Discussion of Kim, Muhn, and Nikolaev (2024)

“From Transcripts to Insights: Uncovering Corporate Risks Using Generative AI”

Discussant: Sangmin Simon Oh (Columbia Business School)

E(uropean)FA 2024

Recap

Objective

- Explore the value of generative AI in helping (investors) uncover dimensions of corporate risk from earnings conference call transcripts

Approach

- Generate risk summaries from transcripts → Convert them into quantitative measures
- Examine how risk measures relate to firm outcomes and financial market variables

Result

- AI-generated risk measures outperform traditional methods in (i) capturing different dimensions of risk and (ii) predicting associated firm responses.

Recap

Objective

- Explore the value of generative AI in helping (investors) uncover dimensions of corporate risk from earnings conference call transcripts

Approach

- Generate risk summaries from transcripts → Convert them into quantitative measures
- Examine how risk measures relate to firm outcomes and financial market variables

Result

- AI-generated risk measures outperform traditional methods in (i) capturing different dimensions of risk and (ii) predicting associated firm responses.

Highly creative approach & opens doors to future applications in finance

- The team's research agenda is indeed promising and definitely worth following closely.

Recap

Objective

- Explore the value of generative AI in helping (investors) uncover dimensions of corporate risk from earnings conference call transcripts

Approach

- Generate risk summaries from transcripts → Convert them into quantitative measures
- Examine how risk measures relate to firm outcomes and financial market variables

Result

- AI-generated risk measures outperform traditional methods in (i) capturing different dimensions of risk and (ii) predicting associated firm responses.

Highly creative approach & opens doors to future applications in finance

- The team's research agenda is indeed promising and definitely worth following closely.

Plan for Discussion

1. Role of information sets in interpreting the LLM output
2. Implementation choices
3. What to do (can be done) with the measured risk exposures?

Point 1. Role of the information set

An underappreciated aspect of Generative AI

- Models like ChatGPT operate with an exceptionally large information set – far exceeding that of an average investor or an analyst facing earnings conference calls.

An underappreciated aspect of Generative AI

- Models like ChatGPT operate with an exceptionally large information set – far exceeding that of an average investor or an analyst facing earnings conference calls.

- Measuring risk exposure can be thought as estimating $f(X_t|I_t)$ where X_t represents the earnings conference calls and I_t is the information set known at time t.
  - Traditional approaches like using bigrams implicitly assumes $X_t \approx I_t$
  - In other words, all relevant information is contained within the call transcript.

An underappreciated aspect of Generative AI

- Models like ChatGPT operate with an exceptionally large information set – far exceeding that of an average investor or an analyst facing earnings conference calls.

- Measuring risk exposure can be thought as estimating $f(X_t|I_t)$ where X_t represents the earnings conference calls and I_t is the information set known at time t.
  - Traditional approaches like using bigrams implicitly assumes $X_t \approx I_t$
  - In other words, all relevant information is contained within the call transcript.

- So far, our focus has been mostly in a flexible modeling of the function f:
  - Predicting expected returns: Gu, Kelly, and Xiu (RFS, 2020)
  - Measuring financial constraints: Linn and Weagley (JFQA, 2023)
  - Microstructure: Easley et al. (RFS, 2021)

An underappreciated aspect of Generative AI

- Models like ChatGPT operate with an exceptionally large information set – far exceeding that of an average investor or an analyst facing earnings conference calls.

- Measuring risk exposure can be thought as estimating $f(X_t|I_t)$ where X_t represents the earnings conference calls and I_t is the information set known at time t.
  - Traditional approaches like using bigrams implicitly assumes $X_t \approx I_t$
  - In other words, all relevant information is contained within the call transcript.

- So far, our focus has been mostly in a flexible modeling of the function f:
  - Predicting expected returns: Gu, Kelly, and Xiu (RFS, 2020)
  - Measuring financial constraints: Linn and Weagley (JFQA, 2023)
  - Microstructure: Easley et al. (RFS, 2021)

- **This Paper:** Generative AI **not only** makes f flexible, **but also** significantly expands I_t

- **This is an underappreciated, yet a very important point.**
  - A fundamental shift in how we approach risk assessment from textual data.

How does an expanded I_t matter in the context of this study?

- A (very) large I_t , from a research perspective, is both exciting and challenging.

- On one hand, it potentially reduces prediction errors by better measuring what we want to measure. In other words, it helps deliver the performance.

How does an expanded I_t matter in the context of this study?

- A (very) large I_t, from a research perspective, is both exciting and challenging.

- On one hand, it potentially reduces prediction errors by better measuring what we want to measure. In other words, it helps deliver the performance.

- On the other hand, it also becomes difficult to pinpoint the source of this improvement when using LLMs. Relatedly, two issues arise:
  - Q1. How much of the improvement is due to better f vs. larger I_t ?
  - Q2. What exactly in $\mathrm{I_t}$ is helping us here?
    - Note that the bigram approach, in contrast, allows for a controlled analysis.

How does an expanded I_t matter in the context of this study?

- A (very) large I_t , from a research perspective, is both exciting and challenging.

- On one hand, it potentially reduces prediction errors by better measuring what we want to measure. In other words, it helps deliver the performance.

- On the other hand, it also becomes difficult to pinpoint the source of this improvement when using LLMs. Relatedly, two issues arise:
  - Q1. How much of the improvement is due to better f vs. larger I_t ?
  - Q2. What exactly in I_t is helping us here?
    - Note that the bigram approach, in contrast, allows for a controlled analysis.

- To connect the usage of LLMs back to economics, it seems crucial to understand the source of improvement. This leads me to two suggestions for the authors.

How does an expanded I_t matter in the context of this study?

- A (very) large I_t , from a research perspective, is both exciting and challenging.

- On one hand, it potentially reduces prediction errors by better measuring what we want to measure. In other words, it helps deliver the performance.

- On the other hand, it also becomes difficult to pinpoint the source of this improvement when using LLMs. Relatedly, two issues arise:
  - Q1. How much of the improvement is due to better f vs. larger I_t ?
  - Q2. What exactly in I_t is helping us here?
    - Note that the bigram approach, in contrast, allows for a controlled analysis.

- To connect the usage of LLMs back to economics, it seems crucial to understand the source of improvement. This leads me to two suggestions for the authors.

Suggestion 1a: It would be helpful to see more side-by-side examples that compare ChatGPT-generated risk summaries vs. those created using traditional methods

- Could potentially help us understand whether improvement is coming from f vs. I_t

Suggestion 1b: Consider iterating with models trained on expanding windows of data (similar to the approach in Sarkar (2024)) to conduct the exercise under a more controlled information environment.

Point 2. Implementation Choices

Implementation Challenge: From Text to Numbers

- Expanding the methodological frontier inevitably comes with baseline implementation choices.

Implementation Challenge: From Text to Numbers

- Expanding the methodological frontier inevitably comes with baseline implementation choices.

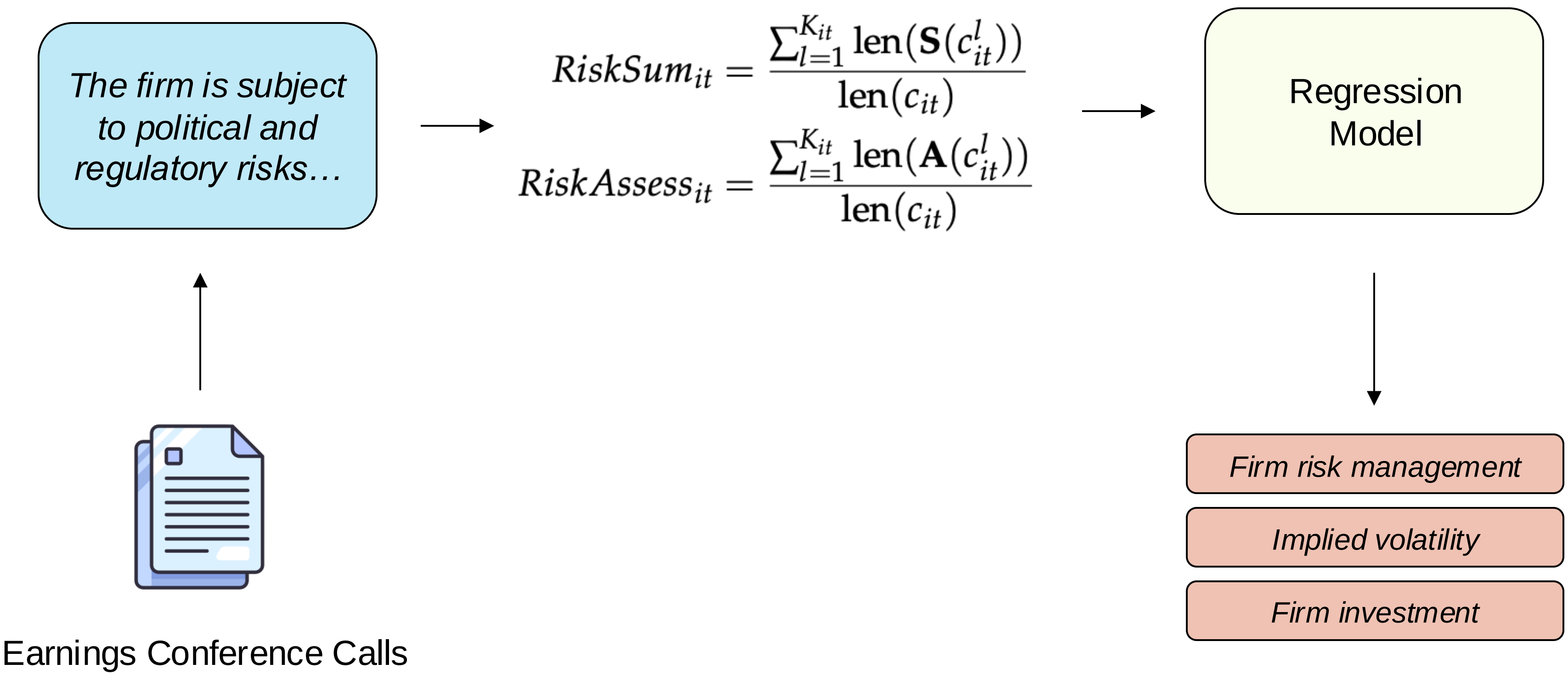
- In my opinion, the biggest hurdle for integrating LLMs into economic analyses is how to convert the text output into a numerical quantity.

Implementation Challenge: From Text to Numbers

- Expanding the methodological frontier inevitably comes with baseline implementation choices.

- In my opinion, the biggest hurdle for integrating LLMs into economic analyses is how to convert the text output into a numerical quantity.

- **Approach 1: KMN Approach (← This paper)**

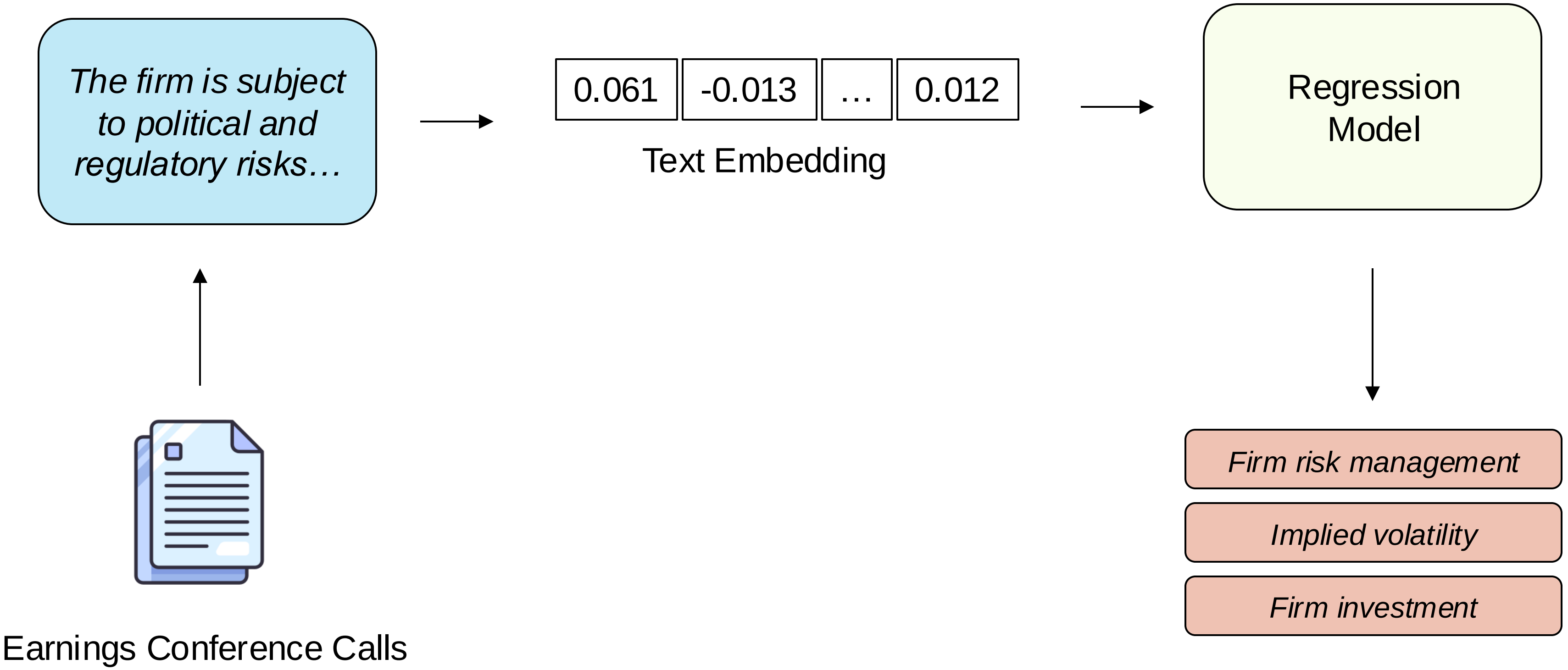


Implementation Challenge: From Text to Numbers

- Expanding the methodological frontier inevitably comes with baseline implementation choices.

- In my opinion, the biggest hurdle for integrating LLMs into economic analyses is how to convert the text output into a numerical quantity.

- **Potential Approach 2: "Embeddings" Approach**



Implementation Challenge: From Text to Numbers

- Expanding the methodological frontier inevitably comes with baseline implementation choices.

- In my opinion, the biggest hurdle for integrating LLMs into economic analyses is how to convert the text output into a numerical quantity.

- **Potential Approach 3: “Just ask the LLM”**

Implementation Challenge: From Text to Numbers

- Expanding the methodological frontier inevitably comes with baseline implementation choices.

- In my opinion, the biggest hurdle for integrating LLMs into economic analyses is how to convert the text output into a numerical quantity.

- **Potential Approach 3: “Just ask the LLM”**

Earnings Conference Calls

Suggestion 2a: It would be beneficial to be more upfront about the limitations of the current metric. For example, the authors could discuss how their measure might not capture the full semantic content or nuance of the risk summary.

Suggestion 2b: If possible, explore sensitivity to alternate ways of quantification would be immensely helpful.

Point 3. What to do with the measured risk exposures?

What to do with the measured risk exposures?

- In Sections 6 through 9, authors examine connections to:
  - Capital market variables (implied volatility, abnormal volatility)
  - Firm decisions (investment, risk mitigation)
  - Asset prices

What to do with the measured risk exposures?

- In Sections 6 through 9, authors examine connections to:
  - Capital market variables (implied volatility, abnormal volatility)
  - Firm decisions (investment, risk mitigation)
  - Asset prices

- While this approach provides a broad view of the potential applications of their methodology, I believe there's an opportunity to focus the analysis more sharply.

What to do with the measured risk exposures?

- In Sections 6 through 9, authors examine connections to:
  - Capital market variables (implied volatility, abnormal volatility)
  - Firm decisions (investment, risk mitigation)
  - Asset prices

- While this approach provides a broad view of the potential applications of their methodology, I believe there's an opportunity to focus the analysis more sharply.

Suggestion 3a:

Rather than explore connections to all of these, I suggest that authors focus on one exercise that *ex ante* can highlight the greatest advantage of using LLMs

- For example, it's possible that LLMs truly uncovers otherwise difficult-to-capture risk dimensions through a more traditional approach. If this is true, it should better forecast firm incidents that are otherwise typically difficult to predict.
- To test this, the authors could leverage a more granular, micro-level dataset, such as RepRisk, to conduct a horse race between traditional approaches and their LLM approach.

What to do with the measured risk exposures?

- In Sections 6 through 9, authors examine connections to:
  - Capital market variables (implied volatility, abnormal volatility)
  - Firm decisions (investment, risk mitigation)
  - Asset prices

- While this approach provides a broad view of the potential applications of their methodology, I believe there's an opportunity to focus the analysis more sharply.

Suggestion 3a:

Rather than explore connections to all of these, I suggest that authors focus on one exercise that *ex ante* can highlight the greatest advantage of using LLMs

- For example, it's possible that LLMs truly uncovers otherwise difficult-to-capture risk dimensions through a more traditional approach. If this is true, it should better forecast firm incidents that are otherwise typically difficult to predict.
- To test this, the authors could leverage a more granular, micro-level dataset, such as RepRisk, to conduct a horse race between traditional approaches and their LLM approach.

Suggestion 3b:

Among the LHS variables examined, I found the risk mitigation efforts to be the most interesting. Are there potentially puzzling firm risk management behavior that only an LLM-based risk exposures can help explain?

Final Thoughts

- Authors explore the use of generative AI (LLMs) to uncover dimensions of corporate risk from earnings call transcripts

- **Punchline**: AI-generated risk measures outperform traditional methods in predicting volatility and firm decisions, demonstrating the potential of LLMs in financial analysis

Final Thoughts

- Authors explore the use of generative AI (LLMs) to uncover dimensions of corporate risk from earnings call transcripts

- **Punchline**: AI-generated risk measures outperform traditional methods in predicting volatility and firm decisions, demonstrating the potential of LLMs in financial analysis

- While the results are impressive, I believe the paper could benefit from delving deeper into the drivers of this outperformance.
    - Using LLMs improves the measurement exercise in multiple dimensions, and it would be useful to know where the outperformance comes from exactly.

Final Thoughts

- Authors explore the use of generative AI (LLMs) to uncover dimensions of corporate risk from earnings call transcripts

- **Punchline**: AI-generated risk measures outperform traditional methods in predicting volatility and firm decisions, demonstrating the potential of LLMs in financial analysis

- While the results are impressive, I believe the paper could benefit from delving deeper into the drivers of this outperformance.
  - Using LLMs improves the measurement exercise in multiple dimensions, and it would be useful to know where the outperformance comes from exactly.

- **That said, I really appreciate the authors' creative and innovative use of LLMs for addressing a central challenge in finance: measuring risk.**

- **Some questions prompted by the paper for the future:**
  - How do/should we control for the information set when working with LLMs?
  - What is the most effective way to integrate LLM outputs into our existing empirical frameworks?
  - Will the introduction of LLMs change how firms disclose information?

Thank you!

Addendum: For Authors Only

- **Q1. Timing:** Some regressions examine contemporaneous relationships while some examine relationship with a lagged risk measure:

$$Volatility_{it+1} = \beta Risk_{it} + \gamma \mathbf{X}_{it} + \delta_x + \varepsilon_{it},$$

$$Investment_{it} = \beta Risk_{it} + \gamma \mathbf{X}_{it} + \delta + \varepsilon$$

$$Response_{it+1} = \beta Risk_{it} + \gamma \mathbf{X}_{it} + \delta + \varepsilon$$

$$\begin{aligned} Implied\_Volatility_{it} =& \beta_{1t} PRiskAssess_{it} + \beta_{2t} CRiskAssess_{it} + \beta_{3t} AIRiskAssess_{it} \\ &+ \gamma_t \mathbf{X}_{it} + \delta_q + \delta_s + \varepsilon_{it} \end{aligned} \tag{6}$$

- **Q2. Factor Structure:** Given the known factor structure in returns, clustering by time seems quite important. Otherwise, we risk overstating the "effective" number of observations.

- **Caveat:** Note that the two questions I have above are independent of how $Risk_{it}$ is measured, which is the main point of the paper.

Addendum: For Authors Only

- **Q3. Authors set the temperature parameter to zero.**
  - This means that the model always chooses the most probable next token, which reduces variability in the generated text.
  - Q. If we vary the temperature, is the LLM better at detecting more nuanced discussion of risks in the earnings calls?

- **Q4. Authors work with OpenAI's GPT3.5-Turbo LLM.**
  - Q. How much can fine-tuning the model improve the risk detection and assessment accuracy?