Measuring Contract Completeness: A Text Based Analysis of Loan Agreements*

Bernhard Ganglmair† Malcolm Wardlaw‡

December 27, 2015

Abstract

Contractual incompleteness is one of the core principles in much of corporate finance theory, but the lack of quantitative measures of completeness has made direct empirical testing difficult. This paper helps fill this gap by proposing several measures of contractual detail using text based analysis. We analyze the default sections of a sample of private loan contracts, generating several measures of contract detail and of the use of common "boilerplate" language. Contracts are more complex when there is greater default risk, more uncertainty, and longer maturities, and an increased likelihood of renegotiation. Default language also shares greater similarities for larger contracts and contracts with more lenders, suggesting a role for standardization at the expense of complexity. We also find evidence that more complex loan contracts are associated with increases in operating performance suggesting that contractual completeness is associated with greater investment efficiency.

Key words: natural language processing; incomplete contracts; contract complexity; contract completeness; loan contracts; text analysis; transaction costs economics

*We thank David Blei, Janet Gao, Bettina Grün, Gerard Hoberg, Price McKay, Meik Michalke, Gordon Phillips, Matt Taddy, and seminar participants at Cavalcade (2015), EFA (2015), FMA Venice (2015), NFA (2015), the Econometric Society World Congress (2015) and the University of Linz for helpful comments and suggestions.

†The University of Texas at Dallas, Naveen Jindal School of Management, 800 W. Campbell Rd. (SM31), Richardson, TX 75080, USA; E: ganglmair@utdallas.edu; P: +1-972-883-4736.

‡The University of Texas at Dallas, Naveen Jindal School of Management, 800 W. Campbell Rd. (SM31), Richardson, TX 75080, USA; E: malcolm.wardlaw@utdallas.edu; P: +1-972-883-5903.

1 Introduction

Debt is often viewed as one of the simplest forms of financing, in which a contract specifies a fixed repayment amount and the failure to repay generates a default condition and a change in control. However, researchers have recognized that corporate debt contracts contain a large number of provisions and contingencies surrounding this somewhat simple transaction. These provisions serve to help complete the contract by spanning more specific states under which the borrower is in default of the debt. Understanding the purpose and consequences of these provisions is therefore necessary to fully understand the role of debt in firm financing.

A large and important literature has helped fill this need, describing the role which accounting based covenants, such minimum coverage ratios and capital expenditure limits play in the design of debt contracts. However, while extremely important, these specific covenants usually make up only a small portion of the debt contract itself. Debt contracts are often extraordinarily detailed, and contain hundreds of detail that result from contract negotiations. Using text based analysis of over 3000 private debt contracts, we propose direct measures of the complexity and detail of debt contracts and the extent to which these contracts contain contingencies or clauses which enrich the contract state space. These metrics capture significant detail about the contract which is not captured by existing metrics such as the number of standardized accounting covenants.

Debt contracts are subject to significant fine-tuning in the negotiation process, and for most large private debt contracts the contingencies specified in the contract are largely firm specific. Using text based measures of dissimilarity between contracts, we find that contracts to the same borrower, but from different lenders, are significantly closer to each other than contracts from the same lender to different firms. Moreover, this difference in increased similarity between common borrower and common lender is extremely large, indicating that the form of the contract is primarily being negotiated at the firm level. The language and complexity specified in the events of default is therefore primarily driven by concerns at the firm level, rather than the bank. This result also strongly rejects the idea that the events of

default are simply boilerplate language added to every contract by the loan department of a given lender.

Somewhat surprisingly, the similarity between contracts overall is actually greater for larger loans and loans with multiple lenders. This suggests that there is some scope for standardization when there is more at stake. However, the increase in common firm similarity is greater for larger loans, while the increase in common lender similarity is greater for smaller loans. This implies the need for more firm specific detail when there is more at stake.

We find that complexity is positively related to measures of ex-ante asymmetric information, consistent with the idea that the cost of contractual incompleteness is significantly higher in the presence of adverse selection. In this setting, the costs of asymmetric information are therefore partially remediated by more efficient contracting. Along similar lines, complexity is significantly higher for loans with multiple lenders, suggesting that the syndication process also increases the cost of incompleteness.

We find no evidence that the operational complexity of the firm contributes to the complexity of the loan contract. Conglomerate firms with more diverse operations and firms with foreign operations do not have more complex contracts, with the effects being insignificant and slightly negative in both cases. We do find evidence that more complex debt structures leads to greater detail, where firms which finance their operations with a number of different types of financial instruments have more complex default provisions. Surprisingly, although contractual detail appears to be very firm specific, the complexity of the contract is not related to the complexity of the underlying operations of the firm. Rather, it appears to be primarily driven by the existing financial structure of the firm and the impact of the lending process on this financial structure.

We find some evidence that the complexity of the default specification is positively related to the likelihood of renegotiation. Loan contracts with more detail are renegotiated more often after initiation. This supports the ideas proposed by Roberts and Sufi (2009) that contract design may be designed to shape renegotiation rather than to simply preclude

it. There is weak evidence that renegotiation is negatively related to the average contract distance, meaning that more "custom" contracts are perhaps less likely to be renegotiated. This would be somewhat supportive of a Hart (2009) style framework in which complex contracts act as a reference point for future renegotiation.

We conclude our analysis by examining the impact of our measures of contractual detail on forward looking firm outcomes. Consistent with the idea that more complete contracts create less holdup and therefore allow for greater investment efficiency, we find that subsequent annual return on assets and sales growth are higher for firms which sign more detailed loan contracts, conditional on other contractual features such as loan size and covenant makeup. The overall evidence suggests that firms which are able to sign more complete loan contracts are better able to exercise their growth opportunities.

Formally, a complete contract is one which specifies the rights and duties of each party in every possible state of the world. Since it is usually infeasible to cover the entire state space, gaps must naturally or strategically arise in which the contract parties are subject to ex-post bargaining (to fill the gaps) or inefficient ex-post transactions (when the gaps remain). Research in incomplete contracts, beginning with early work such as Williamson (1985) and Dye (1985), has produced a great deal of important debate on the form and nature of contracts. Perfect contracting, on the other hand, is often a key assumption in a number of foundational models in finance, perhaps most notably in the Modigliani and Miller (1958) propositions. Defining perfect contracting has been critical to our understanding of the economic frictions that may prevent these economic models form holding, as well as in modeling the potential outcomes.

The literature on contracts has generated a great deal of debate on exactly why contracts may be incomplete and what potential costs may arise which cannot be solved through ex-post renegotiation. While a substantial amount of progress has been made in describing these contracting frictions, little research exists attempting to measure and test the form of contracts themselves. Part of the reason is that contract detail is often not easily quantified.

Contracts are, by definition, complex legal documents which specify duties and contingencies in formal language rather than easily defined data sets. We interpret *more complex* contracts as an attempt to span more of the future state space, arguing that complexity implies a "large number of clauses that are specified in detail" (Ariño and Reuer, 2006, 149) and a "greater [...] specification of promises, obligations, and processes for dispute resolution" (Poppo and Zenger, 2002, 708). In other words, we assume that contract completeness and contract complexity go hand in hand, and that complexity can at least partially be interpreted as an attempt to write a more complete contract.

We propose several measures of contract detail and complexity and apply them to a specific set of contracts, namely private loan contracts between firms and banks. More specifically, we consider the section of a loan contract that specifies the *events of default*. While it is sometimes assumed that default is a simple binary condition on the timely completion of periodic repayments, in reality the default provisions of debt contracts go well beyond a simple statement of non-payment or reference to certain covenant provisions. Instead, default provisions are usually highly detailed and often specify a large number of specific contingencies in an attempt to span many different states of nature. While some are fairly simple, many are several pages long, detailing a large number of provisions and possible outcomes that provide a highly detailed account of types of non-payment, restrictions on formal activities, and specify highly detailed descriptions of cross-default triggers. [1] Importantly for our work, default provisions are also fairly well spelled out in the uniform language of bank debt contracts. The vast majority of these loan agreements contain a well defined section entitled "Events of Default," which formalizes the states of the world in which the borrower is in default and outlines potential remedies.

We provide several sets of metrics for the complexity of these default sections. First, we perform a simple count of the words and sentences in the default sections and posit that longer default sections specify more clauses and contingencies. We use the number of *total words*

[1] Table A1 provides an example of an events of default section from our sample.

to capture this property.[2] We further conjecture that a larger number of different or *unique words* used to describe the events of default—capturing the size or variety of vocabulary—is associated with both more distinct events of default and a more detailed description of these events of default. While these measures are obviously noisy, they are simple to understand and provide a reasonable approximation of the level of detail. We supplement this basic word approach with a dictionary approach, using an extensive dictionary of legal and financial related terms.

Next, we next use each default section to estimate a probabilistic topic model (Blei et al., 2003) to discover the general ideas or themes covered in default sections and provide a more fine-tuned measure for the number of distinct types of clauses. Topic models utilize the natural distribution of words within written language to characterize the occurrence of specific topics within a given document. Generally speaking, a topic model is a latent variable model in which the distribution of topics described in a given document is estimated as a latent variable based on how the distribution of words in a document conforms to the distribution of words within a generic topic in natural language. The output of the model assigns a list of possible topics, as a distribution of words, to each document. These topics can be visually characterized by their most important words or word combinations. The number and concentration of topics within a document then provides a valuable measure of the scope of the contract. We use this approach to predict, for each clause in a default section, the distribution over topics. The most likely topic is then defined as the main topic of that sentence. This procedure then allows us to count the number of unique *main topics* in each default section. Documents with only a few main topics are relatively simple while documents with many main topics are more detailed.

Using the results from both the raw words and the identified main topics, we construct measures of similarity between contract pairs. Specifically, each contract can be thought of as a vector of words or topics over the space of N words or topics. A dissimilarity between two

[2]Kosnik (2014) and Moszoro et al. (2014) use the length of individual articles or the entire contract as a measure for flexibility and complexity, respectively.

contracts can therefore bet estimated as the cosine distance between these two vectors. The measure provides us with two important insights. First we can describe the similarity across different pairs, giving insight into which parties are driving the contract writing process. Second, we gain some insight into what factors drive the adoption of more "boilerplate" clauses.

Taken together, these measures provide a unique set of metrics for measuring the scope of debt contracts and provide empirical evidence for the tradeoffs inherent in writing more complete contracts. By focusing on large bank loan contracts, we are able to relate these measures to various firm characteristics as well as variation in quantifiable loan features such as loan amounts and maturity. This enables us to provide important insight into what types of economic agents choose more or less complete contracts.

In addition to providing insight as to the determinants and outcomes of loan contract detail, a central contribution of this paper to the finance literature is its unique new framework for analyzing broad questions about financial contracting. The availability of textual contract data has been increasing exponentially, and our analysis provides a unique way of analyzing basic contract detail when those contracts are difficult to classify into an item based data set—either because of the sheer volume of documents or a potential researcher bias in their classification. Future research can utilize these tools to provide additional understanding of contractual completeness in other settings beyond loan contracts.

Our analysis is related and contributes to a number of strands of literature in economics and finance. For a comprehensive survey of the growing literature on textual analysis in finance and accounting, see Loughran and McDonald (2015). Masten and Saussier (2000) provide an overview of the general empirical literature of contracting. Saussier (2000) constructs an index of contract completeness (as sum of the number of key clauses included in the contract) to test predictions from transaction cost economics. More recently, Kosnik (2014) (hydroelectric license contracts), Moszoro et al. (2014) (public procurement contracts), and Beuve et al. (2015) (public and private procurement contracts) use textual analysis to study

the tradeoff between flexibility and rigidity in contract language. The literature on probabilistic topic models is ever growing, and topic models have been used on a number of different types of document collections such as emails (McCallum et al., 2007), scientific abstracts (Blei et al., 2003; Griffiths and Steyvers, 2004) and articles (Blei, 2012; Hall et al., 2008), newspaper archives (Wei and Croft, 2006), and U.S. Supreme Court decisions (Livermore et al., 2015).

2 Theoretical Motivation

Ariño and Reuer (2006, 149) define complex contracts as "contracts with a large number of clauses that are specified in detail." We follow their line of reasoning and posit that more detailed and thus more complex contracts are more complete because—as Poppo and Zenger (2002, 708) conclude from survey evidence—they hold a "greater [...] specification of promises, obligations, and processes for dispute resolution." In other words, we assume that contract completeness and contract complexity go hand in hand, and we can measure contractual completeness by using a metric for contractual detail to capture the complexity of the contract.

The economics literature on contracts has presented numerous factors that determine the degree of contractual incompleteness, i.e., lack of detail or complexity. Our approach is that contracts are not incomplete by assumption, but incompleteness is endogenously determined and parties to the contract may indeed find an incomplete contract more favorable. In this spirit, we discuss three sets of factors that have been identified as such determinants. First, drafting costs and benefits as a primary source of transaction costs; second, ex-ante asymmetric information; and third, costs of ex-post renegotiation of the contract.

2.1 Drafting Costs and Benefits

Contractual incompleteness has been broadly linked to transactions costs (Williamson, 1985, 1989). One type of such transaction costs are the costs associated with the drafting of the contract. At the early stage of drafting a contract, "search costs" represent the time and lost value inherent in researching and analyzing contingencies (e.g., Klein, 2002; Tirole, 2009), whereas at a later stage "ink costs" represent the time and lost value inherent in specifying these contingencies. These costs are associated with the actual costs or limitations of drafting a contract (Anderlini and Felli, 1994; Battigalli and Maggi, 2002, 2008; Dye, 1985; Melumad et al., 1997) and increase in the detail, precision, or complexity of the contract (Bajari and Tadelis, 2001).

A second transaction cost is the cost of enforcing and implementing contracts. A direct cost of enforcement is the cost of litigating contracts. Schwartz and Watson (2004) argue that more complex (i.e., complete) contracts may be more costly to enforce (i.e., litigate) because more evidence is required. Another type of implementation costs are monitoring costs. More detailed contracts with more clauses imply higher costs of monitoring in order to detect violations. Given fixed costs of drafting a clause, if ex post monitoring costs increase and parties anticipate that enforcing some of these clauses will be too costly, contracts are predicted to be less complete.

We hypothesize that higher drafting costs result in less complete contracts. As such, more operationally complex companies—associated with higher drafting costs—may have simpler contracts.

Also related to the effect of transaction cost is the transaction benefit. We predict contracts will be more complete when there is greater value at stake since the direct tradeoff between the cost of specifying contingencies is offset by the greater value of specifying a more complete state space. In the context of debt contracts, we hypothesize more detailed contracts will be positively correlated with larger loan amounts and longer maturities, since they both increase the overall value at stake.

2.2 Asymmetric Information

Asymmetric information is an important feature in many models of contractual incompleteness. The relationship between information asymmetry and completeness depends on the specific framework. We take a broad view based on the transactions costs literature, where the costs of writing a complete contract are held fixed, but the costs of incompleteness vary with the potential for possible adverse selection and moral hazard. In these models, problems in asymmetric information can be partially remedied via more detailed contracting. Costly contracting acts as either a disciplining mechanism or as a screening mechanism for borrowers. If detailed contracting is costly in general, contracts will be less complex when information is ex-ante more symmetric. Specifically, we hypothesize that contract complexity will be negatively related to firm age, asset tangibility, whether the borrower has a repeat relationship with the lender, and whether the lender is geographically close to the borrower.

3 Data and Methodology

3.1 Loan Documents

3.1.1 Data Source

For our analysis we construct two collections of documents. First, our initial sample comes from Nini et al. (2009), who extract the texts of a set of private loan contracts from their associated filings in the EDGAR database based upon a manual search starting from the Dealscan loan database.[3] We manually extract the sections that list the events of default for F_S = 3438 of these contracts to obtain our *full sample*. Second, for our discussion of renegotiation we collect a separate set of R_S = 250 contracts which formed the basis for Roberts (2015) examination of dynamic contract renegotiation. These 250 contracts represent 340 unique loan facilities, and are the original loan contracts whose future renegotiations

[3]See Nini et al. (2009) for a detailed description of the search process. The data is available on Amir Sufi's website at http://faculty.chicagobooth.edu/amir.sufi/data.html.

have been tracked. For this *renegotiation sample*, we again extract the sections that specify the events of default. In a final step, we split the documents in both document collections into their individual clauses.[4] For the full sample we obtain $F_C = 69607$ clauses, for the renegotiation sample we end up with $R_C = 5421$ clauses.

3.1.2 Terms and Phrases

We represent each document d (either section or clause) in our document collections as a vector $\vec{w}_d = (w_{d1}, \ldots, w_{dn})$ of the frequencies w_{dg} of each term g of n different terms in document d. For a list of these n terms, we construct a dictionary of terms, phrases, and abbreviations from finance, accounting, and law.[5] The full sample contains $n = 3192$ distinct terms and phrases, the renegotiation sample $n = 1662$ distinct terms and phrases. We parse each document to obtain frequencies w_{dg} for each term g to construct frequency vectors $\vec{w}_d$ for each document d.

Table 1 provides summary statistics for our four document collections. *Terms and phrases* is the total number of terms and phrases in a given document d reflecting the length of the section or clause: $\sum_{g=1}^{n} w_{dg}$. *Unique terms and phrases* is the number of distinct terms and phrases reflecting the size of the dictionary used: $\sum_{g=1}^{n} \mathbf{1}_{+} w_{dg}$ with $\mathbf{1}_{+} = 1$ if $w_{dg} > 0$ and zero otherwise. *Terms and phrases (finance)* and *Terms and phrases (law)* represent the total number of terms and phrases for a subset of the finance and law terms in the dictionary. *Clauses* is the number of clauses per section, and the *Average length of clauses* is the average length of a section's clauses, measured as the number of terms and phrases.

[4] The format of these sections is fairly standardized, allowing us to use periods and semi-colons as delimiters to obtain tokens. We delete tokens that do not contain a verb or verb form. The remaining tokens are the sections' clauses.

[5] The full dictionary comprises 40489 terms. We use Campbell Harvey's finance glossary (`http://people.duke.edu/~charvey/Classes/wpg/glossary.htm`), Black's Law Dictionary (Black et al., 1990), the legal dictionary at the Legal Information Institute (`https://www.law.cornell.edu/wex/all`), and the online version of the New Oxford Companion to Law (Cane and Conaghan, 2009). We stem the terms and phrases, that means, we erase word suffixes to obtain the words' radicals, using the *R* implementation of Porter's stemming algorithm (Porter, 1980).

3.1.3 Topic Models

Because multiple clauses may simply represent extraneous detail on a single event of default, we also count the number of "main topics" in a section as a proxy for distinct events of defaults. These main topics can be uncovered by means of probabilistic topic models that help discover the themes or topics in a sample of text documents. We use the *Latent Dirichlet Allocation (LDA)* model (Blei et al., 2003) which Blei and Lafferty (2009) describe as the "simplest topic model" and "has proven hugely popular" (Taddy, 2012). These models have recently begun to see significant use in finance and accounting related research, see Huang et al. (2015), Ball et al. (2015), Gupta and Israelsen (2015), and), Kogan et al. (2009), as a tool for analyzing language based data.[6]

Probabilistic topic models uncover the latent topical structure of a document by analyzing the co-occurrence of terms and phrases used in the document.[7] The underlying idea is that authors first decide which topics (e.g., events of default) to cover before drafting a document. A document thus becomes a collection of multiple topics. LDA describes such a topic k as a per-topic word distribution $\vec{\beta}_k$ over the vocabulary of terms and phrases. Moreover, for a document collection D, holding documents that cover K topics, each document $d \in D$ will exhibit these K topics with different proportions according to a per-document topic distribution $\vec{\theta}_d$.

The data we observe are the documents in a collection D and the terms and phrases $\vec{w}_d$ used in each document $d \in D$. The topics, however, are not observed. We apply LDA to reverse the process of topic generation and automatically discover the latent topical structure. To obtain a specific event of default for a given clause, we estimate the topic model with $K = 50$ topics on our document collections with clauses (both full sample and renegotiation

[6]Blei and Lafferty (2009), Blei (2012), or Steyvers and Griffiths (2011) provide an introduction to probabilistic topic models. Topic models have been used on a number of different types of document collections such as emails (McCallum et al., 2007), scientific abstracts (Blei et al., 2003; Griffiths and Steyvers, 2004) and articles (Blei, 2012; Hall et al., 2008), newspaper archives (Wei and Croft, 2006), and U.S. Supreme Court decisions (Livermore et al., 2015).

[7]The approach taken is a "bag-of-words" approach in which the order of terms and phrases does not matter. See Blei (2012) for a discussion of some of the assumptions and proposed extensions.

sample). For each clause c in a full section s, we obtain a per-document topic distribution $\vec{\theta}_{cs} = \left(\theta_{1|cs}, \ldots, \theta_{K|cs}\right)$ over K topics. Each $\theta_{k|cs}$ represents the density with which a topic k is covered in clause c, with $\sum_{k=1}^{K} \theta_{k|cs} = 1$. We assume that each clause contains one *main topic*, and use this distribution $\vec{\theta}_{cs}$ to construct a count of main topics (or events of default) for each full section. This means, for each clause c in a full section s, we find the *main topic* k_{cs} as the one with the highest topic density $\theta_{k|cs}$:

$$k_{cs} = \arg \max_{k=1,\ldots,K} \theta_{k|cs}. \tag{1}$$

The union of main topics as the set of topics $k \in K$ that are a main topic for at lease one clause c thus serves as a proxy for the set of events of default specified in a full section s. The number of main topics as the number of events of default is the cardinality of this set:

$$\text{Main topics}_s = \left| \bigcup_{c} k_{cs} \right|. \tag{2}$$

Table 1 provides summary statistics for the number of main topics per section for both the full sample and the renegotiation sample.

In a next step, we characterize the list of *common topics* or events of default as the main topics with the highest frequency with which they appear in a section. For a subset of our results, we will limit our sample to the five most frequent topics. In other words, for each full section s, we keep only those clauses whose main topic is among the top five. By excluding more exotic or special topics, we can focus on the language that is used to describe events of default that are common across a large number of contracts. For each clause c that exhibits one of these common topics as main topic, we obtain the length of the clause as the total number of words.[8] Table 2 summarizes these numbers for the full

[8]For this measure we consider all non "stop word" unigrams in a clause. We first pre-process documents in the following steps: convert all words to lower case, remove punctuation, and delete numbers, number words (1 through 100), roman numerals, and stop words (such as "and", "the", or "that"). For a list of stop words to exclude, we adapt the list provided by the SMART information retrieval system (Salton, 1971) containing 571 words. In a last step we stem the words, that means, we erase word suffixes to obtain the words' radicals,

sample. It also provides a weighted average for the five (ten) most common topics where the length of a clause is weighted by the median length of all clauses of a given common topic. We employ this normalization to avoid our measure to be dominated by the variation in events of default that require a longer description. The numbers in Table 2 illustrate that the contractual language exhibits a sizable variation both across events of default and within events of default.

Table 3 provides additional information for the common topics. First, we describe the topic through a representative list of terms and phrases that define the topic, based on the per-topic word distribution $\vec{\beta}_k$ with the density of a given term $g = 1, \ldots, n$ in topic k denoted by β_{kg}. For each common topic, we list the ten most relevant terms and phrases. Sievert and Shirley (2014) define *relevance* of a term or phrase g in topic k as

$$\text{relevance}_{gk} = \lambda \log \beta_{kg} + (1 - \lambda) \log \left(\frac{\beta_{kg}}{\omega_g} \right) \tag{3}$$

where ω_g is the relative frequency of a term g in the entire document collection (full sample).[9] We also list representative clauses for each of the common topics. We rank each clause (from the full sample) that exhibits a common topic by its length and list the clauses at the 10th as well as the 90th percentile.

3.1.4 Distance Measures

The distance between the loan documents gives us a measure of how similar one loan contract is from another, and by extension, how similar the unique writing of a given contract is relative to the rest of the sample. In other words, the average distance of a document $d = i$ from all other documents provides us with a measure of how customized the language of a

using the *R* implementation of Porter's stemming algorithm (Porter, 1980). The total number of words is the sum of all remaining unigrams.

[9] For $\lambda = 0$ this relevance measure reduces to the *lift* of a term (Taddy, 2012), that is, the ratio of a term's probability within a topic (β_{kg}) to its probability in the entire document collection (ω_g). A value of $\lambda = 1$ ranks the terms and phrases according to their topic-specific probabilities β_{kg}. We follow Sievert and Shirley (2014) who suggest a value of $\lambda = 3/5$.

given document is.

For our distance measures, we use the *cosine similarity* that measures the cosine of the angle between two vectors i and j. These vectors are taken to be representative of the language of the contracts. We use two different approaches to representing a document and thus measuring the "distance" between documents. First, using the vector $\vec{w}_i$ for the frequency counts of terms and phrases, we calculate the *cosine distance* between two documents i and j as follows:

$$\text{word distance}_{ij} = 1 - \frac{\vec{w}_i \cdot \vec{w}_j}{||\vec{w}_i||\ ||\vec{w}_j||}$$

for all $i \neq j$ and $i = 1, \ldots N$ where $N = F_S$ for the full sample or $N = R_S$ for the renegotiation sample. Two documents with vectors $\vec{w}_i$ with the same orientation thus have a cosine distance of zero. The average pairwise distance of a document i from all other documents is then equal to

$$\text{average word distance}_i = \frac{\sum_{j \neq i} \text{word distance}_{ij}}{N - 1}. \tag{4}$$

For the second distance measure, we use the results from the topic models above. For each of the $K = 50$ topics, we determine if a topic k is a main topic for at least one of the clauses in a section i. This yields a vector $\vec{k}_i$ with $K = 50$ elements, each being equal to 1 if topic k is a main topic at least once, and zero otherwise. The *cosine distance* between two documents i and j is then:

$$\text{topic distance}_{ij} = 1 - \frac{\vec{k}_i \cdot \vec{k}_j}{||\vec{k}_i||\ ||\vec{k}_j||}$$

for all $i \neq j$ and $i = 1, \ldots N$. Two contracts with the same set of main topics thus exhibit a cosine distance of zero. The average pairwise distance of a document i from all other

documents is then equal to

$$\text{average topic distance}_i = \frac{\sum_{j \neq i} \text{topic distance}_{ij}}{N-1}. \tag{5}$$

Table 1 provides basic summary statistics for these measures of average pairwise distance.

3.2 Loan and Firm Specific Data

We match the loan contracts to firm level data in CRSP/Compustat and to loan level data in Dealscan. For each loan contract, we match the financial data from the most recent closing quarter prior to the initiation of the loan. We calculate the daily stock return volatility over the previous four quarters up to the most recent closing period. Firm size is the log of total assets, the tangible asset ratio is the net property, plant, and equipment divided by the total assets. Firm age is the number of years the firm has appeared in Compustat. Segment concentration is calculated as the Herfindahl-Hirschman (HHI) index of sales across all business segments in the Compustat segments file as of the most recent closing year. Debt type concentration is the HHI index of debt type (ordinary, convertible, secured, subordinated debt, and preferred stock) as a fraction of total debt and preferred stock. Age is calculated as the number of years the firm has existed in Compustat, and stock return volatility is calculated as the annualized daily volatility over the previous years. Finally, in order to measure default risk more directly, we estimate the "Expected Default Probability" from a Merton (1974) model using the methodology of **?**. This estimates the theoretical default probability for each sample firm prior to the signing of the loan.

For each contract, we match the loan to its associated record in the LPC Dealscan database. Using the Dealscan records, we calculate the total dollar amount of all facilities in the loan package and the average maturity in months. For loan packages with multiple loans, we match the contract section in the primary facility which usually represents the most detailed terms. We match the number of unique lenders for each loan package from

Dealscan records, and we record whether the loan was from a repeat lender, which we define as having borrowed from the same bank within the past 5 years. We calculate a "local bank" indicator which takes a value of one if one of the lead lenders has its headquarters within 100 miles of the firm's headquarters. We also count the number of loan covenants as determined by Nini et al. (2009). Summary statistics for firm and loan level data are provided in Table 4. Correlation coefficients are reported in Table 5.

4 Determinants of Contractual Detail and Similarity

4.1 Who Determines the Contract?

We begin our discussion by first examining our measures of contract similarities. As mentioned in Section 3, we calculate the similarity between two documents as the cosine distance between vectors of main topics or of words in a V-dimensional vocabulary space. Each contract pair generates a cosine distance ranging from 0 to 1, where 0 is identical occurrence of each 1 is maximally different. The difference between contracts can give us a sense of how specific each contract is to a given firm, bank, industry, or time period. Since the distance between two given contracts can be expressed as a function of having the same borrower, lender, or industry, the overall change in distance helps distinguish whether these default provisions are firm specific or whether they are largely determined at the bank level.

For the entire space of contracts, we have 5,502,903 unique contract pairs representing 3318 contracts.[10] For each pair of contracts, we observe whether the pair of contracts shares the same borrower, the same lender, the same industry (4-digit SIC code) or was written in the same year. Each of these indicators indicates the conditional difference in mean distance between contracts as a function of sharing the same characteristics. This gives us a measure of the extent to which shared similarities in the loan counter-parties affects the similarity of the loan contract itself.

[10] We lose 126 contracts due to parsing errors.

We present results for both the topic distance and the overall word distance in Table 6. Column (1) presents the results for the full sample, with indicators for loans made to the same borrower, loans made from the same lead lender, loans made during the same year, and loans made to borrowers in the same industry. While each of these indicators is statistically significant, the *same borrower* indicator is an order of magnitude larger than the *same lead lender* indicator, being about 6 times larger in the case of topic distance and 10 times larger in the case of word distance. This result strongly rejects the idea that the events of default are simply a set of boilerplate terms offered by a given bank. While lenders do exert influence on the common components of the contract, this influence is dwarfed by the influence of the borrower. The firm appears to be actively negotiating similar contracts from different banks, conditional on its own needs, rather than simply acting as a passive "taker" of contract terms set by each individual bank. It is reasonable to conclude therefore that the firm holds significant bargaining power over the form of the contract.

In column (2) we separate contract pairs into groups which are above and below $100 million in total loan amount to understand how loan size affects the customization of default language. Similarly, in column (3) we separate contract pairs which are loans made funded by multiple syndicated lenders and loans made by a single bank. Since we are interested in comparing contract pairs within these groups, this reduces the overall sample size as we exclude, for instance, pairs in which one loan is above $100 million and another is below $100 million. Larger and more widely syndicated loans are significantly "closer" to each other than smaller, single lender loans. Larger loans appear to have less scope for customization in terms of allowable defaults. This may result from issues of enforceability, where larger loans demand more uniformity and give less scope to the borrower for negotiation of terms.

To test this question more directly, we interact the *same borrower* and *same lead lender* terms with the size and syndication dummies in columns (5) and (6). The results support this basic hypothesis, where the influence of having the same lead lender becomes larger for larger, syndicated loans while the same borrower impact is reduced. This indicates that the

lenders appear to exert more influence over the terms of the loan when the loan is larger and funding is more complex. It also indicates that each lender is influenced by a set of unique economic circumstances which drive the negotiating process.

4.2 Determinants of Complexity

We first examine the firm and loan level determinants of our word count measures. We regress the total word count and the unique word count on firm and loan level measures of contracting costs, information, and risk in the quarter just prior to the initiation of each loan. We also add several controls for the financial condition and investment opportunities of the firm, as well as the number of covenant restrictions. Each of our specifications also contains year fixed effects to account for possible time variation in the structure of loan contracts.[11]

The results are presented in Tables 6. Each of the columns presents a different complexity measure for the contracts. Columns (1) and (2) present the number of total (1-gram) words and the number of unique words. Columns (3) and (4) use a dictionary approach, counting the number of law related terms and finance related terms identified by their respective dictionaries. Finally, to avoid simply capturing arbitrary "wordiness" in the contract, column (5) present results for the number of individual clauses, and column (6) uses the number of uniquely identified main topics from our 50 topic LDA topic model. As previously mentioned, this last approach is an attempt to capture "unique clauses" in a meaningful linguistic way.

The results are fairly similar across most of our measures. As might be expected, higher firm leverage higher credit risk, as proxied for by stock return volatility and our Expected Default Frequency measure, leads to a lengthier events of default section in each loan. The detail of each default section is correlated with default risk, in much the same way that the application of covenants is correlated with default risk. However, even after controlling for the number of debt covenants, these items have considerable predictive power, implying that these measures are capturing additional information beyond covenant detail.

[11]Note that we lose a number of observations in our determinant regressions due to missing data for the determinants.

Consistent with our hypotheses regarding stake size, both word count measures are significantly positively related to loan amount and maturity. Simply put, when the monetary stakes are larger overall in terms of both principal and cumulative interest, there is more need for greater detail in specifying the default states. This effect is fairly consistent across all specifications, with the exception of our clause count measure which appears to have slightly worse fit across the board.

When the lead lender is located in close geographic proximity to the borrower (within 100 miles), there is a significant reduction in overall complexity. Consistent with the literature on local finance, contract detail appears to exist in part to contract away adverse selection problems between the borrower and lender. This demonstrates another dimension along which local information impacts the form of firm financing. However, repeat lending relationships do not appear to affect the complexity of the contract. This is perhaps surprising given the existing work on lending relationships over time, but the information benefits of repeat lending relationships do not appear to impact the specificity of defaults. Loans which are syndicated to multiple lenders contain significantly more complex language than loans with a single lender. The syndication process appears therefore to require more complexity in the loan contract in order to resolve information problems between lenders. This also lends context to our contract pair tests, where lenders appear to demand more lender specific provisions when the loan is syndicated to multiple banks.

One further idea which we wish to examine is whether the complexity of the contract is merely a proxy for the complexity of the firm. This is important in light of the fact that the firm, rather than the bank or the industry, appears to be the most important determinant of contract specificity. To answer this question, we construct two measures of operational complexity and one further measure of financial complexity. To proxy for firm complexity we take the segment concentration, which is the HHI index of sales concentration across Compustat business segments and a dummy taking on a value of 1 where the firm reports income from foreign operations. Segment concentration appears to have no effect on

contract complexity, while the presence of foreign operations has a very marginal negative effect, suggesting that the complexity of these default sections do not simply reflect the need to specify states across more operational outcomes.

Conversely, the complexity of the firm's existing financing does appear to impact the complexity of the contract. We measure debt concentration as the HHI index of the book value of the firm's debt and preferred stock financing. Specifically, we calculate the concentration measure of the percentage of financing made up by convertible debt, secured debt, subordinated debt, ordinary debt, and preferred stock. Firms with a lower existing debt concentration show significantly more complexity in their new bank loan contracts. The importance of properly specifying the default space is therefore not limited to the debt extended in the contract itself. Rather it increases as the complexity of the financial claims against the firm increases. Moreover, it appears to do so in a way that is not simply covered by a standard cross-default clause.

4.3 Contract Language Distance and "Boilerplate" Contracting

In addition to measuring the relative amount of detail of these default sections, we can also get some sense of the relative standardization of the contract detail by examining the average distance of the contract to all other contracts in the sample. Specifically, we take the pairwise cosine distance measures described in Section 4.1 and take the average distance between each contract and all other contracts in the sample. This roughly measures how customized the language and terms of a given contract are relative to all other contracts. Contracts with a high average distance contain more language which is not found in the majority of contracts in the sample and omit words or topics which are found in the majority of the sample. While this measure is rough, it does give some sense of the degree to which firms and lenders are writing contracts which are more boilerplate in nature.

Using these measures of average distance as a measure of contract customization, we estimate the determinants of overall contract similarity as a similar function of firm and loan

level variables. In Table 8 we present the results of these regressions. Since any average distance can be biased in one direction or another as a function of the counts, we also control for the number of main topics and words in each regression. As suggested by the pairwise results in Table 6, loans which are syndicated amongst multiple lenders contain more standardized Topics. There is suggestive evidence that repeat lending relationships may have slightly more custom topic usage, though this result is only barely significant.

Average distance appears to also be positively related to the expected default frequency, meaning that firms with greater default risk are more likely to to have default provisions which are less boilerplate in addition to being more detailed. Taken together, this supports theories in which debt standardization allows for greater access to capital, but where such standardization comes at a cost of screening out riskier borrowers. To the extent that riskier borrowers require more contracting detail to complete the state space, this necessary level of detail may be hampered by the information frictions between lenders.

4.4 Variation in Detail Among Standard Clauses

To further mine the details of our topic model, we examine how specific topics vary in their level of detail. A detailed inspection of the assignment of main topics for each clause in our topic model reveals a fairly low level of misclassification for the most frequent topics. That is, taking the highest probability of a topic for each clause yields a topic assignment which only rarely seems to be unrelated to the rest of the identified clauses in other documents. However, the model also tends to assign more general ideas, such as clauses which deal with the specifics of bankruptcy filings into several different topics.

In order to get a sense of how the length of a contract impacts the overall depth of a given topic, we examine the top five most common topics and test which factors determine how detailed the contract clauses which describe that topic are. We use the top five as they are the most well identified across all the individual contracts, appearing as a uniquely main topic in at least 60% of all contracts. Thus these topics, and way in which they describe

detail, are fairly "boilerplate" in their use across all contracts, and the remaining variation is a function of the total depth of the language used in each clause or clauses in the contract.

For each of these topics, we take all clauses in the contract with a uniquely identified main topic. We then count the total number of words used assigned to each topic and regress it against firm and loan level determinants. We also wish to normalize each measure so as to aggregate the relative number of words across these Top 5 topics. To do so we take log ratio of the number of words divided by the median number of words in the sample.

The tests, reported in Table 6 examine topics 23, 20, 29, 26, and 22. Collectively, these topics are uniquely identified in all but 4 of the contracts we examine. The topics, their representative words, and and their representative terse and detailed clauses are examined in Table 3. Roughly speaking, topic 23 describes breaches relating to false representations and warrantees, that is information which the firm asserts is true upon signing the loan. Topic 20 details defaults relating to general failure to repay the loan. Topic 29 concerns cross-default provisions, i.e. defaults which are triggered by the firm defaulting on other loan contracts that the one being described. Topic 26 concerns defaults related to a change in control, and topic 22 concerns legal judgements against the firm.

While the explanatory power of each of our models is somewhat poor, with the R-squared measure shrinking considerably from the full model, we can still see a few trends emerge. Topic 23, detailing false representations and warrantees, for instance, appears to be somewhat unique in that the impact of tangible asset ratio is reversed. While overall detail is negatively correlated with tangible assets, perhaps due to their relative transparency, they appear to rely more on the certification of observable information in specifying a complete contract. Topic 29 which concerns cross-default is far more detailed for loans with multiple syndicated lenders and firms located far from their lead lender. Consequently, the information issues surrounding these loans appear to be particularly important with respect to conflicts with other creditors. The remaining topics are somewhat difficult to classify, with only age being important for Topic 26 (change in control) and Topic 22 (legal judgements.)

4.5 Contract Detail and Renegotiation

Finally, for a subset of contracts we examine whether the complexity of the contract impacts the frequency with which a loan contract is renegotiated. For this exercise, we track down as many of the original loan contracts from Roberts (2015) sample of renegotiated loan contracts and calculate our complexity measures based on the text of the initial contract. Similar to this analysis, we estimate a zero-inflated negative binomial model to estimate the number of times a given loan was amended as a function of our complexity and similarity measures. For our topic models and average distance measures, we calculate their values as compared to the full sample of contracts examined in the rest of the paper, rather than just the renegotiated subsample. Table 10 reports these results.

While our sample is fairly small, we do find some evidence that the number of renegotiations is positively related to the complexity of the contract and negatively related to the customization of the contract. This result implies that more complex contracts do not preclude renegotiation. Rather, they act as a loose framework for shaping renegotiation.

5 Contract Detail and Ex-post Changes in Performance

In this section, we examine the relationship between our contract completeness measures and measures of ex-post performance and cash saving behavior. More complete contracts should be related to the relative efficiency of investment. Further, if firms are willing to bear the costs of greater contract specificity, it is likely that they have an expectation of future performance increases as they exercise their growth options.

To investigate this, we examine the relationship between future return on assets and sales on our measures of contract detail. We regress the annual change in return on assets, summed over the four quarters starting after the initiation of the loan on our measures of contract detail. We also examine sales growth over the same period. Tables 11 and 12 present the results of these tests.

The annual change in ROA is positively related to each measure of complexity. Similarly, annual sales growth is also positively related to these measures. In both cases, greater contract detail is associated with an increase in overall performance. This increase in performance also appears to be distinct from any correlation with credit quality. The change in return on assets is not significantly related to the credit quality of the firm at the invitation of the loan, and the change in sales growth is negatively related to these measures, with lower quality firms having lower overall sales growth. Overall, the signing of more detailed contracts predicts an increase in subsequent performance that is unrelated to the risk of the firm as of the contract signing.

Firms also appear to retain more cash in the year subsequent to signing a debt contract when that contract is more detailed. This is partially the result of the greater returns experienced by the firm over the period, but it does imply that the cash tends to be saved rather than reinvested. It is somewhat difficult to test whether this cash is being used for precautionary savings, but it does at least suggest that these firms are generating more cash relative to future investment needs.

6 Conclusion

We propose several new text based measures of loan contract completeness. We find strong evidence of a cost and benefit to contractual detail, where more complete contracts come at a benefit to investment efficiency, but in which writing detailed contracts has a significant cost. We find that renegotiation and information costs play a significant role in the writing of debt contracts, and that firms which are able to write more detailed and complete contracts see greater future returns and sales growth.

Our measures provide a direct method for analyzing text based contracts without the need to categorize the details of the contract into potentially arbitrary categories. By applying analytical measures directly to the text of contract, we also open up new possibilities for

research by eliminating the need to manually categorize these complex textual documents. This should open up new avenues for future research in analyzing issues in contractual completeness and the large number of detailed contractual forms which bind firms to their various stakeholders.

References

Anderlini, L. and L. Felli (1994, November). Incomplete written contracts: Undescribable states of nature. *Quarterly Journal of Economics 109*(4), 1084–1124.

Ariño, A. and J. J. Reuer (2006). Alliance contractual design. In O. Shenkar and J. J. Reuer (Eds.), *Handbook of Strategic Alliances*, Chapter 9, pp. 149–169. Thousand Oaks, CA: SAGE Publications.

Bajari, P. and S. Tadelis (2001, Autumn). Incentives versus transaction costs: A theory of procurement contracts. *RAND Journal of Economics 32*(3), 387–407.

Ball, C., G. Hoberg, and V. Maksimovic (2015, August). Disclosure, Business Change and Earnings Quality. SSRN Scholarly Paper ID 2260371, Social Science Research Network, Rochester, NY.

Battigalli, P. and G. Maggi (2002). Rigidity, discretion, and the costs of writing contracts. *American Economic Review 92*(4), 798–817.

Battigalli, P. and G. Maggi (2008). Costly contracting in a long-term relationship. *RAND Journal of Economics 39*(2), 352–377.

Beuve, J., M. Moszoro, , and S. Saussier (2015, January). Political contestability and contract rigidity: An analysis of procurement contracts. Working Paper 2475164, Social Science Research Network.

Black, H. C., J. R. Nolan, and J. M. Nolan-Haley (1990). *Black's Law Dictionary: Definitions of the Terms and Phrases of American and English Jurisprudence, Ancient and Modern* (6 ed.). St. Paul, Minn.: West Publishing.

Blei, D., A. Ng, and M. Jordan (2003, January). Latent dirichlet allocation. *Journal of Machine Learning Research 3*, 993–1022.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM 55*(4), 77–84.

Blei, D. M. and J. D. Lafferty (2009). Topic models. In A. N. Srivastava and M. Sahami (Eds.), *Text Mining: Classification, Clustering, and Applications*. CRC Press.

Cane, P. and J. Conaghan (2009). *New Oxford Companion to Law*. Oxford University Press.

Dye, R. A. (1985). Costrly contract contingencies. *International Economic Review 26*(1), 233–250.

Griffiths, T. L. and M. Steyvers (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences 101*(suppl 1), 5228–5245.

Gupta, S. and R. D. Israelsen (2015, April). Hard and Soft Information: Firm Disclosure, SEC Letters, and the JOBS Act. SSRN Scholarly Paper ID 2473509, Social Science Research Network, Rochester, NY.

Hall, D., D. Jurafsky, and C. D. Manning (2008). Studying the history of ideas using topic models. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 363–371.

Hart, O. (2009). Hold-up, asset ownership, and reference points. *The Quarterly Journal of Economics 124*(1), 267–300.

Huang, A., R. Lehavy, A. Zang, and R. Zheng (2015, September). Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach. SSRN Scholarly Paper ID 2409482, Social Science Research Network, Rochester, NY.

Klein, B. (2002). The role of incomplete contracts in self-enforcing relationships. In E. Brousseau and J.-M. Glachant (Eds.), *The Economics of Contracts: Theories and Applications*, pp. 59–71. Cambridge: Cambridge University Press.

Kogan, S., D. Levin, B. R. Routledge, J. S. Sagi, and N. A. Smith (2009). Predicting Risk from Financial Reports with Regression. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, Stroudsburg, PA, USA, pp. 272–280. Association for Computational Linguistics.

Kosnik, L.-R. (2014, March). Determinants of contractual completeness: An environmental regulatory application. *International Review of Law and Economics 37*, 198–208.

Livermore, M., A. B. Riddell, and D. Rockmore (2015, January). A topic model approach to studying agenda formation for the u.s. supreme a topic model approach to studying agenda formation for the u.s. supreme court. Law and Economics Research Paper 2015-2, University of Virginia School of Law.

Loughran, T. and B. McDonald (2015, January). Textual analysis in accounting and finance: A survey. Working Paper 2504147, Social Science Research Network.

Masten, S. E. and S. Saussier (2000). Econometrics of contracts : An assessment of developments in the empirical literature of contracting. *Revue d'Economie Industrielle 92*, 215–237.

McCallum, A., X. Want, and A. Corrada-Emmanuel (2007, September). Topic and role diversity in social networks with experiments on enron and academic email. *Journal of Artificial Intelligence Research 30*(1), 249–272.

Melumad, N., D. Mookherjee, and S. Reichelstein (1997, Summer). Contract complexity, incentives, and the value of delegation. *Journal of Economics and Management Strategy 6*(2), 257–289.

Modigliani, F. and M. Miller (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review 48*(3), 261–297.

Moszoro, M. W., P. T. Spiller, and S. Stolorz (2014, January). Rigidity of public contracts. SSRN Working Paper 2289719, Social Science Research Network.

Nini, G., D. C. Smith, and A. Sufi (2009, June). Creditor control rights and firm investment policy. *Journal of Financial Economics 92*(3), 400–420.

Poppo, L. and T. Zenger (2002). Do formal contracts and relational governance function as substitutes or complements? *Strategic Management Journal 23*, 707–725.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program 14*(3), 130–137.

Roberts, M. R. (2015). The role of dynamic renegotiation and asymmetric information in financial contracting. *Journal of Financial Economics 116*(1), 61 – 81.

Roberts, M. R. and A. Sufi (2009, August). Renegotiation of financial contracts: Evidence from private credit agreements. *Journal of Financial Economics 93*(2), 159–184.

Salton, G. (1971). *The SMART Retrieval System: Experiments in Automatic Document Processing*. Upper Saddle River, NJ, USA: Prentice-Hall.

Saussier, S. (2000). Transaction costs and contractual incompleteness: the transaction costs and contractual incompleteness: The case of électricité de francé. *Journal of Economic Behavior and Organization 42*, 189–206.

Schwartz, A. and J. Watson (2004). The law and economics of costly contracting. *Journal of Law, Economics, and Organization 20*(1), 2–31.

Sievert, C. and K. E. Shirley (2014). Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaxces*, pp. 63–70.

Steyvers, M. and T. L. Griffiths (2011). Probabilistic topic models. In T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*, Chapter 21, pp. 437–448. New York: Routledge.

Taddy, M. A. (2012). On estimation for topic models. In N. Lawrence and M. Girolami (Eds.), *Journal of Machine Learning Research: Workshop and Conference Proceedings*, Volume 22, pp. 1184–1193. http://jmlr.csail.mit.edu/proceedings/papers/v22/taddy12/taddy12.pdf.

Tirole, J. (2009). Cognition and incomplete contracts. *American Economic Review 99*(1), 265–294.

Wei, X. and W. B. Croft (2006). Lda-based document models for ad-hoc retrieval. In *SIGIR 2006 Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 178–185.

Williamson, O. E. (1985). *The Economic Institutions of Capitalism*. New York: Free Press.

Williamson, O. E. (1989). Transaction cost economics. In R. Schmalensee and R. Willig (Eds.), *Handbook of Industrial Organization*, Volume 1, Chapter 3, pp. 135–182. North-Holland.

Tables

Table 1: Summary Statistics - Contracts

This table presents summary statistics for the document collections of *full sample* (both Sections and Clauses) and *renegotiation sample* (both Sections and Clauses). Terms and phrases represents the total number of terms and phrases from our dictionary used in the documents. Unique terms and phrases represents the number of different terms and phrases used in the document. Terms and phrases (finance) and Terms and phrases (law) represent the total number of terms and phrases broken down by finance and law terms. Clauses represents the number of clauses in an events of default section, and Average length of clauses represents the average number of terms and phrases used to describe these clauses. Main topics represents the number of distinct main topics the individual clauses of a section exhibit. Average word distance represents the average pairwise cosine distance based and counts of terms and phrases, Average topic distance represents the average cosine distance based on counts of main topics.

	Mean	SD	Min	Median	Max
Sections					
Full Sample: F_S = 3438 sections with n = 3192 terms					
Terms and phrases	543.1	141.1	14	532	1348
Unique terms and phrases	218.5	40.7	11	217	362
Terms and phrases (finance)	224.0	60.6	7	219	559
Terms and phrases (law)	523.1	136.4	11	412	1302
Clauses	20.24	5.27	1	20	51
Average length of clauses	26.22	4.98	9	25.88	52.18
Main topics	15.01	3.15	1	15	28
Average word distance	.899	.053	.771	.909	.999
Average topic distance	.533	.048	.425	.525	.808
Renegotiation Sample: R_S = 250 sections with n = 1662 terms					
Terms and phrases	580.8	263.0	71	559.5	3628
Unique terms and phrases	226.0	53.2	45	224.5	616
Terms and phrases (finance)	239.2	117.5	29	226	1659
Terms and phrases (law)	560.1	252.6	65	537.5	3477
Clauses	21.68	9.0	4	20	116
Average length of clauses	26.05	4.67	13.57	26.14	38
Main topics	15.4	3.2	6	15	25
Average topic distance	.529	.047	.430	.521	.689
Clauses					
Full Sample: F_C = 69607 clauses with n = 3134 terms					
Terms and phrases	25.84	20.55	0	21	317

Continued on next page

Table 1 (continued from previous page)

	Mean	SD	Min	Median	Max
Unique terms and phrases	19.84	13.02	0	17	113
Terms and phrases (finance)	10.57	8.98	0	8	146
Terms and phrases (law)	24.94	19.86	0	20	276
Renegotiation Sample: R_C = 5421 clauses with n = 1640 terms					
Terms and phrases	25.86	20.21	1	21	201
Unique terms and phrases	19.90	12.88	1	17	110
Terms and phrases (finance)	7.86	5.37	0	7	40
Terms and phrases (law)	19.2	12.46	1	17	107

Table 2: Summary Statistics - Common Topics

This table presents summary statistics for the length of clauses that exhibit the ten most common topics. Main topics are ranked by their frequency (Freq.) of sections whose clauses exhibit a main topic at least once. Average 5 represents the weighted average of the five most common topics (for each section that holds at least one of the common topics) where the length of a clause is weighted by the median length of all clauses of a given common topic. Average five represents the weighted average of the five most common topics. Topic description provides a short description of the respective topic. A list of representative words and representative clauses for each of these topics is provided in Table 3.

Topic #	Topic description	Freq.	Mean	SD	Min	Median	Max
23	False representations and warranties	3394	27.33	10.76	7	26	175
20	Repayment	2350	23.69	9.88	5	22	122
29	Cross-default/acceleration	2342	53.72	28.06	5	51	262
26	Change in control	2084	13.48	17.38	2	3	150
22	Legal judgements against the firm	2055	33.20	13.80	5	30	118
	Average 5	3416	5.91	5.15	0.27	4.33	51.64

Table 3: Description of Common Topics

Topic description	Representative terms
False representations and warranties	made deem represent warranti statement certif connect prove furnish deliv
10th percentile: (b) Any representation or warranty made by or on behalf of the Borrower in any Loan Document or certificate or other writing delivered pursuant thereto shall prove to have been incorrect in any material respect when made or deemed made.	
90th percentile: (c) any representation or warranty made or deemed made by or on behalf of any Loan Party in or pursuant to any Loan Document or any amendment or modification thereof or waiver thereunder, or any material representation or warranty in any report, certificate, financial statement or other document furnished pursuant to or in connection with any Loan Document or any amendment or modification thereof or waiver thereunder, shall prove to have been incorrect in any material respect when made or deemed made.	
Repayment	interest fee due payabl busi.day busi princip amount pay fail
10th percentile: or failure by Borrower to pay any interest on any Loan or any fee or any other amount due under this Agreement within three Business Days after the date due.	
90th percentile: (a) Such Borrower shall fail to pay (i) any principal of any of the Advances when the same becomes due and payable, or (ii) any interest on any of the Advances, or any Facility Fee, other fee or other amount payable by it hereunder (including, in the case of CFSC, any amount payable under the CFSC Guaranty) by the later of (A) five (5) Business Days after such item has become due and (B) two (2) Business Days after receipt of written notice from the Agent that such item has become due.	
Cross-default/acceleration	indebted holder caus beneficiari prior matur state.matur requir redeem state
10th percentile: or any such Indebtedness shall become or be declared to be due and payable, or required to be prepaid (other than by a regularly scheduled required prepayment), or any Loan Party or any of its Subsidiaries shall be required to repurchase or offer to repurchase such Indebtedness, prior to the stated maturity thereof.	

Continued on next page

Table 3 (continued from previous page)

Topic description	Representative terms
90th percentile: (i) The Borrower or any Significant Subsidiary (A) fails to make any payment when due (whether by scheduled maturity, required prepayment, acceleration, demand, or otherwise) in respect of any Indebtedness or Guarantee (other than Indebtedness hereunder and Indebtedness under Swap Contracts) having an aggregate principal amount (including amounts owing to all creditors under any combined or syndicated credit arrangement) of more than $100,000,000, or (B) fails to observe or perform any other agreement or condition relating to any such Indebtedness or Guarantee or contained in any instrument or agreement evidencing, securing or relating thereto, or any other event occurs, the effect of which default or other event is to cause, or to permit the holder or holders of such Indebtedness or the beneficiary or beneficiaries of such Guarantee (or a trustee or agent on behalf of such holder or holders or beneficiary or beneficiaries) to cause, with the giving of notice if required, such Indebtedness to be demanded or to become due or to be repurchased, prepaid, defeased or redeemed (automatically or otherwise), or an offer to repurchase, prepay, defease or redeem such Indebtedness to be made, prior to its stated maturity, or such Guarantee to become payable or cash collateral in respect thereof to be demanded.	
Change in control	director chang control board major board.of.director constitut elect month individu
10th percentile: (p) there occurs a Change of Control.	
90th percentile: (b) during any 24-month period, individuals who at the beginning of such period constituted the Company's Board of Directors (together with any new directors whose election by the Company's Board of Directors or whose nomination for election by the Company's shareholders was approved by a vote of at least two-thirds of the directors who either were directors at beginning of such period or whose election or nomination was previously so approved) cease for any reason to constitute a majority of the Board of Directors of the Company.	
Legal judgements against the firm	judgment stay money order render enforc final appeal final.judgment consecut
10th percentile: Any judgment or order for the payment of money in excess of $50,000,000 shall be rendered against the Borrower and there shall be any period of 60 consecutive days during which a stay of enforcement of such judgment or order, by reason of a pending appeal or otherwise, shall not be in effect.	
90th percentile: (h) any judgment or order for the payment of money in excess of $500,000 shall be rendered against any Credit Party and either (i) enforcement proceedings shall have been commenced by any creditor upon such judgment or order or (ii) there shall be any period of twenty (20) consecutive days during which a stay of enforcement of such judgment or order, by reason of a pending appeal or otherwise, shall not be in effect.	

Table 4: Summary Statistics

This table presents summary statistics for the firm and loan level variables. Stock return volatility represents the standard deviation of daily stock returns for the prior year. Leverage ratio is *dlttq* + *dlcq* divided by *atq*. Tangible asset ratio *ppentq* divide by *atq*. Debt type concentration is the HHI of the firms different debt sources scaled by total debt. Segment concentration is the HHI of firm sales by Compustat business segment. Age is the number of years the firm exists in Compustat. Loan amount is the total loan amount in millions. Maturity is the average maturity of all loans in the package. Debt/ Loan Amt is the ratio of *dlcq* + *dlttq* prior to the loan divided by the total loan amount. Repeat lender is a dummy taking on a value of 1 if the firm has borrowed from the same bank in the previous 5 years. # of lenders is the number of unique lenders participating in the loan package. Lender <100 miles away takes on a value of 1 if the headquarters of at least one of the lead lenders is within 100 miles of the headquarters of the borrowing firm. # of Covenants is the total number of covenants in each loan as gathered by Nini et al. (2009).

	Mean	SD	Min	Median	Max
log(Loan Amount)	5.33	1.44	-1.97	5.4	10.1
log(Maturity)	3.62	.633	.693	3.74	4.82
# of Covenants	2.55	1.23	0	3	10
Tangible Asset Ratio	.351	.245	.0187	.285	.914
log(Age)	2.7	.909	0	2.71	4.01
Repeat lender (last 5 years)	.438	.496	0	0	1
Multiple Lenders	.847	.36	0	1	1
Lender <100 miles away	.163	.37	0	0	1
Segment Concentration	.776	.266	0	1	1
Foreign Operations	.677	.468	0	1	1
Debt Concentration	.81	.211	.254	.912	1
Leverage Ratio	.298	.176	0	.29	.965
Stock Return Vol	.0314	.0161	.00961	.0273	.0944
Expected Default Frequency	.0576	.18	0	8.37e-08	.999
log(Assets)	6.78	1.71	1.45	6.7	12.4
Debt/Loan Amt	2.26	3.9	0	.986	25.9
Repeat lender (last 5 years)	.438	.496	0	0	1
# of Covenants	2.55	1.23	0	3	10

Table 5: Correlation Coefficients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) log(Loan Amount)	1															
(2) log(Maturity)	0.138*	1														
(3) # of Covenants	-0.216*	0.232*	1													
(4) Tangible Asset Ratio	0.0863*	0.00710	-0.109*	1												
(5) log(Age)	0.312*	-0.101*	-0.274*	0.0368	1											
(6) Repeat lender (last 5 years)	0.303*	-0.00973	-0.0733*	0.0588*	0.0988*	1										
(7) Multiple Lenders	0.563*	0.174*	-0.0112	0.0720*	0.172*	0.237*	1									
(8) Lender <100 miles away	-0.0400	-0.0709*	-0.0250	-0.126*	0.0140	0.00578	-0.0607*	1								
(9) Segment Concentration	-0.194*	0.113*	0.141*	0.0875*	-0.344*	-0.0814*	-0.130*	-0.0231	1							
(10) Foreign Operations	0.0181	-0.0178	0.0140	-0.205*	0.0140	0.0184	-0.0150	0.106*	-0.0845*	1						
(11) Debt Concentration	0.0375	-0.154*	-0.145*	-0.00885	0.172*	0.0178	0.0556*	0.00884	-0.0966*	0.0321	1					
(12) Leverage Ratio	0.223*	0.00985	-0.0180	0.249*	0.00582	0.115*	0.0966*	-0.0500*	-0.0135	-0.122*	-0.207*	1				
(13) Stock Return Vol	-0.459*	-0.133*	0.167*	-0.128*	-0.372*	-0.190*	-0.364*	0.0361	0.169*	-0.0219	-0.140*	0.0405	1			
(14) Expected Default Frequency	-0.110*	-0.159*	-0.0455	-0.0138	-0.0787*	-0.0365	-0.132*	0.00505	0.000538	-0.0170	-0.0709*	0.258*	0.492*	1		
(15) log(Assets)	0.803*	-0.131*	-0.372*	0.124*	0.472*	0.299*	0.446*	-0.0249	-0.281*	0.0199	0.103*	0.167*	-0.471*	-0.0676*	1	
(16) Debt/Loan Amt	-0.0125	-0.319*	-0.205*	0.169*	0.260*	0.0671*	-0.0533*	-0.00229	-0.148*	-0.0585*	0.0490	0.321*	-0.0872*	0.133*	0.416*	1

* $p < .01$

Table 6: Contract Pair Topic Distance

This table presents the determinants of the cosine vector distance of each contract pair in terms of the use of main topics or words. The sample contains every unique contract-to-contract pair, where the dependent variable is the cosine distance between the two contracts in the pair. The covariates ">100M" and "Multiple Lenders" indicate that both loans in the contract pair satisfy the criteria or neither. Standard errors are clustered at the contract level.

Panel A: Main Topic Cosine

	(1) Topic Distance	(2) Topic Distance	(3) Topic Distance	(4) Topic Distance	(5) Topic Distance
Same Borrower	-0.213***	-0.227***	-0.215***	-0.158***	-0.175***
	(-36.52)	(-36.57)	(-36.26)	(-10.31)	(-7.97)
Same Lead Lender	-0.0386***	-0.0254***	-0.0346***	-0.0746***	-0.108***
	(-26.07)	(-17.79)	(-22.64)	(-16.67)	(-12.84)
Same Year	-0.00147***	-0.00217***	-0.00126***	-0.00222***	-0.00126***
	(-3.34)	(-4.39)	(-2.81)	(-4.50)	(-2.82)
Same SIC	-0.00997***	-0.0129***	-0.0114***	-0.0130***	-0.0114***
	(-5.45)	(-6.57)	(-5.77)	(-6.61)	(-5.77)
>100M		-0.0314***		-0.0328***	
		(-15.91)		(-16.57)	
>100M × Same Borrower				-0.0782***	
				(-4.77)	
>100M × Same Lead Lender				0.0524***	
				(11.12)	
Multiple Lenders			-0.0376***		-0.0390***
			(-20.49)		(-21.12)
Multiple Lenders × Same Borrower					-0.0411*
					(-1.83)
Multiple Lenders × Same Lead Lender					0.0744***
					(9.08)
Constant	0.536***	0.554***	0.569***	0.555***	0.570***
	(574.53)	(343.37)	(278.87)	(342.98)	(277.51)
N	5502903	2837641	4663396	2837641	4663396
R^2	0.00741	0.0174	0.0107	0.0182	0.0110

Panel B: Word Cosine Distances

	(1) Word Distance	(2) Word Distance	(3) Word Distance	(4) Word Distance	(5) Word Distance
Same Borrower	-0.460***	-0.493***	-0.470***	-0.326***	-0.344***
	(-40.40)	(-40.89)	(-40.49)	(-10.56)	(-7.20)
Same Lead Lender	-0.0469***	-0.0383***	-0.0447***	-0.100***	-0.168***
	(-20.78)	(-16.60)	(-19.73)	(-14.03)	(-9.65)
Same Year	-0.00306***	-0.00224***	-0.00223***	-0.00231***	-0.00223***
	(-4.96)	(-2.97)	(-3.39)	(-3.07)	(-3.40)
Same SIC	-0.0133***	-0.0149***	-0.0145***	-0.0150***	-0.0145***
	(-6.21)	(-5.29)	(-6.33)	(-5.32)	(-6.32)
>100M		-0.0129***		-0.0146***	
		(-5.92)		(-6.70)	
>100M × Same Borrower				-0.188***	
				(-5.68)	
>100M × Same Lead Lender				0.0659***	
				(8.75)	
Multiple Lenders			-0.00334		-0.00551***
			(-1.56)		(-2.60)
Multiple Lenders × Same Borrower					-0.130***
					(-2.66)
Multiple Lenders × Same Lead Lender					0.125***
					(7.22)
Constant	0.902***	0.906***	0.903***	0.907***	0.905***
	(846.73)	(609.88)	(434.81)	(604.67)	(439.87)
N	5502903	2837641	4663396	2837641	4663396
R^2	0.00716	0.00919	0.00727	0.00977	0.00758

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Table 7: Determinants of Default Section Complexity

This table presents the determinants of total words, unique words, law related terms, financial "jargon" related terms, clauses, and main topics in the events of default sections of each contract. Total words represents the total number of non-"stop words" in each section. Unique words represents the counts each unique word only once. Law and financial words represent terms from the specified law and finance dictionaries. Clauses represent the number of discrete clauses in the contract. Main topics represents the number of unique main topics identified from a 50 topic LDA topic model. All models also contain year fixed effects. Standard errors are clustered at the firm level.

	(1) Total Words	(2) Unique Words	(3) Law	(4) Finance	(5) Clauses	(6) Main Topics
log(Loan Amount)	9.565**	3.180**	9.210**	3.957*	0.221	0.222*
	(2.12)	(2.41)	(1.98)	(1.86)	(1.17)	(1.92)
log(Maturity)	17.20***	5.402***	18.68***	7.698***	0.348*	0.381***
	(3.41)	(3.61)	(3.61)	(3.37)	(1.72)	(3.15)
Repeat lender (last 5 years)	-2.289	-0.156	-2.736	1.028	0.154	-0.0406
	(-0.44)	(-0.10)	(-0.51)	(0.44)	(0.74)	(-0.33)
Multiple Lenders	39.61***	9.890***	39.51***	16.44***	0.612*	0.656***
	(4.40)	(3.68)	(4.26)	(4.04)	(1.66)	(2.96)
Lender <100 miles away	-17.29**	-4.298*	-17.23**	-7.430**	-0.869***	-0.513***
	(-2.27)	(-1.91)	(-2.18)	(-2.09)	(-3.02)	(-2.79)
Tangible Asset Ratio	-43.55***	-12.59***	-43.28***	-17.41***	-1.473***	-1.402***
	(-3.03)	(-3.03)	(-2.92)	(-2.64)	(-2.66)	(-4.37)
Segment Concentration	1.456	-0.400	0.530	1.998	0.580	0.186
	(0.12)	(-0.11)	(0.04)	(0.36)	(1.30)	(0.68)
Foreign Operations	-11.07	-1.935	-10.57	-2.452	-0.459*	-0.188
	(-1.59)	(-1.00)	(-1.47)	(-0.75)	(-1.67)	(-1.16)
Debt Concentration	-49.11***	-13.59***	-51.14***	-22.87***	-0.872	-0.687**
	(-3.71)	(-3.54)	(-3.76)	(-3.66)	(-1.57)	(-2.08)
log(Assets)	-5.511	-3.454**	-6.401	-2.506	-0.405**	-0.234**
	(-1.21)	(-2.56)	(-1.37)	(-1.19)	(-2.19)	(-2.08)
log(Age)	-10.09**	-2.904**	-10.59**	-4.152**	-0.336**	-0.228**
	(-2.48)	(-2.48)	(-2.53)	(-2.18)	(-2.18)	(-2.46)
Leverage Ratio	33.26*	7.310	35.74*	15.85*	1.661**	1.293***
	(1.66)	(1.28)	(1.72)	(1.69)	(2.16)	(2.95)
Stock Return Vol	563.4**	132.6*	601.0**	185.8	4.723	-2.926
	(2.25)	(1.86)	(2.34)	(1.62)	(0.47)	(-0.48)
Expected Default Frequency	44.61**	11.20**	44.22**	21.98***	2.090***	0.791*
	(2.42)	(2.22)	(2.30)	(2.59)	(2.90)	(1.90)
# of Covenants	13.45***	3.679***	13.81***	5.697***	0.556***	0.326***
	(4.80)	(4.45)	(4.82)	(4.38)	(4.74)	(4.76)
Debt/Loan Amt	0.170	0.195	0.0667	0.120	0.0367	0.0294
	(0.15)	(0.58)	(0.06)	(0.23)	(0.89)	(1.14)
Constant	425.0***	195.5***	437.4***	185.2***	19.33***	13.63***
	(12.04)	(18.45)	(12.08)	(11.62)	(12.82)	(15.22)
N	2756	2756	2756	2756	2756	2756
R^2	0.115	0.116	0.115	0.0991	0.0902	0.0963

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Table 8: Average Topic and Word Distance

This table presents the determinants of the average topic and word distance for each contract. The dependent variable in each case is the average cosine distance between the contract and all other contracts in the main sample. For topic distance this is the vector of main topics from the 50 topic LDA model. For word distance, it is the vector of words in each contract.

	(1) Avg Topic Dist	(2) Avg Word Dist
log(Loan Amount)	0.000999	0.00140
	(0.68)	(0.69)
log(Maturity)	-0.000554	-0.000765
	(-0.40)	(-0.36)
Repeat lender (last 5 years)	0.00244*	0.000125
	(1.65)	(0.05)
Multiple Lenders	-0.0173***	-0.00375
	(-6.36)	(-1.06)
Lender <100 miles away	0.00240	0.000376
	(1.10)	(0.11)
Tangible Asset Ratio	-0.00372	0.00804
	(-1.00)	(1.32)
Segment Concentration	0.00362	-0.00231
	(1.11)	(-0.41)
Foreign Operations	-0.00471**	-0.0131***
	(-2.40)	(-4.59)
Debt Concentration	-0.00692*	-0.00622
	(-1.74)	(-1.04)
log(Assets)	-0.00290**	-0.00360*
	(-2.05)	(-1.80)
log(Age)	0.00147	-0.00366**
	(1.33)	(-1.99)
Leverage Ratio	-0.00346	-0.0198**
	(-0.62)	(-2.35)
Stock Return Vol	0.162**	0.0458
	(2.22)	(0.46)
Expected Default Frequency	0.0124**	0.0184***
	(2.24)	(2.86)
# of Covenants	-0.00338***	0.00133
	(-4.48)	(1.20)
Debt/Loan Amt	0.000137	0.00137***
	(0.39)	(2.67)
Constant	0.682***	0.983***
	(60.93)	(58.85)
Topic, word, and year controls	Y	Y
N	2756	2749
R^2	0.379	0.0667

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Table 9: Determinants of Individual Topic Detail

This table presents the determinants of the $log(\frac{\text{total words}}{\text{median total words}})$ in each contract in the five most commonly identified main topics. Details on each of the identified topics, along with sample clauses are in Table 2. Standard errors are clustered at the firm level.

	(1) Topic 23	(2) Topic 20	(3) Topic 29	(4) Topic 26	(5) Topic 22	(6) Top 5 Avg
log(Loan Amount)	0.00670	0.0130	-0.0186	0.0467	0.0192	0.0124
	(0.54)	(0.81)	(-0.69)	(0.84)	(1.01)	(0.91)
log(Maturity)	0.0261**	0.0275*	0.0461	-0.0369	0.00398	0.0277**
	(2.01)	(1.79)	(1.60)	(-0.66)	(0.20)	(2.32)
Repeat lender (last 5 years)	-0.0176	-0.00607	-0.0619**	0.0717	0.00723	-0.0177
	(-1.15)	(-0.33)	(-2.32)	(1.23)	(0.32)	(-1.25)
Multiple Lenders	0.0457*	0.0312	0.159***	0.0302	0.0281	0.0786***
	(1.91)	(0.92)	(2.93)	(0.30)	(0.70)	(3.19)
Lender <100 miles away	0.0333	-0.0467	-0.0995**	0.0442	-0.0443	-0.0412**
	(1.48)	(-1.58)	(-2.28)	(0.48)	(-1.26)	(-1.97)
Tangible Asset Ratio	0.0835**	0.0138	-0.0136	-0.184	-0.0326	-0.0134
	(2.29)	(0.31)	(-0.20)	(-1.23)	(-0.53)	(-0.39)
Segment Concentration	-0.0317	-0.00459	0.0828	-0.136	0.00294	-0.0339
	(-0.90)	(-0.12)	(1.28)	(-0.96)	(0.06)	(-1.07)
Foreign Operations	-0.0138	0.0248	-0.0420	-0.0677	0.00688	-0.0137
	(-0.70)	(1.08)	(-1.14)	(-0.90)	(0.22)	(-0.72)
Debt Concentration	0.0103	-0.0627	-0.0893	-0.00340	-0.0682	-0.00718
	(0.27)	(-1.32)	(-1.16)	(-0.02)	(-1.17)	(-0.21)
log(Assets)	-0.00920	-0.00844	-0.0267	-0.00944	0.00281	-0.00930
	(-0.78)	(-0.55)	(-1.02)	(-0.17)	(0.15)	(-0.67)
log(Age)	-0.00554	-0.0195	-0.0213	0.120**	-0.0416**	-0.00407
	(-0.46)	(-1.30)	(-1.02)	(2.49)	(-2.35)	(-0.38)
Leverage Ratio	0.0594	-0.0773	0.159	-0.0213	0.0840	0.0625
	(1.18)	(-1.19)	(1.49)	(-0.09)	(0.95)	(1.14)
Stock Return Vol	1.348**	0.974	-2.335*	-2.233	0.0238	-0.368
	(2.03)	(1.09)	(-1.67)	(-0.78)	(0.02)	(-0.60)
Expected Default Frequency	-0.0345	0.00637	-0.216**	0.0407	0.0636	-0.0288
	(-0.79)	(0.09)	(-2.08)	(0.20)	(0.71)	(-0.61)
# of Covenants	0.0181***	-0.00612	0.0139	-0.0104	0.00594	0.00586
	(2.80)	(-0.64)	(0.88)	(-0.34)	(0.54)	(0.83)
Debt/Loan Amt	-0.00227	-0.00249	-0.0175**	0.00688	0.00171	-0.00419
	(-0.66)	(-0.70)	(-2.10)	(0.35)	(0.36)	(-0.79)
Constant	-0.244**	-0.102	-0.0755	1.075**	-0.0688	-0.0148
	(-2.40)	(-0.77)	(-0.37)	(2.51)	(-0.46)	(-0.15)
N	2730	1924	1902	1659	1665	2751
R^2	0.0292	0.0477	0.0841	0.0552	0.0606	0.0352

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Table 10: Predicting the Number of Renegotiation Rounds

This table presents the results of a negative binomial model on the number of future amendments made to the original contract as a function of the complexity of the original contract. Standard errors are clustered at the contract level.

	(1)	(2)	(3)	(4)
Total Words (1000s)	1.108			
	(0.33)			
Unique Words (1000s)		8.825		
		(1.55)		
Main Topics			1.051***	
			(2.98)	
Average Topic Distance				0.0327**
				(-2.54)
Maturity	1.806***	1.765***	1.746***	1.761***
	(4.98)	(4.72)	(4.78)	(4.67)
Loan Yield	2.771**	2.459**	2.779**	2.815**
	(2.43)	(2.12)	(2.54)	(2.51)
Loan-to-Assets	1.082***	1.078***	1.087***	1.079***
	(3.68)	(3.70)	(4.32)	(3.80)
Constant	0.290*	0.207**	0.167***	2.577
	(-1.94)	(-2.32)	(-2.73)	(0.89)
$log(\alpha)$	0.604***	0.595***	0.578***	0.582***
	(-3.20)	(-3.23)	(-3.34)	(-3.40)
Observations	340	340	340	340
Year FE	Yes	Yes	Yes	Yes
Loan Type FE	Yes	Yes	Yes	Yes
Terminal Event FE	Yes	Yes	Yes	Yes

Exponentiated coefficients; t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 11: Contract Complexity and Future Returns

This table presents the determinants of future changes in return on assets over the four quarters following the initiation of the loan relative to the prior four quarters. Change in return on assets is defined as the sum of four quarters of Compustat item *ibq* divided by beginning of period *atq* minus the quantity over the prior four quarters. All specifications include additional risk controls: log(assets), tangible asset ratio, prior years sales growth, loan amount, repeat lender, and number of lenders, as well as fixed effects for each year and 3 digit SIC code. Standard errors are clustered at the firm level.

	Δ ROA	Δ ROA	Δ ROA	Δ ROA
Total Words (1000s)	0.0495***			
	(2.63)			
Unique Words (1000s)		0.139**		
		(2.08)		
Main Topics			0.00142*	0.00227**
			(1.81)	(2.52)
Average Topic Distance				0.105*
				(1.78)
Leverage Ratio	0.0580***	0.0588***	0.0580***	0.0578***
	(3.31)	(3.36)	(3.31)	(3.30)
Stock Return Vol	0.437*	0.451*	0.464*	0.428*
	(1.71)	(1.77)	(1.82)	(1.67)
Junk	-0.00362	-0.00279	-0.00234	-0.00266
	(-0.53)	(-0.41)	(-0.34)	(-0.39)
Rated	0.00365	0.00322	0.00358	0.00394
	(0.52)	(0.46)	(0.51)	(0.56)
# of Covenants	-0.00519**	-0.00502**	-0.00493**	-0.00454**
	(-2.40)	(-2.33)	(-2.28)	(-2.11)
Observations	2806	2806	2806	2806
R^2	0.119	0.118	0.118	0.119
Addl Risk Controls				

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Table 12: Contract Complexity and Future Returns

This table presents the determinants of sales growth over the four quarters following the initiation of the loan relative to the prior four quarters. Sales growth is defined as the sum of item *saleq* over four quarters divided by the previous four quarters sales minus 1. All specifications include additional risk controls: log(assets), tangible asset ratio, prior years sales growth, loan amount, repeat lender, and number of lenders, as well as fixed effects for each year and 3 digit SIC code. Standard errors are clustered at the firm level.

	Sales Growth	Sales Growth	Sales Growth	Sales Growth
Total Words (1000s)	0.0959			
	(1.41)			
Unique Words (1000s)		0.248		
		(1.11)		
Main Topics			0.00575**	0.00549*
			(2.24)	(1.81)
Average Topic Distance				-0.0320
				(-0.17)
Leverage Ratio	-0.0589	-0.0574	-0.0609	-0.0609
	(-1.08)	(-1.05)	(-1.12)	(-1.12)
Stock Return Vol	-1.535**	-1.504**	-1.478**	-1.467**
	(-2.23)	(-2.19)	(-2.17)	(-2.14)
Junk	-0.0458*	-0.0439	-0.0476*	-0.0475*
	(-1.71)	(-1.63)	(-1.75)	(-1.75)
Rated	0.0524**	0.0514**	0.0545**	0.0544**
	(2.08)	(2.04)	(2.18)	(2.17)
# of Covenants	0.0139*	0.0144**	0.0131*	0.0129*
	(1.90)	(1.98)	(1.81)	(1.81)
Observations	2821	2821	2821	2821
R^2	0.267	0.267	0.268	0.268
Addl Risk Controls				

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Table A1: A Sample Contract f (Full Section)

SECTION 7.01. Events of Default. If any of the following events ("Events of Default") shall occur: (a) any Borrower shall fail to pay any principal of any Loan when and as the same shall become due and payable, whether at the due date thereof or at a date fixed for prepayment thereof or otherwise; (b) any Borrower shall fail to pay any interest on any Loan or any fee or any other amount (other than an amount referred to in clause (a) of this Article) payable under this Agreement, when and as the same shall become due and payable, and such failure shall continue unremedied for a period of five days; (c) any representation or warranty made or deemed made by or on behalf of the Parent or any Subsidiary in or in connection with this Agreement, the Guarantee, any Additional Borrower Agreement or any amendment or modification hereof or thereof, or in any report, certificate, financial statement or other document furnished pursuant to or in connection with this Agreement, the Guarantee, any Additional Borrower Agreement or any amendment or modification hereof or thereof, shall prove to have been incorrect in any material respect when made or deemed made; (d) any Borrower shall fail to observe or perform any covenant, condition or agreement contained in Section 5.02(a), 5.03 (with respect to the Parent's existence), 5.08, 5.09(e), 5.09(f), 5.09(i), 5.09(j) or in Article VI; (e) any Loan Party shall fail to observe or perform any covenant, condition or agreement contained in this Agreement (other than those specified in clause (a), (b) or (d) of this Article) or the Guarantee, and such failure shall continue unremedied for a period of 30 days after the earlier to occur of (i) the date on which a Financial Officer shall have discovered such default and (ii) the date on which written notice thereof has been given to the Parent by the Administrative Agent (at the request of any Lender); (f) the Parent or any Subsidiary shall fail to make any payment (whether of principal or interest and regardless of amount) in respect of any Material Obligations, when and as the same shall become due and payable beyond the applicable grace period therefor; (g) any event or condition occurs that results in any Material Obligations (other than Project Finance Indebtedness which is not guaranteed by the Parent or any Subsidiary (other than a Project Finance Company)) becoming due prior to its scheduled maturity or that enables or permits (with or without the giving of notice, the lapse of time or both) the holder or holders of such Material Obligations or any trustee or agent on its or their behalf to cause all of such Material Obligations to become due, or to require the prepayment, repurchase, redemption or defeasance thereof, prior to its scheduled maturity (unless waived); provided that this clause (g) shall not apply to secured Indebtedness that becomes due as a result of the voluntary sale or transfer of the property or assets securing such Indebtedness; (h) an involuntary proceeding shall be commenced or an involuntary petition shall be filed seeking (i) liquidation, reorganization or other relief in respect of the Parent or any Material Subsidiary or any Additional Borrower or Bidco or its debts, or of a substantial part of its assets, under any Federal, state or foreign bankruptcy, insolvency, receivership or similar law now or hereafter in effect or (ii) the appointment of a receiver, trustee, custodian, sequestrator, conservator or similar official for the Parent or any Material Subsidiary or any Additional Borrower or Bidco or for a substantial part of its assets, and, in any such case, such proceeding or petition shall continue undismissed for 60 days or an order or decree approving or ordering any of the foregoing shall be entered; (i) the Parent or any Material Subsidiary or any Additional Borrower or Bidco shall (i) voluntarily commence any proceeding or file any petition seeking liquidation, reorganization or other relief under any Federal, state or foreign bankruptcy, insolvency, receivership or similar law now or hereafter in effect, (ii) consent to the institution of, or fail to contest in a timely and appropriate manner, any proceeding or petition described in clause (h) of this Article, (iii) apply for or consent to the appointment of a receiver, trustee, custodian, sequestrator, conservator or similar official for the Parent or any Material Subsidiary or any Additional Borrower or Bidco or for a substantial part of its assets, (iv) file an answer admitting the material allegations of a petition filed against it in any such proceeding, (v) make a general assignment for the benefit of creditors or (vi) take any action to authorize any of the foregoing; (j) the Parent or any Material Subsidiary or any Additional Borrower or Bidco shall become unable, admit in writing or fail generally to pay its debts as they become due, including in respect of any Subsidiary organized under the laws of the United Kingdom for the purposes of Section 123 of the Insolvency Act 1986 (other than Section 123(1)(a), (b), (c) and (d), provided that, for purposes of this paragraph, the words "to the satisfaction of the court" shall be deemed to be omitted from Section 123(1)(e) and Section 123(2)); (k) one or more judgments for the payment of money in an aggregate amount in excess of $50,000,000 shall be rendered against the Parent, any Material Subsidiary, any Additional Borrower or any combination thereof and the same shall remain undischarged for a period of 30 consecutive days during which execution shall not be effectively stayed, or any action shall be legally taken by a judgment creditor to attach or levy upon any assets of the Parent or any Material Subsidiary to enforce any such judgment; (l) an ERISA Event shall have occurred that, in the opinion of the Required Lenders, when taken together with all other ERISA Events that have occurred, could reasonably be expected to result in a Material Adverse Effect; (m) a Change in Control shall occur; (n) the guarantee contained in Section 2 of the Guarantee shall cease, for any reason, to be in full force and effect in accordance with its terms or any Loan Party or any Affiliate of any Loan Party shall so assert; or (o) the Acquisition Agreement ceases to be in full effect in all material respects prior to the completion of the Asset Divisions;

Table A2: A Sample Contract f (Individual Sentences $s = 1, \ldots, f_s$)

If any of the following events, ("Events of Default") shall occur:
(a) any Borrower shall fail to pay any principal of any Loan when and as the same shall become due and payable, whether at the due date thereof or at a date fixed for prepayment thereof or otherwise;
(b) any Borrower shall fail to pay any interest on any Loan or any fee or any other amount (other than an amount referred to in clause (a) of this Article) payable under this Agreement, when and as the same shall become due and payable, and such failure shall continue unremedied for a period of five days;
(c) any representation or warranty made or deemed made by or on behalf of the Parent or any Subsidiary in or in connection with this Agreement, the Guarantee, any Additional Borrower Agreement or any amendment or modification hereof or thereof, or in any report, certificate, financial statement or other document furnished pursuant to or in connection with this Agreement, the Guarantee, any Additional Borrower Agreement or any amendment or modification hereof or thereof, shall prove to have been incorrect in any material respect when made or deemed made;
(d) any Borrower shall fail to observe or perform any covenant, condition or agreement contained in Section 5.02(a), 5.03 (with respect to the Parent's existence), 5.08, 5.09(e), 5.09(f), 5.09(i), 5.09(j) or in Article VI;
(e) any Loan Party shall fail to observe or perform any covenant, condition or agreement contained in this Agreement (other than those specified in clause (a), (b) or (d) of this Article) or the Guarantee, and such failure shall continue unremedied for a period of 30 days after the earlier to occur, of (i) the date on which a Financial Officer shall have discovered such default and (ii) the date on which written notice thereof has been given to the Parent by the Administrative Agent (at the request of any Lender);
(f) the Parent or any Subsidiary shall fail to make any payment (whether of principal or interest and regardless of amount) in respect of any Material Obligations, when and as the same shall become due and payable beyond the applicable grace period therefor;
(g) any event or condition occurs that results in any Material Obligations (other than Project Finance Indebtedness which is not guaranteed by the Parent or any Subsidiary (other than a Project Finance Company)) becoming due prior to its scheduled maturity or that enables or permits (with or without the giving of notice, the lapse of time or both) the holder or holders of such Material Obligations or any trustee or agent on its or their behalf to cause all of such Material Obligations to become due, or to require the prepayment, repurchase, redemption or defeasance thereof, prior to its scheduled maturity (unless waived);
provided that this clause (g) shall not apply to secured Indebtedness that becomes due as a result of the voluntary sale or transfer of the property or assets securing such Indebtedness;
(h) an involuntary proceeding shall be commenced or an involuntary petition shall be filed seeking (i) liquidation, reorganization or other relief in respect of the Parent or any Material Subsidiary or any Additional Borrower or Bidco or its debts, or of a substantial part of its assets, under any Federal, state or foreign bankruptcy, insolvency, receivership or similar law now or hereafter in effect or (ii) the appointment of a receiver, trustee, custodian, sequestrator, conservator or similar official for the Parent or any Material Subsidiary or any Additional Borrower or Bidco or for a substantial part of its assets, and, in any such case, such proceeding or petition shall continue undismissed for 60 days or an order or decree approving or ordering any of the foregoing shall be entered;
(i) the Parent or any Material Subsidiary or any Additional Borrower or Bidco shall (i) voluntarily commence any proceeding or file any petition seeking liquidation, reorganization or other relief under any Federal, state or foreign bankruptcy, insolvency, receivership or similar law now or hereafter in effect, (ii) consent to the institution of, or fail to contest in a timely and appropriate manner, any proceeding or petition described in clause (h) of this Article, (iii) apply for or consent to the appointment of a receiver, trustee, custodian, sequestrator, conservator or similar official for the Parent or any Material Subsidiary or any Additional Borrower or Bidco or for a substantial part of its assets, (iv) file an answer admitting the material allegations of a petition filed against it in any such proceeding, (v) make a general assignment for the benefit of creditors or (vi) take any action to authorize any of the foregoing;
(j) the Parent or any Material Subsidiary or any Additional Borrower or Bidco shall become unable, admit in writing or fail generally to pay its debts as they become due, including in respect of any Subsidiary organized under the laws of the United Kingdom for the purposes of Section 123 of the Insolvency Act 1986 (other than Section 123(1)(a), (b), (c) and (d), provided that, for purposes of this paragraph, the words "to the satisfaction of the court" shall be deemed to be omitted from Section 123(1)(e) and Section 123(2));

(k) one or more judgments for the payment of money in an aggregate amount in excess of $50,000,000 shall be rendered against the Parent, any Material Subsidiary, any Additional Borrower or any combination thereof and the same shall remain undischarged for a period of 30 consecutive days during which execution shall not be effectively stayed, or any action shall be legally taken by a judgment creditor to attach or levy upon any assets of the Parent or any Material Subsidiary to enforce any such judgment;

(l) an ERISA Event shall have occurred that, in the opinion of the Required Lenders, when taken together with all other ERISA Events that have occurred, could reasonably be expected to result in a Material Adverse Effect;

(m) a Change in Control shall occur.

(n) the guarantee contained in Section 2 of the Guarantee shall cease, for any reason, to be in full force and effect in accordance with its terms or any Loan Party or any Affiliate of any Loan Party shall so assert;

or (o) the Acquisition Agreement ceases to be in full effect in all material respects prior to the completion of the Asset Divisions.