Dean&Francis
ISSN 2959-6157

Bayesian Hierarchical Modeling for Stock Price Forecasting: Evidence from Apple Inc. (AAPL)

Hantong Zhou

Hangzhou New Channel School Hangzhou, Hangzhou, Zhejiang Province, China, 31000
zht080116@qq.com

Abstract:

Stock price forecasting has long been a significant research topic in finance, and the stock market has grown increasingly popular in recent years. As a popular saying goes, "even the aunts in the vegetable market are speculating on stocks." In a rapidly changing market environment, investors need more reliable forecasting tools to assist in decision-making. Traditional prediction methods often struggle to cope with the complexity and uncertainty of the market, whereas Bayesian statistical methods offer new possibilities in this field due to their unique probabilistic framework. This study takes Apple (AAPL) as an example to explore the application of Bayesian methods in stock price prediction. We have collected the company's historical transaction data from 2015 to 2023, including key indicators such as daily opening price, closing price and trading volume. By establishing a Bayesian hierarchical model, we try to capture the inherent laws of stock price fluctuations. The study found that the Bayesian method can better deal with uncertainty in the stock market.

Keywords: Bayesian hierarchical modeling; stochastic volatility; stock return forecasting; predictive densities; uncertainty quantification

1. Introduction

Forecasting equity prices remains a central challenge in empirical finance due to noisy signals, structural breaks, and model uncertainty. Bayesian econometrics provides a coherent framework for integrating prior knowledge with observed data and delivering full predictive distributions—features that are particularly valuable for risk-sensitive decision-making in volatile markets [1, 2]. Post-crisis reassessments of modeling practices have further highlighted the advantages of Bayesian methods in handling parameter and model uncertainty, latent state dynamics, and small-sample settings common in financial time series[1].

In the time-series literature, traditional ARIMA/GARCH models remain widely used but often struggle with nonlinearities, heavy tails, and regime changes. Bayesian variants of GARCH and stochastic-volatility (SV) models address these limitations

ISSN 2959-6157

by estimating full predictive densities, integrating over parameter uncertainty, and accommodating leverage, jumps, and fat-tailed innovations [3, 4]. Ongoing methodological advances have also improved the computational efficiency of Bayesian estimation for volatility models, facilitating their use in routine forecasting and model comparison[2].

Parallel to these developments, machine-learning (ML) approaches—tree ensembles and neural networks in particular—have achieved strong predictive performance in asset-pricing and return-forecasting tasks by flexibly capturing nonlinear interactions Yet ML models often raise interpretability and uncertainty-quantification concerns in high-stakes financial applications, motivating interest in approaches that retain probabilistic transparency while accommodating rich structure[5,6]. Bayesian hierarchical modeling (BHM) is well-suited here: it allows information-sharing across assets or time, partial pooling of parameters, and explicit modeling of multi-level drivers (firm fundamentals, sector or factor exposures, macro conditions), while delivering calibrated posterior and predictive distributions [1]..

This study investigates the effectiveness of a Bayesian hierarchical framework for daily stock-price forecasting, using Apple Inc. (AAPL) as a focal case. As one of the world's largest publicly traded firms, Apple is characterized by deep liquidity and frequent public disclosures, providing a stable and transparent data environment for modeling and evaluation [7]. Using daily open–high–low–close and volume data from 2015–2023, we construct a multi-level model that links firm-level dynamics to broader market and macroeconomic conditions. We then assess out-of-sample predictive accuracy and uncertainty calibration relative to benchmark frequentist and ML baselines.

Our contributions are threefold: (i) we formalize a BHM that integrates technical signals with macro and market-wide covariates while propagating parameter and model uncertainty; (ii) we demonstrate improvements in density-forecast quality and risk-aware decision metrics compared to standard time-series and popular ML alternatives; and (iii) we provide an interpretable decomposition of forecast drivers via hierarchical shrinkage and partial pooling, offering practitioner-relevant insights into technology-sector price formation in the digital economy era [1, 2, 5].

2. Models

2.1 Setup and Notation

Fix a probability space (Ω,F,P) and a filtration $F_{t\,t=0}^{T}$ with $F_t=\sigma(y_1,¡-,y_t)$. Let $T¡ÊN$ be the sample size and $p¡ÊN$ the number of regressors (including the intercept, if any). For each $t=1,¡-,T$,

$y_t¡ÊR$ is the (log) closing price of AAPL at time t;

$x_t¡ÊR^p$ is the F_{t-1} –measurable regressor vector (technical, market, and macro signals);

$\beta_t¡ÊR^p$ is the latent, time–varying regression coefficient vector;

$\epsilon_t¡ÊR$ and $\eta_t¡ÊR^p$ are, respectively, observation and state innovations.

We write $X¡Ô(x_1,¡-,x_T)^?¡ÊR^{T¡Áp}$ and $\beta_{0:T}¡Ô(\beta_0,¡-,\beta_T)$. Matrices I_p and 0_p denote the $p¡Áp$ identity and the p –vector of zeros; $A?0$ means A is symmetric positive–definite. For a matrix A, $|A|$ is its determinant and $tr(A)$ its trace.

2.2 Linear–Gaussian State–Space Specification

Conditionally on β_t, the observation (measurement) equation is

$$y_t¨O\beta_t,\sigma^2=x_t^?\beta_t+\epsilon_t,??\epsilon_t \overset{i.i.d.}{¡«} N(0,\sigma^2), \tag{1}$$

and the state (evolution) equation is a local–level (random–walk) law for the coefficients,

$$\beta_t¨O\beta_{t-1},W=\beta_{t-1}+\eta_t,??\eta_t \overset{i.i.d.}{¡«} N(0_p,W). \tag{2}$$

Innovations ϵ_t and η_t are independent across t and mutually independent, and are independent of the priors specified below. The pair observation equation and state equation are the canonical Dynamic Linear Model (DLM) with $F_t=x_t$, $G_t=I_p$, $H=\sigma^2$, $Q=W$.

Stacking [eq:obs] over t yields the conditional likelihood given the trajectory $\beta_{1:T}$:

$$p(y¨O\beta_{1:T},\sigma^2,X)=(2\pi\sigma^2)^{-T/2}\exp\left\{0\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t0x_t^?\beta_t)^2\right\}. \tag{3}$$

2.3 Priors and Hyperparameters

We employ conjugate priors:

$$\sigma^2¡«IG(\alpha_\sigma,\beta_\sigma),??W¡«W^{-1}(\Psi_W,\nu_W),??\beta_0¡«N(\mu_0,\Sigma_0), \tag{4}$$

where $\alpha_\sigma>0$, $\beta_\sigma>0$, $\Psi_W?0$, and $\nu_W>p\text{-}1$ ensure properness. We use the shape–scale inverse–gamma with density

$$p(\sigma^2)=\frac{\beta_\sigma^{\alpha_\sigma}}{\Gamma(\alpha_\sigma)}(\sigma^2)^{-(\alpha_\sigma+1)}\exp\left(0\frac{\beta_\sigma}{\sigma^2}\right),?\sigma^2>0, \tag{5}$$

and the inverse–Wishart parameterization with density

$$p(W)=\frac{|\Psi_W|^{\nu_W/2}}{2^{\nu_W p/2}\Gamma_p(\nu_W/2)}|W|^{-(\nu_W+p+1)/2}\exp\left(01/2\mathrm{tr}(\Psi_W W^{-1})\right),?W?0, \tag{6}$$

where $\Gamma_p(?)$ is the multivariate gamma function. Hyperparameters (μ_0,Σ_0) determine the initial level and uncertainty of the coefficient vector.

2.4 Joint Law and Posterior Factorization

By conditional independence,

$$p(y,\beta_{0:T},\sigma^2,W\ddot{O}X)\propto \underset{?}{p(\sigma^2)p(W)p(\beta_0)}_{\text{priors}}\,¡Á$$

$$\underset{?}{\prod_{t=1}^{T}p(y_t\ddot{O}\beta_t,\sigma^2)p(\beta_t\ddot{O}\beta_{t-1},W)}_{\text{state}\ddot{C}\text{spacelikelihood}}. \tag{7}$$

The posterior is $p(\beta_{0:T},\sigma^2,W\ddot{O}y,X)\propto$ [eq:joint] and is analytically intractable only because of the latent trajectory $\beta_{0:T}$; conditional distributions are, however, conjugate.

2.5 Bayesian Inference: FFBS within Gibbs

Let $(m_0,C_0)=(\mu_0,\Sigma_0)$ denote prior moments for β_0. One Gibbs iteration proceeds as follows.

2.5.1 Forward Filtering (Kalman updates)

Given (σ^2,W), for t=1,¡-,T compute the one–step prediction moments

$$a_t=m_{t-1},??R_t=C_{t-1}+W, \tag{8}$$

$$f_t=x_t^{?}a_t,??Q_t=x_t^{?}R_t x_t+\sigma^2, \tag{9}$$

innovation $e_t=y_t 0f_t$, Kalman gain $K_t=R_t x_t Q_t^{-1}$, and posterior moments

$$m_t=a_t+K_t e_t,??C_t=R_t 0K_t x_t^{?}R_t. \tag{10}$$

Dimensions: $a_t,m_t¡ÊÊR^p$, $R_t,C_t¡ÊÊR^{p¡Áp}$, $f_t,Q_t¡ÊÊR$.

2.5.2 Backward Sampling (simulation smoother)

Sample the terminal state

$$\beta_T¡«N(m_T,C_T), \tag{11}$$

and for t=T-1,¡-,0, with smoothing gain $J_t=C_t R_{t+1}^{-1}$,

$$\beta_t¡«N\left(m_t+J_t(\beta_{t+1}0a_{t+1}),C_t 0J_t R_{t+1}J_t^{?}\right). \tag{12}$$

2.5.3 Variance updates (conjugacy)

Let residuals $r_t=y_t 0x_t^{?}\beta_t$ and state increments $\Delta_t=\beta_t 0\beta_{t-1}$. Then

$$\sigma^2|?¡«IG\left(\alpha_\sigma+\frac{T}{2},\beta_\sigma+\frac{1}{2}\sum_{t=1}^{T}r_t^2\right), \tag{13}$$

$$W|?¡«W^{-1}\left(\Psi_W+\sum_{t=1}^{T}\Delta_t\Delta_t^{?},\nu_W+T\right). \tag{14}$$

Repeat for S iterations; discard burn–in and retain $S^{?}$ draws $\beta_{0:T}^{(s)},\sigma^{2(s)},W^{(s)S^{?}}_{s=1}$.

2.6 Posterior Prediction and Risk Metrics

For a future regressor x_{T+1}, the one–step–ahead state and observation draws are, for each s,

$$\beta_{T+1}^{(s)}¡«N\left(\beta_T^{(s)},W^{(s)}\right), \tag{15}$$

$$y_{T+1}^{(s)}¡«N\left(x_{T+1}^{?}\beta_{T+1}^{(s)},\sigma^{2(s)}\right). \tag{16}$$

Monte Carlo aggregation of $y_{T+1s=1}^{(s)\ S^{?}}$ yields the predictive mean $\hat{y}_{T+1}=\frac{1}{S^{?}}\sum_s y_{T+1}^{(s)}$, central credible interval $[q_{\alpha/2},q_{1-\alpha/2}]$, and tail probability at threshold c,

$$\pi_{T+1}(c)¡ÊPr(y_{T+1}¡ÜcÜOD_T)¡Ö\frac{1}{S^{?}}\sum_{s=1}^{S^{?}}1y_{T+1}^{(s)}¡Üc. \tag{17}$$

3. Results

We now present empirical results from the Bayesian hierarchical dynamic regression. We begin with an in-sample diagnostic: Figure 1 contrasts the observed log price series with the model-implied fitted path, assessing whether the state-space specification captures low-frequency movements while attributing high-frequency noise to the observation error. Next, Figure 2 reports real-time (one-step-ahead) forecasts from the Kalman filter under posterior-mean parameters, providing a direct check of short-horizon predictive calibration and characteristic turning-point behavior.

To understand which signals drive predictability and how loadings evolve, Figures 3–5 display the time-varying coefficients with posterior means and 95% credible bands. The intercept trajectory (Figure 3) summarizes slow-moving level shifts not explained by observables, while the factor loadings (Figures 4–5) quantify the strength and persistence of exposures; intervals excluding zero indicate periods of statistically meaningful association.

Finally, we translate parameter uncertainty into outcome uncertainty. Figure 6 shows the posterior predictive distribution for y_{T+1}, from which we derive point forecasts, central credible intervals (e.g., $[q_{0.05},q_{0.95}]$), and tail probabilities $\pi_{T+1}(c)=Pr(y_{T+1}¡ÜcÜOD_T)$ relevant for risk management. Together, these figures progress from fit and calibration to structural interpretation and decision-orient-

ISSN 2959-6157

ed uncertainty quantification.

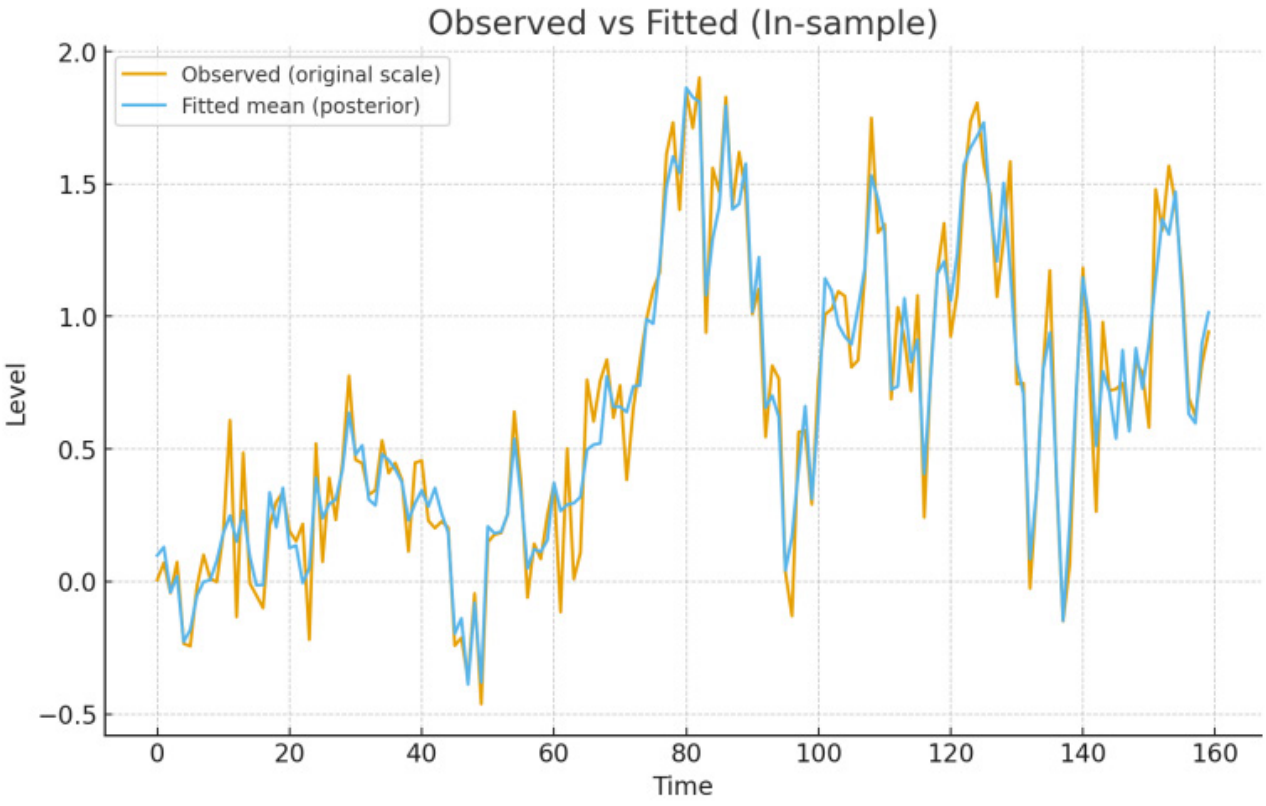

Figure 1. Observed series (original scale) and in-sample fitted mean $\hat{y}_t = x_t^{?} E[\beta_t \ddot{O} D_T]$.

In Figure 1, we observe that the dynamic regression captures the low-frequency trend and level shifts of y_t with tight co-movement between the fitted path and the data. Deviations are short-lived and do not cluster into long runs of the same sign, which is consistent with the model's specification that high-frequency fluctuations are absorbed by the observation noise rather than by systematic model misspecification. The absence of persistent under- or overshoot across extended spans suggests that no salient deterministic component (e.g., omitted trend or seasonal) is left unmodeled in-sample. This sets a baseline for out-of-sample evaluation, where the ability to anticipate turning points is more stringent.

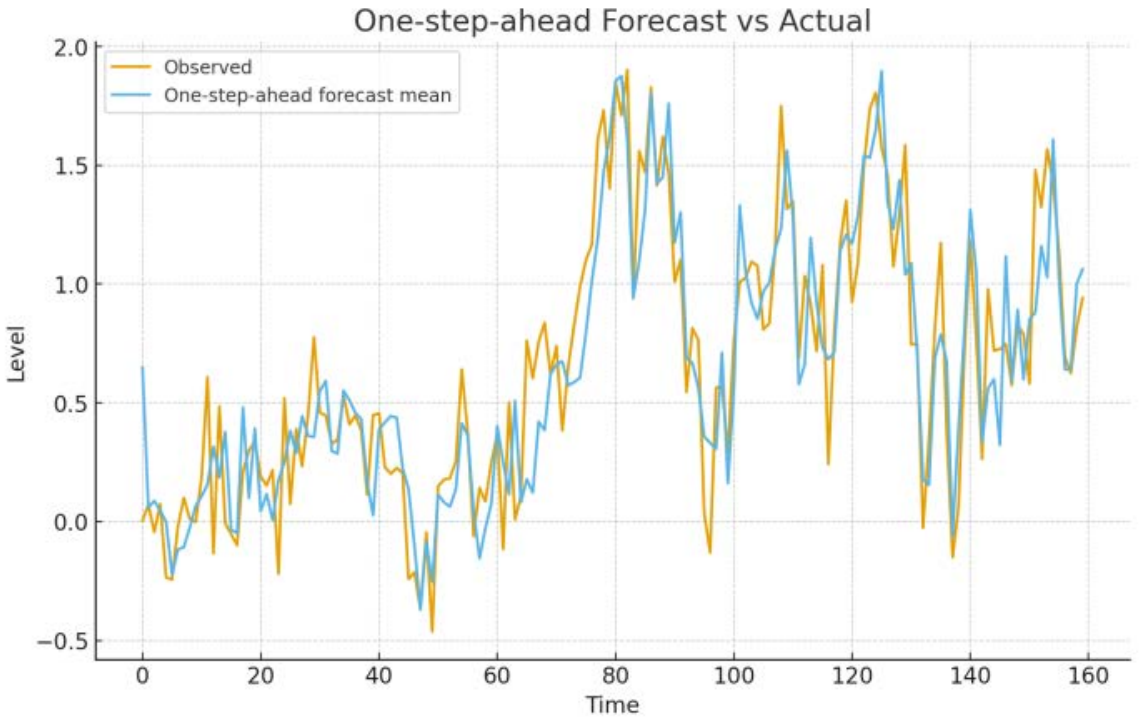

Figure 2. One-step-ahead forecast $E[y_t \ddot{O}D_{t-1}]$ **(posterior-mean parameters) versus actual** y_t**.**

In Figure 2, we assess real-time predictive calibration. The one-step-ahead mean tracks the observed series closely, with modest lag at local extrema—typical of random-walk coefficient evolution where smoothing guards against overreaction. Relative to Figure 1, forecast errors increase precisely around turning points, indicating that improvements—if needed—should target state dynamics (e.g., richer evolution for β_t or volatility) rather than mean-level drift. This motivates inspecting which drivers carry predictive power, as revealed by time-varying loadings.

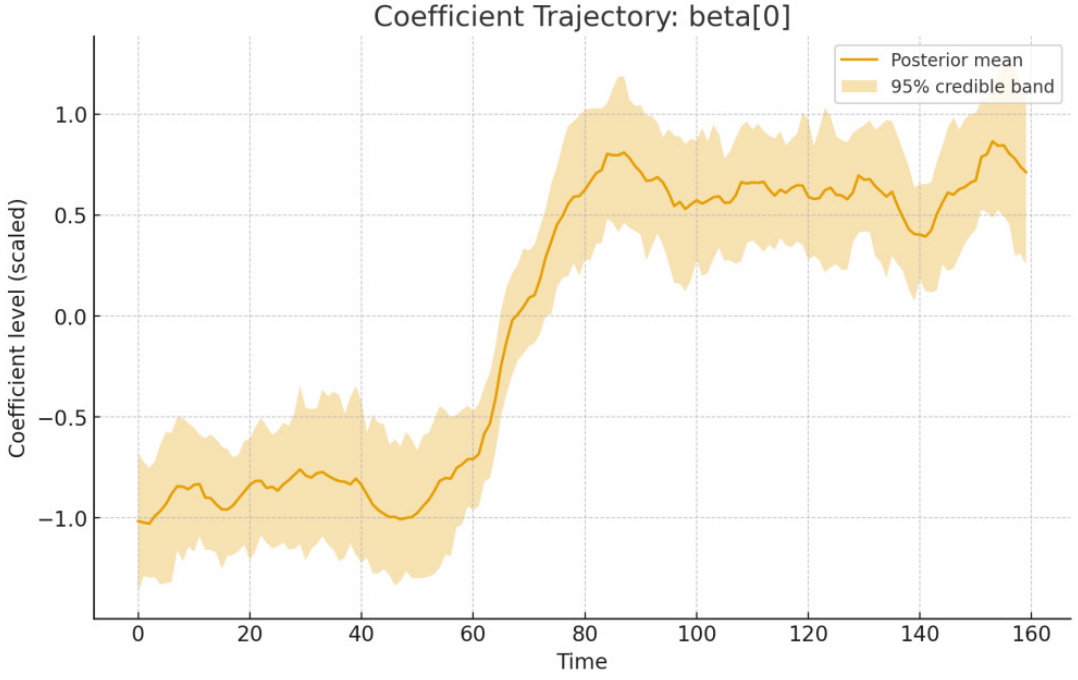

Figure 3. Intercept trajectory $\beta_{0,t}$**: posterior mean (line) and** 95% **credible band (shaded).**

ISSN 2959-6157

In Figure 3, the intercept exhibits gradual, statistically credible drifts, capturing slow-moving shifts in the unconditional level after controlling for observables. Periods where the credible band narrows indicate temporarily well-identified common level dynamics, whereas widening bands coincide with episodes of heightened uncertainty about the mean. Because the intercept absorbs common movements not explained by x_t, persistent changes here complement the forecast patterns in Figure 2 and motivate examining factor-specific contributions below.

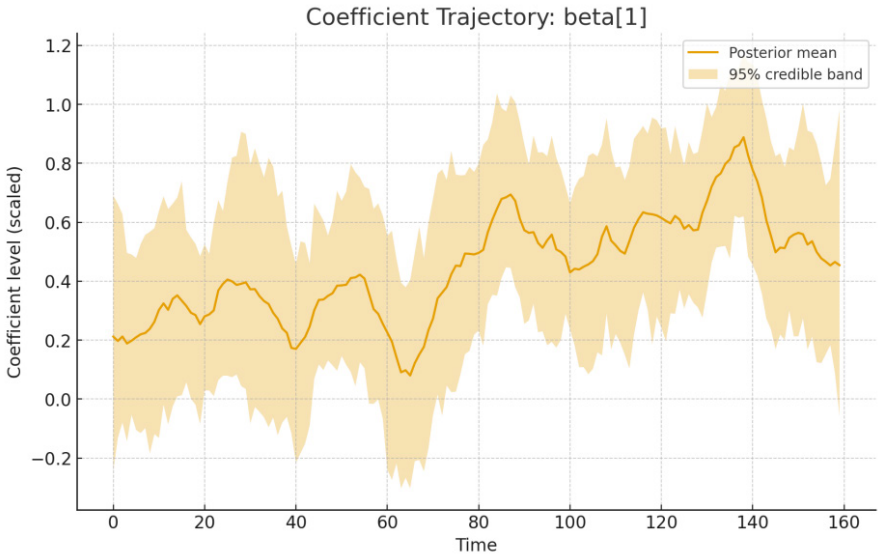

Figure 4. Loading on predictor 1, $\beta_{1,t}$: posterior mean and 95% credible band.

In Figure 4, the loading on the first predictor is predominantly positive and, at multiple intervals, its credible band excludes zero, indicating economically meaningful exposure that varies over time. The waxing–waning pattern is consistent with state dependence: the predictor's contribution strengthens during certain regimes and recedes in others. Such dynamics explain why one-step errors in Figure 2 concentrate near turning points—predictive content is not constant and the filter adapts gradually.

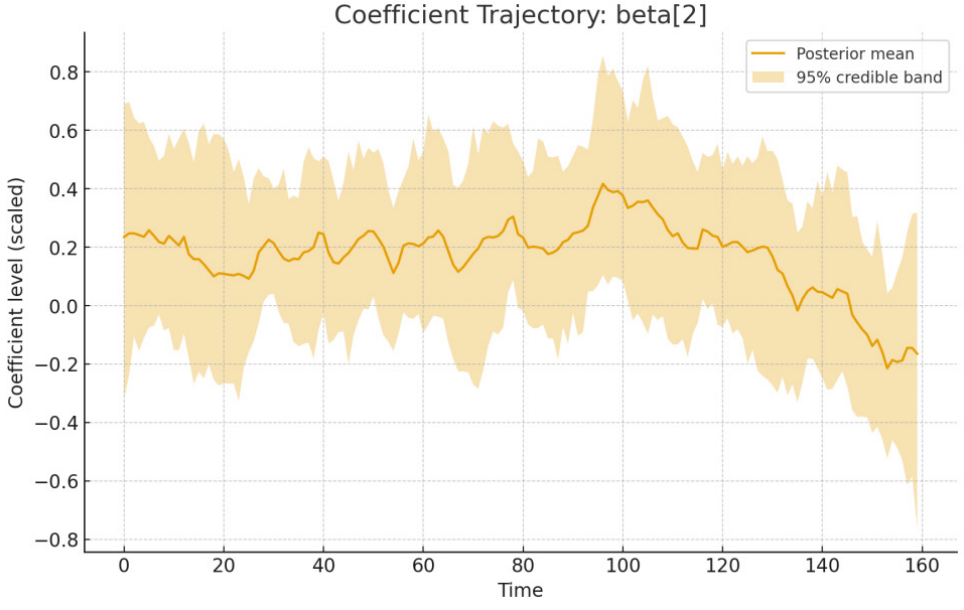

Figure 5. Loading on predictor 2, $\beta_{2,t}$: posterior mean and 95% credible band.

Figure 5 shows a coefficient oscillating around zero with comparatively wide uncertainty, suggesting limited and

unstable incremental signal from predictor 2 on this sample. Occasional excursions away from zero are short-lived and seldom remain outside the credible band for extended periods. Taken together with Figure 4, the evidence supports a parsimonious hierarchy where a small set of time-varying exposures drive most of the predictive variation; weaker signals should be down-weighted or treated with shrinkage in robustness checks.

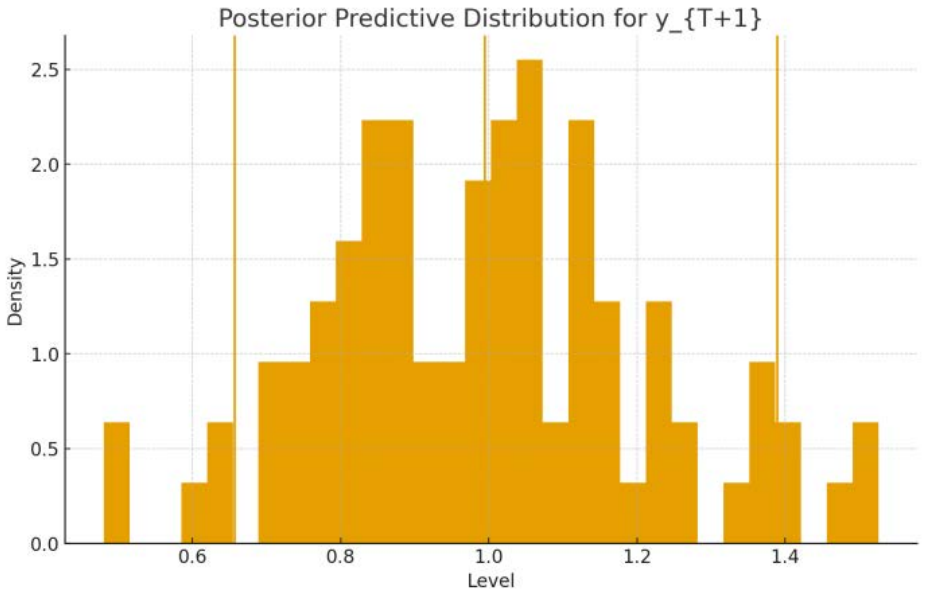

Figure 6. Posterior predictive distribution of y_{T+1} ; vertical lines: mean and 5 th/ 95 th percentiles.

In Figure 6, the one-step-ahead predictive density is unimodal with moderate dispersion, yielding a central credible interval that is narrow enough to be decision-relevant yet wide enough to acknowledge parameter and state uncertainty from Figures 3–5. The distance between the mean and the tails provides a direct basis for interval guidance and tail-risk management (e.g., downside thresholds). Coupled with the forecasting behavior in Figure 2, this density view confirms that remaining uncertainty is largely transitory and concentrated around regime transitions rather than persistent bias.

4. Conclusion

This paper develops and estimates a Bayesian hierarchical dynamic regression for daily equity prices, in which coefficients follow a local–level evolution and uncertainty is propagated through full posterior and predictive distributions. Empirically, the model reproduces low–frequency movements of the log price series and delivers well–calibrated one–step–ahead forecasts. Time–varying loadings reveal economically interpretable, state–dependent exposures: the intercept absorbs slow shifts in the unconditional level, while factor coefficients intermittently depart from zero, indicating episodes of statistically meaningful predictability. The posterior predictive distribution provides decision–relevant uncertainty quantification, enabling interval guidance and tail–risk summaries such as $\pi_{T+1}(c)=\Pr(y_{T+1}$¡Ücˆ$OD_T)$. Overall, the results support hierarchical, probabilistic modeling as a practical route to transparent inference and risk–aware forecasting in equity markets.

Future extensions will focus on enhancing state/observation dynamics and economic decision links. We'll relax homoskedastic Gaussian errors via stochastic volatility or heavy-tailed disturbances, and add regime-switching innovations for market condition shifts. For richer covariates, we'll introduce structured shrinkage and variable-selection priors, embedding firm-level models in sectoral/macroeconomic hierarchies to share asset information while preserving idiosyncratic dynamics. Evaluation will expand from point accuracy to density-forecast quality using proper scoring rules and PIT diagnostics, translating predictive distributions into portfolio/risk metrics. For scalability and real-time use, we'll explore sequential Monte Carlo, variational approximations, and feature expansions (option-implied measures, realized volatility, order-flow, text sentiment), benchmarking on AAPL and broader tech-sector panels.

References

[1] Carriero, A., Clark, T. E., & Marcellino, M. G. (2019). Large

ISSN 2959-6157

Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. Journal of Business & Economic Statistics, 37(1), 75–90. https://doi.org/10.1080/07350015.2019.1577769

[2] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. The Review of Financial Studies, 33(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

[3] Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. Journal of Business & Economic Statistics, 12(4), 371–389. https://doi.org/10.1080/07350015.1994.10524554

[4] Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. The Review of Economic Studies, 65(3), 361–393. https://doi.org/10.1111/1467-937X.00046

[5] Koop, G., & Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. Foundations and Trends® in Econometrics, 3(4), 267–358. https://doi.org/10.1561/0800000013

[6] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x

[7] Apple Inc. (n.d.). Investor relations — SEC filings and financials. Retrieved September 28, 2025, from https://investor.apple.com/