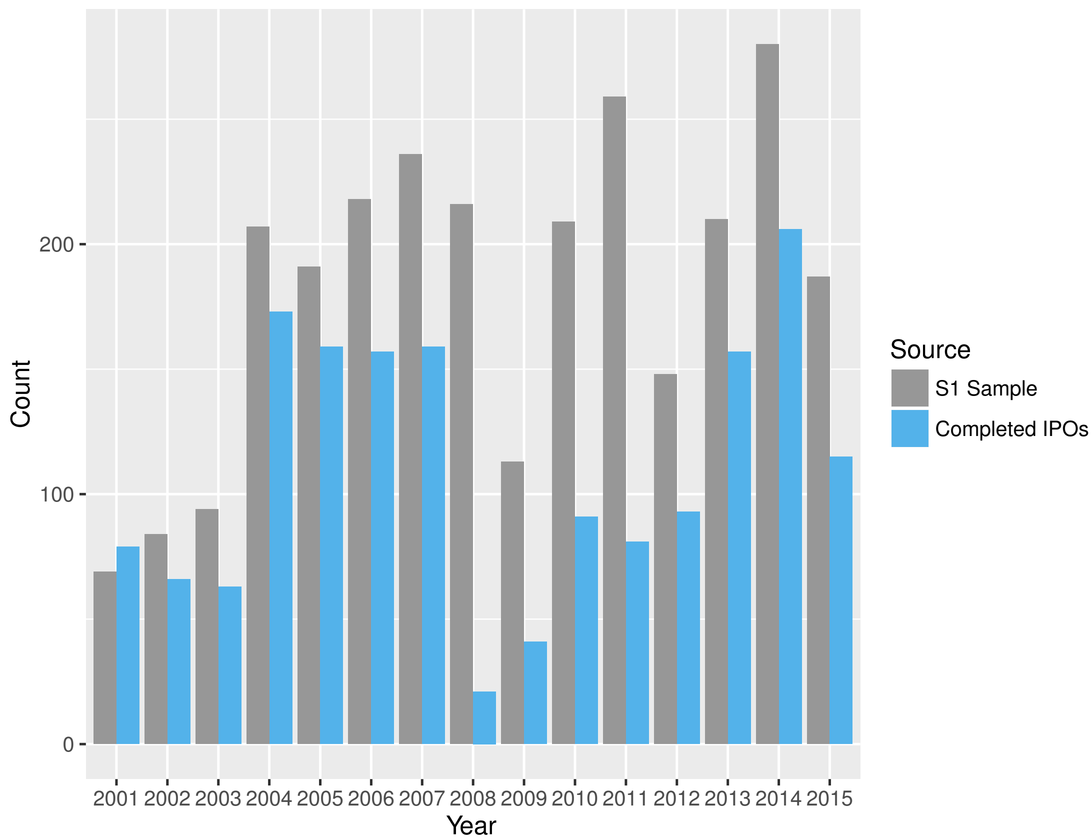

Figure 2: Location and Number of Firms that File a Registration Statement in the Complete Sample

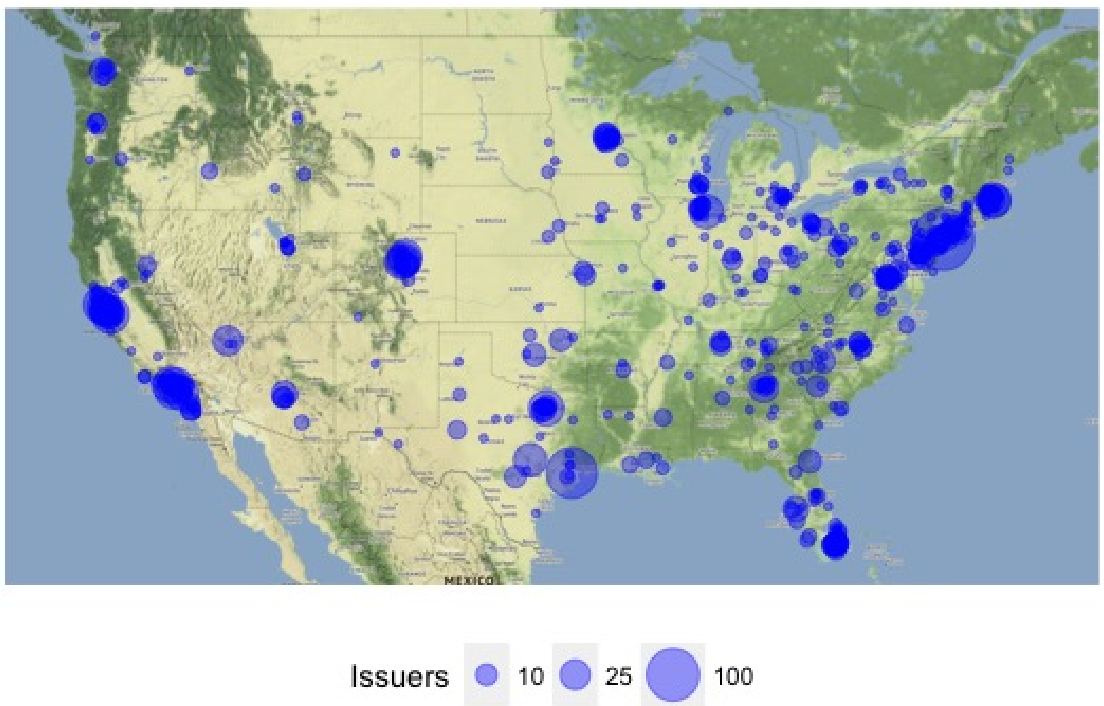

Table 1: Top 20 Locations of Law Firm Offices Representing S-1 Filers in the Complete Sample

City	S-1s Prepared
New York	962
Boston	204
Palo Alto	192
Washington DC	109
Chicago	103
Houston	101
Los Angeles	76
Menlo Park	74
San Diego	71
Philadelphia	50
San Francisco	48
Atlanta	36
Mountain View	36
Denver	33
Dallas	32
Seattle	31
Minneapolis	28
Trenton	28
Costa Mesa	25
Austin	20

Table 2: Count of S-1s Prepared by the Top 25 Law Firms in the Complete Sample

Law Firm	S-1s Prepared
Latham & Watkins	147
Cooley	121
Wilson Sonsini Goodrich & Rosati	114
Sichenzia Ross Friedman Ference	101
Goodwin Procter	72
Skadden Arps Slate Meagher & Flom	69
Kirkland & Ellis	67
Graubard Miller	65
Wilmer Cutler Pickering Hale and Dorr	57
Simpson Thacher & Bartlett	50
Vinson & Elkins	48
DLA Piper	47
Fenwick & West	46
Greenberg Traurig	41
Ropes & Gray	40
Morgan Lewis & Bockius	35
Ellenoff Grossman & Schole	34
Richardson & Patel	34
Davis Polk & Wardwell	31
Paul Weiss Rifkind Wharton & Garrison	31
Weil Gotshal & Manges	30
Foley & Lardner	29
K&L Gates	29
Bingham Mccutchen	27
Fried Frank Harris Shriver & Jacobson	27

Table 3: Means for Binary Variables in Pairwise Dataset

Statistic	N	Mean
First Listed Lawyer is the Same	3,700,560	0.002
Law Firm is the Same	3,700,560	0.016
Same Law Firm, Different Office	3,700,560	0.009
Same Lawyer, Different Law Firm	3,700,560	0.0002
Different Law Firm, Same Law Firm Zip Code	3,700,560	0.015
S-1s Filed within 90 Days of Each Other	3,700,560	0.038
Same Fama-French 48 Industries	2,201,851	0.076
Same Fama-French 30 Industries	2,201,851	0.123
Same Fama-French 12 Industries	2,201,851	0.156
Same Two-digit SIC (as listed in S-1)	1,290,421	0.073

Each sample contains a pairwise comparison of every unique pair of documents. The means report the percentage of unique pairs that share the indicated statistic.

Table 4: Mean and Standard Deviation of Textual Similarity Measure

Subset	N	Mean	St. Dev.
All Observations	3,700,560	0.016	0.038
All Observations (Compustat-Matched)	2,201,851	0.018	0.040
All Observations (SIC Listed in S-1)	1,290,421	0.018	0.041
Same Fama-French 12 Industry	343,646	0.057	0.076
Same Fama-French 30 Industry	270,218	0.068	0.082
Same Fama-French 48 Industry	167,026	0.081	0.094
Same Two-Digit SIC (as listed in S-1)	94,165	0.074	0.095
S-1s Filed within 90 Days of Each Other	138,950	0.020	0.048
Same Law Firm	61,058	0.041	0.082
Same Law Firm, Same Office	26,676	0.050	0.100
Same Law Firm, Different Office	34,382	0.034	0.064
Different Law Firm, Same Law Firm Zip Code	55,086	0.024	0.057
Lawyer Overlap > 0	8,788	0.087	0.147
Lawyer Overlap >0, Same Firm	8,123	0.090	0.150
Lawyer Overlap >0, Different Firm	665	0.045	0.089
First Listed Lawyer is the Same	8,128	0.087	0.148
Same First Listed Lawyer, Same Firm	7,543	0.091	0.151
Same First Listed Lawyer, Different Firm	585	0.043	0.087

This table reports the average of the textual similarity measure (cosine similarity) for each subset of unique pairs that shares the indicated feature.

Table 5: Lawyer Overlap Regression Results

	Dependent variable: Textual Similarity Measure				
	(1)	(2)	(3)	(4)	(5)
Same Law Firm	0.013**	0.009**	0.009**	0.009**	0.010**
	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Lawyer Overlap	0.032**	0.026***	0.025***	0.026***	0.015**
	(0.013)	(0.006)	(0.006)	(0.006)	(0.007)
Lawyer Overlap X Same Law Firm	0.042***	0.029**	0.031**	0.028**	0.036***
	(0.016)	(0.014)	(0.014)	(0.013)	(0.012)
Same Zip Code	0.010***	0.009***	0.009***	0.009***	0.007***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.001)
Same Law Firm, Same Office	−0.001	−0.002	−0.002	−0.003	−0.002
	(0.006)	(0.004)	(0.004)	(0.005)	(0.004)
Filed w/in 90 Days of Each Other	0.003***	0.001***	0.001***	0.001***	0.003***
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0004)
Same Fama-French 48 Industries		0.065***			
		(0.004)			
Same Fama-French 30 Industries			0.054***		
			(0.004)		
Same Fama-French 12 Industries				0.044***	
				(0.003)	
Same Two-Digit SIC Code					0.057***
					(0.004)
DV Mean	.016	.018	.018	.018	.018
Observations	3,700,560	2,201,851	2,201,851	2,201,851	1,290,421
R^2	0.086	0.284	0.290	0.261	0.237
S-1 Fixed Effects	Y	Y	Y	Y	Y

The dependent variable in the OLS regressions in this table is the measure of textual similarity (cosine similarity) between a pair of registration statements (S-1s). For the first regression there are 2,721 statements, which allows for 3,700,560 unique pairwise comparisons. For the second, third, and fourth regressions there are 2,099 S-1s that have been matched to Compustat for a total of 2,201,851 unique pairwise comparisons. For the final regression there are 1,607 S-1s that provide an SIC code with the S-1 for a total of 1,290,421 unique pairwise combinations. The independent variables measure similarities between the pairs of S-1s. All regressions include IPO fixed effects and robust standard errors are clustered by law firm. Statistical significance is denoted by: *p<0.1; **p<0.05; ***p<0.01.

Table 6: Same First Lawyer Regression Results

	Dependent variable: Textual Similarity Measure				
	(1)	(2)	(3)	(4)	(5)
Same Law Firm	0.013**	0.009**	0.009**	0.009**	0.010**
	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Same First Lawyer	0.027***	0.019***	0.017***	0.019***	0.013**
	(0.011)	(0.006)	(0.006)	(0.005)	(0.005)
Same First Lawyer X Same Law Firm	0.027**	0.016	0.019*	0.016	0.023***
	(0.014)	(0.011)	(0.011)	(0.010)	(0.009)
Same Zip Code	0.010***	0.009***	0.009***	0.009***	0.007***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.001)
Same Law Firm, Same Office	−0.001	−0.002	−0.001	−0.003	−0.002
	(0.006)	(0.004)	(0.004)	(0.005)	(0.004)
Filed w/in 90 Days of Each Other	0.003***	0.001***	0.001***	0.001***	0.003***
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0004)
Same Fama-French 48 Industries		0.065***			
		(0.004)			
Same Fama-French 30 Industries			0.054***		
			(0.004)		
Same Fama-French 12 Industries				0.044***	
				(0.003)	
Same Two-Digit SIC Code					0.057***
					(0.004)
DV Mean	.016	.018	.018	.018	.018
Observations	3,700,560	2,201,851	2,201,851	2,201,851	1,290,421
R^2	0.086	0.283	0.290	0.261	0.237
S-1 Fixed Effects	Y	Y	Y	Y	Y

The dependent variable in the OLS regressions in this table is the measure of textual similarity (cosine similarity) between a pair of registration statements (S-1s). For the first regression there are 2,721 statements, which allows for 3,700,560 unique pairwise comparisons. For the second, third, and fourth regressions there are 2,099 S-1s that have been matched to Compustat for a total of 2,201,851 unique pairwise comparisons. For the final regression there are 1,607 S-1s that provide an SIC code with the S-1 for a total of 1,290,421 unique pairwise combinations. The independent variables measure similarities between the pairs of S-1s. All regressions include S-1 fixed effects and robust standard errors are clustered by law firm. Statistical significance is denoted by: *p<0.1; **p<0.05; ***p<0.01.

Table 7: Regression Analysis of Registration Statements Prepared by Firms that Merge or Acquire

	Dependent variable: Textual Similarity Measure				
	(1)	(2)	(3)	(4)	(5)
Documents Prepared by Law Firm that Merges or Acquires	−0.010	−0.005	−0.003	−0.004	−0.006
	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)
Comparison of Pre- and Post-Merger S-1 from Same Law Firm	−0.002	−0.004	−0.006	−0.004	−0.006
	(0.006)	(0.007)	(0.007)	(0.007)	(0.004)
Comparison of Post-Merger S-1s from Same Law Firm	0.024***	0.011**	0.011*	0.014**	0.017***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.003)
DV Mean	.016	.018	.018	.018	.018
Observations	3,700,560	2,201,851	2,201,851	2,201,851	1,290,421
R^2	0.086	0.284	0.290	0.261	0.237
Ind. Control	None	FF 48	FF 30	FF 12	SIC
S-1 Fixed Effects	Y	Y	Y	Y	Y

The dependent variable in these OLS regressions in this table is the measure of textual similarity (cosine similarity) between a pair of registration statements (S-1s). For the first regression there are 2,721 statements, which allows for 3,700,560 unique pairwise comparisons. For the second, third, and fourth regressions there are 2,099 S-1s that have been matched to Compustat for a total of 2,201,851 unique pairwise comparisons. For the final regression there are 1,607 S-1s that provide an SIC code with the S-1 for a total of 1,290,421 unique pairwise combinations. The table omits the following independent variables: Same Law Firm, Lawyer Overlap, Lawyer Overlap X Same Law Firm, Same Zip Code, Same Law Firm, Same Office, Filed within 90 Days of Each Other, and the Industry Controls indicated in the last row of the table. All regressions include S-1 fixed effects and robust standard errors are clustered by law firm. Statistical significance is denoted by: *p<0.1; **p<0.05; ***p<0.01.