Discussion: Corporate Earnings Calls and Analyst Beliefs

Giuseppe Matera (SFI @ EPFL)

Emanuele Luzzi

SFI @ USI Lugano

October 17, 2025

Question

- Do narratives in earnings calls affect analyst beliefs beyond fundamentals?
- **Claim:** Textual features have an impact on analysts' beliefs.
- **Result:** Text adds 2–15% incremental R^2 beyond fundamentals, out-of-sample.

Overall Assessment

- Creative methodology with broad potential applications.
- Important question, useful to better understand economic behavior.
- Fun to read!
- My comments will revolve around the experiment.

How does the experiment work?

Ideal experiment:

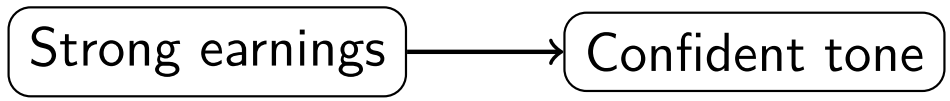
- Rewind time, run earning calls with different language/tone.
- Observe real analysts' reactions.

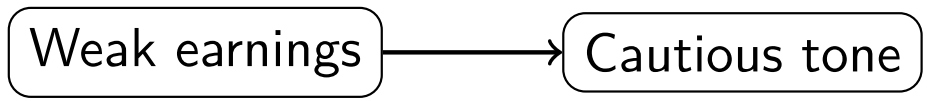
Paper:

1. ML model trained on text + fundamentals $\rightarrow$ Analyst
2. LLM morphes the tone of the call $\rightarrow$ Counterfactual
3. Feed morphed call to the artificial analyst **fixing fundamentals** $\Rightarrow$ get reaction on text features.

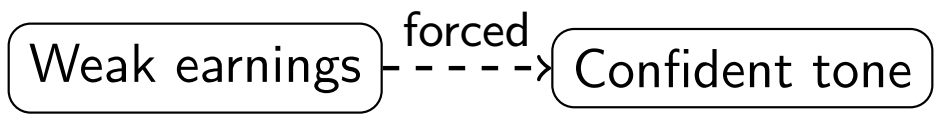
Comment 1: Out-of-Distribution

Training: Artificial analyst learns from natural correlations.





Morphing: Breaks these correlations by fixing fundamentals and changing text:



$\Rightarrow$ Artificial analyst might not be able to assess these new combinations correctly under this regime.

Comment 1: Out-of-Distribution

- Training on the natural covariance of real data.
- Counterfactual inputs break natural correlations observed during training for the prediction.
- ML models learn complex interactions and shortcuts that are **not necessarily** causal ones, so the artificial analyst could break too.

Comment 2: Who's the judge?

- LLM-as-a-judge validates if the synthetic transcript morphing was successful.
- Resembles a discriminator for a GAN (generative adversarial network), but without the adversary.
- Assumes LLM-judge is an oracle.
- Risk of systematic blind spots and redundant hallucinations.
- Might still miss subtle inconsistencies that humans would catch.

Comment 3: R^2 is 10 times the literature

- Chen et al. (2022) uses ML to get $R^2 = 5$–8% (against random walk).
- This paper: $R^2 = 71\%$ $\leftarrow 10\times$ better! (against mean)
- Maybe persistence? (check with a different benchmark).

Target	R^2 **(Numeric)**	R^2 **(+ Text)**	**Text adds**
Change in earnings	71.3%	72.2%	**+0.9pp**
Forecast Error	11.1%	12.3%	**+1.2pp**

With $R^2 = 71\%$, absolute changes are important to assess magnitude.

Suggestions

For Comment 1 (OOD):

- Train data on morphed calls too.
- Disentangle text effects from other varying factors.

For Comment 2 (Judge):

- Human experiment: Show morphed vs. original calls to MBAs
- More rigorous adversarial framework.

For Comment 3 (R^2):

- Benchmark against AR(1), random walk, industry trends.
- Rolling-window validation accounting for structural breaks.
- More detail on incremental R^2 decomposition (text as info vs. proxy).

Thank you!