Shades of Sunshine: Adaptation of Contract Completeness in Florida Public Procurement

Shades of Sunshine: Adaptation of Contract Completeness in Florida Public Procurement

a thesis presented
by
Chris Huang
to
The Department of Applied Mathematics

in partial fulfillment of the requirements
for the degree of
Bachelor of Arts
in the subject of
Applied Mathematics

Harvard University
Cambridge, Massachusetts
March 2025

Shades of Sunshine: Adaptation of Contract Completeness in Florida Public Procurement

Abstract

Public procurement contracts often fail, leading to cost overruns and amendments negotiated ex post. We examine how government agencies modify successive contract completeness—the degree to which all contingencies are specified in a contract—in response to contract failures. Utilizing a novel preprocessing methodology that extends previously developed measures of completeness to produce more granular insights, we measure the completeness of 7,725 construction and maintenance contracts from the Florida Department of Transportation between 2008 and 2024. The findings show a rise in contract completeness and a corresponding decline in contract failures in the long term. No adaptation of contract completeness to failures is found in the short term. Furthermore, we observe evidence of a discounting mechanism for transaction costs associated with writing complete contracts, which allows completeness levels to remain high even during months with a high volume of contract drafting.

Contents

Acknowledgments

As customary, I will try to be as brief as I am sincere. My utmost gratitude to:

Professor Alisha Holland for welcoming and cheering on a lost but eager student.

Juan Dodyk for his ceaseless guidance and great ideas at every step of this project.

Professor Daniel Carpenter for his generosity in agreeing to read this thesis and his delightful lectures.

Io, Gabe, and Camilla for their friendship, encouragement, and discerning eye for edits.

Jess and Pongo for being my solace.

And all my friends and family for their love and understanding support as I disappeared for the last few weeks before submission.

Funding for computation cost was supported by a Harvard College Research Program Grant.

1
Introduction

Each year, government agencies around the world allocate $10 trillion dollars of public procurement contracts to private firms. In the United States, public procurement accounts for 19–24% of GDP across all levels of government [Hafsa et al., 2021]. This massive procurement endeavor is as important as it is difficult. Agencies face canonical challenges typical of principal-agent relationships. Government agencies (principals) must select and constrain private firms (agents) at an asymmetrical informational and expertise disadvantage. In this pursuit, agencies write out explicit expectations and constraints in the form of a contract to define their relationship.

Frequently, contracts fail—either by incurring cost overruns or amendments outside the original contract design. These ex post changes can lead to costly renegotiation which suffers from moral hazard. Beyond minor contract failures and corrections, large-scale, catastrophic failures also occur. A prominent example is Boston's "Big Dig" tunnel project, which required over 1,500 amendments and saw costs escalate from an initial estimate of $2.58 billion to $14.8 billion, largely due to poor design [Greiman, 2010]. Given these prevalent failures, an important question arises: How do agencies adapt their contracts in response to past failures?

This thesis examines the adaptation of contract completeness in reaction to past rates of contract failure (cost overruns, amendments). Completeness is the degree to which a contract covers all possible contingencies—disambiguating the responsibilities of each party in each state of the world. This thesis demarcates two camps of arguments for ideal levels of completeness in contracts: classic completeness theory and strategic ambiguity theory.

Classic completeness theory posits that greater contract completeness is preferable as it increases efficiency and reduces ex post opportunism. The primary limitation is the transaction costs for drafting complete contracts involved in finding and covering all contingencies.

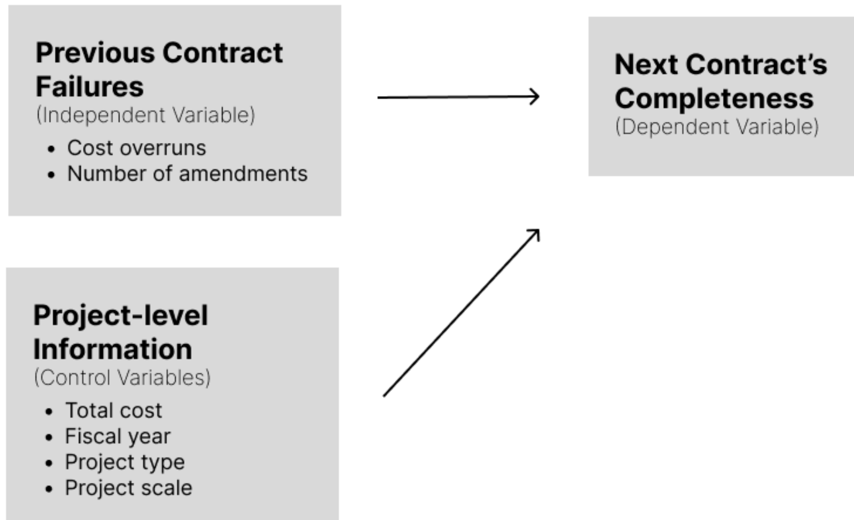

Figure 1.1: Diagram of the Target Analysis

Strategic ambiguity theory argues that deliberate contractual incompleteness can be advantageous, even in the absence of drafting costs. This perspective argues that overly complete contracts incentivize contractors to focus solely on fulfilling narrowly defined and verifiable contingencies, potentially neglecting other important but non-verifiable aspects of the agreement. Additionally, these strictly defined contingencies can be inflexible, increasing the likelihood of ex post renegotiation.

Both of these theories generate opposing suggestions for how agencies should adapt completeness in response to contract failures. Is the failure an issue of too much or too little completeness? To provide an empirical basis, we measure contract completeness and contract failures over time, controlling for project-level factors.

Deriving a measure of completeness is difficult. Ideally, all contingencies present in a contract are compared relative to all possible contingencies. This process can be inherently subjective and prohibitively costly as the same transaction costs needed for writing a complete contract are incurred. In practice, completeness is measured with rough proxies, including manual coding, count-based methods (the number of words, sentences, or restrictive clauses present), and topic-modeling approaches.

This thesis advances current methodology by introducing a preprocessing pipeline of chunking, embedding, and clustering prior to computing completeness measures. It permits the calculation of a completeness measure for the entire contract, alongside more granular measures that capture the completeness of specific sections, such as project details, bidding requirements, or insurance protocols.

For our data, we turn to the "Sunshine State" of Florida. The Florida Accountability Contract Tracking System (FACTS) offers exceptional access to extensive public procurement data for the state. Our analysis focuses on high-value (>$100,000) construction and maintenance projects by the Florida Department of Transportation from 2008 to 2024. We web-scraped 7,725 full-length contract PDFs, along with project-level and amendment data. Furthermore, a text annotation algorithm was applied to the contracts to extract controlling covariates, including project type (e.g., construction or maintenance, bridge or highway) and scale (e.g., miles of work).

Completeness measures for all 7,725 contracts were computed using the advanced methodology proposed earlier. Analysis examining the relationship between past contract failures and successive contract completeness, controlling for project-level details [see fig.1], reveals five major findings: (1) Overall, contract completeness increased and contract failures decreased in the last 16 years at the Florida Department of Transportation. (2) Most variation in completeness is unexplained by project-level details. (3) Short-term adaptation of completeness to previous contract failure is not observed. (4) Granular completeness scores are weakly and positively correlated, but all increase over time. (5) Months with a high volume of contracts are not associated with less complete contracts.

The rest of the thesis is structured as follows: Chapter two details the theories of completeness and the opposing camps of arguments for how complete a contract should be. Chapter three evaluates previous methods for quantifying completeness along with their challenges, then it presents an advancement in methodology for more granular measures. Chapter four combines the theoretical questions and practical methodologies from the previous two chapters to conduct an empirical study on 7,725 Florida DOT contracts. Chapter five reveals the most important findings for the relationship between contract failures and completeness. Finally, chapter six discusses the implications of the work done and the most promising lanes of further research.

2

The Logic of Completeness

2.1 Defining Completeness

Contract completeness is traditionally defined as the specification of contingencies for all possible states of the world. Though, common with abstract concepts, there are alternative definitions. The degree of completeness can also be understood in terms of the specificity of each contingency. Prominent scholars in the field of Contract Theory, such as Jean Tirole, define completeness by the "resources expended to identify the appropriate design" or incompleteness by the "probability that the design specified in the contract needs to be altered ex post" [Tirole, 2009]. Each of these definitions captures some aspect of how comprehensive a contract is.

Regardless of the precise definition, the literature on contract completeness is broadly divided into two normative perspectives: Classic completeness theory and strategic ambiguity theory. Classic completeness theory posits that greater contract completeness is preferable, limited only by the costs of drafting the contract. Conversely, strategic ambiguity theory argues that deliberate contractual incompleteness can be advantageous, even in the absence of drafting costs. These opposing theoretical frameworks generate divergent recommendations for addressing contract failure. This analysis will outline these contrasting outcomes and explore the empirical question of which theory prevails in practice.

2.2 Classic Completeness Theory

Kenneth Arrow and Gérard Debreu set the foundation for complete contract theory with the conception of the complete market. In a complete market, also known as an Arrow-Debreu market, a trade exists for every possible contingency or state of the world. Following, there exists complete contracts which can allocate prices for all possible contingencies. These assumptions of complete markets and complete contracts build into their highly-influential and

Nobel Prize winning theory of General Equilibrium which allows markets to maximize utility in a Pareto optimal manner.

Carefully-written and complete contracts constructed through optimal contract design were seen as a way to maximize utility for all parties—especially the principal writing the contract. Laffont and Martimort offer a theory of incentives to design contracts that align the interests of principals and agents even when information is asymmetric, examining the solutions to mitigate adverse selection and moral hazard [Laffont and Martimort, 2002]. The lack of complete contracts is viewed as a source of inefficiency. Oliver Williamson argued that incomplete contracting leaves gaps that allow for opportunism—requiring the emergence of other institutions to cover the inefficiency [Williamson, 1979]. Similar arguments of incomplete contracts inviting ex post opportunism are argued by Goldberg and Masten [Goldberg, 1985, Masten, 1988].

Incomplete contract theories emerged to explain why complete contracts are rarely, if ever, seen in practice. The prevailing argument is that of transaction costs: complete contracts are costly to write [Williamson, 1979]. In a world with many possible contingencies, nontrivial time and cognitive resources ("search costs") are required to discover each contingency and formulate an optimal response [Klein, 2002]. Consequently, contracting parties deal with their limited cognition by using heuristics and writing incomplete contracts [Tirole, 2009].

As such, contracting parties face a tradeoff between the benefits of contract completeness—efficiency in allocation, alignment of incentives in all possible contingencies, and reduction of opportunism—and the costs of contract completeness from transaction costs due to bounded rationality and search costs. Critically, full contract completeness is an ideal that cannot be practically achieved in classic contracting theory.

2.3 Strategic Ambiguity Theory

Incomplete contract theory initially focused on explaining the inevitability and consequences of contracts falling short of completeness. However, the field has since expanded to explore the strategic advantages of incomplete contracts, suggesting that deliberate opacity can be beneficial. These arguments can be broadly classified into three rationales:

Undesired Information Sharing

By definition, complete contracts require all contingencies to be specified. However, scholars point out cases in which revealing complete information may be disadvantageous. In Tirole, the knowledgeable party revealing their preference for state A as opposed to state A' gives away information and prevents the knowledgeable party from fully benefitting from its information advantage [Tirole, 2009]. In Spier, the act of revealing adverse contingencies by the knowledgeable party makes it appear more likely to occur [Spier, 1992]. In Ederer, the less knowledgeable party may introduce opacity to prevent better-informed agents from gaming incentives schemes by focusing on the least costly tasks [Ederer et al., 2018].

Enforcement of nonverifiable aspects

Complete contracts can only meaningfully cover contingencies that are verifiable when disputed (such as in a court of law) even if the contingency is observable. There is a family of arguments that suggests leaving verifiable contingencies unspecified to improve performance in nonverifiable contingencies. In Holmstrom and Milgrom, contracts that tie an agent's compensation to verifiable measures diverted attention from more important but harder-to-measure tasks [Holmstrom and Milgrom, 1991]. In Bernheim and Whinston, leaving verifiable aspects of performance unspecified created informal influence to encourage performance in nonverifiable aspects of performance [Bernheim and Whinston, 1998].

Complete contracts are too rigid

Incomplete contracts are more flexible in nature by not strictly defining the protocols in each contingency. In Scott and Triantis, vague standards in contracts can encourage parties to fulfill the spirit of the contract rather than only satisfying the narrowly defined stipulations [Scott and Triantis, 2006]. In Bajari and Tadelis, designing strict and comprehensive contracts ex ante may produce strong incentives to reduce costs but leads to higher adaptation costs ex post due to costly renegotiation; naturally, these adaptation costs dominate for more complex projects [Bajari and Tadelis, 2001]. Note that this is the inverse of the definition provided by Tirole, where incompleteness is constructed as the probability that ex post negotiations occur. As such, strategic ambiguity theory suggests that the contracts ideally have some level of incompleteness even in the absence of transaction costs to write complete contracts.

2.4 Response to Contract Failures

In practice, contracts are rarely written optimally in accordance with either theory and often result in contract failures where there are significant cost overruns and amendments to the contract ex post. A salient question emerges for how contracting actors respond to contract failures and modify subsequent contracts and their level of completeness. There is minimal research exploring this question of practical adaptation as discussion of contract completeness tends to be discussed in single acts. However, each theory suggests different solutions for strategically adapting future contracts.

Classic completeness theory is extended by the field of relational contracts. By contracting with a partner in a long-term relationship, cooperation is enforced by repeated interactions [Macchiavello and Morjaria, 2023]. As successful contracts are completed, reputational capital accrues through building a history of not engaging in opportunistic behavior [MacLeod, 2007]; the demand to reduce opportunistic behavior decreases, and contract completeness is less necessary. Inversely, contract failures may increase the perception of opportunism and inadequate design, leading to more contract completeness as a response.

Strategic ambiguity theory offers a different response to contract failure. Based on the rigidity camp of arguments, contract failures may be the result of too much completeness as rigid standards shift agents towards checking off narrow stipulations opposed to fulfilling the spirit of the contract. Extending Bajari and Tadelis, if contract design is a tradeoff between providing incentives through strict design and reducing ex post negotiations, then the response to high levels of ex post negotiations is to reduce contract completeness in favor of a looser contract.

Theory Used	Previous Contract Status	Next Contract Change in Completeness	Reason
Classic completeness theory	Success	Weakly decrease	Relational capital decreases need for costly completeness to enforce behavior
	Failure	Increase	Failure demonstrates need for more complete design to reduce opportunism, even if completeness is costly
Strategic ambiguity theory	Success	Weakly increase	Lack of ex post negotiations allow more incentives to be added
	Failure	Decrease	Failure demonstrates need for more flexible design that encourages the spirit of the contract to be fulfilled instead of only rigid and narrowly defined stipulations

Table 2.1: Theoretical Explanations for Changes in Contract Completeness

The existing literature on contract completeness offers valuable theoretical arguments for optimal completeness, exploring the ideal level of contingencies to mitigate risks and align incentives under various assumptions. However, the practical relevance of normative arguments depends on a foundational understanding of how contracting agencies actually operate and adapt their practices based on real-world experiences. Currently, the literature lacks comprehensive descriptive accounts of how agencies update their contract practices, particularly in response to events like contract failures.

This thesis directly addresses this demand. Fundamentally, this research aims to understand whether, in the face of contract failure, agencies treat the issue as arising from insufficient contractual detail or from an excess of it. By exploring this descriptive dimension, this thesis provides a crucial empirical foundation for evaluating and refining normative theories of contract completeness.

3

Operationalizing Completeness

Deriving a measure of completeness is tricky. The ideal is to count the number of present contingencies and the number of possible contingencies for a given contract. In practice, this is difficult. To measure completeness, one must pay the search costs to discover all possible contingencies—a cost that was not even made in the construction of the contract. For added complexity, there is subjectivity in differentiating contingencies and ranking the specificity of each contingency. Alternative completeness measures, such as the degree of resources expended to identify appropriate design, are dependent on the individual efficiency of contract writers and are not observable directly in contract text. In practice, completeness is measured with rough proxies of varying quality. The following chapter outlines and evaluates common methods to measure completeness, then proposes an advancement to achieve more granular measurements.

3.1 Previous Methods to Measure Completeness

Manual Coding

The most straightforward method is to create a coding scheme for levels of completeness and manually categorize contracts. Crocker and Reynolds analyzed Air Force engine procurement by coding contracts based on the degree of flexibility in the incentive and adjustment scheme[Crocker and Reynolds, 1993]. Aubert et al. used a survey to code the presence of 14 common contractural clauses for performance, adjustments, and coordination mechanisms [Aubert et al., 2017]. Saussier as well as Anderson and Dekker similarly used a survey to code the inclusion of prespecified contractual clauses [Saussier, 2000, Anderson and Dekker, 2005].

Low scalability and low granularity are the primary drawbacks of manual methods. The human component of manual coding increases costs—leading to smaller datasets coded. To ensure consistency between codings, the levels of completeness are demarcated into broad buckets, limiting the precision and granularity of the measure. For example, Aubert et al.'s completeness score only includes integers that range from 0 to 14.

Count-based Methods

Count-based methods offer a simple and scalable proxy for completeness. John D. Huber accomplished his seminal work on how much discretion policymakers give bureaucrats using a simple word count [Huber and Shipan, 2002]. The larger the word count, the greater the amount of detail present in a contract. D'Acunto et al. propose a three-fold count-based measure of completeness by using word count, sentence count, and unique word count [D'Acunto et al., 2020]. Both Kosnik and Moszoro et al. use a dictionary of restrictive words such as "must" and "cannot" to count the number of defined obligations in a contract as a proxy for completeness [Kosnik, 2014, Moszoro et al., 2016].

While count-based methods are tractable and easy to implement at scale, they are noisy proxies for completeness. Count-based methods are susceptible to picking up "wordiness" of contract text, which is dependent on the style of the contract writer rather than directly measuring the number of contingencies present.

Topic Modeling

Topic modeling is an advancement of simpler count-based methods. Instead of counting simple occurrences such as words or sentences, topic modeling attempts to determine the topics in a text. Theoretically, the number of topics in a contract corresponds directly with the number of contingencies specified. Commonly, a Latent Dirchelet Allocation (LDA) is used. Developed by Blei et al., LDA is a generative probabilistic model that assumes each document has a topic distribution, and each topic has a word distribution that is randomly drawn from to create the document [Blei et al., 2003]. LDA iteratively determines topics and their word distribution in a document based on word cooccurrences in a larger corpus of texts. Both Ganglamair and Wardlaw, and D'Acunto et al. use LDA to determine the number of topics in a contract as a measure of completeness [Ganglmair and Wardlaw, 2017].

The drawbacks of topic modeling approaches stem from the limitations of LDA in discovering topics. Fundamentally, it infers topics by the co-occurrences of words without consideration for semantic meaning and other contextual information. The topics discovered are a collection of words that co-occur but may not map onto a human interpretation of relevant topics. Additionally, the output of LDA can be arbitrarily altered by the random seeds and the number of prespecified topics searched for. In practice, utilizing LDA is partly an art rather than a consistent and deterministic procedure.

Distance Measurements

An interesting offshoot approach is determining contract completeness using vector distance measurements. The raw text of each contract is embedded into some numeric vector that represents semantic features of the contract, then the dissimilarity of any two contracts is determined by taking the cosine distance between the corresponding vectors. Ganglmair and Wardlaw attempted this by representing contracts as vectors of word and topic frequencies, then the average distance of a vector with all other vectors is used as a measure of how custom the contract is. The assumption is that more custom contracts include less boilerplate and are more complete.

This type of approach was briefly experimented with in this thesis. However, its validity as a completeness measure relies on two assumptions. First, a contract being more dissimilar is more complete as it takes more resources to write a unique contract. Second, most incomplete contracts look very similar to each other due to boilerplate. This approach was abandoned by this thesis as these assumptions could easily be invalidated. For example, uniquely incomplete contracts can be dissimilar to most other contracts, and incomplete contracts (which are likely to be shorter) can have more variance due to their small amount of content.

3.2 Proposal to Achieve More Granular Completeness Measures

This thesis develops a methodology to generate more granular measures of contract completeness, extending the functionality of previous approaches. Prior methodologies provide a singular completeness value for a contract. However, we may wish to examine how complete different components of a contract are. For instance, there may be a more meaningful signal in the completeness of project details compared to the completeness of insurance protocols. By disaggregating completeness into these component-level measures, it becomes

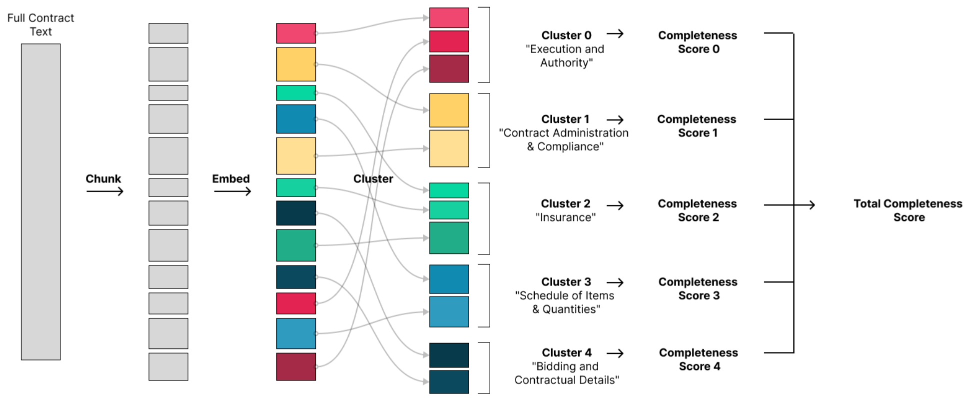

Figure 3.1: Granular Completeness Score Pipeline

possible to observe more nuanced shifts in completeness from one aspect to another. Furthermore, this approach enables the isolation of components where the measurement of completeness is possibly less influenced by noise. This approach of disaggregating completeness is compatible with any of the previously described methods to determine completeness once the original text has been partitioned into components.

As an illustration of the proposed methodology, imagine taking a full contract text printed on a long, continuous sheet of paper. As you read through the text, mark and cut the paper whenever there is a shift in topic. Once the entire text has been cut into chunks, organize all the chunks into a predetermined number of groups based on similar semantic meaning. For example, one may group all the chunks related to technical specifications into one pile whereas insurance chunks and deadline chunks go in different piles. Tape together the pieces of paper in each group and treat each group as an individual document. Use one of the previously specified methods to measure completeness on each group text independently. Now, we have the component-level completeness scores and can aggregate them if desired to also get the same result as the previous approaches.

This procedure is operationalized at scale through the following steps. (1) Chunking text based on semantic breaks using a Large Language Model. (2) Convert each chunked text into a vector representing its semantic meaning via an embedding model. (3) Cluster the vector embeddings to form *k* groups of similar chunks. (4) Aggregate all the chunk texts in a cluster. (5) Calculate a completeness measure on each cluster's aggregated text to get k separate scores. (6) Optionally, sum the *k* scores to generate the total completeness measure of the entire contract—which is also achievable by skipping steps 1-4 and performing 5 on the entire contract text directly. See Figure 3.1 for a visual illustration of this process and Chapter 4 for a more detailed implementation of this procedure on data.

4

Empirical Application: Florida DOT Public Procurement

4.1 Data

Though Florida's reputation as "The Sunshine State" stems from its inviting climate, its "Sunshine Laws" are even more compelling. Florida boasts an exceptional suite of laws to guarantee access to public records and government activity. Interestingly, the popular perception of the "Florida Man" as a uniquely strange individual who frequently appears in odd news headlines is partially a byproduct of the state's expansive public record laws. On the public contracting side, passed in 2010 by the state legislature, the Transparency Florida Act grants public access to data over 2 million public procurement contracts over the last two decades. This outstanding transparency provides the rich, long-run data needed to examine the relationship between contract failure and successive contract completeness.

This analysis selects high-value (>$100,000) construction and maintenance contracts by the Florida Department of Transportation. The Department of Transportation represents the State's second largest spender by total contracting dollars disbursed. Construction-type contracts were selected due to their high value and political salience. Practically, construction-type contracts are useful as each construction project is unique and could warrant uniquely complete and differentiated contracts. In comparison, procurement for commodities is less likely to have meaningful variance in completeness and instead rely on model contracts.

Through the Florida Accountability Contract Track System (FACTS), projects funded or ordered by the Florida Department of Transportation categorized as Building and Facility Construction and Maintenance Services are retrieved. Nearly all (90%) of these contracts are competitively procured through an open bid. Basic info such as date, vendor names, total contract amount,

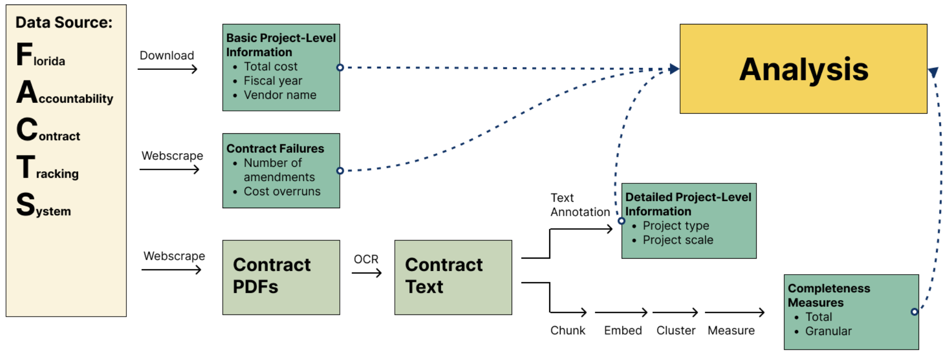

Figure 4.1: Diagram of Data Flow for Empirical Application on Florida DOT

and method of procurement are freely available to download. Distinctly outstanding from all other states' procurement transparency, Florida releases full original contract PDFs for most procurement activity. A script was executed to webscrape full contract PDFs from the FACTS dashboard. Additionally, data on contract amendments and cost overruns are scrapped from FACTS.

4.2 Methodology

Generative Models Use Cases

The field of natural language processing (NLP) took a tremendous leap after the emergence of large language models (LLMs). Trained on extremely large datasets, LLMs can flexibly perform tasks that require understanding human language, including context, semantics, and structure. A cautious approach to using these models is necessary due to questions of output validity and hallucination. Done carefully, these models exhibit solid performance in simple "zero-shot" or "one-shot" tasks—done without extensive training as a traditional machine learning model requires. This leads to faster iteration and expands the scope and scale achieved by this thesis. For this methodology, the use of generative LLMs—all carried out with API calls to Google's Gemini 2.0 Flash—are relegated to the following three mundane and well-defined tasks:

1. **Optical Character Recognition (OCR):** Used to convert PDF documents to machine-readable text. LLMs with integrated vision excel at reading text, even for handwriting and dynamic layouts. Practically, this method achieves high-quality OCR outputs at lower costs. However, the outputs are not differentiable from other traditional OCR methods.

2. **Chunking:** Used to split full text into chunks whenever there is a semantic shift in content. This approach is known as agentic chunking and generates chunks with more contextual understanding compared to naive methods such as fixed-size chunking [Setty et al., 2024]. These chunks will later be embedded and clustered in the pipeline.

3. **Text annotation:** Used to code a text based on a predetermined schema. This is traditionally a costly and labor-intensive process carried out by a team of humans reading and categorizing text. LLMs demonstrate impressive performance at annotation, and this use case constitutes a quickly growing body of practitioner-based recommendations for reliable results and application in political science research [Alizadeh et al., 2025]. In this thesis, text annotation is performed on contracts to derive a better classification of project type and scale than the given data.

Application of Proposed Methodology for Granular Completeness Measures

First, contract PDFs are converted into text files via optical character recognition. Then, these contract texts are agentically chunked into semantically distinct sections. Chunking is successfully completed—producing 399,506 chunks. On average, each contract has 52 chunks, and each chunk has 252 words.

Then, each chunk is mapped to a vector that represents its semantic meaning. This is accomplished using OpenAI's text-embedding-3-small model. Through the process of embedding, each chunk of text is converted into a 1,536-dimensional vector. Intuitively, chunks with embedded vectors that are closer together have more semantic similarity.

Following, the embedded vectors of all 399,506 chunks are clustered into $k = 5$ groups using the k-means cluster algorithm. Each cluster represents a group of chunks unified by semantic similarity. For readability, a random sample of each cluster is taken and qualitatively assessed to create labels for the cluster's general topic.

For each contract, chunks of common cluster membership are aggregated to create five sections of subtext (i.e., a subtext for "Execution and Authority," "Contract Administration & Compliance," "Insurance Coverage," "Schedule of Items & Quantities," and "Bidding and Contractual Details," respectively). All section subtexts are measured for completeness separately. For simplicity,

Cluster Label	# of chunks	freq. in documents %	Variance	% of words
Execution and Authority	43,172	93.2%	0.4922	14.7%
Contract Administration & Compliance	115,024	81.9%	0.6132	24.6%
Insurance Coverage	36,467	79.0%	0.4616	9.7%
Schedule of Items & Quantities	94,100	93.6%	0.4079	15.8%
Bidding and Contractual Details	110,678	99.2%	0.3732	35.3%

Table 4.1: Cluster Statistics Summary

a count-based method of total words and unique words was used to create a summed completeness score for each section subtext. We can now observe the granular completeness scores. However, it is helpful to additionally compute the summed completeness score to simplify certain analyses.

Acquiring controlling variables

To isolate the effect of past contract failures and successive contract completeness, this thesis controls for the following variables that could affect contract completeness: fiscal year, total cost, project type, and project scale. The fiscal year and total cost (which control for time-dependent factors and general measures of stakes, respectively) are both found in the basic info provided by FACTS.

The project type and project scale also intuitively affect the baseline completeness of a contract. Type matters as a bridge construction contract is notably different from a pothole repair contract. Likewise, a high construction contract will require a different baseline completeness depending on whether it procures a minor extension to an existing highway or the construction of a large interstate highway. The basic info provided by FACTS include a rough and overly broad type classification and no scale classification.

To acquire these important controlling covariates, I run a text annotation procedure on the contract text to code three aspects: the project type, project scale, and unit of project scale. Every project type was successfully classified, with the five most common being "Roadway Maintenance", "Traffic Infrastructure (Signals, Road Lighting, Signage)", "Highway Construction", "Landscaping & Roadside Maintenance", and "Drainage & Stormwater Management". Measures of scale coding had a lower yield, with only 60% of contracts assigned to a specified scale unit. The most common units of scale were "Undetermined/Other", "Miles", "Locations", "Installations", and "Days".

Setup of Analysis

Our data, derived from the preceding steps, includes a set of dependent, independent, and controlling variables. The dependent variables are a total completeness score and five granular completeness scores for each contract. Contract failures, our independent variables, are represented by cost overruns and the number of amendments, measured at both the vendor (based on the vendor's last contract) and agency (average over the 90 days prior to execution) levels. The analysis also controls for total project cost, fiscal year, project type, and project scale. The analytical approach is outlined below:

First, we examine descriptive statistics of completeness, specifically exploring the distribution of completeness scores and how average completeness changes over time. This initial overview also investigates its relationship with other variables on a contract level and temporally, both across years and within the annual cycle. This initial, basic assessment utilizes exploratory plots, correlations, and time series visualizations of the relevant variables.

Second, we isolate the effect of past contract failure on successive contract completeness using regressions. By construction on a vendor's prior contract or the agency's performance in the preceding 90 days, the contract failure indicators are already time-lagged. Therefore, we directly regress completeness scores on past failures and project controls. To account for potential non-linearities, we include the logarithm of cost overrun and total price. We also incorporate interaction terms between project type, scale, and scale unit, as well as between total cost and project type. The specific regression setup is detailed in the following chapter. The regression outputs address two key questions: What is the significance of past contract failure as a predictor of completeness, and what is the direction of this effect? A positive relationship would support classic completeness theory (completeness increases after failure), while a negative relationship would align with strategic ambiguity theory (completeness decreases after failure).

Lastly, we examine descriptive statistics of the granular completeness scores generated by our novel preprocessing methodology. This analysis investigates the correlations between these distinct granular scores and explores their individual trends over time. The utility of these granular measures lies in their potential to reveal variations beyond those captured by overall completeness scores.

5

Results

There are five main findings from the data exploration and regression analysis.

5.1 Overall trend of contract completeness increasing and contract failures decreasing in the last 16 years.

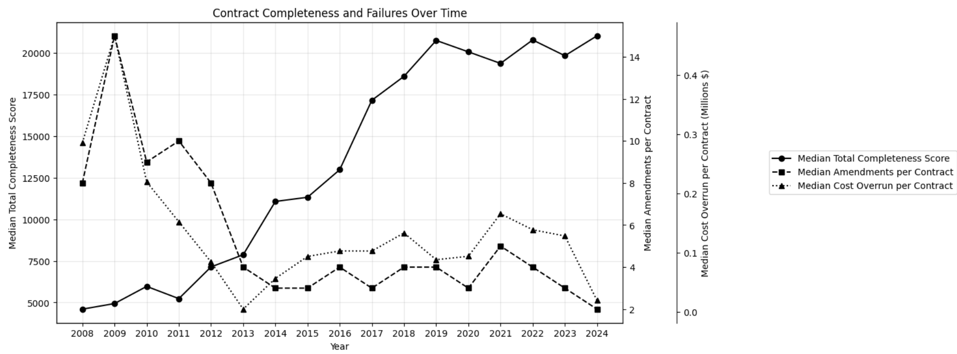

Figure 5.1: Contract Completeness and Failures Over Time

Over the 16-year period with substantial data in FACTS, median contract completeness has increased. This trend is associated with a general decline in contract failures, defined as median ex post amendments and cost overruns. Pearson correlation analysis reveals a moderate negative relationship between yearly median contract completeness and contract failures, with coefficients of -0.730 for median amendments and -0.447 for median cost overruns.

Interestingly, levels of contract completeness continue to steadily increase even as change in contract failures plateaus after 2013. The broadly negative correlation between contract completeness and contract failures is not sufficient to infer strategic behavior. However, if strategic behavior existed, it would be consistent with classic contract theory and must justify steadily raising completeness despite diminishing returns in lowering contract failures.

5.2 Most variation in completeness is unexplained by contract-level details

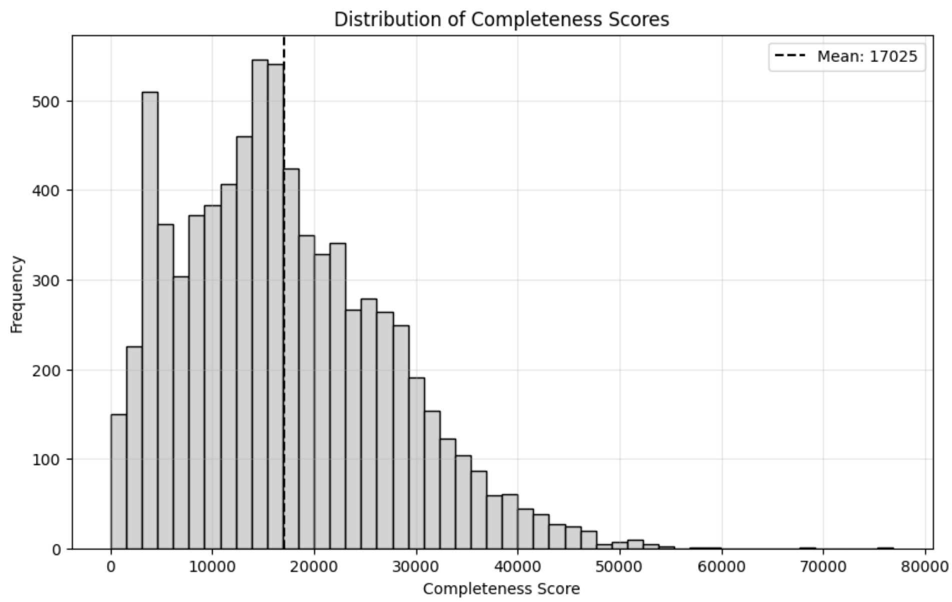

Figure 5.2: Distribution of Completeness Scores

The 7,725 contracts exhibit a wide distribution of completeness scores. Intuitively, knowing project-level information should help narrow the expected range of completeness. For instance, higher-priced contracts might be more complete, different project types (e.g., construction, maintenance) likely have varying baseline completeness levels, and larger-scale projects could be more complete. To test this, we used the following regression model to predict contract completeness:

$$
\begin{aligned}
\text{ContractCompleteness}_i = \beta_0 &+ \beta_1 \text{FiscalYear}_i + \beta_2 \log(\text{TotalPrice}_i) \\
&+ \beta_3 \log(\text{TotalPrice}_i) \times \text{ProjectType}_i \\
&+ \beta_4 \text{ProjectType}_i \times \text{ProjectScale}_i \times \text{ProjectUnit}_i + \varepsilon_i
\end{aligned}
\tag{5.1}
$$

This fairly well-specified regression yields an r-squared value of 0.331, indicating only 33.1% of the variation in completeness can be explained by these project-level predictors. This suggests that the majority of variation is likely due to unobserved factors or inherent noise.

5.3 Short-term adaptation of completeness to previous contract failure is not observed.

To examine the impact of past contract failures on completeness, I expanded the baseline regression model, which initially included only project-level details (fiscal year, total cost, project type, and project scale). I added four measures of previous contract failure: two at the vendor level (the number of contract amendments and total cost overruns for the vendor's previous contract) and two at the agency level (the agency's average number of amendments and cost overruns processed in the 90 days prior to the current contract's execution). The expanded regression model is as follows:

$$
\begin{aligned}
\text{ContractCompleteness}_i = \beta_0 &+ \beta_1 \text{FiscalYear}_i + \beta_2 \log(\text{TotalPrice}_i) \\
&+ \beta_3 \log(\text{TotalPrice}_i) \times \text{ProjectType}_i \\
&+ \beta_4 \text{ProjectType}_i \times \text{ProjectScale}_i \times \text{ProjectUnit}_i \\
&+ \beta_5 \text{NumPreviousAmendments}_i + \beta_6 \log(\text{PreviousCostOverruns}_i) \\
&+ \beta_7 \text{MeanAgencyAmendments}_i^{(90 \text{ days})} + \beta_8 \text{MeanAgencyOverruns}_i^{(90 \text{ days})} \\
&+ \varepsilon_i \qquad (5.2)
\end{aligned}
$$

Variable	Coefficient	Std. Error	t-Statistic	p-Value
Log.Past_90d_Cost_Overrun	-2.274	37.282	-0.061	0.951
Past_90d_Num_Amendments	0.357	0.073	4.904	9.60e-07 ***
Prev.Vendor.Number.of.Amendments	0.108	1.896	0.057	0.955
Log.Prev.Vendor.Cost.Overrun.Amount	7.826	5.589	1.400	0.162

Table 5.1: Regression Results. Non-contract failure terms omitted

The regression table presents the coefficients and p-values for each measure of past contract failure. While one of the four measures are statistically significant, none demonstrate a practically significant effect (the average completeness score is 17,025). This suggests that past contract failure does not have a discernible impact on subsequent contract completeness. In addition, the R-squared value of this expanded regression is 0.340, a minimal increase of 0.9% compared to the baseline regression's R-squared of 0.331, which only used project-level information.

5.4 Granular completeness scores are weakly positively correlated, but all increase over time.

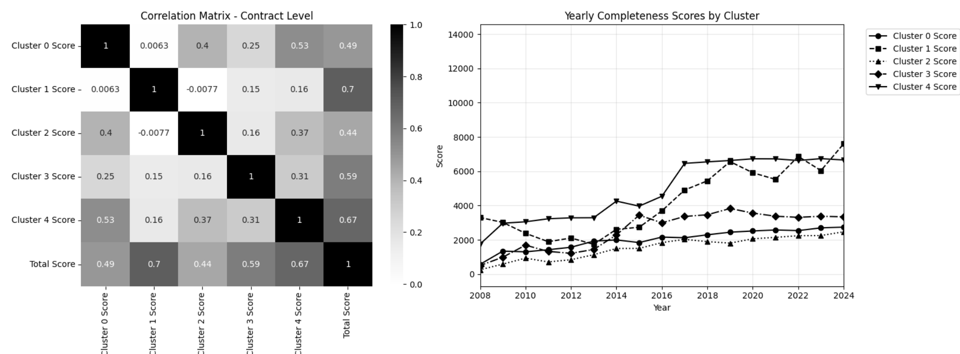

Figure 5.3: Granular Completeness Scores Correlation Matrix and Over Time

The completeness scores for different sections within a contract show weak correlations. For example, although a more complete insurance section tends to be associated with a more complete project details section, the relationship is not strong. Despite these weak correlations at the contract level, all section completeness scores exhibit an upward trend over time, mirroring the increase in the total completeness score.

In addition, the regressions presented in Results 2 and 3 were replicated using each section's completeness score as the dependent variable instead of the total score. These analyses yielded the same null results regarding adaptation to contract failures as observed with the total completeness score.

5.5 Months with high volume of contracts are not associated with less complete contract

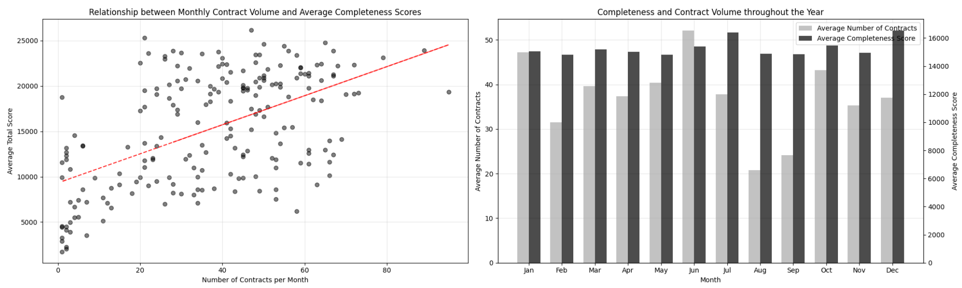

Figure 5.4: Relationship between Contract Completeness and Monthly Volume

Counterintuitively, average contract completeness for a given month tends to increase with the number of contracts processed during that month. This association is partially attributable to the rising contract volume at the Florida DOT over the past decade, which coincided with the increase in completeness documented in Result 1. However, even after controlling for the fiscal year by examining monthly data aggregated across years, the average contract completeness does not exhibit a significant relationship with the number of contracts processed in the same month.

This result is particularly interesting as it challenges a fundamental assumption of completeness theory: that writing complete contracts is costly. Given the practical constraints on government agencies' ability to perfectly adjust human resources to procurement demand, especially in the short term, we would expect higher contract volume in a month to lead to fewer resources per contract and, consequently, lower overall completeness. Surprisingly, this does not occur. This suggests an alternative mechanism maintains high contract completeness despite increased volume. Potential explanations include the prevalent use of boilerplate contracts or the efficient sharing of common clauses and contingencies among contracts written concurrently.

6

Discussion

This analysis is successfully completed on 7,725 contracts over 280 hours of computation—the largest sample size for any analysis done on completeness in public procurement. It begins to fill a practical field of study on if and how contract completeness is adjusted by government agencies and advances methodological approaches to measuring completeness. Furthermore, it sheds light on interesting implications and promising paths for additional research.

6.1 Implications for Current Theory

Implications of no adaptation to contract failures found.

This null result of no adaptation does not provide evidence for either classic completeness theory or strategic ambiguity theory. It suggests that agencies cannot and do not wish to adapt completeness in response to history. One plausible explanation is a lack of bureaucratic capacity, which can prevent accounting for past history or perceiving completeness as a feature to strategically adapt.

Notably, this exposes a blind spot not commonly addressed in theory: Debates in optimal completeness must be preceded by the assumption that completeness is controlled. While this somewhat obvious assumption is more likely true for private firms contracting a small number of agents, it is less tenable for state government agencies that procure thousands of contracts a year. This implies that contract design should be studied with respect to characteristics of the contracting body—especially at a public level.

Implications of completeness scores not decreasing in high-volume months

Both classic completeness theory and strategic ambiguity theory assume that complete contracts are costly, whether in ink costs or search costs. Fascinatingly, this assumption is contradicted in the data. Following this, contract completeness not decreasing in more cost-constrained environments suggests alternative mechanisms for how transaction costs are calculated.

One plausible explanation is text sharing between contract provisions written concurrently. If an agency is writing multiple contracts at the same time, it is virtually costless to copy and paste contract provisions from a similar contract. This repeated use of the same contract provisions may also drive the long-term trend of rising completeness scores in the Florida Department of Transportation—newer contracts may be more complete due to the increasing access to boilerplate provisions gained in previous years.

Another plausible explanation may be the use of "catch-all" contract provisions during times of high-volume. Instead of paying high search costs to discover relevant contingencies, agencies may use generic provisions that indiscriminately cover a wide range of contingencies.

These behaviors have important implications both theoretically and methodologically. It highlights that not all contingencies are equally important. Capturing weak or irrelevant contingencies can inflate completeness de jure but not fulfill the spirit of completeness. In a sense, catch-all or copy-paste contingencies may make a contract no more complete, even if the contract is longer and more detailed. Broadly, it implies a need to differentiate between the quality of contingencies specified, especially as a glaring flaw in current completeness measures.

6.2 Next Steps

Better methods of measuring completeness

The granular methodology developed in this thesis is compatible with any previous approaches to measuring completeness (manual, count-based, topic models, etc.). It works by preprocessing the contracts into groups of unified subjects before computing completeness for each group individually. Fundamentally, it is still limited by the validity of the completeness measured used in the final step. The most promising approach to measuring completeness is one that can differentiate boilerplate provisions. This type of measure relies on Tirole's definition of completeness as the amount of resources expended to identify appropriate design. Approaches that rely on identifying the number of contingencies present (count-based, LDA) suffer from not differentiating the quality or relevance of the contingencies included, as shown by our discussion on catch-all and copy-paste provisions.

Intuitively, a human can get a good sense of what is boilerplate and what is custom-written from inspecting a contract. Visually, there often are obvious templates with blank lines filled in with handwriting or a text entry. Dense, small text filled with "legalese" is more likely to be copy-paste text instead of custom-written text. Less perfectly detailed and edited provisions are counterintuitively a sign of novelness and greater effort expended into writing new text.

This process can likely be automated by a multi-modal LLM, given advancements in visual and contextual understanding. However, the question of output validity is far more critical for such a complex task. If attempted, this approach would require the careful fine-tuning of a generative model through manually creating a training and validation dataset instead of relying on zero-shot or one-shot performance. Though assuredly an arduous task, the technology for this method has only been practically possible in the last year and would result in likely the most insightful completeness measure to date.

Refinements to the analysis of Florida DOT data

This thesis offers original descriptive statistics for the procurement contracting of the Florida Department of Transportation, marking the first analysis of this data. The calculation yielded 7,725 overall completeness scores and a substantial 38,625 granular completeness scores. This rich dataset provides a strong foundation for analysis, yet further exploration is warranted.

The granular completeness methodology developed in this thesis is adaptable to any contract corpus, automatically clustering chunk topics into five groups based on semantic similarity. While this inherent flexibility is beneficial, deeper exploration could involve clustering chunks according to pre-defined categories. For example, one may want to group all environmental regulation-related chunks within a contract to assess its completeness in particular. This can be accomplished by adjusting the data pipeline to substitute the clustering algorithm for an embedding-based retrieval system.

Qualitative analysis involving procurement officials could complement the quantitative results presented. Employing surveys and interviews with agency personnel engaged in contract processes could reveal their preferences and practices. Such qualitative methods can reveal underlying intentions, even when they are not fully reflected in observed outcomes.

6.3 Conclusion

This thesis successfully processed 7,725 contracts, yielding important descriptive statistics on the evolution of Florida's public procurement within the Department of Transportation through the application of a novel preprocessing methodology. We found that contract completeness rose and contract failures fell in the long-term, though completeness still steadily increased while failures decline stagnated. In the short-term, no adaptation of completeness to contract failures was found. The findings stress the importance of establishing whether agencies are capable of strategically adjusting completeness before considering normative questions about how they ideally should do so. Lastly, this thesis invites further methodological research on assessing the quality of contingencies in completeness measures and theoretical work on how agencies economize on transaction costs during contract drafting.

References

[Alizadeh et al., 2025] Alizadeh, M., Kubli, M., Samei, Z., Dehghani, S., Zahedivafa, M., Bermeo, J. D., Korobeynikova, M., and Gilardi, F. (2025). Open-source LLMs for text annotation: a practical guide for model setting and fine-tuning. *Journal of computational social science*, 8(1):17–. Place: Singapore Publisher: Springer Nature Singapore.

[Anderson and Dekker, 2005] Anderson, S. W. and Dekker, H. C. (2005). Management Control for Market Transactions: The Relation Between Transaction Characteristics, Incomplete Contract Design, and Subsequent Performance. *Management Science*, 51(12):1734–1752. Publisher: INFORMS.

[Aubert et al., 2017] Aubert, B. A., Houde, J. F., Rivard, S., and Patry, M. (2017). Determinants of contract completeness for information technology outsourcing. *Information Technology and Management*, 18(4):277–292.

[Bajari and Tadelis, 2001] Bajari, P. and Tadelis, S. (2001). Incentives versus Transaction Costs: A Theory of Procurement Contracts. *The RAND Journal of Economics*, 32(3):387–407. Publisher: [RAND Corporation, Wiley].

[Bernheim and Whinston, 1998] Bernheim, B. D. and Whinston, M. D. (1998). Incomplete Contracts and Strategic Ambiguity. *The American Economic Review*, 88(4):902–932. Publisher: American Economic Association.

[Blei et al., 2003] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

[Crocker and Reynolds, 1993] Crocker, K. J. and Reynolds, K. J. (1993). The Efficiency of Incomplete Contracts: An Empirical Analysis of Air Force

Engine Procurement. *The RAND Journal of Economics*, 24(1):126–146. Publisher: [RAND Corporation, Wiley].

[D'Acunto et al., 2020] D'Acunto, F., Xie, J., and Yao, J. (2020). Trust and Contracts: Empirical Evidence.

[Ederer et al., 2018] Ederer, F., Holden, R., and Meyer, M. (2018). Gaming and strategic opacity in incentive provision. *The RAND Journal of Economics*, 49(4):819–854. Publisher: [RAND Corporation, Wiley].

[Ganglmair and Wardlaw, 2017] Ganglmair, B. and Wardlaw, M. (2017). Complexity, Standardization, and the Design of Loan Agreements.

[Goldberg, 1985] Goldberg, V. (1985). PRICE ADJUSTMENT IN LONG-TERM-CONTRACTS. *Wisconsin law review*, (3):527–543. Place: MADISON Publisher: Univ Wisconsin Law School.

[Greiman, 2010] Greiman, V. (2010). The Big Dig: Learning from a Mega Project | APPEL Knowledge Services.

[Hafsa et al., 2021] Hafsa, F., Darnall, N., and Bretschneider, S. (2021). Estimating the True Size of Public Procurement to Assess Sustainability Impact. *Sustainability*, 13(3):1448.

[Holmstrom and Milgrom, 1991] Holmstrom, B. and Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization*, 7:24–52.

[Huber and Shipan, 2002] Huber, J. D. and Shipan, C. R. (2002). *Deliberate Discretion?: The Institutional Foundations of Bureaucratic Autonomy*. Cambridge Studies in Comparative Politics. Cambridge University Press, Cambridge.

[Klein, 2002] Klein (2002). The Role of Incomplete Contracts in Self-enforcing Relationships. *The Economics of Contracts Theory and Applications*, pages 59–71. Publisher: Cambridge University Press.

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

[Laffont and Martimort, 2002] Laffont, J.-J. and Martimort, D. (2002). The Theory of Incentives: The Principal-Agent Model. ISBN: 9780691091846.

[Macchiavello and Morjaria, 2023] Macchiavello, R. and Morjaria, A. (2023). RELATIONAL CONTRACTS: RECENT EMPIRICAL ADVANCEMENTS AND OPEN QUESTIONS.

[MacLeod, 2007] MacLeod, W. B. (2007). Reputations, Relationships, and Contract Enforcement. *Journal of Economic Literature*, 45(3):595–628.

[Masten, 1988] Masten, S. E. (1988). Equity, Opportunism, and the Design of Contractual Relations. *Journal of institutional and theoretical economics*, 144(1):180–195. Place: TUBINGEN Publisher: J.C.B. Mohr Paul Siebeck.

[Moszoro et al., 2016] Moszoro, M., Spiller, P. T., and Stolorz, S. (2016). Rigidity of Public Contracts.

[Saussier, 2000] Saussier, S. (2000). Transaction costs and contractual incompleteness: the case of Électricité de France. *Journal of Economic Behavior & Organization*, 42(2):189–206.

[Scott and Triantis, 2006] Scott, R. E. and Triantis, G. G. (2006). Anticipating Litigation in Contract Design. *the yale law journal*.

[Setty et al., 2024] Setty, S., Thakkar, H., Lee, A., Chung, E., and Vidra, N. (2024). Improving Retrieval for RAG based Question Answering Models on Financial Documents. arXiv:2404.07221 [cs].

[Spier, 1992] Spier, K. E. (1992). Incomplete Contracts and Signalling. *RAND Journal of Economics*, 23(3):432–443. Publisher: The RAND Corporation.

[Tirole, 2009] Tirole, J. (2009). Cognition and Incomplete Contracts. *The American Economic Review*, 99(1):265–294. Publisher: American Economic Association.

[Williamson, 1979] Williamson, O. E. (1979). Transaction-Cost Economics: The Governance of Contractual Relations. *Journal of Law and Economics*, 22(2):233–261.