

# Risk Factors that Matter: Textual Analysis of Risk Disclosures for the Cross-Section of Returns

by Alejandro Lopez-Lira

Discussant: Asaf Manela

January 2022

## Why do we care?

- ▶ Most asset pricing “risk factors” are latent representations designed to explain common variation in returns
  - ▶ For example, “value premium” has many potential underlying reasons
- ▶ But investors and managers must understand which risks exactly to optimize their portfolios and their firms
  - ▶ How do you reduce your “momentum risk”?
- ▶ We would like to understand what are the underlying economic sources of risk for which investors require compensation

## What the paper does?

- ▶ Use LDA to allocate risk factors discussions in firm annual reports into 25 topics
- ▶ Argues using a model that firms more exposed to a risk allocate more text to warn about it
- ▶ Finds that firms with similar risk disclosures have correlated returns
- ▶ Then, form factor mimicking stock portfolios based on topic weights (attentions)
- ▶ Finds the portfolios are priced in the cross section of stock returns
- ▶ Portfolios are not spanned by commonly used factor models

## What did I learn?

- ▶ Textual analysis has been used elsewhere by prior literature and for somewhat similar purposes

*“[News implied volatility] variation is interpretable and provides insight into the origins of risk variation ... allows us to identify which kinds of risk were important to investors.”*  
*(Manela-Moreira, 2017 JFE)*
- ▶ Main innovation here is the claim that
  1. Firms know better than anyone the risks to their cash flows and disclose them in their 10-K's
  2. LDA successfully allocates these disclosures to 25 topics
  3. Topic-based factor mimicking portfolios are as good as leading factor models
- ▶ Not much progress on interpretation, but an important step in the right direction

## Suggestion 1: Interpretation vs. statistics

- ▶ Hard to reconcile these excerpts:
  1. Cochrane (2005) warns that: "it is probably not a good idea to evaluate economically interesting models with statistical horse races against models that use portfolio returns as factors..."
  2. Hence, it is especially surprising that the model appears to have a statistical fit at least as good as the leading models in the literature
  3. I use machine learning and the information revealed by the firms in the economy to answer some of the essential questions in asset pricing: What are the significant risks in the economy according to firms themselves? Which ones are systematic? Are they priced?
  4. labels and keywords are just for illustrative purposes. There is no need to manually label the risks in the paper or define the keywords since the risks will arise naturally using the LDA algorithm

# Suggestion 1: Interpretation vs. statistics



A word cloud centered around the word "china". Other prominent words include "legal", "chinese", "rmb", "subsidiary", "tax", "foreign\_currency", "safe", "offshore", and "chinese\_government". Smaller words surrounding the center include "economy", "approval", "procedure", "transfer", "relevant", "shareholder", "registration", "currency", "foreign\_exchange", "dollar\_entity", "enterprise", "court", "authority", and "safe".



A word cloud centered around the word "client". Other prominent words include "government\_agency", "u\_government", "contractor", "professional", "agency", "procurement", "funding", "subcontractor", "government\_contract", "bid", "budget", "work", "defense", "award", "perform", "engagement", "backlog", "profit", "audit", "spending", "federal\_government", and "equipment". Smaller words surrounding the center include "terminate", "termination", and "bid".



A word cloud centered around the word "institution". Other prominent words include "advertiser", "television", "video\_medium", "medium", "satellite", "user", "internet", "channel", "mobile", "radio", "cable", "online", "station", "network", "fcc", "digital", "license", "entertainment", "content", "advertising", "education", "school", "institution", "lose", "proprietary", "private", "enrollment", "loan", "online", "loan", "educational", "funding", "federal", "university", "fiscal\_year", "eligibility", "participate", "title", "course", "authorization", and "commission". Smaller words surrounding the center include "employment", "participation", "default\_rate", "agency", "grant", and "private".



A word cloud centered around the word "advertising". Other prominent words include "advertiser", "television", "video\_medium", "medium", "satellite", "user", "internet", "channel", "mobile", "radio", "cable", "online", "station", "network", "fcc", "digital", "license", "entertainment", "content", "advertising", "cable", "online", "station", "network", "fcc", "digital", "license", and "entertainment". Smaller words surrounding the center include "internet", "channel", "mobile", "radio", "cable", "online", "station", "network", "fcc", "digital", "license", and "entertainment".

Part of Figure 4: Risk Topics

## Suggestion 2: Theoretical model ...

- ▶ From Footnote 9:

*"Firms are legally required to discuss "the most significant factors that make the company speculative or risky" (Regulation S-K, Item 105(c), SEC 2005) in a specific section of the 10-K annual reports (Section 1A). They could face legal action if they fail to obey the regulation and be vulnerable to lawsuits from investors."*

## Suggestion 2: Theoretical model ...

$$\min_{L_i} \sum_{i=0}^I p(L_i) C(L_i, r_{ij}) + h(L_i)$$

## Suggestion 3: Explain better the key contribution

- ▶ A central challenge in macro finance is to move beyond explaining prices with prices, toward explaining prices with economic fundamentals
- ▶ The paper makes progress toward that goal:
  - ▶ “A model that uses only firm-identified risk factors performs at least as well as traditional factor models when pricing a broad set of assets, despite not using any information from past prices or returns.”

## My take

- ▶ Really nice contribution to our understanding of firm-identified risks
- ▶ Finds that risk factors discussion in annual reports is informative about priced risks
- ▶ Overcomes several interesting methodological issues and I encourage you all to read it

## Other suggestions / minor points

- ▶ Figure 4 should include the topic labels below each word cloud
- ▶ The methodology section says you analyze the text as bigrams, which makes a lot of sense, but then the word clouds have unigrams too
- ▶ See Bybee-Kelly-Manela-Xiu (2021) for a couple of data driven approaches to pick the optimal number of LDA topics