Large Language Models in Fundamental Equity Research

Xizhi Fang Yilu Pan Christelle Saad Nicholas Wong
Master of Finance 2026 – Finance Lab 2025

MIT MANAGEMENT SLOAN SCHOOL

PanAgora

Project Objective

Use LLM to predict EPS change and surprise. Compare performance with XGBoost and analysts forecasts. Build a portfolio based on these predictions.

Data Collection

	Financial Statements	Global Factors	Analysts Forecast
Use	Feeding LLM and XGBoost	Feeding XGBoost	Comparison with LLM and ML performance
Frequency	Annual	Monthly	Monthly
Source	WRDS - Compustat	WRDS – Contributed Data Forms	WRDS – IBES

LLM's Models

Meta-Llama-3-8B-Instruct	DeepSeek-R1-Distill-Llama-8B	DeepSeek-R1-Distill-Qwen-14B

Chain-of-Thought Prompting

Set Scenario — You are a financial analyst specialized in analysing accounting data…

Explain Input Format — The format of the two tables is as follows….

Detailed Intructions — Step 1: Begin by a trend analysis…
Step 2: Perform a ratio analysis…
Step 3: Finally, predict whether the EPS is likely to increase or decrease over the next fiscal year

Explain & Demo Output Format — Structure your output as follows:
[OUTPUT STRUCTURE START]
Trend analysis: [TREND ANALYSIS HERE]
Ratio analysis: [RATIO ANALYSIS HERE]
Direction: [DIRECTION HERE]
Magnitude: [MAGNITUDE HERE]
Confidence: [CONFIDENCE HERE]
Rationale: [RATIONALE HERE]
[OUTPUT STRUCTURE END]

Data — Anonymized Balance Sheet and Income Statement

Example Output

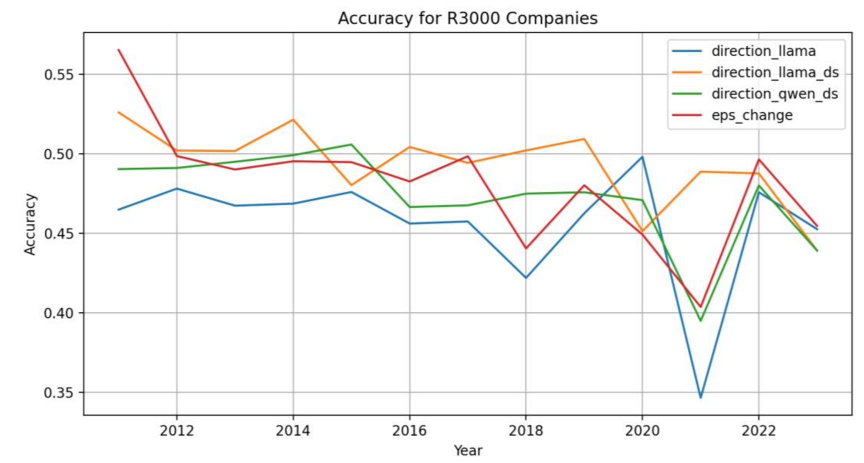
Trend analysis: The company has shown consistent growth in sales and net income over the past three years. Cash and short-term investments have increased significantly, indicating improved liquidity. […]

Ratio analysis: Key financial ratios reveal a healthy company with strong liquidity and profitability. The current ratio and quick ratio are within acceptable ranges, indicating good short-term solvency. Profitability ratios such as gross, operating, and net profit margins have increased, demonstrating improving profitability. The debt-to-equity ratio is moderate, suggesting manageable leverage.

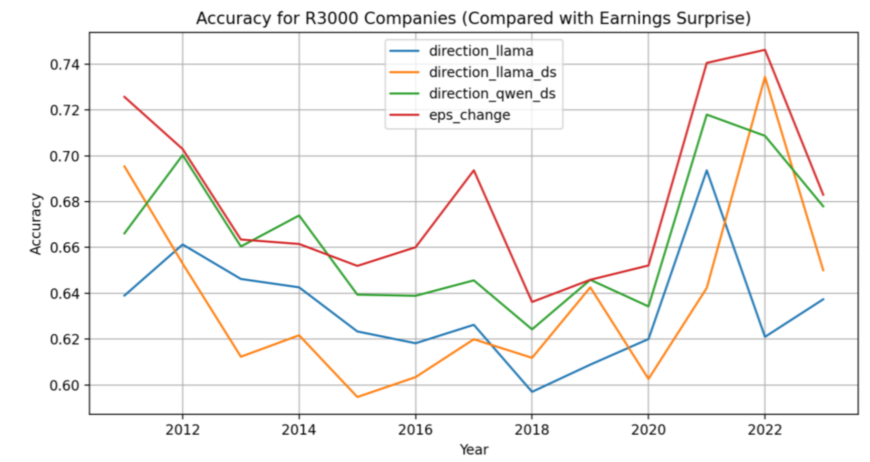
Direction: Increase **Magnitude:** Moderate **Confidence:** High

Rationale: The company's consistent growth in sales, net income, and EPS, along with improved liquidity and profitability ratios, supports the prediction that EPS will increase in the next fiscal year. […]

LLM's Performance



EPS Change Prediction: Low Accuracy, no significant difference across models



Surprise Prediction: Higher Accuracy, no significant difference across models

Correlation Matrix of Model Prediction

	direction_llama	direction_llama_ds	direction_qwen_ds	eps_change
direction_llama	1	0.42	0.5	0.47
direction_llama_ds	0.42	1	0.54	0.49
direction_qwen_ds	0.5	0.54	1	0.59
eps_change	0.47	0.49	0.59	1

Correlation between models is not very high => LLM's predictions results can still be useful

XGBoost Targets

Target 1: Binarized EPS change
Flags whether a company's EPS (earnings per share) has increased compared to the previous period

A value of 1 means EPS went up; 0 means it declined or stayed the same

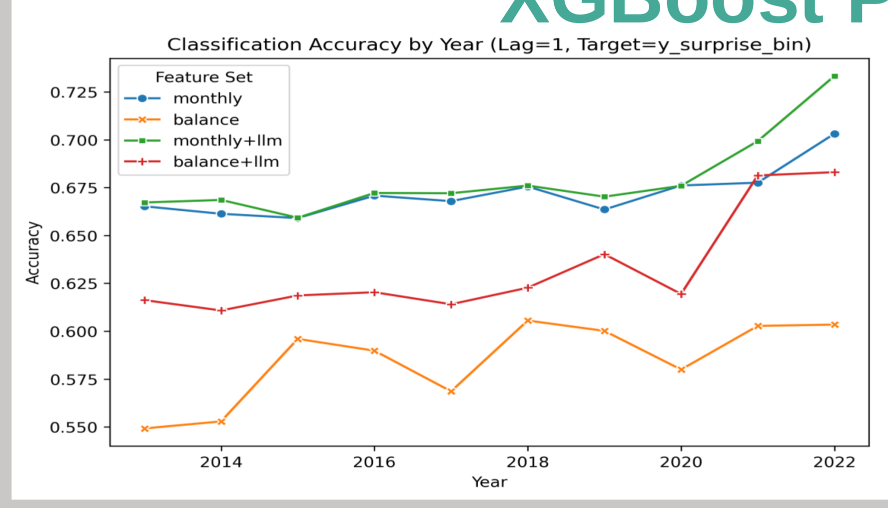
Target 2: Binarized earnings surprise
Flags if the reported earnings exceed market expectations (Earnings Surprise = Actual EPS - Expected EPS)

A value of 1 means the firm's earnings surprised on the upside; 0 means they met or fell below forecasts

XGBoost Features

Monthly Factors	Balance Sheet/Income Statement
Monthly Factors + LLM embeddings	Balance Sheet/Income Statement + LLM embeddings

XGBoost Performance



The highest accuracy is approximately 68%, reached using the global monthly factors (Base and LLM-enhanced features).

Trading Strategy

The strategy ranks stocks based on predictive signals:

- Top 33% (highest signal values) → Long
- Bottom 33% (lowest signal values) → Short
- Middle 33% → No position (neutral)

Each stock in the portfolio gets an equal share of the capital (equal weighted portfolios).
Positions are held for one month and then re-evaluated at the end of the month.

- At the end of each month, the strategy re-ranks stocks based on updated signals.
- If a stock remains in the same tercile, the position is maintained.
- If a stock moves out of the long/short tercile, it is closed.

Strategy Performance

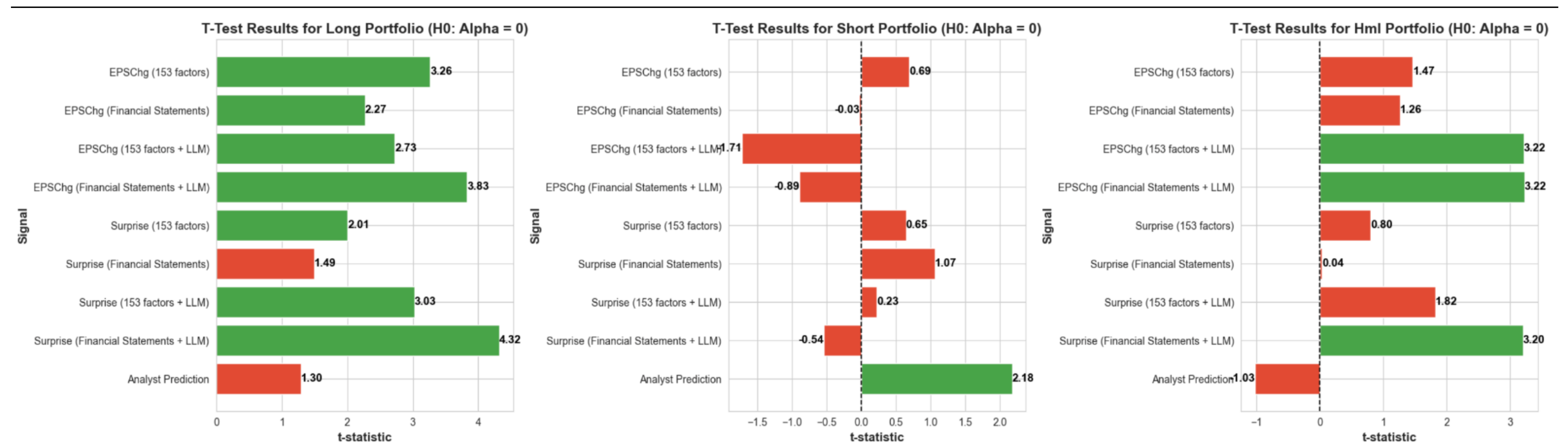
Metrics:
Alpha = (Portfolio Return) - (Equal-Weighted Universe Return)
Information Ratio IR = Mean(Alpha) / Std(Alpha)

		Average Alpha			Information Ratio		
Side		Long	Short	HML	Long	Short	HML
EPS Change	EPSChg (153 factors)	0.12	0.03	0.09	1.09	0.23	0.49
	EPSChg (Financial Statements)	0.09	0	0.09	0.76	-0.01	0.42
	EPSChg (153 factors + LLM)	0.10	0.08	0.18	0.91	-0.57	1.08
	EPSChg (Financial Statements + LLM)	0.16	-0.04	0.20	1.28	-0.30	1.08
Surprise	Surprise (153 factors)	0.09	0.04	0.05	0.67	0.22	0.27
	Surprise (Financial Statements)	0.06	0.06	0	0.50	0.36	0.01
	Surprise (153 factors + LLM)	0.14	0.01	0.13	1.01	0.08	0.61
	Surprise (Financial Statements + LLM)	0.17	-0.02	0.20	1.45	-0.18	1.07
	Analysts prediction	0.05	0.11	-0.07	0.43	0.73	-0.34

Long portfolios perform overall better than short and HML. Short portfolios perform very weakly.

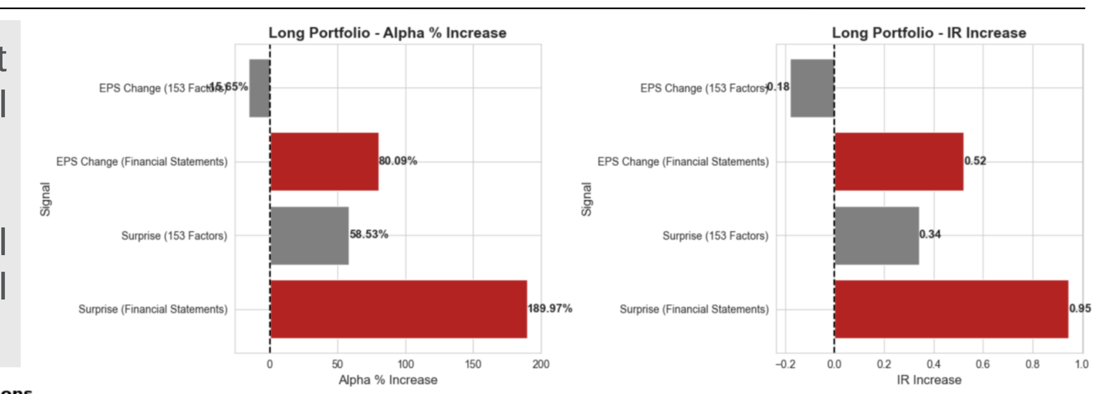
The signal surprise + financial statements + LLM embeddings seems to perform the best.

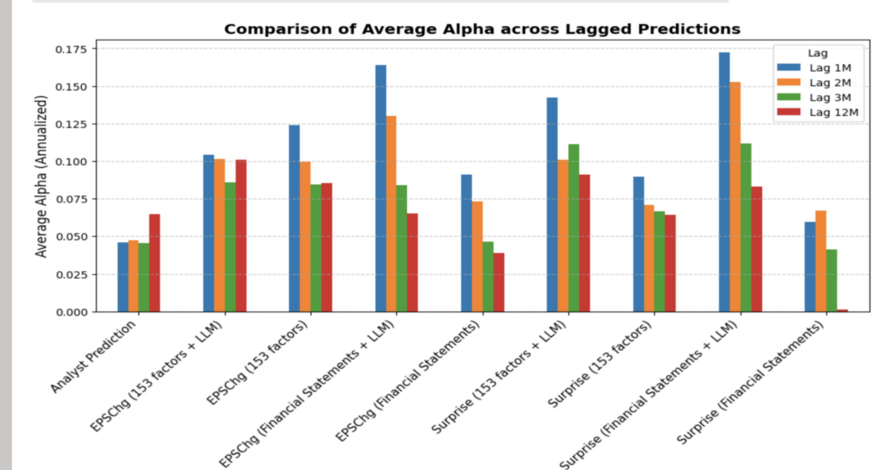
Analysts tend to perform the worst.



LLM inputs have more significant enhancement when used with financial statements than with factors.

LLM 'thoughts' about financial statements work well with financial statement predictors.

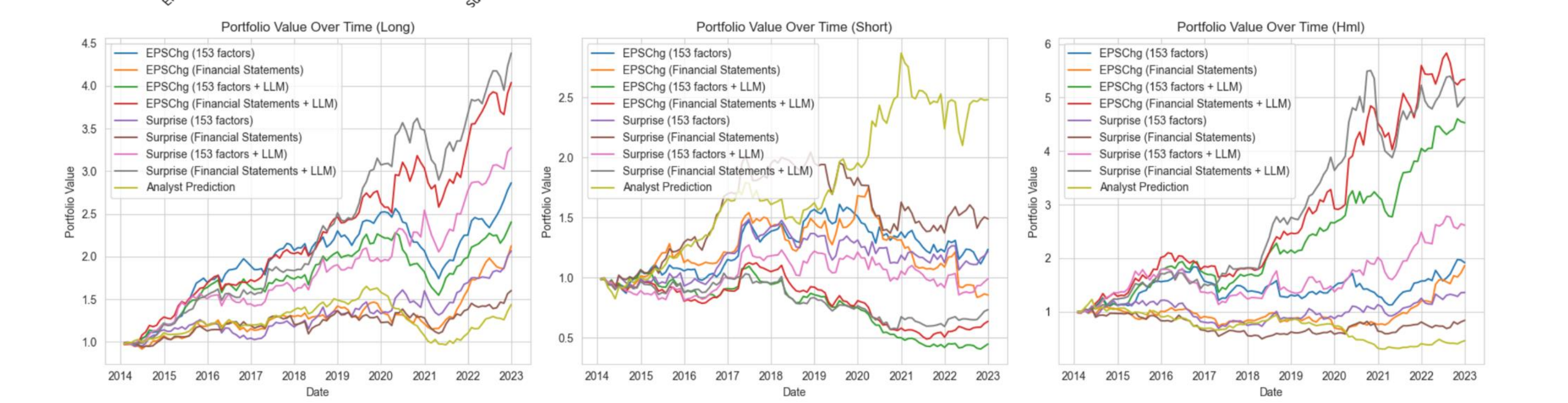




Prediction lag reduces alpha, however, strategies are still robust.

Non-LLM signals struggle more with lags.

Strategies that are based on analysts estimates are consistently weak.



Long portfolios have the best risk-adjusted returns, while short portfolios perform significantly worse. Analyst-based predictions only perform well for short portfolios. LLM-enhanced models consistently outperform their basic counterparts.

LLMs incorporate common economic sense and contextual understanding, allowing ML models to interpret financial statements in a way that aligns with real-world economic reasoning. By integrating LLM embeddings, ML models can enhance generalization in out-of-sample scenarios, leading to more robust and reliable financial predictions.