The Language of Contract: Promises and Power in Union Collective Bargaining Agreements

Elliott Ash, W. Bentley MacLeod, Suresh Naidu*

March 30, 2019

Preliminary version: Comments welcome.

Abstract

We document determinants of control rights in union contracts using a new corpus of 30,000 collective bargaining agreements from Canada from 1986 through 2015. Using ideas and methods from computational linguistics, we extract measures of rigidity and worker control from the text of the contract clauses. Motivated by a model of efficient contract design, we analyze how rigidity and authority in contracts varies according to firm-level factors and external factors. We document that contracts impose obligations equally on firms and workers but give entitlements mostly to workers. An increase in personal income tax rates is associated with an increase in worker entitlements, consistent with substitution away

*Elliott Ash: ashe@ethz.ch, Department of Social Sciences, ETH Zurich; W. Bentley MacLeod: bentley.macleod@columbia.edu, Department of Economics, Columbia University; Suresh Naidu, sn2430@columbia.edu, Columbia University, Economics Department/SIPA. We thank Yisehak Abraham, Alex Ahmadi, Sungmin An, Eli Ben-Michael, Jakob Brounstein, Zoey Chopra, Leslie Cordero, Sydney Egnasko, Jeff Jacobs, Vidushi Jayathilak, Lorenzo Lagos, Ji-Yoon Lee, Allison Peck, Jyotirmayee Sabat, Valentin Saitarli, Divya Singh, Eileen Suh, Anna Vladymyrska, Erin Walsh, Yong Wang, Grace Zheng, and Jon Zytnick for helpful research assistance. We thank Yuki Shiraito, Rosa Ferrer, Eric Helland, Scott Masten, the participants at the Princeton University Conference on Real-World Impacts of Legal Text, Bonn Economics Seminar, Conference on Empirical Legal Studies in Europe, Nanterre Law and Economics Workshop, Barcelona Graduate School of Economics Summer Forum, Conference on Empirical Legal Studies, and Columbia Data Science Institute for helpful comments. Columbia University's Program for Economic Research, the National Science Foundation, and Washington Center for Equitable Growth provided financial support for this research.

from taxed compensation (income) and toward untaxed compensation (control rights as amenities). Lower local sectoral unemployment rates and provincial governance by pro-labor parties is associated with increased worker authority, consistent with effects of changed bargaining power for workers. Worker control also reduces labor conflict: while unanticipated real wage shocks from COLA clauses increase the probability of a strike, this effect is attenuated in contracts with a high degree of worker control.

1 Introduction

Economics has produced an extensive body of theory of optimal contract design, where the allocation of decision rights depends on economic parameters such as asymmetric information and investment specificity. These theories generate testable predictions about quantitative features of contracts. Lawyers, however, traffic in real-world contracts, which are extensive bodies of text, themselves governed by statute (also text) and judicial precedent (again, text). In this paper, we take contract theory to data using a large corpus of collective bargaining agreements produced in the relatively homogeneous environment of Canadian collective bargaining in the last three decades. We use tools from computational linguistics to operationalize ideas from the economic theory of contracts – specifically, the degree of specified contingency, and the extent of agent delegation – and examine economic and political determinants of these contractual features.

Our innovation on this literature is a data-driven analysis of contract language, with the goal of empirically operationalizing long-standing theoretical notions of control rights. We extend natural language processing techniques to extract the contractual obligations to workers and restrictions on managerial prerogatives embedded in union contracts. We unpack the details of "what unions do" by analyzing the text of collective bargaining agreements. The goal is to provide field evidence of some of the predictions of recent models of optimal contract design (Battigalli and Maggi, 2002a; Hart and Moore, 2007).

We begin with a model of labor contracting to motivate the approach to the data. We show that when efficient assignment of workers to tasks depends upon unobserved worker characteristics, it may be efficient to add contract terms that constrain firm choice and provide on-the-job amenities for workers. We then embed this model into a standard union-firm bargaining model to derive the implications of variation of the

exogenous parameters upon employment, number of clauses, and wages.

The empirical setting is nearly all collective bargaining agreements for unionized firms in Canada for the years 1986 through 2015. We use natural language processing tools to extract norms, commitments, and entitlements from the text of contracts, where we document a number of stylized facts. Overall, contracts can be understood as bundles of obligations and entitlements. We see about the same number of obligations on employers and employees. For entitlements, almost all of them are granted to workers.

To analyze the determinants of worker authority, we link the contracts data to economic and institutional variables such as employment, strikes, political control, and provincial income tax changes. We have panel data on firm-union bargaining pairs, which means we can explore the relationship between changes in economic and political conditions and changes in the contract terms. First, we look at an an increase in labor income tax rates, which can be seen as a change in the relative prices between wage compensation and non-wage compensation. These tax increases are associated with an increase in worker entitlements, consistent with a substitution effect away from taxed compensation (income) and toward untaxed compensation (amenities).

A second interesting factor determining contract terms is the relative bargaining power of employers and employees. Control of province government by the labor-supporting New Democratic Party is associated with higher worker authority, consistent with an effect of higher bargaining power for workers due to political support. Relatedly, increases in local sectoral unemployment rate are associated with decreases in worker entitlements.

We are also interested in the effects of worker authority on firm outcomes. In particular, we investigate the role of the contract in mitigating conflict, which can be measured in the union contex by the frequency and intensity of strikes. First, we find that negative wage shocks (due to cost-of-living clauses not keeping up with inflation) are associated with higher strike rates and intensity. However, this effect is reduced when the contract gives relatively high authority to workers through entitlements. This differential effect is consistent with worker rents from control rights reducing strike incentives.

The findings attest to the important role of collective bargaining in the labor market. Economic and political conditions have an impact on the text-based measures of worker authority. That authority matters for firm outcomes. Future work in labor economics

would benefit from integrating these text-based measures of the granular details of union contracts.

This research adds to a large literature arguing that unions were an important force in compressing the wage (and income) distribution in the twentieth century (DiNardo et al., 1996; Card, 2001). For example, Jaumotte and Osorio (2015) show that union density reduces top income shares in a panel of OECD country-years (building on Western, 1999). In parallel work Farber, Herbst, Kuziemko, and Naidu (2017) find a similar effect of union density on top income shares across U.S. states, using new micro-data on unions extracted from historical polling data. More generally, a large amount of recent research has suggested that rent-sharing within the firm is an important component of the wage distribution (Card et al., 2016). Unionization may also be related to historical socioeconomic disparities related to race and gender (Blau and Beller, 1988); indeed, Farber et al. (2017) find that nonwhite workers are more likely to be union members and enjoy larger union premia than comparable white workers.

This previous work invites investigation into how unions managed to capture those rents. What trade-offs were made in terms of workplace amenities, firm-specific human capital, and worker autonomy? Even though each day, parties enter into countless contracts, very little is known regarding the economic effects of contract design. Workplace authority and amenities, such as scheduling, job security, training, and seniority, are embedded in contractual language. By treating each firm as a jurisdiction, and each contract as a workplace constitution, we obtain fine, granular evidence that the "rules of the game" matter for economic outcomes. Even in an era of weakening unions, the lessons from these contracts will help policymakers design labor-market rules to govern workplace amenities, rent-sharing, and control rights within the firm.

The rest of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the metadata, while Section 4 describes the text data and methods. Section 5 articulates the econometric approach. Section 6 reports the results, while Section 7 concludes.

2 Literature Background

The modern theory of union contracts, building on the work of Grossman and Hart (1986a), recognizes that an important function of labor institutions and contracts is the efficient allocation of authority and decision rights within a relationship. Most

economists agree that labor law and labor unions affect the relative bargaining power of workers (Svejnar, 1986; Abowd and Lemieux, 1993a), but the standard model sees them as merely redistributing rents. On this view, any allocation of bargaining power that results in prices diverging from competitive levels is inherently inefficient. In contrast, the modern contract literature views authority as an instrument for mitigating transaction costs due to asymmetric information and holdup. Because transaction costs are significant, labor protections and labor unions may enhance productive efficiency (Freeman and Medoff, 1984; Svejnar, 1986). In a similar vein, collective bargaining by a union can mitigate problems related to employer monopsony power (Manning, 2010). The hysteresis in contract terms documented in Card (1986a) suggests that observed union contracts cannot be viewed as achieving the first best, and hence transaction costs are a necessary ingredient for understanding the observed structure of negotiated employment contracts.

These papers on labor contracts are a part of the broader theory literature on of contract design and contract writing, reviewed in detail by Kornhauser and MacLeod (2012). An important entry in this literature is Battigalli and Maggi (2002a), where contracts are modeled "from the ground up" as a set of statements mapping events to actions. Writing contract statements is costly, so incompleteness arises endogenously. Discretion is the case where an event does not have an associated action, so the agent chooses what to do. Rigidity is the case where an action is always performed, regardless of events. Battigalli and Maggi (2008) extend this model to multiple periods. Other theory papers on costly contract writing include Schwartz and Watson (2004), Posner (2005), and Shavell (2006).

The empirical literature on unions asks whether unionization of a workforce affects productive efficiency or firm profits. Two recent meta-analyses using a large number of union studies conclude that while unions enhance firm productivity (Doucouliagos and Laroche, 2003), they also reduce firm profits (Doucouliagos and Laroche, 2009). But the evidence is mixed and the effects of unions seem to be sensitive to context. Abowd (1989), among others, shows that an unexpected increase in union wages results in a dollar-for-dollar transfer from shareholders to workers, with little evidence of net loss or net gain. DiNardo and Lee (2004) use union certification elections as an instrument for unionization in a regression discontinuity design (RDD), finding no local average treatment effects on firm stock price of being unionized. Lee and Mas (2012) replicate the zero LATE of unionization at the RD cutoff, but also find a large

negative average treatment effect on firm equity value using diffs-in-diffs. The decrease in value is delayed, however, occurring over the 18 months following the unionization vote. A relevant detail from Jordan and Bruno (2005) is that only 57 percent of new bargaining units achieve a first contract within a year, and only 70 percent achieve a first contract at all. Unions with weak support – that is, near the RD cutoff – may be less likely to achieve a contract. Using Spanish data, Card et al. (2014) provide evidence that increasing profits are shared with union workers, although the sharing does not significantly reduce the return on capital.

There is evidence that these transfers of rents from shareholders to unions have induced firm responses. Holmes (1998) finds that manufacturing companies located near U.S. state borders will locate factories in the state with Right-to-Work laws, consistent with firms responding to the negative effect of unions on profits. Similarly, the large decline in unionization since 1970 is consistent with the hypothesis of excessive rent extraction by unions, with firms subsequently turning to nonunion alternative investments (Farber and Western, 2001). Machin (2000) documents a similar trend in the United Kingdom.

In Canada, unions have declined more slowly than in the United States (Kuhn, 1998). This again suggests that the fitness of unionization is sensitive to institutional and economic context. Abowd and Lemieux (1993a) analyze the impacts of trade shocks on union wage provisions in Canada. Budd and Wang (2004) analyze the effects of strikes on investment.

MacLeod (2011) discusses the literature on labor unions as part of a broader review of employment contracts, noting that there is little systematic evidence that unions reduce the efficiency of the employment relationship. There are a few papers showing that labor/employment contracts matter for firm and work outcomes (e.g., Card and De La Rica, 2006; Garloff and Guertzgen, 2012); these papers focus on wage provisions, although Freeman and Kleiner (1990) argue that unionization affects non-wage employment conditions more than it affects wages. Research that focuses on hand-coded features of union contracts includes Juravich et al. (2006) and Strunk and Grissom (2010). Empirical work on contract terms in other contexts include Masten and Crocker (1985), Joskow (1987), Leffler and Rucker (1991), Allen and Lueck (1992), Ackerberg and Botticini (2002), Gulati and Scott (2012), and Matvos (2013).

Most recently, empirical researchers have begun to apply computational techniques from natural language processing to the text of written contracts (Talley and O'Kane,

Table 1: Summary Tabulations for Contracts Metadata

Province	Freq.	Percent
Alberta	3,541	11.87
British Columbia	3,693	12.38
Manitoba	1,658	5.56
Multiprovince	1,461	4.90
New Brunswick	789	2.64
Newfoundland / Labrador	552	1.85
Northwest Territories	424	1.42
Nova Scotia	1,140	3.82
Nunavut	171	0.57
Ontario	14,414	48.30
Prince Edward Island	147	0.49
Quebec	490	1.64
Saskatchewan	1,165	3.90
Yukon Territory	196	0.66
Total	29,841	100.00

Industry Group	Freq.	Percent
Construction	1,645	5.51
Educational/health	10,148	34.01
Entertainment	782	2.62
Finance / Real estate	829	2.78
Information / culture	1,216	4.07
Manufacturing	4,979	16.69
Primary industries	459	1.54
Public admin	3,731	12.50
Transportation	4,696	15.74
Utilities	533	1.79
Wholesale / retail	823	2.76
Total	29,841	100.00

2012; Sanga, 2014; Moszoro et al., 2016; Ganglmair and Wardlaw, 2017). This is part of a growing area of empirical research using text data by economists and other social scientists (Gentzkow and Shapiro, 2010; Roberts et al., 2013; Taddy, 2013; Ash, 2016; Ash et al., 2017b,a). Gentzkow et al. (2017) provide a recent survey of this literature.

Besides these papers in economics, there is a large literature in labor and legal history, sociology, and political science concerned with the interaction of labor laws and labor contracts. Stepan-Norris and Zeitlin (2003) argue that certain contract provisions, such as strong stewards, were a key distinctive demand of particularly politicized unions in the post-war period. McCammon (1990) and Pope (2004), among others, have argued that court interpretations of the Wagner Act have been decisive in weakening the strike provisions in contracts.

3 Data on Contracts

Our data source for Canadian contracts is Employment and Social Development Canada, from which we obtained 28,848 contracts in the English language for the years 1986 through 2015. This section describes the metadata (that is, non-text numerical data) used in the analysis.

Our data on union contracts is more comprehensive than that used in the previous literature. There are 29,848 contracts, 6,004 companies (~5 contracts per company),

Table 2: Summary Statistics for Contracts Metadata

Variable	Mean	Std. Dev.	Min	Max
Private-Sector	.4860	.499	0	1
Number of Employees	655.87	2721.	0	170000
Effective Year	1999.79	7.89	1986	2015
Duration (Years)	2.584	1.1	0	20
Has COLA	.2731	.445	0	1
Annual Wage Increase (%)	2.61	1.843	-7.560	19.836
Inflation (%)	5.77	3.347	-.8643	31.62
Negative Shock	.362	.4807	0	1

14 provinces, 794 cities, 11 industry groupings, and 551 industry codes. Table 1 reports summary tabulations for the provinces and industry groups. Quebec has relatively few contracts in our sample because most of them are in French.

For each contract, we have the company, union, location, industry, public/private status, and number of employees. We have a set of related dates (signing, effective, and expiry), which allows us to compute contract duration, and match up to short-term changes in economic and political conditions.

For some contracts, we have a wage-adjustment schedule, which gives the planned wage increases over the course of the contract. We matched this data with realized inflation over the course of the contract, with the idea that a COLA clause wage increase, minus the realized inflation, results in an unanticipated real wage shock at the start of bargaining over the subsequent contract (Card, 1986b). In our results we focus on "Negative Shock," an indicator equaling one when inflation beats the COLA.

Table 2 provides summary statistics on these variables. Figure 1 provides histograms for a selection of the real-valued variables. About half of the firms are in the private sector. The number of employees is widely dispersed. There is significant variation in contract duration, with bunching at 12-month increments. A good number of firms have COLA clauses, with meaningful variation in the gap between COLA and inflation.

Our second data set from Employment and Social Development Canada is the history of strikes among Canadian unions since 1945. We have data on 28,471 strike events, for 22,163 companies. Summary statistics are reported in Table 3. These are

Figure 1: Summary Figures for Contracts Metadata

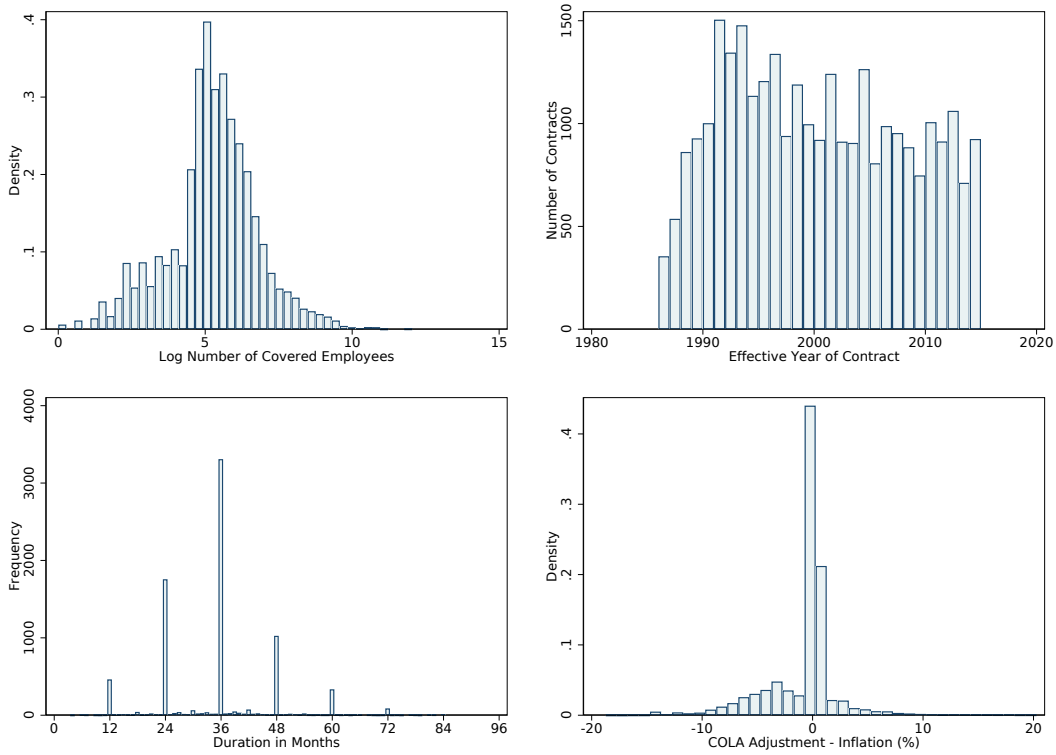


Table 3: Summary Statistics on Strikes

Variable	Mean	Std. Dev.	Min	Max
Workers (Max)	553.95	6452.4	1	830000
Work Days	33.210	75.57	0	2144
Person Days	8445.50	51786.4	0	2156980
Private Sector	.8126	.3901	0	1
Lockout	.0754	.264	0	1
Rotating Strike	.0220	.1469	0	1
Issue: Wages	.595	.49	0	1

Table 4: Summary Statistics on Population and Employment

Variable	Mean	Std. Dev.	Min	Max
Population (1000s)	5795.22	3746.8	96.9	11684
Labor Force (1000s)	266.08	210.57	.2	855.3
Employment (1000s)	254.69	202.75	.2	838.40
Unemployment Rate (%)	5.209	3.729	1.08	49.9
Personal Income Tax Rate (%)	22.3	1.458	16.11	25.62

big events, with over 500 workers on average, and over 8,445 person-days lost. Over 80% are private-sector strikes, and in about 60% of cases, wages are a reported issue over which the workers are striking. Of these cases, the only reported issue in about half of the strikes. In about 70% of strikes, there are non-wage issues reported, including fringe benefits, working conditions, negotiation delays, hours, dismissals, suspensions, job security, and sympathy for other unions.

Next, we have local labor market data. We have numbers by province, sector, and year, for population, labor force, full-time employment, and part-time employment. We use these numbers to compute unemployment rates by province, sector, and year. In addition, we have data on the implicit personal income tax rate, by province and year, from the Center for the Study of Living Standards.[1] Summary statistics for these data are reported in Table 4.

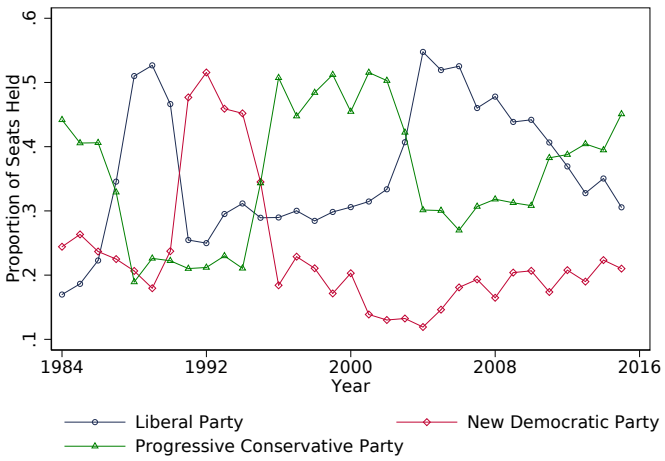
We are interested not just in economic conditions, but also the local political conditions in each province. We collected data on all of the provincial elections in Canada during our time period. This includes the date of these elections, the number of seats to be filled, and the allocation of seats to Liberal, New Democratic, and Progressive Conservative parties.[2] Figure 2 shows that the relative control of these political parties has changed substantially and repeatedly in our time period of interest. The New Democratic Party is known to favor labor unions politically (Jansen and Young, 2009).[3]

[1] Available at `http://www.csls.ca/reports/csls2011-17appendixtables.pdf`. This is the implicit personal income tax rate by province, computed from income tax revenues as a share of income. Union-worker earnings are 1.5 times that of non-union workers, and tend to locate in the middle income tax brackets.

[2] There are three smaller regional parties: Quebecois, Saskatchewan, and Social Credit.

[3] We also have data on trade deals that are relevant to the labor market in Canada. First, we have

Figure 2: Provincial Party Control Over Time



4 Measuring Power and Rigidity in Union Contract Text

This section describes our methods for measuring features of union contracts. In addition to the aforementioned metadata by contract, we have access to the full corpus of 29,841 contracts (this is after removing about 10,000 French-language contracts). We begin by describing our framework for estimating power and rigidity, and then describe the details of how we processed the contracts.The data pipeline is summarized in Figure 3.

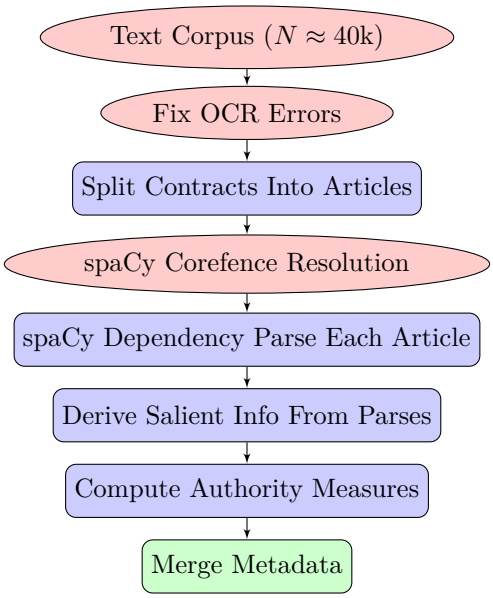
4.1 Pre-Processing

The contracts database arrived as scanned PDFs. The first step was to convert them to machine-readable text using OCR software. We excluded wage schedules, exhibits, appendices, and other miscellaneous materials.

We have introduced a number of tools for detecting and fixing OCR errors. First, we use three separate OCR engines: ABBBY FineReader, Adobe, and Tesseract. We

the Canada-United States Free Trade Agreement (CUSDTA), signed on January 2, 1988. Second, we have data the North American Free Trade Agreement (NAFTA), which entered into force on January 1, 1994. NAFTA included an ancillary agreement, the North American Agreement on Labor Cooperation (NAALC), which implemented uniform rules on labor contracting.

Figure 3: Data Pipeline



can show that agreement is high. When there was disagreement in OCR, we picked the word that was chosen by two of the three OCR engines. If all three chose differently, we went with ABBBY FineReader which has the lowest error rate.

Next, we split the contracts into sections. We tried a number of off-the-shelf legal section splitters. But we got better performance with a custom-built splitter, which used the relatively standard legal style to detect and segment section headers.

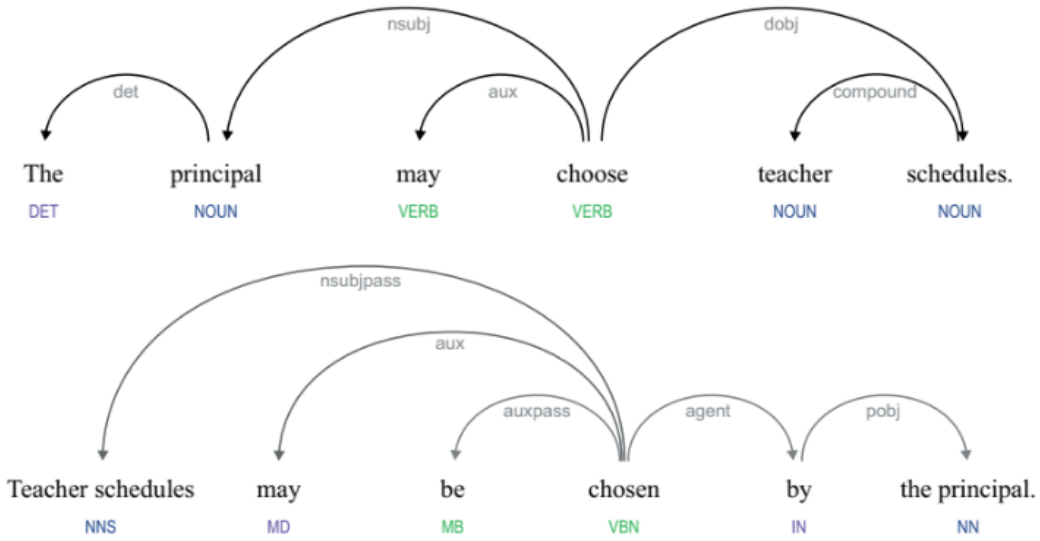
Within each section, we ran co-reference resolution using the spaCy plugin neuralcoref. This tool replaces pronouns with the associated entity.

We then used a sentence tokenizer to split each section into a list of sentences. The resulting corpus consists of 980,909 contract sections (32.9 per contract) and 10.8 million sentences (11.06 per section).

4.2 Measuring the Intent of Union Contract Terms

This section describes our approach for extracting legal provisions from union contracts. We build on recent methods using natural language processing to automate the interpretation of laws and contracts by extracting commitments, entitlements, and the like. This small literature includes Francesconi and Passerini (2007) and Ceci et al.

Figure 4: Syntactic Dependency Parse for Deontic Modal Verb Structures



(2011). These recent works draw on an older conceptial literature in law seeking to construct typologies of legal statements .

Each contract sentence is parsed using a syntactic dependency parser called spaCy (`spacy.io`). This package uses the ClearNLP dependency schema and has proven accuracy and efficiency relative to other parsers.[4] The parser transforms sentences into parse trees, which represent the relations between words in a recursive hierarchical structure. Figure 4 shows the dependency parse for two example sentences.

Besides spaCy, we ran two additional parsers on our data: the Stanford Parser and the Google Parser. The results were very similar across parsers. We have also checked that our results are robust to averaging the results of the parsers.

4.3 Conceptual Framework and Estimation

Like most natural language processing applications, our framework simplifies and abstracts from much of the syntactic and semantic richness of a collective bargaining agreement. We consider a collection of contracts I . Each contract, indexed by i , is a collection of N i independent clauses. Each clause j has 3 components: a subject, an

[4]See `http://www.mathcs.emory.edu/~choi/doc/clear-dependency-2012.pdf`.

Table 5: Summary Tabulations: Subjects, Modals, and Verbs

Subject	Freq.
employee	32465
who	12633
it	7198
employer	6431
company	5666
which	5404
he	5101
party	4044
they	3997
there	3081
union	2735
that	2649
teacher	2598
member	2501
leave	2303
board	2247
grievance	2092
dans	1960
nurse	1809
hour	1690
hospital	1626
rate	1612
time	1596
period	1572
he/she	1485
she	1460
committee	1350
day	1346
work	1301
agreement	1299
provision	1278
seniority	1267
notice	1233
position	1224

Modal	Freq.
shall	57263
will	30974
may	13491
must	3465
should	1954
would	1788
can	1702
could	206
might	130
ought	79
need	27

Verb	Freq.
be	35265
have	6212
agree	5900
be_pay	5400
receive	4236
work	4035
be_require	3656
apply	3468
provide	3045
be_make	2955
be_entitle	2694
be_grant	2663
continue	2355
be_give	2301
pay	2237
be_consider	1945
include	1639
make	1570
become	1553
mean	1518
be_provide	1495
occur	1486
complete	1420
be_understand	1402
leave	1301
require	1293
take	1224
be_agree	1212
recognize	1202
be_deem	1188
meet	1142
give	1102
notify	1092
commence	1063

Most Frequent Subject-Modal-Verb Tuples

Subject - Modal - Verb	Subject - Modal - Verb	Subject - Modal - Verb
agreement_shall_be	employee_shall_be	employee_shall_receive
arbitrator_shall_have	employee_shall_be_allow	employee_shall_retain
board_shall_have	employee_shall_be_consider	employee_will_be
case_may_be	employee_shall_be_entitle	employee_will_be_allow
committee_shall_meet	employee_shall_be_give	employee_will_be_entitle
company_shall_pay	employee_shall_be_grant	employee_will_be_give
company_shall_provide	employee_shall_be_lay_off	employee_will_be_grant
company_will_pay	employee_shall_be_pay	employee_will_be_pay
company_will_provide	employee_shall_be_require	employee_will_be_require
decision_shall_be	employee_shall_continue	employee_will_have
employee_may_request	employee_shall_lose	employer_shall_grant

object, and a verb.

- **Subject**: One of Company/Manager (the employer) or Employee/Union (the worker). The subject is assigned using a dictionary of synonyms to one of four agent categories: worker, union, owner, and manager (or other).

- **Verb**: We are most interested in deontic modal verb structures. The modal verb is distinguished as strict (*shall*, *will*, *must*) or permissive (*may*, *can*). Formally speaking, modality prescribes a favored action within a possible world (Kratzer, 1991). In contracts, these statements create legal obligations and entitlements, featuring the modal verbs shall, will, may, must, and can. Statements are tagged as negative ("shall not" rather than "shall"), and tagged as active ("shall hire") or passive ("shall be hired"). Table 5 provides tabulations for the most frequent subjects, modals, and verbs encountered in our data set. In the bottom panel, we have listed the most frequent subject-model-verb tuples (starting with msot frequent). We classify modal verbs as restrictive (shall) or permissive (may). We can see in the first column a focus on obligations for the company. In the second and third columns, we see a focus on entitlements for the worker.

- We identify a handful of special verbs that appear often in the contracts and delineate obligations and entitlements: Obligation Verbs (be required, be expected, be compelled, be obliged, be obligated, have to, ought to), Prohibition Verbs (be prohibited, be forbidden, be banned, be barred, be restricted, be proscribed), Permission Verbs (be allowed, be permitted, be authorized), and Entitlement Verbs (have, receive, retain). We define Action Verbs as all non-special active-tense words, including "be" by itself. Passive Verbs are all non-special passive-tense verbs. Now, we use these grammatical features to assign statements to one of four types of contract statements (or "other"). The formal requirements, plus some examples, are included in Table 6. An *Obligation* requires that the subject perform an action or set of actions. A *Prohibition* requires that the subject not perform an action or set of actions. A *Permission* gives the subject permission or authority over an action or set of actions. An *Entitlement* gives the subject an entitlement. We calculate frequency counts for each statement type and each agent. Table 7 reports summary statistics on these frequencies. As can be seen in the bottom set of variables, contracts in the main can be understood as a bundle of obligations and entitlements. We understand obligations and prohibitions as

reducing an agent's authority. We understand permissions and entitlements as expanding an agent's authority.

- **Object:** the remainder of the clause, specifying the action described by the verb, We suppose this is a noun phrase described by a "bag of words" drawn from vocabulary L and thus are represented by a vector of frequencies $\{f_{ij}\}_L$ associated with each clause.

We represent whether or not a clause is giving workers more power with a binary variable $X_{ij} = 1$ if combination is worker-permissive or employer-restrictive. Suppose each contract i is indexed by two latent variables: worker power $\beta_i \in [0, 1]$ and rigidity $r_i \in [0, \infty)$. We can then write the likelihood of the data given the parameters as:

$$\Pr(\text{data}|\text{parameters}) = \Pi_I \Pr(N_i, \{X_{ij}\}_{j \leq N_i}, \{\{f_{ij}\}_L\}_{j \leq N_i} | r_i, \beta_i).$$

To identify the parameters r_i and β_i we make a set of assumptions.
Assumptions:

1. The object of a clause is independent of the subject-verb prefix.

2. The number of clauses in a contract is independent of the content of clauses, and has a Poisson distribution with parameter r_i

3. The probability that a clause is granting control to a worker is i.i.d across clauses with probability given by the latent degree of worker power, so$X_{ij} \sim Binomial(\beta_i)$.

4. Object phrases in the same section of a contract are governed by a LDA (Latent Dirichlet Allocation (see e.g. Blei, 2012)) topic model with K topics (where topics can be understood as types of actions), with parameters θ_k for multinomial distribution over L noun phrases. Then can assume object is described by K-vector of topic proportions $\{\alpha_{ij}\}_K$ over the noun phrases. We used the "action" segment of the clause, which includes the other pieces of the parse tree besides the subject, modal, special verbs, and stopwords. To train the model, we treated each contract section as a document. We obtained interpretable results with 20 topics.

Under these assumptions, the log-likelihood function becomes:

$$Log(\Pr(\text{data}|\{\alpha_{ij}\}_K, \theta_1 ... \theta_K, \beta_i, r_i) = \underbrace{-r_i - log(N_i!) + log(r_i)N_i}_{r_i = Rigidity} +$$

$$\underbrace{\sum_{j=1}^{N_i} X_{ij} log(\beta_i) + (1 - X_{ij}) log(1 - \beta_i)}_{\beta_i = \text{WorkerAuthority}} + \underbrace{Log \Pr(f_{ij}|\{\alpha_{ij}\}_K, \theta_1 ... \theta_K)}_{LDA}$$

Clearly, the maxima of this function are the estimated parameter of Poisson distribution over number of clauses in contract ($r_i = N_i$) and the estimated parameter of Binomial distribution $\beta_i = \frac{\sum_j X_{ij}}{N_i}$, the share of pro-worker subject-verb combinations, along with the LDA topic parameters, α_{ij} and θ_k. The first two parameters are very simple to compute and require no estimation, but the derivation from the likelihood shows that our framework can be extended to include more complicated dependencies and priors over β and r. Table 8 provides a list of the estimated LDA topics, with the associated words. A handful of junk topics (0, 3, 6, 11, 12, and 13) have been excluded, leaving 14 interpretable topics to help us understand the content of collective bargaining agreements. Table 9 gives the distribution across topics in our data set. There is a relatively even distribution over topics across contracts. The most frequent topics are Topic 8 (Grievances) and Topic 18 (Insurance/Benefits). These topics get the most text dedicated to them in our sample of contracts. We looked at the topic shares over time and did not see any notable changes in our sample period.

Figure 5 plots our authority measures by topic, and by agent group. The top panel includes statements for worker and union, and the bottom panel includes statements for firm and manager. Employees receive entitlements and permissions – which can be understood as amenities and authorities. In turn, firms have obligations and prohibitions imposed, but do not receive entitlements/permissions. This is consistent with these contracts being designed to protect employees from unemployment risk, from work-related disutility, and potential abuse by managers. These types of protections could be efficiency-enhancing in labor markets characterized by monopsony, asymmetric information, or holdup.

The distribution of entitlement shares is depicted in Figure 6. One can see that it is approximately normal and has a higher mean for workers.

Table 6: Contract Statement Typology

Categorization Logic	Examples
Obligations	
Positive, Strict Modal, Active Verb	shall be, shall provide, shall include, shall notify, shall continue
Positive, Strict Modal, Obligation Verb	shall be required, shall be expected, shall be obliged
Positive, Non-Modal, Obligation Verb	is required, is expected
Prohibitions	
Negative, Any Modal, Active Verb	shall not exceed, shall not use, shall not apply, shall not discriminate
Negative, Permission Verb	shall not be allowed, is not permitted
Positive, Strict Modal, Constraint Verb	shall be prohibited, shal be restricted
Permissions	
Positive, Non-Modal, Permission Verb	is allowed, is permitted, is authorized
Positive, Strict Modal, Permission Verb	shall be allowed, shall be permitted
Positive, Permissive Modal, Active Verb	may be, may request, may use, may require, may apply
Negative, Any Modal, Constraint Verb	shall not be restricted, shall not be prohibited
Entitlements	
Strict Modal, Passive Verb	shall be paid, shall be given, shall not be discharged
Positive, Strict Modal, Entitlement Verb	shall have, shall receive, shall retain
Negative, Any Modal, Obligation Verb	may not be required

Table 7: Summary Statistics: Statements Per Contract

Variable	Mean	Std. Dev.	Min	Max
Active Verbs	441.01	374.07	0.00	8501.00
Passive Verbs	221.88	156.74	0.00	2053.00
Modal Verbs	332.49	229.33	0.00	2797.00
Special Verbs	72.11	50.06	0.00	820.00
Obligation Verbs	11.55	10.38	0.00	190.00
Constraint Verbs	0.27	0.66	0.00	14.00
Permission Verbs	4.45	4.94	0.00	96.00
Entitlement Verbs	32.88	24.01	0.00	412.00
Promise Verbs	22.97	18.92	0.00	381.00
Obligations	427.77	367.22	0.00	8443.00
Constraints	23.48	18.35	0.00	235.00
Permissions	4.09	4.53	0.00	83.00
Entitlements	241.24	168.49	0.00	2248.00
Total Statements	718.73	519.15	1.00	9626.00

Table 8: LDA Topic Words in Collective Bargaining Agreement Clauses

1 -- "**Sick Leave**" -- period month sick leave six probationary credit three complete employment twelve absent completion accumulate date exceed consecutive professional
2 -- "**Parental Leave**" -- leave absence pay request date grant prior week parental commencement pregnancy write maternity duty witness advance approve notice
4 -- "**Payroll**" -- change due result deduction amount status deduct monthly payroll reduction affect cheque technological fee employment orientation statement
5 -- "**Bargaining Unit**" -- unit bargaining person appointment appoint employ outside activity membership represent agent terminal sole select exercise ontario bargain behalf
7 -- "**Overtime**" -- hour shift work schedule overtime period call rest meal half minute start end break duty sunday weekend saturday two friday
8 -- "**Grievances**" -- grievance party procedure arbitration writing decision write step matter arbitrator committee complaint submit final dispute request name process
9 -- "**Job Training**" -- requirement operation training require equipment individual meet service responsibility provide program area manner performance" business duty operational
10 -- "**Vacation Leave**" -- year vacation service pay date employment week continuous effective two annual entitlement percent january salary earn termination period follow
14 "**Medical Leave/Injuries**" medical reasonable illness reason certificate unable duty injury course require due provide information circumstance accident personal condition examination reasonably
15 -- "**Discipline/Firing**" -- school act safety committee health action discharge labour cause discipline disciplinary file application canada public relations suspension regulation authority accordance
16 -- "**Seniority**" -- seniority lay position list layoff vacancy recall transfer post temporary qualification permanent job hire fill date provide ability copy basis
17 -- "**Work-Related Deaths**" – article accordance law child spouse pursuant family death include immediate parent purpose require city office paragraph funeral
18 -- "**Insurance/Benefits**" -- benefit plan insurance payment cost premium eligible provide receive compensation disability pay coverage pension receipt term amount
19 -- "**Scheduling**" -- work hour day week schedule two return perform normal regular report normally excess regularly require notice eight teaching available emergency

Table 9: Summary Statistics on Topic Proportions

Variable	Mean	Std. Dev.	Min	Max
Topic 0	.0434	.0248	0	.6100
Topic 1	.0309	.0192	0	.6833
Topic 2	.0465	.0325	0	.6620
Topic 3	.024	.0279	0	.8403
Topic 4	.0547	.0308	0	.6387
Topic 5	.0423	.0268	0	.6833
Topic 6	.0730	.0440	0	.8943
Topic 7	.0329	.0276	0	.8522
Topic 8	.0719	.0509	0	1
Topic 9	.0689	.0407	0	.794
Topic 10	.0429	.0338	0	.81
Topic 11	.0415	.0267	0	.668
Topic 12	.0488	.0297	0	.593
Topic 13	.0423	.0251	0	.847
Topic 14	.0562	.0318	0	.670
Topic 15	.059	.0332	0	1
Topic 16	.0624	.0369	0	.7625
Topic 17	.0405	.0259	0	.525
Topic 18	.0714	.0545	0	1
Topic 19	.0446	.0290	0	.683

Table 10: Relative Worker Control, By Industry Grouping

Industry Grouping	Relative Worker Control	
	Mean	*Standard Error*
Construction	-.627	.028
Educational, Health	-.071	.014
Entertainment/Hospitality	-.124	.062
Finance, real estate	-.001	.044
Information and culture	.152	.039
Manufacturing	.148	.015
Primary industries	.112	.049
Public administration	.228	.022
Transportation	-.013	.02
Utilities	.341	.044
Wholesale/Retail Trade	-.092	.052

Figure 5: Permissions and Entitlements by Agent and Topic

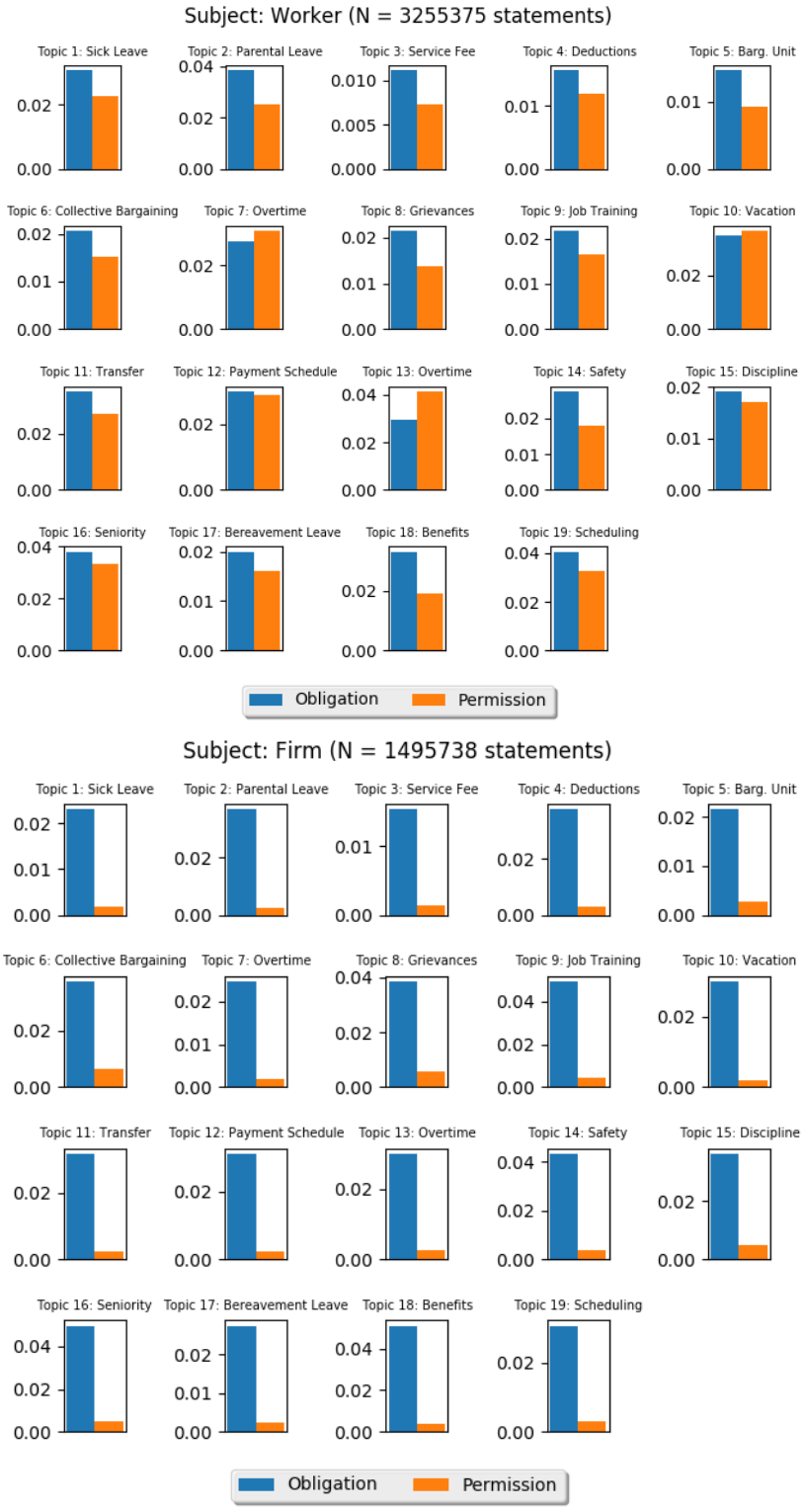
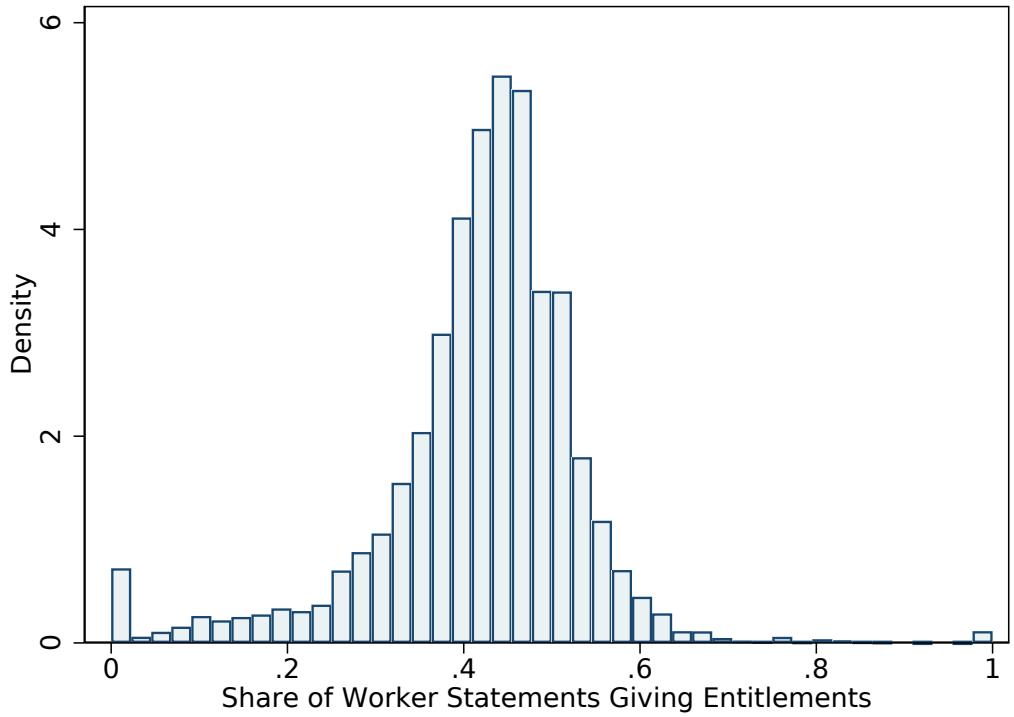


Figure 6: Distribution of Entitlement Shares for Workers and Employers



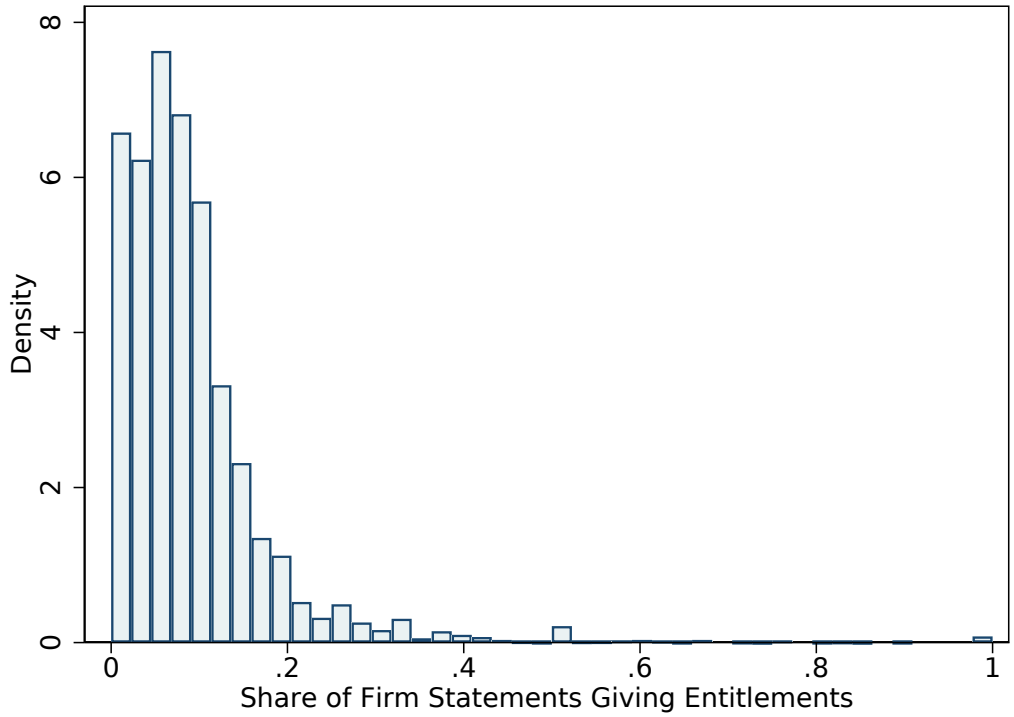


Figure 7: Relative Worker Control Over Time, and By Contract Length

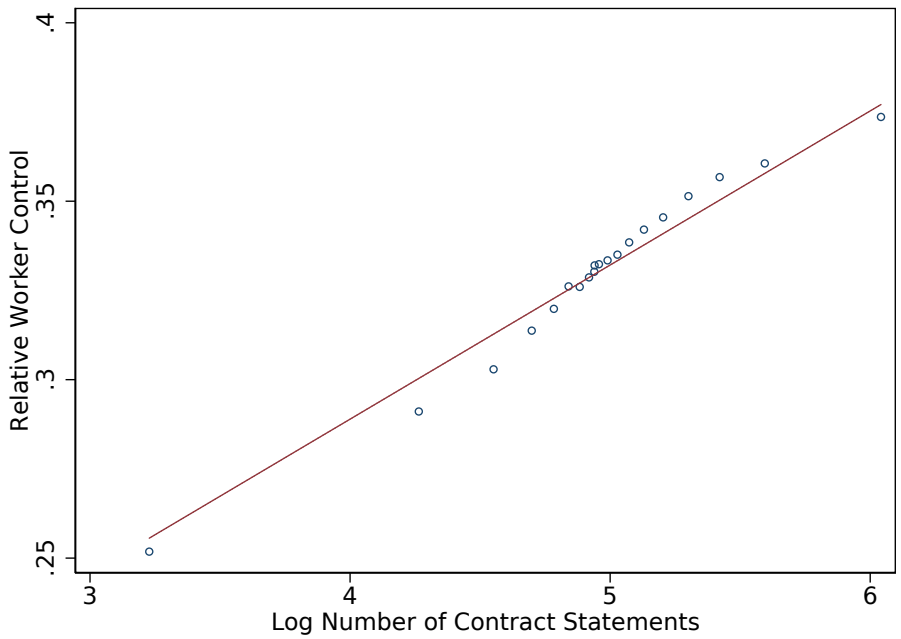


Table 11: Relative Worker Control, By Union

Union	# of Contracts		Relative Worker Control
	Private-Sector	Public-Sector	
Alberta Union of Provincial Employees	64	323	0.3770774
Unifor	141	27	0.3714532
Communications, Energy and Paperworke..	1160	84	0.3710064
Industrial Wood and Allied Workers of..	109	0	0.3707015
Public Service Alliance of Canada	323	898	0.3706359
Ontario Nurses' Association	27	1295	0.3657097
Nova Scotia Government and General Em..	9	93	0.362997
British Columbia Government and Servi..	61	66	0.360454
Canadian Union of Public Employees	282	3655	0.3599985
Ontario Public Service Employees Union	12	430	0.3575886
Office and Professional Employees Int..	71	56	0.3561289
United Food and Commercial Workers Ca..	315	26	0.3537522
National Automobile, Aerospace, Trans..	1402	262	0.3537032
International Association of Machinis..	482	11	0.3528805
International Brotherhood of Boilerma..	99	0	0.3511368
Elementary Teachers' Federation of On..	0	149	0.3488342
United Steel, Paper and Forestry, Rub..	443	28	0.3464268
Professional Institute of the Public ..	7	176	0.3461352
British Columbia Teachers' Federation	0	173	0.343406
United Steelworkers of America	929	39	0.3419327
Teamsters Canada	611	30	0.3405123
International Brotherhood of Electric..	303	206	0.3354559
National Automobile, Aerospace and Ag..	229	6	0.3311382
Canadian Merchant Service Guild	259	51	0.3291438
United Food and Commercial Workers In..	659	104	0.3290809
International Union of Operating Engi..	277	95	0.3277051
International Brotherhood of Teamsters	949	52	0.3269495
Seafarers' International Union of Can..	187	0	0.3222763
Service Employees International Union	167	1031	0.321059
Canadian Paperworkers Union	119	0	0.320908
International Association of Fire Fig..	6	328	0.3115588
Hotel Employees and Restaurant Employ..	101	12	0.3109177
Ontario Secondary School Teachers' Fe..	7	615	0.3090633
Ontario Public School Teachers' Feder..	0	102	0.3064964
Ontario English Catholic Teachers' As..	0	316	0.3001441

The measure of authority captured β_i is the proportion of statements that serve as entitlements for workers (or obligations for firms). Figure 7 plots the correlation between r_i(contract length) and β_i(Relative Worker Control) across contracts, showing quite a strong correlation (values residualized on company fixed effects and sector-year fixed effects). Longer contracts give more control to workers, consistent with a world in which firms have the default authority within the employment relationship, and reflecting the idea that the primary function of contracts is to protect workers; longer contracts provide more protections.

Table 10 reports the mean relative worker control by major industry grouping. Workers have relatively more control in Utilities and Public Administration. They have relatively less control in Construction and Entertainment/Hospitality (restaurants).[5] These ideas are echoed in the ranking in Table 11, which gives the unions with the highest relative worker control.

As some validation for our measure, we merged our dataset with survey data from a sample of firms from the World Management Survey (Bloom et al., 2012). We were able to link data on 52 Canadian manufacturing firms. In these surveys, they ask about human resource management practices on a 1-5 scale. We constructed an index for pro-worker HR practices as follows. Items that increase our index are "managers care about workers," "promotes good workers," and "employees are valued." Items that decrease the index are "focus on top talent," "incentives," and "fire poor performers." A binscatter of our authority measure against the index is reported in Figure 8.

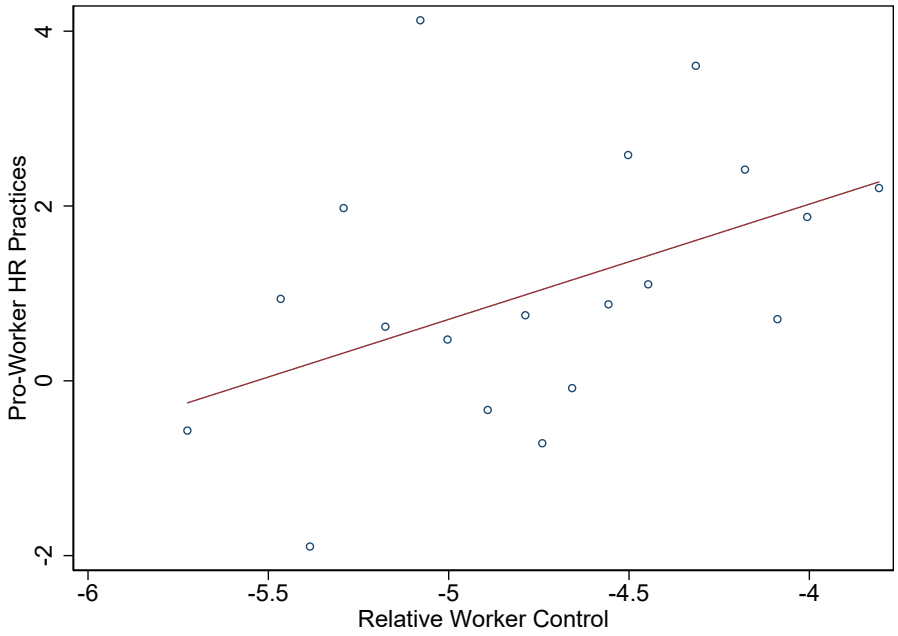
5 Regression Approach

We are interested in measuring how observed worker authority in Canadian collective bargaining agreements responds to changes in economic, legal, and political conditions. Here we describe our econometric analysis in more detail.

In our data an observation is a contract, indexed by province s, firm i, and effective year t. For each contract we have a set of outcomes, represented by y_{sit}. We use a

[5] In parallel work we are exploring a more principled measure, motivated by the idea that relative agent authority would generate a correlation between agent type and modal verb type at the statement level. To summarize, worker authority in a contract could be given by the within-topic correlation coefficient (or OLS coefficient) across statements between agent indicators and indicators for a permissive vs, restrictive modal verb. More generally, there are promising avenues for structural modeling of the contract drafting process.

Figure 8: Worker Authority and Pro-Worker HR Practices



Binscatter diagram of worker authority (horiztonal axis) and an index for Pro-Worker HR Practices, constructed from the World Management Survey.

Table 12: List of Variables Used in Regression Analysis

Variable Label	Description
Contract Feature Outcome Variables	
Relative Control	Employee Entitlements plus Employer Obligations
Endogenous Descriptive Variables	
Log Employees	Log number of employees covered by contract
Log Duration	Log of expiry month minus effective motnh
Has COLA Clause	Indicator equaling one if contract has COLA
COLA Amount	Conditional on having COLA, average annualized change
Exogenous Treatment Variables	
Negative Wage Shock	Indicator equaling one if inflation beats COLA during previous contract
Unemployment Rate	Province-sector-year unemployment rate
Log Tax Rate	Log of the province-year- implicit personal income tax rate
NDP Control	Indicator: New Democratic Party controls provincial government

linear model

$$y_{sit} = \rho z_{sit} + \alpha_{sit} + X'_{sit}\beta + \epsilon_{sit},$$

with the components described as follows. First, z_{sit} is the explanatory variable of interest, with ρ giving the coresponding OLS coefficient. Depending on the specification and associated assumptions, $\hat{\rho}$ may or may not estimate a causal relationship. Second, α_{sit} includes a set of fixed effects, which may include indicators for year, province, sector, or company. It may also include interacted fixed effects. Third, X_{sit} includes a set of time-varying controls, for use in assessing robustness of $\hat{\rho}$. Finally, ϵ_{sit} is an error term. In all regression results, we cluster standard errors by province (e.g. Bertrand et al., 2004).

Our outcome variables and treatment variables, as labeled in the tables, are listed in Table 12. We provide a description as well.

6 Results

This section reports our results. We report two sets of results. First, we look at how contract features respond to local economic and political conditions that affect outside

Table 13: Regression Estimates: Local Sectoral Unemployment Rate

	Effect on Worker Control							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Unemp.Rate	-0.112**	-0.116**	-0.116**	-0.102*	-0.107**	-0.140**	-0.144**	-0.106*
	(0.0409)	(0.0412)	(0.0311)	(0.0402)	(0.0387)	(0.0379)	(0.0518)	(0.0448)
Province-Sector FEs	X							
Sector-Year FEs	X	X	X	X	X	X	X	X
Firm FE's		X	X	X	X	X	X	X
Cluster by Province			X					
Province Trends				X				
Contract Controls					X			
Topic Controls						X		
Economy Controls							X	
Government Controls								X

options, relative pricing of amenities, and worker bargaining power. Second, we look at the strike response to unexpected wage cuts, and how that varies according to contract features.

6.1 Effects of Economics and Political Conditions on Contract features

This section looks at external influences of contract features. We look at a set of factors that, conditional on the fixed effects, are exogenous to the features of the contract.

The first economic variable that we look at is the unemployment rate at the time of contract negotiation. This can be seen as a negative shock to worker outside options. These coefficients are reported in Table 13. A higher unemployment rate is associated with a statistically significant decrease in measured worker control. The estimate is robust to a range of alternative speciticications.

Next we look at province-wide shocks to wages due to tax policy. In Table 14, we look at whether changes to the within-province tax rate are associated with changes to text features. We find that a higher income tax is associated with higher worker authority. This is consistent with a substitution away from the taxed income (wages) toward untaxed income (amenities). The result is robust to a range of specifications.

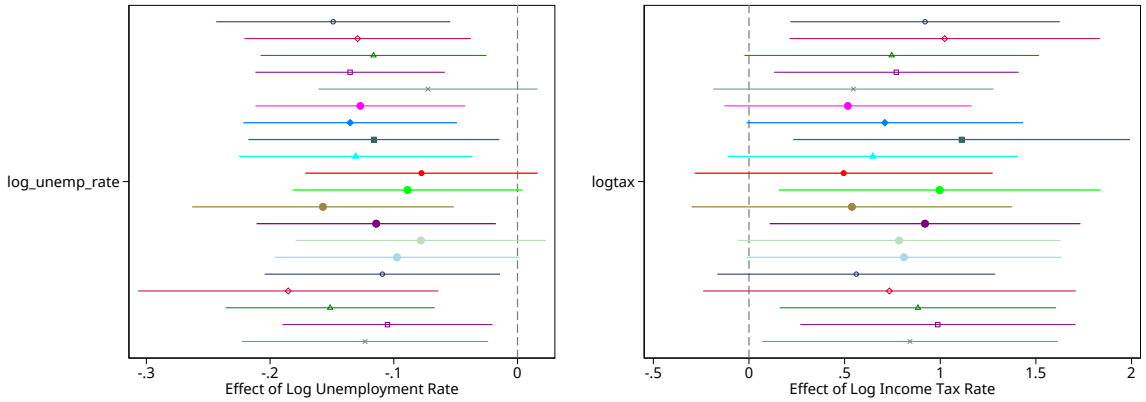
Table 14: Cross-Sectional and Panel Differences: Personal Income Tax Rate

	Effect on Worker Control							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Income Tax Rate	1.651**	0.772*	0.772*	1.053**	0.812*	0.912**	1.030*	0.764*
	(0.352)	(0.355)	(0.338)	(0.384)	(0.340)	(0.329)	(0.467)	(0.378)
Province-Sector FEs	X							
Sector-Year FEs	X	X	X	X	X	X	X	X
Firm FE's		X	X	X	X	X	X	X
Cluster by Province			X					
Province Trends				X				
Contract Controls					X			
Topic Controls						X		
Economy Controls							X	
Government Controls								X

Table 15: Cross-Sectional and Panel Differences: New Democratic Party Control

	Effect on Worker Control							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NDP Control	0.113**	0.0511*	0.113**	0.0998**	0.125**	0.110**	0.0809*	0.117**
	(0.0146)	(0.0224)	(0.0146)	(0.0264)	(0.0261)	(0.0237)	(0.0325)	(0.0272)
Sector-Year FEs	X	X	X	X	X	X	X	X
Firm FE's	X	X	X	X	X	X	X	X
All firms		X						
Cluster by Province			X					
Province Trends				X				
Contract Controls					X			
Topic Controls						X		
Economy Controls							X	
Government Controls								X

Figure 9: What Topics are Changing?



Topic 7 (turquoise, square) is "Payment Schedule"; Topic 9 (bright red) is "Job Training"; Topic 10 (bright green) is "Vacations"; Topic 16 (dark red, diamond) is "Seniority".

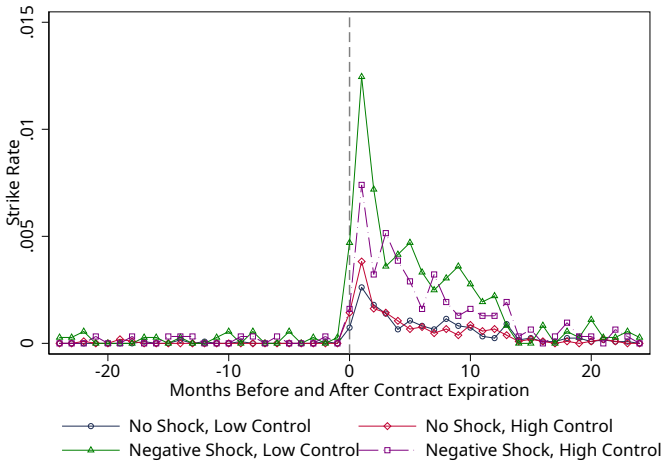
Our third set of results looks at how contracts respond to political conditions (Table 15). We find that relative to other parties, control of province government by NDP (New Democratic Party, which is known to support the labor movement) has empowering effects on labor unions in the private sector. NDP control is associated with contracts that give more authority to employees. This is consistent with stronger worker bargaining power due to political support.

As an additional application of our topics data, we look at heterogeneous impacts of unemployment and labor income taxes on worker control by topic. We see that higher taxes are associated with more entitlements related to vacation, consistent with a substitution to an un-taxed amenity.

6.2 Worker Authority and Strike Responses to a Wage Cut

This section looks at the impacts of an unexpected wage cut on strike rates. We follow Card (1986) and use the (log) gap between actual inflation and the cost-of-living adjustment (COLA) clause in the last agreement as a real wage shock at the time of contract expiration. The basic empirical pattern we illustrate is shown in Figure 10. This graph shows the strike rates before and after contract expiration. As expected, strikes tend to happen around contract expiration dates. We can see that, as in the regression, the strikes are more likely when there is a negative wage shock due to a low

Figure 10: COLA-Inflation Wage Shock, Worker Control, and Strike Rates



COLA. But we can also see that the spike is smaller when the workers have relatively more control encoded in the contract.

Figure 11 shows the correlation between the real wage shock and the number of strikes, residualizing the variables on province-sector-year fixed effects. We find that strikes are less likely when COLA does better than inflation; that is, negative wage shocks increase the probability of a strike. In a regression, this coefficient is significant with $p = .022$.

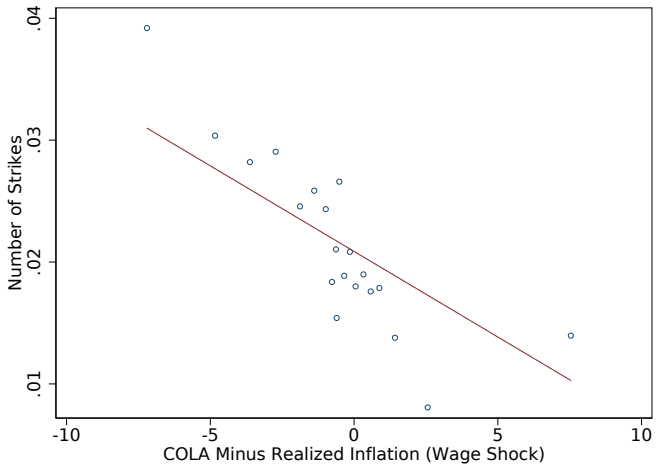
Table 16 reports a set of regressions with the number of strikes and the number of work days on strike as outcome variables. As treatment variables, these regressions include High Worker Control (an indicator for above-median worker control in a province-sector-year), Negative Wage Shock (an indicator for a COLA clause not keeping up with inflation in the previous contract), and the interaction between the two. We report results with province-sector-year fixed effects, and with these fixed effects plus firm fixed effects.

The regressions show the following. While there is no real difference in strike rates or intensity according to the level of worker control, there is an increase in strikes due to negative wage shocks. However, with high worker control, there is a significantly smaller effect of the negative wage shock on strikes. This interaction effect is only seen in the private sector. Note that we ran the same regression with contract length (rather than relative worker control) and found no effect.

Table 16: COLA-Inflation Wage Shock, Worker Control, and Strike Intensity

Private Sector Firms				
	(1)	(2)	(3)	(4)
	Has Strike		Work Days on Strike	
High Worker Control	0.00206	0.00410	0.153	0.0405
	(0.00272)	(0.00551)	(0.105)	(0.249)
Negative Wage Shock	0.0179**	0.0151*	0.808**	0.655
	(0.00371)	(0.00684)	(0.182)	(0.529)
High Control * Negative Shock	-0.00883*	-0.0200*	-0.785**	-1.273*
	(0.00403)	(0.00861)	(0.218)	(0.425)
Public Sector Firms				
	(1)	(2)	(3)	(4)
	Has Strike		Work Days on Strike	
High Worker Control	0.00138	-0.0205	0.00689+	0.120
	(0.00257)	(0.0904)	(0.00323)	(0.0945)
Negative Wage Shock	0.0178*	0.337+	0.0140**	0.325+
	(0.00742)	(0.177)	(0.00450)	(0.172)
High Control * Negative Shock	-0.00323	-0.0955	-0.00643	-0.224
	(0.00507)	(0.109)	(0.00533)	(0.168)
Province-Sector-Year FEs	X	X	X	X
Firm FE's		X		X

Figure 11: Effect of COLA-Inflation Wage Shock on Strike Intensity



7 Conclusion

This paper has provided empirical evidence of how labor union contracts respond to changes in the economic and political environment. We showed that in the main, labor contracts impose obligations on both workers and firms, and give entitlements to workers. The strength of those worker entitlements varies across firms and over time in response to a range of factors.

In addition to uncovering "what unions do," this research aims to uncover "what unions want" – and inform what future collective bargaining institutions might look like. By comparing contracts with strong unions to contracts with weak unions, we can produce statistical evidence on what types of clauses – amenities, obligations, entitlements, and protections – unions tend to bargain for. These dimensions of workplace autonomy are difficult to measure with traditional datasets, but may be an important component of well-being on the job. Indeed, while unions almost certainly compressed the income distribution, our project aims to document their further effects on workplace control rights and amenities. The lessons from these contracts will help policymakers design labor-market rules that efficiently govern workplace amenities, rent-sharing, and control rights within the firm. Given the recent emphasis on heterogeneity in firms as a source of wage inequality, understanding the firm-specific institutions that govern pay practices is important for unpacking the income distribution.

In the domain of law, while there is extensive theorizing about contract language, there is little credible empirical evidence. A fundamental identification problem is that the terms used in contracts are rarely tested in court, so it is difficult to decide what is boilerplate and what is probative. Our measurement and identification approach provides one way to gain some traction on this problem. Further, for practitioners of labor law, having an annotated database of the kinds of clauses unions have demanded in the past will likely be of some value in designing and negotiating future collective bargaining agreements.

In addition, we can provide evidence on what clauses increase the quality of the firm-employee relationship and increase efficiency. While a now considerable literature has measured firm productivity, comparatively little has gone into measuring firm amenities. Collective bargaining agreements might provide one way to get a sense of historical variation in quality of the workplace.

These findings will be relevant to ongoing debates within the labor movement, both public and private, about what unions have to offer workers in the 21st century. It might be that many of the contractual provisions that unions offered are no longer demanded by workers because they are now protected by law.

References

Abowd, J. A. and Lemieux, T. (1993a). The effects of product market competition on collective bargaining agreements: The case of foreign competition in canada. *The Quarterly Journal of Economics*, 108(4):983–1014.

Abowd, J. M. (1989). The effect of wage bargains on the stock-market value of the firm. *American Economic Review*, 79(4):774–800.

Abowd, J. M. and Lemieux, T. (1993b). The effects of product market competition on collective bargaining agreements: the case of foreign competition in Canada. *Quarterly Journal of Economics*, 108(4):983–1014.

Ackerberg, D. A. and Botticini, M. (2002). Endogenous matching and the empirical determinants of contract form. *The Journal of Political Economy*, 110(3):564–591.

Alchian, A. and Demsetz, H. (1972). Production, information costs, and economic organization,. *American Economic Review*, 62(5):777–795.

Allen, D. and Lueck, D. (1992). Contract choice in modern agriculture: cash rent versus cropshare. *Journal of Law and Economics*, pages 397–426.

Altonji, J. G. and Devereux, P. J. (1999). The extent and consquences of downward rigidity. Technical Report 7236, NBER.

Anderlini, L. and Felli, L. (1994). Incomplete written contracts: Undescribable states of nature. *The Quarterly Journal of Economics*, 109(439):1085–1124.

Ash, E. (2016). The political economy of tax laws in the u.s. states. Technical report.

Ash, E., Morelli, M., and Osnabrugge, M. (2017a). Electoral systems, polarization, and party cohesion: Evidence from new zealand's electoral reform. Technical report.

Ash, E., Morelli, M., and Van Weelden, R. (2017b). Elections and divisiveness: Theory and evidence. *Journal of Politics*.

Ashenfelter, O. and Johnson, G. E. (1969). Bargaining theory, trade unions, and industrial strike activity. *American Economic Review*, 59(1):35–49.

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

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

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

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Blau, F. and Beller, A. (1988). Trends in earnings differentials by gender, 1971-1981. *Industrial & Labor Relations Review*, 41(4):513–529.

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

Bloom, N., Genakos, C., Sadun, R., and Reenen, J. V. (2012). Management practices across firms and countries. NBER Working Papers 17850, National Bureau of Economic Research, Inc.

Brown, J. N. and Ashenfelter, O. (1986). Testing the efficiency of employment contracts. *Journal of Political Economy*, 94:S40–S87.

Budd, J. W. and Wang, Y. (2004). Labor policy and investment: Evidence from canada. *ILR Review*, 57(3):386–401.

Card, D. (1986a). Efficient contracts with costly adjustment: Short-run employment determination for airline mechanics. *American Economic Review*, 76(5):1045–1071.

Card, D. (1986b). An empirical model of wage indexation provisions in union contracts.

The Journal of Political Economy, 94(3):S144–S175.

Card, D. (2001). The effect of unions on wage inequality in the us labor market. *ILR Review*, 54(2):296–315.

Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2016). Firms and labor market inequality: Evidence and some theory. Technical report, National Bureau of Economic Research.

Card, D. and De La Rica, S. (2006). Firm-level contracting and the structure of wages in spain. *Industrial and Labor Relations Review*, pages 573–592.

Card, D., Devicienti, F., and Maida, A. (2014). Rent-sharing, holdup, and wages: Evidence from matched panel data. *The Review of Economic Studies*, 81(1):84–111.

Ceci, M., Lesmo, L., Mazzei, A., Palmirani, M., and Radicioni, D. P. (2011). Semantic annotation of legal texts through a framenet-based approach. In *International Workshop on AI Approaches to the Complexity of Legal Systems*, pages 245–255. Springer.

Coase, R. (1937). The nature of the firm. *Economica*, 4:386–405.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044.

DiNardo, J. and Lee, D. S. (2004). Economic impacts of new unionization on private sector employers: 1984-2001. *Quarterly Journal of Economics*, 119(4):1383–1441.

Doucouliagos, C. and Laroche, P. (2003). What do unions do to productivity? a meta-analysis. *Industrial Relations*, 42(4):650–691.

Doucouliagos, H. and Laroche, P. (2009). Unions and profits: A meta-regression analysis. *Industrial Relations*, 48(1):146–184.

Dye, R. A. (1985). Optimal length of labor contracts. *International Economic Review*, 26(1):251–270.

Farber, H., Herbst, D., Kuziemko, I., and Naidu, S. (2017). Unions and inequality: A historical perspective. Technical report.

Farber, H. S. and Western, B. (2001). Accounting for the decline of unions in the private sector, 1973-98. *Journal of Labor Research*, 22(3):459–85.

Farnsworth, E. A. (2004). *Contracts, 4th edition*. Aspen Publishers, New York.

Francesconi, E. and Passerini, A. (2007). Automatic classification of provisions in legislative texts. *Artificial Intelligence and Law*, 15(1):1–17.

Freeman, R. B. and Kleiner, M. M. (1990). The impact of new unionization on wages

and working-conditions. *Journal of Labor Economics*, 8(1):S8–S25.

Freeman, R. D. and Medoff, J. L. (1984). *What do unions do?* Basic Books, New York, NY.

Ganglmair, B. and Wardlaw, M. (2017). Complexity, standardization, and the design of loan agreements. Technical report.

Garloff, A. and Guertzgen, N. (2012). Collective wage contracts, opt-out clauses, and firm wage differentials: Evidence from linked employer–employee data. *Industrial Relations: A Journal of Economy and Society*, 51(3):731–748.

Gentzkow, M., Kelly, B. T., and Taddy, M. (2017). Text as data. Technical report, National Bureau of Economic Research.

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71.

Grossman, S. J. and Hart, O. D. (1986a). The costs and benefits of ownership: A theory of vertical and lateral integration. *The Journal of Political Economy*, 94(4):691–719.

Grossman, S. J. and Hart, O. D. (1986b). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4):691–719.

Grout, P. (1984). Investment and wages in the absence of binding contracts: A Nash bargaining approach. *Econometrica*, 52(2):449–460.

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

Hart, O. and Moore, J. (1999). Foundations of incomplete contracts. *Review of Economic Studies*, 66(1):115–138.

Hart, O. D. and Moore, J. H. (1990). Property rights and the nature of the firm. *Journal of Political Economy*, 98:1119–58.

Hart, O. D. and Moore, J. H. (2007). Contracts as reference points. *Quarterly Journal of Economics*, 123(1):1–48.

Holmes, T. (1998). The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of Political Economy*, 106:667–705.

Jansen, H. J. and Young, L. (2009). Solidarity forever? the ndp, organized labour, and the changing face of party finance in canada. *Canadian Journal of Political Science/Revue canadienne de science politique*, 42(3):657–678.

Jaumotte, M. F. and Osorio, M. C. (2015). *Inequality and labor market institutions*. International Monetary Fund.

Jordan, L. and Bruno, R. (2005). Do the organizing means determine the bargaining

ends? *Advances in Industrial & Labor Relations*, 14:101–125.

Joskow, P. (1987). Contract duration and relation-specific investments: Empirical evidence from coal markets. *American Economic Review*, 77:168–185.

Jurafsky, D. and Martin, J. H. (2014). *Speech and language processing*, volume 3. Pearson London.

Juravich, T., Bronfenbrenner, K., and Hickey, R. (2006). An analysis of union first contracts. *Justice on the Job: Perspectives on the Erosion of Collective Bargaining in the United States*, page 87.

Klein, B., Crawford, R., and Alchian, A. (1978). Vertical integration, appropriable rents, and the competitive contracting process. *Journal of Law and Economics*, 21(2):297–326.

Kornhauser, L. and MacLeod, W. B. (2012). Contracts between legal persons. In Gibbons, R. and Roberts, J., editors, *Handbook of Organizational Economics*, chapter 23. Princeton University Press.

Kratzer, A. (1991). Modality. In *Semantics: An international handbook of contemporary research*. Berlin: De Gruyter.

Kreps, D. M. and Wilson, R. B. (1982). Sequential equilibria. *Econometrica*, 50:863–894.

Kuhn, P. (1998). Unions and the economy: what we know; what we should know. *Canadian Journal of Economics-Revue Canadienne D'Economique*, 31(5):1033–1056.

Lee, D. S. and Mas, A. (2012). Long-run impacts of unions on firms: New evidence from financial markets, 1961–1999. *The Quarterly Journal of Economics*, 127(1):333–378.

Leffler, K. B. and Rucker, R. R. (1991). Transactions costs and the efficient organization of production: A study of timber-harvesting contracts. *Journal of Political Economy*, pages 1060–1087.

Machin, S. (2000). Union decline in Britain. *British Journal of Industrial Relations*, 38(4):631–645.

MacLeod, W. B. (2002). Complexity, bounded rationality and heuristic search. *Contributions to Economic Analysis and Policy*, 1(1).

MacLeod, W. B. (2011). Great expectations:law, employment contracts, and labor market performance. In Ashenfelter, O. and D., C., editors, *Handbook of Labor Economics*, volume 4. Elsevier.

MacLeod, W. B. and Malcomson, J. M. (1993). Specific investment and wage profiles in labour markets. *European Economic Review*, 37(2/3):343–354.

Manning, A. (2010). Imperfect competition in labor markets. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics, Volume 4*, volume 4. Elsevier.

Maskin, E. and Tirole, J. (1999). Two remarks on the property-rights literature. *Review of Economic Studies*, 66(1):139–150.

Masten, S. E. and Crocker, K. J. (1985). Efficient adaptation in long-term contracts: Take-or-pay provisions for natural gas. *American Economic Review*, 75(5):1083–93. ISSN: 0002-8282 Journal Article.

Matvos, G. (2013). Estimating the benefits of contractual completeness. *Review of Financial Studies*, page hht060.

McCammon, H. J. (1990). Legal limits on labor militancy: Us labor law and the right to strike since the new deal. *Social Problems*, pages 206–229.

McDonald, I. M. and Solow, R. M. (1981). Wage bargaining and employment. *American Economic Review*, 71(5):896–908.

Milgrom, P. (1988). Employment contracts, influence actitivities, and efficient organizational design. *Journal of Political Economy*, 96:42–60.

Milgrom, P. and Shannon, C. (1994). Monotone comparative statics. *Econometrica*, 62(1):157–180.

Moszoro, M., Spiller, P. T., and Stolorz, S. (2016). Rigidity of public contracts. *Journal of Empirical Legal Studies*, 13(3):396–427.

Nickell, S. J. and Andrews, M. (1983). Unions, real wages and employment in britain 1951-79. *Oxford Economic Papers*, 35:183–206.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Pope, J. G. (2004). How american workers lost the right to strike, and other tales. *Michigan Law Review*, pages 518–553.

Posner, R. A. (2005). The law and economics of contract interpretation. *Texas Law Review*, 83:1581.

Roberts, M. E., Stewart, B. M., Tingley, D., Airoldi, E. M., et al. (2013). The structural topic model and applied social science. In *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation*.

Rothstein, M. A. and Liebman, L. (2003). *Employment Law*. Thomson West, New York, NY.

Sanga, S. (2014). Choice of law: An empirical analysis. *Journal of Legal Studies*

(forthcoming).

Savage, L. J. (1972 (first published 1954)). *The Foundations of Statistics*. Dover Publications, New York, N.Y.

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

Shavell, S. (2006). On the writing and the interpretation of contracts. *J Law Econ Organ*, 22(2):289–314.

Stepan-Norris, J. and Zeitlin, M. (2003). *Left out: Reds and America's industrial unions*. Cambridge University Press.

Strunk, K. O. and Grissom, J. A. (2010). Do strong unions shape district policies? collective bargaining, teacher contract restrictiveness, and the political power of teachers' unions. *Educational Evaluation and Policy Analysis*, 32(3):389–406.

Svejnar, J. (1986). Bargaining power, fear of disagreement, and wage settlements - theory and evidence from united-states industry. *Econometrica*, 54(5):1055–1078.

Taddy, M. (2013). Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, 108(503):755–770.

Talley, E. and O'Kane, D. (2012). The measure of a mac: A machine-learning protocol for analyzing force majeure clauses in m&a agreements. *Journal of Institutional and Theoretical Economics JITE*, 168(1):181–201.

Townsend, R. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 22:265–293.

Western, B. (1999). *Between class and market: Postwar unionization in the capitalist democracies*. Princeton University Press.

A Theoretical Framework

This section makes two contributions. First, we begin with the observation that the purpose of an employment contract is to set worker wages, and then the firm can decide how to allocate the worker's time to tasks within the scope of the job for which the worker is hired. In some cases the efficient assignment of workers to tasks depends upon unobserved worker characteristics. We show that in those cases it maybe efficient to add contract terms that constrain firm choice. This implies that under the appropriate conditions, increasing the number of contract terms can increase on-the-job amenities for workers.

The second contribution is to embed this model into a standard union-firm bargaining model (McDonald and Solow, 1981). This allows us to derive the implications of variation of the exogenous parameters upon employment, number of clauses, and wages.

A.1 Model Setup

Formal contract models in economics begin with Savage's (1954) setup.[6] One begins with a "model of the world" given by states $\omega \in \Omega$.[7] As Savage points out, one never directly observes states, but only *events* $E \subset \Omega$. For example, in an insurance case related to the 9/11 attacks a court had to determine whether the aircrafts hitting the buildings consisted of two events or one event (a coordinated attack). The court decided that it was in fact a single event for the purposes of the insurance contract.[8] The point here is that as a practical matter events in a contract may be imperfectly or incompletely defined, even though standard models of union contracting assume completeness.

[6] See Kreps and Wilson (1982) on the integration of Savage's model with economic models of games, of which contract theory can be viewed as a subfield.

[7] To reduce the technical jargon, let us suppose that Ω is finite, but very large. The fact that large state spaces imply large transactions costs should now be familiar to economists. Consider the example of Bitcoin mining, which entails nothing more that guessing a single number from a finite (but large) set. Nevertheless the cost of Bitcoin mining was approximately equal to the total energy needs of Ireland (at time of writing).

[8] The coverage by Allianz was for $78 million per event while the French firm SCOR was on the hook for $355 million per occurrence. It should be pointed out that having an aircraft hit a tower is a foreseeable event. The buildings were designed to survive such an event. What was unforeseen is two, fully fueled commercial aircraft hitting the towers in a coordinated attack. It was the burning of the jet fuel that led to the collapse of the building.

We suppose that a contract can be represented by a document that specifies obligations when certain events occur (Rothstein and Liebman, 2003; Farnsworth, 2004). Formally, a contract is defined by n obligations:

$$K = \{(E_i, o_i) \,|i = 0, 1, 2, ..., n-1\}.$$

Each obligation i corresponds to a clause in a written contract that specifies the obligation o_i in event E_i. Each clause must describe the event (for example: "if the worker is ill. . . ") and associated obligation (". . . then they will not be dismissed if they do not come to work that day.").

There is a literature that attempts to define precisely what is meant by a complete contract.[9] We sidestep this issue by observing that any contract can be made complete by defining the default event:

$$E_0 = \Omega \backslash \cup_{i=1}^{n-1} E_i,$$

and then requiring the use of a third party to resolve disputes in this event.[10] For commercial contracts, when obligation o_0 is missing from a contract, then normally the courts will be asked to adjudicate. In the case employment where new events occur day by day, the use of the courts is prohibitively expensive. In the case of union contracts the standard rule is the "right to manage" clause that gives the firm the right to ask the employee to carry out any task within the scope of the employment relationship. This assumption is at the heart of the modern theory of the firm, beginning with Coase (1937), Alchian and Demsetz (1972), Klein et al. (1978), Grossman and Hart (1986b), and Hart and Moore (1990).

Our analysis will provide a formal test of this theory by discussing its implications for the evolution of the number clauses over time. To this end, let K_{jt} be the contract for bargaining pair j at time t. Given a written document K, we define $n_{jt} = n(K_{jt})$ as the number of clauses in a contract. We can then use our data to address the question: How does n_{jt} vary over time and between bargaining pairs?

Notice that in addition to the number of clauses, we can also define:

$$l_{jt} = \frac{conditions_{jt}}{n_{ji}},$$

[9] See for example Anderlini and Felli (1994), Hart and Moore (1999) and Maskin and Tirole (1999).

[10] Commercial contracts use a variety of solutions, include specifying a legal jurisdiction such as New York State, or requiring the use of binding arbitration.

where $conditions_{jt}$ is the number of conditions or contingencies in a contract, which we measure below. The statistic l_{jt}, the number of conditions per clause, provides a measure of contingency or specificity of the contract. A clause can be expected to have more conditions if the event it is describing is more complex or more valuable. This statistic is not perfect since contracts has a section with definitions that are designed to provide compact descriptions of events and actors. The length of the definitions section will be used as an alternative measure of contract richness.

A.2 Why add a clause, and for whom?

Let's consider a union and firm who have agreed to the minimal contract, which specifies employment, L, and wage per worker, w. All workers are on full-time fixed-hours (40 per week) contracts. The question now is: Why would the parties wish to add any more clauses to the contract?

Consider the following motivating example. An airline faces the problem of assigning airline stewards to routes A or B.[11] The airline is indifferent to the allocation of these employees to the two routes, but the employees care a great deal due to family constraints that vary over time. Formally, the firm needs to allocate tasks A and B to workers 1 and 2. It is assumed that the two workers have identical productivity; however, they may have different private valuings to each task. For $i \in \{1,2\}, j \in \{A,B\}$ let

$$U_i(j) = w - \theta_{ij},$$

where $\theta_{ij} \sim N(m_i, \sigma^2)$ is a random draw.

Under a right-to-manage agreement, the airline might use a random allocation of individuals to tasks because it is cheap to implement and the airline has no information with which to distinguish the preferences of the stewards. This arrangement generates total expected welfare:

$$W^0 = E\left\{\frac{U_1(A) + U_2(B) + U_1(B) + U_2(A)}{2}\right\} = 2w - m_1 - m_2,$$

the outcome in the absence of a clause on task assignment.

The union at some point may realize that worker welfare could be improved with choice. In the airline steward case a common rule is to let more senior workers have

[11] See Milgrom (1988) for an extensive discussion of this point.

first choice over routes. Suppose that worker 1 is the senior worker, and the union negotiates a clause that allows the more senior employees to choose their preferred routes first. Let $\eta_i = \theta_{iA} - \theta_{iB} \sim N(0, 2\sigma^2)$ be the difference in the shocks and let

$$j_1(\eta_1) = \begin{cases} A, & if\ \eta_1 \geq 0, \\ B, & if\ \eta < 0, \end{cases}$$

notate the optimal task for for employee 1 (correspondingly the outcome for employee 2 is $j_2(\eta_1) = \{A, B\} \setminus j_1(\eta)$). Under this rule the total expected welfare is:

$$W^1 = E\left\{\frac{U_1(j_1(\eta_1)) + U_2(j_2(\eta_1))}{2}\right\} \tag{1}$$

$$= W^0 + 2\sigma\sqrt{\frac{2}{\pi}} \tag{2}$$

where $\sqrt{\frac{2}{\pi}} = \int_{-\infty}^{\infty} |x| f(x)\, dx > 0$ is the expected value of the absolute value of a normally distributed random variable.

This result has a number of implications. First, when there is private information on the side of employees, then efficiency can be increased with a reallocation of *ex post* decision rights. Second, the benefit of this decision right is increasing with the amount of uncertainty regarding the preferences of the employees. In large, diverse firms there are likely to be more opportunities for such rule changes, and hence we should expect efficient contracts in larger firms to have more clauses to cover the larger number of contingencies where there can be efficiency gains.

It is worth highlighting that the motivation for adding a clause is to enhance the value of the relationship for some, and not necessarily all, workers. In the motivating example, the seniority-choice rule increases the average utility for the senior worker, but does not affect the utility of the junior worker. This type of insight on subsets of workers is different from traditional union models, which focus on homogeneous measures such as wages and employment.

In addition, the motivation of this example for adding clauses for *firm-specific* events, that in turn result in tailored clauses for that firm's contract. Given that the events that lead to clauses are idiosyncratic, that implies heterogeneity in clauses across firms. It is *impossible* to have a general theory for the content of contracts. Instead, we focus upon what can be measured, such as the number or types of clauses

in contracts.

The theory implies that increasing the number of clauses can increase the rents that workers obtain from the relationship. Let n^W be the number of clauses that provide the worker rights/rents in the employment relationship. Given that the default rule is right-to-manage, the most basic union contracts should give workers new rights. However, the provision of just worker rights may not be optimal. In our example, the airline might want certain routes to have stewards with particular characteristics, such as language skills. In that case, the airline would add a clause limiting the free choice of stewards that ensures each international flight has enough crew with the needed language skill. Thus, as the number of worker clauses, n^W, rises, we might expect a corresponding rise in clauses that give the firm rights, denoted by n^F. Let $\vec{n} = \left\{n^W, n^F\right\}$ be the vector of counts for worker-rights and firm-rights clauses.

A.3 The Optimal Union Contract

Having established an economic motivation for the addition of contract terms, we next consider the implications for negotiated contracts. Standard contract theory supposes that there are no transactions costs, and hence parties add clauses if and only if this is efficient (McDonald and Solow, 1981; Svejnar, 1986; Brown and Ashenfelter, 1986)). . Under an efficient contract, all terms, including wage and employment, are set before production begins.

In the right-to-manage case (see Nickell and Andrews, 1983), it is assumed that parties negotiate over terms that determine contract length, $\vec{n}$, and wages, w, before production begins. Once production starts the firm sets employment as function of the wage and other parameters.

The exogenous parameters of the model are given by $\beta = \{\gamma, \bar{w}, \tau, \alpha\}$.[12] The parameter γ is the productivity of the firm. For larger values of γ, it is efficient to have a larger firm. Hence it will do double duty as a product demand shock and firm size parameter. The second parameter, $\bar{w}$, is the market wage. Both of these parameters are correlated with unemployment rates, and depending upon the market, changes in unemployment rates may affect one of these more than the other. The parameter τ is the wage income tax. The worker's after-tax wage is $(1-\tau)\,w$. Next is $\alpha \in [0,1]$, the bargaining power of the union. This can be affected by the political environment and laws that increase or decrease worker power.

[12]Below we will allow these to vary over time: at date t we will have $\beta_t = \{\gamma_t, \bar{w}_t, \tau_t, \alpha_t\}$.

Finally, we have the cost of adding a contract term, c. We do not include this as an exogenous parameter since it is not observed. However, we will have evidence that this cost is positive and probably not linear in the number of terms. This possibility is discussed in the final subsection.

The payoff of the union is given by:

$$U^W = \left(B\left(\vec{n}\right) + \left(1-\tau\right)\left(w-\bar{w}\right)\right) L\left(w\right) - c\left\|\vec{n}\right\|, \tag{3}$$

where $L(w)$ is employment and $\left\|\vec{n}\right\| = n^W + n^F$. The per-worker surplus is the value of contract terms, $B\left(\right)$ and the net-of-tax wage premium over the market wage, $\bar{w}$. Here we let $L\left(w\right)$ vary with the wage so that the model can nest both the efficient-contract model and the right-to-manage model. For the efficient-contract model, one simply adds the requirement that $L'\left(w\right) = 0$.

The payoff for the firm is given by:

$$U^F = \left(\frac{\gamma R\left(L\right)}{L} - C\left(\vec{n}\right) - w\right) L - c\left\|\vec{n}\right\|, \tag{4}$$

where $\gamma R\left(L\right)$ is total revenue and $C\left(\vec{n}\right)$ is the cost of contract terms to the firm. Note that $\frac{\gamma R(L)}{L}$ gives the average revenue per worker. In order to have empirically interpretable effects, let:

$$R\left(L\right) = \frac{L^{1-\eta}}{1-\eta}, \tag{5}$$

where $\eta \in (0,1)$ is the elasticity of revenue with respect to employment. This in turn implies that marginal revenue satisfies:

$$\gamma R'\left(L\right) = \gamma L^{-\eta} \tag{6}$$

All functions are assumed to be differentiable and bounded.

The discussion of the previous section implies that there is a benefit to adding clauses to a labor contract. Let the benefit of the contract gross of the contracting costs be

$$W\left(\vec{n}\right) = B\left(\vec{n}\right) - C\left(\vec{n}\right). \tag{7}$$

It is assumed this function satisfies the following condition:

Assumption-1: The writing benefit function $W\left(\vec{n}\right)$ is bounded, twice-differentiable, and quasisupermodular in $\vec{n}$,where $\vec{n} \in \Re_+^2$, with the natural ordering on $\Re^2$,

with $W\left(\vec{0}\right)=0$.

The fact that W is bounded and differentiable, and $W\left(\vec{0}\right)=0$, ensures that there is a solution $\vec{n}^{*}\left(c\right)$ to

$$\max_{\vec{n}\geq\vec{0}}W\left(\vec{n}\right)-c\left\|\vec{n}\right\|.$$

The fact that W is quasisupermodular and the single-crossing condition in c is satisfied ensures that a decrease in c leads to more clauses – from Theorem 4 of Milgrom and Shannon (1994).

To simplify the model further, we take firm clauses (those giving the firm authority) to be a response to adding worker clauses (those giving control rights to workers). Suppose that the optimal number of firm clauses as a function of the worker clauses

$$n^{F*}\left(n^{W}\right)=\text{argmax}_{n}B\left(n^{W},n\right)-C\left(n^{W},n\right),$$

is well-defined and increasing and differentiable in n^{W}. This ensures that $n\left(n^{W}\right)=n^{W}+n^{F*}\left(n^{W}\right)$ is strictly increasing and continuous in n^{W}. Without loss of generality we can then take n as the choice variable, from which n^{W} and n^{F} can be derived.

Going forward we deal with n only. If there are no transactions costs ($c=0$), the optimal contract length is infinite. $W\left(n\right)$ is assumed to be bounded and strictly concave.

Given the exogenous parameters $\beta=\{\gamma,\bar{w},\tau,\alpha,c\}$, the union and firm engage in Nash bargaining and choose the optimal contract $k^{*}\left(\beta\right)=\{L^{*}\left(\beta\right),w^{*}\left(\beta\right),n^{*}\left(\beta\right)\}$ to satisfy (as in Abowd and Lemieux (1993b)):

$$k^{*}\left(\beta\right)=\text{argmax}_{k}V\left(k,\beta\right), \tag{8}$$

$$=\text{argmax}_{k}\alpha\log\left(U^{W}\left(k,\beta\right)\right)+\left(1-\alpha\right)\log\left(U^{F}\left(k,\beta\right)\right). \tag{9}$$

where as mentioned, α is the relative bargaining power of the union.

A.4 Efficient Contracts Case

Consider first the efficient contract case in which wages, employment, and contract terms are set simultaneously. We begin by solving for the wage w. Set $\frac{\partial V}{\partial w}=0$ to get:

$$0=\frac{\alpha}{U^{W}}\left(1-\tau\right)-\frac{1-\alpha}{U^{F}}.$$

This immediately implies that wage is set such that:

$$\frac{U^W}{U^F} = \frac{\alpha(1-\tau)}{(1-\alpha)}. \tag{10}$$

The relative utility of workers is increasing in bargaining power and decreasing in the tax.

From this expression we can substitute from the utility functions and define the *quasi-surplus*

$$\begin{aligned} S(L,n,\beta) &= \frac{U^W}{(1-\tau)} + U^F \\ &= \gamma R(L) + \left(\frac{B(n)}{1-\tau} - C(n) - \bar{w}\right) L - c\left(1 + \frac{1}{(1-\tau)}\right) n \end{aligned}$$

which is independent of the wage. Notice that at the bargained wage:

$$\begin{aligned} U^W/(1-\tau) &= \alpha S(L,n,\beta), \\ U^F &= (1-\alpha) S(L,n,\beta). \end{aligned}$$

Therefore, the optimal level of employment and contract clauses maximize $S()$. We have the following proposition:

Proposition 1. *At an efficient employment contract* $k^*(\beta) = \{L^*(\beta), n^*(\beta), w^*(\beta)\}$ *we have:*

$$\begin{aligned} \gamma R'(L^*) &= \bar{w} - \left(\frac{B(n^*)}{(1-\tau)} - C(n^*)\right), \\ \frac{B'(n^*)}{(1-\tau)} - C'(n^*) &= \frac{c}{L^*}\left(1 + \frac{1}{(1-\tau)}\right) \end{aligned}$$

$$w^* = \bar{w} + \frac{\alpha S^*}{L^*} - \left(\frac{1}{(1-\tau)}\left(B(n^*) - \frac{cn^*}{L^*}\right)\right), \tag{11}$$

where S^* is the quasi-surplus at the optimal contract.

From this proposition we can explore the effect of the exogenous parameters on contract terms.

Corollary 2. *At an optimal contract we have the following effects:*

- *Increasing firm productivity* (γ) *increases employment* (L^*) *and increases the number of clauses* (n^*).
- *Increasing the income tax rate* (τ) *decreases employment and increases the number of clauses.*
- *Increasing contracting costs* (c) *reduces employment and reduces the number of clauses.*
- *An increase in union bargaining power* (α) *has no effect on employment or on the number of clauses.*
- *Increasing the market wage* ($\bar{w}$) *reduces employment and the number of clauses in a contract.*

Proof. At an optimal contract the Hessian for the quasi-surplus is negative definite. Using the first order conditions for the off diagonal, the Hessian is:

$$H = \begin{bmatrix} \gamma R''(L^*) & \frac{c}{L^*}\left(1+\frac{1}{(1-\tau)}\right) \\ \frac{c}{L^*}\left(1+\frac{1}{(1-\tau)}\right) & \left(\frac{B''(n)}{1-\tau} - C''(n)\right) L^* \end{bmatrix}.$$

The diagonal entries are negative, while the off diagonals are positive. Notice that $h = \det(H) > 0$. The comparative static results for each parameter x follow from:

$$\begin{aligned} \begin{bmatrix} \frac{\partial L^*}{\partial x} \\ \frac{\partial n^*}{\partial x} \end{bmatrix} &= H^{-1} \begin{bmatrix} -\frac{\partial^2 S^*}{\partial L \partial x} \\ -\frac{\partial^2 S^*}{\partial n \partial x} \end{bmatrix}, \\ &= \frac{1}{h} \begin{bmatrix} \left(-\frac{\partial^2 S^*}{\partial L \partial x} S^*_{nn} + \frac{\partial^2 S^*}{\partial n \partial x} S^*_{Ln}\right) \\ \left(\frac{\partial^2 S^*}{\partial L \partial x} S^*_{Ln} - \frac{\partial^2 S^*}{\partial n \partial x} S^*_{LL}\right) \end{bmatrix} \end{aligned}$$

With some tedious calculation, one can derive the results in the proposition. □

The effects of the exogenous parameters upon wages need to be considered separately. The reason is that the introduction of a contract creates some countervailing effects. We can see this by writing bargained wages in the following form:

$$w^*(\beta) = \bar{w} + \alpha S^*(\beta) - A^*(\beta),$$

where $A^*(\beta) = \frac{1}{(1-\tau)}\left(B(n^*(\beta)) - cn^*(\beta)/L^*(\beta)\right) \geq 0$ denotes the *on-the-job amenities* (OJA) due to the union contract. The effect of contract costs upon wages is given

by:

$$\frac{\partial w^*}{\partial c} = \alpha \frac{\partial S^*}{\partial c} - \frac{\partial A^*}{\partial c}. \tag{12}$$

Notice that $\frac{\partial S^*}{\partial c} < \frac{\partial A^*}{\partial c} < 0$, since raising the cost of contracting reduces total surplus and amenities. However, the effect on wages is ambiguous. When the union has a great deal of power (α is close to 1) then an increase in contracting costs reduces wages. Conversely, when union power is low, then raising contracting costs can *increase* wages to compensate workers for a loss of amenities.

A.4.1 Modeling Strikes

The next question is to ask whether there is a relationship between contract length and strike incidence. Grout (1984) showed that under U.K. labor law, the rule that wages have to be negotiated from scratch can lead to inefficient investment. A solution that is applied in North America is to suppose that contracts are designed not to be renegotiated. MacLeod and Malcomson (1993) extend the Grout model to allow for dynamics and relationship-specific investments, and show that in some circumstances it is efficient to have indexing clauses (see Altonji and Devereux (1999) for a direct test of this model).

In practice indexing is not perfect, and the union and the firm may need to adjust terms. The question we address here is how does contract length affect this process. We consider two strike models that take different viewpoints on the source of a strike. Ashenfelter and Johnson (1969) provide a simple and elegant model that views a strike as a way to get workers to accept a lower wage. The starting point is that at the time of negotiation, the current contract wage, w_{t-1}, is focal, and that the union may wish to negotiate to a new wage w_t. The firm can resist these demands by allowing a strike that "softens" the position of workers.

Let us consider the case of two bargaining units that are identical, except that firm C has higher contracting costs than firm B, and hence the number of terms is $n^B > n^C$. This also implies that employment at B is higher, as are the level of amenities, $A^B > A^C$. Let us suppose that the bargaining power is such that the negotiated wage at each period is the same, with $w_{t-1}^{*A} = w_{t-1}^{*B} = w^*$. Now suppose that firms have imperfect indexing, so that in period t the real value of contracted wages falls to $\delta w^* < w^*$, where δ is the value of a period $t-1$ dollar today. Thus the desired wage increase is $y^0 = (1-\delta)\, w^0$.

Ashenfelter and Johnson (1969) then ask: what is the lowest wage that the union could accept with zero bargaining power? In the context of our model this would be the outside option less the amenity value of the job: namely, $y^i = \bar{w} - A^i(n) - w^*, i \in \{B, C\}$. Notice that this might be negative, so in some cases firm B has a longer contract than firm C, then we have $A^B > A^C$. Ashenfelter and Johnson (1969) then suppose that after a strike of length S, the union will accept the increase:

$$y^i(s, n) = y^i(n) + \left(y^0 - y^i(n)\right) e^{-rs}. \tag{13}$$

The cost of a strike to the firm is $F(s)$, while the cost of agreeing to a wage increase y is Ly - employment times the wage increase/decrease. Given that strikes last normally days to months, we ignore time discounting. The firm then decides on a strike length, that in turn determines the acceptable wage increase. The chosen strike length is:

$$s^{i*} = \text{argmax}_{S \geq 0}\{R^i - \delta w^* - F(s) - L^i y^i(s), 0\}, \tag{14}$$

where R^i is firm revenue. Notice that the strike cost function does not have to be continuous. In particular, we can allow for a fixed cost of a strike by supposing $F'(s) \geq 0, F''(s) > 0$ for $s > 0$ and $\lim_{s \to 0+} F(0) = k > 0$. Our assumptions imply the following:

Proposition 3. *Suppose that firms B and C have the same desired wage, but firm B has a longer contract than firm C. If there is inflation, and both firms strike at contract renegotiation, then strike length is shorter at the firm with the longer contract.*

The proof follows from the fact that employment is higher at B, and due to the higher level of amenities, the union *softens* at a faster rate under the Ashenfelter-Johnson model than with firm C. We interpret this result as also illustrating why strike incidence with a more detailed contract is likely to be lower: with a more detailed contract, workers are getting more on-the-job rents, and hence are "softer" with regard to wage negotiations.

A.5 Contract Dynamics

In our data we can observe the evolution of clauses over time. Consider first the implications of standard agency theory for contract length. In any relationship the number of possible states is unlimited, and so clearly one cannot have a contract of

unlimited length. In his classic book on statistics, Savage (1954) recommends solving this problem by building a small, finite model of the environment. Townsend (1979) and Dye (1985) provide formal models of this procedure that suppose that there is a fixed cost c for each contingency.

Any contract requires oversight and enforcement. These costs rise with the number of clauses in a contract. Thus, let us suppose that c is a cost that is paid *each* period to enforce an agreement (it is a flow). When the contract is signed, parties agree upon $n^{W*}(\beta) + n^{F*}(\beta)$ clauses that remain fixed over the duration of the contract. As we outlined above, we have a number of predictions on how the size of the contract varies with the exogenous parameters β. It should be emphasized that the prediction of a fixed contract size is an implication of essentially all modern theories of contract design.

As we shall see, the prediction is clearly rejected by the data. What we observe is that the number of terms increase over time. A natural reason for this to occur is due to the cost of considering all the possible cases given that bargaining must be completed in finite time.[13] Let $n_t = n_t^W + n_t^F$ be the total number of clauses in the contract at time t, denoted by k_t. We can suppose that the cost of negotiating over a new term is $\hat{C}\left(n_t^W + n_t^F - \left(n_{t-1}^W + n_{t-1}^F\right)\right)$, where $\hat{C}(0) = 0$, $\hat{C}(x) \geq 0, \hat{C}''(x) < 0$ for $x \in \Re$. This function captures the cost of increasing or decreasing the number of contract terms. Thus in period t the parties choose contract length to solve:

$$\max W\left(n_t^W, n_t^F\right) - c\left(n_t^W + n_t^F\right) - \hat{c}\left(n_t^W + n_t^F - \left(n_{t-1}^W + n_{t-1}^F\right)\right). \tag{15}$$

The solution to this problem results in an easily solvable difference equation.

Proposition 4. *Suppose a contract is negotiated starting at date $t = 1$ (so that $n_0 = 0$). The optimal contract lengths at date t are characterized by a decreasing sequence c_t such that $c_1 = c + \hat{c}$, $\lim_{t\to\infty} c_t = c$, and*

$$n_t^F = n^{F*}(\{\gamma, \bar{w}, c_t\}),$$
$$n_t^W = n^{W*}(\{\gamma, \bar{w}, c_t\}).$$

The proof of this is straightforward. The difference equation can can solved recursively. In period t simply set $c_t = c + \hat{C}'\left(n_t^W + n_t^F - \left(n_{t-1}^W + n_{t-1}^F\right)\right)$, and observe that this sequence converges to the static optimum. All the comparative static results now

[13] For models of contract complexity see Anderlini and Felli (1994), MacLeod (2002), and Battigalli and Maggi (2002b).

apply to the long-run equilibrium.

B Labor Law in Canada

Canadian labor law is generally protective of workers' rights. "Unlike the United States, the labour relations jurisdiction of the Canadian federal government is much more extensive than that of the state governments" (L.M. Farrell and G.F. Marcil, Collective Bargaining in Canada, National Center for the Study of Collective Bargaining 1 (April 2008).). In Health Services and Support – Facilities Subsector Bargaining Assn. v. British Columbia, 2007 SCC 27 (Supreme Court of Canada), the Canadian Supreme Court extended the definition of freedom of association to include protection for employees to engage in collective bargaining.

In addition, "[a]rbitration is available in the major jurisdictions upon a showing that an impasse has occurred because bargaining has become dysfunctional" (Id.). Unlike collective agreement arbitration in the United States, collective agreement arbitration in Canada has both public and private elements (Id. (citing Mitchnick & Etherington, Labour Arbitration in Canada (Lancaster House, 2006), pp. 3, 76.). "Arbitrators can and, where relevant, must consider and apply external statutes" (Id.).[14]

C Similarity Metrics

An alternative approach that we use for analyzing union contracts is to measure the stability of contract terms within a firm-union bargaining pair over time. We do this by computing distance measures between consecutive negotiated contracts. We have three specifications for contract similarity, each of which is well established in previous works from natural language processing and information extraction. Each of our metrics is based on cosine similarity, which gives the cosine of the angle between the vectorized documents. First, we use the standard cosine similarity between the word frequencies (Jurafsky and Martin, 2014). Second, we have GloVe similarity, which gives the cosine similarity between the contract vectors in a word embeddings space (Pennington et al., 2014). Third, we have LDA similarity, which gives the similarity between the LDA

[14]For introductions and overviews to Canadian labor law, see `https://www.americanbar.org/content/dam/aba/administrative/labor_law/meetings/2009/ac2009/125.authcheckdam.pdf` and `http://irc.queensu.ca/sites/default/files/articles/adams-overview-of-labour-law-in-canada.pdf`.

Table 17: Cross-Sectional and Panel Effects of Larger Work Force

Private Sector Firms						
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Total Clauses			Relative Control		
Log Employees	0.0986**	0.102**	0.0986**	-0.00156+	-0.000968	0.00348*
	(0.00467)	(0.00508)	(0.00859)	(0.000872)	(0.000948)	(0.00153)
Public Sector Firms						
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Total Clauses			Relative Control		
Log Employees	0.126**	0.117**	0.124**	-0.00219*	-0.00594**	-0.0135**
	(0.00555)	(0.00588)	(0.00908)	(0.000996)	(0.00108)	(0.00163)
Province FEs	X			X		
Sector-Year FEs	X			X		
Prov.-Sect.-Year FE's		X	X		X	X
Firm FE's			X			X

topics of the contracts (Blei, 2012). These variables are highly correlated, so for the main analysis we use the average of the three metrics.

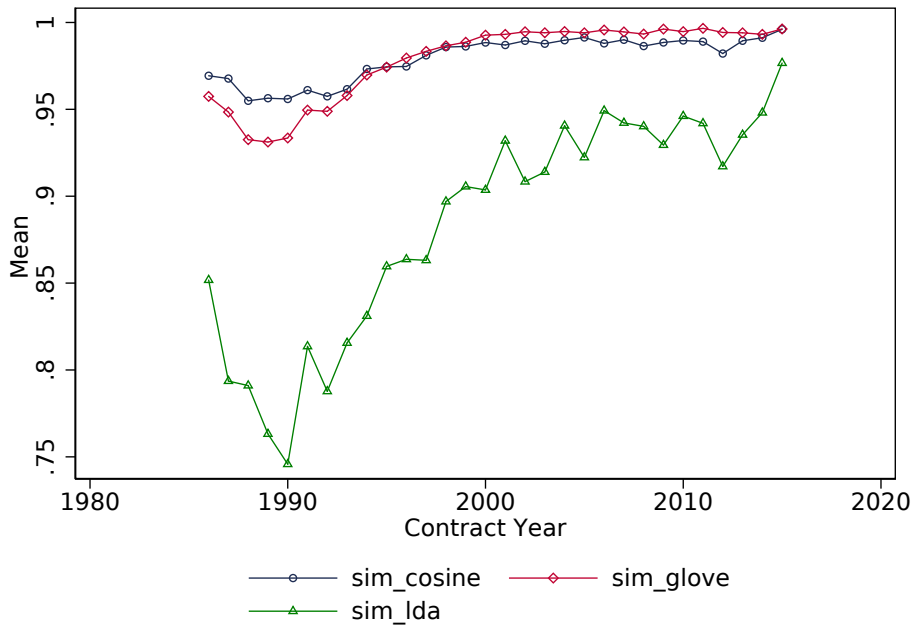
Figure 12 (top panel) shows the time trend for this metric (similarity between current contract and previous contract) across the time period in our data set. Contract terms appear to have become more stable in recent years. Figure 12 (bottom panel) shows that more contracts with higher similarity to the preceding contract also tend to have higher worker control (values residualized on company fixed effects and sector-year fixed effects). This is consistent with workers putting value on higher contract-term stability.

D Firm-Level Determinants of Contract Features

This appendix looks at how a set of firm-level variables are related to features of the contract. These are endogenous, and these regressions should be understood as descriptive statistics rather than causal estimates.

We begin by looking at variation in contracts by firm size, measured by the number

Figure 12: Sequential Contract Similarity Over Time



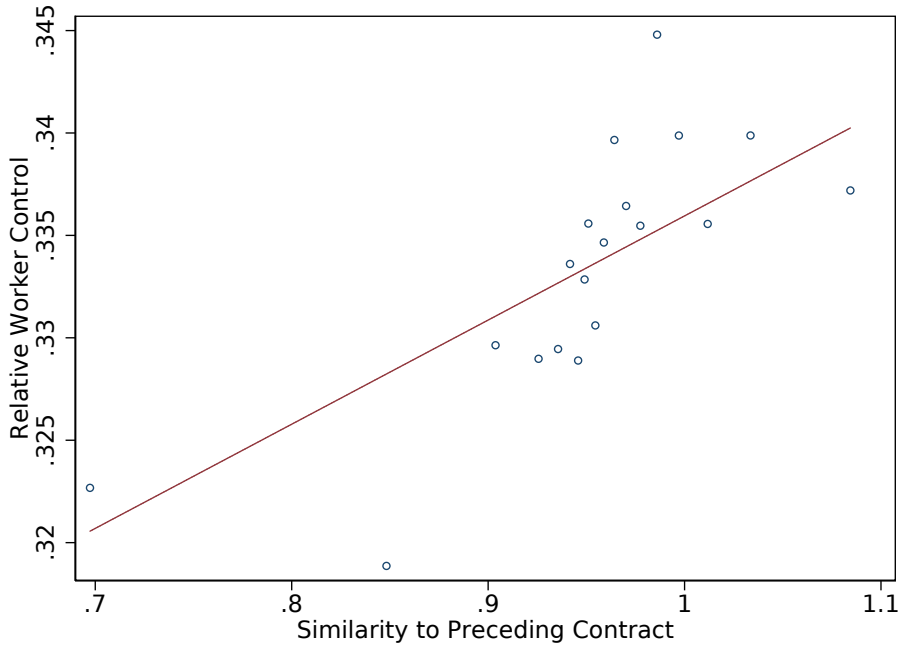


Table 18: Cross-Sectional and Panel Differences: COLA-Inflation Wage Shock

Private Sector Firms						
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Total Clauses			Relative Control		
Negative Shock	0.0250	0.0153	-0.0381+	0.000845	0.00483	0.00336
	(0.0176)	(0.0196)	(0.0229)	(0.00323)	(0.00358)	(0.00403)
Public Sector Firms						
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Total Clauses			Relative Control		
Negative Shock	-0.0132	-0.0294	-0.0364+	0.0119**	0.00832**	0.00667*
	(0.0178)	(0.0185)	(0.0190)	(0.00300)	(0.00320)	(0.00318)
Province FEs	X			X		
Sector-Year FEs	X			X		
Prov.-Sect.-Year FE's		X	X		X	X
Firm FE's			X			X

of employees. Table 17 reports these coefficients for private-sector and public-sector firms. We see that in both sectors, firms with more employees have longer contracts. We see divergent effects in terms of relative control. In the private sector, a larger workforce is associated with a lower employee entitlement share, and there a higher relative wortker control. There is an opposite effect for the public sector, where larger workforces are associated with lower worker entitlements, higher employer entitlements, and lower relative worker control.

This appendix reports some additional regression results.

We begin with exogenous variation in wages due to COLA clauses mis-predicting inflation, reported in Table 18. These firm-level wage shocks do not appear to have large effects. There may be an associated decrease in contract detail. In the public sector, there is an increase in relative worker control.

Next we look at descriptive evidence of differences between contracts that are longer or shorter in duration (Table 19). Longer-duration contracts are more detailed. In the public sector, longer-term contracts are associated with greater worker control.

Table 20 provides descriptive statistics on how contracts differ depending on whether

Table 19: Cross-Sectional and Panel Differences: Contract Duration

Private Sector Firms						
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Total Clauses			Relative Control		
Log Duration	0.202**	0.205**	0.179**	-0.00848**	-0.00849*	0.000599
	(0.0163)	(0.0185)	(0.0234)	(0.00302)	(0.00343)	(0.00418)
Public Sector Firms						
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Total Clauses			Relative Control		
Log Duration	0.175**	0.111**	0.0729**	0.0163**	0.0132**	0.00608+
	(0.0182)	(0.0199)	(0.0214)	(0.00310)	(0.00348)	(0.00360)
Province FEs	X			X		
Sector-Year FEs	X			X		
Prov.-Sect.-Year FE's		X	X		X	X
Firm FE's			X			X

or not they have a schedule for cost-of-living adjustments to wages. First, we see that contracts with COLAs tend to be longer in both the public and private sectors. In the private sector, we see an increase in employee entitlements and control for COLA contracts. In the public sector, we see the opposite; when public firms add a COLA clause, that is associated with reduced authority. This is consistent with a tradeoff in the public sector, and stronger bargaining power in the private sector where strong unions get both types of compensation.

If we zoom in on the contracts that have COLA clauses, we can see if there is variation in contract features according to the size of the COLA (average annualized change over the course of the next contract). These regressions are reported in Table 21. There are no effects.

Table 20: Cross-Sectional and Panel Differences: Has COLA Adjustment

Private Sector Firms						
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Total Clauses			Relative Control		
Has COLA Clause	0.282**	0.270**	0.159**	0.0112**	0.0115**	0.00522
	(0.0147)	(0.0160)	(0.0221)	(0.00275)	(0.00299)	(0.00396)
Public Sector Firms						
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Total Clauses			Relative Control		
Has COLA Clause	0.319**	0.290**	0.167**	-0.00314	-0.0122**	-0.0207**
	(0.0163)	(0.0171)	(0.0232)	(0.00281)	(0.00302)	(0.00392)
Province FEs	X			X		
Sector-Year FEs	X			X		
Prov.-Sect.-Year FE's		X	X		X	X
Firm FE's			X			X

Table 21: Cross-Sectional and Panel Differences: Higher COLA Adjustment

Private Sector Firms						
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Total Clauses			Relative Control		
COLA Amount	-0.0159	-0.0245+	0.00535	0.00176	0.000115	0.000320
	(0.00979)	(0.0129)	(0.0170)	(0.00155)	(0.00203)	(0.00252)
Public Sector Firms						
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Total Clauses			Relative Control		
COLA Amount	0.0623**	0.00869	0.0163	-0.00253	-0.00440+	-0.00352
	(0.0114)	(0.0151)	(0.0142)	(0.00185)	(0.00255)	(0.00244)
Province FEs	X			X		
Sector-Year FEs	X			X		
Prov.-Sect.-Year FE's		X	X		X	X
Firm FE's			X			X