Summary

Zhenting Hu

April 7, 2024

1. What are the research questions?

The main analysis examine whether GPT-based proxies are effective at measuring firm-level risks and the risk exposure measures can explain future stock price volatility and firm behavior.

2. Why are the research questions interesting?

Unlike traditional methods that analyze risks based on a single document with dictionaries, such as a conference call transcript, generative language models are trained on vast corpora that enable them to leverage general knowledge acquired from similar documents or documents featuring related topics, therefore it leaves a lot of research to explore.

3. What is the paper's contribution?

(1) Previous literature

i. Former studies have shown that AI tools are effective at distilling disclosures to extract information about diverse risk categories(see Bernard et al., 2023; Lopez-Lira and Tang, 2023; Jha et al., 2023; Eisfeldt et al., 2023; Kim et al., 2023; Chen et al., 2023).

ii. Existing studies rely on topic-based bigram dictionaries to construct firm-level risk exposures in cluding political risk (Hassan et al., 2019), country risk (Hassan et al., 2021), climate risk (Sautner et al., 2023), inflation risk (Chava et al., 2022), and pandemic risk (Hassan et al., 2020).

(2) Marginal contribution of the paper

i. The paper probe the economic usefulness of AI-powered large language models in risk assessment.

ii. The paper complement and build on the textual analysis work by adopting AI-based technology to analyze risks.

iii. The paper show that LLMs successfully leverage their general knowledge to derive insights about corporate risks from a given context, which goes beyond the information discussed in the processed document.

4. What hypotheses are tested in the paper?

While the exact wording of the hypotheses isn't provided in the text snippets, based on the context and main research questions, the hypotheses likely include:

Hypothesis1:

The GPT-based risk measures have significant predictive power regarding firm-level outcomes, such as abnormal firm-level volatility, investment, and innovation activities, surpassing that of conventional risk measures.

Hypothesis2:

AI-generated measures of emerging risks are not only effectively identified by generative AI models but are also reflected in equity market pricing, indicating their relevance and informativeness to investors.

The hypotheses are developed and tested based on the research questions, they answer the main question of the validation of the GPT-based risk measures and make additional analysis on the investment value of the measures.

5. Sample: Comment on the appropriateness of the sample selection procedures.

The most important data are US firms' transcripts available between January 2018 and March 2023.

i. Selecting the period because (1) generating risk summaries and assessments for each risk metric is costly and time-consuming, (2) a considerable part of the sample is outside of GPT's training window (allowing for pure out-of-sample tests), and (3) this time period is characterized by significant changes in political, climate and AI uncertainty.

ii. The paper exclude very short calls and calls without a discussion session, calls that are conducted in languages other than English, including the ones that are machine-translated into English. These filtering processes drop much noise from the corpus.

6. Dependent and Independent Variables.

Dependent variables:

Implied Volatility: The paper use the implied volatility derived from the 90-day at-the-money options measured as of the end of each fiscal quarter.

Abnormal Volatility: The ratio of Post-call-RMSE and Pre-call-RMSE minus 1, where the RMSE is the root mean squared errors from the market model residuals. The Pre-interval is [-257,-6] while the Post-interval is [+6,+28].

The above variables are used following previous literature(Engle, 2004, Loughran and McDonald (2014)) and can serve as credible measures.

Independent variables:

$$RiskSum_{it} = \frac{\sum_{l=1}^{K_{it}} len(\mathbf{S}(c_{it}^{l}))}{len(c_{it})}$$

$$RiskAssess_{it} = \frac{\sum_{l=1}^{K_{it}} len(\mathbf{A}(c_{it}^{l}))}{len(c_{it})}$$

Where $\boldsymbol{S}$ and $\boldsymbol{A}$ correspond to the GPT-based function. c_{it} is earnings call transcript for a company i in quarter t divided into K_{it} chunks. This simple method is developed from the paper and can distinguish between different types of AI-based measures.

7. Regression/prediction model specification.

The main model used in the paper is OLS regression with fixed effects which is a typical and well-suited econometric model

8. What difficulties arise in drawing inferences from the empirical work?

(1) Generative AI models like GPT can generate outputs based on vast amounts of data, but the "black box" nature of these models can make it challenging to interpret how they arrive at certain conclusions.

(2) Establishing the predictive power of these measures concerning firm-level outcomes and market behavior involves complex statistical analyses that can be difficult to interpret and validate not only from previous study.

(3) The findings may not be easily available across different sectors, regions, or time periods without additional validation and adaptation of the model.

9. Describe at least one publishable and feasible extension of this research.

(1) Can newer versions of generative AI models (e.g., GPT-4 or beyond) more effectively identify and analyze sector-specific risks, including emerging risks not previously captured?

(2) How do political, climate, AI-related or other type of risks differently impact specific sectors like technology, finance, and manufacturing instead of the whole market.