TIME SERIES FOUNDATION MODELS IMPROVE LLM DECISIONS: A CASE STUDY IN STOCK TRADING

Shifeng Xie[1,*] Ziwei Li[2,*] Themis Palpanas[1] Peilin Zhao[3] Chenghao Liu[4]

[1] Université Paris Cité, France [2] King Abdullah University of Science and Technology (KAUST), Saudi Arabia
[3] Shanghai Jiao Tong University, China [4] Datadog

Key Findings

- TSFM forecasts significantly improve LLM-only trading decisions.
- This gain is robust across forecast formats.

1, Motivation & Core Hypothesis

- TSFMs: strong at numerical forecasting
- LLMs: strong at high level reasoning
- **Hypothesis:** better time series signals ⇒ better sequential decisions

2, Trading-Agent Setup

- Universe: MAG7 + CASH
- Decision agent: Qwen-30B
- TSFM: Chronos2, TimesFM2.5, Moirai2, Toto
- Inputs: Fundamentals + price history
- Output: daily weights over 8 assets

3, Evaluation Protocol

- 3 non overlapping periods
- 30 day horizon for each run
- Observed close price execution
- No transaction costs or market impact
- Metrics: cumulative return, annualized return, Sharpe, Sortino, max drawdown

4, TSFM Signal Formats

Format	Meaning
1	Raw 30-day prices
2	30-day return ratios
3	Multi-horizon returns
4	30-day price quantiles
5	30-day return quantiles
6	Multi-horizon quantile returns

5, Result: TSFM Improve LLM Decisions

+0.47% mean improvement over LLM-only

- Paired samples: 72
- 95% CI (t): [0.21, 0.74]
- 95% CI (bootstrap): [0.22, 0.73]
- One-sided t-test: $t = 3.57,\ p = 3.22 \times 10^{-4}$
- Effect size: Cohen's $d_z = 0.42$

TSFM forecasts yield statistically significant and practically meaningful gains.

6, Format Robustness

- One-way ANOVA: $F = 0.85,\ p = 0.52$
- **No significant difference** across the 6 presentation formats
- Improvement comes primarily from numerical predictive content
- Not from prompt-level formatting tricks

Format 3 has the best point estimate, but not significant after Holm correction.

7, No Look-Ahead Safeguards

Leakage prevention

- Fundamentals available at the decision time
- Price history truncated at date t
- TSFM context truncated at date t
- Evaluation windows start after the Qwen release date

Time Series in the Age of Large Models Workshop at ICLR 2026

Contact: shifeng.xie@telecom-paris.fr

eosc | DataGEMS HAR.S.H. HARdware-aware Similarity searcH European Commission Ελλάδα 2.0 ΕΘΝΙΚΟ ΣΧΕΔΙΟ ΑΝΑΚΑΜΨΗΣ ΚΑΙ ΑΝΘΕΚΤΙΚΟΤΗΤΑΣ Με τη χρηματοδότηση της Ευρωπαϊκής Ένωσης NextGenerationEU ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ Υπουργείο Παιδείας, Θρησκευμάτων και Αθλητισμού