

# Measuring Corporate Culture Using Machine Learning

Kai Li, Feng Mai, Rui Shen, Xinyan Yan  
Review of Financial Studies, 2020.7

吕漫妮

2020. 10. 3

# Contents

- Data, Preprocessing and Parsing, and Learning Phrases
- Word Embedding, word2vec, and Model Training
- Measuring Corporate Culture Using word2vec
- Validating Our Measure of Corporate Culture
- Implications of Corporate Culture
- Conclusion

# Backgrounds & Motivation

- Corporate culture matters because employees will inevitably face choices that cannot be properly regulated ex ante.
- Ample extant literature about culture - performance link is focus on survey/interview evidence, but has limited large sample evidence.
- The notion of corporate culture is somewhat nebulous and thus raises numerous measurement issues in empirical research.

# Research Problem

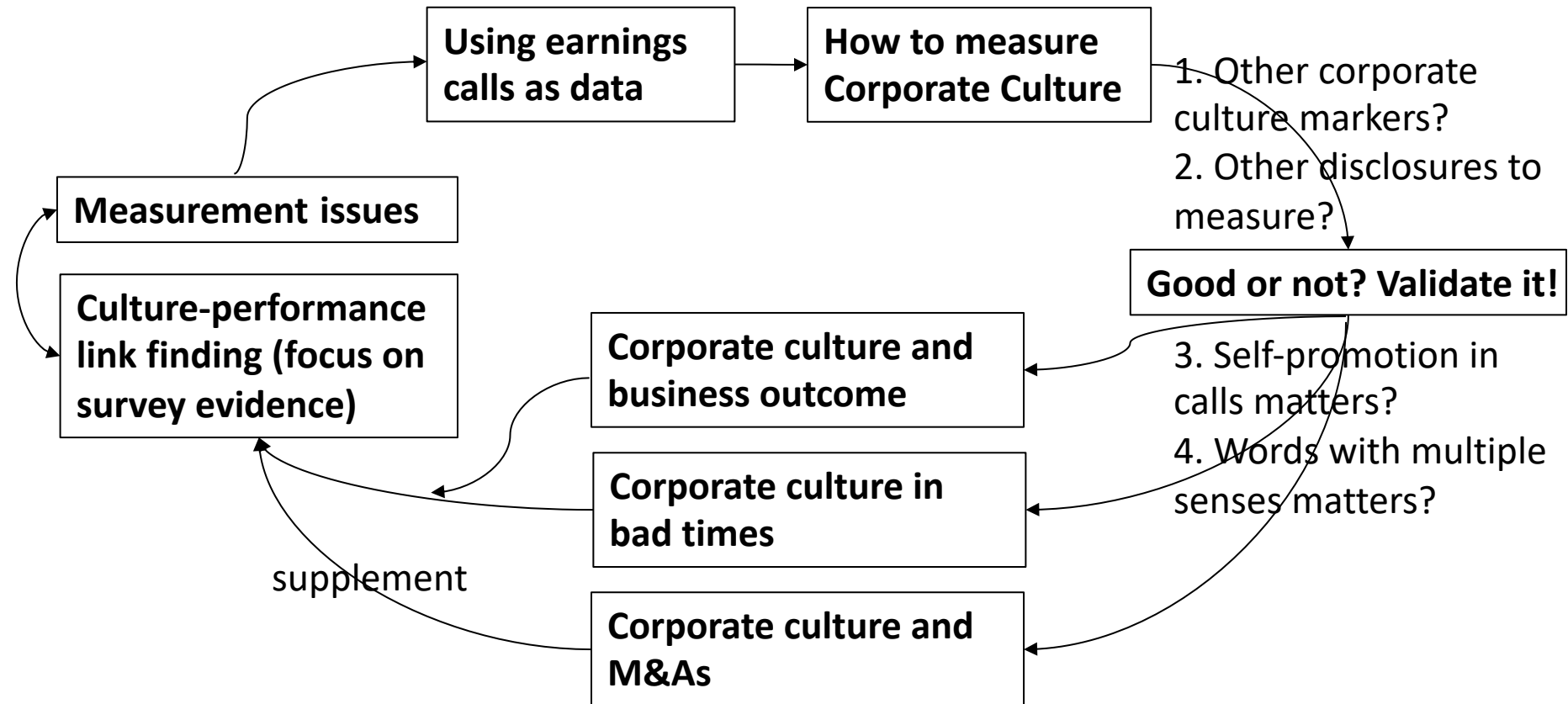
- Can we propose a more appropriate approach to measure corporate culture with large sample evidence?
  - We propose a semi-supervised machine learning approach to measuring corporate culture.
- If yes, can we validate some existing findings base on our new measure?
  - We show that corporate culture correlates with all aspects of business outcomes and the culture-performance link is more pronounced in bad times.
- Can we find out some other evidence about corporate culture?
  - We show that corporate culture plays an important role in deal incidence and merger pairing, and corporate culture is shaped by major corporate events.

# Contribution

- Our paper puts forward a new measure of corporate culture which is easily applied to a large sample of firms over time, extending earlier literature on corporate culture.
- We introduce the word embedding model to quantify meaning of expressions in finance and accounting text, provides an effective way to quantify the semantics, rather than merely the syntactic, at the expression level.
- Our paper is the first to apply neural network language models to the analysis of qualitative corporate disclosures.

# Outline

1. Preprocessing and Parsing
2. Model train
3. Score



# Data, Preprocessing and Parsing, and Learning Phrases

- Use what to score corporate culture? Earnings calls! QA section!

① CEO attributes and behaviors matter to corporate culture.

Earnings calls is a communication channel involving mostly CEOs.

② QA section is more spontaneous and offers far less of an opportunity for managers to engage in window dressing during calls.

- Data: Thomson Reuters' StreetEvents (SE) database to obtain the earnings call data set for the period January 1, 2001, to May 25, 2018, consisting of 209,480 QA sections that can be mapped to 64,511 firm-year observations

# Data, Preprocessing and Parsing, and Learning Phrases

- Preprocess and parse: Stanford CoreNLP package

Sentence segmentation and tokenization → Lemmatization → Named Entity Recognition (replace named entities with a predefined tag)

- Learning Phrases: Phraser module of the gensim library

→ Identify fixed multiword expressions (e.g., with respect to, rather than) and compound words (e.g., intellectual property, healthcare provider)

→ hold those terms that are specific to our corpus, concatenate them using the symbol “\_” and treat them as a single word



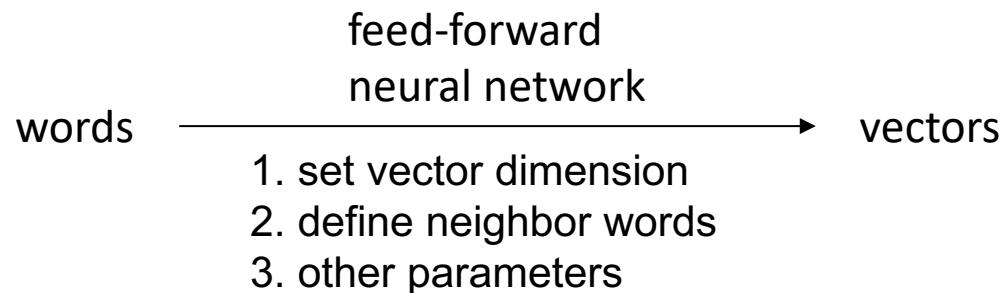
# Word Embedding, word2vec, and Model Training

- Why not create dictionary directly?
  - ① Corporate culture is often discussed in a subtle fashion (abbr., phrases, or idioms that make sense only in a particular context).
  - ② And it can be a multidimensional construct (difficult to categorize).
  - ③ Words enter and drop out of business vocabulary as technologies and industries evolve (e.g. artificial intelligence).
- Why word embedding?

Word embedding starts with seed words that define each cultural value and automatically creates a high-quality dictionary from qualitative corporate disclosures adaptively.

# Word Embedding, word2vec, and Model Training

- What is word embedding?
  - Co-occur with same neighboring words have similar meanings
    - Identifies synonyms from common neighboring words (construct a count vector that tallies the number of times other words appear near, and use the cosine similarity of vectors to describe words' relationship)
- Possible neighboring words is enormous! How? *Word2vec*!!!
  - *Word2vec* employs a neural network to efficiently learn dense and low-dimensional vectors that can represent the meaning of words.



# Measuring Corporate Culture Using word2vec

- | <i>Culture value</i> | <i>Seed words</i> |
|----------------------|-------------------|
| Innovation           | creativity        |
|                      | efficiency        |
|                      | excellence        |
|                      | innovation        |
- Seed words: Guiso, Sapienza, and Zingales (2015) provide seed words for innovation, integrity, quality, respect, teamwork—five most-often mentioned values by S&P 500 firms on their websites.
    - .Applicable? Manually adjust!
  - Generating the dictionary with trained model
    - ① Compute the mean of the vectors of seed words for a certain value.
 
$$\overline{V}^{\{teamwork\}} = \frac{1}{7} \sum_{i=1}^7 \left[ x_1^{\{i\}}, x_2^{\{i\}}, \dots, x_{300}^{\{i\}} \right]$$
    - ② Compute the cosine similarity between each unique word with it.
    - ③ Select top-500 words with the closest associations as the dictionary
    - ④ Manually inspect

### C. The correlation matrix

	Innovation	Integrity	Quality	Respect	Teamwork
Innovation	1.000				
Integrity	0.109***	1.000			
Quality	0.490***	0.023***	1.000		
Respect	0.321***	0.269***	0.317***	1.000	
Teamwork	0.371***	0.276***	0.271***	0.258***	1.000
Firm size	-0.186***	-0.010**	-0.261***	-0.255***	-0.309***
Leverage	-0.282***	0.024	-0.276***	-0.170***	-0.199***
ROA	-0.105***	-0.130***	-0.069***	-0.093***	-0.352***
Sales growth	0.008*	-0.047***	0.017***	0.033***	-0.025***
Top-5 institutions	0.059***	-0.096***	0.018***	0.033***	-0.081***

#### ③ Correlations:

Firm size, leverage and Operating performance (ROA) are negatively and significantly associated with almost all five cultural values.

→ Having a strong culture calls for investment, resulting in forgone short-term profit.

# Validating Our Measure of Corporate Culture

- Corporate culture and its markers
  - ① Innovation →  $\ln(\text{Patent})$ , R&D spending, innovation strength
  - ② Integrity → restatement (if a firm restates its financial statements), backdating (if option grants to a firm's CEO are backdated)
  - ③ quality → product quality, product safety, top brand (standards)
  - ④ respect → diversity and “best employer” status
  - ⑤ teamwork → the level of employee involvement, the number of joint ventures (JVs) and strategic alliances (SAs)

There remains expected associations between the cultural value and any of its markers, after controlling for size, ROA, industry and year fixed effects and all four other cultural values.

# Validating Our Measure of Corporate Culture

- What about other ways of measuring corporate culture?

① our dictionary to full transcript of calls → “\_full”

② seed words (plus value word) to the QA section of calls → “\_seed”

③ our dictionary to the MD&A section of 10-Ks → “\_10k”

## *D. Validating alternative measures for respect*

Variable	Diversity (1)	Best employer (2)	Diversity (3)	Best employer (4)	Diversity (5)	Best employer (6)
Respect_full	0.098*** (0.035)	0.200* (0.106)				
Respect_seed			0.389 (0.296)	2.461*** (0.690)		
Respect_10k					0.211*** (0.054)	0.275** (0.136)
Size/ROA	Yes	Yes	Yes	Yes	Yes	Yes
Ind FE/yr FE	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	19,385	52,930	19,385	52,930	16,679	37,719
$R^2$ /pseudo $R^2$	.307	.189	.306	.192	.293	.219
Incremental $R^2$	.0008		.0001		.0026	
Incremental pseudo $R^2$		.0030		.0068		.0037

Main measure outperforms!

① > ③ > ②

QA section of calls is the better alternative for measuring corporate culture!  
Good dictionary!

# Validating Our Measure of Corporate Culture

- Excessive self-promotion during calls make our approach does not fully capture corporate culture?
  - remove the top quartile of paragraphs with the greatest emotion score
- The model combine multiple meanings of a word into a single vector. Does it matter?
  - use an algorithm developed by Pelevina et al. (2016) to identify the words with multiple senses and remove them

The correlation between the measures with and without adjustment is high( $>0.86$ ), suggesting the limited significance.

# Implications of Corporate Culture

- Corporate culture and business outcomes

## 1. The implications of having a strong culture on business outcomes

	Assets turnover (1)	Inventory turnover (2)	Stock return volatility (3)	Discretionary accrual (4)	ln(Delta) (5)	ln(Vega) (6)	CEO pay duration (7)	Tobin's q (8)
Strong culture <sub>(t-1)</sub>	0.052*** (0.014)	6.741*** (1.582)	0.005*** (0.001)	-0.917** (0.404)	0.080** (0.040)	0.202*** (0.045)	1.037*** (0.349)	0.043*** (0.009)
Firm-level controls	operational efficiency		earnings management		CEO wealth-performance sensitivity/sensitivity of CEO wealth to stock volatility/CEO pay duration		firm value	
Ind FE/yr FE								
Intercept								
Obs.	48,763	55,197	48,614	59,200				48,750
R <sup>2</sup>	.445	.054	.429	.007				.687

- ① Consistent with interview evidence in Graham et al. (2018, 2019) and other intuition.
- ② The strong association remains even use different lead-lag specifications up to 5 years apart.



# Implications of Corporate Culture

- Link is more apparent in bad times!

We focus on the effects of the financial crisis on companies in the financial industry and on the effects of British Petroleum's (BP)

Deepwater Horizon oil spill on oil companies

## *B. The implications of having a strong culture in bad times*

	Abnormal return (1)	Abnormal return (2)	Abnormal return (3)	Abnormal return (4)
Strong culture	-0.012*** (0.003)	-0.004 (0.004)	-0.006* (0.003)	0.000 (0.007)
Strong culture × Financial crisis	0.028*** (0.005)	0.024*** (0.005)		
Strong culture × BP oil spill			0.017*** (0.005)	0.018*** (0.005)

Firms with a strong culture are associated with a 2.4%/1.8% higher monthly return during the crisis/oil spill period than their counterparts without a strong culture.

# Implications of Corporate Culture

- Corporate culture and M&As

Cultural fit hypothesis: Fewer deals between firms with conflicting corporate cultures. Differences in corporate cultures of firm-pairs are a key determinant of deal incidence.

Acculturation hypothesis: Merging firms with different cultures will develop a jointly determined culture.

## ① How to measure cultural fit and conflict?

Cultural similarity: the cosine similarity between two five-by-one vectors capturing the cultural values of a firm-pair

Cultural distance: the square root of the sum of squared differences between a firm-pair across all five cultural values

# Implications of Corporate Culture

② Firm's cultural values and its probability of being an acquirer

Sample: all U.S. deals completed from January 1, 2003, to December 31, 2018, reported in Thomson Reuters' SDC database

Variable	Full sample	Industry and size-matched		Industry, size, and B/M-matched	
	LPM (1)	LPM (2)	Clogit (3)	LPM (4)	Clogit (5)
Innovation	0.004** (0.002)	0.005 (0.003)	0.028 (0.020)	0.007** (0.003)	0.044** (0.020)
Integrity	-0.045*** (0.005)	-0.053*** (0.008)	-0.330*** (0.047)	-0.059*** (0.008)	-0.368*** (0.047)
Quality	-0.008*** (0.003)	-0.014*** (0.004)	-0.083*** (0.024)	-0.014*** (0.004)	-0.084*** (0.025)
Respect	0.015*** (0.002)	0.032*** (0.004)	0.183*** (0.022)	0.033*** (0.004)	0.191*** (0.023)
Teamwork	-0.000 (0.003)	-0.001 (0.005)	-0.010 (0.032)	-0.003 (0.005)	-0.021 (0.033)

(The dependent variable is equal to one for the acquirer)

Firms scoring high on innovation and respect are more likely to be acquirers, whereas ones scoring high on integrity and quality are less.

# Implications of Corporate Culture

## ③ Cultural fit and merger pairing

Variable	Industry and size-matched		Industry, size, and B/M-matched	
	Clogit (1)	Clogit (2)	Clogit (3)	Clogit (4)
Cultural similarity	4.305*** (0.902)		4.210*** (0.943)	
Cultural distance		-0.496*** (0.075)		-0.515*** (0.075)

(The dependent variable is equal to one for the acquirer-target pair)

Firms closer in cultural values are more likely to do a deal together, vice versa (Cultural fit hypothesis).

# Implications of Corporate Culture

## ③ Post-merger acculturation

Can the merger itself shape corporate culture?(Acculturation hypothesis)

Target-specific culture is obtained by regressing each target cultural value on corresponding acquirer cultural value in the year prior to deal announcement and keeping the residual

	Innovation <sub>t+1</sub>	Innovation <sub>t+3</sub>	Integrity <sub>t+1</sub>	Integrity <sub>t+3</sub>
	(1)	(2)	(3)	(4)
Acquirer innovation	0.854*** (0.039)	0.905*** (0.053)	0.030** (0.014)	0.042** (0.020)
Target-specific innovation	0.108*** (0.034)	0.108** (0.052)	0.003 (0.014)	0.022 (0.021)

The acquirer's cultural values are significantly related to both the acquirer's and the target's premerger cultural values. Mergers help acquirers create a new jointly determined culture.

# Conclusion

- We use a word embedding model to generate a culture dictionary then quantify earnings call transcripts to obtain scores for corporate culture.
- We conduct a large number of tests to validate our measure and demonstrate the advantages of our method.
- We show that corporate culture correlates with business outcomes, including operational efficiency, risk-taking, earnings management, executive compensation design, and firm value.
- The culture performance link is more pronounced in bad times.
- Corporate culture plays an important role in deal incidence and merger pairing, and can be shaped by major corporate events.

# Reflection

- a new semi-supervised machine learning approach for textual analysis
- how to measure