The Ghost in the Machine: Generating Beliefs with Large Language Models

Leland Bybee (2025)

Discussion by Aditya Chaudhry

The Ohio State University

19th March 2025

This Paper in Context

Growing interest in deviations from full information rational expectations (FIRE)

This Paper in Context

Growing interest in deviations from full information rational expectations (FIRE)

- **Critique:** "Without rational expectations, anything goes"
  - Deviations from FIRE introduce free parameters
  - Need new moments/data to discipline those new free parameters

This Paper in Context

Growing interest in deviations from full information rational expectations (FIRE)

- **Critique:** “Without rational expectations, anything goes”
  - Deviations from FIRE introduce free parameters
  - Need new moments/data to discipline those new free parameters
- **Solution:** Collect beliefs data
  - Households, executives, professional forecasters, equity analysts, etc.

This Paper in Context

Growing interest in deviations from full information rational expectations (FIRE)

- **Critique:** “Without rational expectations, anything goes”
  - Deviations from FIRE introduce free parameters
  - Need new moments/data to discipline those new free parameters
- **Solution:** Collect beliefs data
  - Households, executives, professional forecasters, equity analysts, etc.

Limitation: Surveyed beliefs data are sparse

- Cover only a small portion of the space of beliefs we care about
  - Economic variables (e.g. hundreds of variables in FRED-QD)
  - Time periods (less data going further back in time)
  - Frequencies (less data at higher frequencies)
  - Moments (less data on second & higher moments)

This Paper in Context

Growing interest in deviations from full information rational expectations (FIRE)

- **Critique:** "Without rational expectations, anything goes"
  - Deviations from FIRE introduce free parameters
  - Need new moments/data to discipline those new free parameters
- **Solution:** Collect beliefs data
  - Households, executives, professional forecasters, equity analysts, etc.

Limitation: Surveyed beliefs data are sparse

- Cover only a small portion of the space of beliefs we care about
  - Economic variables (e.g. hundreds of variables in FRED-QD)
  - Time periods (less data going further back in time)
  - Frequencies (less data at higher frequencies)
  - Moments (less data on second & higher moments)

This paper: Extract beliefs from news text using large language models (LLMs)

- **Upshot:** If this works, can fill more of space of beliefs we want to measure

Summary

Feed in WSJ headlines from time period *t* into ChatGPT 3.5

- Ask: Based on this news, will [a certain macro series] increase or decrease?
- Aggregate responses across all headlines in period t to obtain signed belief measure $\in [-1, 1]$

Summary

Feed in WSJ headlines from time period *t* into ChatGPT 3.5

- Ask: Based on this news, will [a certain macro series] increase or decrease?
- Aggregate responses across all headlines in period t to obtain signed belief measure $\in [-1, 1]$

Main focus: Comparing generated beliefs to standard beliefs data

- Generated beliefs for return & macro expectations correlate reasonably strongly with common surveys
- Generated beliefs predict forecast errors to a similar degree as common surveys
  - Suggests generated beliefs may reflect similar deviations from FIRE

Summary

Feed in WSJ headlines from time period t into ChatGPT 3.5

- Ask: Based on this news, will [a certain macro series] increase or decrease?
- Aggregate responses across all headlines in period t to obtain signed belief measure $\in [-1, 1]$

Main focus: Comparing generated beliefs to standard beliefs data

- Generated beliefs for return & macro expectations correlate reasonably strongly with common surveys
- Generated beliefs predict forecast errors to a similar degree as common surveys
  - Suggests generated beliefs may reflect similar deviations from FIRE

Main application: Predicting bubble crashes

- Extract aggregate "sentiment" factor from generate beliefs
- Measure industry betas wrt. aggregate sentiment
- Finds that sentiment betas predict industry stock price crashes

Overall: Very Interesting Paper

Innovative methodology to tackle difficult & important problem

Overall: Very Interesting Paper

Innovative methodology to tackle difficult & important problem

Paper does a good job of comparing generated beliefs vs. standard data

Overall: Very Interesting Paper

Innovative methodology to tackle difficult & important problem

Paper does a good job of comparing generated beliefs vs. standard data

Where do these beliefs come from?

- Want to understand how to model deviations from FIRE
- Some discussion of this, but have opportunity to go deeper

Overall: Very Interesting Paper

Innovative methodology to tackle difficult & important problem

Paper does a good job of comparing generated beliefs vs. standard data

Where do these beliefs come from?

- Want to understand how to model deviations from FIRE
- Some discussion of this, but have opportunity to go deeper

Discussion: Suggestions on how to open black box & reveal how these beliefs are formed

- What is the model's prior?
- How exactly does the model update in response to different signals?

Simple Model

Want to forecast some variable x

$$x \sim \underbrace{N\left(\overline{x}, \sigma^2\right)}_{\text{Prior}}$$

Simple Model

Want to forecast some variable x

$$x \sim \underbrace{N\left(\overline{x}, \sigma^2\right)}_{\text{Prior}}$$

Observe some signal s

$$s = x + \epsilon$$
$$\epsilon \sim N\left(0, \omega^2\right)$$

Simple Model

Want to forecast some variable x

$$x \sim \underbrace{N\left(\overline{x}, \sigma^2\right)}_{\text{Prior}}$$

Observe some signal s

$$s = x + \epsilon$$
$$\epsilon \sim N\left(0, \omega^2\right)$$

Standard expression for posterior expectation

$$\mathbb{E}\left[x\right] = \theta \cdot s + (1 - \theta)\,\overline{x}$$

- Bayesian chooses optimal gain θ
- Biases can lead to too-large (overreaction) or too-small (underreaction) θ

Simple Model

Want to forecast some variable x

$$x \sim \underbrace{N\left(\overline{x}, \sigma^2\right)}_{\text{Prior}}$$

Observe some signal s

$$s = x + \epsilon$$
$$\epsilon \sim N\left(0, \omega^2\right)$$

Standard expression for posterior expectation

$$\mathbb{E}\left[x\right] = \theta \cdot s + (1 - \theta)\,\overline{x}$$

- Bayesian chooses optimal gain θ
- Biases can lead to too-large (overreaction) or too-small (underreaction) θ

To understand how generated beliefs are formed, need to understand

- What is ChatGPT's prior: $\overline{x}$
- How does ChatGPT's update to different signals: θ

What is ChatGPT's Prior?

Feed placebo headlines to ChatGPT, ask it to forecast macro variables

- E.g. Celebrity gossip headlines
- Force an up/down response, shut down "uncertain" option
- Faced with objectively uninformative signals, it should return prior expectations
- E.g. Forecasts S&P 500 increases 60% of the time

What is ChatGPT's Prior?

Feed placebo headlines to ChatGPT, ask it to forecast macro variables

- E.g. Celebrity gossip headlines
- Force an up/down response, shut down "uncertain" option
- Faced with objectively uninformative signals, it should return prior expectations
- E.g. Forecasts S&P 500 increases 60% of the time

Compare responses to empirical distributions of macro variables

What is ChatGPT's Prior?

Feed placebo headlines to ChatGPT, ask it to forecast macro variables

- E.g. Celebrity gossip headlines
- Force an up/down response, shut down "uncertain" option
- Faced with objectively uninformative signals, it should return prior expectations
- E.g. Forecasts S&P 500 increases 60% of the time

Compare responses to empirical distributions of macro variables

Is ChatGPT (unconditionally) overly optimistic or pessimistic?

- Compare to surveys from other agents
- Is ChatGPT more or less biased?

How does ChatGPT Update?

General challenge: Difficult to identify why beliefs change

- Many correlated variables may explain a particular update

How does ChatGPT Update?

General challenge: Difficult to identify why beliefs change

- Many correlated variables may explain a particular update

- E.g. Return vs. fundamental extrapolation
  - Observe: High past returns $\leftrightarrow$ High return expectations

How does ChatGPT Update?

General challenge: Difficult to identify why beliefs change

- Many correlated variables may explain a particular update
- E.g. Return vs. fundamental extrapolation
  - Observe: High past returns $\leftrightarrow$ High return expectations
  - Return Extrap.: High past returns $\rightarrow$ High return expectations

How does ChatGPT Update?

General challenge: Difficult to identify why beliefs change

- Many correlated variables may explain a particular update
- E.g. Return vs. fundamental extrapolation
  - Observe: High past returns $\leftrightarrow$ High return expectations
  - Return Extrap.: High past returns $\rightarrow$ High return expectations
  - Fundamental Extrap.: High past returns $\leftrightarrow$ High past cash flow growth $\rightarrow$ High growth expectations $\rightarrow$ High return expectations
    - Due to heterogeneity or neglect of general equilibrium

How does ChatGPT Update?

General challenge: Difficult to identify why beliefs change

- Many correlated variables may explain a particular update
- E.g. Return vs. fundamental extrapolation
  - Observe: High past returns $\leftrightarrow$ High return expectations
  - Return Extrap.: High past returns $\rightarrow$ High return expectations
  - Fundamental Extrap.: High past returns $\leftrightarrow$ High past cash flow growth $\rightarrow$ High growth expectations $\rightarrow$ High return expectations
    - Due to heterogeneity or neglect of general equilibrium

LLM-unique solution: "Belief counterfactuals"

- Feed in headlines describing high past returns and low cash flow growth, and vice versa
- See which signal leads model to expect higher future returns
- More direct complement to DAG approach in paper

How does ChatGPT Update?

General challenge: Difficult to identify why beliefs change

- Many correlated variables may explain a particular update
- E.g. Return vs. fundamental extrapolation
  - Observe: High past returns $\leftrightarrow$ High return expectations
  - Return Extrap.: High past returns $\rightarrow$ High return expectations
  - Fundamental Extrap.: High past returns $\leftrightarrow$ High past cash flow growth $\rightarrow$ High growth expectations $\rightarrow$ High return expectations
    - Due to heterogeneity or neglect of general equilibrium

LLM-unique solution: “Belief counterfactuals”

- Feed in headlines describing high past returns and low cash flow growth, and vice versa
- See which signal leads model to expect higher future returns
- More direct complement to DAG approach in paper

More generally, can compute counterfactual generated belief series

- Drop articles on certain topics (e.g. inflation, growth, etc.) when constructing generated belief series
- See if alternative series correlate more/less strongly with forecast errors
- Reveals which types of signals model misreacts to most strongly

Minor Comments

Additional way to mitigate lookahead bias

- Project ChatGPT output on embeddings from time-stamped BERT models (Sarkar (2025))
- Similar to equation (6), bust using BERT models without lookahead bias
- In principle, fitted value from projection uses only variation without lookahead bias

When comparing generated belief series, univariate regression results would be useful

- E.g. Regress GPT-3.5 on BERT (or WSJ vs. NYT) and report coefficient & R^2
- Tells us if if series capture same variation
  - Complementary to telling us if they have same correlations with target series

Sharpe ratios for sentiment trading strategy would be useful to gauge economic significance

Conclusion

Very interesting paper

Innovative methodology to tackle difficult & important problem

- Suggestions on how to open black box and shed more light on how generated beliefs are formed