Improving S&P 500 Volatility Forecasting through Regime-Switching Methods

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK
BARNARD

Ava Blake Nivika Gandhi Anurag Jakkula

Motivations

Market volatility forecasts are essential for portfolio management, risk control, derivatives pricing, and regulatory decisions. Traditional models like the Heterogeneous Autoregressive (HAR) model use past volatility over different time horizons (daily, weekly, monthly) to predict future volatility, assuming stable market dynamics with fixed parameters.

However, financial markets often move through distinct regimes, such as calm, crisis, and recovery phases, that cause volatility behavior to change over time. In these settings, the statistical relationship between past and future volatility is not constant, and fixed-parameter models may fail to capture such evolving patterns, reducing forecast accuracy.

Our research addresses these challenges by developing regime-aware forecasting frameworks that detect structural breaks and adapt to changing market conditions. By combining historical realized volatility with forward-looking indicators like the Implied Volatility Index (VIX), our models adjust dynamically to different volatility regimes, improving predictive performance.

Ultimately, this work enhances the robustness and reliability of volatility forecasts, supporting better-informed real-time decisions in risk management and hedging.

Background

The original HAR model, introduced by Corsi [1] in 2009, was designed to capture realized volatility (RV) behavior across multiple time scales (daily, weekly, and monthly) reflecting the activity of different types of investors:

$$RV_t = \beta_0 + \beta_d \cdot RV_{t-1} + \beta_w \cdot \overline{RV}^{(w)}_{t-1} + \beta_m \cdot \overline{RV}^{(m)}_{t-1} + \varepsilon_t \tag{1}$$

where $\beta_0, \beta_d, \beta_w, \beta_m$ are the HAR parameters to be estimated, and ε_t is the error term. Subsequent research has shown that while the HAR model captures important long-memory features, volatility dynamics often shift across distinct market regimes, exhibiting structural breaks and regime-dependent behavior. To better capture these time-varying patterns, regime-switching extensions of HAR have been proposed. For instance, Zhang [2] demonstrates that incorporating regime switching enhances forecasts of Chinese stock market volatility by accounting for structural shifts driven by international market conditions.

Data Collection

Our dataset consists of high-frequency intraday price data for the S&P 500 index (SPX), spanning eleven years from June 2, 2014 to April 29, 2025, sourced via Bloomberg Terminal. Using 5-minute closing prices, we calculate intraday log-returns as:

$$r_{t,i} = \ln\left(\frac{P_{t,i}}{P_{t,i-1}}\right) \tag{2}$$

where $P_{t,i}$ is the price at the i^{th} 5-minute interval on day t. These returns are then aggregated to compute daily realized volatility (RV), adjusted to account for shorter trading days:

$$RV_t = \sqrt{\frac{N}{n}\sum_{i=1}^{n} r_{t,i}^2} \tag{3}$$

Here, n is the number of intraday returns observed on day t (which may vary due to holidays), and N is the standard number of returns in a full trading day—78 for intraday data only, or 79 including overnight returns. This scaling ensures comparability of RV across days with varying lengths.

Feature Engineering

We extend the HAR model by introducing a dual-memory structure that captures both historical volatility patterns and forward-looking market sentiment. Our model applies HAR-style lags to implied volatility features (VIX), allowing the model to respond to shifts in investor expectations.
To build this structure, we engineered:

- Lagged VIX values (daily, 5-day, 22-day): market fear over varying time horizons
- Short-term reversal factor (STR): captures recent return reversals and mean reversion
- Realized kurtosis: measures tail risk and extreme return behavior
- Jump variation: isolates large discontinuous price moves from continuous volatility

This feature design allows the model to adapt across regimes by integrating both behavioral and structural signals, improving forecast accuracy under changing market conditions.

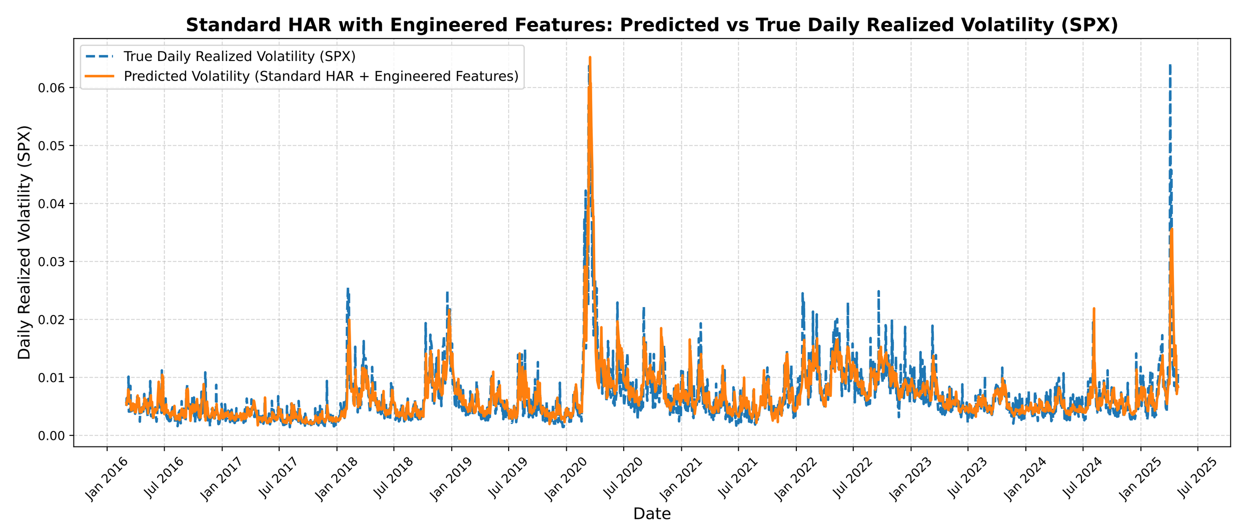

Figure 1. Feature-engineered standard HAR model over the eleven-year period.

Incorporating these additional features with a regime-switching model takes the following form:

$$\begin{aligned} RV_t = \beta_{0,S_t} + \beta_{d,S_t} \cdot RV_{t-1} + \beta_{w,S_t} \cdot \overline{RV}^{(w)}_{t-1} + \beta_{m,S_t} \cdot \overline{RV}^{(m)}_{t-1} + \gamma_{d,S_t} \cdot VIX_{t-1} \\ + \gamma_{w,S_t} \cdot VIX^{(w)}_{t-1} + \gamma_{m,S_t} \cdot VIX^{(m)}_{t-1} + \zeta_{S_t} \cdot KTS_{t-1} + \eta_{S_t} \cdot JMP_{t-1} + \varepsilon_t \end{aligned} \tag{4}$$

where VIX is the Chicago Board of Exchange's implied volatility, KTS is realized kurtosis, and JMP is the jump variation. The model coefficients depend on the unobserved regime state S_t.

Methodology Overview

Traditional regime switching models use techniques such as Markov regime-switching, implementing a Hidden Markov Model (HMM) to capture volatility behavior as the market evolves through differing regimes. Our models build off of existing frameworks and introduce new regime-switching techniques.

1. Soft Markov Regime Switching: We extend HAR by fitting a Gaussian HMM to smoothed volatility and estimating soft regime probabilities. These probabilities are used to weight regime-specific WLS or Ridge regressions within an Expectation-Maximization (EM) loop. Forecasts are blended using forward-propagated regime weights for smooth, robust transitions.
2. Distributional Clustering with Spectral-XGBoost: We detect regime shifts using the Mood test and cluster segments via Wasserstein distances and spectral clustering. Each cluster has an HAR model, with XGBoost assigning regimes at test time. This enables structural break adaptation based on feature distributions.
3. Coefficient-Based Soft Clustering. We extract HAR coefficients from mood-based segments and cluster them using PCA and BGMM to obtain soft regime weights. These weights inform WLS regressions per regime. XGBoost predicts regime probabilities, allowing for smooth, probabilistic forecasts.

Model Comparisons

We evaluated each model's performance across three time periods: Pre-COVID (June 2014-January 2020), COVID (January 2020-September 2020), and Post-COVID (September 2020-April 2025).

Model	MAPE	MSE	# Reg.
HAR	27.09	3.40	1
Markov Soft EM	24.33	3.12	3
Distributional Clustering	27.36	3.87	2
Coefficient Clustering	24.26	3.00	3

Table 1. Model performance during the Pre-COVID period. Divide MSE values by 1,000,000.

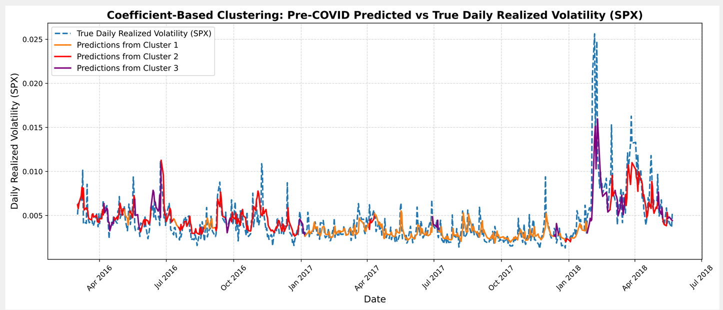

Figure 2. Coefficient-Based Clustering on Pre-COVID Data.

Model	MAPE	MSE	# Reg.
HAR	30.13	35.35	1
Markov Soft EM	31.93	36.22	2
Distributional Clustering	27.27	33.89	2
Coefficient Clustering	38.63	39.54	2

Table 2. Model performance during the COVID period. Divide MSE values by 1,000,000.

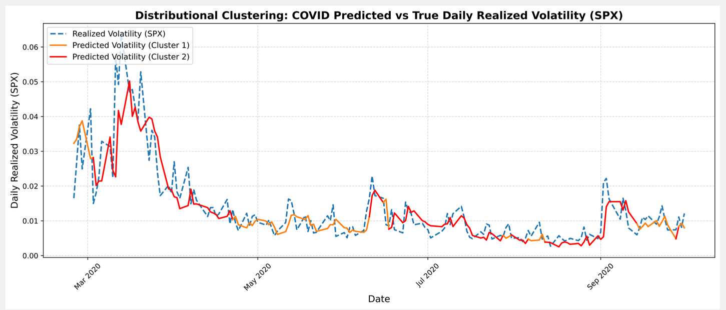

Figure 3. Distributional Clustering on COVID Data.

Model	MAPE	MSE	# Reg.
HAR	23.31	8.63	1
Markov Soft EM	22.45	7.66	2
Distributional Clustering	24.95	8.80	2
Coefficient Clustering	22.58	7.56	2

Table 3. Model performance during the Post-COVID period. Divide MSE values by 1,000,000.

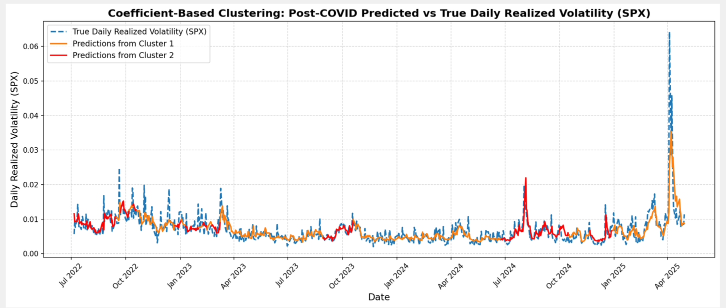

Figure 4. Coefficient-Based Clustering on Post-COVID Data.

Significance

Our empirical results demonstrate that regime-aware HAR extensions consistently yield lower forecasting errors than the standard HAR model, confirming the value of incorporating time-varying dynamics. Specifically, coefficient-based soft clustering effectively captures gradual shifts and identifies structural breaks in volatility distributions, achieving superior performance before and after the COVID time period. Meanwhile, distributional clustering better captures volatility behavior during the highly volatile COVID-19 time period. Inclusion of the VIX as a forward-looking feature enhances model responsiveness to shifts in market sentiment, thereby refining predictive accuracy.

These outcomes illustrate that volatility dynamics exhibit regime-dependent statistical properties, and that flexible models adapting regime-specific parameters provide a meaningful advantage. Future research could explore hybrid frameworks that integrate clustering and Markov switching, as well as advanced sequential models like LSTMs to capture more complex temporal dependencies.

References

[1] F. Corsi.
A simple approximate long-memory model of realized volatility.
Journal of Financial Econometrics, 7(2):174–196, 2009.

[2] Y. Zhang, L. Lei, and Y. Wei.
Forecasting the chinese stock market volatility with international market volatilities: The role of regime switching.
The North American Journal of Economics and Finance, 52(101145), 2020.