Determinants of Contract Completeness

Lea Kosnik
Department of Economics
University of Missouri-St. Louis
St. Louis, MO 63121-4499
Tel: 1-314-516-5564
Fax: 1-314-516-5352
kosnikl@umsl.edu

Keith Taylor
Research Division
Federal Reserve Bank of St. Louis*
St. Louis, MO 63106
Tel: 1-314-444-4211
Fax: 1-314-444-8731
Keith.G.Taylor@stls.frb.org

Abstract

There is a tradeoff that must be addressed any time a contract is written; whether to make a contract flexible but incomplete or rigid but comprehensive. This paper investigates the completeness of hydroelectric license contracts over three decades and finds that as environmental concerns increase, so does contract flexibility, ultimately confirming the predictions of transaction cost theory. The paper offers a look at the development of the U.S. hydroelectric license as it ages over time and responds to growing environmental concerns. It also, in a novel empirical application, combines traditional regression analysis with insights from textual analysis and computational linguistics.

JEL Codes: K2; Q4
Keywords: hydro, transaction cost theory, electricity, renewable energy, word cloud, textual analysis, environmental regulation

---

Acknowledgements: We gratefully acknowledge the input and assistance of John Whitehead, Lynne Lewis, Trudy Ann Cameron, Ian Lange, Sarah Tucker, and numerous participants and colleagues at meetings and presentations around the U.S.
* The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Introduction

There is a tradeoff that must be addressed every time a contract is written; whether or not to make a contract flexible but incomplete or rigid but comprehensive. A flexible contract allows room for adaptation to future unforeseen circumstances, but at the same time allows for potential opportunism by the contracting parties as states of the world evolve. This paper investigates the factors that determine the degree of flexibility chosen in a long-term contract to balance these competing concerns.

The fact that contracts exist at all in a market-based free exchange system was first elaborated on by Coase in his classic 1937 paper *The Nature of the Firm*, where he explained the existence of vertical integration of exchange, from contracts to mergers all the way up to government production, as a result of transaction costs. Over the years other seminal authors have operationalized the factors that underlie transaction costs, including asset specificity, uncertainty, duration, and probity (Klein et al., 1978; Williamson, 1979; 1999), while others have attempted empirical tests of the theory (Joskow, 1987; 1990; Crocker and Masten, 1988; 1991; Crocker and Reynolds, 1993; Saussier, 2000; Kerkvliet and Shogren, 2001; Neumann and von Hirschhausen, 2008; Kozhevnikova and Lange, 2009). A look through the literature, however, fails to uncover papers that have tested contract completeness per se. Data limitations have generally led to indirect or partial tests of contract completeness. This paper utilizes a unique dataset on U.S. hydroelectric license contracts that spans over thirty years, and tests whether the factors suggested in transaction cost theory indeed influence the actual completeness of a given contract.

The contributions of this paper are three-fold. First, it is a direct empirical test of contract completeness based on the theoretical predictions of transaction cost theory. Second, it offers an interesting historical look at the evolution of hydroelectric license contracts as they respond over a thirty year time span to changing environmental conditions and morphing societal welfare functions. And finally, and perhaps most novel empirically, this paper arrives at its results through a combination of traditional regression analysis and computational linguistics, including textual analysis, word clouds and new forms of digitized quantitative analysis that are only just beginning to be utilized in the wider academic literature (Grimmer and King, 2010; Michel et al., 2011; Evans and Foster, 2011).

Our results indicate that hydroelectric license contracts have grown more incomplete over the last three decades, confirming the general predictions of transaction cost theory. As environmental uncertainty regarding the implications of hydroelectric power production has increased, so too has the incompleteness and flexibility of hydropower license contracts. In the tradeoff between flexibility and rigidity, flexibility has won out as environmental concerns have dominated asset-specific hold-up fears. The regression results provide the primary evidence for this result; the textual analysis is more nuanced. Textually, increasing contract flexibility is mirrored in increasing average sentence lengths, however frequency analysis of the words themselves do not show obvious support for a changing flexibility interpretation of the contract provisions. This result appears to highlight the importance of clear contractual terms *stating* contract flexibility, rather than subtle word choice changes that could, in fact, be prone to future legal misinterpretation anyway. Ultimately, the results provided in this paper provide direct evidence that the relative completeness of a given hydroelectric license contract can be explained by such transaction cost theory based factors as uncertainty, asset specificity, and reputation.

The rest of this paper is organized as follows: in section 2 we give an extended literature review on transaction cost theory and its implications for contractual completeness, in section 3 we develop our model and primary hypothesis, section 4 gives background on the hydroelectric license contracting process, section 5 presents our data, section 6 the results from the regression analysis, section 7 the results from the textual analysis, and section 8 concludes.

Literature Review

The early literature on contract design was pioneered by Coase (1937) who mulled over the rational existence of the firm in a world of free market-based exchange. He came to the conclusion that firms exist because the vertical integration of exchange they embody can save on costs; in other words, it may be cheaper to produce some items within a firm than through a series of individual market-based exchanges. Goldberg (1976) added a vociferous justification of such vertical integration, all the way up to the government level, when the difficulties embodied in production are so great they may be inherently unmanageable any other way. Klein et al (1978) and Williamson (1979; 1999) began the task of operationalizing the key variables that led to the need for some form of vertical integration, be it at the firm or government level. Klein et al (1978) in particular stressed the hold-up problem, where one party to an exchange makes production-related investments whose value is tied to the exchange relationship proceeding, but then finds that the other party may try to take advantage of the situation by holding up production and expropriating production related rents. Williamson (1979; 1999) identified other key factors, including the uncertainty and complexity of the trading environment, the time duration of the exchange relationship, and the probity of the production good under consideration. These four categories – asset specificity, uncertainty, duration, and probity - are the primary categories of transaction costs that are regularly referred to when discussing transaction cost theory and its effects on contractual exchange relationships.

There are a number of predictions that stem from this work on transaction cost theory, including that as the asset specificity embodied in a contract increases, the rigidity or completeness of the contract should increase as well in order to protect the parties from any potential hold-up problems in the future. Another prediction is that as the uncertainty surrounding a contracting environment increases, so too should the ultimate flexibility (or incompleteness) of a contract in order to allow for an efficient response to unexpected, unfolding states of the world. Similarly, as the duration of a contract increases, the prediction is that flexibility should also increase in order to allow for efficient responses to increasingly distant and unknown future states of the world. And finally probity implies that contracts can be written with greater flexibility as the contracting parties increasingly trust one another.[1]

Empirically, these specific hypotheses have not been subject to robust tests in the literature. To date, the empirical literature on transaction cost theory has primarily investigated tests of contract duration. For example, Joskow (1987), Crocker and Masten (1988), Kerkvliet and Shogren (2001), Neumann and von Hirschhausen (2008), and Kozhevnikova and Lange (2009) have all utilized contract duration as the dependent variable and tested whether or not contract length is affected by things such as asset specificity, degree of opportunism, and degree of regulatory interference. The context of these empirical tests varies from coal to natural gas, but the duration dependent variable remains consistent. A few studies do exist which get closer to literal tests of contract completeness (Joskow 1990; Crocker and Masten 1991; Crocker and Reynolds 1993), but the dependent

[1] Probity was initially highlighted in its relation to government-based contracts, but in a more general sense it is related to integrity and reputation, which has been stressed as an important determinant of contractual outcomes in, for example, Banerjee and Duflo (2000).

variables in all of these studies relies heavily on pricing provisions. Pricing provisions for exchange of the good in question is only one possible aspect of contract flexibility. Using pricing provisions alone to test for contract completeness in a comprehensive sense is specious. The only empirical study in the literature that we are aware of which tests for contract completeness with a dependent variable that is more broadly based is Saussier (2000). The main problem with this study, however, is that the empirical results are based on just 29 observations, which is a rather small sample size. Our paper appears to be the first in the literature that tests the flexibility/completeness predictions of transaction cost theory with a dependent variable that measures direct flexibility contractual provisions. The results are then given an additional empirical check with textual analysis and other computational linguistic techniques.

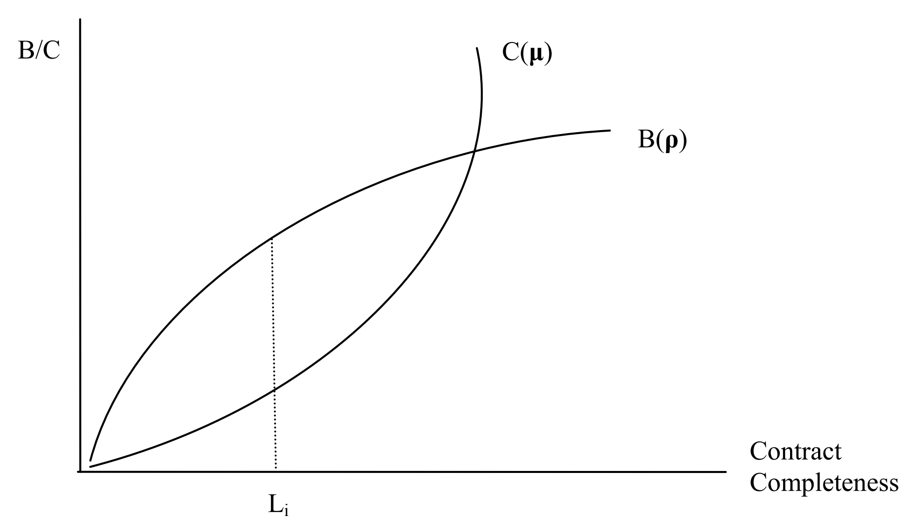
Model

The level of completeness embodied in a given contract, *i*, depends on the tradeoff chosen between contract flexibility (which allows parties to respond more efficiently to events as they unfold in the future) and the potential for opportunism (which stems from the hold-up problem elaborated on by Klein et al (1978)). This can be represented, as in Figure 1 below, as a benefit-cost tradeoff calculation that results in a given level of contract completeness, L_i.

Specifically, the benefit curve depends on a vector of characteristics, $\boldsymbol{\rho}$, that indicate the degree of asset specificity involved in a contracting relationship. It includes large capital costs, human capital investments, and geographic specificity that make the trading partners reliant on each other and subject to potential rent extraction from opportunism later. The higher the level of asset specificity, the greater the value of the benefit from contract completeness due to reduced opportunism later. This is why the benefit curve is increasing but at a decreasing rate. Should the overall level of asset specificity in a trading environment increase (decrease), this would cause a pivot of the benefit curve and a subsequent increase (decrease) in the level of contract completeness, L_i.

The cost curve depends on a vector of characteristics, **μ**, that indicate the measure of uncertainty inherent in a contracting environment. They include such things as product quality uncertainty, political regulatory uncertainty, and technological and environmental uncertainty. The more complete a contract (the higher the level of L_i), the greater the contracting costs from these variables will be, as there will be less room for flexible responses to them as the contract plays out. This is why the cost curve is increasing and at an increasing rate. As well, should the level of uncertainty embodied in these factors increase (decrease), this would cause a pivot of the cost curve and a subsequent decrease (increase) in the level of contract completeness, L_i.

Figure 1: The Contracting Completeness Decision



This model of the contracting completeness decision results in the following testable hypotheses:

> *Hypothesis 1: As the uncertainty in a contracting environment increases (decreases), the level of contract completeness should fall (rise).*

> *Hypothesis 2: As the asset specificity in a contracting environment increases (decreases), so too should the level of contract completeness.*

These two hypotheses relate directly to predictions from transaction cost theory. Transaction cost theory, however, also has something to say about duration and probity, namely, that as the duration of a contract increases, the level of contract completeness should fall, and that as the probity implicit in a contracting environment increases, the level of contract completeness should fall. These two additional predictions can be embodied in the hypotheses stated above. Duration in a hydroelectric licensing contract, as will be described below, is determined in a perfunctory manner on capital costs, therefore in this context in particular, duration is akin to asset specificity, which implies a direct relationship to hypothesis 2. Probity, in this context, can be interpreted as reputation and a strong reputation, the literature suggests, reduces uncertainty. Probity, therefore, is akin to certainty, which implies a direct relationship to hypothesis 1. In our particular empirical context, therefore, we can test the four traditional operationalizations of transaction cost theory (asset specificity, uncertainty, duration, and probity) as embodied in the two hypotheses stated above.

Hydroelectric License Contracts

In the United States today there are approximately 84,000 dams.[2] These dams serve a variety of purposes, including flood control, irrigation, navigation, and recreational purposes. Less than 4% of these dams, or only about 3,000 in total, support hydroelectric production.[3] These hydroelectric dams have a variety of owners, including the federal government, non-federal public municipalities, private utilities, private generation companies (also called "non-utilities"), private industrial owners, and electric cooperatives. All dams not directly owned by the federal government (i.e. "nonfederal") must be periodically licensed by the Federal Energy Regulatory Commission (FERC) in order to legally operate. This licensing process evolved in the early 20th century in response to the tradeoff posed in constructing large hydroelectric dams. Traditional hydropower dams have large fixed costs, which, by definition, require a number of years of positive revenues in order to recoup. At the same time, hydropower dams are, by necessity, built along public streams and rivers over which the public likes to maintain a degree of control. In the early 20th century the country was growing rapidly and new sources of power to satisfy its mounting energy needs were in great demand. The government sought to encourage hydroelectric power production, but private investors were hesitant to build without assured control of their assets. The conundrum in Congress, therefore, was how to encourage private hydroelectric power development, while at the same time not relinquishing societal input and ultimate control over the nation's river systems.

In 1920 Congress passed the Federal Water Power Act (FWPA), the first piece of legislation to formally institutionalize the hydropower licensing

[2] As documented by the National Inventory of Dams: http://geo.usace.army.mil/pgis/f?p=397:1:1670802410045182::NO:::

[3] Either as the sole purpose of the dam, or as an ancillary purpose to something else, for example, flood control.

responsibilities of FERC.[4] It achieved a balance between public and private objectives by creating lengthy licensing terms of at least 30, but not more than 50, years. Essentially, all nonfederal hydropower dams became subject to intense public oversight once every 30-50 years, but after they were licensed, they were essentially left alone from new contracting authority for decades at a time. This was considered sufficient to maintain a degree of public control over the river system, while at the same time allowing enough independence to encourage private hydropower development.

The licensing process initially instituted by FERC had a well-defined procedural schedule (FERC 2004), involved the dam owner and public interest groups in the input process, and resulted in a formal contractual license that was in effect for 50 years for brand new dams, 30 years for relicenses of existing dams, and 40 years for relicenses of existing dams that had undergone major capital improvements. The duration of a license varied little within these set 50, 40, or 30 year parameters,[5] and licenses themselves were almost always issued once a project had begun construction. In the dataset this paper covers (all licenses and relicenses issued by FERC between 1977-2007), approximately fourteen percent of the observations are for original licenses, while the vast majority are for relicenses.[6] Of the 1,343 projects to have come up before FERC over this time span, only one was ever denied. In other words, our dataset does not suffer from bias due to observations that never completed the licensing process – nearly all projects that come before FERC *do* get a new license. The licensing process is not pro forma, however, it is a complex and at times contentious process, but that is because of what goes *into* a license, not because

[4] Which, at that time, was called the Federal Power Commission (FPC).
[5] The duration of a license could be affected by acceleration requests due to transfers of ownership or basin-wide management concerns, but this has been, and still is, infrequent.
[6] Most hydropower dams in the U.S. were built just after WWII, and very few are being built now. There was a small spurt of new hydropower facility construction in the early 1980s after the first oil shock, but by the 1990s new dam construction had again fallen off.

there is generally a great deal of debate over whether or not to issue a license at all.

The issues that are generally debated in a licensing application have changed over time. Early in FERC's licensing days, the important points over which licenses varied concerned safety and structural integrity issues of the dam, its electric power facilities, and the reservoir. Rarely were environmental concerns over water quality, ecosystem management, aesthetic appearances, and other more contemporary environmental issues considered. This began to change in the 1980s as the modern environmental movement in the U.S. gained momentum. Beginning in the 1980s, when nonfederal dams came up for licensing, the debate increasingly emphasized environmental management issues, reflecting society's changing preferences for more environmental protection. Research in the law literature suggests this change may have been accompanied by increasingly flexible license provisions (Pollak 2007). Safety concerns are relatively straightforward and can be addressed by closed-ended provisions targeting engineering and technical integrity checks. Environmental issues, however, are more complex. The biological processes behind many environmental factors are still poorly understood, and adequately protecting things like fishery resources and water quality is difficult to implement at best. Adaptive management, a common methodological tool for implementing environmental protection measures, is by definition based on flexibly responding to the changing and unpredictable biological requirements of an ecosystem over time. At first blush, therefore, we would expect hydroelectric power licenses to have become increasingly flexible over the past thirty years in response to the increasing emphasis on environmental issues, but this isn't the end of the story.

An understanding of transaction cost theory tells us that uncertainty (biological or otherwise) is only one factor that can affect contract completeness. Asset specificity, in particular, is another important determinant and hydropower projects in particular involve significant levels of asset specificity. Hydropower dams are large, capital intensive, immovable projects that imply a substantial degree of illiquid investment on the hydropower owner's part. Indeed, the marginal costs of hydropower production and management are relatively minor compared to the large costs of hydropower plant construction, including the dam and the power generation facilities. Hydropower owners, then, would be expected to push for a high degree of contract completeness in any specific licensing procedure in order to avoid the rather significant hold-up problems that could occur later.

In essence, therefore, there are conflicting motivations in the hydropower licensing process concerning contract completeness, and they stem primarily from asset specificity on the dam owners' side, and biological uncertainty on the environmental side. These conflicting motivations would imply that the level of contract completeness in any particular licensing application is dependent on the tradeoff between these two forces. We investigate the level of actual contract completeness in hydropower license contracts that resulted over the last thirty years in the following empirical section.

Data

Our data come from a number of sources. The primary information on hydropower projects comes from a FERC maintained database, the Hydropower Resource Assessment database, or HPRA. It lists all the hydropower projects that have been licensed or relicensed by FERC since its inception in 1977.[7] The

[7] Before 1977 FERC was organized as the FPC (the Federal Power Commission). To avoid any impact of this organizational change on our results, our dataset begins with FERC's creation in October, 1977.

HPRA includes not just a list of project numbers, but a range of key descriptive characteristics for each of the projects listed including, for example, dam height, project owner, and physical location of the project. It is from this dataset that we get most of the variables which represent asset specificity in our model. Specifically, *kW* measures the available kW production capacity at a project. *Multiple Dams* is a dummy variable indicating whether a project (and thus a project license) is composed of more than one hydroelectric dam. The majority of projects consist of only one dam (83%), but there are some projects that are large, sprawling facilities with multiple dams and these receive a positive *Multiple Dams* value. *Max Dam Height*, measured in feet, is a measure of the tallest (if there is more than one) dam on site. *Plant Type* is a dummy variable indicating whether a project is a run-of-river facility. Run-of-river facilities (as opposed to peaking facilities) are restricted to generating power as the river flows, rather than at peak times of day when energy demand (and thus energy prices) are highest, and are therefore less profitable. As run-of-river facilities have less financial value they may be viewed as having less asset specificity. And finally, we include the following geographic variables: *Portland, San Francisco, Atlanta, New York,* and *Chicago.* These are the locations of the five regional offices of FERC and there is no real organizational difference between them, but they do process the paperwork and license filings for the projects within their district (Figure 2), and as such serve as a convenient indicator variable for geographic region.

Our next set of variables is designed to test for environmental concern, or uncertainty, in a licensing process. They include *Wild Scenic*, which indicates whether the river upon which the project is built has been listed under the federal Wild and Scenic Rivers Act of 1968 anytime before (and including) the year in which the project filed for a new license. *Endangered Species* is a count of the number of endangered species (as determined under the federal Endangered

Species Act of 1973) listed in the basin of a project in the year in which it applied for a new license.[8] *Government Land* is a dummy variable indicating whether any part of a project is located on government (federal, state, or city) land. Presumably, projects not entirely on private property may undergo greater environmental scrutiny. Finally, *New License* is a dummy variable indicating whether a license is for an entirely new project, as opposed to a relicense for existing hydropower facilities. As mentioned earlier, the majority of licenses (86%) are actually relicenses and few of our observations are for entirely new projects, but it may be the case that those few new projects are held to a higher environmental quality standard, and so we test for this in the empirical analysis.

We also include a few variables in our empirical analysis that test for a reputation effect on contract completeness. The theory is that stronger reputations and greater trust among contracting parties can lead to more flexible contracts as the parties have less fear of opportunism in the future. *Multiple Owner* is a dummy variable indicating whether the owner of the project seeking a license owns more than one FERC regulated hydroelectric project or not. It indicates a recurring role in repeated licensing processes over time and thus implies that a positive reputation may be relatively more valuable to a repeat player than to a player that only engages in the licensing process once every forty years or so. Approximately 56% of project owners are multiple owners. *Industrial* and *Utility* are indicators for the type of owner of a hydroelectric project (the alternatives being nonutility, municipal, and cooperative). Industrial owners (for example, pulp and paper plants or chemical factories) and utilities, whether justified or not, are often viewed by environmental groups involved in the licensing process as

---

[8] We broke this variable down into All Endangered Species and Mammals Only Endangered Species in order to determine if such a distinction made any difference to our results. There is some evidence in the literature that people respond to the fate of (often cute, furry) mammals more than they do to all species in general. In fact, the distinction made no difference to our results so in the empirical section which follows *Endangered Species* refers to the All Endangered Species variable.

more market-driven and profit-focused than other types of owners, and therefore, potentially less trustworthy. *ALP-ILP* indicates the type of regulatory procedure a license underwent at FERC. There are three licensing procedures available at FERC, the traditional licensing process (TLP) which was used historically by all projects prior to 1998, the alternative licensing process (ALP) which came into use on a voluntary basis after 1998, and the integrated licensing process (ILP), which became the new default licensing process at FERC after 2005. A more detailed description of the background and distinctions between these licensing procedures can be found in Kosnik (2010), but what is important to understand here is that both the ALP and ILP processes are designed to be cooperative and to encourage friendly relations among the involved negotiating parties at an early stage in the licensing process. The ALP and the ILP processes were developed in part *because* the TLP was often seen as contentious and unhelpful in fostering communicative relationships among the negotiating parties. Parties that are involved in an ALP or ILP process, relative to a TLP process, may therefore be viewed as more friendly and trustworthy. Finally, *OI* is a count of the number of official intervenors involved in a project licensing. Anyone can become an official intervenor to a licensing process if they file the paperwork with FERC in a timely manner. Official intervenorship status grants a party the legal right of appeal if they disagree with the final relicensing decision of FERC. Official intervenors tend to be environmental groups with a stake in the licensing outcome and the more of them there are in any particular licensing process, the more contentious that licensing procedure can be expected to be.

Duration is another variable frequently tested with respect to its relationship on contract completeness. As discussed earlier in the background section, the duration of a hydropower license is for the most part restricted to predetermined 50, 40, or 30 year time spans, depending solely on the level of recent capital investments in the project. Although this variable is not therefore determined by the negotiating parties (as traditionally assumed in transaction cost theory), we include it anyway as a test of the theory that the longer a license is expected to last, the greater the degree of flexibility it will embody.

We include three decade dummy variables, *1980s*, *1990s*, and *2000s* to indicate whether a project was licensed in the 1980s, the 1990s, or the 2000s. These variables account for possible time effects on contract completeness, which could include anything from changing societal preferences affecting license contract provisions, to changing FERC administrators over the decades. We have no a priori predictions on the signs of these variables in the final regression analyses.

In addition to the HPRA database and the public information we gathered on endangered species numbers and wild scenic rivers status, we also gathered together a hard copy of every one of the actual licenses issued by FERC from 1977-2007.[9] A hydroelectric license is generally structured so that the background information and licensing process discussion is at the beginning of the license, and the end of the license contains the mandatory requirements, or "Articles," a license is subject to in satisfaction of its terms. As mentioned earlier in the background section, whether or not to issue a license at all is rarely up for debate, what is up for discussion between the negotiating parties are the mandatory Articles at the end; what they will contain and how strict they will be. Our first dependent variable, *Reopener Clauses*, was created by reading through the environmental Articles at the end of every license, and coding each one as flexible or not depending on whether it contained a reopener clause, and then adding up all the reopener clauses in a particular license to arrive at a final

[9] FERC records indicate that 1,343 licenses were issued between 1977 and 2007. We were able to get copies of 1,334 (or more than 99%) of them.

variable count.[10] A reopener clause was always easy to identify. After an Article laid out its terms (be it to require a plan to improve water quality by so many parts per billion, or a requirement to implement specific ramping rates to limit water flow variability), an Article either said something to the effect of, "The Commission reserves the right to require changes to the plan," or it didn't. The language was always straightforward and always clear, and an article either included a reopener clause to this effect or it didn't. In other words, there was very little subjectivity in identifying and coding these reopener clauses. The summary information in Table 1 indicates that the average license had 4.5 reopener clauses, but that a license could have as few as zero, or as many as 28 distinct reopener clauses.[11]

Our second dependent variable, *Average Length,* begins our foray into computational linguistics by using the average word length of a license's Articles as a proxy for contract completeness. The variable is created by counting the number of words in each Article in a license and then averaging to come up with the average word length of the particular license's Articles.[12] Table 1 indicates that the average article length was a bit over 179 words, but that there were articles as terse as 40 words, and others quite verbose at 1,088 words. The assumption behind the use of *Average Length* as a dependent variable is that inflexible Article provisions tend to be terse and straightforward, while flexible Article provisions tend to be wordier in order to get the contingencies across as to

[10] Note that we did not include in our *Reopener Clauses* counts the standard, pro forma articles included in every license on things like annual payments to FERC, terms and conditions affecting interstate or foreign commerce, and the granting of authority to use project lands for common purposes.

[11] We also created a dependent variable, *Percent Reopener Clauses*, which was the percentage of total Articles that contained reopener clauses, and the results were not fundamentally different from what is reported in Table 2.

[12] When a license wasn't available to us in Word or txt formats (where one can easily use installed word-count features), a license was turned into text using OCR (optical character recognition) software.

when the Article applies and in what particular states of the world. We state clearly that this is an assumption - that the wordier the license Articles are the more flexible the contract is - but after reading tens of thousands of license Articles we believe it to be an accurate assumption in this context.

Regression Analysis

Our regression analysis is of the form:

$$L_i = f(\boldsymbol{\rho}_i, \boldsymbol{\mu}_i) + \epsilon_i$$

where L_i is defined (as in the model section above) as the level of contract completeness embodied in a particular hydroelectric license. $\boldsymbol{\rho}_i$ is a vector representing the set of variables indicating asset specificity and reputation effects in a particular licensing process, and $\boldsymbol{\mu}_i$ represents the set of variables that indicate uncertainty, as well as duration, in a license. Table 1 presents the summary statistics for the variables used in the regression analysis.

Table 2 presents the OLS regression results for both dependent variables, *Reopener Clauses* and *Average Length*.[13] The results across both dependent variables are remarkably similar.[14] The first, somewhat surprising result, is that the size asset specificity (ρ) variables are for the most part insignificant. *kW*, *Multiple Dams*, and *Plant Type* are all insignificant, across both regressions. Only *Max Dam Height* shows some significance, and that in only one of the

[13] We ran *Reopener Clauses* with a poisson regression as well and the results were fundamentally similar. Only OLS results are reported to facilitate comparisons between the two dependent variables, *Reopener Clauses* and *Average Length.* We also ran the regressions with the dependent variables as natural logs – to avoid any potential outlier effects – and the results were again fundamentally similar.
[14] The correlation between *Reopener Clauses* and *Average Length* is high at 0.696.

regressions. This implies that size asset specificity is not an influential factor in contract completeness, despite the predictions of transaction cost theory. Why that might be so is unclear. One possibility is that while hydroelectric dams are large, capital intensive projects, the ones analyzed in this paper are for the most part over half a century old. The majority of the licenses issues by FERC were, again, for relicenses, not brand new projects. By the time they came up for relicensing in our dataset the initial capital costs had likely been recouped for these projects and their owners may have felt less concern about threats from potential opportunism on their bottom line.[15]

The second interesting finding from our empirical results is that the environmental uncertainty (μ) variables most definitely do matter. The more environmental concern and uncertainty there is with respect to a particular project (as measured by increased presence of endangered species, *Endangered Species,* location of project on a Wild or Scenic River, *Wild Scenic*, or location of a project at least partially on government land, *Government Land*) the greater the flexibility a license contract eventually embodies. This accords with transaction cost theory and Hypothesis 1 of our model which states that as uncertainty rises, contract completeness should fall (i.e. flexibility should rise). The one μ category variable that does not show any significance, across both regressions, is *New License.* FERC is in fact legislatively mandated to treat relicense applications as equivalent to new license applications (*Yakima Indian Nation vs. The Federal Energy Regulatory Commission*, 746 F. 2d. 466), the idea being that a relicense, since it only comes once every 30-50 years, should be treated as a blank slate and approached with a fresh outlook on regulatory and management practices related to the hydropower facilities as well as the river system. It appears from the

[15] For long term owners of relicenses, in other words, the capital costs may have been viewed as "sunk," and thus unimportant in forward-looking thinking and decision-making. Geographic asset specificity would still hold, but for some owners at least capital based asset specificity may have been of less importance.

insignificance of *New License* that FERC is indeed not making any kind of a noticeable distinction between relicenses and new licenses, at least when it comes to the degree of contract completeness the final license issued embodies.[16]

The effect of reputation on contract completeness is a little less straightforward. For the most part the reputation variables are insignificant, with *Multiple Owner* and *Utility* showing insignificance across both regressions, and *Industrial* and *ALP-ILP* coming up insignificant in at least one regression each. Only *OI* is significant across both regressions, but it is strongly significant at the 1% level and implies that the degree of contract completeness is affected by the number of intervenors involved in a licensing process. The more official intervenors there are in a process, the more flexible the final license issued ultimately is. As official intervenors are for the most part environmental interest groups, and environmental interests embody uncertainty (Hypothesis 1), this makes sense.

The *Duration* coefficient is also strongly significant across both regressions, indicating that as the length of a license increases, so does its flexibility. This conforms with the predictions of transaction cost theory.

Finally, the decade dummy variables, *1980s*, *1990s*, and *2000s* as well as the geographic region location variables, *Portland, San Francisco, Atlanta* and *New York*, are all strongly significant across the regressions. The decade variables indicate that contract flexibility steadily increased from the 1970s, through the 1980s, and into the 1990s and 2000s.[17] This mirrors society's preferences for environmental activism and protection which has increased from the 1970s through to today. FERC, therefore, in issuing hydroelectric licenses does appear to be responsive to the period context in which they are issued. The

---

[16] This indistinction result between new licenses and relicenses does not always hold, however. Evidence presented in Kosnik (2010) seems to show a grandfathering effect by FERC in relation to the number and type of environmental regulations mandated per license.

[17] The same trend is evident using decade dummies or individual year dummies.

mechanism through which this is achieved is uncertain (be it through the changing of FERC administrations over the years, changing political pressure and oversight from Congress, or something else), but the effect appears strong. The reason for the significance of the geographic location variables is also somewhat unclear, but one hypothesis is that coastal regions (the comparison region in the regressions is Chicago) may end up with less flexible contracts (all of the coefficients are negative) because coastal states prefer less ambiguity when it comes to environmental protection measures. The environmental movement in the U.S. did stem from coastal regions like Portland, San Francisco, and New York, and has generally been less prevalent than in the Midwest regions around Chicago.[18]

Overall, the result that stands out most clearly from this empirical analysis is that environmental concerns have dominated hydroelectric license contract provisions over the years from 1977-2007. Other predictions of transaction cost theory also hold some ground, in the case of duration and reputation effects, but by far the most important factor in influencing contract completeness in this context is environmental concern and uncertainty. We continue to explore this effect in the textual analysis which follows.

Textual Analysis

There has been an explosion in the availability of data in recent years. Creative researchers from such disparate fields as psychology, medieval literature, and finance have mined resources that include Google clicks (Choi and Varian, 2009), twitter feeds (Bollen et al., 2011), and YouTube videos (Patel et al., 2009; Schachner et al., 2009) to investigate a wide range of behavioral and historical

[18] Alternative geographic specifications isolating West coast, East coast, midland, and other groupings all show the coastal states having less flexible contracts than the inland states.

questions. One particular form of analysis this new data deluge allows is what is called "textual analysis." Textual analysis involves the accumulation of a large amount of text (from digitized books, online message boards, or twitter feeds, for example), cleaning and parsing the text with unique algorithms, and then turning the text into a database where the words themselves are statistically analyzed for trends and correlative patterns (Grimmer and King, 2010; Michel et al., 2011; Evans and Foster, 2011).

Interesting examples of recent textual analyses include an investigation of culture from Top Ten song lyrics (DeWall et al., 2011), gender identification in literary styles (Koppel et al., 2011), media slant in newspapers (Gentzkow and Shapiro, 2010), and bargaining power in US-American Indian treaties (Spirling, 2010).

For the most part, published academic studies involving textual analyses have been in the humanities, political science, and computer science fields. While there have been a few notable finance papers, such as on the ability of stock message boards to predict the Dow Jones Industrial Average (Antweiler and Frank, 2004) and the effect of negative words in the *Wall Street Journal* on stock market returns (Tetlock, 2007), and some entertaining non-academic economic analyses (Krugman, 2011; Gerow and Keane, 2011), as far as we are aware, textual analysis in the economics literature more broadly is still very much in its infancy. This paper appears to be one of the first to utilize robust textual analysis as a significant methodological component in its empirical analysis.

Our textual analysis is based on a corpus of the complete set of licenses issued by FERC between 1977-2007. Digital copies of each of the licenses issued in that timespan were gathered and turned into a text file. The text was then cleaned, for example by removing page numbers, copyright information from the supplying agency (i.e. Westlaw or LexisNexis), and unnecessary addendums and attachments until we had the plain license itself.[19] Then we identified individual, unique words in the textual documents and entered them into a database. This parsing of the text was based on the techniques described in Michel et al. (2011), with only a few deviations.[20,21] We also maintained an association of the text of each license with metadata about the license itself, such as where the project was located (*Portland, San Francisco, Atlanta, New York,* or *Chicago*), the year the license was issued, or the size of the project (*kW*, *Multiple Dams*, and *Max Dam Height*). This way, we were able to analyze the license text, limited by particular metadata values. For example, we could generate word frequencies for all the licenses issued to projects in San Francisco, or, for all the licenses that were for projects having a capacity greater than 2,000 kW.

[19] While we maintained the full license text, the frequency analyses described below are based only on the text at the end of the licenses, in the Articles sections.

[20] Specifically, we deviated in two ways. First, we did not include contractions in our corpus. A list of contractions and their unambiguous full forms were identified and converted. Contractions with ambiguous full forms, such as "what's" were identified and converted by hand using the context of the sentence. Second, we used the Porter stemmer (Porter, 1980: http://tartarus.org/martin/PorterStemmer/) to stem the words, while also retaining their original form in the database.

[21] Note that, in line with other papers in the academic literature, we also did not include into the database common words such as "a," "of," or "the," and we maintained the 10% threshold whereby the absolutely least common words (generally those that just appeared once) were deleted in order to make the database more manageable.

Figure 3: Frequency Analysis on Entire Corpus

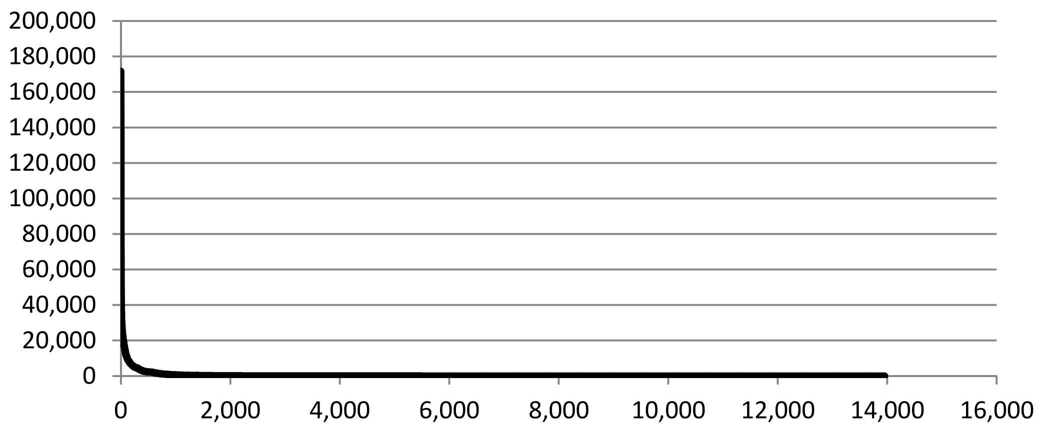


Figure 3 presents a frequency analysis on the entire body of text from all of the licenses put together. Frequency distributions of most texts seem to follow a power law, whereby there is a long tail of words (to the right of the graph) that appear very few times, and a few words that dominate (the left side of the graph), and this appears to be true for our text as well. The most common word in the entire corpus of text is "shall" and it appears 171,634 times. There are nearly fourteen thousand distinct words in the corpus itself. Frequency distributions on portions of the text, for example text limited by year of license issuance or by project geographic area, all also exhibit distinct power law distributions.[22] What this tells us is that all inquiries into the text are dominated by key words, and that the majority of the words in any given body of text are actually used rather infrequently. This is helpful with regards to the textual analysis as it allows us to focus on a smaller body of words - the ones that occur with more frequency – in the analysis.

[22] Individual graphs available upon request.

What are we looking for in our textual analysis of hydroelectric power licenses? First and foremost, we are looking for evidence of flexibility/rigidity changes in word choice over time, perhaps a decrease in the use of words such as "must" or "shall," and a concomitant increase in the use of the words such as "may" or "might." We are also interested in teasing out environmental emphasis, by noting the relative frequency over time of words such as "environmental," "fish," and "pollution." Ultimately, through frequency tests and other tools we hope to get a feel for how the body of text in hydroelectric power licenses may have subtly changed over time and in relation to specific project characteristics, if at all.

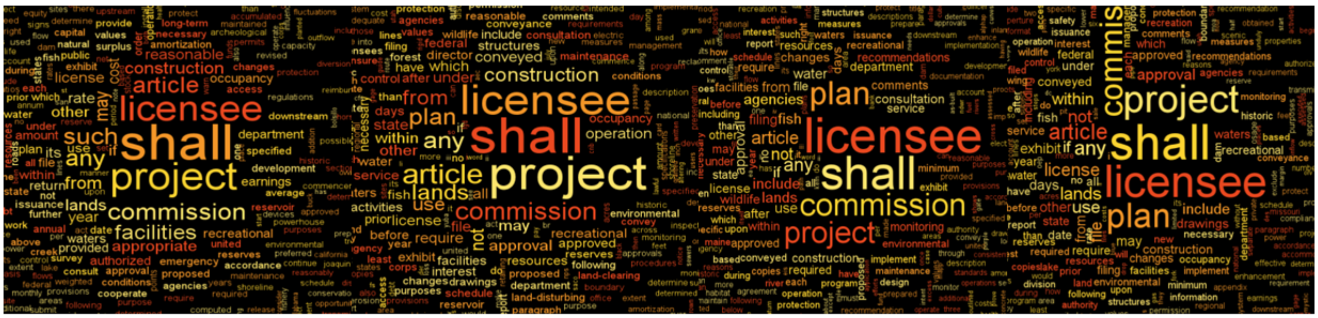
Figure 4 is a word cloud generated from the text in all the licenses issued in 1978. A word cloud is a visual representation of word frequency, with the most frequent words appearing in large font in the center of the figure. As words decrease in frequency they are pushed out from the center of the figure and appear in increasingly smaller font. Word clouds allow for quick visualization of relative word importance in a particular body of text. For 1978 it is clear that formal words, such as "licensee," "project," "shall," and "commission" appear with greatest frequency. Words such as "environmental, "fish," and "pollution" are hard to even locate.

Figure 4: Word Cloud for Licenses Issued in 1978

Figure 5 displays the word clouds for licenses issued in 1978, 1988, 1998, and 2007 next to one another for comparative visual reference. What stands out is the lack of obvious change. While our econometric analysis showed an increase in flexibility over time, such flexibility does not appear to be embedded in word choice in any obvious way. “Shall” still dominates, and “environmental” continues to be hard to find.[23]

[23] Similar word clouds are generated for the other years in our database as well.

Figure 5: Word Cloud for 1978, 1988, 1998, 2007



1978 1988 1998 2007

Table 3 lists the top ten most frequent words in the entire corpus of text, as well as for the particular licenses issued per year. Again, the most frequent words are remarkably similar, reinforcing the notion that general word choice over time has not obviously changed in hydroelectric license contracts issued by FERC. After reading many hundreds of licenses, it does become apparent that many license Articles from the early years serve as templates for similar Articles in later years. This tendency to use previous Articles as templates for future Articles is a possible reason for the minor changes in word choice over time.

At the same time, a close perusal of Table 3 does identify a few key changes. First is the addition of the word “plan” in the top ten lists for the years from the mid-1980s on. After passage of the Electric Consumers Protection Act (ECPA) in 1986, a key piece of legislation that mandated FERC begin including more environmental “balance” in the licenses that they issued, the “plan” became a common method for implementing new environmental requirements. For example, FERC would add an Article requiring the licensee to implement a “plan” to study water quality over time, or to implement a “plan” to increase fishery resources. The addition of the word "plan" is in fact an indication that environmental considerations increased their presence in licenses after 1986.

Another interesting trend in Table 3 is the increase in the use of the word "not" from the 1990s on. Many license Articles appeared to transition around that time from being positive mandates to do something, for example build new recreational facilities, or implement nature-blending aesthetic improvements, to being negative mandates "not" to do something, such as not to increase ramping rates in the winter months, or not to construct new facilities without first consulting a historian or an archaeologist. What this change from positive to negative tone implies for flexibility is unclear, but it is an intriguing trend nonetheless.

Finally, while the top seven or eight words appear remarkably consistent over the entire time period under study, there is a clear drift at the bottom of the table showing that, however slowly, word choice frequency does appear to be changing for words at the end of the top ten frequency distribution.[24] This may say more about linguistic trends than anything, but it would be interesting in the future to extend this table out for many more decades to see if the slow change in some of the top words utilized continues.

Ultimately, what Table 3 indicates is that if flexibility is increasing over time in these license contracts, as the regressions imply, then it is primarily through the one sentence reopener clauses attached to particular Articles, and not through subtle word choices in the body of the text of the Articles themselves. Legally, this may indeed be the most straightforward approach for inserting flexibility into a contract anyhow.

Table 4 and Figure 6 together give an indication of the relative frequency of particular environmental words in the licenses over time. Our hypothesis is

[24] This is related to the concept of "entropy" in the textual analysis literature.

that any increase in flexibility of hydroelectric power licenses over time is due to an increase in uncertainty, most likely as related to environmental concerns. We would expect, therefore, to see an increasing focus on environmental issues in the licenses over time as society's preferences for environmental protection increased. However, most of the words in Table 4 and Figure 6 show little to no frequency changes over time. The only words to show a significant non-constant trend are "recreational," which decreases over time, and "fish," which increases. The ECPA, mentioned previously, did signal out fishery resources as an environmental concern that FERC should pay greater attention to, so perhaps the increase in the prevalence of the word "fish" results from a greater effort by FERC to comply with the ECPA legislative mandate. The decrease in "recreational" is surprising, though it may indicate that FERC began to focus less on recreational mandates in hydropower licenses from the late 1980s on, as a sort of trade-off to including more fishery resource requirements. At this point, however, this is pure speculation.

Finally, it is worth mentioning that an analysis of the text based on the metadata in the licenses (essentially, all of the independent variables listed in Table 1, sans the decade dummies) shows very little change in top word choice or environmental word use. With respect to top word choice, nearly half of the variables have the exact same top ten word list as the corpus as a whole (i.e. the "All" column in Table 3), and from all the variables put together, only 14 distinct words exist in the top ten lists (as opposed to 24 in Table 3). With respect to the environmental words, the only distinct outlier is the word "fish," which in licenses that have *Wild Scenic* rivers nearly doubles in frequency usage.[25] Otherwise, the prevalence of environmental words in licenses constrained by metadata variables remains remarkably consistent.

[25] Tables and data available from the author upon request.

Finally, we ran a number of regressions with particular word counts as dependent variables,[26] in order to test whether asset specific or environmental uncertainty variables predicted the prevalence of certain word choices in the hydroelectric license Articles. The main result was that, similar to the regressions in Table 2, the asset specific variables were for the most part insignificant, while the environmental uncertainty variables were significant with greater consistency across the regressions. All of the regressions had similar levels of significance and adjusted R^2s so they were uninformative as to the relative choice of particular words (i.e. may versus must or should versus shall) across licenses. These additional regressions are also available from the author upon request.

Conclusions

In conclusion, there are three main results from this research. The first is that, by providing a truly direct test of the flexibility/rigidity tradeoff embodied in long term contracts, it offers a robust confirmation of the main predictions of transaction cost theory. The previous literature testing the operationalized factors underlying transaction cost theory (i.e. asset specificity, uncertainty, duration, and probity) relied on indirect or incomplete dependent variables. This paper utilizes a large data set and a set of dependent variables that embody comprehensive measures of contract completeness.

Second, this paper offers specific insights on the evolution of U.S. hydroelectric license contracts from 1977-2007. In the tradeoff between flexibility and rigidity in these specific long-term contracts, flexibility has won out as the environmental concerns embodying hydropower production have

[26] The words were: must, may, shall, should, recreational,fish, wildlife, environmental, pollution, not.

dominated any potential hold-up fears due to asset specificity and the large capital costs of hydroelectric dams. Mirroring broader social preferences from this time period where environmental concerns have grown, the flexibility associated with dealing with environmental concerns have come to dominate hydropower license contract provisions.

Finally, the textual analysis utilized in the empirical section of this paper allows us to make some conclusions regarding the form this increase in contractual flexibility has taken. Word choice across the licenses has been relatively consistent over time, implying that the increase in flexibility is not embodied in subtle wording choices, but much more straighforwardly in the number of well-defined, legally unambiguous reopener provisions. This is an interesting result. It begs the question of what *is* the efficient language choice when writing long-term contracts? Note that optimality is not being examined in this paper, only process. Future research investigating optimality, both in language word choice and in levels of contract flexibility or rigidity, would be both an interesting and informative direction to pursue.

Figure 2: The Five FERC Regional Offices and their Coverage

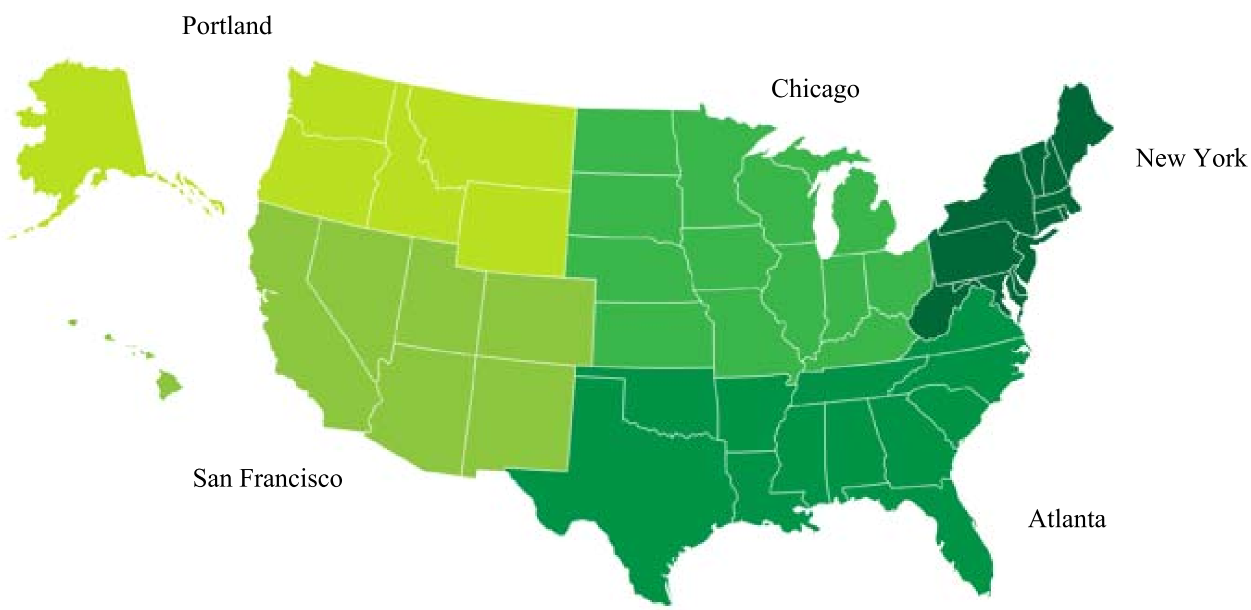


Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Dependent Variables:				
Reopener Clauses	4.47	4.60	0	28
Average Length	179.35	77.54	40	1,088
Independent Variables:				
$kW/10^3$	20.45	108.20	.07	2,755.50
Multiple Dams	.17	.38	0	1
Max Dam Height	52.87	77.94	1	773
Plant Type	.70	.46	0	1
Chicago	.18	.38	0	1
Portland	.17	.38	0	1
San Francisco	.17	.37	0	1
Atlanta	.13	.33	0	1
New York	.36	.48	0	1
Wild Scenic	.03	.17	0	1
Endangered Species-All	1.92	2.73	0	31
Endangered Species-Mammals	.58	.96	0	8
Government Land	.39	.49	0	1
New License	.13	.34	0	1
Relicense	.87	.34	0	1
Multiple Owner	.56	.50	0	1
Industrial	.10	.30	0	1
Utility	.25	.43	0	1
Nonutility	.38	.48	0	1
Municipal	.24	.43	0	1
Cooperative	.03	.18	0	1
ALP-ILP	.04	.18	0	1
TLP	.96	.18	0	1
OI	2.66	5.23	0	50
Duration	41.01	7.83	18	50
1970s	.03	.18	0	1
1980s	.57	.50	0	1
1990s	.24	.43	0	1
2000s	.155	.36	0	1

Table 2: OLS Regression Results

	Variable	Reopener Clauses Coefficient	Reopener Clauses Std. Error	Average Length Coefficient	Average Length Std. Error
	Constant	-1.80**	0.85	94.39***	13.63
Asset Specificity Variables (ρ):	kW/10^3	-0.00	0.00	-0.02	0.01
	Multiple Dams	0.17	0.26	1.90	4.09
	Max Dam Height	0.00	0.00	0.11***	0.02
	Plant Type	0.16	0.23	-4.32	3.75
Environmental Uncertainty Variables (μ):	Wild Scenic	3.75***	0.57	51.72***	9.01
	Endangered Species	0.10***	0.04	2.51***	0.58
	Government Land	0.45*	0.23	6.20*	3.71
	New License	0.12	0.34	-0.98	5.35
Reputation Variables:	Multiple Owner	-0.03	0.21	2.52	3.38
	Industrial	-1.16***	0.33	1.13	5.24
	Utility	-0.26	0.26	-2.12	4.15
	ALP-ILP	-0.18	0.55	-17.06*	8.81
	OI	0.25***	0.02	2.73***	0.34
Duration Variable:	Duration	0.07***	0.02	0.73***	0.25
Decade Variables:	1980s	1.54***	0.53	21.82**	8.69
	1990s	6.80***	0.56	97.32***	9.05
	2000s	5.43***	0.60	112.23***	9.73
Geographic Variables:	Portland	-0.55	0.35	-32.73***	5.67
	San Francisco	-2.04***	0.36	-31.90***	5.75
	Atlanta	-1.44***	0.35	-15.39***	5.56
	New York	-0.94***	0.28	-10.05**	4.45
		Adj. R^2=0.53		Adj. R^2=0.54	

Table 3: Top Ten Most Frequent Words

Top 10 Words	All	'78	'79	'80	'81	'82	'83	'84	'85	'86	'87	'88	'89	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05	'06	'07
shall	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
licensee	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
project	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
commission	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
plan	x										x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
article	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
any	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
lands	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			x		x	x	x			x	x
use	x		x	x	x	x	x	x	x	x				x	x	x	x						x		x		x			x	
not	x														x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
such		x																													
facilities		x																													
may		x		x																											
from		x	x	x			x	x	x	x	x	x	x																		
construction			x			x	x	x	x	x	x	x	x	x																	
conveyed					x	x																									
other					x																										
service																		x													
if																			x	x		x				x					
agencies																					x										
include																					x	x		x							
approval																								x					x		
file																												x	x		
license																												x			x

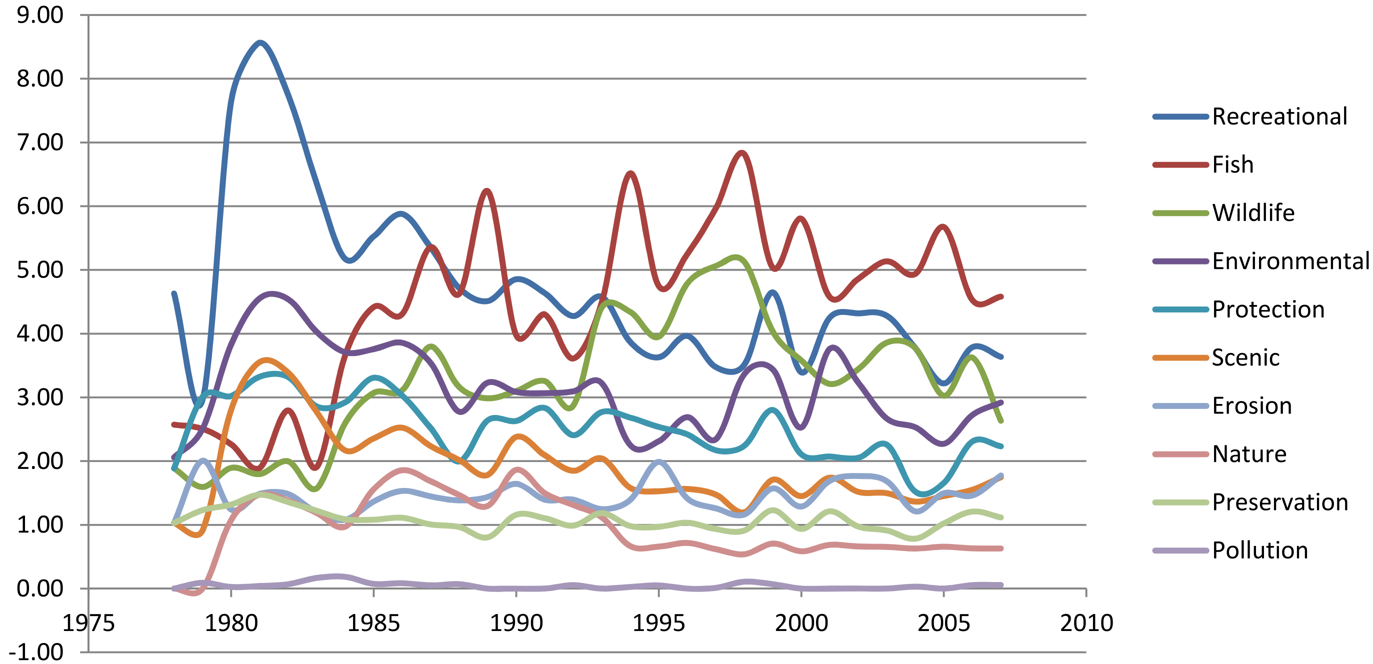
Table 4: Frequency of Relevant Environmental Words*

Words	All	'78	'79	'80	'81	'82	'83	'84	'85	'86	'87	'88	'89	'90	'91	'92
recreational	4.76	4.63	2.96	7.64	8.57	7.75	6.36	5.17	5.53	5.88	5.35	4.72	4.51	4.85	4.64	4.28
fish	4.66	2.57	2.51	2.27	1.89	2.80	1.91	3.66	4.42	4.31	5.36	4.62	6.24	3.98	4.30	3.61
wildlife	3.37	1.89	1.60	1.89	1.80	2.00	1.57	2.60	3.07	3.11	3.80	3.16	2.99	3.10	3.26	2.87
environmental	3.16	2.06	2.51	3.83	4.55	4.54	4.03	3.71	3.76	3.86	3.54	2.78	3.23	3.08	3.07	3.09
protection	2.53	1.89	3.01	3.02	3.32	3.32	2.86	2.92	3.31	3.03	2.52	1.99	2.64	2.63	2.83	2.41
scenic	1.97	1.03	0.91	2.80	3.55	3.39	2.78	2.17	2.35	2.52	2.24	2.02	1.78	2.38	2.09	1.85
erosion	1.42	1.03	2.01	1.24	1.49	1.48	1.20	1.08	1.37	1.53	1.45	1.38	1.44	1.64	1.40	1.40
nature	1.08	0.00	0.00	1.08	1.47	1.38	1.19	0.97	1.56	1.86	1.68	1.47	1.30	1.86	1.50	1.31
preservation	1.05	1.03	1.23	1.32	1.47	1.36	1.22	1.09	1.08	1.11	1.01	0.97	0.80	1.16	1.10	0.99
pollution	0.05	0.00	0.09	0.03	0.04	0.07	0.17	0.18	0.07	0.08	0.05	0.07	0.00	0.00	0.00	0.05

		'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05	'06	'07
recreational		4.58	3.87	3.63	3.97	3.47	3.51	4.65	3.39	4.25	4.32	4.28	3.77	3.22	3.79	3.64
fish		4.49	6.52	4.75	5.25	5.97	6.82	5.03	5.80	4.57	4.86	5.14	4.94	5.67	4.53	4.58
wildlife		4.41	4.34	3.95	4.78	5.06	5.13	4.03	3.58	3.21	3.44	3.86	3.77	3.03	3.62	2.63
environmental		3.23	2.25	2.31	2.69	2.34	3.37	3.44	2.53	3.76	3.23	2.67	2.53	2.27	2.72	2.92
protection		2.77	2.68	2.53	2.42	2.17	2.24	2.80	2.11	2.07	2.05	2.26	1.52	1.66	2.31	2.23
scenic		2.04	1.58	1.53	1.56	1.47	1.20	1.71	1.45	1.74	1.52	1.50	1.36	1.45	1.55	1.75
erosion		1.25	1.39	1.99	1.41	1.26	1.16	1.57	1.29	1.69	1.77	1.68	1.21	1.50	1.46	1.77
nature		1.12	0.67	0.66	0.72	0.62	0.54	0.71	0.59	0.69	0.66	0.66	0.63	0.66	0.63	0.63
preservation		1.18	0.98	0.97	1.03	0.93	0.91	1.23	0.94	1.21	0.97	0.91	0.78	1.02	1.21	1.12
pollution		0.00	0.03	0.05	0.00	0.01	0.11	0.07	0.00	0.00	0.00	0.00	0.03	0.00	0.05	0.06

***** Numbers are percent frequencies, raised to the 10^3.

Figure 6: Frequency of Relevant Environmental Words



References

1. Antweiler, Werner and Murray Z. Frank. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *The Journal of Finance* 59(3):1259-1294.

2. Banerjee, Abhijit V. and Esther Duflo. 2000. "Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry." *The Quarterly Journal of Economics* 115(3):989-1017.

3. Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science* 2:1-8.

4. Choi, Hyunyoung and Hal Varian. 2009. "Predicting the Present with Google Trends." *Working Paper.*

5. Coase, Ronald. 1937. "The Nature of the Firm." *Economica* 4(16):386-405.

6. Crocker, Keith J. and Scott E. Masten. 1988. "Mitigating Contractual Hazards: Unilateral Options and Contract Length." *RAND Journal of Economics* 19(3):327-343.

7. Crocker, Keith J. and Scott E. Masten. 1991. "Pretia Ex Machina? Prices and Process in Long-Term Contracts." *Journal of Law and Economics* 34(1):69-99.

8. Crocker, K., and K.J. Reynolds. 1993. "The Efficiency of Incomplete Contracts: An Empirical Analysis of Air Force Engine Procurement." *The RAND Journal of Economics* 24 (1):126-146.

9. DeWall, C. Nathan, Richard S. Pond, Jr., W. Keith Campbell, and Jean M. Twenge. 2011. "Tuning in to Psychological Change: Linguistic Markers of Psychological Traits and Emotions Over Time in Popular U.S. Song Lyrics." *Psychology of Aesthetics, Creativity, and the Arts* 5(3):200-207.

10. Evans, James A. and Jacob G. Foster. 2011. "Metaknowledge." *Science* 331:721-725.

11. Federal Energy Regulatory Commission, Office of Hydropower Licensing. 2004. *Handbook for Hydroelectric Project Licensing and 5MW Exemptions from Licensing.* Washington, D.C.

12. Gentzkow, Matthew and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica* 78(1):35-71.

13. Gerow, Aaron and Mark T. Keane. 2011. "Mining the Web for the 'Voice of the Herd' to Track Stock Market Bubbles." *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence* 2244-2249.

14. Goldberg, Victor P. 1976. "Regulation and Administered Contracts." *The Bell Journal of Economics* 7(2):426-448.

15. Grimmer, Justin and Gary King. 2010. "General Purpose Computer-Assisted Clustering and Conceptualization." *Proceedings of the National Academy of Sciences* 1:1-8.

16. Joskow, Paul L. 1987. "Contract Duration and Relationship-Specific Investments: Empirical Evidence from Coal Markets." *American Economic Review* 77(1):168-185.

17. Joskow, Paul. 1990. "The Performance of Long-Term Contracts: Further Evidence from Coal Markets." *RAND Journal of Economics* 21(2):251-274.

18. Kerkvilet, Joe and Jason Shogren. 2001. "The Determinants of Coal Contract Duration for the Powder River Basin." *Journal of Institutional and Theoretical Economics* 157:608-622.

19. Klein, Benjamin, Robert Crawford & Armen Alchian. 1978. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process." *Journal of Law and Economics*, 21(2): 297-326.

20. Koppel, Moshe, Shlomo Argamon, Anat Rachel Shimoni. 2011. "Automatically Categorizing Written Texts by Author Gender." *Working Paper.*

21. Kozhevnikova, Maria and Ian Lange. 2009. "Determinants of Contract Duration: Further Evidence from Coal-Fired Power Plants." *Review of Industrial Organization* 34:217-229.

22. Krugman, Paul. April 28, 2011. "The Intimidated Fed." *The New York Times*.

23. Michel, Jean-Baptisti, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331:176-182.

24. Neumann, Anne and Christan von Hirschhausen. 2008. "Long Term Contracts and Asset Specificity Revisited - An Empirical Analysis of Producer-Importer Relations in the Natural Gas Industry." *Review of Industrial Organization* 32(2):131-143.

25. Patel, Aniruddh D., John R. Iversen, Micah R. Bregman, and Irena Schulz. 2009. "Experimental Evidence for Synchronization to a Musical Beat in a Nonhuman Animal." *Current Biology* 19(10):827-830.

26. Pollak, Daniel. 2007. "*S.D. Warren* and the Erosion of Federal Preeminence in Hydropower Regulation." *Ecology Law Quarterly* 34:763-800.

27. Porter, M.F. 1980. "An Algorithm for Suffix Stripping." *Program* 14(3):130-137.

28. Saussier, Stephane. 2000. Transaction Costs and Contractual Incompleteness: the Case of Electricite de France. *Journal of Economic Behavior & Organization*. 42(2):189-206.

29. Schachner, Adena, Timothy F. Brady, Irene M. Pepperberg, and Marc D. Hauser. 2009. "Spontaneous Motor Entrainment to Music in Multiple Vocal Mimicking Species." *Current Biology* 19(10):831-836.

30. Spirling, Arthur. 2010. "Bargaining Power in Practice: U.S. Treaty-Making with American Indians 1784-1911." *Working Paper*.

31. Tetlock, Paul C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance* 62(3):1139-1168.

32. Williamson, Oliver E. 1979. "Transaction-Cost Economics: The Governance of Contractual Relations." *Journal of Law and Economics* 22(2):233-261.

33. Williamson, Oliver E. 1999. "Public and Private Bureaucracies: A Transaction Cost Economics Perspective." *Journal of Law, Economics, and Organization* 15(1):306-342.