Risk exposures from risk disclosures: What they said and how they said it

Rahul Mazumder
MIT

Seth Pruitt
ASU

Landon J. Ross
Tulane

Management's disclosure of risks

Officially part of Form 10-K Section 1-A – most firms started writing in 2006

> *Provide any discussion of risk factors in plain English in accordance with Rule 421(d) of the Securities Act of 1933 (§230.421(d) of this chapter)*

- Are managers providing valuable information to investors?
- Can we extract new and economically useful data from 10-Ks?
- Can we simplify natural language processing without compromising results?

Questions

Does Section 1-A predict **future beta**?

- Betas are important to investors, cost of capital calculations, etc.
- Simple features of language might help structure map from text to betas.

Do different methods of *natural language processing* recover different information?

- What they said $\rightarrow$ **Topic** modeling
- How they said it $\rightarrow$ **Context** modeling

Summary

Topic and **context** are complementary:

- Statistically significant and distinct in forecasts.
- Certainty-equivalent value of 0.3% per annum.

Novel evidence

- Different NLP methods $\rightarrow$ Different data from text.
- For market beta, only **context** is important; for other betas, **both** are important

Results hold while controlling for other firm characteristics.

Main Idea

$$\beta_{i,t} = \delta_0 + \delta x_{i,t-1}$$

We forecast firm i beta for year t with information x from year $t-1$.

- Factor index is suppressed.

$\beta_{i,t}$ is from regression of year t daily stock returns on factor realizations:

- $r_{i,t}^{daily} = cons + \beta_{i,t} f_t^{daily} + error$

We estimate δs with first half of sample.

- We show forecast results for second half of sample.

Data

Standard Datasets

- Stock returns from CRSP. Daily factor returns from CRSP, Ken French, FRED.
- Firm characteristics from CRSP, Compustat, S&P Global (via Jensen, Kelly, Pedersen (2022 JF) codes).

Risk Disclosures

- Risk Disclosures from 10-K between 2006-2022. Downloaded from Edgar.
- Use hyperlinks in table of contents to extract most disclosures.
- Use page numbers and regular expressions for remaining disclosures.

Forecasts with Firm Characteristics

Lagged beta

- $\beta_{i,t} = cons + \delta\beta_{i,t-1} + error$
- Standard benchmark.

Lagged characteristics (use in robustness checks)

- $\beta_{i,t} = cons + \delta' x_{i,t-1} + error$
- x is vector of 36 characteristics from Kelly, Moskowitz, and Pruitt (2021).

Topic Modeling Setup

We use Latent Dirichlet allocation (LDA) to factor documents word counts, into a collection of latent topics.

LDA assumes the following specification generates a document's words from K latent topics:

1. Choose $N \sim \text{Poisson}(\xi)$.
2. Choose $\theta \sim \text{Dirichlet}(\alpha)$.
3. For each of the words w_n in the document:
   3.1 Choose a topic $z_n \sim \text{Multinomial}(\theta)$.
   3.2 Choose a word w_n from multinomial distribution $p(w_n|z_n, B)$.

Topic Modeling Intuition

LDA and topic modeling are similar to PCA.

- Latent topics $\approx$ latent factors
- θ for given document $\approx$ factor loadings.
- Caution: Topics do not have natural ordering like factors.

LDA and topic modeling are similar to dictionary methods.

- Topics $\approx$ Word lists.
- Fraction of document about topic $k \approx$ Normalized word counts.
- Caution: Topic word lists are data driven.

Topic Modeling Characteristics

The main object of interest for our study is $\theta_{i,t}$.

- $\theta_{i,t}$ is a K-dimensional vector.
- Component k of $\theta_{i,t}$ measures **how much** of a risk disclosure is about topic k.

Forecast and Interpretation

- $\beta_{i,t} = cons + \delta' \theta_{i,t-1} + error$
- Forecast posits relation between betas and **quantity** of text about each topic.

Context Model

$$\beta_{i,t} = cons + \sum_{w \in keywords} \delta'_w x^w_{i,t-1} + e_{i,t} \quad (1)$$

Keywords

- Words where context might contain information about betas.
- We use nouns occurring in at least 5% of documents. About 2,500.
- Intuition: nouns $\approx$ risk factors. So noun contexts should summarize betas.

Vector $x^w_{i,t-1}$

- $x^w_{i,t-1}$ is a <u>vector</u> measuring average context of word w in the firm i risk disclosure.
- We use seven dimensional vectors to represent $x^w_{i,t-1}$. Chosen via cross-validation.
- Beta forecasts include about 17,500 variables before regularization.

Context Model Intuition

Suppose we see the sentence below in a 10-K:

- "The Company uses derivative instruments, such as foreign currency forward and option contracts, to hedge certain exposures to fluctuations in foreign currency exchange rates."

How exchange rate fluctuations effect the company is right next to the risk itself!

- **"How"** is usually close to **"what."** This is a feature of well written English.

How do we systematically and quantitatively represent both **what** and its neighboring **how**?

- Keywords $\rightarrow$ **what** is discussed.
- Context $\rightarrow$ **how** the keyword is discussed.

Other Ingredients for Context Model

- Word Embeddings $\rightarrow$ Quantitative representation of context.
- Group Lasso $\rightarrow$ Select keywords where context explains betas.

Context Model Estimation

$$\min_{\delta_w} \frac{1}{2} \sum_{i,t} \left\| \beta_{i,t} - \left(cons + \sum_{w \in keywords} \delta_w' x_{i,t-1}^w \right) \right\|_2^2 + \lambda \sum_{w \in keywords} \|\delta_w\|_2 \qquad (2)$$

Group Lasso Regression Model

- Group lasso penalty selects words where context explains **how** keyword effects firm.
- One group = context vectors for one keyword.
- Interpretation: Only active keywords' context explains how keyword effects betas.
- λ penalty chosen via cross validation.

Context Model Vectors

$$x_{i,t}^{w} = \sum_{j \in \{-10,\ldots,10\}} h_{i,t}^{w+j} v^{w+j} \tag{3}$$

- $v_{i,t}^{w+j}$ is the embedding vector for the word at position j relative to word w.
- $h_{i,t}^{w+j}$ is weight of embedding vector $w+j$.
- If word w occurs more than once in a risk disclosure, then average.

Example

- "We sell WTI futures to reduce the volatility of our <u>oil</u> extraction revenues..."
- $x^{oil} = \frac{1}{21}\left(v^{we} + v^{sell} + v^{WTI} + v^{futures} + \ldots\right)$.

Word Embeddings

Goal

- Associate words with vectors that produce useful features for downstream tasks.
- Vectors are trained on unsupervised tasks, e.g. masked word prediction.

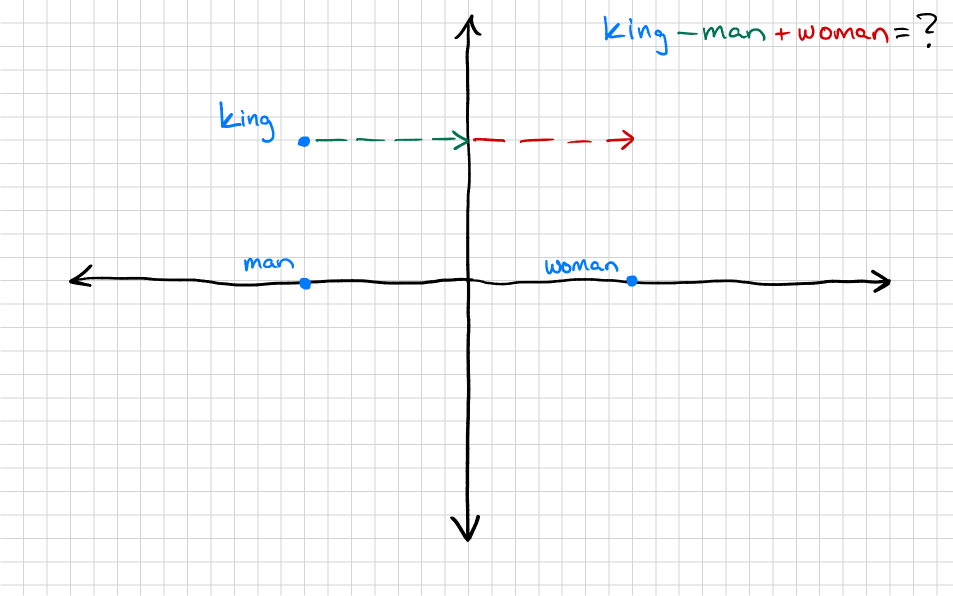
Examples

- Vector addition and subtraction can solve word analogies and classification questions.
- Inner products can express synonymy, similarity, and magnitude.

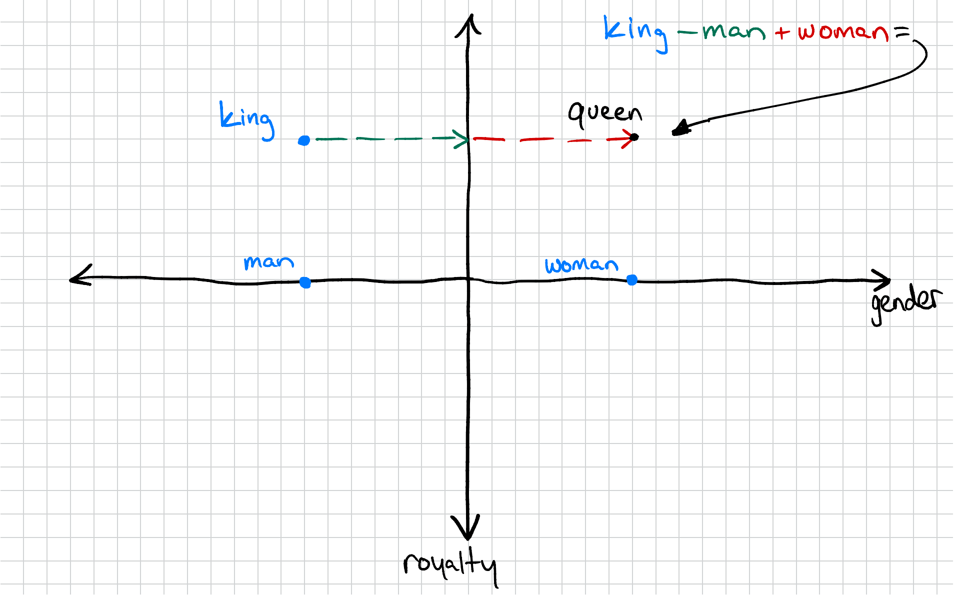
Usage

- We use FastText embedding trained on Wikipedia with post processing from Khodak *et al.* (2018) and additional dimensional reduction. More in paper.

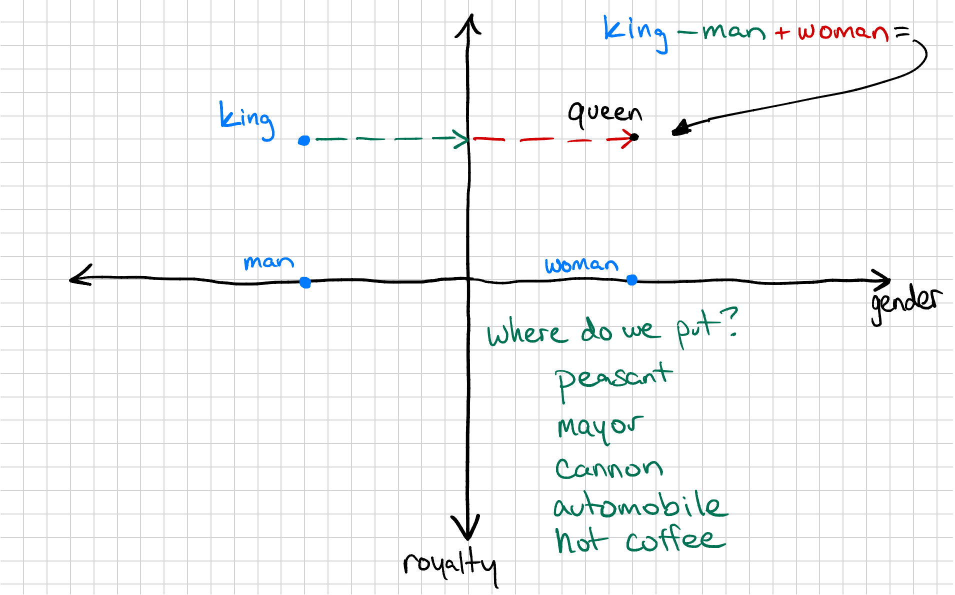
Word Embedding Example



Word Embedding Example



Word Embedding Example



Out-of-sample R^2

Factor	Lagged beta	Topic	Context
Panel A: Portfolios			
Market	−0.054	−0.034	0.072
Size	0.156	0.016	0.082
Value	0.294	0.151	0.159
Investment	0.140	0.292	0.320
Profitability	0.599	0.103	0.104
Momentum	−0.584	0.230	0.236
Panel B: Macroeconomic			
Exchange Rate	−0.169	0.112	0.098
Credit Spread	−0.304	0.015	0.028
Term Spread	−0.165	0.006	0.038

Out-of-sample R^2

Factor	Lagged beta	Topic	Context
Panel A: Portfolios			
Market	−0.054	−0.034	0.072
Size	0.156	0.016	0.082
Value	0.294	0.151	0.159
Investment	0.140	0.292	0.320
Profitability	0.599	0.103	0.104
Momentum	−0.584	0.230	0.236
Panel B: Macroeconomic			
Exchange Rate	−0.169	0.112	0.098
Credit Spread	−0.304	0.015	0.028
Term Spread	−0.165	0.006	0.038

- Topics do not forecast market beta
  Management's information about firm exposure to the market, an aggregation of everything going on, is not related to their choice of topics

Out-of-sample R^2

Factor	Lagged beta	Topic	Context
Panel A: Portfolios			
Market	-0.054	-0.034	0.072
Size	0.156	0.016	0.082
Value	0.294	0.151	0.159
Investment	0.140	0.292	0.320
Profitability	0.599	0.103	0.104
Momentum	-0.584	0.230	0.236
Panel B: Macroeconomic			
Exchange Rate	-0.169	0.112	0.098
Credit Spread	-0.304	0.015	0.028
Term Spread	-0.165	0.006	0.038

- Text does forecast market beta

  Context revealed the valuable information

Out-of-sample R^2

Factor	Lagged beta	Topic	Context
	Panel A: Portfolios		
Market	−0.054	−0.034	0.072
Size	0.156	0.016	0.082
Value	0.294	0.151	0.159
Investment	0.140	0.292	0.320
Profitability	0.599	0.103	0.104
Momentum	−0.584	0.230	0.236
	Panel B: Macroeconomic		
Exchange Rate	−0.169	0.112	0.098
Credit Spread	−0.304	0.015	0.028
Term Spread	−0.165	0.006	0.038

- Text is more valuable for future investment risk
  This exposure probably has the most to do with management's own investment decisions
  $\rightarrow$ both topics and context reveal this information

Out-of-sample R^2

Factor	Lagged beta	Topic	Context
Panel A: Portfolios			
Market	−0.054	−0.034	0.072
Size	0.156	0.016	0.082
Value	0.294	0.151	0.159
Investment	0.140	0.292	0.320
Profitability	0.599	0.103	0.104
Momentum	−0.584	0.230	0.236
Panel B: Macroeconomic			
Exchange Rate	−0.169	0.112	0.098
Credit Spread	−0.304	0.015	0.028
Term Spread	−0.165	0.006	0.038

- Macroeconomic (nontradable) risks exposures harder to forecast, particularly bond-spreads

Out-of-sample R^2

Factor	Lagged beta	Topic	Context
	Panel A: Portfolios		
Market	−0.054	−0.034	0.072
Size	0.156	0.016	0.082
Value	0.294	0.151	0.159
Investment	0.140	0.292	0.320
Profitability	0.599	0.103	0.104
Momentum	−0.584	0.230	0.236
	Panel B: Macroeconomic		
Exchange Rate	−0.169	0.112	0.098
Credit Spread	−0.304	0.015	0.028
Term Spread	−0.165	0.006	0.038

- Both context and topic predict exchange-rate exposure
  Like initial example; but note these are LDA-derived topics, and not one you might construct *a priori*

Forecast significance

What drives out what? What is redundant info?

- Forecast combination regressions—estimated on second half
  - No overfit concern: have 3 regressors at most
  - Usual inference (clustered std errs)
  - Holds constant the first-half-estimated parameters (to continue to be as out-of-sample as possible)

If topics and contexts are *both* statistically significant, they are uncovering *distinct, complementary* information from the same text

Forecast significance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: Market			
Context	0.903***			0.527***	0.781***		0.454***
	(34.03)			(23.26)	(23.90)		(17.93)
Lagged beta		0.542***		0.410***		0.474***	0.406***
		(41.85)		(28.72)		(34.82)	(28.24)
Topic			1.011***		0.284***	0.532***	0.179***
			(21.76)		(5.34)	(17.00)	(5.12)
intercept	0.0464	0.507***	-0.0852	0.0301	-0.147**	-0.0506	-0.0915**
	(1.47)	(31.25)	(-1.54)	(1.45)	(-2.98)	(-1.61)	(-2.89)
R^2	0.217	0.286	0.126	0.343	0.223	0.316	0.345

Optimal combination is about 45-40-15 context-beta-topic

Forecast significance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel G: Exchange Rate			
Context	0.595***			0.413***	0.398***		0.277***
	(23.27)			(19.99)	(9.81)		(8.69)
Lagged Beta		0.337***		0.293***		0.298***	0.290***
		(37.99)		(34.42)		(34.06)	(33.75)
Topic			0.734***		0.328***	0.506***	0.230***
			(22.28)		(6.72)	(19.20)	(5.86)
intercept	-0.00205***	-0.00757***	-0.000116	-0.00222***	-0.000230	-0.000889*	-0.000947**
	(-5.88)	(-74.01)	(-0.26)	(-8.14)	(-0.55)	(-2.57)	(-2.82)
R^2	0.064	0.119	0.058	0.148	0.069	0.145	0.150

Optimal combination is approximately equal-weighted

Forecast significance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel D: Investment			
Context	0.752***			0.424***	0.812***		0.471***
	(17.62)			(13.21)	(12.55)		(11.28)
Lagged beta		0.439***		0.381***		0.409***	0.381***
		(41.21)		(32.40)		(32.71)	(32.98)
Topic			0.821***		-0.112	0.426***	-0.0882
			(8.38)		(-0.84)	(6.47)	(-1.02)
intercept	-0.146***	-0.172***	-0.120***	-0.0485***	-0.165***	-0.0318*	-0.0638***
	(-10.81)	(-15.41)	(-4.59)	(-4.96)	(-6.44)	(-2.03)	(-3.92)
R^2	0.102	0.219	0.056	0.247	0.102	0.233	0.247

Context drives out topic

Discussion: Economic Models of Language

Question

What would an economic model of text look like?

- Statistical: LDA, HDP, etc.
- Economic: Text is function of management choices and information.

Example: Boilerplate Risk Disclosures

- Boilerplate risk disclosures sound like a pooling equilibria.
- Boilerplate text tells us about something about firm beyond text itself.

Model Sketch

- Manager observes private signal of market beta $\tilde{\beta} = \beta + \varepsilon$.
- Manager writes document $D = \begin{bmatrix} w^1 & \dots w^N \end{bmatrix}$
- Manager and investors have vocabulary V with where vector $v^w \in V$ represents word w.

Discussion: Economic Models of Language

Manager Choice Problem

$$\max_{w^1,\ldots,w^N} u(f(v^{w_1},\ldots v^{w_N}) - \lambda\mathcal{L}(g(v^{w_1},\ldots v^{w_N};\ \tilde{\beta}) \tag{4}$$

Manager chooses document $D = [w^1 \ldots w^N]$ to maximize some utility function u, e.g. firm value, less some cost $\lambda\mathcal{L}$ for writing a document where the reported beta $g(v^{w_1},\ldots v^{w_N})$ is too far from the manager's private signal $\tilde{\beta}$.

Discussion: Economic Models of Language

Does This Help?

Can analysis of these games facilitate economically principled development of:

- Word embeddings?
- Natural language processing?
- Language models?

What results can we borrow from other elsewhere to help study these games?

- Information Theory
- Computer Science
- Electrical Engineering

Conclusion

- **Topic** and **context** are complementary—statistically significant and distinct in forecasts
  $\rightarrow$ certainty-equivalent value of 0.3% per annum
- For market beta, only **context** is important; for other betas, **both** are important
- This holds while controlling for current beta, other firm characteristics
- Novel evidence: same text provides distinct data using different NLP methods

Thank you