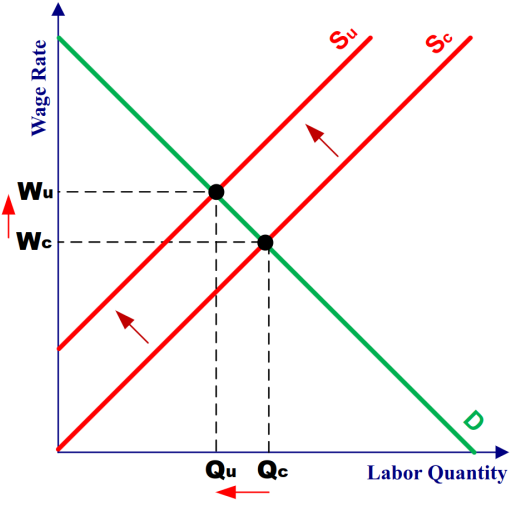
Do Words Matter?

The Value of Collective Bargaining Agreements

Benjamin W Arold, ETH Zürich & Cambridge
Elliott Ash, ETH Zürich & CEPR
W Bentley MacLeod, Princeton, Columbia, & NBER
Suresh Naidu, Columbia & NBER

NBER Summer Institute
22 July 2024

Collective Bargaining, for Economists



Collective Bargaining, for Lawyers

What is all this text for?

"The difference between a one-page teaching contract [South Carolina] and a fifty-page teaching contract [New York] is that one of them has forty-nine extra pages of things that are good for teachers."

– Hamilton Nolan, *The Hammer* (2024, p. 47).

What We Do: Measurement

- New data:
  - Corpus of ∼30k collective bargaining agreements (CBAs) from Canada, 1986-2015.
    - Much larger and more systematic than U.S. CBA collections.
  - Detailed metadata by province over time (e.g. tax rates), contract (e.g. workers covered, sector, duration), and worker (Labour Force Survey microdata).
- Novel natural language methods applied to CBAs:
  - Unsupervised text algorithm to extract contract agents and associated rights and duties, based on legal theory of contract language (e.g. Hohfeld 1913, Balkin 1990).
  - Validate grammar-based measure of **worker rights** against human/LLM-coded clauses, World Management Survey.

What We Do: Research Design

- Empirical analysis of how taxes and employment affect contracted worker rights.
- Differences-in-Differences – changes coinciding with contract renegotiation:
  - Labour income taxes by province over time
  - Employment by province and sector over time
- Instrumental Variables for exogenous shifts:
  - Province tax shifter based on federal tax structure (Akcigit et al 2021).
  - Bartik employment shifter with leave-one-out by sector.

What We Learn

- Union contracts consist of worker obligations ("workers shall do X"), firm obligations ("firms shall do X"), and worker rights ("workers shall *have* X").
- Increase in labour tax rates or outside options $\rightarrow$ increase in worker rights.
- Largest effects on clauses related to scheduling (worker time use).
- Consistent with amenity interpretation for worker rights:
  - Workers value rights and they go up when their relative price decreases (labour income taxes) or when workers get more bargaining power (outside options).
  - Union contracts allow firms to commit to providing worker rights and protections.
  - Based on estimates from regressing (union) wages on tax changes, we calculate that a one-S.D. change in worker rights is worth about 5.7% of wages.

Contribution -- Natural Language Processing

- Active literature in economics applying tools from NLP to economic problems (e.g. Gentzkow, Shapiro, and Taddy 2017; Ash and Hansen 2023).
- Legal documents becoming an important data source for social science (e.g. Ash, Chen, Naidu 2023; Ash, Chen, Ornaghi 2023; Ash, Morelli, Vannoni, 2024).
- Extending work that moves past standard text tools (e.g. Ash, Jacobs, MacLeod, Naidu, Stammbach 2020; Ash, Gauthier, Widmer 2023), we use formal grammar to identify which party benefits from a clause.

Contribution – Labor Economics

- Non-wage compensation:
  - Standard model of labor contracts: wage/hour bundle (e.g. Simon 1951, MacLeod 2011).
  - Recent interest in non-wage amenities in frictional labour markets (e.g. Mas & Pallais 2017; Sorkin 2018; Dube, Naidu, and Reich 2022; Sockin 2022; Rousille & Scuderi 2023).
- Unions:
  - Large literature on union effects on wages/firms/inequality (e.g. Dinardo and Lee 2004; Lee and Mas 2012; Farber, Herbst, Kuziemko, Naidu 2021), but less on non-wage benefits (e.g. Buchmueller, DiNardo, Valletta 2004; Knepper 2020;).
  - Corradini, Lagos, and Sharma (2023) use NLP to show that when unions started prioritizing women's issues $\rightarrow$ increase in female-centric amenities (using methods from Lagos 2024).

Collective Bargaining in Canada

(e.g. Card 1983; Abowd & Lemieux 1993; Saggio, Beauregard, Lemieux, & Messacar 2024)

- Similar to U.S. system:
  - common-law foundation, now administrative-law-based.
  - decentralized bargaining at the firm level.
- But:
  - No right-to-work laws and generally stronger strike protections.
  - Persistently higher union density and broader public support for unions.
- CBAs are legally binding, but cannot override statutory employment rights.
- Reserve Rights:
  - Employers get residual control rights, but disputes are resolved based on contract's text.
  - Encourages more detailed contracts to explicitly define workers' rights.

Data on Union Contracts

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

▸ Summary Statistics

What a Union Contract Looks Like (first 3 pages)

TABLE OF CONTENTS

2005 – 2006 calendar

AGREEMENT

This Agreement ratified December 16,2005 is made and entered into between ST. CLAIR TECHNOLOGIES INC., Wallaceburg, 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

1. 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.

2. Wherever the male noun or pronoun is used, it shall also mean the female.

3. The Company recognizes the Union as the sole bargaining agent of all its employees at Wallaceburg, 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.

4. The word "employee" or "employees" wherever used in this Agreement shall mean only the employees in the bargaining unit defined above unless the context otherwise provides.

5. 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

UNION SECURITY

1. 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.

2. 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.

3. 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 fixed exclusively in the Company and, without restricting the generality of the foregoing, the Union acknowledges that it is the exclusive function of the Company to:

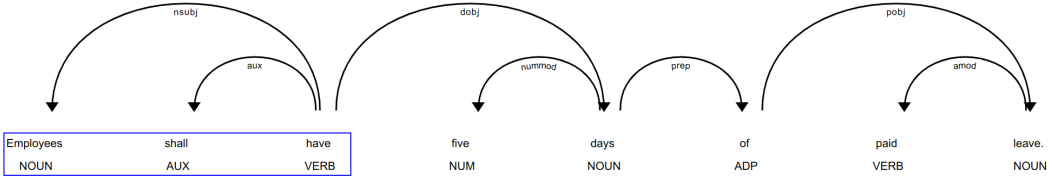
1. Maintain order and efficiency

Representing Contract Text as Data

1. Convert contract PDFs to machine-readable text.
2. Extract legal content, exclude table of contents / wage tables / appendices, etc.
3. Co-reference resolution: replace pronouns with referent entity, so that each sentence is a complete legal clause.
4. Run grammar parser and count clause types at sentence level.

▸ Text Pipeline Details

Structural Text Analysis of Contract Grammar



- Grammar parsing (spaCy):
  - Output: Parse tree, giving functional relations between words in a sentence.
  - Identify subject, verb, object, and associated modifiers.

Grammar Parse based on Legal Theory of Contracts

"Classic" legal linguistic indicators: e.g. must/shall indicate obligations, may/can indicate permissions (e.g. Hohfeld 1913, Balkin 1990).

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

Contract Statement Typology (Simplified)

Based on human (lawyer) annotation, machine assignments have precision of 91-99% (Ash et al, 2020).

Categorization Logic	Examples
Obligations	
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
Prohibitions	
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
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
Rights	
Strict Modal & Passive Verb	**shall be paid, shall be given, shall not be discharged**
Positive & Any Modal & Rights Verb	**shall have, shall receive, shall retain**
Negative & Any Modal & Obligation Verb	**shall not be required**

Summary Stats: Statement Type Shares

Subject	Clause Type				
	Obligation (%)	Prohibition (%)	Permission (%)	Right (%)	*Total (%)*
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.

"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. Where an employee is prevented by circumstances beyond his control from returning to work on time, he **shall be paid for the holidays**.
4. 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.
5. An employee terminated during his probationary period would be **entitled to review under the grievance procedure** up to and including Step 3.

▸ Worker / Firm Obligations Examples

What do Worker-Rights Clauses Consist of?

Label	Frequency
Scheduling	0.26
Vacation	0.17
Health & Wellness	0.14
Seniority	0.12
Payments	0.11
Family Issues	0.10
Termination	0.10

Note: Clause topics constructed from embedding worker-rights clauses using MPNet, applying k-means clustering with $k = 32$, and aggregating up to 7 more interpretable topics. Other/miscellaneous topic (11%) omitted.

LLM Validation of Worker-Rights Clauses

- 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.
- LLM coding (gpt-3.5-turbo-0613):
  - 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.]"
- Dataset:
  - 100 randomly sampled sentences for each of 16 clause types
    4 agents (worker, firm, union, manager) $\times$ 4 provisions (rights, obligations, prohibitions, permissions)
  - form across-clause-type pairs: $16 \times 15 \times 100$ clauses = 24,000 pairs
- **For each clause type (e.g. worker rights), compute % probability of being more pro-worker than other clause types.**

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

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: Clause 1.

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: Neither.

- 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.

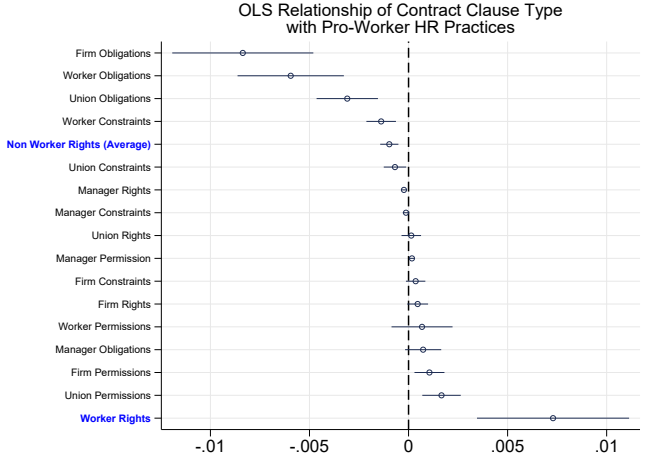
Ranking of Clause Types by Pair-Wise Pro-Worker Frequency

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"; decreases in "focus on top talent", "incentives", "fire poor performers"
  - Matched to 127 contracts by firm name and time.

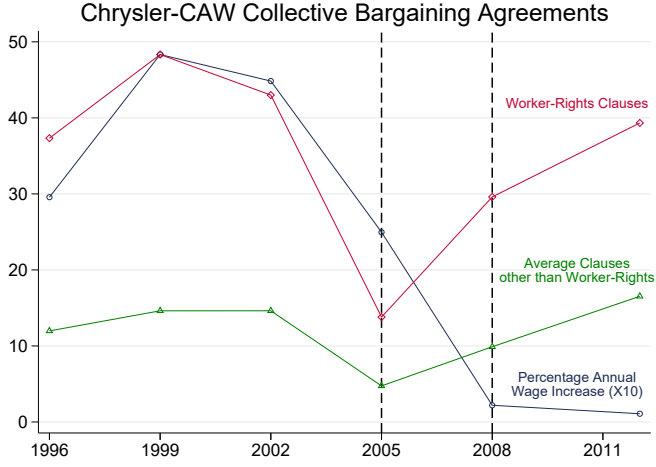


Note: Figure presents coefficients and 95% confidence intervals of regression of contract clause types on index for Pro-Worker HR Practices. Outcome: Clause type, defined as share of clauses of given type (number of clauses of type in question over the number of all clauses) . Treatment: Standardized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

Effect of 2000's Concession Bargaining on Auto Workers

Canadian Auto Workers president Buzz Hargrove on 2005 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."



Theory Predictions

- Stylized model extending Gruber & Poterba (1994), Dube, Naidu, & Reich (2022):
  - contract specifies pre-tax wage and rights (amenities), which are costly to draft.
  - workers maximize utility, firm maximizes profit and offers contract with wage and amenities
- Predictions:
  - rights increase with firm productivity
  - rights increase with taxes (pre-tax wages decrease $\rightarrow$ rights and wages are *substitutes* in this case)
  - rights increase with outside option (so do wages $\rightarrow$ rights and wages are *complements* in this case)

▸ Model

Fixed-Effects Specification

$$Y_{psit} = \rho Z_{pst} + \alpha_{ps} + \alpha_{st} + X'_{psit}\beta + \epsilon_{psit}, \quad (1)$$

- Y_{psit} = Text outcome (i.e. share of worker rights clauses) of contract adopted in province p, sector s, firm i, year t.
- Z_{pst} = Economic treatment variable of interest:
  - log labor income tax rate τ at average income $\overline{y}_{p,t}$ in province p at year t:

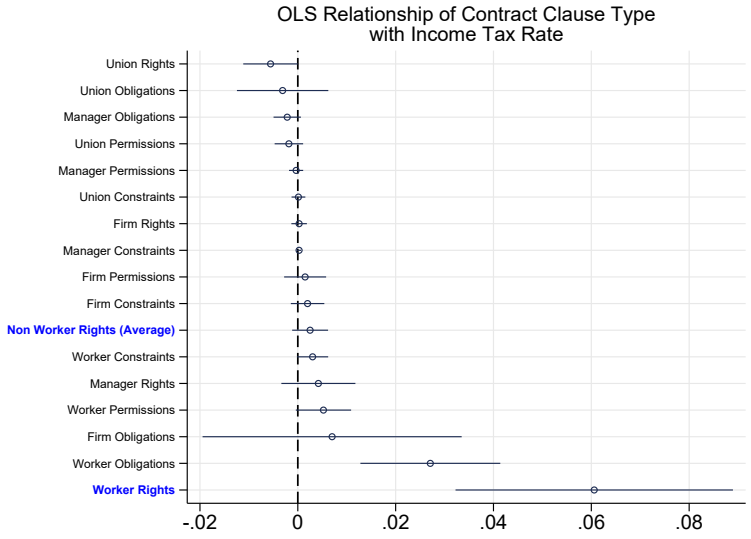
$$Z_{pt} = \log(\tau_{p,t}(\overline{y}_{p,t}) + \tau_{\text{federal},t}(\overline{y}_{p,t}))$$

  - Z_{pst}= log employment rate in sector s of province p at year t
- α_{ps},α_{st} = province×sector and sector×year fixed effects.
- X_{psit}: Time-varying controls and additional FE, for robustness checks.

→ Identification assumption: No time-varying province×sector-level confounders affecting both economic treatment and contract outcome.

- exogenous timing motivated by pre-determined contract negotiation schedule.
- (in the data: treatment variables are unrelated to firm exits, the number of employees, and whether the employees have a COLA clause).

Higher Income Tax → More Worker Rights



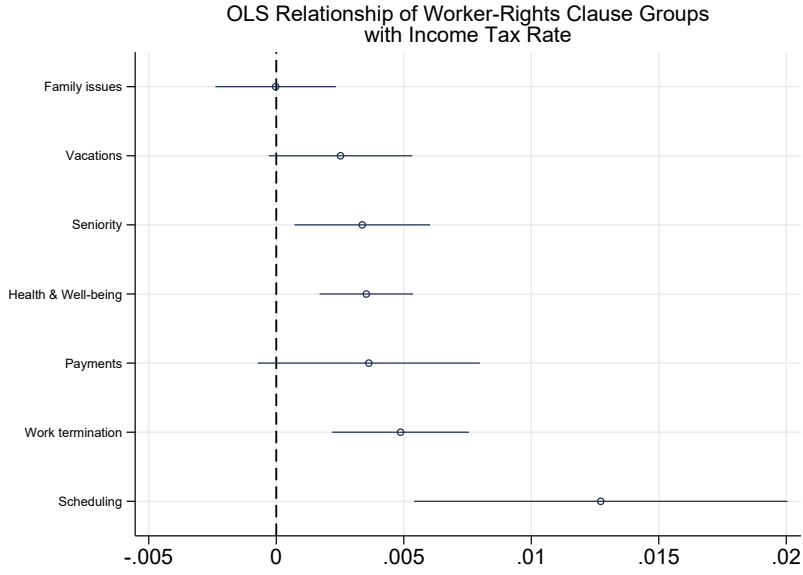
Note: Figure presents coefficients and 95% confidence intervals of effect of labour tax rate on contract clause types. Each coefficient is from a separate OLS regression. Outcome: Clause type, defined as the share of clauses of given type (number of clauses of type in question over the number of all clauses). Treatment: Labour tax rate, defined as logarithmized implicit personal income tax rate. Controls: Province-by-sector fixed effects and year-by-sector fixed effects. Inference: Standard errors clustered at the province-by-sector level. Data sources: Employment and Social Development Canada, Center for the Study of Living Standards.

Effect of Income Tax Rate Change on Worker Rights

	Worker-Rights Clauses											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log Income Tax Rate	0.060***	0.037***	0.041***	0.049***	0.060***	0.058***	0.060***	0.059***	0.060***	0.046***	0.035***	0.041***
	(0.014)	(0.011)	(0.014)	(0.018)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)	(0.011)	(0.012)
R-Squared	0.15	0.16	0.55	0.34	0.15	0.15	0.15	0.15	0.15	0.30	0.47	0.16
Number of Observations	24,826	24,826	22,554	10,841	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	X	X
Sector-Year FEs	X	X	X	X	X	X	X	X	X	X	X	X
Province Trends		X										
Firm Fixed Effects			X									
Union Fixed Effects				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

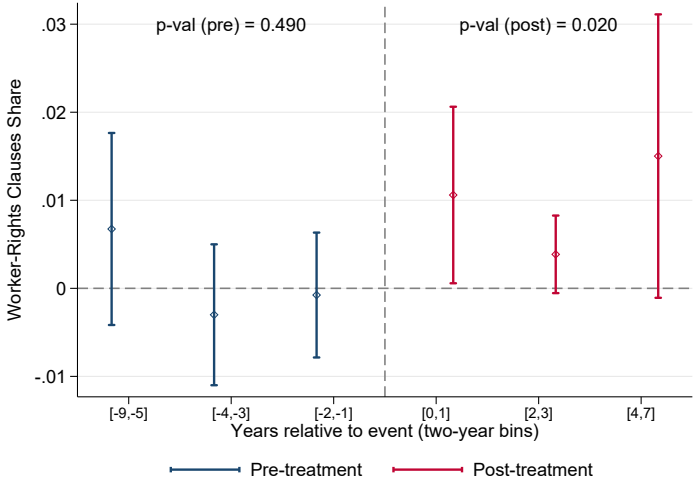
Note: Coefficients and standard errors of effect of labour tax rate on worker rights clauses, for different specifications as indicated in table footer. Outcome: Share of worker rights clauses, defined as number of worker rights clauses over the number of all clauses. Treatment: Labour tax rate is defined 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.

Effect of Income Taxes on Worker Rights, by Topic



▸ Assigning Clauses to Topics

Event Study: Largest Discrete Increase in Tax Rates by Province



Note: Figure presents coefficients and 95% confidence intervals for time indicators before and after labour tax rate increase on share of worker-rights clauses. Callaway & Sant'Anna (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 as the largest labour tax increase in a given province in the 1990s, where labour tax rate is defined as implicit personal income tax rate. Inference: Standard errors clustered at the province-by-sector level.

▸ Event-Study: Tax Decrease

Instrumental Variables Strategy

- Idea for instrument: Exogenous changes in *province* income tax due to changes in *federal* tax rates, driven by associated deductions/credits (see Gruber and Saez, JPubE 2002; Akcigit, Grigsby, Nicolas, Stantcheva, QJE 2021).
- Recall first Z_{pt} variable in fixed effects specification: :

$$Z_{pt} = \log(\tau_{p,t}(\overline{y}_{p,t}) + \tau_{\text{federal},t}(\overline{y}_{p,t}))$$

- Instrument with variable constructed from Kevin Milligan's CTaCS tax calculator:

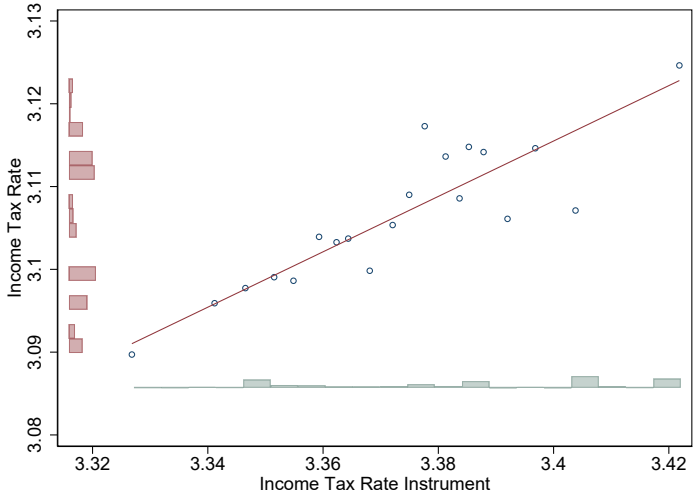
$$Z_{pt}^{\text{IV}} = \log(\tau_{p,t-k}(\overline{y}_{p,t-k}) + \tau_{\text{federal},t}(\overline{y}_{p,t-k}))$$

where $k \in \{1, 3\}$ is the lag in rates/income.

  - income is lagged $\rightarrow$ no endogenous income responses to taxes.
  - province rates are lagged $\rightarrow$ no endogenous tax responses to province-level confounders.
  - federal tax rates are not lagged $\rightarrow$ instrument isolates province tax changes due to changes in *federal* rates.

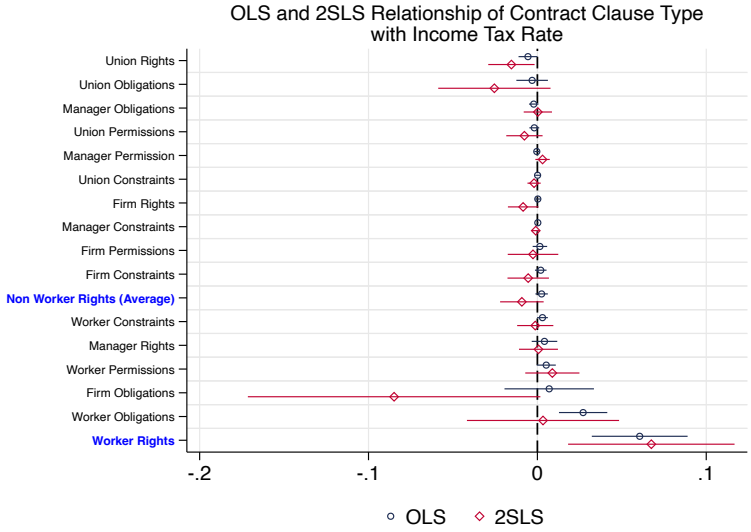
Tax IV First Stage

$$Z_{pt} = \gamma Z_{pt}^{IV} + \alpha_{ps} + \alpha_{st} + X'_{psit}\beta + \eta_{psit}$$



Note: Figure presents first stage binscatter of labour income tax rate (vertical axis) and predicted income tax rate based on lagged rates and exemptions (horizontal axis). Kleibergen-Paap First Stage F-Statistic = 65.08. Controls: Province-by-sector fixed effects and year-by-sector fixed effects. Data sources: ESDC, Center for the Study of Living Standards, Statistics Canada.

Higher Income Taxes → More Worker Rights (OLS & IV)

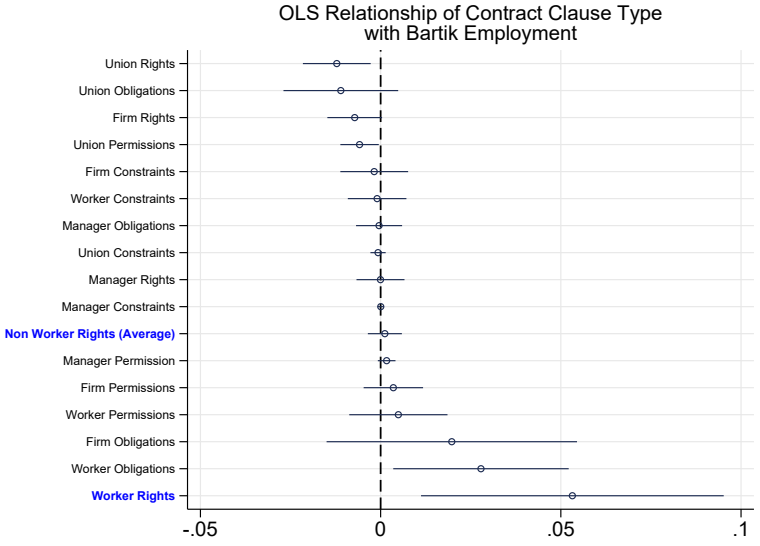


Note: Figure presents OLS and 2SLS coefficients and 95% confidence intervals of effect of instrumented labour tax rate on contract clause types. Outcome: Clause type, defined as the share of clauses of given type (number of clauses of type in question over the number of all clauses). Treatment: Log Labour tax rate in province s and year t, for 2SLS instrumented as the log of the sum of federal income tax rate of year t, calculated for the average income of province s and year t-k, and the province income tax rate of province s and year t-k, calculated for the average income of province s and year t-k, for k=1. Each coefficient is from a separate regression. Controls: Province-by-sector fixed effects and year-by-sector fixed effects. Inference: Standard errors clustered at the province-by-sector level. Data sources: Employment and Social Development Canada, Center for the Study of Living Standards, Statistics Canada.

Labour Demand Shock

- Employment rate in sector $\times$ province $\times$ year is a measure of workers' outside option:
  - costliness of strike to employers – more difficult to hire replacements.
  - also a measure of labour demand.
- Use leave-one-out sectoral employment rate ($\times$ province by year) as Bartik instrument – helps to isolate outside-option component.
- Positive labour demand shock improves bargaining position of unions relative to firms: We expect an increase in worker rights.

Better Outside Option → More Worker Rights



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.

Better Outside Option → More Worker Rights

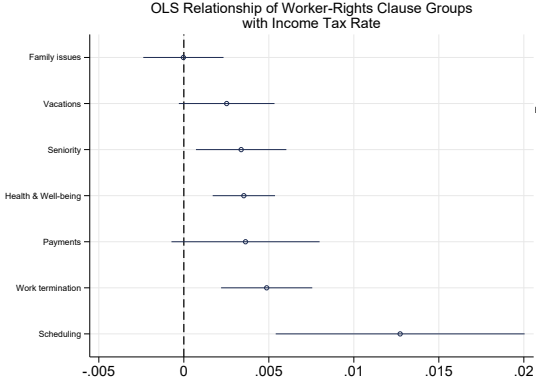
	Worker-Rights Clauses											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log Emp. Rate	0.053**	0.050***	0.040**	0.055**	0.053*	0.078***	0.056***	0.050**	0.049**	0.037**	0.035**	0.052***
	(0.021)	(0.019)	(0.018)	(0.027)	(0.024)	(0.019)	(0.017)	(0.020)	(0.021)	(0.018)	(0.014)	(0.017)
R-Squared	0.15	0.16	0.56	0.36	0.15	0.15	0.15	0.15	0.15	0.31	0.47	0.16
Number of Observations	29,157	29,157	26,669	13,735	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	X	X
Sector-Year FEs	X	X	X	X	X	X	X	X	X	X	X	X
Province Trends		X										
Firm Fixed Effects			X									
Union Fixed Effects				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.

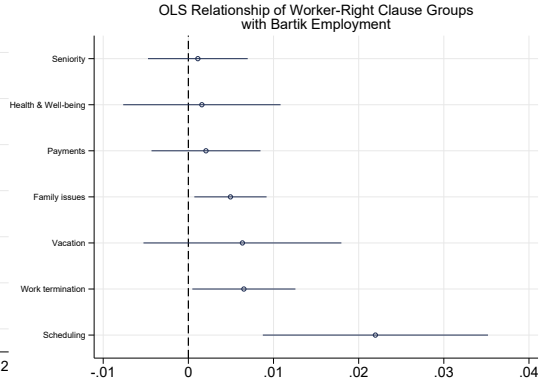
▸ Event Study ▸ Wage Regression

Taxes and Employment Shift the Same Topics

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.

▸ Assigning Clauses to Topics

Valuing Worker Rights in Terms of Wages

	(1) Share Worker Rights (S.D.)	(2) Log Wages
Log Tax Rate	**2.34*** **	**0.23*** **
	(0.55)	**(0.06)**
Union		1.34***
		(0.50)
Log Tax Rate * Union		**-0.36** **
		(0.16)
R-Squared	0.15	0.31
Number of Obs	24,826	4,877,128
Province-Sector FEs	X	X
Sector-Year FEs	X	X
Dataset:	Union Contracts	Labour Force Survey

- In response to a 10% increase in income taxes:
  - share of worker-rights clauses increases by 0.23 standard deviations ($\frac{10}{100} \times 2.34$)
  - union wages fall by 1.3% ($\frac{10}{100} \times (0.23 - 0.36)$)

→ **One std deviation increase in share of worker-rights clauses is worth about** $1.3\% \times \frac{1}{0.23} =$ **5.7% of wages**.

Amenity Value of Worker Rights

One standard deviation increase in share of worker-rights clauses is worth about 5.7% of wages.

- Compare to:
  - Mas and Pallais (2017): option for remote work worth 8% of wages.
  - 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.

Summary and Substantive Implications

- We demonstrate that the value of collective bargaining agreements is in worker rights clauses:
  - Personal Income Tax $\nearrow$ or Outside Option $\nearrow$: Increase in worker-rights clauses.
  - Substitution of wage and non-wage compensation.
- Evidence in support of Simon (1951): employment is an authority contract with constraints on employer power.
  - contracts allow employers to commit to providing protections and amenities that they otherwise would not provide ex post.
  - these constraints on employer discretion have real value for employees.
  - future work can go further in implementing and testing theories of decision rights (see MacLeod 2022, ch. 8).

Methodological Implications

- We show how to measure ***worker rights***, a dimension of worker surplus missed by existing datasets.
- Structural text analysis is important for studying law, and specifically measuring the economic value of legal texts.
- Upshot: 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.

Do Words Matter?

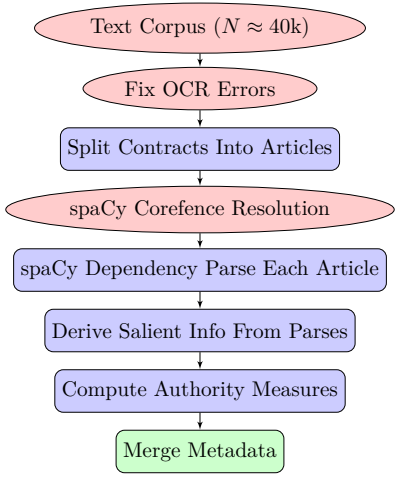
The Value of Collective Bargaining Agreements

Benamin Arold, Elliott Ash, Bentley MacLeod, Suresh Naidu

NBER Summer Institute, July 2024

Pipeline Overview

(Ash, Jacobs, MacLeod, Naidu, Stammbach 2020)



Back

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.
- Co-reference resolution by section: replace pronouns with referent entity
- Split the contracts into sections (RegEx) and sentences (spaCy):
  - 980,909 contract sections (33 per contract), 10.8 million sentences (11 per section)

Back

Examples: Worker / Firm Obligations

Worker Obligations

1. When a driver reports a defect in equipment on the repair report, he must tag or mark the vehicle involved in such a manner so that other 'employees will notice the defective equipment.
2. Each new employee, at the time of hire, shall sign an authorization for Union dues deductions which shall be sent to the Union Office.
3. An Employee who becomes aware of an Occupational Health or Safety concern at the Employee's worksite shall immediately notify the Employee's Supervisor or designate.
4. In a like manner, such teachers shall accept the professional responsibility of completing all activities connected with school closing.
5. In order to qualify for the overtime premiums provided above, the employee must work all scheduled hours in the work week.

Firm Obligations

1. The Hospital agrees to pay members of the Negotiating Committee for time spent during regular working hours in negotiations with the Hospital for a renewal agreement up to, but not including, arbitration.
2. The Employer will make space available for such functions subject to normal scheduling restrictions.
3. The Employer shall provide sanitary drinking water facilities on all jobs, when conditions require ice-water if available.
4. The Company shall maintain a file of such applications for transfer and shall consult such file when considering hiring new employees.
5. The Employer, its representatives and agents agree that there shall be no discrimination, interference, restriction or coercion exercised or practiced with respect to any employee within the bargaining unit.

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.

Back (Tax) Back (Employment)

All Worker-Rights Topics ($k = 32$)

Topic Label	Frequency
Work Hours	0.058
Workplace Safety	0.054
Payment Rules	0.051
Vacations	0.045
Leaves of Absence	0.039
Seniority-Based Benefits	0.039
Seniority-Based Vacation	0.038
Holiday Pay	0.037
Position Classification	0.036
Recall	0.032
Grievance & Discipline	0.031
Job Security	0.03
Seniority & Promotion/Transfer	0.03
Scheduling	0.03
Bereavement Leave	0.028

Topic Label	Frequency
Notice Requirements	0.027
Parental Leave	0.027
Termination	0.026
Overtime	0.026
Holiday Work Pay	0.026
Shift Premiums	0.025
Sick Leave	0.025
Personnel Records	0.024
Workplace Injuries	0.024
Part-Time Employment	0.023
Reimbursements	0.022
Probation Period	0.015
Meals	0.015
Breaks	0.013
Jury Duty	0.009

Interpreting the Topics

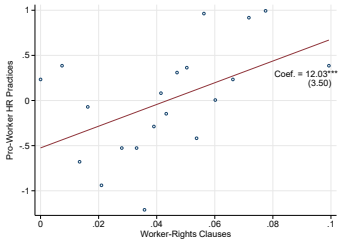
- Summarization:
  - Sample 10 clauses from a topic; generate summaries with GPT-4 ("I will give you a list of clauses sampled from collective bargaining agreements. These clauses are on a related topic in terms of giving similar rights to workers. Summarize in one sentence what types of rights the clauses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

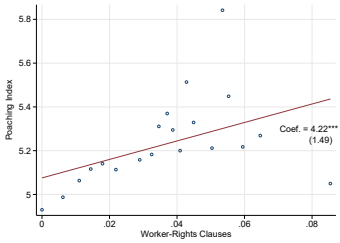
Comparison to Supervised-Learning-Based Method

- 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.

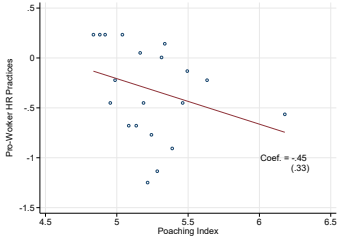
A. Pro-Worker HR Practices vs. Worker-Rights Clauses



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).

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 / labour 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
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 labour 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,

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).
  - $\rightarrow$ associated language representations are not optimized for specific task of contract design.
  - $\rightarrow$ 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.

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.

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

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

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

  - 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 labour demand.
  - $\epsilon \sim 1 - F(V - V^0)$, $F(\cdot)$ = cdf for probability of no strike, increasing and concave.
- Firm profit

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

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

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):

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

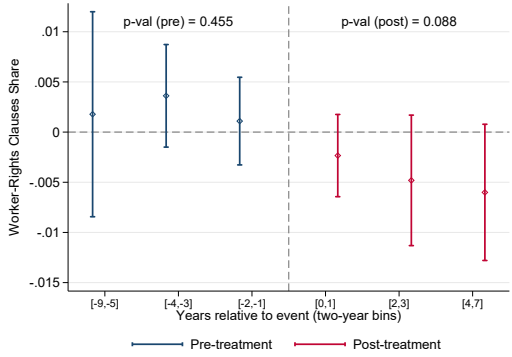
- Taking logs and re-arranging, we have

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

  - 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.

Back

Event Study: Discrete Decrease in Tax Rates



Note: Figure presents coefficients and 95% confidence intervals for time indicators before and after labor tax rate decrease (Panel B) on share of worker-rights clauses. Callaway & Sant'Anna (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 as the largest labor tax decrease in a given province in the 1990s, where labour tax rate is defined as implicit personal income tax rate. Inference: Standard errors clustered at the province-by-sector level.

Return

IV Strategy: Details

- Idea: Exploit changes in overall income tax that are driven by changes in federal-level taxes rather than province taxes (similar to Gruber and Saez (2002), and Akcigit et al. (2021)).
- Data: Federal and Province Level Tax Rate Data from Tax Calculator (CTaCS 2021 by Kevin Milligan), Income data from Statistics Canada,
- First stage: Log labour tax rate in province s and year t instrumented as the log sum of federal income tax rate of year t (for average income of province s and year t-k), and the province income tax rate of province s and year t-k (for average income of province s and year t-k), for k=1, 3:

$$\hat{z}_{sit} = ln(z^{federal}_{si(rate(t),income(t-k))} + z^{province}_{si(rate(t-k),income(t-k))})$$

- Second Stage: Regress contract feature (i.e. share of worker rights) of contract adopted in province s, firm i, and becoming effective in year t, on instrumented labour income tax rate z_{sit}, year-by-sector and province-by-sector fixed effects, α_{sit}, other time-varying controls X_{sit} in robustness checks, and an error term ϵ_{sit}:

$$y_{sit} = \rho\hat{z}_{sit} + \alpha_{sit} + X'_{sit}\beta + \epsilon_{sit}$$

Back

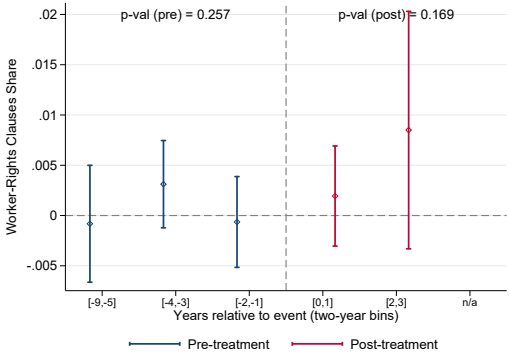
Income Taxes and Worker Rights: IV Regressions

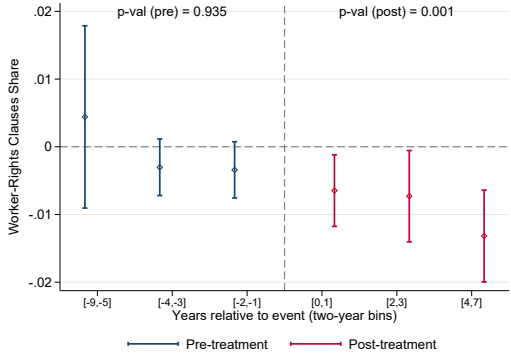
	Lag = 1 years			Lag = 3 years		
	First Stage	Reduced Form	2SLS	First Stage	Reduced Form	2SLS
Outcome:	Log Tax Rate	Worker Rights	Worker Rights	Log Tax Rate	Worker Rights	Worker Rights
	(1)	(2)	(3)	(1)	(2)	(3)
Instrumented Log Tax Rate	0.335***	0.023**	0.068***	0.212***	0.020*	0.092**
	(0.042)	(0.010)	(0.025)	(0.038)	(0.011)	(0.041)
Kleibergen-Paap rk Wald F-Statistic			65.08			31.67
R-Squared	0.92	0.15	0.003	0.92	0.15	0.002
Number of Observations	24,826	24,826	24,826	24,826	24,826	24,826
Province-Sector FEs	X	X	X	X	X	X
Sector-Year FEs	X	X	X	X	X	X

Note: Figure presents the first stage, reduced form, and 2SLS coefficients of estimation of instrumented labour tax rate on worker rights. Outcome: Clause type, defined as the share of clauses of given type (number of clauses of type in question over the number of all clauses). Treatment: Log Labour tax rate in province s and year t, instrumented as the log of the sum of federal income tax rate of year t, calculated for the average income of province s and year t-k, and the province income tax rate of province s and year t-k, calculated for the average income of province s and year t-k, for k=1, and k=3. Controls: Province-by-sector fixed effects and year-by-sector fixed effects. Inference: Standard errors clustered at the province-by-sector level. Single, double, and triple asterisks indicate statistical significance at the 10%, 5%,and 1% levels, respectively. Data sources: Employment and Social Development Canada, Center for the Study of Living Standards, Statistics Canada.

Back

Event-Study: Outside Options Change

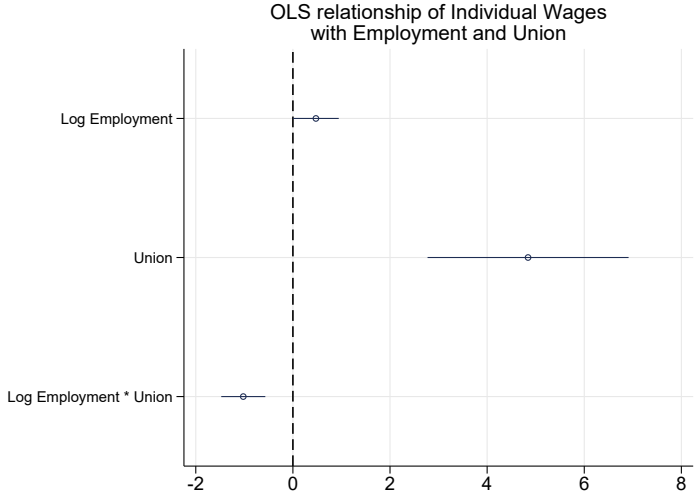




Note: Figure presents coefficients and 95% confidence intervals for time indicators before and after employment increase/decrease on share of worker-rights clauses. Callaway and Sant'Anna (2021) estimator, accounting for heterogeneous treatment effects and staggered treatment timing. Dynamic aggregation/event 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 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). Clustering at province-by-sector level.

Back

Effect of Employment Rates and Union Status on Wages



Back

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

Back