

Large Language Models for Financial and Economic Prediction

Introduction

Generative AI and large language models (LLMs) are driving a significant technological shift in finance research. Recent studies in academic finance and economics have begun leveraging state-of-the-art LLMs (such as OpenAI's GPT-3.5/ChatGPT, GPT-4, and related transformer models) to predict economic and financial outcomes from unstructured data like news text and financial reports. Early findings are striking: LLM-based text interpretations can predict stock price movements out-of-sample better than traditional text analysis methods, forecast corporate fundamentals (like earnings changes) with accuracy exceeding human analysts, and even improve macroeconomic forecasts such as inflation trend predictions relative to professional forecasters.

LLMs for Stock Market Prediction

One of the earliest applications of LLMs in finance has been using textual news data to predict stock returns. Traditional “text-as-data” approaches relied on bag-of-words counts or sentiment dictionaries (e.g. Loughran-McDonald sentiment word lists) to gauge news tone, but those methods ignore context and often misinterpret nuanced language. Recent research shows that transformer-based LLMs can read news with human-like comprehension and extract signals that better forecast stock performance. Several notable studies illustrate this progress:

Lopez-Lira & Tang (2023)

In one of the first papers on this topic, the authors demonstrate that ChatGPT can interpret news headlines to predict individual stock price movements. Without any special financial training, ChatGPT's inferred “sentiment scores” from headlines significantly predict next-day stock returns out-of-sample. These LLM-derived signals outperform traditional sentiment metrics, and the predictability is strongest for harder-to-arbitrage stocks (e.g. smaller firms)

and after negative news. The results suggest that advanced LLMs can decipher subtle news context (including bad news that investors underreact to) and translate it into profitable forecasting signals.

Chen, Kelly & Xiu (2024)

This Chicago Booth-led study leverages ChatGPT and LLaMA embeddings to analyze full news articles (not just headlines) across 16 global equity markets. By obtaining contextualized representations of news text, they show that stock prices respond slowly to bearish or bullish news content – indicating market inefficiency – and that return predictions based on LLM text embeddings strongly outpace those based on technical signals (like past returns) or simpler NLP methods. Notably, the LLM-based approach shines on articles with complex language (for example, pieces containing negation or nuanced narratives), where it correctly interprets tone when simpler bag-of-words models would fail. This comprehensive evidence across 13 languages confirms that news content worldwide has predictive power for returns when understood in context by an LLM.

Chen, Tang, Zhou & Zhu (2024)

Focusing on the aggregate stock market, this paper asks whether LLMs can extract signals from news to predict overall market returns and macro trends. The authors use ChatGPT to read Wall Street Journal front-page articles and summarize the mix of “good news” vs “bad news” each day. They find that a higher ratio of positive news identified by ChatGPT signals improving economic conditions and predicts higher monthly stock market returns, consistent with theories that investors underreact to good news (especially during downturns or high uncertainty). In contrast, while negative news content correlates with contemporaneous market drops, it had no predictive power for future returns – perhaps because bad news is quickly absorbed by the market.

Importantly, ChatGPT’s news-based market signal outperforms other methods: a domestic Chinese LLM (“DeepSeek”) and standard sentiment models both underperformed ChatGPT, likely because ChatGPT’s extensive English training allows it to better grasp economically relevant content. At present, ChatGPT appears to be the only model (among those tested) that can consistently capture the news information linked to the market risk premium. This underscores the unique advantage of frontier LLMs in understanding financial news.

Together, these studies provide compelling evidence that LLMs can convert textual news into superior return-prediction signals. By comprehending context, negation, and narrative nuance, LLMs like ChatGPT avoid the pitfalls of simpler sentiment metrics. The result is that news-driven strategies based on LLM analysis achieve higher predictive accuracy and Sharpe ratios, indicating economically significant improvements in forecasting stock returns. This line of

work – spanning both individual stock predictions and broad market timing – shows LLMs adding value in an area central to asset pricing and investment management.

LLMs in Macroeconomic Forecasting and Policy Analysis

Researchers are also exploring LLM applications in macroeconomics – from forecasting aggregate economic indicators to interpreting central bank communications. The high dimensionality and unstructured nature of economic data (e.g. news, policy statements) make it a ripe area for LLM-driven innovation. Recent studies offer mixed but encouraging evidence that LLMs can capture signals that improve macroeconomic predictions or at least help explain economic expectations:

Bybee (2023)

This study introduces a novel approach to generate a “survey of economic expectations” using an LLM. Essentially, Bybee uses GPT-3.5 to read historical news articles (Wall Street Journal, 1984–2021) and asks the model to act as if it were forecasting various macroeconomic and financial variables based on the news. The resulting AI-generated expectations turn out to closely track real-world survey forecasts, such as the Survey of Professional Forecasters and the American Association of Individual Investors sentiment.

Intriguingly, the LLM’s biases mirror human forecasters’ biases: for example, the AI’s macro forecasts exhibit the same under-reaction patterns that professional consensus forecasts do (failing to fully adjust to new information). Likewise, the LLM’s stock return expectations are extrapolative and overly optimistic (and consequently negatively correlated with future realized returns), much like the documented biases in investor surveys.

These results suggest that large language models can be used to understand and even replicate human expectation formation in economics. By “surveying” an LLM, researchers can generate plausible expectation series without directly polling humans. Moreover, because the LLM’s outputs remain correlated with actual survey data even on news samples from after the model’s training period, the evidence indicates the model is genuinely generalizing patterns rather than just memorizing past news.

Faria-e-Castro & Leibovici (2024)

Analysts at the Federal Reserve Bank of St. Louis tested whether LLMs could improve inflation forecasting. They employed Google’s PaLM, a 540-billion-parameter transformer model, to produce inflation projections and compared them to professional forecasters’ predictions. The LLM’s forecasts were surprisingly strong: the Fed researchers report that the LLM estimates inflation trends more accurately than professional analysts in most years and across nearly all forecast horizons during 2019–2023. In some cases, the AI model picked up signals of

persistent inflation sooner than human forecasters did, providing an early warning of rising price pressures.

This suggests that LLMs (with their ability to ingest vast amounts of text and data) can discern macroeconomic undercurrents that experts might overlook or downplay. That said, the study also flags typical LLM challenges: the models are “black boxes” and prone to instability or reasoning errors, so their predictions must be interpreted with caution. Nonetheless, this work – along with parallel efforts at central banks in Europe – points to the potential of LLMs to augment traditional macroeconomic nowcasting and forecasting tools.

Conclusion

In summary, the emerging body of academic research indicates that large language models are powerful new tools for prediction in finance and economics. From forecasting stock returns using news text to analyzing accounting reports and macroeconomic news, LLMs have demonstrated an ability to extract predictive signals that often beat traditional benchmarks.

The common theme is that LLMs can interpret complex, contextual information – whether it’s a news article, a financial statement, or a central bank statement – far more effectively than earlier techniques, thus improving predictions of economic outcomes.

And as seen with ChatGPT’s success in parsing news and statements, these models can capture qualitative nuances (tone, narrative, context) that humans and simpler algorithms often miss, converting them into quantitative forecasts. Going forward, we can expect LLMs to become increasingly integrated into financial economics research. Experts see generative AI as a major technological shock for the field⁴, opening new avenues for analyzing qualitative data and perhaps even altering how we understand market efficiency and investor behavior.