Expected Returns and Large Language Models

Summarized by： Weihao Zhao

04.18.2025

1 Research questions

Can LLMs improve the predictive ability of expected stock returns by mining financial news texts?

2 Why are the research questions interesting?

• Financial market information mainly comes from text, but it has not been fully utilized in financial economics.
• Traditional text methods have obvious limitations:
    - BoW ignores context and cannot handle complex phenomena;
    - The modeling and prediction capabilities of the word vector model are limited.
• LLMs are widely successful in NLP tasks, but their evaluation in financial forecasting is still insufficient.

3 What is the paper's contribution?

Literature on Financial Text Research
Prior: Focusing on specific data sources (10-K files), feature expression is rough/frequency driven.
    - High dimensional sparse representation leads to low statistical efficiency;
    - Lack of cross corpus and cross task transfer ability(BoW, Word2Vec, SESTM)
    - The technical threshold for researchers is relatively high (NLP)
This: Extracting contextual embeddings using LLMs while preserving semantic structure;
    - Using pre-trained models to enhance predictive ability;
    - Simplify modeling using embeddings;
    - Using open-source APIs to lower application barriers.

4 What hypotheses are tested in the paper?

H1: The tone of news text (positive/negative) can predict short-term stock price changes.
H2: Using news embedding vectors can directly predict short-term cross-sectional expected returns of stocks.

a) Do these hypotheses follow from and answer the research questions?
• LLMs can be widely deployed in the financial industry, such as fund companies, investment banks.
Do these hypotheses follow from theory or are they otherwise adequately developed?
• Challenge hypo of market efficiency and provide evidence to support 'limited attention' and 'reaction delay'.
• Promoting upgrading of financial text analysis: From 'word frequency driven' to 'semantic understanding'.

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

Market coverage: wide range of countries and languages; The text content, timestamp, and associated stock code.

6 Dependent and independent variables: comment on the appropriateness.

There is noise in the labels, but paper effectively guide model to recognize text emotions in large samples.

7 Regression model specification: comment on the appropriateness.

Logistic/ridge regression, combined with supervised learning ideas; In line with practical operations.

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

Without considering actual execution factors such as transaction costs, slippage, liquidity, etc.

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

- Introducing multimodal data and factor fusion:
  - Combining text features with structural factors such as financial indicators, factor model beta, ESG ratings;
  - Check if text embedding enhance traditional asset pricing models.

10 Summarize the similarities, differences, and correlations between literature.

关联维度	具体说明
方法互补	《Uncovering Information》基于信息论提出衡量“文本信息量”的理论框架，可为《Expected Returns》和《FSA-LLM》中构建信号或嵌入提供原理支持。
目标互补	《Expected Returns》关注收益预测，《Uncovering Information》关注信息来源与反应机制，《FSA-LLM》关注基本面盈利预测，分别对应资产定价、信息处理、公司分析。
实证补充	三篇文章均展示 LLM 在金融语境中的强预测性与解释力，从收益预测、市场反应到盈利能力形成全方位验证，具有互为支持的实证效果。
市场有效性挑战	三者均发现 LLM 可识别出市场未即时反映的重要信息，提供对强式市场有效性假说（EMH）的系统挑战。
共同对比对象	三文均对比 LLM、专业分析师与传统机器学习模型（如 Logit、ANN），发现 LLM 兼具类人推理力与模型通用性，在复杂任务中具有显著优势。
理论建构互补	《Expected Returns》验证了 LLM 提取的信息具有投资价值；《Uncovering Information》定义了信息的本质与位置；《FSA-LLM》强调 LLM 可以模拟分析师的思维流程。

维度	Expected Returns	Uncovering Information	FSA with LLMs
研究对象	新闻文本（多语言/市场）	公司公开披露（10-K）	财务报表中结构化数字
核心任务	预测未来股票回报、构建多空投资组合	衡量披露文件中“新信息”的位置和强度	预测未来盈利方向（EPS 上升或下降）
输入数据类型	非结构化新闻文本 + 时间戳	结构化与非结构化文本（全类型披露文件）	标准化的纯财务数字（无文本）
输出变量	横截面回报率、情绪、夏普比率等	信息得分、市场反应（如绝对收益、交易量）	EPS 变动方向（二元分类）与预测准确率
技术亮点	使用 GPT/LLaMA 等嵌入建构收益预测因子	基于信息论与 LLM 概率测量“出乎意料性”	设计 Chain-of-Thought 提示模拟分析师思维流程
预测模型	岭回归、逻辑回归、多空组合策略	不直接预测收益，分析信息的形成与反应过程	GPT/ANN 对比，重点考察推理路径/预测效果
实证发现	LLM 构建信号具有显著超额收益与高夏普比率	大量重要信息来自 8-K 及其附录，且持续释放	CoT 提示下 GPT 预测准确率超越分析师与 NN
研究边界	挑战 EMH，强调新闻预测收益的可交易性	信息提取/注意力限制/披露形式结构性影响	LLM 是否具备类人推理能力，可否替代分析师