


---

This is the **published version** of the bachelor thesis:

Barrado Díaz, Adrián. *Can LLMs predict the market? ChatGPT vs DeepSeek in financial sentiment analysis*. Treball de Final de Grau (Universitat Autònoma de Barcelona), 2025 (Administració i Direcció d'Empreses)

---

This version is available at <https://ddd.uab.cat/record/321177>

under the terms of the  license.



# 1. Introduction

- ✓ **News effect:** headlines can trigger price movements.
- ✓ **Traditional sentiment analysis:** reliance on lexicons.
- ✓ **LLMs understand context:** nuance, irony & ambiguity.
- ✓ **Research gaps:** *can LLMs predict price movements given the news published?*

*Every day, investors react to news but, can an AI beat them?*

**Extract sentiment from financial news in a way that it correlates with short-term stock movements.**

- ✔ Quantify sentiment
- ✔ Assess correlation
- ✔ Compare models



## 4. Null hypothesis

No positive correlation between daily/weekly sentiment scores and subsequent stock returns.

## Alternative hypothesis

*Positive correlation exists.*

## Tetlock (2007)

## Giving content to investor sentiment: *The role of media in the stock market*

Media pessimism predicts short-term downward price pressure: predictive signal.

**Loughran & McDonald (2011)**

## When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks.

Developed a domain-specific financial sentiment dictionary.

**López-Lira & Tang  
(2023)**

# Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models

Using GPT-3.5 to classify sentiment as good or bad

## 5. Methodology

## Data collection

Google News, “Repsol”  
keyword, 2024

## Relevance filter with DeepSeek

From 5.244 to 2.555 articles

## Models

ChatGPT: GPT-4o  
DeepSeek: V3

**Technique:** zero-shot prompting

- Sentiment (-1 to 1)
- Keyword & Reasoning

## Variables

Dependent: subsequent returns  
Independent: sentiment scores.  
Control: IBEX35, Brent, Dividend.

## Analysis

Correlations, OLS regressions, abnormal returns, interaction effects, binary variables.

## Core findings

- **No** statistically significant **correlation**
- Low  $R^2$ , **minimal explanatory power**
- Control variables influence **as expected**
- **Isolated** alignment in DeepSeek interactions

## Key discussion

- No predictive power from sentiment alone **in this setting**
- **Validity:** control variables worked, sentiment didn't
- Interaction effects are **weak and isolated**
- López-Lira & Tang (2023) vs Zhang et al. (2023): **task-based**

## 7. Conclusion

- ✓ LLM-generated sentiment **does not reliably predict** short-term returns in this setting.
- ✓ The methodology is **valid**, but sentiment **lacks independent signal**, predictive value may depend on model tuning, context, and datasets

## References

- ✓ López-Lira, A & Tang, Y (2023), Can ChatGPT forecast stock price movements? return predictability and large language models. SSRN Electronic Journal
- ✓ Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance