Bayesian Estimation of Sentiment Impact on Stock Prices

ACM40960 – Projects in Maths Modelling

Saipavan Narayanasamy (24233785) & Shreemadhi Babu Rajendra Prasad (24207575)
M.Sc. in Data and Computational Science

PROBLEM & MOTIVATION

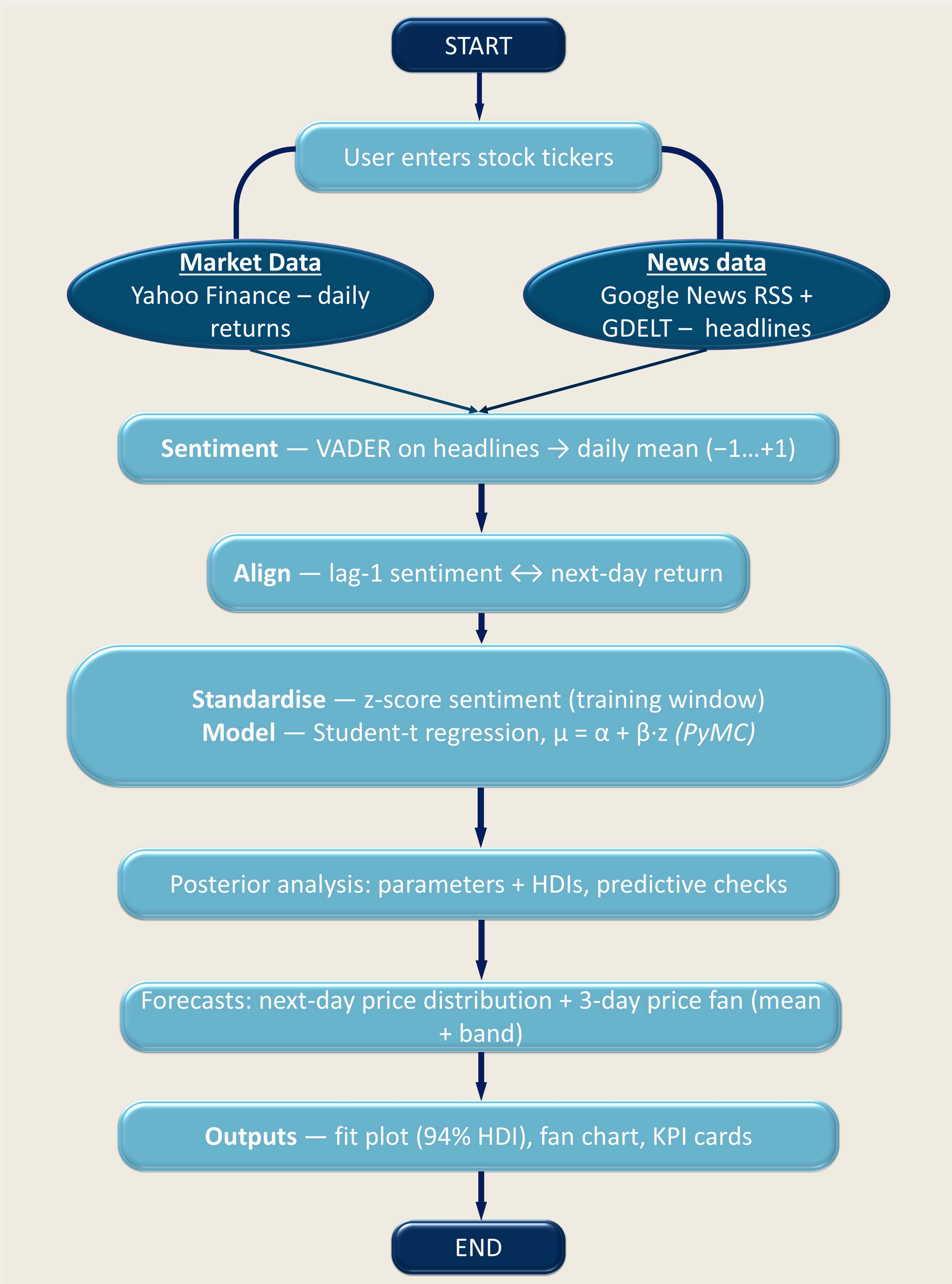
Can daily news-headline sentiment help predict a stock's next-day return?

- Markets react to information flow. Headlines are a fast, public signal.
- Prediction should expose uncertainty (not just point estimates).

GOAL:

- Transform daily headlines into a sentiment score, then measure how yesterday's sentiment (lag-1) influences today's log-return using a Bayesian model.

WORKFLOW: FROM HEADLINES TO PRICE PREDICTION



MODEL: BAYESIAN STUDENT-T REGRESSION

We model daily log-returns rt with heavy-tailed noise:

$$r_t \sim \text{Student-t}(\nu, \mu_t, \sigma), \quad \mu_t = \alpha + \beta z_{t-1}$$

where z t−1 = z-scored lag-1 sentiment.

Priors :

α N(0, 0.02),

β N(0, 0.05),

σ HalfNormal(0.02),

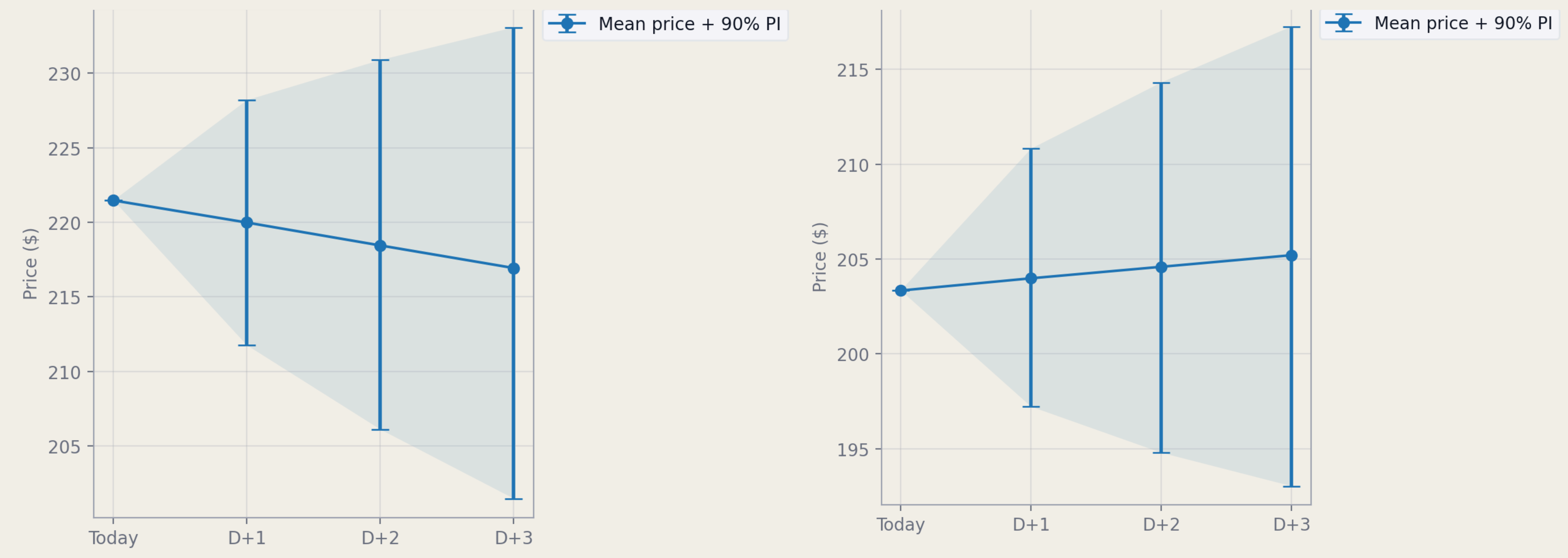
ν Exponential(0.1).

Inference: PyMC (NUTS), target_accept ≈ 0.92; report posterior means & 94% HDIs.

Why Student-t?

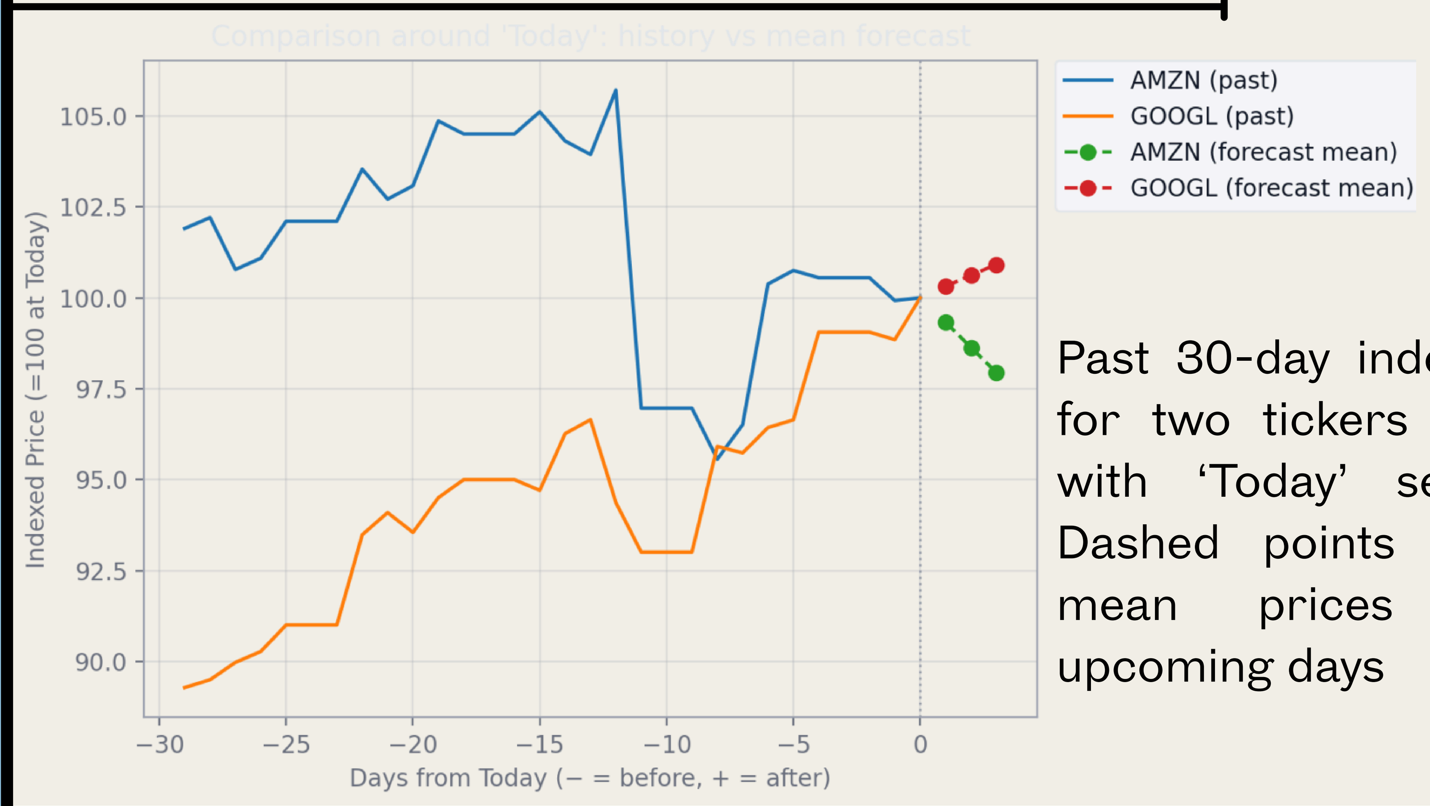
Handles heavy-tailed returns better than Gaussian, improving robustness to price jumps.

PRICE FORECAST



Blue line = mean forecast price over next 3 days; vertical bars = 90% prediction intervals from Bayesian Student-t regression, showing forecast uncertainty.

COMPARISION



Past 30-day indexed prices for two tickers are shown with 'Today' set to 100. Dashed points = forecast mean prices for the upcoming days

COMPARISON TABLE

Ticker	β mean	β HDI low	β HDI high	D+1 mean	D+1 p05	D+1 p95
AMZN	-0.002	-0.006	0.002	220.290	211.790	228.640
GOOGL	0.001	-0.002	0.004	204.140	197.240	211.240

- β (beta) - effect of yesterday's daily sentiment on today's return (posterior mean).
- β HDI low / high - 94% credible range for β.
- D+1 ret. mean - expected return for the next trading day.
- D+1 price mean - implied next-day price ($) from that return.
- D+1 p05 / p95 - 90% prediction interval for next-day price.

LIMITATIONS & FUTURE WORK

- **Data coverage:** News volume and timing vary across tickers, which can bias daily sentiment averages.
- **Model enhancement**: Integrate finance-specific NLP models (e.g., FinBERT, LLM embeddings) using full articles or summaries, not just titles.
- **Advanced modelling:** Explore time-varying or hierarchical Bayesian models to capture changes in sentiment effects across sectors and market.

REFERENCES

- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science.
- Hutto, C. & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for SentimentAnalysis of Social Media Text.
- Salvatier, J., Wiecki, T.V., & Fonnesbeck, C. (2016). Probabilistic Programming in Python using PyMC3. PeerJ CS.