Measuring Contract Completeness: A Text Based Analysis of Loan Agreements

March 14, 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 fills 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. Larger syndicates and repeat lending relationships increase this impact, suggesting an important role for renegotiation costs. We also find evidence that more complex loan contracts are associated with increases in operating performance and cash holdings, 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; property rights theory; transaction costs economics

1 Introduction

The majority of modern theories in debt finance are founded on the basis of incomplete contracts. Contractual incompleteness is one of the fundamental benchmarks describing wide ranging theories of renegotiation, under/over-investment, and a host of other agency based models. In spite of the importance of this idea, empirical researchers have very few useful measures of contractual completeness when analyzing contracting theory. This paper works to fill that gap. First, using text based analysis of over 3000 debt contracts, we propose direct measures of the complexity and detail of debt contracts and the extent to which these contracts contain contingencies which enrich the contract state space. These metrics capture significant detail about the contract which is not captured by existing, coarser metrics such as a covenant count. Second, we provide insights as to the empirical determinants of contractual completeness and relate our results to the existing theoretical literature. Finally, we examine the impact of our measures of contractual completeness on forward looking firm outcomes.

We use our completeness metrics to test several hypotheses from the theoretical literature. Provided that greater contractual detail is costly (e.g., Battigalli and Maggi, 2002; Dye, 1985), we hypothesize that when the stake size in a given contract is larger, then benefits of greater contractual completeness should also be larger and counter the costs of greater completeness. Confirming this hypothesis, we find that contractual complexity is significantly positively related to measures of the overall stake size. Complexity is also higher for firms which are in greater danger of default, suggesting that debt contracts specify more detail when the probability of entering costly future default states is higher.

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. We also find some evidence that complexity is higher when the costs of renegotiation are higher.

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 is 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 Modiglianni and Miller (1958) propositions. Understanding the form and nature of perfect contracting has been critical to our understanding of the economic frictions that may or may not allow these economic models to hold, 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 being *more complete*, 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 we can measure contractual completeness by using metrics for amount of contractual detail to capture the complexity of the 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 three sets of metrics for the completeness of contracts (i.e., 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* to capture this property.[2] We further conjecture that a larger number of different or *unique words* used to describe the events of defaults—capturing the size or variety of vocabulary—is associated with both more distinct events of defaults and a more detailed description of these events of default. While these measures are obviously noisy, they are

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

[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.

simple to understand and provide a reasonable approximation of the level of detail. To address the worry that word counts may simply be characterizing the excessive use of words by a particular writer rather than actually capturing a larger number of ideas, we also use the number of *sentences* as a rough proxy for the number of clauses written into the default section.

Secondly, we use each default section to estimate a probabilistic topic model (Blei et al., 2003) to discover the general ideas or themes covered in default section—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 sentence 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.

Finally, we calculate the average word distance of each contract from the space of all other contracts. This is accomplished by creating a vector of word counts for each contract and calculating the cosine distance between each combination of word vectors. The average distance of each contract from the rest of the sample provides an intuitive measure of how much custom language is placed into the contract. Contracts which contain largely boiler-plate language will have a fairly low average cosine distance, whereas contracts with more

customized provisions will have a high average cosine distance.

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. Moreover, much of the language is "boilerplate", meaning that similar language is found is many other contracts and the language is regarded as somewhat of a formality across many debt contracts. However, some language is quite custom to the contracting parties, and our text based approach allows to quantify this idea in a large scale way.

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 prediction 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).

The remainder of the paper proceeds as follows. Section 2 lays out our main hypotheses in the context of existing contracting literature. Section 3 lays out our empirical methodology. Section 4 examines the empirical determinants of contractual complexity. Section 5 examines the impact of contract detail on future output. Section 6 concludes.

2 Theoretical Literature and Hypotheses

We review the theoretical literature behind contractual complexity and completeness from which we derive our hypotheses. We then discuss the implications for contractual similarity.

2.1 Contractual Detail and Completeness

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 uncertainty; and third, costs of ex-post renegotiation of the contract.

2.1.1 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. Similarly, in Bustos (2008) vague clauses are cheaper, and more precise clauses are more expensive to enforce. 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.

2.1.2 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.

2.1.3 Renegotiation Costs

Another factor that determines the degree of contractual completeness is the extent to which the contract can be renegotiated. On the one hand it is argued that higher costs of renegotiation (or more restrictive renegotiations) result in more complete contracts. The results in Kahn and Huberman (1988) or Aghion and Rey (1994), for instance, suggest that complex contracts are replicated by simple contracts when renegotiation is possible and the details

are specified ex post. Bajari and Tadelis (2001) draw a similar conclusion finding that fixed price contracts are more likely to be complete because they are more costly to renegotiate. The literature on ex post holdup, on the other hand, suggests that if contracts are renegotiated ex post, then ex ante contracts tend to be more complete to alleviate the problem of ex post opportunism.

In line with this, contract complexity should be positively related to proxies for the cost of renegotiation. The total cost of renegotiation is related to the overall likelihood of renegotiation. As demonstrated by Roberts and Sufi (2009), longer maturity contracts are mechanically more likely to require renegotiation. Consequently, we hypothesize that proxies for renegotiation costs will have a larger effect on contract complexity for longer maturity contracts. We consider two proxies for these costs. First, the number of syndicate partners, since the renegotiation costs tend to increase with the number of counter-parties. Second, we use a measure of product market fluidity, suggested by Hoberg et al. (2014). If the threat of entry in product markets is greater, the cost of holdup should also be greater, suggesting larger potential renegotiation costs.

3 Data and Methodology

3.1 Contract Data

For our analysis we use a set of private loan contracts collected from SEC filings from 1996 to 2005. Our initial sample comes from Nini et al. (2009), who extract the text of each contract from its associated filing in the EDGAR database based upon a manual search starting from the Dealscan loan database.[3] We focus on the sections that specify the events of defaults. Our sample consists of these sections from 3,444 contracts; 2,968 of these sections (86.18%) also include remedies, among other things, in addition to the events of default. We drop

[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.

these additional parts specifying the remedies and extract only the list of events of defaults.

We go through a number of steps of cleaning our sample. For our later analysis we want to be able to use both the full documents as well as individual sentences. We first split each document into its individual sentences (with '.', ':', ';', '!', and '?' as sentence delimiters; but a dot in, e.g., "Section 5.1" is not) and drop the resulting "sentences" that do not include a verb.[4]t The latter step eliminates section and subsection headers from our sample. To illustrate this step, in Table A1 we list a sample section of events of default and in Table A2 the list of the same section's individual sentences.

We obtain two distinct document collections: (1) A collection of n_F = 3,444 full sections, D_F. We refer to an individual document from this collection as *contract* and denote it by $f = 1, \ldots, 3{,}444$. (2) A collection of n_S = 69,324 individual sentences. We refer to an individual document from this collection as *sentence* and denote it by $s = 1, \ldots, 69{,}324$. Each contract (full section) comprises f_s sentences and represents the set of its sentences.

To prepare the documents for our textual analysis, we pre-process documents in the following steps:[5] convert all words to lower case, remove punctuation, and delete numbers, number words (1 through 100), roman numerals, and stopwords (such as "and," "the," or "that"). For stopwords we use the list provided by the SMART information retrieval system (Salton, 1971) containing 571 words. In a last step we stem the words, i.e., we erase word suffixes to obtain the words' radicals in order to reduce complexity, using the implementation of Porter's stemming algorithm (Porter, 1980).

After pre-processing we are left with a vocabulary of 3,999 words that occur a total of 1,692,595 times in our document collections. The descriptive statistics for the pre-processed document collections are provided in Table 1.

[4]We use the sentence tokenizer and part-of-speech tagger provided in the `openNLP` package (Hornik, 2014) for *R*.

[5]We use the `tm` package (Feinerer and Hornik, 2014) for *R* documented in Feinerer et al. (2008).

3.2 Measuring Contractual Detail and Completeness

We propose three measures for contractual detail: (i) simple word counts to capture length and richness of the contracts; (ii) sentence counts as a proxy for the number of clauses in the default sections of the contract; and (iii) probabilistic topic models to obtain a rough count of the number of distinct topics (or themes) that are covered.

3.2.1 Simple Words Counts

For our first measure of contract detail we use counts for words and phrases to capture how detailed the events of default are described. We use the number of *total words* in a default section as a measure of contract length, positing that longer default sections specify more clauses and contingencies. We also use the number of *unique words* in a default section to capture the size or variety or vocabulary. We conjecture that a larger number of different (or, unique) words used to describe the events of defaults is associated with both more distinct events of defaults and a more detailed description of these events of default. Panel (a) in Table 2 provides the summary statistics for our word based measures of contract detail.

3.2.2 Sentence Counts

Our word counts provide simple, easy to understand metrics of contract detail. However, there is a worry that these word counts may simply be characterizing the excessive use of words by a particular writer rather than actually capturing a larger number of ideas. To validate our results, we therefore also use the number of *sentences* in a default section. This represents a rough proxy for the number of clauses written in a contract and is meant to capture the detail of contract in terms of the breadth of the topics that are covered in the contract clauses. Panel (b) in Table 2 provides the summary statistics for the sentences metric.

3.2.3 Probabilistic Topic Models

Because multiple sentences may simply represent extraneous detail on a simple topic, we also count the number of "main topics" in a default section, as described by a probabilistic topic model. Probabilistic topic models 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). Probabilistic topic models uncover the latent topical structure of a document by analyzing the words used in the document.[6] The underlying idea is that authors of documents first decide which topics to cover before drafting the document. Documents therefore arise from (potentially) multiple topics where a topic is a distribution over a fixed vocabulary of words. If a document collection D is comprised of documents that cover K topics then each document $d \in D$ will exhibit these K topics with different proportions. Moreover, each word in the document is from a topic selected according to this distribution over topics.

The observed data are the words used in a document, however, the topics covered by the document are not observed. Probabilistic topic models are developed to reverse the process of topic generation (choose a topic as distribution over words from a distribution over topics; then choose a word according to that distribution) and automatically discover the latent topical structure. In other words, we infer the topical structure from observed words in our document collection.

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

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

et al., 2015). We add loan contracts to this list.

As a measure of contract detail we construct a topics count using the per-document distribution over topics for the document collection of sentences, D_S. When considering the sentence collection the structure of our sample, we assume that each individual sentence holds one event, or at most a small number of events. The topics count is constructed as follows. For each sentence $s \in D_S$ we find the main topic k_s, i.e., the topic with the highest topic proportion,

$$k_{s|f} = \arg\max_k \vec{\theta}_{s|f,k}. \tag{1}$$

We then determine the number of different *main topics* for each full section:

$$main\ topics_f = \left| \bigcup_{s=1}^{f_s} k_{s|f} \right|. \tag{2}$$

We use Taddy's implementation of LDA as part of the `maptpx` package (Taddy, 2012, 2013) for R which comes with a method for model selection. The optimal number of topics K^* is the one corresponding to the model with the highest Bayes factor. We limit our vocabulary to the 1,500 most frequent words/unigrams in the sample. For the individual sentences we estimate the model on 69,307 non-empty sentences. The optimal number of topics is $K_S^* = 55$.

In Table A3 we present some of the results for the topics from the sentences D_S. We cluster the topics (i.e., distributions of words), and for each topic we list the ten most relevant terms. Sievert and Shirley (2014) define *relevance* of a word g in topic k as

$$relevance_{g|k} = \lambda \log \beta_{k,g} + (1 - \lambda) \log \left(\frac{\beta_{k,g}}{\omega_g} \right) \tag{3}$$

where ω_g is the probability of a word g in the entire document collection. For $\lambda = 0$ this measure reduces to the *lift* of a term (Taddy, 2012), the ratio of a term's probability within a

topic ($\beta_{k,g}$) to its probability in the entire document collection (ω_g). A value of $\lambda = 1$ ranks the words according to their topic-specific probabilities $\beta_{k,g}$. Sievert and Shirley (2014) suggest a value of $\lambda = 3/5$. The topics are ranked according to their *usage*, i.e., average proportions in the document collection

$$usage_k = \frac{1}{n}\sum_{i=1}^{n}\theta_{i,k}. \tag{4}$$

Table 2 provides the summary statistics for our measures of contract detail: *total* word count, *unique* word count, a simple *sentence count*, and the number of *main topics*.

3.3 Measuring Contract Similarities

In addition to our measures of contractual detail, we also calculate the distance between documents. 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 the writing of a given contract is relative to the rest of the sample.

Each document f in D_F can be represented by a vector $\vec{\omega}_f$ consisting of the set of n-grams and the number of times each of these occurs. A widely used measure for the distance between two documents (i.e., the distance between their respective vectors $\vec{\omega}_i$ and $\vec{\omega}_j$) is the dissimilarity based on the cosine similarity. We denote this distance by

$$count_{cos} = \left(1 - \cos_{ij}\right) = 1 - \frac{\vec{\omega}_i \cdot \vec{\omega}_j}{||\vec{\omega}_i||\ ||\vec{\omega}_j||} = 1 - \frac{\sum_{g=1}^{V}\omega_{i,g}\omega_{j,g}}{\sum_{g=1}^{V}\left(\omega_{i,g}\right)^2\sum_{g=1}^{V}\left(\omega_{j,g}\right)^2} \tag{5}$$

with V the size of the vocabulary (the number of unique n-grams). We calculate the pairwise distance between all contracts.

3.4 Loan and Firm Specific Data

We match the loan contracts to firm level data in CRSP/Compustat. 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. We calculate Tobin's Q as the sum of the market value of equity plus the book value of short and long-term debt divided by the book value of assets. 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 HHI index of sales across all business segments in the Compustat segments file as of the most recent closing year.

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 it's 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 3. Correlation coefficients are reported in Table 4.

4 Determinants of Contractual Detail and Similarity

4.1 Total and Unique Word Counts

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 level measures of risk, uncertainty, and information in the quarter just prior to the initiation of each loan. We subsequently add loan level variables such as the amount and maturity of each loan and features of the lender detail and covenant structure of the loan. Each of our specifications also contains year fixed effects to account for possible time variation in the structure of loan contracts.

The results are presented in Table 5. Columns (1) and (3) present the determinants of the number of total and unique words in each default section as a function of firm level determinants. Columns (2) and (4) add loan level variables to each specification. Results are similar across each specification, with stock return volatility becoming statistically significant after the inclusion of loan level variables. As might be expected, higher firm leverage and higher stock return volatility 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, which implies that our measures are capturing additional information beyond covenant detail.

Consistent with our hypotheses regarding stake size, both word count measures are significantly positively related to loan amount and maturity. Word counts are also significantly related to leverage and volatility, suggesting that contracts are more detailed when firms are close to the default boundary. Age, asset tangibility, and our "close" lender indicator are all negatively related to word counts, suggesting that the dominant effect is for borrowers and lenders to attempt to contract away the costs of asymmetric information. The number of lenders is negatively related to word count however, suggesting that the overall effect of larger lending syndicates is to decrease complexity. This may suggest that contracts with

larger lending syndicates are less likely to be renegotiated for other reasons.

4.2 Sentences and Probabilistic Topic Models

Our word count results provide a simple, easy to understand test of contract detail. However, there is a worry that these word counts may simply be characterizing the excessive use words by a particular writer rather than actually capturing a larger number of ideas. To validate our results, we therefore turn to two more broad measures which attempt to capture the breadth of topics which are being captured in the contract clauses. The first tests involve the sentence count, which represents a rough proxy for the number of clauses written into a contract. Since multiple sentences may simply represent extreanous detail on a simple topic, we also use our count of the number of "main topics" which should give us a rough count of the number of distinct topics which are covered in the default section of the contract.

The results of these tests are reported in Table 6. For each column, we test the same model as before with the number of sentences and the number of main topics as dependent variables. To determine whether we are capturing anything additional, we also include the number of words in models (2) and (4). While the number of words is highly significant, the sentence and main topic counts appear to contain additional information beyond number of words.

The results are similar to our previous results using total and unique words, with a couple of exceptions. Segment concentration, i.e., the HHI of the sales fractions of each firm by Compustat segment, appears to be positively and significantly correlated with this measure of contract complexity. This is somewhat unexpected as firms with a larger number of segments (a lower HHI) are likely to be more diverse and have more complex operations. Such firms may have diversification benefits which make bankruptcy events less likely, but the failure to specify greater detail in default means that there must be some set of costs to specifying greater contractual detail.

Surprisingly, the impact of stock return volatility becomes negative in both cases after

controlling for the number of words. This seems to imply that firms with higher stock return volatility may enter contracts with a larger number of sentences with fewer words, though the exact meaning is somewhat unclear.

4.3 Contract Language Distance and Boilerplate

We now turn our attention to our measures of contract similarities. As mentioned in Section 3, we calculate the similarity between two documents as the cosine distance between vectors of n-grams 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 word or n-gram and 1 is maximally different.

For the entire space of contracts, we have 5,649,841 contract pairs representing 3362 contracts.[7] Table 7 presents the results of a linear regression of contract pair distance on several dummy variables indicating whether the contract pair has the same borrower, the same lender, the same industry, and the same year. We note similarities across contracts for each of these factors, with the same borrower being of first order importance.

In order to measure the relative similarity of each contract to the entire space of all contracts, we need to calculate the average distance for each contract. To do so, we first wish to control for the natural similarities which occur between contracts to the same firm, bank, industry, and time period which often have similar language. We regress the contract distance of each pair on a dummy indicating whether the pair of contracts is to the same borrower, from the same lender, within the same 2-digit SIC code, or within the same year. We then calculate the residual from this regression, and average the value of this residual distance for each contract. This measure then provides an estimate of the relative similarity of each contract to the entire space of contracts.

Using this average residual distance as our dependent variable, we estimate the determinants of overall contract similarity as a similar function of firm and loan level variables. In

[7] We lose 82 contracts due to parsing errors.

Table 8 we present the results of these regressions, using the average distance as both single words only and bigrams (one and two word combinations). The higher the average distance, the more dissimilar the language of the default section is from the rest of the sample. We interpret this as a measure of how "custom" the contract language is. Higher values indicate that the language used is more unique while lower values indicate that the language used is less unique and more boilerplate.

Contract distance is negatively related to to Tobin's Q and positively related to asset tangibility across all specifications. Loan contracts to firms with high growth opportunities appear to have less boilerplate language while firms with high asset tangibility have less. This provides evidence that enforcement costs and imperfections are related to boilerplate language. Firms with high growth opportunities are more likely to face greater enforcement costs and imperfections given the significant fraction of firm value tied up in future growth. Conversely, firms with more tangible assets face fewer deadweight default costs, and therefore show less boilerplate language.

Contract distance is also negatively related to the number of lenders, suggesting a possible role for drafting costs in boilerplate language, although the effect is only marginally significant. Contrary to our initial hypothesis, loan amount is negatively related to contract distance, suggesting more boilerplate language. Though the evidence is mixed, it appears that more specialized loan provisions are more common for smaller loans and loans which make up a smaller portion of existing debt. Contract standardization may therefore play a role in assuring enforceability for larger contracts. Specifically, as the stake size increases, the cost of litigating non-boilerplate contract features increases.

4.4 Credit Risk and Complexity

To further explore what our measures are capturing, we examine how they are related to credit risk and what additional information they may possess. We examine a subsample of loans for which we have information on the firms credit rating prior to the loan. For this

subsample, we break credit ratings into five categories, AA or above, A, BBB, BB, and B or below. We regress total words, unique words, average bigram cosine distance, and dispersion on the credit rating dummies and our remaining controls. The results are reported in Table 9.

Total and unique words are monotonically related to credit quality, with lower credit quality firms having more total and unique words in their default sections. After controlling for credit quality, leverage becomes insignificant, though stock return volatility remains a significant predictor of total words.

4.5 Interaction with Loan Maturity

To further examine the impact of total renegotiation costs on contract complexity, we interact each measure of renegotiation costs with loan maturity. Loan maturity mechanically increases the likelihood that a loan will be renegotiated, since the loan will be "alive" for a longer period of time. Consequently, we would expect measures of renegotiation costs to have a positive interaction effect with loan maturity.

The results are presented in Table 10. Interactions with the number of lenders provides a consistent positive relationship across all interactions, suggesting that the direct cost of renegotiating with a larger syndicate increases overall complexity. The remaining results are mixed, with product market fluidity providing only a marginally significant positive relationship. While repeat lending relationship has the expected negative sign, it is not at all significant.

5 Contract Detail and Ex-post Changes in Performance and Financing

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 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 and the change in cash holdings over the same period. Tables 11, 12, and 13 present the results of these tests.

The annual change in ROA is positively related to the 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

Aghion, P. and M. D. P. Rey (1994). Renegotiation design with unverifiable information. *Econometrica 62*, 257–282.

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.

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.

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.

Bustos, A. E. (2008). Litigation and the optimal combination of vague and precise clauses in contracts. Northwestern Law & Economics Research Paper 08-11, Northwestern University.

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

Feinerer, I. and K. Hornik (2014). `tm`: Text mining package. R package version 0.5-10, http://CRAN.R-project.org/package=tm.

Feinerer, I., K. Hornik, and D. Meyer (2008). Text mining infrastructure in r. *Journal of Statistical Software 25*(5), 1–54.

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

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.

Hoberg, G., G. Phillips, and N. Prabhala (2014, February). Product Market Threats, Payouts, and Financial Flexibility. *The Journal of Finance 69*(1), 293–324.

Hornik, K. (2014). `openNLP`: Apache opennlp tools interface. R package version 0.2-3, http://CRAN.R-project.org/package=openNLP.

Kahn, C. and G. Huberman (1988). Two-sided uncertainty and "up-or-out" contracts. *Journal of Labor Economics 6*(4), 423–444.

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.

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). Measuring readability in financial disclosures. *Journal of Finance*.

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.

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. 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.

Taddy, M. A. (2013). `maptpx`: Map estimation of topic models. R package version 1.9-1, http://CRAN.R-project.org/package=maptpx.

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. North-Holland.

Tables

Table 1: Two Document Collections: Full Sections and Sentences

	Sample Size	Total Words	Unique Words	Per Document $d = f, s$: Mean	Median	Min	Max	Std.Dev
Full D_F	3,444	1,692,595	3,999	491.5	485.0	9	1,396	133.5
Sentences D_S	69,324	1,692,595	3,999	24.4	20.0	1	278	19.1

Table 2: Measures of Contract Detail

Measure	Mean	Median	Min	Max	Std.Dev.
(a) Word Based Measures					
words – Total	491.5	485.0	9	1,396	133.57
words – Unique	205.2	205.0	8	341	39.12
(b) Topic Based Measures					
$sentences_f$	20.2	20.0	1	54	5.45
$main\ topics_f$	15.45	15.00	1	28	3.471

(c) Correlation Coefficients

	Total	Unique	$sentences_f$	$main_f$
Total	1.000			
Unique	0.923	1.000		
$sentences_f$	0.730	0.688	1.000	
$main\ topics_f$	0.683	0.731	0.837	1.000

Table 3: 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. Tobin's Q is Compustat item *prccq* × *cshoq* + *dlcq* + *dlttq* + *pstkq* divided by *atq*. Leverage ratio is *dlttq* + *dlcq* divided by *atq*. Tangible asset ratio *ppentq* divide by *atq*. 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
Stock Return Vol	.0322	.0175	.00685	.0278	.211
Tobin's Q	1.47	1.03	.331	1.16	6.34
Leverage Ratio	.295	.195	0	.283	.965
log(Assets)	6.7	1.71	1.45	6.6	12.4
Tanible Asset Ratio	.346	.245	.0179	.279	.915
log(Age)	2.71	.88	0	2.64	4.03
log(Loan Amount)	5.28	1.46	-1.97	5.3	10.1
log(Maturity)	3.62	.63	.693	3.72	4.88
Debt/Loan Amt	2.42	6.85	0	.929	148
Repeat Lender	.437	.496	0	0	1
# of Lenders	9.96	10.4	1	7	110
Lender <100 miles away	.167	.373	0	0	1
# of Covenants	2.56	1.23	0	3	10

Table 4: Correlation Coefficients

	Stock Return Vol	Tobin's Q	Leverage Ratio	log(Assets)	Tangible Asset Ratio	Segment Concentration	log(Age)	log(Loan Amount)	log (Maturity)	Debt/Loan Amt	Repeat Lender	# of Lenders	# of Covenants
Stock Return Vol	1												
Tobin's Q	-0.000656	1											
Leverage Ratio	0.0326*	-0.211***	1										
log(Assets)	-0.451***	-0.143***	0.172***	1									
Tangible Asset Ratio	-0.126***	-0.139***	0.230***	0.137***	1								
Segment Concentration	0.170***	0.159***	-0.0234	-0.339***	0.0795***	1							
log(Age)	-0.352***	-0.215***	0.00480	0.468***	0.0533***	-0.378***	1						
log(Loan Amount)	-0.439***	-0.0682***	0.244***	0.804***	0.101***	-0.227***	0.314***	1					
log(Maturity)	-0.136***	0.0176	0.0373**	-0.117***	0.00820	0.123***	-0.0891***	0.151***	1				
Debt/Loan Amt	-0.0443**	-0.123***	0.196***	0.261***	0.114***	-0.126***	0.169***	-0.0653***	-0.237***	1			
Repeat Lender	-0.191***	-0.0505***	0.133***	0.305***	0.0450**	-0.0760***	0.112***	0.313***	0.00927	0.0127	1		
# of Lenders	-0.276***	-0.0464**	0.220***	0.561***	0.0689***	-0.139***	0.158***	0.692***	0.138***	-0.0142	0.283***	1	
# of Covenants	0.131***	0.0729***	-0.0233	-0.360***	-0.109***	0.156***	-0.266***	-0.200***	0.240***	-0.138***	-0.0676***	-0.0960***	1

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

Table 5: Determinants of Word Counts

This table presents the determinants of total and unique words 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. Standard errors are clustered at the firm level.

	(1) Total Words	(2) Total Words	(3) Unique Words	(4) Unique Words
Stock Return Vol	202.1	629.9***	11.49	134.1**
	(0.94)	(3.06)	(0.18)	(2.20)
Tobin's Q	-3.849	-4.553*	-0.966	-1.160
	(-1.46)	(-1.74)	(-1.25)	(-1.50)
Leverage Ratio	85.07***	53.56***	22.19***	13.32***
	(5.11)	(3.23)	(4.41)	(2.66)
log(Assets)	4.144	-4.639	-0.414	-2.836**
	(1.56)	(-1.27)	(-0.52)	(-2.56)
Tanible Asset Ratio	-44.03***	-35.63***	-13.72***	-11.15***
	(-3.35)	(-2.74)	(-3.49)	(-2.83)
log(Age)	-19.63***	-14.34***	-5.265***	-3.723***
	(-4.45)	(-3.36)	(-4.11)	(-3.00)
Segment Concentration	11.34	4.512	3.283	1.313
	(0.99)	(0.40)	(0.99)	(0.40)
log(Loan Amount)		24.34***		6.633***
		(5.69)		(5.42)
log(Maturity)		19.75***		5.775***
		(4.00)		(3.82)
Debt/Loan Amt		0.411		0.0809
		(1.03)		(0.60)
Repeat Lender		-4.268		-0.840
		(-0.82)		(-0.54)
# of Lenders		-0.741**		-0.176*
		(-2.02)		(-1.65)
Lender <100 miles away		-20.73***		-5.245**
		(-2.86)		(-2.42)
# of Covenants		15.60***		4.433***
		(5.77)		(5.42)
Constant	487.7***	300.9***	216.1***	162.6***
	(17.95)	(8.96)	(27.90)	(16.27)
Observations	2918	2916	2918	2916
R^2	0.044	0.108	0.039	0.099
Year FE	Yes	Yes	Yes	Yes

t statistics in parentheses

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

Table 6: Determinants of the the Number of Sentences and Main Topics

This table presents the determinants of the number of sentences and the number of unique main topics in the events of default sections of each contract. Sentences represent all non-empty sentences. Main topics represent the number of unique main topics resulting from the probabilistic topic model. Standard errors are clustered at the firm level.

	(1) Sentences	(2) Sentences	(3) Sentences	(4) Main Topics
Stock Return Vol	7.350	-11.34**	-0.801	-11.74***
	(0.87)	(-2.15)	(-0.14)	(-2.90)
Tobin's Q	-0.209*	-0.0735	-0.0711	0.00801
	(-1.91)	(-0.89)	(-0.92)	(0.14)
Leverage Ratio	2.131***	0.542	1.343***	0.412
	(3.34)	(1.16)	(3.29)	(1.27)
log(Assets)	-0.423***	-0.285**	-0.420***	-0.340***
	(-2.69)	(-2.57)	(-3.98)	(-4.29)
Tanible Asset Ratio	-1.620***	-0.562	-1.131***	-0.513**
	(-3.05)	(-1.56)	(-3.27)	(-2.06)
log(Age)	-0.471***	-0.0449	-0.265**	-0.0156
	(-2.77)	(-0.38)	(-2.42)	(-0.19)
Segment Concentration	0.806*	0.672**	0.568**	0.490**
	(1.82)	(2.07)	(2.00)	(2.20)
log(Loan Amount)	0.704***	-0.0184	0.453***	0.0306
	(3.95)	(-0.15)	(3.90)	(0.36)
log(Maturity)	0.243	-0.344**	0.237*	-0.106
	(1.19)	(-2.38)	(1.84)	(-1.16)
Debt/Loan Amt	0.0126	0.000426	0.0114	0.00426
	(0.52)	(0.03)	(0.72)	(0.37)
Repeat Lender	0.0544	0.181	0.0904	0.165
	(0.24)	(1.10)	(0.66)	(1.62)
# of Lenders	-0.0453***	-0.0233**	-0.0148*	-0.00193
	(-3.27)	(-2.38)	(-1.66)	(-0.27)
Lender <100 miles away	-1.005***	-0.390*	-0.780***	-0.420***
	(-3.53)	(-1.73)	(-4.09)	(-2.74)
# of Covenants	0.523***	0.0601	0.348***	0.0774
	(4.57)	(0.81)	(4.92)	(1.55)
# of Words		0.0297***		0.0174***
		(37.18)		(32.09)
# of Unique Words				
Constant	17.78***	8.850***	14.34***	9.118***
	(12.99)	(9.14)	(15.97)	(13.03)
Observations	2916	2916	2916	2916
R^2	0.073	0.545	0.092	0.492
Year FE	Yes	Yes	Yes	Yes

t statistics in parentheses

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

Table 7: Determinants of Contract Pair Distance

This table presents the determinants of the cosine vector distance of each contract pair for words and 1-to-2 word combinations. The cosine distance represents the dissimilarity of each pair of contracts in terms of word or bigram (1 and 2 word combination) occurrence. Standard errors are clustered at the contract level.

	Word Distance	Word Distance	Bigram Distance	Bigram Distance
Matching Borrower	-0.469***	-0.474***	-0.490***	-0.492***
	(-31.34)	(-32.21)	(-32.75)	(-33.03)
Matching Lender	-0.0474***	-0.0474***	-0.0461***	-0.0461***
	(-17.59)	(-17.60)	(-19.65)	(-19.66)
Matching Year	-0.00290***	-0.00289***	-0.00263***	-0.00262***
	(-4.53)	(-4.53)	(-4.69)	(-4.68)
Matching SIC4	-0.0131***		-0.0106***	
	(-4.25)		(-4.01)	
Matching SIC2		-0.00787***		-0.00810***
		(-6.40)		(-6.95)
Constant	0.902***	0.902***	0.945***	0.945***
	(614.27)	(614.04)	(761.26)	(760.71)
Observations	5649841	5649841	5649841	5649841
R^2	0.008	0.008	0.011	0.011

t statistics in parentheses

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

Table 8: Determinants of Average Language Distance

This table presents the determinants average residual cosine distance of each contract from the space of all contracts. The average residual distance is calculated by averaging the residuals from a regression of contract distance on indicators for the same borrower, lender, year, and 2 digit SIC code. Standard errors are clustered at the firm level.

	(1) Word Distance	(2) Word Distance	(3) Bigram Distance	(4) Bigram Distance
Stock Return Vol	0.198***	0.156**	0.216***	0.156**
	(2.85)	(2.19)	(3.58)	(2.53)
Tobin's Q	-0.00449***	-0.00435***	-0.00353***	-0.00327***
	(-3.75)	(-3.62)	(-3.37)	(-3.10)
Leverage Ratio	-0.00451	-0.00257	-0.0144***	-0.00914*
	(-0.72)	(-0.41)	(-2.67)	(-1.68)
log(Assets)	-0.00104	0.0000886	-0.00347***	-0.000367
	(-0.98)	(0.06)	(-3.91)	(-0.28)
Tanible Asset Ratio	0.0107*	0.00907	0.0145***	0.0123**
	(1.94)	(1.64)	(3.01)	(2.53)
log(Age)	-0.00350**	-0.00441***	0.000936	-0.000175
	(-2.07)	(-2.60)	(0.63)	(-0.12)
Segment Concentration	0.00227	0.00281	0.00466	0.00523
	(0.48)	(0.60)	(1.11)	(1.25)
log(Loan Amount)		-0.000770		-0.00359**
		(-0.49)		(-2.56)
log(Maturity)		-0.000512		0.000425
		(-0.25)		(0.25)
Debt/Loan Amt		0.000364**		0.000138
		(2.41)		(0.92)
Repeat Lender		-0.000572		0.000595
		(-0.27)		(0.32)
# of Lenders		-0.000312**		-0.000314**
		(-2.09)		(-2.40)
Lender <100 miles away		0.00187		0.00206
		(0.57)		(0.76)
# of Covenants		-0.00124		-0.00168*
		(-1.22)		(-1.92)
Constant	0.0104	0.0182	0.0170*	0.0242**
	(1.05)	(1.37)	(1.92)	(2.14)
Observations	2916	2916	2916	2916
R^2	0.028	0.037	0.062	0.077
Year FE	Yes	Yes	Yes	Yes

t statistics in parentheses

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

Table 9: Credit Ratings and Complexity Measures

This table presents the determinants of total words, unique words, sentences, main topics, and average bigram cosine distance for the subset of firms which have public debt ratings as of the quarter prior to the loan date. Each credit rating category is a dummy taking on a value of one if the credit rating is in the assigned category. The baseline category is AA and above. All models contain additional firm and loan level controls from previous models. Standard errors are clustered at the firm level.

	(1) Total Words	(2) Unique Words	(3) Sentences	(4) Main Topics	(5) Cosine Distance
A	83.12***	19.32**	2.869***	0.286	0.00127
	(3.09)	(2.30)	(2.64)	(0.49)	(0.10)
BBB	92.16***	21.98**	3.045**	0.360	-0.00690
	(3.28)	(2.53)	(2.56)	(0.63)	(-0.54)
BB	134.8***	32.11***	4.212***	0.447	-0.0101
	(4.62)	(3.59)	(3.39)	(0.72)	(-0.76)
B and C	156.8***	39.94***	4.962***	0.243	-0.00934
	(4.99)	(4.18)	(3.61)	(0.36)	(-0.67)
Stock Return Vol	1029.9**	151.9	19.04	-15.08**	0.152
	(2.34)	(1.22)	(0.98)	(-2.11)	(1.18)
Leverage Ratio	-1.984	-1.583	1.248	1.141**	0.00268
	(-0.07)	(-0.19)	(1.07)	(2.15)	(0.28)
log(Assets)	-12.23**	-4.564***	-0.431	-0.159	0.00370
	(-1.99)	(-2.60)	(-1.62)	(-1.29)	(1.56)
Observations	1391	1391	1391	1391	1391
R^2	0.203	0.204	0.147	0.584	0.058
Year FE	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

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

Table 10: Interactions with Maturity and Return Volatility

This table presents the determinants of total words, unique words, sentences, and main topics, with additional interaction effects. Loan maturity is interacted with the number of lenders in each loan syndicate, whether the lead lender has made another loan to the firm in the past 5 years, and the product market fluidity as measured by Hoberg et al. (2014). All models contain additional firm and loan level controls from previous models. Standard errors are clustered at the firm level.

	(1) Total Words	(2) Unique Words	(3) Sentences	(4) Main Topics
log(Maturity)	-5.477	-0.236	-0.478	-0.0979
	(-0.50)	(-0.07)	(-1.12)	(-0.35)
# of Lenders	-6.376***	-1.809***	-0.198***	-0.106**
	(-3.13)	(-2.72)	(-2.67)	(-2.20)
Product Mkt Fluidity	-9.407	-2.208	-0.264	-0.0930
	(-1.53)	(-1.23)	(-1.09)	(-0.66)
Repeat Lender	15.65	6.009	0.703	0.241
	(0.51)	(0.63)	(0.50)	(0.30)
# of Lenders × log(Maturity)	1.427***	0.413***	0.0391**	0.0229*
	(2.93)	(2.61)	(2.19)	(1.95)
Product Mkt Fluidity × log(Maturity)	2.637*	0.591	0.0739	0.0259
	(1.67)	(1.27)	(1.17)	(0.70)
Repeat Lender × log(Maturity)	-5.386	-1.878	-0.177	-0.0422
	(-0.65)	(-0.74)	(-0.48)	(-0.20)
Constant	361.5***	176.5***	19.24***	14.82***
	(7.16)	(11.63)	(9.84)	(11.35)
Observations	2805	2805	2805	2805
R^2	0.115	0.106	0.073	0.088
Year FE	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes

t statistics in parentheses

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

Table 11: Contract Complexity and Future Outcomes

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	Δ ROA
Total Words (1000s)	0.0544***				
	(2.76)				
Unique Words (1000s)		0.166**			
		(2.45)			
Sentences			0.00134***		
			(2.80)		
# of Main Topics				0.00156**	
				(1.99)	
Avg Cosine Distance					-0.000612
					(-0.01)
Leverage Ratio	0.0600***	0.0606***	0.0598***	0.0602***	0.0609***
	(3.04)	(3.07)	(3.03)	(3.05)	(3.07)
Stock Return Vol	0.373	0.384	0.381	0.400	0.397
	(1.32)	(1.35)	(1.34)	(1.41)	(1.40)
Junk	-0.00253	-0.00191	-0.00132	-0.000947	0.00127
	(-0.34)	(-0.26)	(-0.18)	(-0.13)	(0.17)
Rated	0.00547	0.00528	0.00569	0.00550	0.00411
	(0.69)	(0.66)	(0.72)	(0.69)	(0.51)
# of Covenants	-0.00477**	-0.00465**	-0.00456*	-0.00444*	-0.00393*
	(-2.01)	(-1.97)	(-1.94)	(-1.89)	(-1.66)
Observations	2793	2793	2793	2793	2791
R^2	0.099	0.098	0.099	0.098	0.096
Addl Risk Controls	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

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

Table 12: Contract Complexity and Future Outcomes (Continued)

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	Sales Growth
Total Words (1000s)	0.171***				
	(3.05)				
Unique Words (1000s)		0.559***			
		(2.84)			
Sentences			0.00395***		
			(2.77)		
# of Main Topics				0.00588***	
				(2.69)	
Avg Cosine Distance					-0.168
					(-1.09)
Leverage Ratio	-0.0595	-0.0575	-0.0597	-0.0590	-0.0596
	(-0.86)	(-0.83)	(-0.86)	(-0.86)	(-0.86)
Stock Return Vol	-1.191	-1.159	-1.162	-1.104	-1.092
	(-1.33)	(-1.30)	(-1.31)	(-1.24)	(-1.23)
Junk	-0.0667**	-0.0655**	-0.0626**	-0.0633**	-0.0560**
	(-2.55)	(-2.51)	(-2.38)	(-2.41)	(-2.08)
Rated	0.0593**	0.0589**	0.0598**	0.0602**	0.0560**
	(2.37)	(2.35)	(2.38)	(2.40)	(2.22)
# of Covenants	0.0185**	0.0187**	0.0192***	0.0192***	0.0201***
	(2.53)	(2.55)	(2.63)	(2.63)	(2.78)
Observations	2807	2807	2807	2807	2805
R^2	0.259	0.259	0.259	0.259	0.257
Addl Risk Controls	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

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

Table 13: Contract Complexity and Future Outcomes (Continued)

This table presents the determinants of the future change in cash holdings, divided by assets, over the four quarters following the initiation of the loan. Cash holdings is defined as Compustat item *cheq* 4 quarters after the initiation of the loan, minus beginning of period *cheq* divided by beginning of period assets. 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 Cash	Δ Cash	Δ Cash	Δ Cash	Δ Cash
Total Words (1000s)	0.0477***				
	(2.69)				
Unique Words (1000s)		0.171***			
		(2.65)			
Sentences			0.000786**		
			(2.38)		
# of Main Topics				0.00147***	
				(2.67)	
Avg Cosine Distance					-0.0413
					(-0.78)
Leverage Ratio	-0.0214	-0.0209	-0.0212	-0.0213	-0.0208
	(-1.30)	(-1.27)	(-1.29)	(-1.29)	(-1.26)
Stock Return Vol	-0.223	-0.215	-0.211	-0.200	-0.196
	(-1.16)	(-1.12)	(-1.09)	(-1.04)	(-1.03)
Junk	-0.00471	-0.00465	-0.00298	-0.00352	-0.00202
	(-0.87)	(-0.85)	(-0.55)	(-0.65)	(-0.37)
Rated	0.0109*	0.0109*	0.0108*	0.0111*	0.0102
	(1.71)	(1.70)	(1.70)	(1.73)	(1.63)
# of Covenants	0.00242	0.00242	0.00277	0.00266	0.00313
	(1.06)	(1.06)	(1.20)	(1.14)	(1.37)
Observations	2811	2811	2811	2811	2809
R^2	0.080	0.081	0.079	0.079	0.078
Addl Risk Controls	Yes	Yes	Yes	Yes	Yes

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.

Table A3: Probabilistic Topic Model Results for Sentences

This table presents the results from the probabilistic topic model with $K^* = 55$ topics using the sentences corpus D_S with 69,324 sentences. Topics are listed by clusters. Kmeans clustering with 20 clusters (i.e., average number of sentences per default section) is applied. Words within a topic are ranked by relevance (Sievert and Shirley, 2014); topics within a cluster are ranked by usage; and clusters are ranked by aggregate usage of topics. The list of words is stem-completed using most frequent word.

Cl.	Rk.	Words	Usage
1	39	provision bankruptcy hereof pursuant hereunder accordance neglect circumstance promptly omit	1.2
2	10	plan erisa liability affiliate multiemployer pension incur withdrawal contribution notified	2.5
2	15	plan termination pbgc administrative reportable entities single erisa trustee appointment	2
2	30	member group erisa plan employer complete funding partial current intent	1.4
3	32	continue period unremedied unwaived addressed mentioned orrower subclause private enumerated	1.3
3	34	parties contract swap guarantee affected defined termination earlier greater result	1.3
3	41	agent lender administrative requisite favor representation production absence ratable disadvantaging	1.2
3	42	obligation thereunder thereto lease contingent balance bound listed sheet capital	1.2
3	44	condition exist relating entitled obtain adjudicate decree reason hcc unrelated	1.1
3	51	credit letter revolving account drawing issuer deposit arising indemnity page	0.9
3	55	time including limitation corrected omission supplement passage true misrepresentation basic	0.7
4	4	made deemed representation warranties certificate statement connection prove furnished incorrect	3.4
4	7	receiver appointment custodian substantial party trustee liquidation properties similar official	2.9
4	21	indebtedness holder give beneficiaries permit behalf prior evidencing maturity lapse	1.7
4	28	constitute court portion governmental competent jurisdiction therefor agency authority condemnation	1.4
4	29	lien collateral security created priorities perfected purported pledge interest document	1.4
4	45	occurred consolidated subsection determined good faith net sum contest fiscal	1.1
4	49	cure breach remedied advance facility waived compliance solely specifically additional	0.9
5	9	debts write pay generally admit unable due fail solvent	2.6
6	17	subsidiaries significant immaterial unrestricted insignificant brightpoint pesco aceto dormant minimis	1.9
6	35	section inclusive negative sentence maintenance preserve reference proviso appraisal consensual	1.3
6	36	material viacom infinity apria target wli worldwide appraisal heretofore induce	1.3
7	23	adverse result expected effect event opinion financial imposition prospects impairment	1.7
8	24	parent stock action vote exchange holder directors beneficiary ownership power	1.6
8	38	license operations loss claim environmental renew revocation violation suspension damage	1.2
8	53	demand commitment owed present cash immediately automatically thereon accrued relevant	0.8
8	54	excluding rate equal hedging management equivalent collectively currency longer greater	0.7
9	47	guarantor obligor guaranties restricted future disaffirm merger resources ppl consummation	1
10	5	perform observe contained covenant fail article compliance party affirmative	3.3
11	26	make creditors benefit assignment possession generally takes consent meeting called	1.5
12	33	days sequestrator dismissed ryder eurocurrency uncontested exceptions begun mentioned unconditionally	1.3
12	46	respect carpenter assumption consensual interest sepracor duplication appraisal conformity parenthesis	1
13	48	provided clause paragraph sale transfer type applicable assets foreclosure security	1
14	18	aggregate excess amount payment exceeds individually dollars threshold million combination	1.9
15	16	agreement term intercreditor promissory shorter sepracor subpart access interest concurrently	2
16	8	judgment insurance stayed appeal money order rendered final covered pending	2.9
16	11	cease full reason force valid enforcement denies binding effect deliveries	2.3
16	12	petition filed federal stated foreign answer law united proceeding	2.2
16	14	seeking relief reorganization insolvency law debtors arrangement winding dissolution composition	2.1
16	27	execution attachment levied warranty process writing assets bonded issued vacated	1.5
16	52	subject granted transaction tax code extension base standard penalties forfeiture	0.9
17	22	action set corporate foregoing acquiescence authority consent furtherance approving partnership	1.7
17	40	remain unstayed undischarged referred undismissed period unbonded nature consecutive suffers	1.2
17	50	note subordinated indenture issued permit senior defined mortgage charge	0.9
18	2	default event happening called cross unmatured number subdivision quot subpart	4.3
18	6	occur control change sufficient rise wilmar fundamental import pacific texas	3.1
18	13	loan document remain australian eurocurrency asi enumerated defect wireless conformity	2.1
18	31	thereof date fixed prepayment disbursement rollover mentioned component orrower ryder	1.3
18	37	companies business conduct restrained prevented enjoined ordinarily suspend discontinue	1.3
19	1	due interest principal pay payable fee fail amount payment reimbursement	4.3
19	3	case commenced order proceeding involuntarily entered relief decree law	3.4
19	19	applicable grace acceleration instrument period maturity indebtedness outstanding expiration principal	1.8
19	20	required scheduled prepayment purchase prior stated prepaid maturity redeem repurchase	1.8
19	43	person directors board majority election calendar nomination individually month consist	1.1
20	25	failure notice written earlier official request responsible knowledge aware receipt	1.6