Literature Review: Forecasting through Calm and Crisis: A Regime-Aware HAR–LSTM Ensemble for Volatility Prediction

Ava Blake, Nivika Gandhi, Anurag Jakkula

June 2025

1 Introduction

Volatility forecasting remains a central research topic in the field of financial economics, as it is driven by the need to anticipate risk for effective portfolio management, derivatives pricing, and regulatory policy. While traditional models like the GARCH model have long been used for forecasting, they often fall short in accounting for varying market conditions. The Heterogeneous Autoregressive (HAR) model introduced by Corsi (2009) improves on this by incorporating high-frequency data across multiple time horizons—daily, weekly, etc.—to capture more granular volatility dynamics. However, recent studies show that the fixed dynamic structure does not fully capture the behavior of volatility in the market, suggesting that financial markets often switch between calm and turbulent regimes. As a result, models which incorporate regime-switching dynamics and nonlinear learning mechanisms are being explored with growing interest.

The purpose of this literature review is to provide the context and motivating studies behind our research. This project aims to develop a regime-switching HAR model that adjusts their parameters based on market regimes identified via the following methods: a Markov chain, or using a K-Means clustering algorithm to identify the regimes and classification algorithms (Logistic Regression, XGBoost) to detect them during forecasting. Then, it aims to integrate the best model with an LSTM network (comparing horizontal and vertical ensembling of the LSTM), and compare their performances to the standalone regime-switching HAR. We will quantify comparisons using metrics such as MSE on a test set of 1,500 daily observations.

By accounting for regime changes and complex temporal patterns, the proposed hybrid model seeks to capture nonlinear and dynamic market behaviors to improve volatility forecasting accuracy, thereby providing financial institutions with more informed and accurate information regarding future market behavior. Fund and risk managers, traders, and the government all make decisions based on volatility predictions. Such decisions include adjusting or hedging positions in assets, options pricing, and monetary and regulatory policy to stabilize the economy. Important market measures such as the Value at Risk (VaR) and derivatives are also informed by market volatility (Gunnarsson et al. 2024). Because of a low volatility's association with a rising market and volatility's sensitivity to tax and interest rate policies (Wagner 2022), the government has an interest in predicting future volatility to inform current economic policy decisions.

2 Data

The data for this project consists of high-frequency (5-minute intervals) intraday price data from the Standard & Poor's 500 (or S&P 500) Index accessed via Bloomberg Terminal. The data spans across the past six years from May 28th, 2019 till May 27th, 2025. We used the closing prices for each 5-minute interval in order to calculate the returns of the Index (SPX) throughout the day, as well as the overnight returns between the end of one trading day and the start of the next. The returns are then used to calculate daily realized volatility, following calculation and normalization techniques widely accepted and used by

Zhang, Lei, and Wei (2020), Luo et al. (2022), Ding, Kambouroudis, and McMillan (2025), Li et al. (2024), N. Hu, Yin, and Yao (2025), and many others. Additionally, we incorporated features into the model including implied volatility, for which we collected daily price data from the VIX index which is the Chicago Board Options Exchange's measure of 30-day forward-looking implied volatility. Additionally, we plan to include other feature data such as volume.

3 Key Research

3.1 Regime Switching Models

Previous studies have provided a framework and justification for a regime switching volatility prediction model. Previous literature provides evidence that volatility prediction patterns follow a nonlinear structure, justifying the claim that models such as the HAR take different forms depending on current market regimes.

Zhang, Lei, and Wei (2020) implement a Markov regime-switching model, in which HAR model parameters depend on the unobservable states. Their approach incorporates both global and domestic volatility indicators, and they demonstrate that the switching model significantly outperforms traditional HAR models, even when alternate realized volatility constructions are tested. The parameters of their HAR model depend on an unobserved regime variable which is determined by following a Markov chain process that assumes two different market regimes. Transition probabilities in their Markov transition matrix are statistically significant, and distinct regression coefficients between high- and low-volatility regimes reveal that regime-aware modeling provides meaningful improvements in capturing market behavior. The claim that these two distinct states of the market correlate with high- and low-volatility regimes is based on the work of Hamilton (1989), who justifies Markov switching for regime detection and supports the idea of volatility clustering. According to Otranto and Gallo (2007), the exact number of regimes, or the dimension of the Markov matrix, can be determined through a Bayesian nonparametric approach.

Additionally, the work of Gallo and Otranto (2020) explores the idea of regime switching by comparing Markov and smooth transition models within a multiplicative error model (MEM) framework. Their analysis, based on S&P 500 volatility, uses a 3-state Markov structure rather than 2. This approach was validated by a nonparametric Bayesian technique. In this study, results show that while Markov models dominate for one-step-ahead forecasts, smooth transition models excel over longer horizons.

Another method of regime switching which we plan to implement in our research is clustering. Clustering seeks to identify the volatility regime labels of our data for an eventual classification of the regime during inference. As such, some questions that must be answered before clustering are how time-series data is treated, how the number of regimes is determined, what clustering methods should be compared, and how clustering can be validated beyond final regression accuracy.

A consideration regarding time-series clustering is the breaking of the iid assumption due to the fact that financial volatility time-series data depend on previous entries. However, clustering in this scenario is intended to capture these time-related dependencies, because regimes occur over time and are internally dependent. Prakash et al. (2021) use the mood test to segment time series data, which compares medians and variances between time-series segments. Returns are assumed to be generated by a Dahlhaus locally stationary process. Because the Mood test tests for differences in variance, it is well-suited for regime clustering as regimes are typically characterized by variance shifts. This ensures that each cluster is drawn from a homogeneous volatility regime. Prakash et al. use Wasserstein space to create a distance matrix for the segmented data, which roughly measures the effort required to transform one time-series segment into another. This method proves more successful for time-series data than Euclidean distance, in which distributions may evolve over time.

The authors use spectral clustering as well, given that it captures nonlinear structures and sparse, irregular data well. However, spectral clustering is sensitive to noise, which volatility data tends to have. Empirical validity was found in comparing clusters to historical events (e.g. testing whether a cluster captures the 2008 financial crisis).

3.2 Implementation of Neural Networks

According to Bucci (2019), volatility is difficult to predict because it is highly persistent and nonlinear. HAR models, though persistent to an extent due to lag variables, do not capture nonlinear dynamics. On the other hand, neural networks can capture nonlinearity and, especially in recurrent typed, persistence to a greater degree than HAR models can.

To select features apart from the standard realized volatility features, the paper uses LASSO regression. The features identified are dividend-price ratio (S&P dividends/price), market excess return (S&P 500 return 3-month T-Bill rate), short term reversal factor (low prior return high prior return), and default spread (BAA AAA corporate bond yields). We will implement these features into our HAR model, as this paper found them to be the most influential features in the volatility forecasting.

MSE, QLIKE, Model Confidence Set, and the Diebold-Mariano tests are used as evaluation criteria. We intend to use MSE as our metric because it is a standard, theoretically grounded metric that effectively penalizes large forecast errors and allows for consistent comparison across different volatility models. Among the neural network architectures tested, the LSTM and NARX neural networks performed the best. This is unsurprising because they have recurrent architectures.

Further studies aim to leverage the nonlinear learning capacity of machine learning methods with the long memory power of regressive models by introducing hybrid volatility forecasting models. Although this is a more novel approach, recent literature contributes promising evidence in support of the predictive accuracy of hybrid models.

Wang (2022) explores the integration of the GARCH and HAR models with various machine learning methods, including Long Short Term Memory (LSTM). Families of HAR models, comprising the baseline models and extensions to account for jump component, are first constructed and evaluated on the realized volatility forecast of four major stock indices: SPX, DJI, IXIC, and HSI. More notably, Wang uses the outputs of these models as input variables in the machine learning methods. For the purposes of our research project, we will particularly examine Wang's construction and evaluation of a hybrid HAR-LSTM model. Their ensemble strategy uses standard HAR-generated features (daily, weekly, and monthly lagged volatilities) as explanatory variables in the LSTM, thus aiming to preserve the interpretability of the HAR model while allowing the model to learn more intricate patterns. While the approach is methodologically sound, Wang finds that the LSTM-HAR model does not strongly outperform the standalone HAR model when modeling SPX data. Nevertheless, among the hybrid models, the LSTM and Transformer models consistently outperformed the other hybrid models for all four major indices. Moreover, most of the hybrid models constructed by Wang tend to underestimate extreme periods of high volatility, such as in 2020.

The work of G. Hu, Ma, and Zhu (2025) further explores the application of a hybrid modeling framework to forecasting the volatility of China's oil futures. The authors specifically ensemble an HAR model that incorporates a measure for signed jumps (HAR-SJ model) with four machine learning methods, including LSTM. High-frequency 5-minute data from China's Shanghai crude oil futures market was collected and manipulated to calculate the lagged realized volatities. As in Wang's dissertation, the authors use the HAR-generated features as input variables in the machine learning models. Empirical results demonstrate that the hybrid ML-HAR models achieve significantly lower error values in comparison to the standalone HAR models across extended time horizons. These findings further solidify Wang's conclusion in the predictive potential of ensembling machine learning methods and traditional regressive models to forecast volatility.

4 Synthesis

Across all the previous literature, certain trends arise which motivate our methodology in this project. Zhang, Lei, and Wei (2020), Gallo and Otranto (2020), and Ding, Kambouroudis, and McMillan (2025) show that the standard HAR model is consistently outperformed by a regime-switching model, though the specific type of regime-switching model varies depending on the time scale. For the purposes of this project, we are focusing on daily forecasts with the construction of the regime-switching HAR model, and then furthering this by implementing recursive methods in order to forecast weekly and monthly future volatility. This is supported by the findings of Gallo and Otranto (2020), as they showed Markov regime-switching models to perform best on the daily scale. We will identify the dimension of the Markov matrix using the Bayesian nonparametric approach used by Otranto and Gallo (2007). We will compare this method of regime-classification with clustering methods, which have not been previously explored in this specific context. Although, spectral clustering was used for volatility regime classification by Prakash et al. (2021). We plan to build off of this approach by applying spectral clustering to identify regimes, as well as testing the performance of K-means clustering.

Studies such as the ones conducted by Sullivan (2018) and Bucci (2019) also show that deep learning models, especially LSTM architectures, have proven effective in capturing nonlinear relationships and long-range dependencies in financial time series. However, these models are not able to capture the probabilistic regime switching patterns.

Wang (2022) and G. Hu, Ma, and Zhu (2025) make notable contributions by integrating HAR models with these deep learning models. Both studies use HAR model outputs - daily, weekly, and monthly lagged realized volatilites - as additional input features for a range of machine learning algorithms, including LSTM. While the authors find that these hybrid models can improve predictive accuracy over their standalone components, Wang observes that gains are limited when applied to SPX data and that most hybrid models still underestimate periods of extreme volatility. The work of Hu, Ma, and Zhu further suggests that hybrid models do not fare as successful over shorter time horizons. These limitations underscore the need for hybrid models that explicitly account for structural shifts to improve forecasting accuracy under extreme market conditions.

Our research looks to address these gaps by incorporating a regime-switching framework directly into the model framework. Unlike Wang (2022) or G. Hu, Ma, and Zhu (2025), we introduce probabilistic regime detection using both Hidden Markov Models and unsupervised clustering methods. These regimes are then used to adjust the parameters of the HAR model and additionally included as features in the LSTM. This integration approach allows us to address the issue of regime dependency that previous models overlook, while still leveraging nonlinear learning.

Thus, we hope to unify two previously disjoint efforts: the use of regime-informed parameters in volatility forecasting, and nonlinear learning of temporal patterns by neural networks. Moreover, the clustering and classification pipeline we propose addresses concerns raised by Gallo and Otranto (2020) about error propagation in Markov switching models by offering an alternative that is potentially more effective than smooth transition models.

5 Conclusion

This literature review surveys key developments in volatility forecasting, with a focus on regime-aware and hybrid modeling approaches. While traditional econometric models like HAR remain widely used for their interpretability and performance under stable conditions, they are limited in handling structural breaks and nonlinear dynamics. Regime-switching models, particularly those based on hidden Markov models (HMMs), improve forecast accuracy by adapting to latent states of market volatility. However, they are constrained by strong probabilistic assumptions and often suffer from error propagation.

Clustering-based methods present a promising alternative by identifying regimes through observable patterns in the data, without relying on predefined transition structures. Approaches such as spectral clustering and Wasserstein distance-based segmentation offer greater flexibility and potentially more accurate regime detection.

Additionally, LSTM networks have proven effective in modeling the persistence and nonlinearity of realized volatility. Recent hybrid models attempt to combine HAR and LSTM architectures, but most do not integrate regime classification explicitly, nor do they explore how different methods of combining HAR and LSTM components affect forecast outcomes.

Our proposed model aims to address these gaps by introducing a clustering-based classification mechanism for regime detection, which is then incorporated into a hybrid HAR–LSTM ensemble. This literature review thus provides both the theoretical motivation and empirical context for our project.

References

Bucci, A. (2019). "Realized Volatility Forecasting with Neural Networks". In: *Munich Personal RePEc Archive* 95443.

Corsi, F. (2009). "A Simple Approximate Long-Memory Model of Realized Volatility". In: *Journal of Financial Econometrics* 7.2, pp. 174–196.

Ding, Y., D. Kambouroudis, and D. G. McMillan (2025). "Forecasting realised volatility using regime-switching models". In: *International Review of Economics & Finance* 101.104171.

Gallo, G. M. and E. Otranto (2020). "Forecasting realized volatility with changing average levels". In: *International Journal of Forecasting* 52.3, pp. 620–634.

Gunnarsson, E. S. et al. (2024). "Prediction of realized volatility and implied volatility indices using AI and machine learning: A review". In: *International Review of Financial Analysis* 93.103221.

Hamilton, J. D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle". In: *Econometrica* 52.2, pp. 357–384.

Hu, G., X. Ma, and T. Zhu (2025). “Forecasting volatility of China’s crude oil futures based on hybrid ML-HAR-RV models”. In: *The North American Journal of Economics and Finance* 78, p. 102428. ISSN: 1062-9408. DOI: https://doi.org/10.1016/j.najef.2025.102428. URL: https://www.sciencedirect.com/science/article/pii/S1062940825000683.

Hu, N., X. Yin, and Y. Yao (2025). “A novel HAR-type realized volatility forecasting model using graph neural network”. In: *International Review of Financial Analysis* 98.103881.

Li, X. et al. (2024). “Forecasting the volatility of educational firms based on HAR model and LSTM models considering sentiment and educational policy”. In: *Heliyon* 10.19.

Luo, J. et al. (2022). “Forecasting realized volatility of agricultural commodity futures with infinite Hidden Markov HAR models”. In: *International Journal of Forecasting* 38.1, pp. 51–73.

Otranto, E. and G. M. Gallo (2007). “A Nonparametric Bayesian Approach to Detect the Number of Regimes in Markov Switching Models”. In: *Econometric Reviews*.

Prakash, A. et al. (May 2021). “Structural Clustering of Volatility Regimes for Dynamic Trading Strategies”. In: *Applied Mathematical Finance* 28.3, pp. 236–274. ISSN: 1466-4313. DOI: 10.1080/1350486x.2021.2007146. URL: http://dx.doi.org/10.1080/1350486X.2021.2007146.

Sullivan, J. C. (2018). “Stock Price Volatility Prediction with Long Short-Term Memory Neural Networks”. In: *Stanford University*.

Wagner, H. (2022). *Why Volatility Is Important for Investors*. URL: https://www.investopedia.com/articles/financial-theory/08/volatility.asp.

Wang, X. (2022). “Hybrid Volatility Forecasting Models Based on Machine Learning of High-Frequency Data”. PhD thesis. Florida State University College of Arts and Sciences.

Zhang, Y., L. Lei, and Y. Wei (2020). “Forecasting the Chinese stock market volatility with international market volatilities: The role of regime switching”. In: *The North American Journal of Economics and Finance* 52.101145.