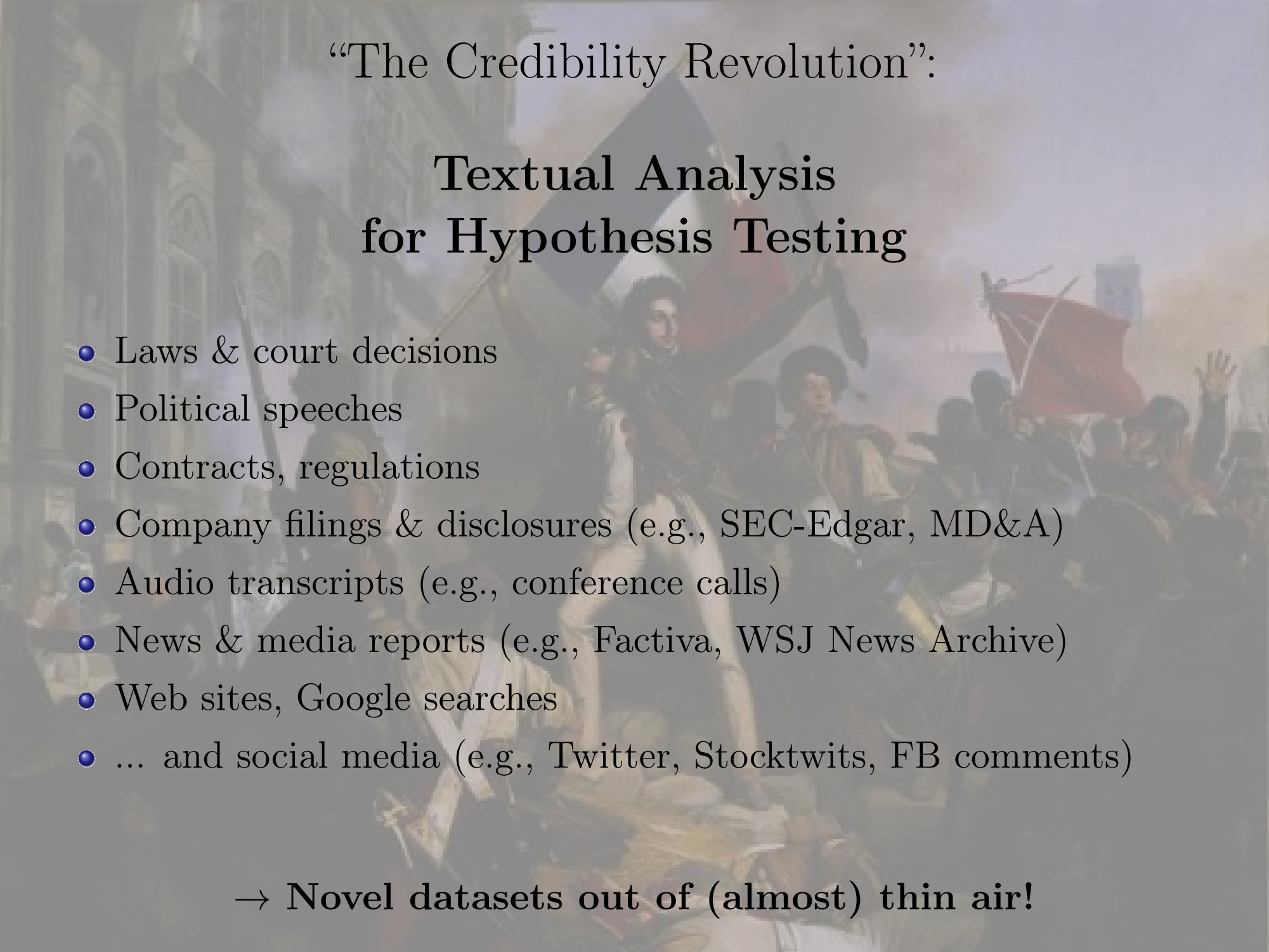


"The Credibility Revolution":

Textual Analysis for Hypothesis Testing

- Laws & court decisions
- Political speeches
- Contracts, regulations
- Company filings & disclosures (e.g., SEC-Edgar, MD&A)
- Audio transcripts (e.g., conference calls)
- News & media reports (e.g., Factiva, WSJ News Archive)
- Web sites, Google searches
- ... and social media (e.g., Twitter, Stocktwits, FB comments)

→ Novel datasets out of (almost) thin air!

Textual Analysis for Hypothesis Testing

Marian Moszoro[1]

[1]University of California, Berkeley and Kozminski University

European School on New Institutional Economics
Cargèse, May 18–22, 2015

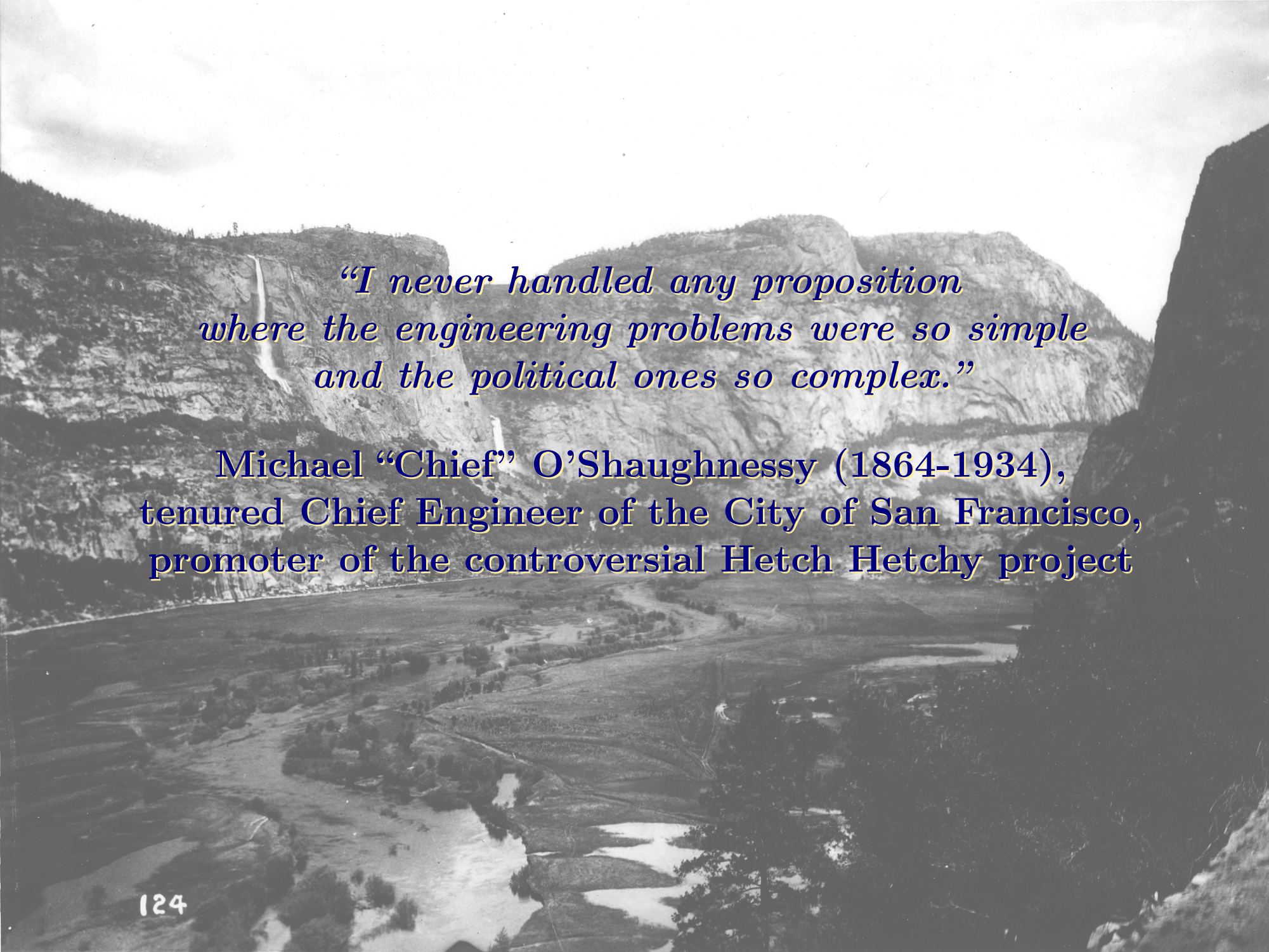
Techniques

1. Dictionaries
2. Identification & clustering
3. Parsing & comparison
4. “Fogginess”

Dictionaries

"Political Contestability and Contract Rigidity:
An Analysis of Procurement Contracts"

(Beuve, Moszoro & Saussier 2014)


"I never handled any proposition
where the engineering problems were so simple
and the political ones so complex."
Michael "Chief" O'Shaughnessy (1864-1934),
tenured Chief Engineer of the City of San Francisco,
promoter of the controversial Hetch Hetchy project
124

Scholarly Progression: A Research Agenda

Theory Development

↓

Formalization & Testable Hypotheses

↓

Operationalization & Test of Hypotheses

Spiller 2008: *"An Institutional Theory of Public Contracts"*

Moszoro & Spiller 2012: *"Third-Party Opportunism and the Nature of Public Contracts"*
Moszoro & Spiller 2014: *"Political Contestability, Scrutiny, and Public Contracting"*

Moszoro, Spiller & Stolorz 2013: *"Rigidity of Public Contracts"*
Aneja, Moszoro & Spiller 2014: *"Political Bonds: Political Hazards and the Choice of Municipal Financing Instruments"*
Beuve, Moszoro & Saussier 2014: *"Political Contestability and Contract Rigidity: An Analysis of Procurement Contracts"*

Perception of Public Contracts

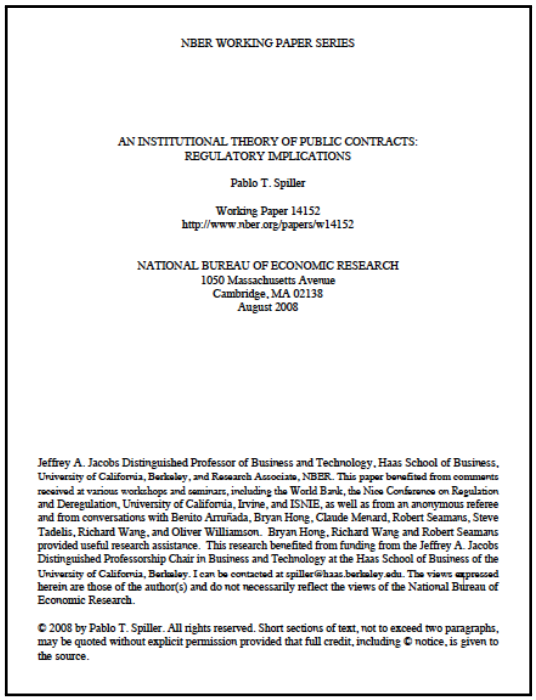
- inefficient
- low quality
- delays
- expensive
- corruption, favoritism
- bureaucratic, red tape
- politics
- intricate, convoluted
- scrutiny, regulation
- controls, inspections
- protests, courts
- specific, rule-based . . . **rigid**

Explanations of Contracts' Rigidity

1. **Transaction costs and contractual hazards**
   Crocker & Reynolds 1993, Saussier 2000, Bajari, Houghton & Tadelis 2014, ...
2. **Incentive & commitment**
   Bajari & Tadelis 2011, Guasch, Laffont & Straub 2008, ...
3. **Incompleteness, verifiability, and reference points**
   Hart, Fehr & Zhender 2011, Hart & Halonen-Akatwijuka 2014, ...
4. **Relational contracting & value of future business**
   Macaulay 1963, Baker & al. 2002, Gil & Marion 2013, ...

...but few ingredients germane to public contracts

Public Contract Rigidity and Third Party Opportunism



NBER WORKING PAPER SERIES

AN INSTITUTIONAL THEORY OF PUBLIC CONTRACTS:
REGULATORY IMPLICATIONS

Pablo T. Spiller

Working Paper 14152
http://www.nber.org/papers/w14152

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2008

Jeffrey A. Jacobs Distinguished Professor of Business and Technology, Haas School of Business, University of California, Berkeley, and Research Associate, NBER. This paper benefited from comments received at various workshops and seminars, including the World Bank, the Nice Conference on Regulation and Deregulation, University of California, Irvine, and ISNIE, as well as from an anonymous referee and from conversations with Benito Arruñada, Bryan Hong, Claude Menard, Robert Seamans, Steve Tadelis, Richard Wang, and Oliver Williamson. Bryan Hong, Richard Wang and Robert Seamans provided useful research assistance. This research benefited from funding from the Jeffrey A. Jacobs Distinguished Professorship Chair in Business and Technology at the Haas School of Business of the University of California, Berkeley. I can be contacted at spiller@haas.berkeley.edu. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

"A fundamental difference between private and public contracts is that public contracts are in the public sphere, and thus, although politics is normally not necessary to understand private contracting, it becomes fundamental to understanding public contracting" (Spiller 2008, "An Institutional Theory of Public Contracts," NBER Working Paper 14152, p. 3)

- Third-party opportunism prevents the use of relational contracts for public-private transactions
- Political contestability is an issue for public authorities

"Political Contestability" in a Nutshell

- Mainstream contract theory deals with incentives and frictions of the contracting parties

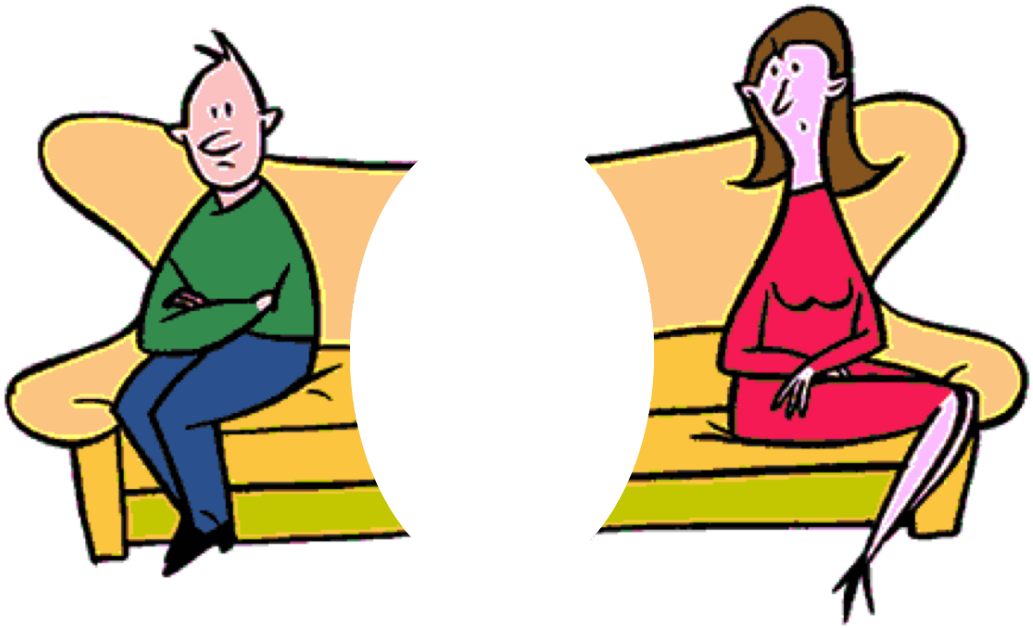


Figure: Monster-in-Law

"Political Contestability" in a Nutshell

- Mainstream contract theory deals with incentives and frictions of the contracting parties
- But there may be **third parties**...

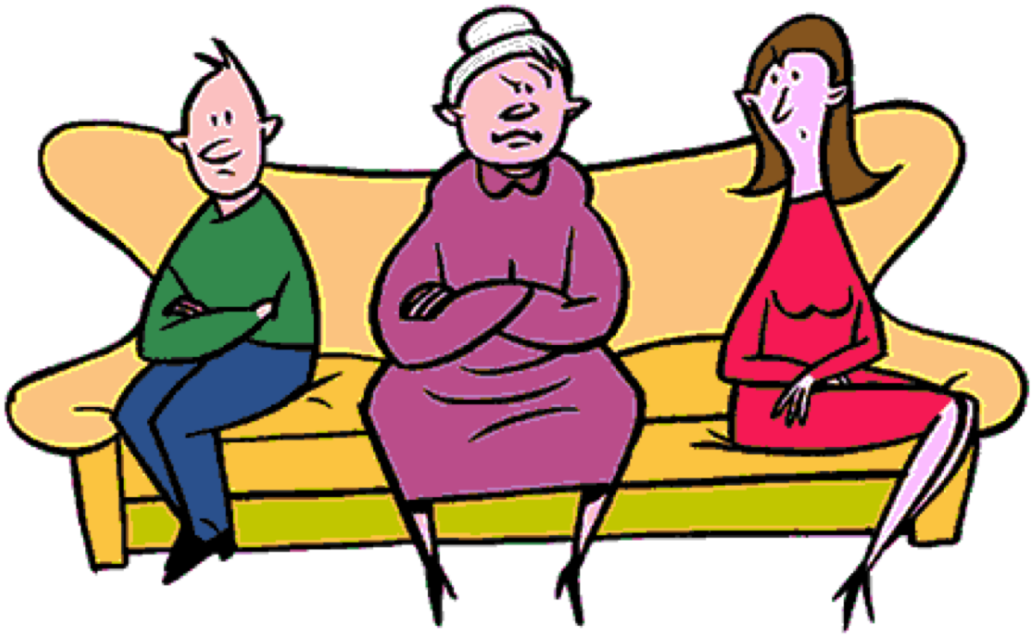


Figure: Monster-in-Law

... not necessarily interested in the success of the relationship (political opponents, excluded bidders, and interest groups)

A Model of Public Procurement under Political Contestability

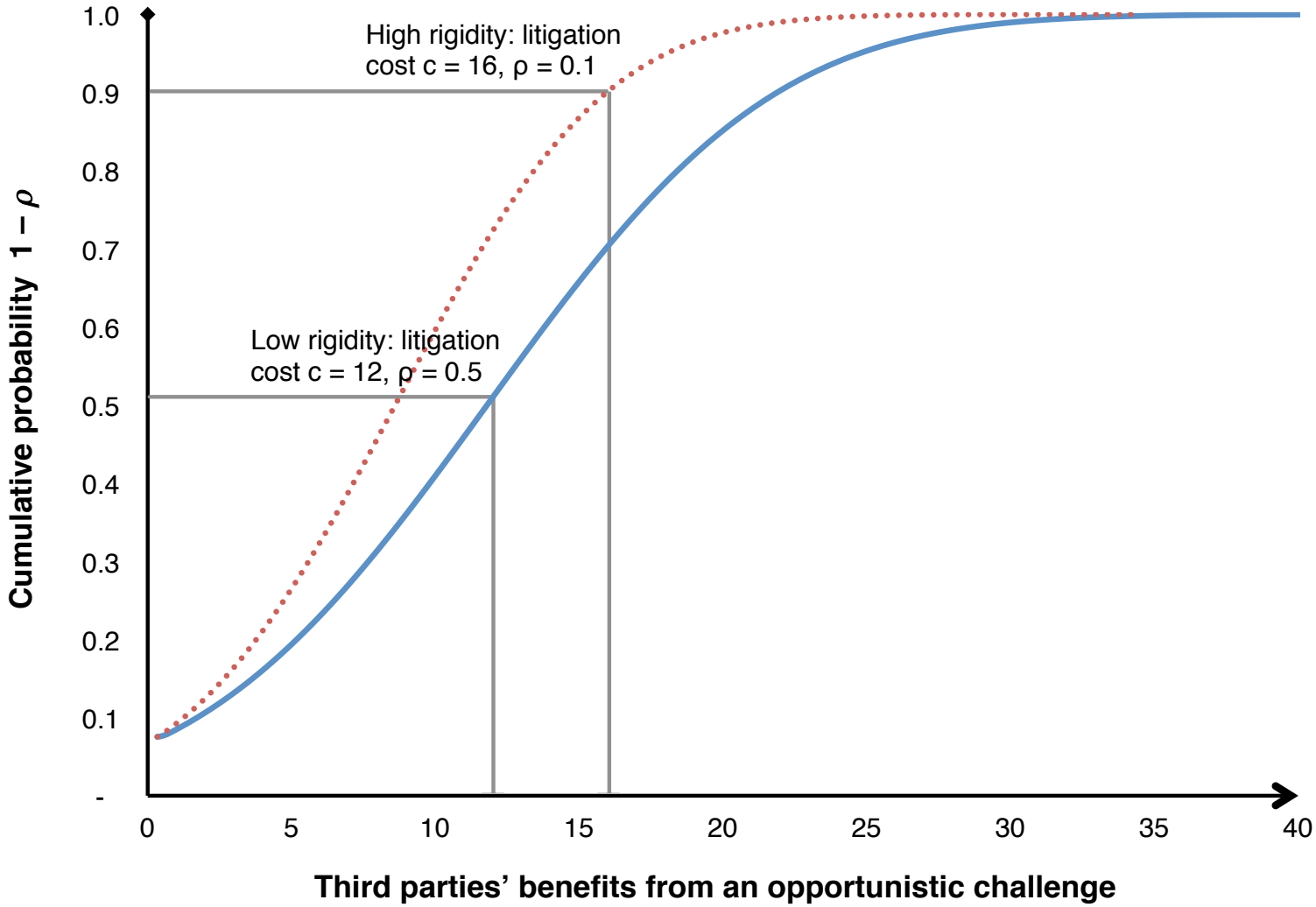
Preliminaries:

- Public agent's perspective
- Simple contract for standard good/service

Four agents explicitly and implicitly involved in public contracting:

1. Incumbent public agent who minimizes political risks and contracting costs
2. Private contractor who maximizes profit
3. Third-party challengers, i.e., political opponents to the incumbent public agent, competitors to the contractor, and interest groups ("anti-arbitrators") who extract political rents
4. Public at large, i.e., voters and courts who allocate political power

Endogeneity of Opportunistic Challenge



Moszoro & Spiller (2012)—Results

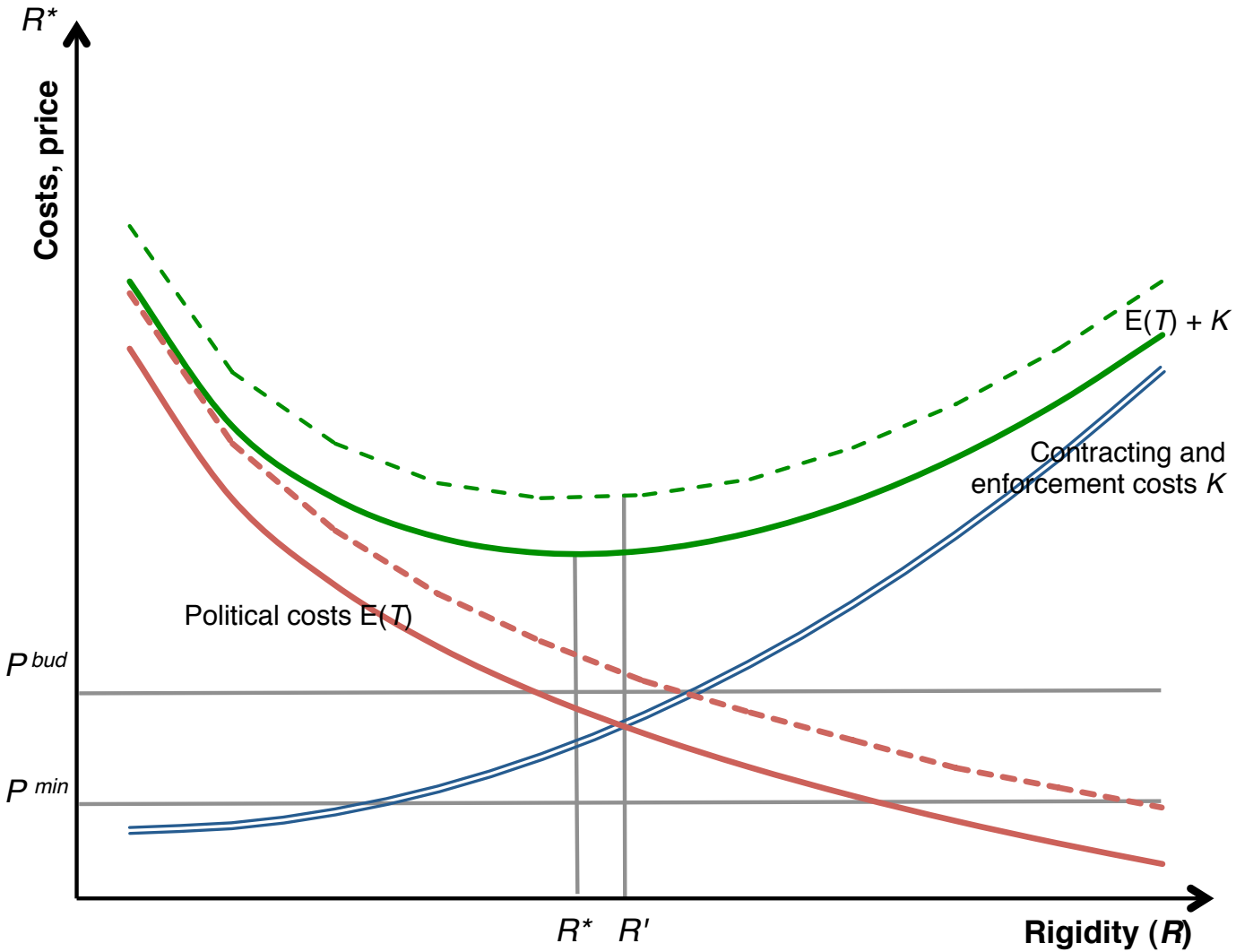
- Opportunistic challenges as **key differential hazard** of public transactions
- Rigidity in public contracting is a **political risk adaptation** by public agents
  - Public agents **limit** the risk of third parties' challenges through formalities and rules
  - ... **externalizing** the associated costs to the public at large
- Testable predictions:
  1. Contracts subject to public scrutiny show more rigidity clauses than purely private (i.e., relational) contracts
  2. In the sub-sample of public contracts, rigidity increases with political contestability

Moszoro & Spiller (2012)—Results

- Opportunistic challenges as **key differential hazard** of public transactions
- Rigidity in public contracting is a **political risk adaptation** by public agents
  - Public agents **limit** the risk of third parties' challenges through formalities and rules
  - ... **externalizing** the associated costs to the public at large
- Testable predictions:
  1. Contracts subject to public scrutiny show more rigidity clauses than purely private (i.e., relational) contracts
  2. In the sub-sample of public contracts, rigidity increases with political contestability

Public Oversight $\times$ **Political Contestability** $\uparrow$ $=$ **Contractual Rigidity** $\uparrow$

Opportunistic Challenge & Optimal Contract Rigidity



Empirical Approach

Research Question & Propositions

Are **public contracts more rigid** than private contracts?

Propositions

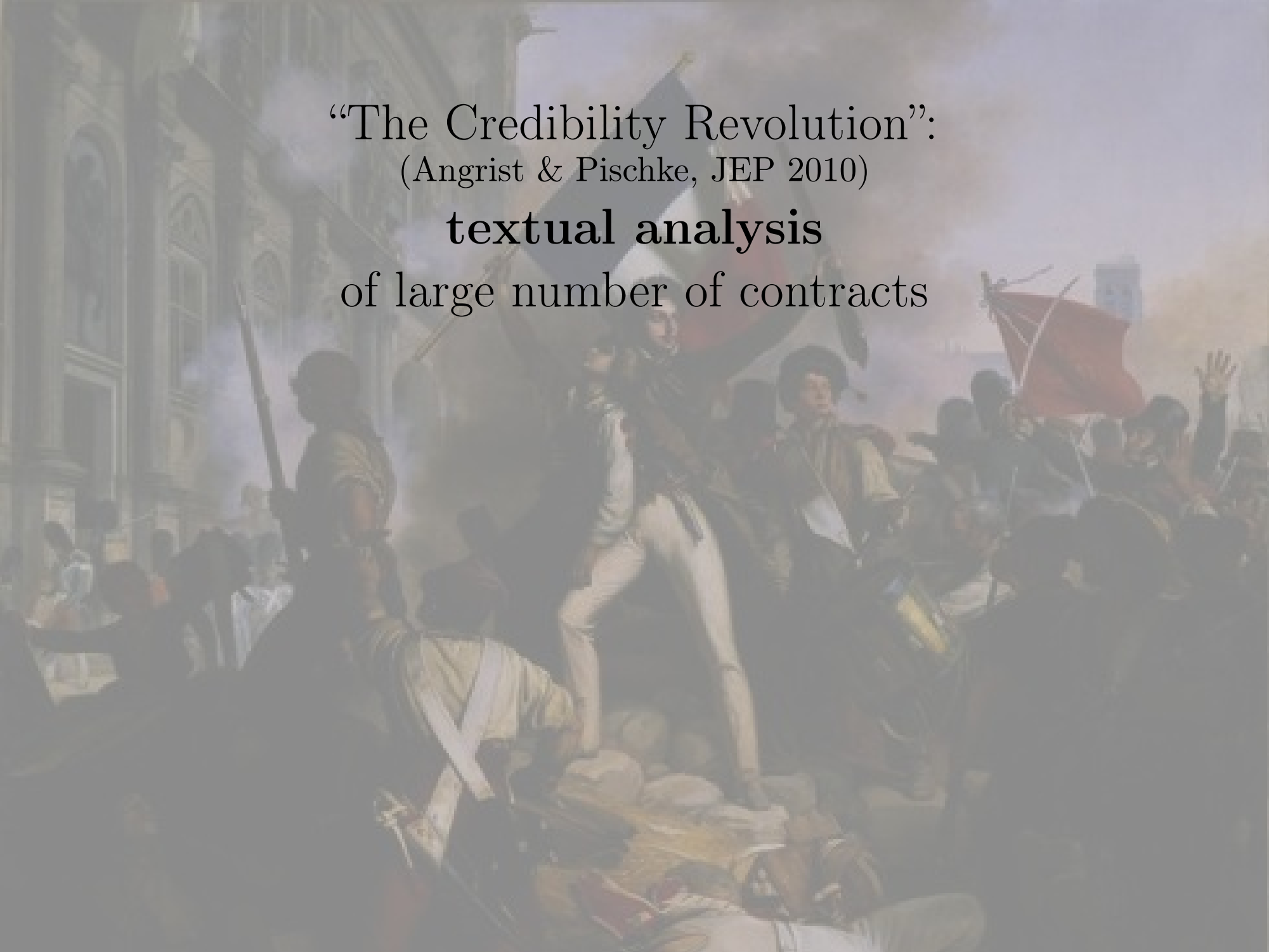
1. Public contracts have more rule-based clauses than private contracts
2. Public contracts' rigidity rises in political contestability
3. Public contracts are more often renegotiated through formal amendments

Ideal Experiment

- Sequence of contracts between A and B on comparable product
- Variation over time in "oversight" and "contestability" associated with one contracting party
- See how these contracts vary with this oversight and contestability

High Bar on Data Requirements

- Data on many contracts:
  (a) Standard product/service
  (b) Over time
  (c) Across contractees
- Measures of contractual features
- Measures of public oversight and political contestability



"The Credibility Revolution":
(Angrist & Pischke, JEP 2010)

textual analysis

of large number of contracts

"The Credibility Revolution":
(Angrist & Pischke, JEP 2010)

textual analysis

of large number of contracts

- Scanned & OCR-ed car parking contracts from 1963 to 2009
- Merged with election outcomes in French municipalities
- Contract classification: concession, provision of services, operating
- Standard product, one contractor: heterogeneity from public vs. private (i.e., politically vs. non-politically contestable) procurers and in-sample time-varying political contestability

$\rightarrow$ **Novel dataset of $\approx$1.3K contracts**

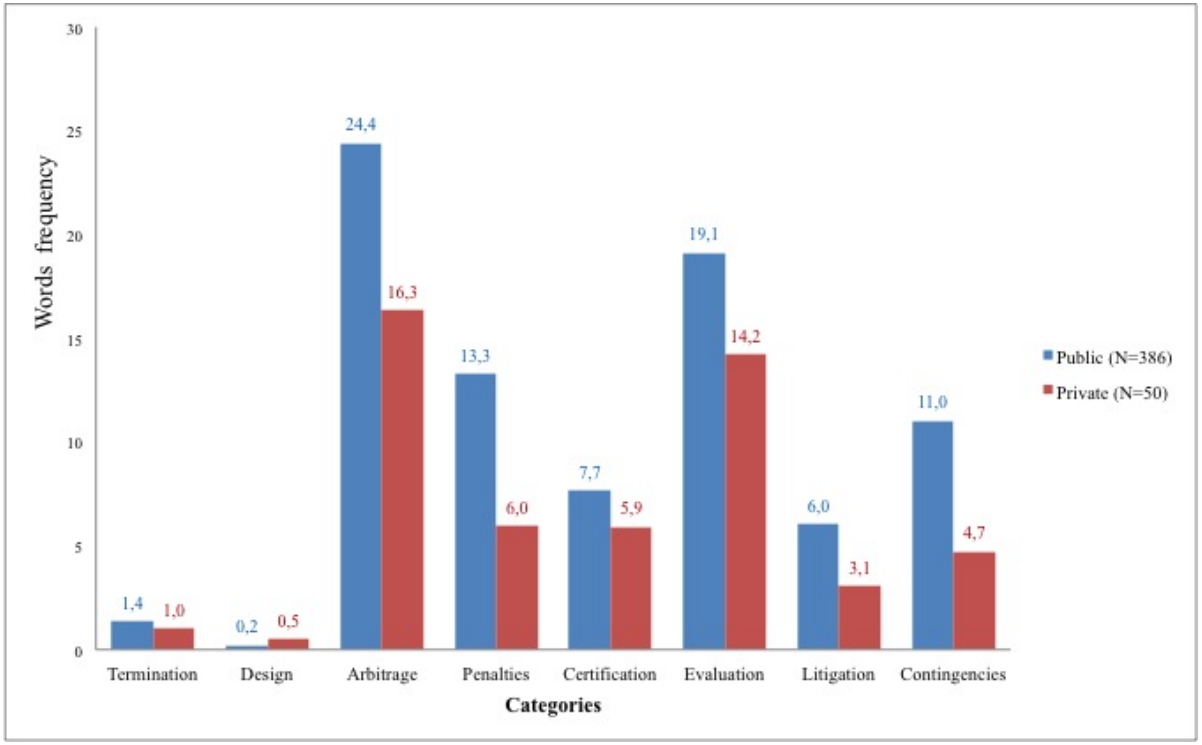
Dependent Variable: Contract Rigidity

- Follow on Schwartz & Watson (2012): number of arbitration clauses and cost of contract design
- Build Moszoro, Spiller & Stolorz (2014): several dimensions of contract rigidity
- Apply algorithmic data reading and textual analysis to compare the complexity of public contracts subject to public scrutiny with relational private contracts

Keywords in Rigidity Categories: "Dictionaries"

Arbitration	appeal, arbitration, conciliation, guarantee, intervention, mediation, settlement, warranty, whereas	10,241
Certification	certification, permit, regulation	3.236
Evaluation	accountability, control, covenant, obligation, quality, specification, scrutiny	8,090
Litigation	court, dispute, indictment, jury, lawsuit, litigation, pleading, prosecution, trial	2,479
Penalties	damage, fine, indemnification, penalty, sanction	5,431
Termination	breach, cancel, dissolution, separation, termination, unilateral	580
Contingencies	contingent, if, provided that, providing that, subject to, whenever, whether	4,488
Design	anticipation, event, scenario, plan	109

Contract Rigidity at First Glance



We normalized rigidity variables: e.g.,

$$zArbitration = \frac{Arbitration - \mu}{\sigma}$$

and focus on total rigidity defined as:

$$zRigidity = zArbitration + zCertification + zEvaluation + zLitigation + zPenalties + zTermination + zContingencies + zDesign$$

Explanatory Variables

- Dummy Public versus Private
- Political Contestability
  (a) **HHI**: Herfindahl-Hirschman Index of the first round of elections preceding the date of signature
  (b) **Residual HHI**: Concentration of all non-winning parties
  (c) **Number of lists**: Number of competing lists during prior election
  (d) **Win_Margin**: Margin of victory between the winner and the runner-up party
  (e) **Distance**: Time between the date of signature and the date of future election

Control Variables

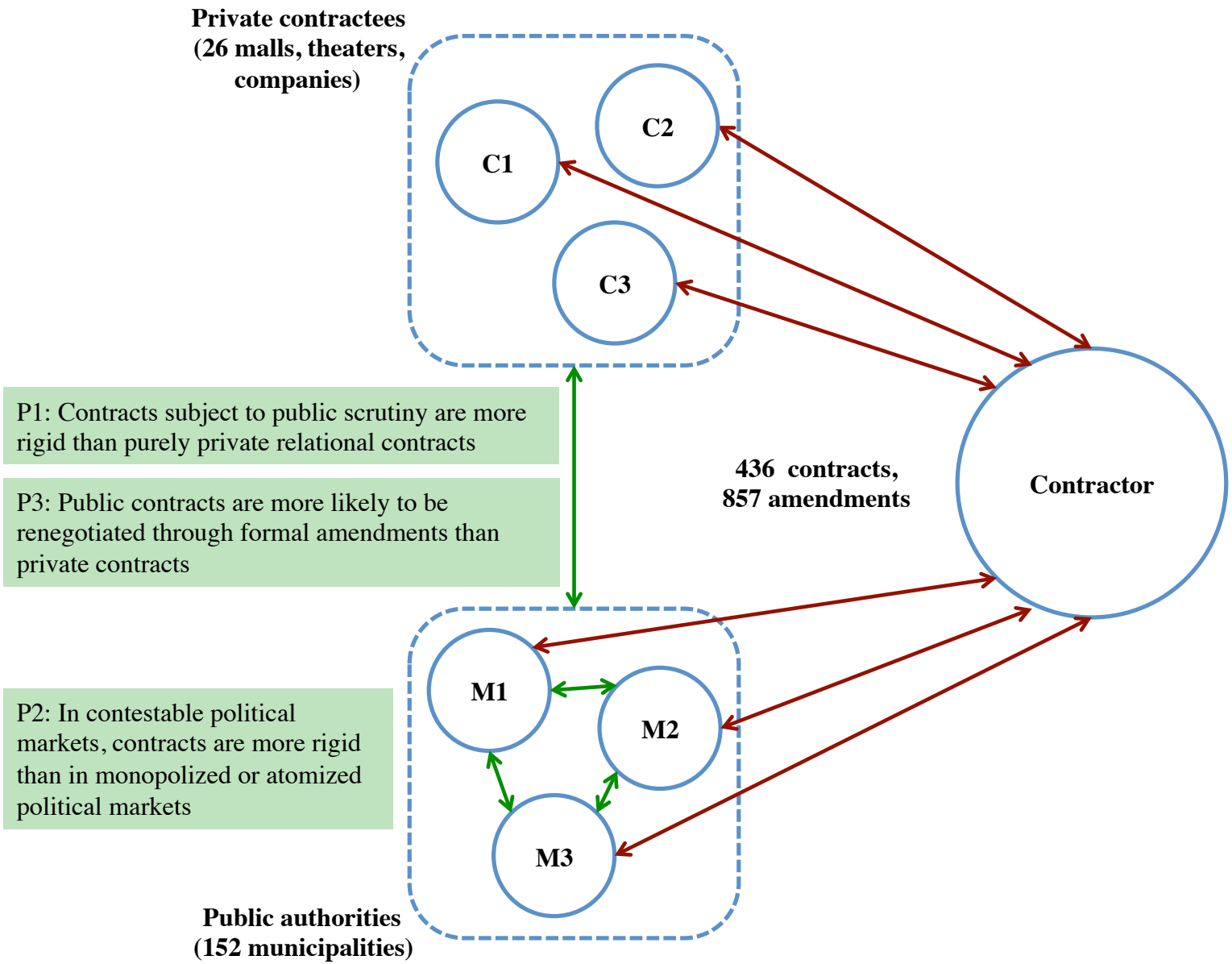
- Type of contracts (Operating, Provision of Services, Concession)
- Size of the city (number of voters, inhabitants)
- Political color of the mayor (left wing versus right wing)
- Renewed contract (dummy)
- Trend
- Year dummy variables
- Department dummy variables

High Bar on Data Requirements: Car Parks

- Data on many contracts:
  - (a) Standard product/service
  - (b) Over time
  - (c) Across contractees
- Measures of contractual features
- Measures of public oversight and political contestability

✓ 436 contracts + 857 amend.
✓ one country/provider/product
✓ 1985-2009
✓ 26 firms, 152 municipalities
✓ *zRigidity*
✓ private vs. public
✓ HHI, win margins, # of lists, ...

Propositions



Results

P1: Public versus Private Contract Rigidity

Dependent variables:	***zRigidity***	***zDesign***	***zTermination***	***zArbitrage***	***zPenalties***	***zCertification***	***zEvaluation***	***zLitigation***	***zContingencies***
Public	12.045***	-0.625	1.246*	2.189**	3.306***	0.309	1.849+	2.596***	0.921**
	(4.364)	(0.553)	(0.701)	(0.829)	(0.505)	(0.659)	(1.235)	(0.548)	(0.439)
Renewed	-4.990*	-0.336	-0.727	-0.697+	-0.279	0.235	-1.099*	-1.099*	-0.830**
	(2.632)	(0.244)	(0.506)	(0.477)	(0.417)	(0.494)	(0.607)	(0.606)	(0.400)
Provision_ of_ Services	-15.204***	-0.701**	-2.207***	-1.721***	-3.034***	-1.522**	-2.768***	-1.621***	-0.886**
	(2.972)	(0.329)	(0.542)	(0.506)	(0.375)	(0.597)	(0.750)	(0.476)	(0.362)
Concession	-0.969	0.014	-0.431	0.046	0.229	0.377	-0.819+	-0.500	0.327
	(1.719)	(0.216)	(0.478)	(0.350)	(0.497)	(0.395)	(0.524)	(0.499)	(0.436)
Inhabitants	-0.038	-0.099+	-0.008	0.015	-0.109	-0.010	0.123	0.118	0.003
	(0.640)	(0.063)	(0.106)	(0.141)	(0.095)	(0.127)	(0.214)	(0.173)	(0.125)
Left_ Wing	-1.511	0.020	-0.765	-0.582	0.738	-0.241	-0.714	-0.342	0.416
	(2.832)	(0.173)	(0.811)	(0.585)	(0.596)	(0.587)	(0.719)	(0.724)	(0.488)
Right_ Wing	3.506	0.376	1.101+	0.094	0.412	1.175***	0.189	0.344	-0.463*
	(2.587)	(0.328)	(0.664)	(0.439)	(0.446)	(0.380)	(0.804)	(0.535)	(0.272)
Trend	1.256***	0.031	0.173***	0.244***	0.051	0.199***	0.280***	0.142**	0.094*
	(0.419)	(0.068)	(0.046)	(0.069)	(0.094)	(0.062)	(0.102)	(0.066)	(0.054)
Year Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Constant	-2,515.520***	-58.567	-345.601***	-490.111***	-103.749	-398.061***	-560.806***	-287.003**	-187.776*
	(831.354)	(134.959)	(92.086)	(136.395)	(188.939)	(122.400)	(201.751)	(131.371)	(108.003)
N	436	436	436	436	436	436	436	436	436
r^2	0.222	0.136	0.138	0.156	0.313	0.114	0.168	0.142	0.111

P2: Political Contestability Effect (Public Contracts)

Dependent variable:	***zRigidity***			
Renewed	-3.908	-4.406+	-4.235	-3.879
	(2.968)	(2.911)	(2.966)	(2.994)
Provision_of_Services	-17.455***	-17.298***	-17.158***	-17.537***
	(2.434)	(2.391)	(2.464)	(2.434)
Concession	-0.540	-0.275	-0.865	-0.803
	(2.693)	(2.590)	(2.724)	(2.635)
Inhabitants	0.660	-0.396	1.317	1.566+
	(1.168)	(1.291)	(1.161)	(1.062)
Left_Wing	1.227	0.236	1.699	1.116
	(3.787)	(3.935)	(3.900)	(3.807)
Right_Wing	2.136	2.081	1.069	1.587
	(2.756)	(2.693)	(2.789)	(2.852)
Trend	0.682*	0.606+	0.746*	0.707*
	(0.387)	(0.380)	(0.381)	(0.363)
Political Contestability Variables				
HHI	-14.390+	.	.	.
	(9.075)	.	.	.
Residual_HHI	.	19.953+	.	.
	.	(12.348)	.	.
Number_of_Lists	.	2.208***	.	.
	.	(0.684)	.	.
Win_Margin	.	.	0.215*	.
	.	.	(0.126)	.
Win_Margin^2	.	.	-0.003***	.
	.	.	(0.001)	.
Distance	.	.	.	-1.478
	.	.	.	(1.913)
Year Dummies	yes	yes	yes	yes
Constant	-1,356.617*	-1,218.703+	-1,501.107*	-1,412.583*
	(779.453)	(763.015)	(767.244)	(719.988)
N	331	331	331	331
r^2	0.251	0.262	0.254	0.245

P3: Frequency of Contract Renegotiations

- Public contract should be renegotiated more frequently through formal amendments than private ones (i.e., no relational contract)

  "When faced with unforeseen or unexpected circumstances, private parties, as long as the relation remains worthwhile, adjust their required performance without the need for costly renegotiation or formal recontracting" (Spiller 2008, page 1)

  "The risk of third-party opportunism means that 'relational' contracting is less likely to evolve in the public sphere" (Spiller 2008, page 13)

H3: Frequency of Contract Renegotiations

Dependent Variable	*AverageAmendments*	
Public	0.084*	0.090*
	(0.046)	(0.053)
Renewed	-0.116**	-0.122**
	(0.045)	(0.053)
Duration	-0.010***	-0.010***
	(0.002)	(0.002)
Provision_ of_ Services	0.078**	0.093**
	(0.037)	(0.041)
Concession	-0.046	.
	(0.044)	.
Inhabitants	-0.010	0.010
	(0.013)	(0.014)
Left_ Wing	-0.022	-0.047
	(0.049)	(0.071)
Right_ Wing	0.035	0.019
	(0.065)	(0.071)
Trend	-0.028***	-0.021***
	(0.008)	(0.003)
Year dummies	yes	yes
Constant	55.423***	42.531***
	(15.421)	(6.871)
N	436	339
r^2	0.224	0.245

Robustness Checks

1. Contract duration
   - Results are changed by introducing contract duration
2. Contract type
   - Very few concessions in our private contracts sample (2%)
   - Results not changed without concession contracts
3. Geographical FE & Clustering
   - Results robust to department fixed effects and clustering

Discussion

Note on Corruption

Minor concern to our setting:

- TPO $\sim$ endogenizing anti-corruption rules by increased rigidity:
  if corruption $\rightarrow$ less, not more rigidity of public contracts
- Signed contracts, not bid specifications; no use for the corrupted public agent to restrict the favored bidder at the contracting stage:
  if corruption $\rightarrow$ estimates biased towards less rigidity
- Single contractor, largest car park company in France:
  if corruption $\rightarrow$ reputational spillovers $>$ potential gains

Limitations and Alternative Stories

- **Algorithmic textual analysis $\sim$ imperfect interpretation**
  Strong results indicate not spurious correlations; better algorithms and "dictionaries" will corroborate findings
- **Corruption as confounding factor**
  Not much relevant to our setting; if absent, even stonger results
- **Omitted variables: demand stochasticity and prices**
  No data; year and geographic fixed effects

Identification & Clustering

"Rigidity of Public Contracts"

(Moszoro, Spiller & Stolorz 2014)

Projects using Political Contestability Framework

1. *Contracting and Political Economy: "Rigidity of Public Contracts"* (Moszoro, Spiller & Stolorz)
   - SEC-Edgar database, ∼1 million scanned & OCR-ed filings
   - Machine reading, data scraping, and word clustering from Exhibits 10 (contracts) in 10-K's
   - Company identification by CIK and SIC
   - Document classification by keywords in first 100 lines
   - H: public contracts (utilities) are lengthier and feature more rigidity clauses than private contracts

Data Treatment

Step	Treatment	Count
1	Readable filings	206,677
	Filing companies	14,043
	Average filings per company	15
	Average filing length (geometric average of "the", "and", "of")	285
2	Sample industry diversity: identified different 4-code SIC	443
	Dropped non-readable filings	3,670
	Dropped files with no CIK or SIC codes identified	26,282
	Dropped files SIC 6*** (Finance) and SIC 9*** (Administration)	32,982
	Public utilities contracts (SIC 4900–4999)	11,657
	Quasi-regulated industries contracts (SIC 4000–4499 & 4800–4899)	8,543
	Total public contracts	**20,200**
	Total private contracts	**123,543**
4	Keywords count overall	5,644,668
	Arbitration	396,178
	Certification	872,843
	Evaluation	1,304,934
	Litigation	289,750
	Penalties	773,392
	Termination	1,940,419
	Design	67,152
5	Filings with identified categories (categories may overlap)	126,913
	Amendment	96,552
	Comercial contracts	54,344
	Compensation/Employment	88,238
	Consulting	4,559
	Finance	50,492

Keywords Clustered in Rigidity Categories: "Dictionaries"

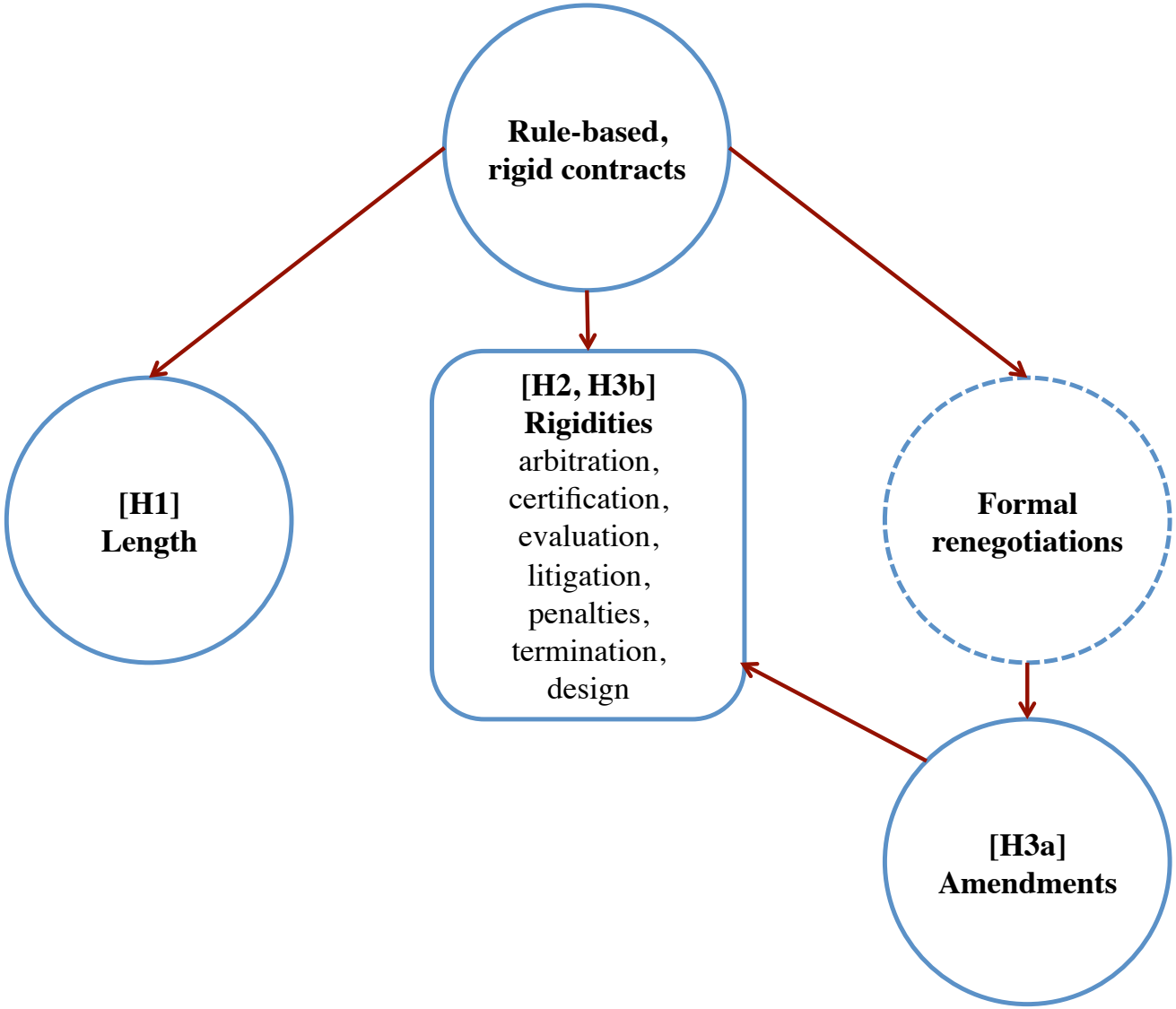
Arbitration	**Certification**	**Evaluation**	**Litigation**
arbitration, conciliation, settlement, whereas	certification, regulation	obligation, quality, scrutiny	dispute, indictment, jury, litigation
Penalties	**Termination**	**Design**	
fine, penalty, sanction	dissolution, termination	anticipation, planning, scenario	

$$Length_i = ln \sqrt[3]{\sum \text{“the”} \times \sum \text{“and”} \times \sum \text{“of”}} \quad \text{(geometric average)}$$

$$Rigidity_{i,l} = ln \frac{\sum \text{rigidity words } l \text{ in contract } i}{\sqrt[3]{\sum \text{“the”} \times \sum \text{“and”} \times \sum \text{“of”}}} \quad \text{(frequency)}$$

Research Question & Hypotheses

Are **public contracts more rigid** than private contracts?



Rigidity Clauses in Utilities Contracts

	(1) Arbitration	(2) Certification	(3) Evaluation	(4) Litigation	(5) Penalties	(6) Termination	(7) Design
Utilities	0.368**	0.175**	0.353***	0.524***	0.234***	-0.133	0.0810
	(2.55)	(1.98)	(3.91)	(4.34)	(2.76)	(-1.25)	(0.76)
Quasi-regulated	-0.137	0.0123	0.0494	-0.0223	0.0574	-0.0896	0.144
	(-1.03)	(0.14)	(0.60)	(-0.21)	(0.74)	(-0.95)	(1.19)
Length	-0.535***	-0.150***	-0.0667***	-0.456***	-0.337***	-0.270***	-0.754***
	(-16.64)	(-8.17)	(-3.76)	(-17.94)	(-21.58)	(-14.18)	(-28.89)
Assets Total	-0.0180	-0.0201***	-0.0268***	-0.0357***	0.0364***	0.0307***	0.0102
	(-1.47)	(-2.74)	(-3.67)	(-3.64)	(5.16)	(3.69)	(0.98)
Constant	-1.167***	-2.917***	-3.045***	-1.760***	-2.399***	-1.939***	-1.050***
	(-6.06)	(-27.72)	(-29.99)	(-11.89)	(-25.83)	(-17.36)	(-6.34)
Observations	1008	1552	1661	1044	1640	1664	487
Adjusted R^2	0.221	0.048	0.020	0.262	0.233	0.112	0.632

Concluding Remarks (1)

Empirical results:

- Public contracts are **more rigid**: feature more arbitration, evaluation, litigation, penalty, termination, and contingency clauses, and their renegotiation is formalized in **amendments**
- These patterns are reinforced in **political contestability** in the subsample of public contracts

Concluding Remarks (2)

Apart from the empirical results themselves, our paper contributes to the literature in a threefold manner:

(a) Replicable methodology for the analysis of contracts

(b) "Dictionaries" that are descriptive of the multidimentional characteristics of public vs. public contracts

(c) Plausible rationale with testable hypotheses of the difference between public and purely private contracts: the observed higher rigidity of public contracts is a political risk adaptation of public agents by which they lower the likelihood of success of third-party (opportunistic) challengers

Masters of "Dictionaries"

(a) Finance: Loughran & McDonald

- Journal of Finance (2011):
  – Created dictionary of all words occurring in 10-Ks in 1994-2007
  – Classified words occurring in 5% or more of documents
- http://www3.nd.edu/~mcdonald/Word_Lists.html
  Master dictionary, 10-X File Summaries, many resources

(b) Law: Schwartz & Watson

- Regress arbitration words and contract length

Loughran & McDonald (JF 2011) Financial Dictionaries

- **Fin-Neg**—negative words (e.g., loss, bankruptcy, indebtedness, felony, misstated, discontinued, expire, unable); N=2,349
- **Fin-Pos**—positive words (e.g., beneficial, excellent, innovative); N = 354
- **Fin-Unc**—uncertainty words (e.g., ambiguity, approximate, assume, risk); N = 291
- **Fin-Lit**—litigious words (e.g., admission, breach, defendant, plaintiff, remand, testimony); N = 871
- **Modal Strong**—e.g., always, best, definitely, highest, lowest, will; N = 19
- **Modal Weak**—e.g., could, depending, may, possibly, sometimes; N = 27

Parsing & Fogginess

"Words versus Deeds in Corporate Governance:
Event Study of Key Financial Laws
on Corporate Reporting"

Financial Misreporting

- SEC-Edgar database, advanced Boolean search on ∼1 million filings
- Filtered 8,401 "Codes of Conduct" and "Codes of Ethics" from Exhibit 14 in 10-Q, 10-K, and 8-Q
- Company identification by CIK and SIC
- Merged with financial data from Compustat
- H: financial firms adapt "codes," but not actual conduct

Testing (1)—Stemming, Parsing & Comparison

1. Stemming and parsing key financial laws—Sarbanes-Oxley, Basel II, Dodd-Franck—into three-word strings
2. Stemming and parsing "Codes of Conducts" into three-word strings
3. Dummy count of matchings between set (1) and (2)
4. Diff-in-diff of CoC changes before and after key financial laws; control group: firms not affected by SOX, B2, and D-F (non-financial and small financial firms)
5. Regress managerial reporting discretion—e.g., discretionary accruals and off-balance items—on matching count

Testing (2)—Gunning Fog Index

1. Developed by U.S. businessman Robert Gunning
2. Measures readability of English writing
3. Proxies years of formal education needed to understand text on a first reading; e.g., fog index $= 12 \rightarrow$ U.S. high-school senior ($\sim$18 years old)

$$\text{Fog index} = 0.4 \times \left[\underbrace{\left(\frac{\text{words}}{\text{sentences}} \right)}_{\text{avg. sentence length}} + \underbrace{\left(\frac{\text{complex words}}{\text{words}} \right)}_{\text{perc. of complex words}} \right]$$

Complex word: three or more syllables

4. Diff-in-diff of CoC "fogginess" before and after key financial laws
5. Regress discretionary accruals and off-balance items on fogginess

Applications

- Laws & court decisions
- Political speeches
- Contracts
- Company filings & disclosures (e.g., SEC-Edgar, MD&A)
- Audio transcripts (e.g., conference calls)
- News & media reports (e.g., Factiva, WSJ News Archive)
- Web sites, Google searches
- ... and social media (e.g., Twitter, Stocktwits, FB comments)

Textual Analysis of Large Libraries

- Big data is the future..., but we need tools
- Novel datasets from any set of documents—text, web, PDF
- Python is the way to go

☞ mmoszoro@berkeley.edu