

When Words Matter: The Value of Collective Bargaining Agreements

Benjamin W Arold, ETH Zürich

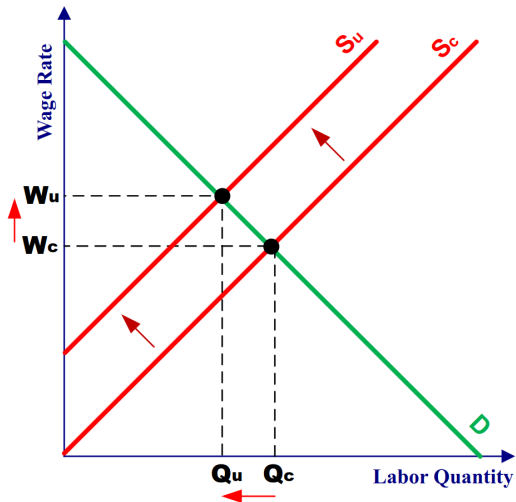
Elliott Ash, ETH Zürich

W Bentley MacLeod, Princeton University

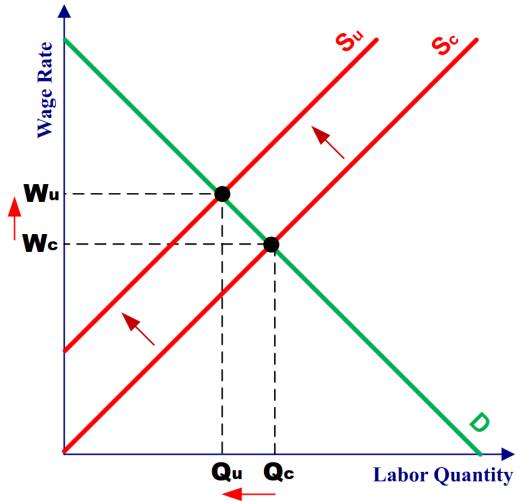
Suresh Naidu, Columbia University

Workshop: NLP and Computational Linguistics in Economics
Princeton University
March 2024

Collective Bargaining, for Economists



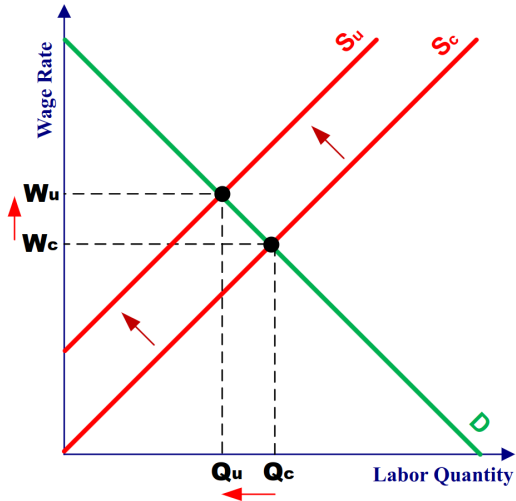
Collective Bargaining, for Economists



Collective Bargaining, for Lawyers



Collective Bargaining, for Economists



Collective Bargaining, for Lawyers



What is all this text for?

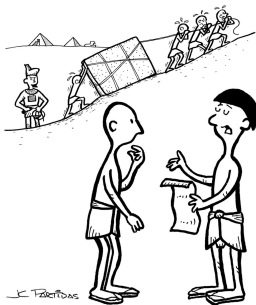
- ▶ Canonical theories of the labor contract (Simon, 1951): workers exchange control rights over their time in exchange for a wage.

- ▶ Canonical theories of the labor contract (Simon, 1951): workers exchange control rights over their time in exchange for a wage.
- ▶ Wages are important, but non-wage features matter as well:



"IN THIS PYRAMID WE ARE UNIONIZED.
WE DON'T ACCEPT MORE THAN 50 LASHES PER DAY."

- ▶ Canonical theories of the labor contract (Simon, 1951): workers exchange control rights over their time in exchange for a wage.
- ▶ Wages are important, but non-wage features matter as well:



"IN THIS PYRAMID WE ARE UNIONIZED.
WE DON'T ACCEPT MORE THAN 50 LASHES PER DAY."

- ▶ This paper: Measure and explain non-wage features of union contracts.

Project Overview

- ▶ New data:
 - ▶ Corpus of ~30k collective bargaining agreements from Canada, 1986-2015.
 - ▶ Much larger and more systematic than US collective agreement collections.
 - ▶ Data on province-year economic variables (income tax +LFS microdata).

Project Overview

- ▶ New data:
 - ▶ Corpus of ~30k collective bargaining agreements from Canada, 1986-2015.
 - ▶ Much larger and more systematic than US collective agreement collections.
 - ▶ Data on province-year economic variables (income tax +LFS microdata).
- ▶ Novel natural language methods:
 - ▶ unsupervised text algorithm to extract contract agents and associated rights and duties.
 - ▶ Validate against human/LLM coded clauses, firm surveys.

Project Overview

- ▶ New data:
 - ▶ Corpus of ~30k collective bargaining agreements from Canada, 1986-2015.
 - ▶ Much larger and more systematic than US collective agreement collections.
 - ▶ Data on province-year economic variables (income tax +LFS microdata).
- ▶ Novel natural language methods:
 - ▶ unsupervised text algorithm to extract contract agents and associated rights and duties.
 - ▶ Validate against human/LLM coded clauses, firm surveys.
- ▶ Main Results:
 - ▶ Worker rights and firm obligations are the most common contract features.
 - ▶ Increases in labor tax rates and outside options that coincide with contract re-negotiation raise worker rights.
 - ▶ Union wage premium falls with tax rate, rises with outside option – consistent with amenity interpretation.

Related Literature

- ▶ Empirical properties of contracts.
 - ▶ Standard model of contracts – wage/hour bundle (e.g. Simon 1951, MacLeod 2011).
 - ▶ Recent interest in non-wage amenities in frictional labor markets (e.g. Sorkin 2019; Dube, Naidu, and Reich 2022; Sockin 2022; Rousille & Scuderi 2023).

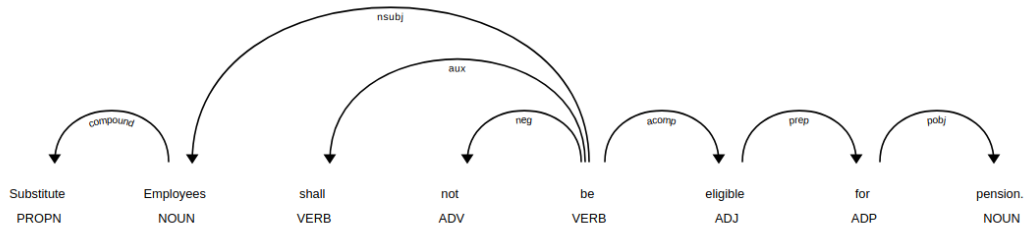
Related Literature

- ▶ Empirical properties of contracts.
 - ▶ Standard model of contracts – wage/hour bundle (e.g. Simon 1951, MacLeod 2011).
 - ▶ Recent interest in non-wage amenities in frictional labor markets (e.g. Sorkin 2019; Dube, Naidu, and Reich 2022; Sockin 2022; Rousille & Scuderi 2023).
- ▶ Unions
 - ▶ Large literature on union effects on wages/firms/inequality (Dinardo and Lee 2004; Lee and Mas 2012; Farber, Herbst, Kuziemko, Naidu 2021), but less on non-wage benefits (Lagos 2020).
 - ▶ Corradini, Lagos, and Sharma (2023) show that when unions started prioritizing women's issues → increase in female-centric amenities.

Related Literature

- ▶ Empirical properties of contracts.
 - ▶ Standard model of contracts – wage/hour bundle (e.g. Simon 1951, MacLeod 2011).
 - ▶ Recent interest in non-wage amenities in frictional labor markets (e.g. Sorkin 2019; Dube, Naidu, and Reich 2022; Sockin 2022; Rousille & Scuderi 2023).
- ▶ Unions
 - ▶ Large literature on union effects on wages/firms/inequality (Dinardo and Lee 2004; Lee and Mas 2012; Farber, Herbst, Kuziemko, Naidu 2021), but less on non-wage benefits (Lagos 2020).
 - ▶ Corradini, Lagos, and Sharma (2023) show that when unions started prioritizing women's issues → increase in female-centric amenities.
- ▶ Natural Language Processing
 - ▶ Active literature in economics applying tools from NLP to economic problems (Gentzkow, Shapiro, and Taddy 2017; Ash and Hansen 2023).
 - ▶ Recent work moves past bag-of-words representations, uses grammatical structure in text (e.g. Ash, Jacobs, MacLeod, Naidu, Stambach 2020; Ash, Gauthier, Widmer 2023).
 - ▶ Legal documents becoming an important data source for social science (e.g. Ash, Chen, Naidu 2023; Ash, Chen, Ornaghi 2023).

What We Do (1): Syntactic Parse for Contract Statements



- Dependency parsing pipeline extracts pivotal information from clauses:
 - Subject = worker, firm, union, manager
 - Verbs encode obligations (e.g. shall), rights (e.g. "have")

What We Do (2): “Worker Rights” Examples

1. Employees who retire as well as current retirees and survivors **will be provided with Life Insurance** in the amount of \$6,000.
2. Where the Company schedules an employee to work in excess of seventy-seven (77) hours in one pay period, the **employee will be paid for the excess hours at the applicable overtime rate.**
3. An employee terminated during his probationary period would be **entitled to review under the grievance procedure** up to and including Step 3.
4. Where an employee is prevented by circumstances beyond his control from returning to work on time, he **shall be paid for the holidays.**
5. However, where practicable, **senior employees in each job shall be given the opportunity to perform any available work** in that job, on their shift, within their Department.

What We Do (3): Validation & Analysis

- ▶ Validations:
 - ▶ qualitatively, worker rights capture amenities and protections at work.
 - ▶ according to “pro-worker” annotations, worker rights score highly.
 - ▶ worker rights clause frequency correlated with pro-worker HR index.

What We Do (3): Validation & Analysis

- ▶ Validations:
 - ▶ qualitatively, worker rights capture amenities and protections at work.
 - ▶ according to “pro-worker” annotations, worker rights score highly.
 - ▶ worker rights clause frequency correlated with pro-worker HR index.
- ▶ Analysis:
 - ▶ Panel regressions and event studies with worker rights clauses as outcome.
 - ▶ Increases in labor income tax rates → increase in worker rights (price effect)
 - ▶ Increase in local labor demand (Bartik shock) → increase in worker rights (outside option / bargaining power effect).

Table of Contents

Background & Data

Measuring Worker Rights in CBAs

- Choice of Methods

- NLP Approach

- Validating Worker Rights

Empirical Analysis

Discussion and Conclusion

Outline

Background & Data

Measuring Worker Rights in CBAs

- Choice of Methods

- NLP Approach

- Validating Worker Rights

Empirical Analysis

Discussion and Conclusion

Collective Bargaining in Canada

- ▶ Canadian collective bargaining shares similarities with the U.S., such as decentralized bargaining at the firm level and common-law foundation.
- ▶ Persistently higher union density than USA or UK.
- ▶ Labor regulations partly at provincial level, with diverse rules on union recognition, strike actions, and dispute resolution.
- ▶ Collective bargaining agreements in Canada are legally binding, but cannot override basic employment rights.
- ▶ Reserve Rights: CBAs give employers residual control rights but require disputes to be resolved based on the contract's text. This encourages more detailed contracts to explicitly define workers' rights.

Union Contract Information

- ▶ Canadian union contracts
 - ▶ 1986 through 2015
 - ▶ restrict attention to English ones for now.
 - ▶ From Employment and Social Development Canada NEGOTECH database.

Union Contract Information

- ▶ Canadian union contracts
 - ▶ 1986 through 2015
 - ▶ restrict attention to English ones for now.
 - ▶ From Employment and Social Development Canada NEGOTECH database.
- ▶ 32,404 contracts:
 - ▶ 7,572 companies (~ 4 contracts per company)
 - ▶ 13 provinces, 906 cities
 - ▶ 11 industry groupings, and 606 industry codes

Union Contract Information

- ▶ Canadian union contracts
 - ▶ 1986 through 2015
 - ▶ restrict attention to English ones for now.
 - ▶ From Employment and Social Development Canada NEGOTECH database.
- ▶ 32,404 contracts:
 - ▶ 7,572 companies (~ 4 contracts per company)
 - ▶ 13 provinces, 906 cities
 - ▶ 11 industry groupings, and 606 industry codes
- ▶ Contract Metadata:
 - ▶ Company, union, location, industry, public/private status, number of employees, COLA, wage.
 - ▶ Timing (signing, effective, and expiry): Compute contract duration, and match economic variables

Union Contract Information

- ▶ Canadian union contracts
 - ▶ 1986 through 2015
 - ▶ restrict attention to English ones for now.
 - ▶ From Employment and Social Development Canada NEGOTECH database.
- ▶ 32,404 contracts:
 - ▶ 7,572 companies (~4 contracts per company)
 - ▶ 13 provinces, 906 cities
 - ▶ 11 industry groupings, and 606 industry codes
- ▶ Contract Metadata:
 - ▶ Company, union, location, industry, public/private status, number of employees, COLA, wage.
 - ▶ Timing (signing, effective, and expiry): Compute contract duration, and match economic variables
- ▶ Economic Data:
 - ▶ Income tax rate, by province and year (Center for the Study of Living Standards)
 - ▶ Employment rates by province, sector, and year (Canadian Labor Force Survey)

▶ Summary Statistics

Union Contract Example

TABLE OF CONTENTS

ARTICLE PAGE

1	AGREEMENT.....	1
2	1 RECOGNITION	1
2	UNION SECURITY	1
3	MANAGEMENT RIGHTS.....	2
4	NO STRIKES-NO LOCKOUTS.....	3
5	REPRESENTATION.....	3
6	GRIEVANCE PROCEDURE.....	5
7	CONFERENCES	8
8	DISCIPLINE	8
9	SENIORITY	9
10	LOSS OF SENIORITY.....	10
11	LAYOFF AND RECALL.....	11
12	TEMPORARY TRANSFERS.....	12
13	JOBPOSTINGS.....	13
14	GENERAL.....	16
15	TEAMCOORDINATOR.....	16
16	LEAVES OF ABSENCE.....	16
17	WORK BY EXCLUDED PERSONNEL.....	20
18	PRODUCTIVITY.....	20
19	BULLETIN BOARDS.....	21
20	HOURS OF WORK AND OVERTIME.....	21
21	REST PERIODS.....	23
22	WAGES.....	23
23	INJURY ON THE JOB.....	24
24	REPORTING FOR WORK	24
25	CALL-IN-PAY.....	24
26	AFTERNOON & MIDNIGHT SHIFT PREMIUM.....	25
27	HOLIDAY PAY.....	25
28	VACATION TIME AND VACATION PAY.....	27
29	COST OF LIVING.....	29
30	PAID EDUCATION LEAVE.....	31
31	TECHNOLOGICAL CHANGE.....	31
32	BENEFIT PROGRAM.....	32
33	HEALTH & SAFETY.....	34
34	OUTSIDE CONTRACTING.....	35
35	LETTERS OF UNDERSTANDING.....	35
36	DURATION OF AGREEMENT.....	35
	SCHEDULE "A" CLASSIFICATIONS AND RATES OF PAY.....	36
	LETTERS OF UNDERSTANDING, INTENT & AGREEMENT.....	38-51

2005 – 2006 calendar

AGREEMENT

This Agreement, dated December 16, 2005 is made and entered into between ST. CLAIR TECHNOLOGIES INC., Wallacburg, Ontario (hereinafter called "the Company"), and the International Union, united Automobile, Aerospace and Agricultural Implement Workers of America (UAW-CLC) and its Local No. 251, (hereinafter called "the Union").

ARTICLE 1 RECOGNITION

- The provisions of this Agreement shall apply to all employees covered by this Agreement without discrimination on account of race, creed, colour, sex, marital status, nationality, ancestry or place of origin.
- Wherever the male noun or pronoun is used, it shall also mean the female.
- The Company recognizes the Union as the sole bargaining agent of all its employees at Wallacburg, Ontario, save and except supervisor, those above the rank of supervisor, office and sales staff, students for not more than twenty-four hours per week and students employed during the school vacation period (May 1st-September 15th). In case of reduction in force, students would be laid off first. Students will be paid at a rate to be determined by the Company, but will not be less than the Employment Standards Act.
- The word "employee" or "employees" whenever used in this Agreement shall mean only the employees in the bargaining unit defined above unless the context otherwise provides.
- The Company will negotiate with the Union for the purpose of adjusting any disputes which may arise concerning sickness and accident, wages, hours and working conditions.

ARTICLE 2

- Hire, promote, demote, classify, transfer, suspend and retire employees, and to discipline or discharge for just cause, any employee provided that a claim by an employee who has acquired seniority that he has been discharged or disciplined without just cause may be the subject of a grievance and dealt with as here in before provided.
- Make, enforce, and alter, from time to time, rules and regulations to be observed by the employees, such rules not to be inconsistent with the provisions of this Agreement. The Company agrees to give a copy of any changes in plant rules to the Union Chairperson prior to posting of same on bulletin boards.
- Determine the nature and kind of business conducted by the Company, the kinds and locations of plants, equipment and material to be used, the control of materials and parts, the use of incentive programs, the methods and techniques of work, the content of jobs, the schedules of production, the number of employees to be employed, the extension, limitation, curtailment or cessation of operations or any part thereof, and to determine and exercise all other function and prerogatives which shall remain solely with the Company except specifically limited by the express provisions of this Agreement.

ARTICLE 4 NO STRIKES - NO LOCKOUTS

- The Union agrees that during the term of this agreement, there shall be no strikes, sit-downs, work stoppage, slowdowns, or suspension of work, either complete or partial, for any reason, by an employee or employees. There shall be no lockout of employees by the Company, for the duration of this Agreement.

ARTICLE 5 REPRESENTATION

- The Union shall elect or appoint, and the Company shall recognize, from those employees who have completed at least one (1) year service with the Company a plant committee of four (4) people, one of whom will be the chairperson and one of whom will be vice chairperson. The committee people shall be employed on the day shift.
- The Company shall also recognize a steward who will be elected or

UNION SECURITY

- All employees covered by this Agreement who are members of the Union at the signing date of this Agreement or who after become members thereof during the term of this Agreement, must retain their membership in the Union for the duration of the Agreement by paying the regular monthly dues levied against all members, as a condition of employment. All employees covered by this Agreement who are not members of the Union shall pay regular monthly dues, the same as the dues that are levied against those who are members of the union as a condition of employment.
- All new employees, upon completion of thirty (30) days employment shall become members thereof in good standing in accordance with the constitution and bylaws of the Union for the life of this Agreement.
- The Company will during the term of the Agreement, deduct initiation fees, monthly dues and assessments on a monthly basis from the pay cheque of all seniority employees, probationary employees and full-time students who have worked or been compensated for forty (40) hours in any one (1) month, or as required by the U.A.W. constitution, full-time student being a student who works all or any time between May 1st and September 15th of the same year). Such deductions shall be credited to the Secretary-Treasurer of Local 251, not later than the tenth (10th) day of the calendar month next following the month in which such deductions are made. The Company and the Union will work out a mutually satisfactory arrangement by which the Company will furnish monthly records to the Financial Secretary of Local 251 of those from whom deductions were made, together with the amount of such deductions.

ARTICLE 3 MANAGEMENT RIGHTS

The Union recognizes and acknowledges that the management of the plant and direction of the working force are held exclusively by the Company and, without restricting the generality of the foregoing, the Union acknowledges that it is the exclusive function of the Company to:

- Maintain order and efficiency
 - appointed by the Union and work on the afternoon or midnight shift during such periods as the Company schedules these shifts and is equal to or greater than five (5) employees. Stewards will have preferred seniority on their shift for lay off and recall purposes only.
 - The Union will inform the Company in writing of the names of the stewards and members of the Grievance Committee and of any subsequent changes in the names of any steward or members of the Grievance Committee. The Company shall not be asked to recognize any steward or member of the Grievance Committee until such notification from the Union has been received.
 - The Union acknowledges that committee persons and stewards have their regular duties as employees to perform and that such persons will not leave their regular duties without first obtaining permission from their supervisor. Such permission shall not be unreasonably withheld. In the application of representation language "such permission shall not be unreasonably withheld" it is understood that a supervisor will not detain a Union representative from performing legitimate representation and by the same token the Union representative will understand the occasional need to complete a job in the interest of continuing production before leaving for legitimate union business. In any event, no such Union representative shall be detained any longer than thirty (30) minutes to perform their union representation duties.
 - The Company shall schedule a meeting, date and time within the time limits prescribed for any grievance submitted to Step 2 and/or Step 3 of the grievance procedure. The grievance committee only shall be compensated at their job rate for any regular scheduled work hours lost during such meeting with Company representatives. Overtime shall be paid when the meeting has been requested by the Company or the meeting goes beyond the Union representatives scheduled shift.
 - The plant committee referred to in Section 1, shall head the seniority list during their term of office for layoff and recall purposes only.
- A committee person will be required where more than nine (9) employees on the day shift at any one plant are required to work on Saturdays, Sundays and Statutory Holidays.

Outline

Background & Data

Measuring Worker Rights in CBAs

- Choice of Methods

- NLP Approach

- Validating Worker Rights

Empirical Analysis

Discussion and Conclusion

Outline

Background & Data

Measuring Worker Rights in CBAs

Choice of Methods

NLP Approach

Validating Worker Rights

Empirical Analysis

Discussion and Conclusion

Why Use Linguistic Features of Contracts?

- ▶ Why not measure wages and maybe some well-defined benefits like health insurance?

Why Use Linguistic Features of Contracts?

- ▶ Why not measure wages and maybe some well-defined benefits like health insurance?
 - ▶ For one: Many things that workers care about are not wages or pecuniary benefits – e.g. flexible scheduling, dignity at work (Dube et al 2022).

Why Use Linguistic Features of Contracts?

- ▶ Why not measure wages and maybe some well-defined benefits like health insurance?
 - ▶ For one: Many things that workers care about are not wages or pecuniary benefits – e.g. flexible scheduling, dignity at work (Dube et al 2022).
 - ▶ Contract terms matter to the parties: Firms and unions are spending a lot of money on contract drafting / labor lawyer services.

Why Use Linguistic Features of Contracts?

- ▶ Why not measure wages and maybe some well-defined benefits like health insurance?
 - ▶ For one: Many things that workers care about are not wages or pecuniary benefits – e.g. flexible scheduling, dignity at work (Dube et al 2022).
 - ▶ Contract terms matter to the parties: Firms and unions are spending a lot of money on contract drafting / labor lawyer services.
- ▶ Why not measure behavioral responses like hiring, strikes, and litigation?

Why Use Linguistic Features of Contracts?

- ▶ Why not measure wages and maybe some well-defined benefits like health insurance?
 - ▶ For one: Many things that workers care about are not wages or pecuniary benefits – e.g. flexible scheduling, dignity at work (Dube et al 2022).
 - ▶ Contract terms matter to the parties: Firms and unions are spending a lot of money on contract drafting / labor lawyer services.
- ▶ Why not measure behavioral responses like hiring, strikes, and litigation?
 - ▶ Litigation is rare – occurring only out of equilibrium. Well-designed contracts have no behavioral outputs, hence litigated contracts are selected sample.
 - ▶ hard to look at impacts of contract terms on strikes/litigation/etc.
 - ▶ still can look at language as outcome – how changes in incentives (e.g. tax rates, outside options) affect the language.

Options for Measuring “Pro-Worker” Clauses

1. Dictionaries: count pro-worker terms, e.g. “health benefits”, “management rights”, etc.
 - ▶ difficult to construct exhaustive list of pro-worker words/phrases
 - ▶ misses important context: directionality, negation

Options for Measuring “Pro-Worker” Clauses

1. Dictionaries: count pro-worker terms, e.g. “health benefits”, “management rights”, etc.
 - ▶ difficult to construct exhaustive list of pro-worker words/phrases
 - ▶ misses important context: directionality, negation
2. Supervised learning 1: hand-code clauses as pro-worker or pro-firm, train a classifier to extrapolate to whole dataset.
 - ▶ difficult to designate observed clauses as pro-worker or not – even labor lawyers are hesitant to do that except for a few special types of clauses.

Options for Measuring “Pro-Worker” Clauses

1. Dictionaries: count pro-worker terms, e.g. “health benefits”, “management rights”, etc.
 - ▶ difficult to construct exhaustive list of pro-worker words/phrases
 - ▶ misses important context: directionality, negation
2. Supervised learning 1: hand-code clauses as pro-worker or pro-firm, train a classifier to extrapolate to whole dataset.
 - ▶ difficult to designate observed clauses as pro-worker or not – even labor lawyers are hesitant to do that except for a few special types of clauses.
3. Supervised learning 2: predict a metadata variable, e.g. firm size – as in Lagos’s (2020) text-based “poaching index”.
 - ▶ relies on strong structural assumptions, especially given that observed text is equilibrium outcome.
 - ▶ classifier doesn’t observe legal rules, brings in other confounding variation.

Options for Measuring “Pro-Worker” Clauses

1. Dictionaries: count pro-worker terms, e.g. “health benefits”, “management rights”, etc.
 - ▶ difficult to construct exhaustive list of pro-worker words/phrases
 - ▶ misses important context: directionality, negation
2. Supervised learning 1: hand-code clauses as pro-worker or pro-firm, train a classifier to extrapolate to whole dataset.
 - ▶ difficult to designate observed clauses as pro-worker or not – even labor lawyers are hesitant to do that except for a few special types of clauses.
3. Supervised learning 2: predict a metadata variable, e.g. firm size – as in Lagos’s (2020) text-based “poaching index”.
 - ▶ relies on strong structural assumptions, especially given that observed text is equilibrium outcome.
 - ▶ classifier doesn’t observe legal rules, brings in other confounding variation.
4. Parser-based approach: combine legal knowledge and grammatical structure.
 - ▶ based on “classic” legal linguistic indicators (Hohfeld 1913, Balkin 1990): e.g. must/shall indicate obligations, may/can indicate permissions.
 - ▶ use syntactic parsers to attach rights/duties to regulated agents (e.g. worker, manager).

Why use grammar? Why not transformers all the way?

- ▶ Transformer-based NLP methods like BERT/GPT learn language by predicting masked tokens.
 - ▶ not informed by real-world inputs or outputs (outside the text).

Why use grammar? Why not transformers all the way?

- ▶ Transformer-based NLP methods like BERT/GPT learn language by predicting masked tokens.
 - ▶ not informed by real-world inputs or outputs (outside the text).
 - ▶ → associated language representations are not optimized for specific task of contract design.
 - ▶ → associated predictions come out of black box, may not be informative about worker rights/duties.

Why use grammar? Why not transformers all the way?

- ▶ Transformer-based NLP methods like BERT/GPT learn language by predicting masked tokens.
 - ▶ not informed by real-world inputs or outputs (outside the text).
 - ▶ → associated language representations are not optimized for specific task of contract design.
 - ▶ → associated predictions come out of black box, may not be informative about worker rights/duties.
- ▶ Our approach: Use legal knowledge on what the syntax means.
 - ▶ in principle, transformer models could be given such information during the training process.

Outline

Background & Data

Measuring Worker Rights in CBAs

Choice of Methods

NLP Approach

Validating Worker Rights

Empirical Analysis

Discussion and Conclusion

Text Pre-Processing Steps

- ▶ Contracts arrived as PDFs, along with matched metadata.
- ▶ Convert PDFs to machine-readable text (best was ABBBY FineReader)

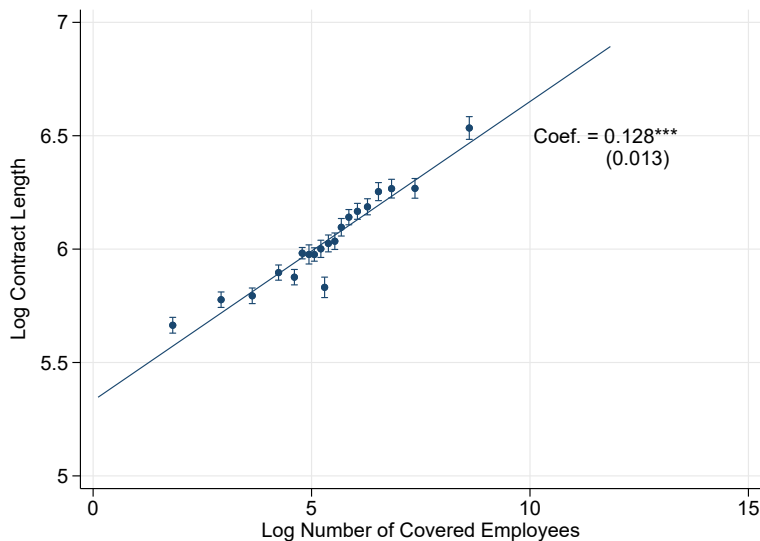
Text Pre-Processing Steps

- ▶ Contracts arrived as PDFs, along with matched metadata.
- ▶ Convert PDFs to machine-readable text (best was ABBBY FineReader)
- ▶ Exclude text for wage schedules, exhibits, appendices, etc.

Text Pre-Processing Steps

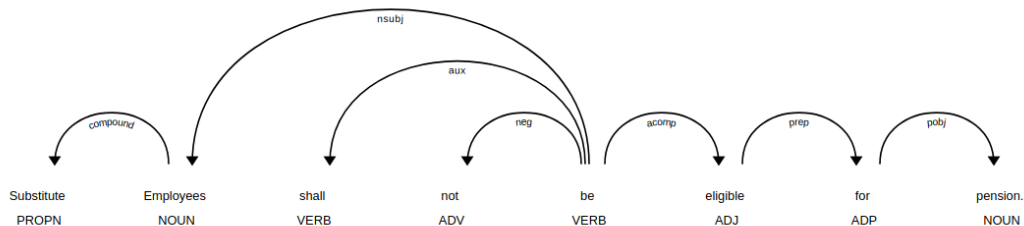
- ▶ Contracts arrived as PDFs, along with matched metadata.
- ▶ Convert PDFs to machine-readable text (best was ABBBY FineReader)
- ▶ Exclude text for wage schedules, exhibits, appendices, etc.
- ▶ Split the contracts into sections (Regex) and sentences (spaCy):
 - ▶ 980,909 contract sections (33 per contract), 10.8 million sentences (11 per section)
- ▶ Co-reference resolution by section: replace pronouns with referent entity

Number of Clauses \propto Log(Firm Size)



Note: Binscatter plot of the logarithmized number of clauses in the contract (Y axis) against the logarithmized number of covered employees (X axis). Data source: Employment and Social Development Canada.

Syntactic Parse for Contract Statements



- ▶ Dependency parsing (spaCy):
 - ▶ Output: Parse tree, giving functional relations between words in a sentence.
 - ▶ Identify syntactic subjects, and form statements around each subject
- ▶ Pipeline extracts clauses of the form: **Subject, Verb, Object**

Parse Information on Subjects and Verbs

- ▶ Subject categories:
 - ▶ worker, firm, union, manager

Parse Information on Subjects and Verbs

- ▶ Subject categories:
 - ▶ worker, firm, union, manager
- ▶ Deontic modal verbs (deontic indicating “duty”) capture necessity/possibility in social freedoms to act:
 - ▶ strict (*shall*, *will*, *must*) modals express necessity
 - ▶ permissive (*may*, *can*) modals express possibility

Parse Information on Subjects and Verbs

- ▶ Subject categories:
 - ▶ worker, firm, union, manager
- ▶ Deontic modal verbs (deontic indicating “duty”) capture necessity/possibility in social freedoms to act:
 - ▶ strict (*shall*, *will*, *must*) modals express necessity
 - ▶ permissive (*may*, *can*) modals express possibility
- ▶ Parser indicates negation (“shall **not**”) and active/passive (“shall provide” vs “shall be provided”)

Parse Information on Subjects and Verbs

- ▶ Subject categories:
 - ▶ worker, firm, union, manager
- ▶ Deontic modal verbs (deontic indicating “duty”) capture necessity/possibility in social freedoms to act:
 - ▶ strict (*shall*, *will*, *must*) modals express necessity
 - ▶ permissive (*may*, *can*) modals express possibility
- ▶ Parser indicates negation (“shall **not**”) and active/passive (“shall provide” vs “shall be provided”)
- ▶ Special verbs:
 - ▶ *Obligation Verbs* (have to, ought to, be required, be expected, be compelled, be obliged, be obligated)
 - ▶ *Prohibition Verbs* (be prohibited, be forbidden, be banned, be barred, be restricted, be proscribed)
 - ▶ *Permission Verbs* (be allowed, be permitted, be authorized)
 - ▶ *Entitlement Verbs* (have, receive, retain)

Contract Statement Typology (Simplified)

Based on human (lawyer) annotation, machine assignments have precision of 91-99%
(Ash, Jacobs, MacLeod, Naidu, Stammbach 2020).

Categorization Logic	Examples
<u>Obligations</u>	
Positive & Strict Modal & Active Verb	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
<u>Prohibitions</u>	
Negative & Any Modal & Active Verb	shall not exceed, shall not use, shall not apply
Negative & Permission Verb	shall not be allowed, is not permitted
Positive & Strict Modal & Constraint Verb	shall be prohibited, shall be restricted
<u>Permissions</u>	
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
<u>Rights</u>	
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

Summary Stats: Statement Type Shares

Subject	Clause Type				Total (%)
	Obligation (%)	Prohibition (%)	Permission (%)	Right (%)	
Worker	20.9	3.1	8.4	22.9	55.3
Firm	24.7	1.5	3.4	0.9	30.5
Union	7.0	0.6	2.0	2.1	11.7
Manager	1.7	0.1	0.4	0.2	2.5
Total	54.4	5.3	14.1	26.2	100.0

- ▶ Contracts consist mostly of worker rights (22.9%), worker obligations (20.9%) and firm obligations (24.7%)
- ▶ Firm rights are rare (0.9%); makes sense as management reserves rights.

Outline

Background & Data

Measuring Worker Rights in CBAs

Choice of Methods

NLP Approach

Validating Worker Rights

Empirical Analysis

Discussion and Conclusion

Worker Rights Clause Examples

1. Employees who retire as well as current retirees and survivors **will be provided with Life Insurance** in the amount of \$6,000.
2. Where the Company schedules an employee to work in excess of seventy-seven (77) hours in one pay period, the **employee will be paid for the excess hours at the applicable overtime rate.**
3. An employee terminated during his probationary period would be **entitled to review under the grievance procedure** up to and including Step 3.
4. Where an employee is prevented by **circumstances beyond his control** from returning to work on time, he **shall be paid for the holidays.**
5. However, where practicable, **senior employees in each job shall be given the opportunity to perform any available work** in that job, on their shift, within their Department.

Getting Worker-Rights Topics

- ▶ Topic method:
 - ▶ Vectorize each worker rights clause with transformer-based context-sensitive sentence encoder (Reimers and Gurevych 2019): Pretrained S(entence)-BERT encoder to represent clauses as 768-dim vectors.
 - ▶ Uses context of sentences tuned to capture similar meanings (rather than word counts like LDA).
 - ▶ Construct topics using k-means clustering applied to the sentence embeddings.
- ▶ Three advantages over LDA:
 - ▶ allows word meanings to be interdependent, rather than independent. For example, our method registers that “employee” and “worker” are synonyms, whereas LDA treats those words as independent.
 - ▶ method learns context-sensitive representations. For example, the word “bank” can have a different meaning for bank tellers than for dock workers. LDA does not make such a distinction.
 - ▶ assigns each individual clause to a single topic, rather than a distribution across topics. That results in a simpler dataset, and makes more sense for short documents (single sentences) rather than long documents.

All Worker-Rights Topics

Topic Label	Frequency	Topic Label	Frequency
Work Hours	0.058	Notice Requirements	0.027
Workplace Safety	0.054	Parental Leave	0.027
Payment Rules	0.051	Termination	0.026
Vacations	0.045	Overtime	0.026
Leaves of Absence	0.039	Holiday Work Pay	0.026
Seniority-Based Benefits	0.039	Shift Premiums	0.025
Seniority-Based Vacation	0.038	Sick Leave	0.025
Holiday Pay	0.037	Personnel Records	0.024
Position Classification	0.036	Workplace Injuries	0.024
Recall	0.032	Part-Time Employment	0.023
Grievance & Discipline	0.031	Reimbursements	0.022
Job Security	0.03	Probation Period	0.015
Seniority & Promotion/Transfer	0.03	Meals	0.015
Scheduling	0.03	Breaks	0.013
Bereavement Leave	0.028	Jury Duty	0.009

Interpreting the Topics

- ▶ Summarization:
 - ▶ Sample 10 clauses from a topic; generate summaries with GPT-4 ("I will give you a list of clauses sampled from collective bargaining agreements. These clauses are on a related topic in terms of giving similar rights to workers. Summarize in one sentence what types of rights the clauses in this topic are giving to workers.").

Interpreting the Topics

- ▶ Summarization:
 - ▶ Sample 10 clauses from a topic; generate summaries with GPT-4 ("I will give you a list of clauses sampled from collective bargaining agreements. These clauses are on a related topic in terms of giving similar rights to workers. Summarize in one sentence what types of rights the clauses in this topic are giving to workers.").
- ▶ Distinguishing wage-like amenities from control rights:
 - ▶ U.S. BLS National Compensation Survey provides a list of benefits and amenities that impose a quantifiable cost to employers.
 - ▶ Extract list as plain-text phrases that would appear in contracts (incentive-based pay, a commission, a production bonus, a piece rate, a cost-of-living allowance, hazard pay, a uniform allowance, a tool allowance, free room and board, subsidized room and board, paid vacation leave, paid holiday leave, paid sick leave, paid personal leave, overtime pay, shift differential pay life insurance, health insurance, disability insurance, retirement benefits).
 - ▶ Construct simulated priced-amenity clauses as "'Employees shall have ... [amenity phrase]", apply transformer sentence encoder to get embedding.
 - ▶ Compute cosine similarity of topic embeddings to priced-amenity embeddings → topic clusters that are closest to the list of priced-amenity clauses on average are the most substitutable with wages, and vice versa.

Topic Label	Sim to Wages	Frequency	Topic Summary
Grievance & Discipline	0.1590	0.031	The clauses provide workers with rights related to disciplinary actions, grievance procedures, and representation, ensuring transparency, due process, and the ability to challenge or appeal employer decisions.
Recall	0.1768	0.032	The clauses provide rights related to job security and recall for workers who have been laid off, including options to accept vacant positions, refuse temporary recalls without penalty, and priority for rehiring in their former or equivalent positions if they become available.
Seniority & Promotion	0.1792	0.03	The clauses are granting workers rights related to job preference, promotion, and transfer based on seniority, qualifications, and experience.
Leaves of Absence	0.2100	0.039	The clauses provide workers with the right to take leaves of absence for union activities, public service, education, retraining, and other approved reasons, with varying conditions regarding pay and benefits.
Position Classification	0.2115	0.036	The clauses are providing workers with rights to receive pay adjustments or increases when they take on duties in higher paying classifications, substitute in higher paying roles, transfer to new positions with higher salary scales, or are temporarily appointed to positions of a higher pay grade.
Workplace Safety	0.2118	0.054	The clauses provide workers with rights related to workplace safety, health protection, and compensation in case of job loss due to technological changes, as well as opportunities for union engagement and training on safety procedures.
Scheduling	0.2146	0.03	The clauses provide workers with rights related to scheduling flexibility, compensation for working during non-standard hours or days off, and benefits during absences or layoffs.
Seniority-Based Benefits	0.2194	0.039	The clauses provide workers with rights related to pro-rated benefits, eligibility for allowances based on employment duration, credit for service and seniority during leaves, cost-sharing for benefits, and entitlements based on continuous service, including adjustments in pay and long-term disability plans.
Vacations	0.2405	0.045	The clauses are granting workers rights related to vacation entitlements, including the timing, duration, and pay during their vacation periods.
Payment Rules	0.2414	0.051	The clauses are giving workers rights related to the timing, frequency, and accuracy of their wage payments.

Validation of Worker-Rights Clauses using LLM Annotations

- ▶ Quite difficult: Scoring a given clause as “pro-worker” or not.
 - ▶ much easier: compare two clauses and say which one is more favorable to workers.

Validation of Worker-Rights Clauses using LLM Annotations

- ▶ Quite difficult: Scoring a given clause as “pro-worker” or not.
 - ▶ much easier: compare two clauses and say which one is more favorable to workers.
- ▶ Dataset:
 - ▶ 100 randomly sampled sentences for each of 16 clause types (4 agents \times 4 provisions)
 - ▶ form across-clause-type pairs: $16 \times 15 \times 100$ clauses = 24,000 pairs

Validation of Worker-Rights Clauses using LLM Annotations

- ▶ Quite difficult: Scoring a given clause as “pro-worker” or not.
 - ▶ much easier: compare two clauses and say which one is more favorable to workers.
- ▶ Dataset:
 - ▶ 100 randomly sampled sentences for each of 16 clause types (4 agents \times 4 provisions)
 - ▶ form across-clause-type pairs: $16 \times 15 \times 100$ clauses = 24,000 pairs
- ▶ LLM Annotation (gpt-3.5-turbo-0613):
 - ▶ System Prompt: "You are a helpful legal assistant."
 - ▶ User Prompt: "Which of these sentences from a union collective bargaining agreement is more likely to be interpreted as an entitlement, benefit, or amenity for workers? Answer 'Definitely 1', 'Probably 1', 'Probably 2', 'Definitely 2', or 'Neither'. 1. [sentence 1]. 2. [sentence 2]."

Validation of Worker-Rights Clauses using LLM Annotations

- ▶ Quite difficult: Scoring a given clause as “pro-worker” or not.
 - ▶ much easier: compare two clauses and say which one is more favorable to workers.
- ▶ Dataset:
 - ▶ 100 randomly sampled sentences for each of 16 clause types (4 agents \times 4 provisions)
 - ▶ form across-clause-type pairs: $16 \times 15 \times 100$ clauses = 24,000 pairs
- ▶ LLM Annotation (gpt-3.5-turbo-0613):
 - ▶ System Prompt: "You are a helpful legal assistant."
 - ▶ User Prompt: "Which of these sentences from a union collective bargaining agreement is more likely to be interpreted as an entitlement, benefit, or amenity for workers? Answer 'Definitely 1', 'Probably 1', 'Probably 2', 'Definitely 2', or 'Neither'. 1. [sentence 1]. 2. [sentence 2]."
- ▶ For each clause type, compute % probability of being more pro-worker than other clause types.

Pair-Wise Comparisons: Which is more Pro-Worker?

1. The Employer and the Union will not tolerate, ignore or condone workplace harassment.

2. The principal should be specific in his/her comments and should base comments on personal observation.

Pair-Wise Comparisons: Which is more Pro-Worker?

1. The Employer and the Union will not tolerate, ignore or condone workplace harassment.

2. The principal should be specific in his/her comments and should base comments on personal observation.

GPT Annotation: Neither.

Pair-Wise Comparisons: Which is more Pro-Worker?

1. The Employer and the Union will not tolerate, ignore or condone workplace harassment.

2. The principal should be specific in his/her comments and should base comments on personal observation.

GPT Annotation: Neither.

1. Employees who retire as well as current retirees and survivors will be provided with Life Insurance in the amount of \$6,000.

2. If the parties mutually agree, the Company may hire temporary employees for short term periods not longer than 30 work days for non-routine work or special projects.

Pair-Wise Comparisons: Which is more Pro-Worker?

1. The Employer and the Union will not tolerate, ignore or condone workplace harassment.

2. The principal should be specific in his/her comments and should base comments on personal observation.

GPT Annotation: Neither.

1. Employees who retire as well as current retirees and survivors will be provided with Life Insurance in the amount of \$6,000.

2. If the parties mutually agree, the Company may hire temporary employees for short term periods not longer than 30 work days for non-routine work or special projects.

GPT Annotation: Clause 1.

Pair-Wise Comparisons: Which is more Pro-Worker?

1. The Employer and the Union will not tolerate, ignore or condone workplace harassment.

2. The principal should be specific in his/her comments and should base comments on personal observation.

GPT Annotation: Neither.

1. Employees who retire as well as current retirees and survivors will be provided with Life Insurance in the amount of \$6,000.

2. If the parties mutually agree, the Company may hire temporary employees for short term periods not longer than 30 work days for non-routine work or special projects.

GPT Annotation: Clause 1.

- ▶ Validation of GPT-3.5 annotations: compare to 102 human-labeled pairs
 - ▶ overall agreement: **62.7%**
 - ▶ agreement when one clause is a worker right: **83.3%**
 - ▶ GPT-4 even better.

Ranking of Clause Types by Pair-Wise Pro-Worker Frequency

Clause Type	Clause Frequency (%)	Pro-Worker Frequency (%)
Worker Right	22.9	80.9
Union Right	2.1	67.8
Worker Permission	8.4	63.08
Manager Right	0.2	59.85
Firm Obligation	24.7	55.63
Worker Prohibition	3.1	55.51
Worker Obligation	20.9	55.33
Union Permission	2	46.33
Manager Prohibition	0.1	44.36
Firm Right	0.9	39.0
Union Obligation	7	38.74
Union Prohibition	0.6	38.73
Manager Obligation	1.7	38.5
Manager Permission	0.4	37.43
Firm Prohibition	1.5	36.17
Firm Permission	3.4	35.56

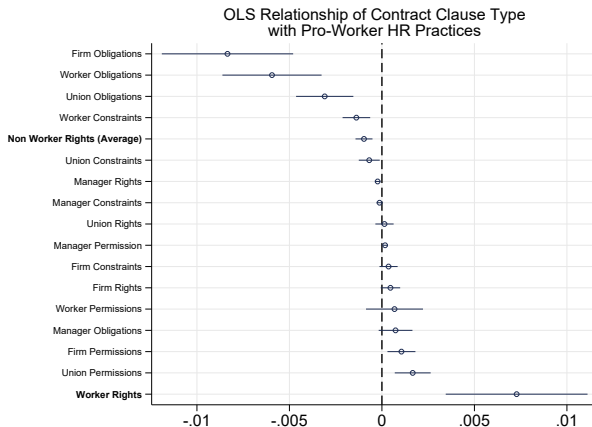
Note: Statistics from pairwise comparisons of clause types with GPT-3.5, as described in the text. Rows indicate clause types. Second column gives the frequency of that clause in the corpus; third column gives the proportion of pairwise comparisons where that category's clause is annotated as more beneficial to workers than the paired clause from another category. Sorted by third column.

Validation Against Pro-Worker HR Index

- ▶ Pro-Worker HR Index based on World Management Survey (Bloom et al, 2012)
 - ▶ Increases in “managers care about workers”, “promotes good workers”, “employees are valued”, and decreases in “focus on top talent”, “incentives”, “fire poor performers”
 - ▶ Matched to 127 contracts by firm name and time.

Validation Against Pro-Worker HR Index

- ▶ Pro-Worker HR Index based on World Management Survey (Bloom et al, 2012)
 - ▶ Increases in “managers care about workers”, “promotes good workers”, “employees are valued”, and decreases in “focus on top talent”, “incentives”, “fire poor performers”
 - ▶ Matched to 127 contracts by firm name and time.



Note: Figure presents coefficients and 95% confidence intervals of regression of contract clause types on index for Pro-Worker HR Practices. Outcome: Clause type, defined as share of clauses of given type (number of clauses of type in question over the number of all clauses) . Treatment: Standardized index of Pro-Worker HR Practices, defined as sum of approval rates to six statements about worker practices”. Controls: None. Heteroscedasticity-robust standard errors.

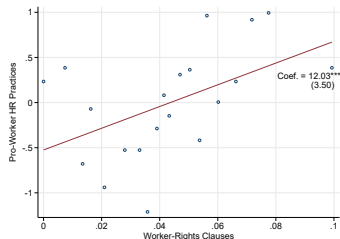
Comparison to Supervised-Learning-Based Method

- ▶ Lagos (2020) introduces a text-based measure of amenities in collective bargaining agreements based on “poaching”:
 1. Vectorize contract clauses (e.g. using LDA topic shares) $\rightarrow \vec{L}$
 2. Fit a regression model to predict higher firm employment N with contract vectors, conditional on wages and FE.
 3. Higher amenities = higher $\hat{N}(\vec{L})$

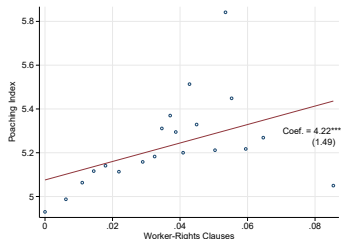
Comparison to Supervised-Learning-Based Method

- ▶ Lagos (2020) introduces a text-based measure of amenities in collective bargaining agreements based on “poaching”:
 1. Vectorize contract clauses (e.g. using LDA topic shares) $\rightarrow \vec{L}$
 2. Fit a regression model to predict higher firm employment N with contract vectors, conditional on wages and FE.
 3. Higher amenities = higher $\hat{N}(\vec{L})$
- ▶ Following Lagos (2020), we train an LDA model on our contracts to get \vec{L} and predict \hat{N} using firm level employment, conditioning on province-sector-year wages and province-sector and sector-year FE.

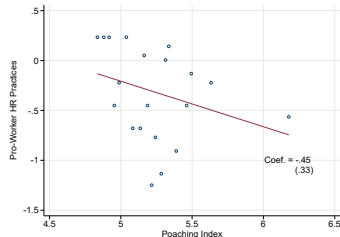
A. Pro-Worker HR Practices vs. Worker-Rights Clauses



B. Poaching Index vs. Worker-Rights Clauses



C. Poaching Index vs. Pro-Worker HR Practices



Note: Panel A: Binscatter plot of worker rights clauses (horizontal axis) and index for Pro-Worker HR Practices (vertical axis). Panel B: Binscatter plot of worker rights clauses (horizontal axis) and poaching index (vertical axis). Panel C: Binscatter plot of poaching index (horizontal axis) and index for Pro-Worker HR Practices (vertical axis). Worker rights clauses is defined as number of worker rights over the number of all clauses. Index of Pro-Worker HR Practices is defined as standardized sum of approval rates to six statements about worker practices; it increases in “managers care about workers”, “promotes good workers”, and “employees are valued,” and decreases in “focus on top talent”, “incentives”, and “fire poor performers”. Poaching index (text-predicted firm size) from Lagos (2020). Worker right measure significantly positively correlated with Pro-Worker HR Practices (Panel A). Poaching index significantly positively correlated with worker rights measure (Panel B), but not correlated with the pro-worker HR practices (Panel C). Data sources: Employment and Social Development Canada, World Management Survey (Bloom et al 2012).

Canadian Auto Workers president Buzz Hargrove on the 2005 concession agreement:

- ▶ "totally unprecedented....there was 'no business as usual' in this round of bargaining"
- ▶ "The companies started bargaining by demanding big concessions: like replacing wage increases with lump sums, abandoning COLA (even for pensioners), 10% co-pays on prescriptions, and giving up a week of paid time off per year."

Canadian Auto Workers president Buzz Hargrove on the 2005 concession agreement:

- ▶ "totally unprecedented....there was 'no business as usual' in this round of bargaining"
- ▶ "The companies started bargaining by demanding big concessions: like replacing wage increases with lump sums, abandoning COLA (even for pensioners), 10% co-pays on prescriptions, and giving up a week of paid time off per year."

Chrysler-CAW Collective Bargaining Agreements



Outline

Background & Data

Measuring Worker Rights in CBAs

Choice of Methods

NLP Approach

Validating Worker Rights

Empirical Analysis

Discussion and Conclusion

Model Predictions

- ▶ Stylized model based on Dube, Naidu, & Reich (2022):
 - ▶ contract specifies pre-tax wage and rights (amenities), which are costly to draft.
 - ▶ workers maximize utility, firm maximize profit and offer contract with wage and amenities

Model Predictions

- ▶ Stylized model based on Dube, Naidu, & Reich (2022):
 - ▶ contract specifies pre-tax wage and rights (amenities), which are costly to draft.
 - ▶ workers maximize utility, firm maximize profit and offer contract with wage and amenities
- ▶ Predictions:
 - ▶ rights increase with firm size
 - ▶ rights increase with taxes (pre-tax wages decrease)
 - ▶ rights increase with outside option (so do wages)

▶ Model

Identification

Baseline Fixed Effects Model:

$$y_{sit} = \rho z_{sit} + \alpha_{sit} + X'_{sit}\beta + \epsilon_{sit}, \quad (1)$$

Identification

Baseline Fixed Effects Model:

$$y_{sit} = \rho z_{sit} + \alpha_{sit} + X'_{sit}\beta + \epsilon_{sit}, \quad (1)$$

- y_{sit} : Contract feature (i.e. share of worker rights) of contract adopted in province s , firm i , and becoming effective in year t

Identification

Baseline Fixed Effects Model:

$$y_{sit} = \rho z_{sit} + \alpha_{sit} + X'_{sit}\beta + \epsilon_{sit}, \quad (1)$$

- ▶ y_{sit} : Contract feature (i.e. share of worker rights) of contract adopted in province s , firm i , and becoming effective in year t
- ▶ z_{sit} : Economic variable of interest (i.e. labour income tax rate or employment rate)

Identification

Baseline Fixed Effects Model:

$$y_{sit} = \rho z_{sit} + \alpha_{sit} + X'_{sit}\beta + \epsilon_{sit}, \quad (1)$$

- ▶ y_{sit} : Contract feature (i.e. share of worker rights) of contract adopted in province s , firm i , and becoming effective in year t
- ▶ z_{sit} : Economic variable of interest (i.e. labour income tax rate or employment rate)
- ▶ α_{sit} : Year-by-sector and province-by-sector fixed effects (and further FE in robustness)
 - ▶ Conduct event-studies using largest change in each province in 1990s and Callaway-Sant'Anna estimator.

Identification

Baseline Fixed Effects Model:

$$y_{sit} = \rho z_{sit} + \alpha_{sit} + X'_{sit}\beta + \epsilon_{sit}, \quad (1)$$

- ▶ y_{sit} : Contract feature (i.e. share of worker rights) of contract adopted in province s , firm i , and becoming effective in year t
- ▶ z_{sit} : Economic variable of interest (i.e. labour income tax rate or employment rate)
- ▶ α_{sit} : Year-by-sector and province-by-sector fixed effects (and further FE in robustness)
 - ▶ Conduct event-studies using largest change in each province in 1990s and Callaway-Sant'Anna estimator.
- ▶ X_{sit} : Time-varying controls in robustness checks

Identification

Baseline Fixed Effects Model:

$$y_{sit} = \rho z_{sit} + \alpha_{sit} + X'_{sit}\beta + \epsilon_{sit}, \quad (1)$$

- ▶ y_{sit} : Contract feature (i.e. share of worker rights) of contract adopted in province s , firm i , and becoming effective in year t
- ▶ z_{sit} : Economic variable of interest (i.e. labour income tax rate or employment rate)
- ▶ α_{sit} : Year-by-sector and province-by-sector fixed effects (and further FE in robustness)
 - ▶ Conduct event-studies using largest change in each province in 1990s and Callaway-Sant'Anna estimator.
- ▶ X_{sit} : Time-varying controls in robustness checks
- ▶ ϵ_{sit} : Error term. Standard errors clustered at province-sector level (robustness: province level)

Identification

Baseline Fixed Effects Model:

$$y_{sit} = \rho z_{sit} + \alpha_{sit} + X'_{sit}\beta + \epsilon_{sit}, \quad (1)$$

- ▶ y_{sit} : Contract feature (i.e. share of worker rights) of contract adopted in province s , firm i , and becoming effective in year t
- ▶ z_{sit} : Economic variable of interest (i.e. labour income tax rate or employment rate)
- ▶ α_{sit} : Year-by-sector and province-by-sector fixed effects (and further FE in robustness)
 - ▶ Conduct event-studies using largest change in each province in 1990s and Callaway-Sant'Anna estimator.
- ▶ X_{sit} : Time-varying controls in robustness checks
- ▶ ϵ_{sit} : Error term. Standard errors clustered at province-sector level (robustness: province level)

→ Identification assumption: Economic variables affect contract features, without confounding variables influencing both.

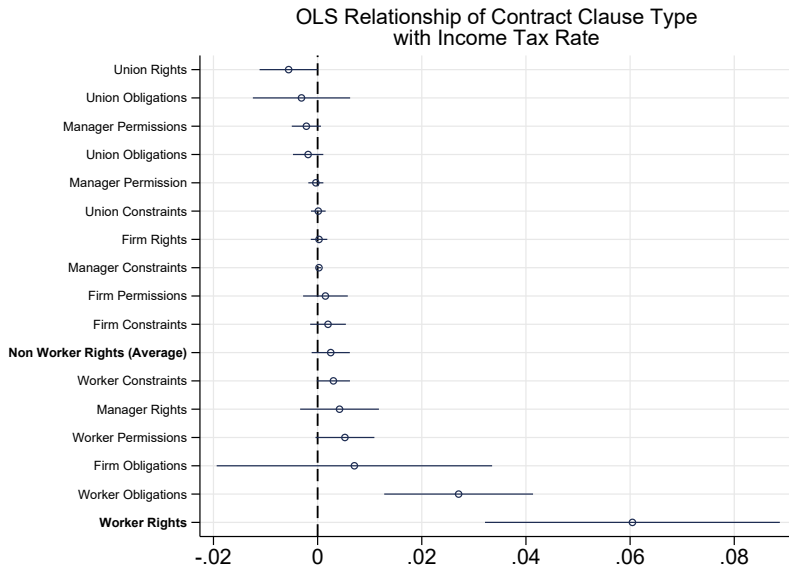
(consistent with that, treatment variables are unrelated to firm exits, the number of employees, and whether the employees have a COLA clause).

Main Result: Income Tax Rate

- ▶ Increase of income tax makes non-wage forms of income cheaper relative to wages:
We expect an increase in worker rights clauses and a decrease in wages

Main Result: Income Tax Rate

- Increase of income tax makes non-wage forms of income cheaper relative to wages:
We expect an increase in worker rights clauses and a decrease in wages

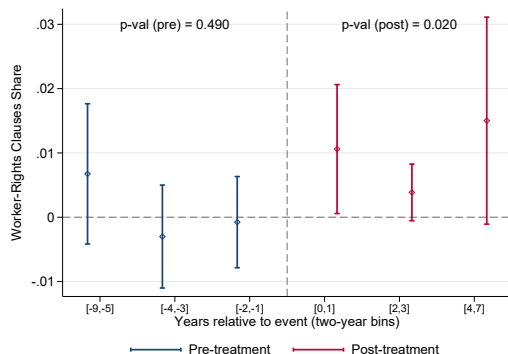


Robustness: Income Tax Rate

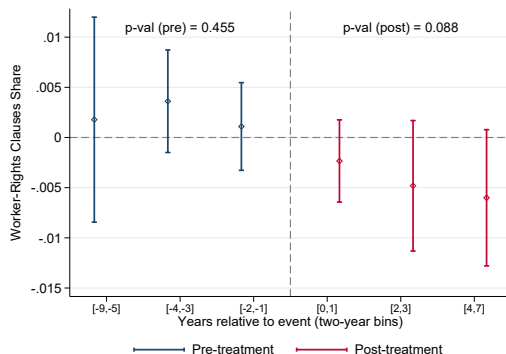
	<u>Worker-Rights Clauses</u>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log Income Tax Rate	0.060*** (0.014)	0.037*** (0.011)	0.060*** (0.015)	0.058*** (0.015)	0.060*** (0.015)	0.059*** (0.015)	0.060*** (0.015)	0.046*** (0.014)	0.035*** (0.011)	0.041*** (0.012)
R-Squared	0.15	0.16	0.15	0.15	0.15	0.15	0.15	0.30	0.47	0.16
Number of Observations	24,826	24,826	24,826	24,826	24,826	24,826	24,549	24,826	24,826	23,043
Province-Sector FEs	X	X	X	X	X	X	X	X	X	X
Sector-Year FEs	X	X	X	X	X	X	X	X	X	X
Province Trends		X								
Cluster by Province			X							
Pro-Union Law Controls				X						
Anti-Union Law Controls					X					
NDP Party Control						X				
Employment Control							X			
Worker and Firm Obligation Control								X		
Share Parsed Clauses Control									X	
Drop Zero-Worker-Rights Clauses										X

Note: Coefficients and standard errors of effect of labor tax rate on worker rights clauses, for different specifications as indicated in table footer. Outcome: Share of worker rights clauses, defined as number of worker rights clauses over the number of all clauses. Treatment: Labor tax rate is defined as logarithmized implicit personal income tax rate. Controls: Pro-Union (Anti-Union) Law Controls includes set of separate indicator variables for whether a given law favorable (unfavorable) to unions is in place. Inference: Standard errors clustered at the province-by-sector level, unless noted otherwise. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

A. Discrete Increase in Tax Rates



B. Discrete Decrease in Tax Rates



Note: Figure presents coefficients and 95% confidence intervals for time indicators before and after labor tax rate increase (Panel A) or decrease (Panel B) on share of worker-rights clauses. Callaway & Santanna (2021) estimator, accounting for heterogeneous treatment effects and staggered treatment timing. Dynamic aggregation/event study effects, using doubly robust inverse probability weighting. Outcome: Worker rights share, defined number of worker rights over the number of all clauses. Controls: Not-yet-treated observations. Numbers on horizontal axis refer to final year of respective two-year bins; i.e., -1 = last two years prior to event. Event is defined the largest labor tax increase or decrease, respectively in a given province in the 1990s, where labor tax rate is defined as logarithmized implicit personal income tax rate. Inference: Standard errors clustered at the province-by-sector level.

Effect of Income Tax Rates on Worker Rights, by Topic Group

Effect of Income Tax Rates on Worker Rights, by Topic Group



Effect of Tax Rates on Worker Rights, by Substitutability with Wages

Effect of Tax Rates on Worker Rights, by Substitutability with Wages



Note: Figure presents coefficients and 95% confidence intervals of effect of labor tax rate on worker-right topical categories (by substitutability with wages). Substitutability with wages of given topic calculated as the dot-product similarity between the centroid of the topic and BLS Compensation Survey statements. 16 topics most (least) close to these statements are grouped in categories of close (distant) substitutes of wages.

Effect of Labor Income Tax Rates and Union Status on Wages (LFS)

Effect of Labor Income Tax Rates and Union Status on Wages (LFS)



Note: Figure presents coefficients and 95% confidence intervals of effect of labor tax rate, union status, and the interaction of labor tax rate and union status on individual wages. Outcome: Individual wages, defined as worker's logarithmized hourly wage (before taxes and other deductions, but including tips, commission and bonuses). Treatments: Labor tax rate, defined as logarithmized implicit personal income tax rate; union status, defined as indicator variable that equals one if worker is member of a union, and zero otherwise. Controls: Province-by-sector fixed effects, and year-by-sector fixed effects. Standard errors clustered at the province level. Sample: 1999-2006 (excludes years from the financial crisis 2007 onward).

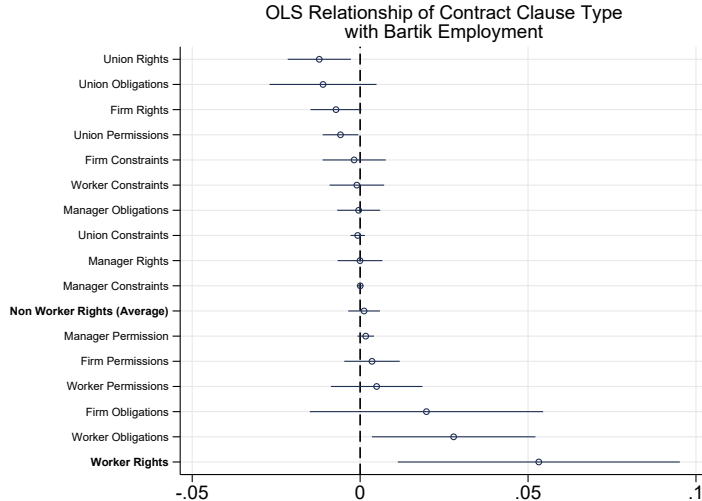
Labor Demand Shock

- ▶ Employment rate in sector X province X year is a measure of workers' outside option:
 - ▶ costliness of strike to employers – more difficult to hire replacements.
 - ▶ also a measure of labor demand.
- ▶ Predicts more pro-worker contracts during high labour demand.

Labor Demand Shock

- ▶ Employment rate in sector X province X year is a measure of workers' outside option:
 - ▶ costliness of strike to employers – more difficult to hire replacements.
 - ▶ also a measure of labor demand.
- ▶ Predicts more pro-worker contracts during high labour demand.
- ▶ Use leave-one-out sectoral employment rate (X province by year) as Bartik instrument.
 - ▶ Isolate outside option component.
- ▶ Positive labor demand shock improves bargaining position of unions relative to firms: We expect an increase in worker rights and wages

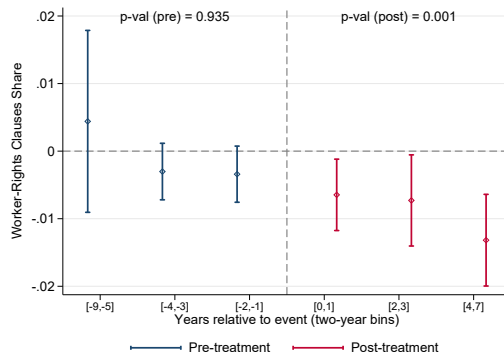
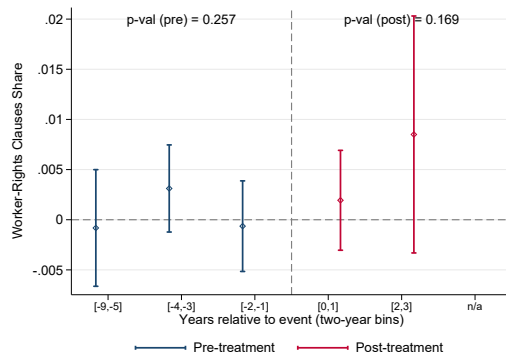
Effect of Outside Option on CBA Clauses



Note: Figure presents coefficients and 95% confidence intervals of effect of Bartik-style leave-one-out employment rate on contract clause types. Outcome: Clause type, defined as share of clauses of given type (number of clauses of type in question over the number of all clauses). Treatment: Bartik-style leave-one-out employment rate in a given sector, defined as the logarithmized average over the employment rates in other sectors. Controls: Province-by-sector fixed effects and year-by-sector fixed effects. Inference: Standard errors clustered at the province-by-sector level.

Outside Option Event Study

Outside Option Event Study



Note: Figure presents coefficients and 95% confidence intervals for time indicators before and after employment increase (Panel A) or decrease (Panel B) on share of worker-rights clauses. Callaway and Santanna (2021) estimator, accounting for heterogeneous treatment effects and staggered treatment timing. Dynamic aggregation/event study effects, using doubly robust inverse probability weighting. Outcome: Worker rights share, defined number of worker rights over the number of all clauses. Controls: Not-yet-treated observations. Numbers on horizontal axis refer to final year of respective two-year bins; i.e., -1 = last two years prior to event. Event is defined the largest employment increase or decrease, respectively in a given province in the 1990s, where employment is defined as Bartik-style leave-one-out employment rate in a given sector (logarithmized average over the employment rates in other sectors; there were no such observations for the 4-7 window for employment rate increases). Clustering at province-by-sector level.

Robustness: Effect of Outside Options

	<u>Worker-Rights Clauses</u>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log Emp. Rate	0.053** (0.021)	0.050*** (0.019)	0.053* (0.024)	0.078*** (0.019)	0.056*** (0.017)	0.050** (0.020)	0.049** (0.021)	0.037** (0.018)	0.035** (0.014)	0.052*** (0.017)
R-Squared	0.15	0.16	0.15	0.15	0.15	0.15	0.15	0.31	0.47	0.16
Number of Observations	29,157	29,157	29,157	27,603	27,603	29,157	29,157	29,157	29,157	27,108
Province-Sector FEs	X	X	X	X	X	X	X	X	X	X
Sector-Year FEs	X	X	X	X	X	X	X	X	X	X
Province Trends		X								
Cluster by Province			X							
Pro-Union Law Controls				X						
Anti-Union Law Controls					X					
NDP Party Control						X				
Employment Control							X			
Worker and Firm Obligation Control								X		
Share Parsed Clauses Control									X	
Drop Zero-Worker-Rights Clauses										X

Note: Coefficients and standard errors of effect of Bartik-style leave-one-out employment rate on worker rights clauses, for different specifications as indicated in table footer. Outcome: Share of worker rights clauses, defined as number of worker rights clauses over the number of all clauses. Treatment: Bartik-style leave-one-out employment rate in a given sector, defined as the logarithmized average over the employment rates in other sectors. Controls: Pro-Union (Anti-Union) Law Controls includes set of separate indicator variables for whether a given law favorable (unfavorable) to unions is in place. Employment control controls for logarithmized employment rate (own sector). Inference: Standard errors clustered at the province-by-sector level, unless noted otherwise. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Effect of Tax Rates and Employment Rates on Worker Rights, by Topic

Effect of Tax Rates and Employment Rates on Worker Rights, by Topic

A. Effect of Tax Rates



B. Effect of Employment Rates



Note: Figure presents coefficients and 95% confidence intervals of effect of log tax rate (panel A) and Bartik-style leave-one-out log employment rate (Panel B) on worker right topics. Outcome: Worker-rights topic, defined as share of worker rights clauses that belong to given topic (number of clauses of topic in question over the number of all clauses). Controls: Province-by-sector fixed effects and year-by-sector fixed effects. Standard errors clustered at the province-by-sector level.

Outline

Background & Data

Measuring Worker Rights in CBAs

- Choice of Methods

- NLP Approach

- Validating Worker Rights

Empirical Analysis

Discussion and Conclusion

CBAs and Tax Avoidance

- ▶ In response to a 10% income tax increase:
 - ▶ share of worker-rights clauses in contracts increase by roughly 0.6 p.p. (0.23 standard deviations).
 - ▶ union wages fall by 1.25%

CBAs and Tax Avoidance

- ▶ In response to a 10% income tax increase:
 - ▶ share of worker-rights clauses in contracts increase by roughly 0.6 p.p. (0.23 standard deviations).
 - ▶ union wages fall by 1.25%
- ▶ CBAs offer additional tax avoidance margin:
 - ▶ Unionized firms can bargain around untaxed amenities in addition to wages.
 - ▶ In response to a tax increase, there is more movement among unions on rights clauses rather than wages.
- ▶ Estimates of taxable income elasticity may be biased if amenities are endogenous (eg Chetty 2009; Chetty et al 2011).

Valuing Text as Amenities

- ▶ In response to a 10% income tax increase:
 - ▶ share of worker-rights clauses in contracts increase by roughly 0.6 p.p. (0.23 standard deviations).
 - ▶ union wages fall by 1.25%

Valuing Text as Amenities

- ▶ In response to a 10% income tax increase:
 - ▶ share of worker-rights clauses in contracts increase by roughly 0.6 p.p. (0.23 standard deviations).
 - ▶ union wages fall by 1.25%
- ▶ One std deviation increase in share of worker-rights clauses is worth about 5.4% of wages.
- ▶ Compare to:
 - ▶ Lagos (2020): CBA employment protection worth 4% of wages.
 - ▶ Dube, Naidu, Reich (2021): one s.d. of “workplace dignity” worth 6% of wages.
 - ▶ Anelli and Koenig (2023): reducing workplace fatality risk by 1 in 100,000 is worth 9% of wages.
 - ▶ Roussille and Scuderi (2023): a one S.D. increase in amenities (in job posts) worth about 12% of wages.

Conclusion

- ▶ NLP is opening up new dimensions of language for empirical social science research
 - ▶ in particular, NLP lets economists study high-stakes legal agreements at scale.
 - ▶ detailed parsing of language particularly important in rarely litigated legal documents like collective bargaining agreements.

Conclusion

- ▶ NLP is opening up new dimensions of language for empirical social science research
 - ▶ in particular, NLP lets economists study high-stakes legal agreements at scale.
 - ▶ detailed parsing of language particularly important in rarely litigated legal documents like collective bargaining agreements.
- ▶ We find that economic conditions determine legal content of union contracts.
- ▶ In particular, worker rights clauses behave as a contracted amenity:
 - ▶ Personal Income Tax + Outside Option \uparrow : Increase in worker rights clauses
 - ▶ Substitution of wage and non-wage compensation
 - ▶ Allows valuation of contract language in empirical economics.

Conclusion

- ▶ NLP is opening up new dimensions of language for empirical social science research
 - ▶ in particular, NLP lets economists study high-stakes legal agreements at scale.
 - ▶ detailed parsing of language particularly important in rarely litigated legal documents like collective bargaining agreements.
- ▶ We find that economic conditions determine legal content of union contracts.
- ▶ In particular, worker rights clauses behave as a contracted amenity:
 - ▶ Personal Income Tax + Outside Option ↑: Increase in worker rights clauses
 - ▶ Substitution of wage and non-wage compensation
 - ▶ Allows valuation of contract language in empirical economics.
- ▶ Thank you!

Benjamin Arold | aroldb@ethz.ch

Appendix Slides

Model Ingredients (based on Dube, Naidu, & Reich 2022)

- ▶ Worker CES Utility with $\rho < 1$:

$$V(\{a\}, w) = \left(\left(\int_0^1 a_i di \right)^\rho + ((1 - \tau)w)^\rho \right)^{1/\rho}$$

- ▶ wage w , tax rate τ
- ▶ mass of potential rights (to amenities) $a_i \in [0, 1]$, $\forall i \in [0, 1]$.

Model Ingredients (based on Dube, Naidu, & Reich 2022)

- ▶ Worker CES Utility with $\rho < 1$:

$$V(\{a\}, w) = \left(\left(\int_0^1 a_i di \right)^\rho + ((1 - \tau)w)^\rho \right)^{1/\rho}$$

- ▶ wage w , tax rate τ
 - ▶ mass of potential rights (to amenities) $a_i \in [0, 1]$, $\forall i \in [0, 1]$.
- ▶ Firm offers wage w and CBA with length $T \in [0, 1]$ describing rights $\{a_i^*\}$.
 - ▶ workers have right to each contracted amenity: $a_i = a_i^*$ for $i \in [0, T]$
 - ▶ management reserves rights for non-contracted amenities: $a_i = 0$ for $i \in (T, 1]$.

Model Ingredients (based on Dube, Naidu, & Reich 2022)

- ▶ Worker CES Utility with $\rho < 1$:

$$V(\{a\}, w) = \left(\left(\int_0^1 a_i di \right)^\rho + ((1 - \tau)w)^\rho \right)^{1/\rho}$$

- ▶ wage w , tax rate τ
- ▶ mass of potential rights (to amenities) $a_i \in [0, 1]$, $\forall i \in [0, 1]$.
- ▶ Firm offers wage w and CBA with length $T \in [0, 1]$ describing rights $\{a_i^*\}$.
 - ▶ workers have right to each contracted amenity: $a_i = a_i^*$ for $i \in [0, T]$
 - ▶ management reserves rights for non-contracted amenities: $a_i = 0$ for $i \in (T, 1]$.
- ▶ Workers can strike if $V(\cdot) + \epsilon > V^0$.
 - ▶ V^0 = outside option, increases with local sectoral labor demand.
 - ▶ $\epsilon \sim 1 - F(V - V^0)$, $F(\cdot)$ = cdf for probability of no strike, increasing and concave.

Model Ingredients (based on Dube, Naidu, & Reich 2022)

- ▶ Worker CES Utility with $\rho < 1$:

$$V(\{a\}, w) = \left(\left(\int_0^1 a_i di \right)^\rho + ((1 - \tau)w)^\rho \right)^{1/\rho}$$

- ▶ wage w , tax rate τ
- ▶ mass of potential rights (to amenities) $a_i \in [0, 1]$, $\forall i \in [0, 1]$.
- ▶ Firm offers wage w and CBA with length $T \in [0, 1]$ describing rights $\{a_i^*\}$.
 - ▶ workers have right to each contracted amenity: $a_i = a_i^*$ for $i \in [0, T]$
 - ▶ management reserves rights for non-contracted amenities: $a_i = 0$ for $i \in (T, 1]$.
- ▶ Workers can strike if $V(\cdot) + \epsilon > V^0$.
 - ▶ V^0 = outside option, increases with local sectoral labor demand.
 - ▶ $\epsilon \sim 1 - F(V - V^0)$, $F(\cdot)$ = cdf for probability of no strike, increasing and concave.
- ▶ Firm profit

$$\Pi(\{a\}, w, T) = \underbrace{\left(p - w - \int_0^1 c a_i di \right)}_{\text{net worker value}} \underbrace{F(V(\{a\}, w) - V^0)}_{\text{prob. no strike}} - C(T)$$

- ▶ p = worker marginal product, c = firm amenity cost
- ▶ $C(T)$ = drafting cost of CBA, $C(\cdot)$ is increasing and convex, with $C'(0) = 0$ and $C'(1) = \infty$.

Equilibrium

- Firm problem:

$$\max_{a^*, T, w} (p - w - cTa^*)F(V(Ta^*, w) - V^0) - C(T)$$

- In equilibrium, worker MRS is equal to firm MRT (net of tax):

$$\left(\frac{w}{Ta^*}\right)^{\rho-1} = \frac{1}{c(1-\tau)^\rho}$$

Equilibrium

- ▶ Firm problem:

$$\max_{a^*, T, w} (p - w - cTa^*)F(V(Ta^*, w) - V^0) - C(T)$$

- ▶ In equilibrium, worker MRS is equal to firm MRT (net of tax):

$$\left(\frac{w}{Ta^*}\right)^{\rho-1} = \frac{1}{c(1-\tau)^\rho}$$

- ▶ Taking logs and re-arranging, we have

$$\log\left(\frac{w}{Ta^*}\right) = c_0 - \sigma \log(1 - \tau)$$

- ▶ where $\sigma = \frac{\rho}{\rho-1}$ is the MRS and $c_0 = \frac{\log c}{\rho-1}$ is a constant.
- ▶ i.e., an increase in the tax rate (decrease in net-of-tax rate) increases the ratio of amenities to wages.

Equilibrium

- ▶ Firm problem:

$$\max_{a^*, T, w} (p - w - cTa^*)F(V(Ta^*, w) - V^0) - C(T)$$

- ▶ In equilibrium, worker MRS is equal to firm MRT (net of tax):

$$\left(\frac{w}{Ta^*}\right)^{\rho-1} = \frac{1}{c(1-\tau)^\rho}$$

- ▶ Taking logs and re-arranging, we have

$$\log\left(\frac{w}{Ta^*}\right) = c_0 - \sigma \log(1 - \tau)$$

- ▶ where $\sigma = \frac{\rho}{\rho-1}$ is the MRS and $c_0 = \frac{\log c}{\rho-1}$ is a constant.
- ▶ i.e., an increase in the tax rate (decrease in net-of-tax rate) increases the ratio of amenities to wages.
- ▶ Other results:
 - ▶ amenities and wages increase with the outside option V^0 or with firm productivity p .

NLP in Legal Contexts: Text corpora

- ▶ Legislation
 - ▶ The statutes enacted by legislators, which are then added to a compiled code
 - ▶ Hierarchical structure, extensively cross-referenced
- ▶ Regulations
 - ▶ The more specific rules to implement legislation, decided by more technocratic agencies.
 - ▶ E.g., tax agency should decide whether a gift counts as income
- ▶ Judicial opinions
 - ▶ When a dispute arises over the meaning of a statute or regulation, a judge decides
 - ▶ Judge will write an opinion, citing statutes and previous caselaw, explaining the interpretation

NLP in Legal Contexts: Potentials (Robot clerk)

- ▶ In general: Legal documents tend to have more structure, legal language tends to be more precise than other corpora
- ▶ Annotation tasks
 - ▶ Categorize documents into topics (Osnabruegge, Ash, and Morelli 2021), tag slant/sentiment in opinions (Ash, Chen, and Galletta 2021)
- ▶ Document Comparison and Retrieval
 - ▶ Finding similar precedents to a given case (Ostendorff, Ash, et al 2021)
 - ▶ Compare international tax treaties to understand influential tax systems (Ash and Marian 2020).
- ▶ Legal Summarization and Drafting (powered up by neural nets and language models)
 - ▶ Generate coherent legal language (Peric, Mijic, Stammbach and Ash 2020), extractive summarization: highlight the relevant portions of long texts (Gu, Ash, and Hahnloser 2022; Bauer, Stammbach, Gu, and Ash 2023)

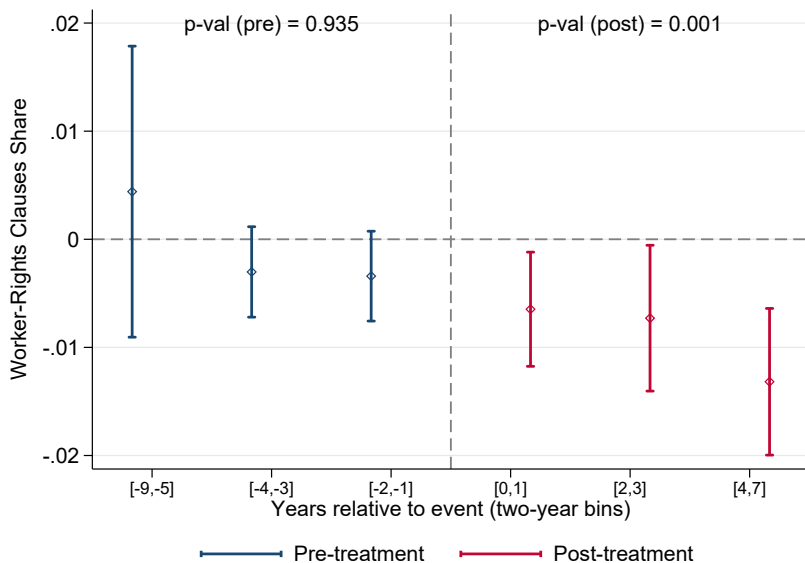
NLP in Legal Contexts: Issues and Limitations

- ▶ Text complexity:
 - ▶ Definitions are often specified elsewhere in the document
 - ▶ Extensive and pivotal references to other documents
- ▶ Text ambiguity
 - ▶ bounded cognition and time; strategic ambiguity
 - ▶ failed efforts to put law on a formal-logic basis, or to say “law is code”
- ▶ Context
 - ▶ Legal texts are embedded in a complex social system, e.g. parliamentary debates, proposed bills etc.

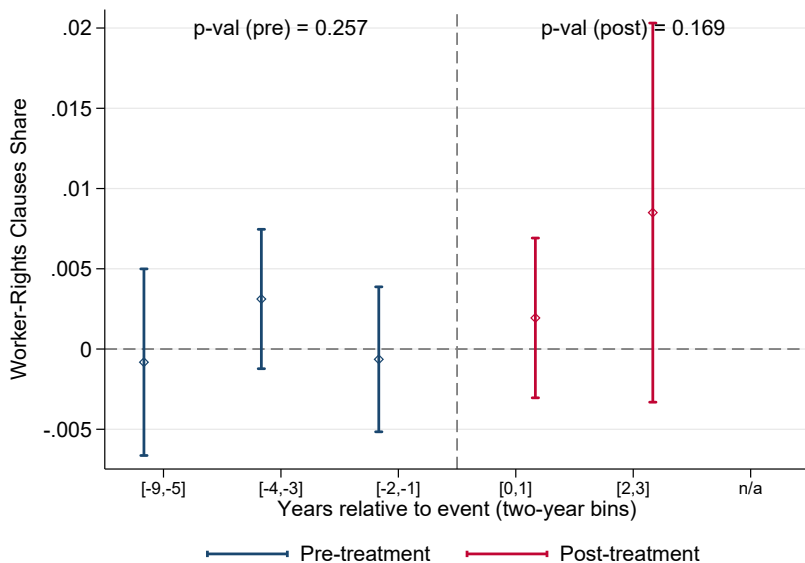
Labor Demand Shock

- ▶ Employment rate in sector X province X year measure of outside option of workers and costliness of strike to employers.
 - ▶ E.g. more difficult to hire replacement workers.
- ▶ But also a measure of labor demand.
 - ▶ Also predicts more pro-worker contracts during high labor demand.
- ▶ Use leave-one-out sectoral employment rate (X province by year) as Bartik instrument.
 - ▶ Isolate outside option component.

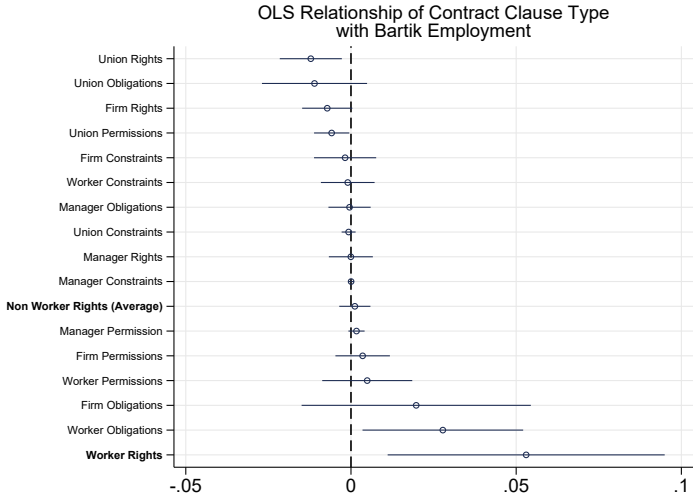
Event-Study: Outside Employment Rate Decrease



Event-Study: Outside Employment Rate Increase

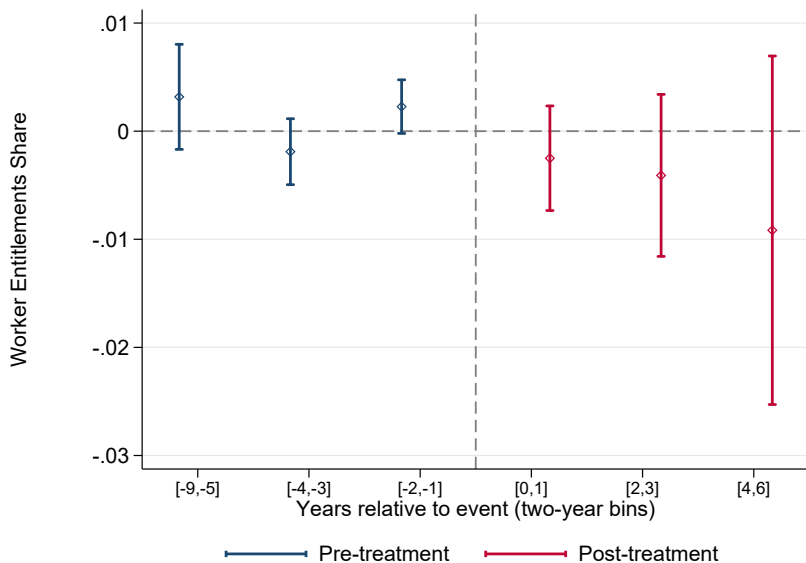


Outside Option on Worker Rights

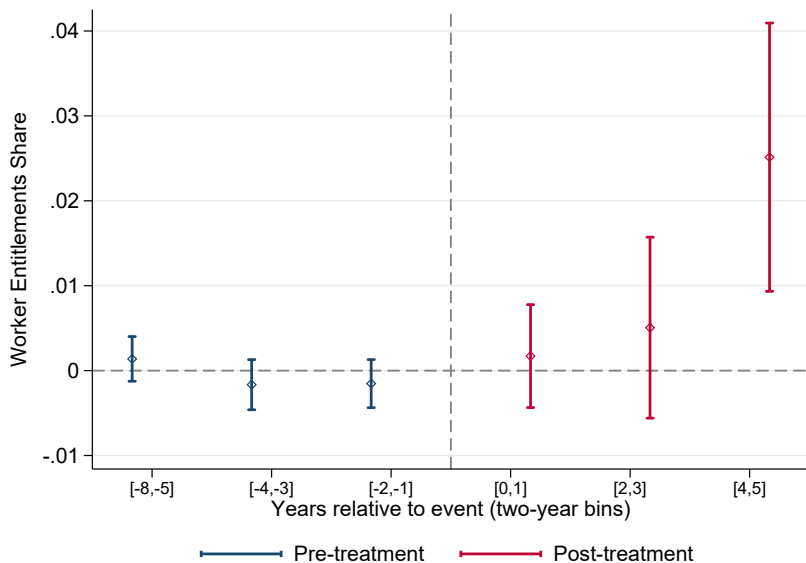


Note: Effect of Bartik employment rate on contract clause types. Bartik employment calculated as leave-one-out specification, where employment in a given province, year, and sector equals the average employment in this province and year for all other sectors besides the one of the contract. [▶ Back](#)

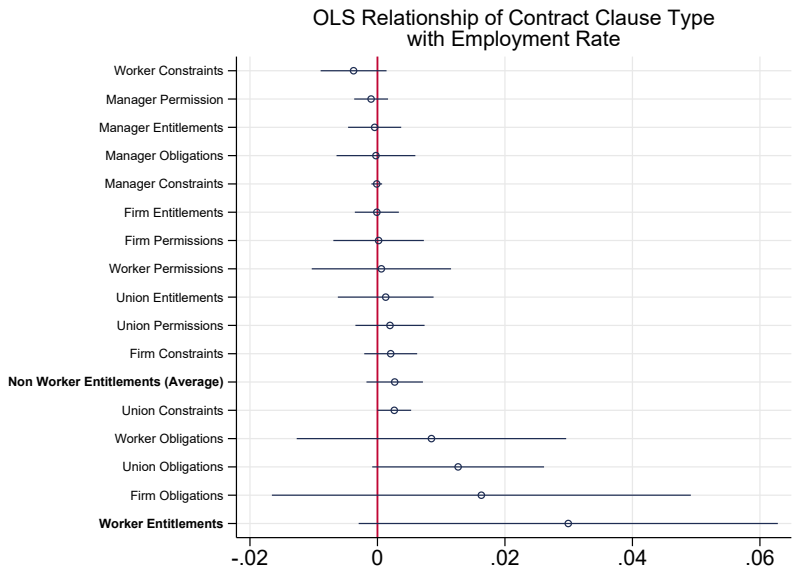
Event-Study: Own Employment Rate Decrease



Event-Study: Own Employment Rate Increase



Employment Rate Results



Robustness: Labour Demand

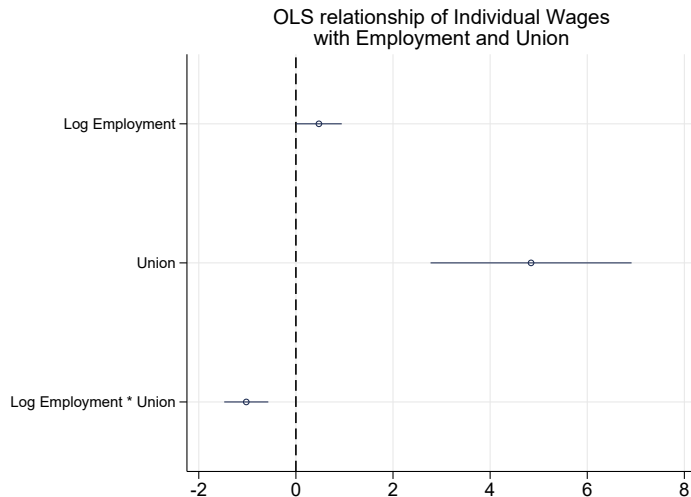
	<u>Worker Rights Clauses</u>					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Emp. Rate	0.030*	0.031**	0.019	0.025	0.030*	0.029*
	(0.017)	(0.015)	(0.014)	(0.015)	(0.013)	(0.017)
Province-Sector FEs	X	X	X	X	X	X
Sector-Year FEs	X	X	X	X	X	X
Province Trends		X				
Firm FE			X			
Union FE				X		
Cluster by Province					X	
Contract Controls						X

Note: Table presents coefficients and standard errors for effect of logarithmized employment rate on inverse hyperbolic sine of the share of worker rights clauses (number of worker rights clauses over the number of all clauses). Inference: Clustering at province-by-sector level. 14 / 21

Effect of Employment Rate on Worker Rights by Topic Group



Effect of Employment Rates and Union Status on Wages



► Event-Study: Employment Decrease

Appendix: Summary Statistics for Contracts Metadata

Variable	Obs	Mean	Std. Dev.	Min	Max
Private-Sector	29848	.4860	.499	0	1
Number of Employees	29841	655.87	2721.	0	170,000
Year	29503	1999.79	7.89	1986	2015
Duration (Years)	29503	2.584	1.1	0	20
Has COLA	29848	.2731	.445	0	1
Annual Wage Increase (%)	8152	2.61	1.843	-7.560	19.836
Inflation (%)	20429	5.77	3.347	-.8643	31.62
Union Has Strike	32,402	.0328375	.1782138	0	1
Income Tax Rate (%)	24,910	22.38973	1.447889	16.11	25.62
Unemployment Rate (%)	29,200	5.086423	3.544908	1.08	49.92
NDP Province Govt Control	32,402	.2127338	.4092472	0	1

Appendix: Text Pre-Processing Steps

- ▶ Contracts arrived as PDFs, along with matched metadata

Appendix: Text Pre-Processing Steps

- ▶ Contracts arrived as PDFs, along with matched metadata
- ▶ Convert PDFs to machine-readable text:
 - ▶ Compared three different OCR engines. Best was ABBBY FineReader (by misspelling rate), followed by Adobe's and then Tesseract

Appendix: Text Pre-Processing Steps

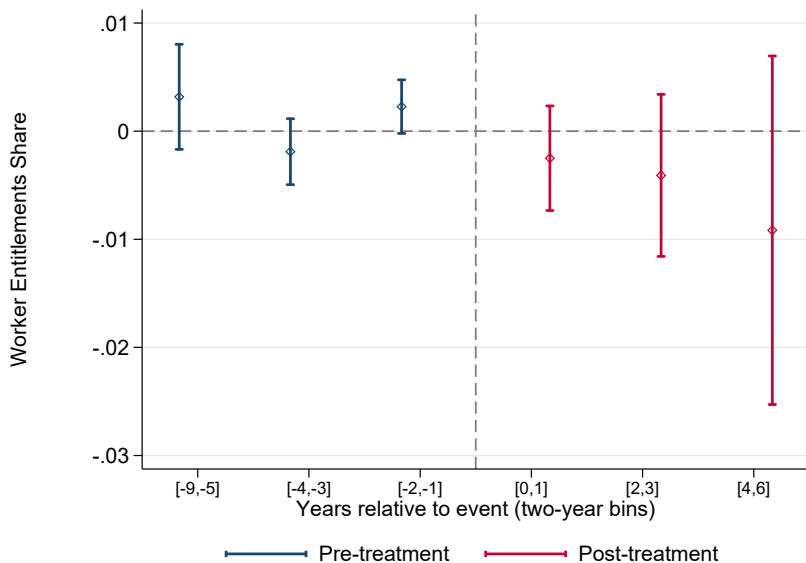
- ▶ Contracts arrived as PDFs, along with matched metadata
- ▶ Convert PDFs to machine-readable text:
 - ▶ Compared three different OCR engines. Best was ABBBY FineReader (by misspelling rate), followed by Adobe's and then Tesseract
- ▶ Exclude text for wage schedules, exhibits, appendices, etc.

Appendix: Text Pre-Processing Steps

- ▶ Contracts arrived as PDFs, along with matched metadata
- ▶ Convert PDFs to machine-readable text:
 - ▶ Compared three different OCR engines. Best was ABBBY FineReader (by misspelling rate), followed by Adobe's and then Tesseract
- ▶ Exclude text for wage schedules, exhibits, appendices, etc.
- ▶ Split the contracts into sections and sentences:
 - ▶ Use custom-built splitter (based on regular expression), and SpaCy tokenizer
 - ▶ 980,909 contract sections (33 per contract), 10.8 million sentences (11 per section)

▶ Back

Appendix: Event-Study: Employment Rate Decrease



Assigning Clauses to Topics

- ▶ Use pretrained S(entence)-BERT encoder to represent clauses as 768-dim vectors.
 - ▶ Uses context of sentences tuned to capture similar meanings (rather than word counts like LDA).
- ▶ Apply k-means clustering within clause type (e.g. within set of worker-rights sentences) to produce $k = 32$ topics.
- ▶ Aggregate clusters into 7 more interpretable topics: Scheduling, work termination, health & well-being, vacations, family issues, payments, and seniority

▶ [Back \(Tax\)](#)

▶ [Back \(Employment\)](#)

Summary Statistics on Topic Clusters of Worker Rights

Label	Frequency
Scheduling	0.182
Work Termination	0.069
Health & Wellbeing	0.051
Vacation	0.117
Family Issues	0.068
Payments	0.079
Seniority	0.089

Note: Cluster topics from “Other” category not reported.