



The Language of Rules: Textual Complexity in Banking Reforms

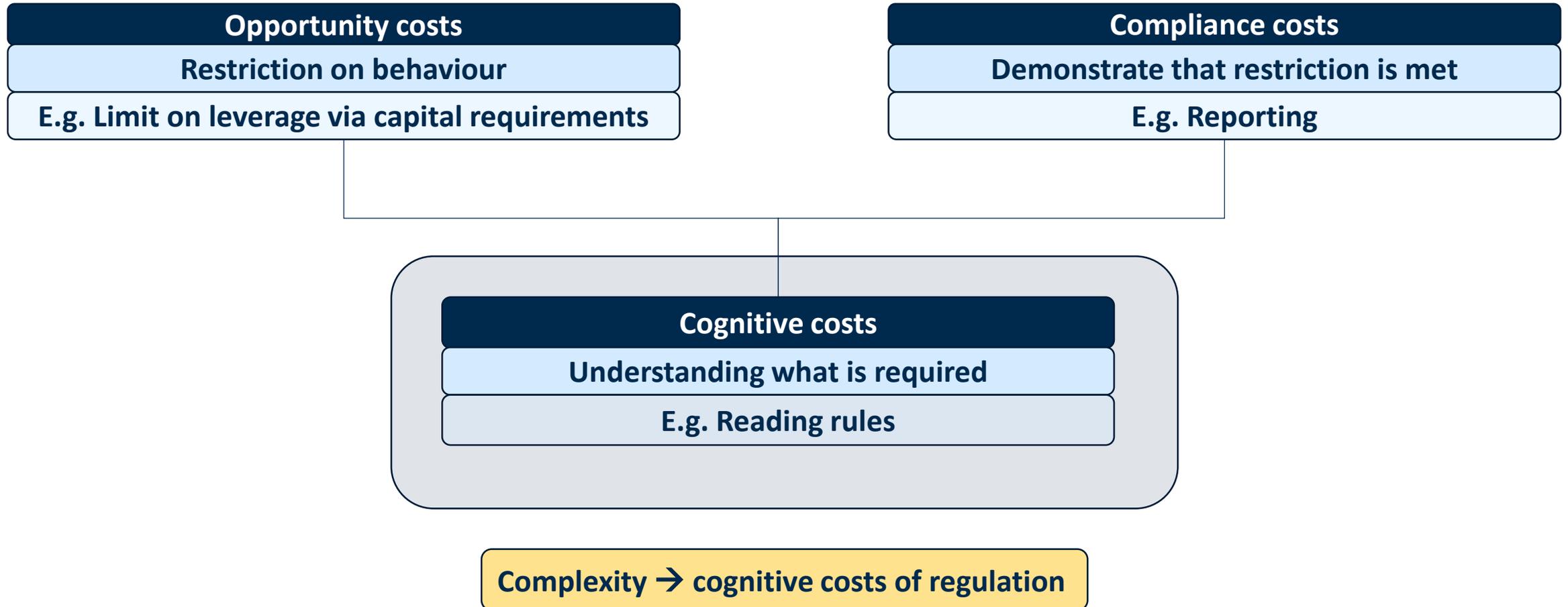
Zahid Amadxarif, James Brookes, Nicola Garbarino, Rajan Patel and **Eryk Walczak**
Bank of England

ESCoE EM 2020

17 September 2020

The views in this presentation are the authors' and do not necessarily reflect those of the Bank of England or its policy committees.

Costs of regulation



Why textual analysis of rules?

1. Data

Rules are embedded in texts:

- Data on regulatory costs are scarce, but texts are readily available
- Natural Language Processing—analyse large texts as data

2. Measures

Measures of cognitive costs of processing texts grounded in:

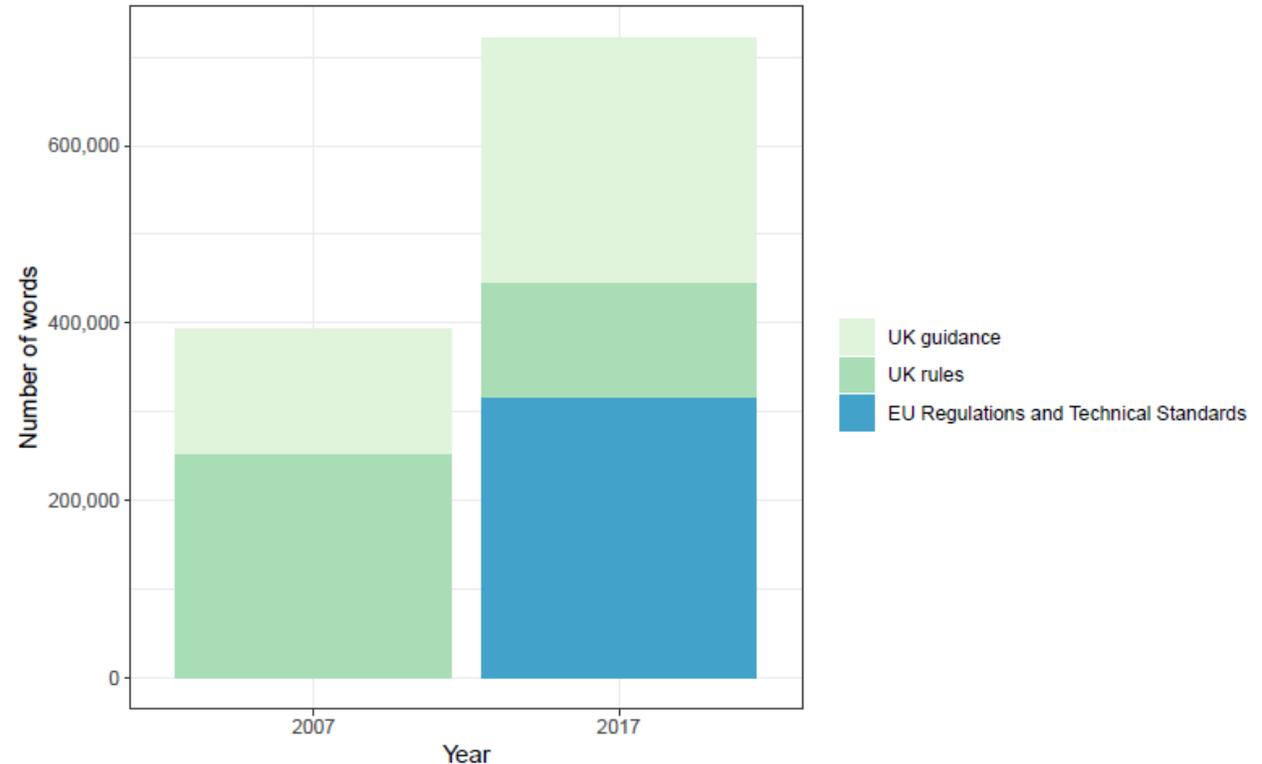
- Psycholinguistics: how people process language (e.g. concepts, operations)
- Computational linguistics: measures to capture language processing costs

Textual complexity ↔ cognitive costs

This paper: Post-crisis change in textual complexity

Between 2007 and 2017, for rules applying to UK banks:

- **Volume** (rules and guidance): increases from 400k to 700k words
- **Sources**: some EU rules become directly binding (CRR, Technical Standards)



Note:

UK guidance: PRA Supervisory Statements (2017), guidance in FSA Handbook (prudential guidance only, 2007).

UK rules (including rules implementing EU Directives): PRA Rulebook (2017), FSA Handbook (prudential rules only, 2007).

EU Regulations and Technical Standards: Capital Requirements Regulation (2013) and related EBA Technical Standards.

Results: four facts on textual complexity of post-crisis reforms

1. A tighter core emerges in the **network** of cross-references, centred around CRR
2. **Legal style** limits complexity of language in individual rules
3. At least 1/3 of rules contain vague terms (e.g. “adequate”) that require substantial **interpretation**
4. We **validate** our measures using EBA Q&A and a case study on definition of capital

Agenda

1. Data

2. Measures

3. Post-crisis reforms

4. Machine-readable rules

5. Conclusions

Objectives for the dataset

- 1. Comprehensive for post-crisis (2017).** Capture all: 1) prudential regulation that applies directly to UK banks; and 2) supporting guidance.
- 2. Comparable pre- and post-crisis (2007 v. 2017).** Capture 'equivalent scopes' of regulation to identify post-crisis changes robustly.
- 3. Facilitate network analysis.** Capture the entire structure of cross-references within the regulatory framework.

Data sources

- Web scraping
- Text extraction from pdf documents
- Link extraction
- Mostly automated
- Code: <https://github.com/bank-of-england/PRArulebook>



BANK OF ENGLAND
PRUDENTIAL REGULATION
AUTHORITY

View Rulebook as at:

16/11/2017  

You are viewing the Rulebook in the past on 16/11/2017

Allocation of Responsibilities - Application and Definitions

Application provision

1.1 Unless otherwise stated, this Part applies to every *firm* that is: 

ring-fenced body prescribed responsibility

means the responsibility in **4.2(4)**.

senior management regime

means the requirements of the *regulatory system* which apply to *relevant authorised persons* insofar as they relate to *approved persons* performing *PRA senior management functions* and *FCA designated senior management functions*, including those set out in **Senior Management Functions**, and **Allocation of Responsibilities and Fitness and Propriety**.

small CRR firm

Path	Capital Requirements Regulation > PART ONE > TITLE II > CHAPTER 2 > Section 3 > Article 20 (Copy link to article)
Title	Article 20
Description	Joint decisions on prudential requirements
Main content	<p>1. The competent authorities shall work together, in full consultation:</p> <p>(a) in the case of applications for the permissions referred to in Article 143(1), Article 151(4) and (9), Article 283, Article 312(2) and Article 363 respectively submitted by an EU parent institution and its subsidiaries, or jointly by the subsidiaries of an EU parent financial holding company or EU parent mixed financial holding company, to decide whether or not to grant the permission sought and to determine the terms and conditions, if any, to which such permission should be subject;</p>

Agenda

1. Data

2. Measures

3. Post-crisis reforms

4. Machine readable rules

5. Conclusions

Our measures of complexity are derived from linguistics ...

LINGUISTIC MEASURES

Measure	Calculated as...	Cognitive costs...
Lexical diversity	Relative frequency of unique words	Concepts
Conditionality	Relative frequency of conditional statements (e.g. “if”, “but”)	Operations
Length	Number of words	All of the above

These measures capture “local” complexity, i.e. cognitive costs incurred while reading a rule.

... network science ...

STRUCTURAL MEASURES (NETWORK)

Measure	Calculated as...	Cognitive costs...
Degree (In)	Number of direct cross-references TO a rule	Centrality
PageRank	Chain of cross-references leading TO a rule, weighted by their importance	Centrality
Degree (Out)	Number of direct cross-references FROM a rule	Context
Reverse PageRank	Chain of cross-references starting FROM a rule, weighted by their importance	Context



... and law.

VAGUENESS/PRECISION

Measure	Calculated as...	Cognitive costs...
Vagueness	Ratio of very vague terms (e.g. “reasonable”, “adequate”)	Need for interpretation
Precision	Ratio of very precise numerical signs / words (e.g. GBP, %)	“Bright-line” thresholds

Vagueness captures a different type of “global” complexity, i.e. additional contextual information that is not available in the rulebook, but may refer to legal precedents or market practices.

Are complex rules more likely to get requests for clarification?

List of Q&As

Final Q&As **1818**

Rejected Q&As **48**

All **1866**

Export results ▾

Showing 1 - 20 of 1818 results

1 2 3 4 5 6 7 8 9 ... NEXT › LAST » 

Clarification on what is considered insured deposits 

When determining the covered deposits, should netting with customer loans be applied before determining the amount of deposits which are not covered by the DGS?

COM Delegated or Implementing Acts/RTS/ITS/GLs: Not applicable

ID: 2015_2324 | Topic: Rights of depositors (Eligibility, coverage level, repayable amount, repayment, claim against DGS, depositor information) | Date of submission: 25/09/2015

Risk weights for the core indicator set 

EBA/GL/2015/10 Article 58, bullet point 2 suggests that there is no flexibility of the risk weights when the core risk indicator set are applied (with no additional indicators or any indicators left out). Would it be in compliance with the guidelines to use the core indicator set but to distribute the 25 % flexible weights somewhat different than what article 58, bullet point 2 suggests?

COM Delegated or Implementing Acts/RTS/ITS/GLs: EBA/GL/2015/10 - Guidelines on methods for calculating contributions to deposit guarantee schemes

ID: 2017_3096 | Topic: Funding provisions (financing means, use of funds, borrowing, calculation of contributions) | Date of submission: 12/01/2017

Definition of payee for dynamic linking 

Article 5 of the RTS on strong customer authentication and secure communication requires the authentication code to be specific to the amount of the payment transaction and the payee. Does it suffice to include a meaningful part of the identifier into the calculation of the authentication code? For instance, would it suffice to include only numeric characters of the IBAN in the calculation of the authentication code?

Banks (and others) can submit clarifying questions to the EBA

Are complex rules more likely to get requests for clarification? (2)

- We expect banks to ask more questions on rules that are more costly
- We run a (ridge) logistic regression model:

$$\text{logit}(\Pr[Y_i = 1]) = X_i + \gamma_t + \varepsilon_{it}$$

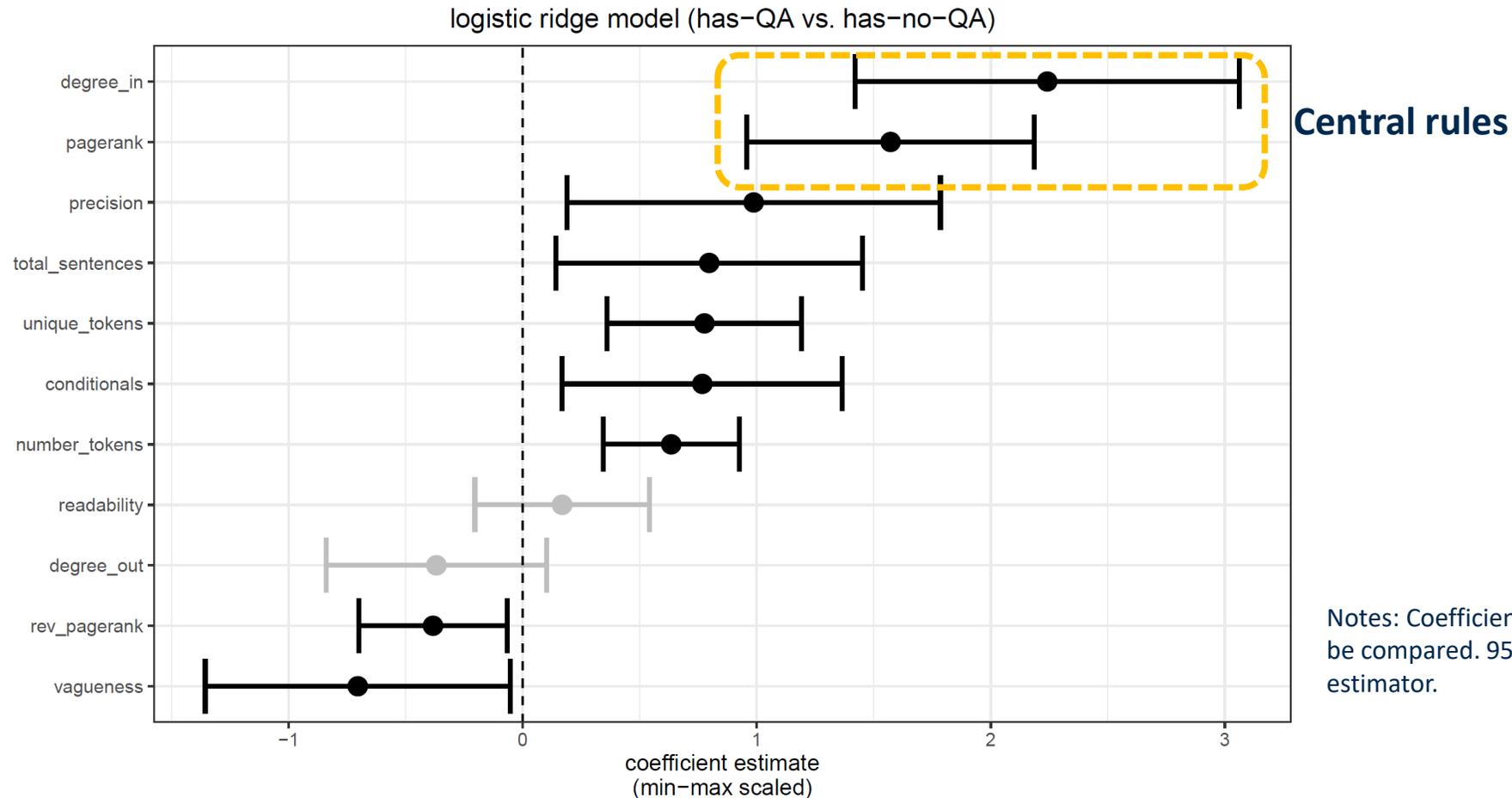
Y_i : dummy = 1 if rule i has a Q&A attached

X_i : vector of complexity measures for rule i

γ_t : dummy for topic t (e.g. capital), to control for non-cognitive costs

- Sample: 1,450 Q&As related to 261 (out of 519) CRR articles

Are complex rules more likely to get requests for clarification? (3)



Q&As focus on rules that are central, long, precise and contain multiple concepts and operations.

Agenda

1. Data

2. Measures

3. Post-crisis reforms

4. Machine-readable rules

5. Conclusions

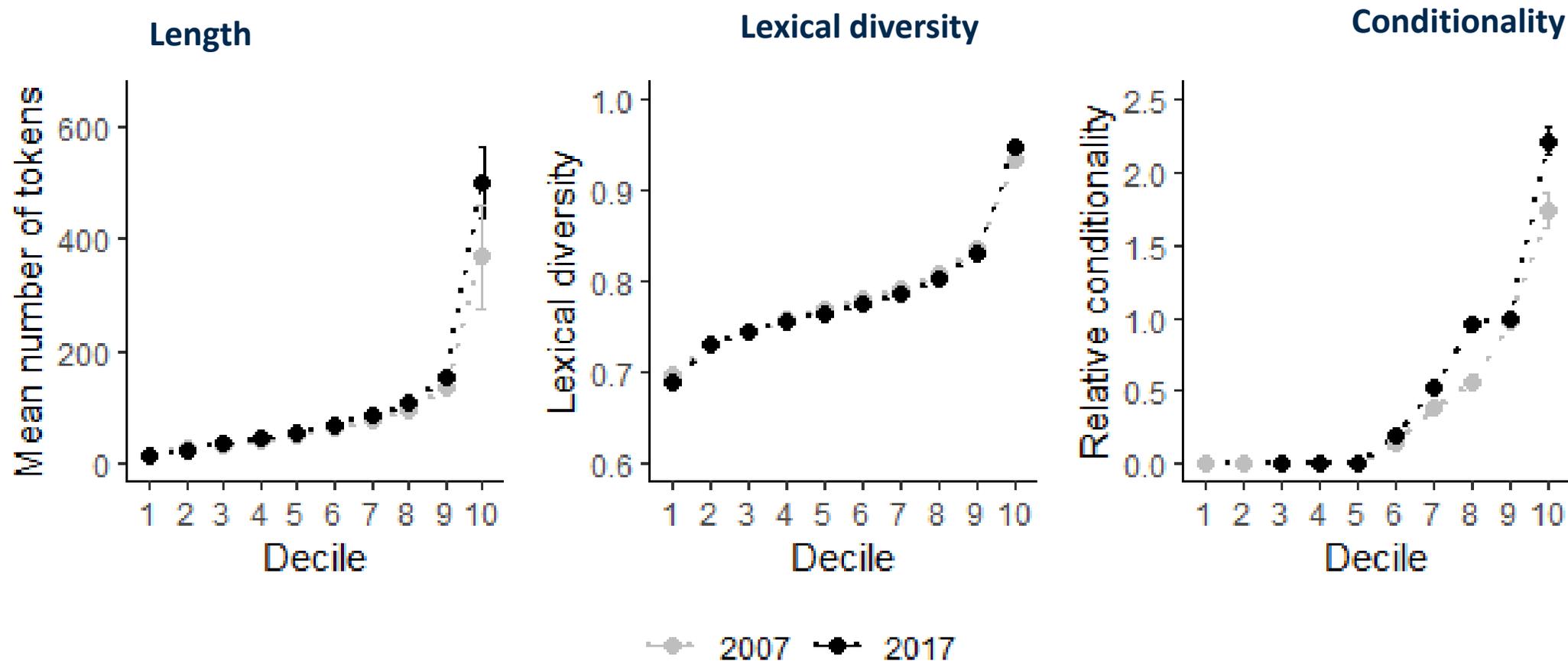
2007 vs 2017: All rules and guidance

	2007	2017	Change %
Panel A: Number of words			
<i>Total</i>	393,290	721,642	83.5%
Rules	221,912	445,710	100.8%
Guidance	171,378	275,932	61.0%
Panel B: Unique words			
<i>Total</i>	11,799	13,420	13.7%
Rules	8,657	9,671	11.7%
Guidance	7,558	8,765	16.0%
Panel C: Conditional words			
<i>Total</i>	3,930	6,320	60.8%
Rules	2,250	3,802	69.0%
Guidance	1,680	2,518	49.9%
Panel D: Provision level network (rules only)			
Nodes (Provisions)	2,440	3,961	62.3%
Edges (Cross-references)	2,569	4,289	67.0%

Note: Does not include titles and footnotes.

Post-reform regulation is longer, contains more concepts and more operations. It contains more rules, and more cross-references.

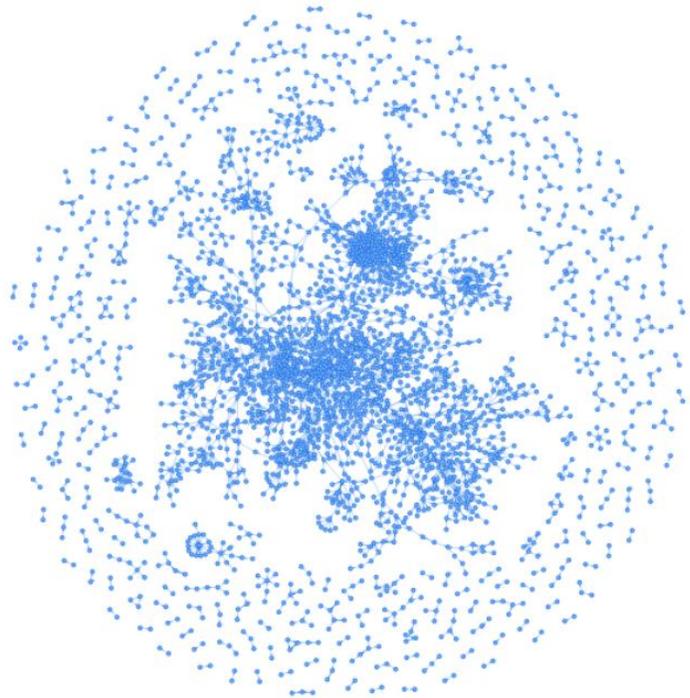
2007 vs 2017: Linguistic measures for individual rules



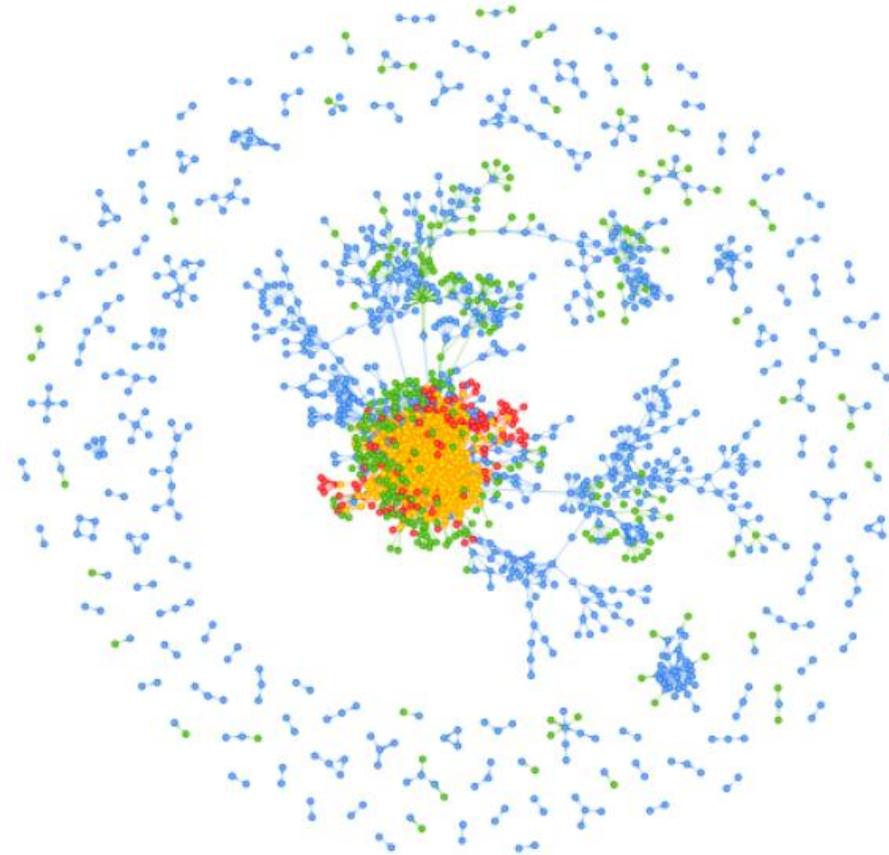
“Legal style” ensure that linguistic complexity is contained in small pockets.

2007 vs 2017: Network of cross-references

2007



2017

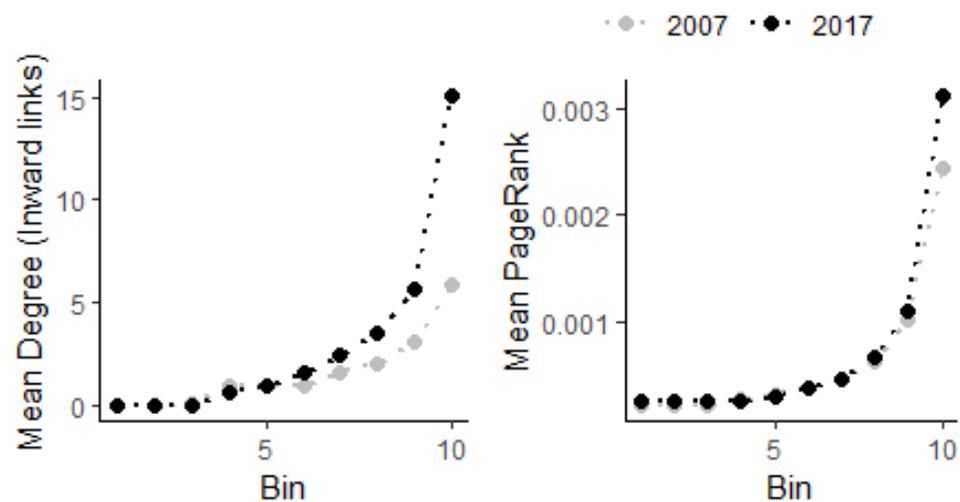


CRR
Technical Standards
Rulebook
Supervisory Statements

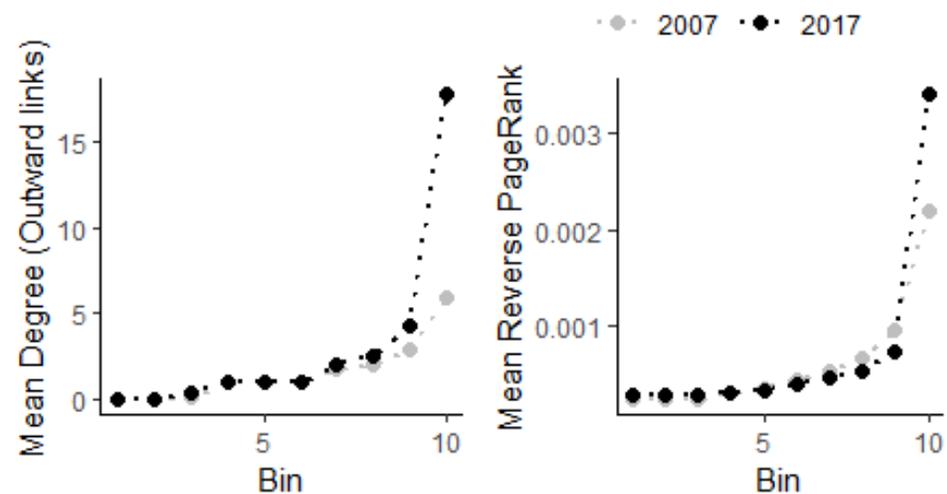
The network structure has a tightly connected core around CRR (but also a larger isolated periphery).

2007 vs 2017: network centrality and context

Network centrality



Network context



... resulting in highly interconnected cores post-crisis.

Agenda

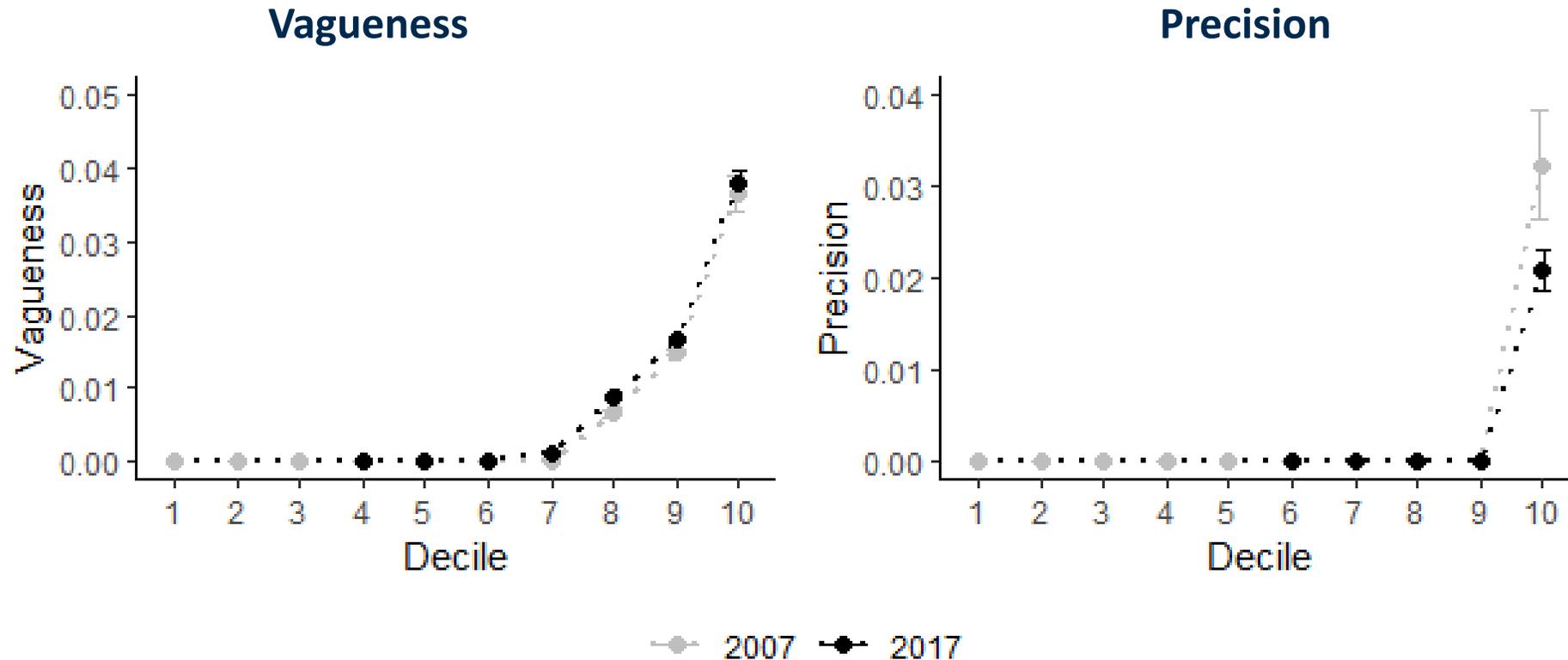
1. Data
2. Measures
3. Post-crisis reforms
4. Machine-readable rules
5. Conclusions

Rule simplification vs regulatory technology

- Traditional solution to regulatory complexity is **simplification** (Epstein, 1997; Sunstein, 2013)
- Some argue that RegTech means that simplification is no longer necessary (Casey, 2016)
- But law is **interpretation** (Dworkin, 1982) ...
- ... and supervision requires **discretion** (Black, 1997).
- Non-digitized, contextual information → difficult for machines.

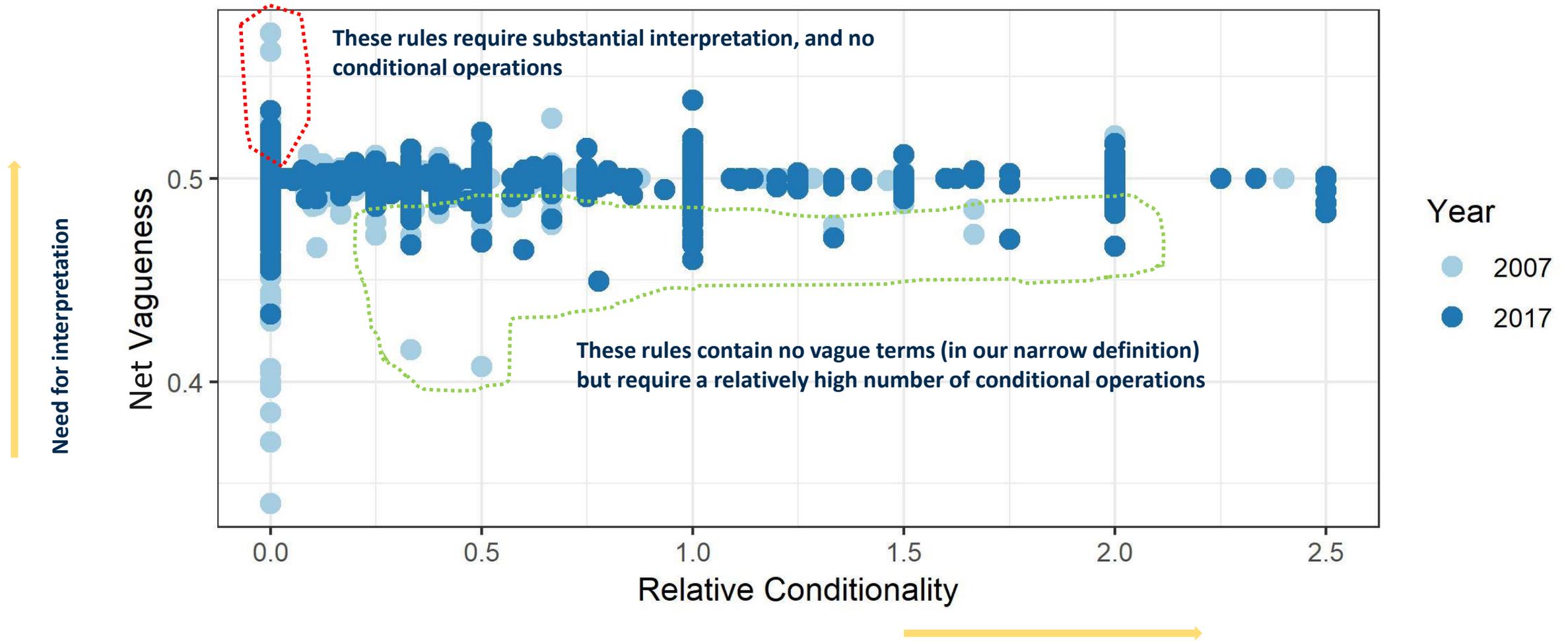
→ Concepts with **vague** (e.g. “adequate”) vs **specific** (e.g. “EUR”) interpretation

Distribution of vague and precise terms



Vague terms are common, specific numerical values are not (and in relative decline post-crisis)

Which rules can be successfully automated?



Net Vagueness = $0.5 + (\text{vague terms} - \text{specific terms}) / \text{number of words}$
Relative conditionality = $\text{conditional terms} / \text{number of sentences}$

Benefits of machine-readable rules (cognitive costs for humans)

Agenda

1. Data
2. Measures
3. Post-crisis reforms
4. Machine-readable rules
5. Conclusions

Conclusions

1. Using textual analysis to measure post-crisis increase in complexity:
 - Larger and more interconnected ...
 - ... but language of individual rules did not become more complex.
2. Approach to identify where machine-readable rules could succeed
3. Creation of a new data set



Thank you

Email: erykwalczak@gmail.com

Twitter: [@eryk_walczak](https://twitter.com/eryk_walczak)

Working Paper:

Zahid Amadjarif, James Brookes, Nicola Garbarino, Rajan Patel and Eryk Walczak (2019). [The language of rules: textual complexity in banking reforms](#). *Bank of England Staff Working Paper No. 834*.