Adaptive Level-Set Fuzzy Heterogeneous Autoregressive Model for Realized Volatility Forecasting

Leandro Maciel, Gabriel Navarro, and João Pedro Canabrava
School of Economics, Business, Accounting and Actuarial Science
University of São Paulo
São Paulo, Brazil
{leandromaciel, gbnavarro, joaopedro.canabrava}@usp.br

Fernando Gomide
School of Electrical and Computer Engineering
University of Campinas
Campinas, Brazil
gomide@unicamp.br

***Abstract*—This paper suggests an Adaptive Level Set Fuzzy (ALSM) extension of the Heterogeneous Autoregressive (HAR) model for realized volatility forecasting. The ALSM-HAR model preserves the interpretability of the HAR structure while introducing nonlinear, regime-dependent, and adaptive dynamics through level set–based fuzzy inference and recursive estimation. The model is used to forecast the volatility of eight major cryptocurrencies one-step-ahead using high-frequency data. Out-of-sample results show that ALSM-HAR consistently performs better than standard HAR and its main extensions in both forecast accuracy and statistical robustness, particularly under mean squared error criteria. ALSM-HAR advances the current state of the art in the area since it significantly improves the effectiveness of forecasting in highly volatile and uncertain markets.**

***Index Terms*—Adaptive fuzzy systems, HAR models, realized volatility, cryptocurrency markets, level sets.**

I. Introduction

Volatility is not directly observable in asset price time series data, but is perceived by investors and plays a central role in financial markets. Volatility is widely used as a risk measure and its accurate forecasting is crucial for portfolio allocation, asset pricing, and risk management [1]. As a result, reliable volatility forecasting models have been a longstanding topic in the literature, fostering a wide range of econometric and statistical approaches.

Early contributions focused on Generalized Autoregressive Conditional Heteroskedasticity (GARCH)-type models, which capture conditional heteroskedasticity and volatility clustering using daily return data [2], [3]. With the increasing availability of intraday data, realized variance (RV) emerged as a viable proxy for latent volatility, motivating the use of time-series models for volatility forecasting [4]. Among these approaches, the Heterogeneous AutoRegressive (HAR) model developed in [5] became a benchmark. HAR decomposes realized volatility into daily, weekly, and monthly components to reflect the heterogeneous investment horizons of the market participants.

The widespread adoption of HAR models is justified by their strong out-of-sample accuracy, parsimony, and interpretability. Unlike GARCH models, HAR-based specifications directly exploit intraday information and treat volatility as an observable process, yielding superior forecast accuracy [1], [6]. Several extensions have further enhanced the HAR approach, including the incorporation of jumps, leverage effects, and signed volatility components [7]–[9].

Machine learning-based approaches have been increasingly adopted to forecast volatility [1], [6], [10]. Interestingly, existing evidence suggests that nonlinear models do not consistently outperform linear HAR-type models in a statistical sense [6]. Properly specified HAR models remain difficult to beat, while offering superior economic interpretability [11]. Recent advances combining machine learning techniques with HAR structures show promising results [1].

Cryptocurrencies constitute a particularly challenging environment for volatility forecasting due to their high variability and distinct market microstructure. In addition, cryptocurrencies are assets with inherently uncertain and ambiguous fundamentals, as their valuation is strongly influenced by investor sentiment, attention, and speculative behavior [12]. As a result, modeling approaches capable of explicitly accommodating such uncertainty and regime-dependent dynamics become crucial for an accurate measurement of risk. Prior studies have assessed the HAR framework in the cryptocurrency market by incorporating sentiment indicators, jumps, higher-order moments, regularization techniques and machine learning models [10], [13].

This paper uses the Adaptive Level Set Fuzzy Modeling (ALSM) framework to extend the HAR framework for realized volatility forecasting. ALSM, originally developed in [14], provides a data-driven approach to compute the output of fuzzy rule-based models based on the concept of level sets. While linguistic fuzzy models typically rely on centroid-based defuzzification and functional fuzzy models to compute outputs as weighted averages of the functions in the consequents of the active rules, the level set-based approach determines the output through a mapping between rule activation levels and the output space [15]. ALSM combines level set modeling with recursive correntropy-based parameter estimation and processes data sequentially within an adaptive framework.

More precisely, the ALSM framework is embedded into the HAR structure to originate the novel ALSM-HAR model introduced in this paper. This formulation preserves the heterogeneous autoregressive decomposition of realized volatility while introducing nonlinear, regime-dependent, and adaptive dynamics through fuzzy inference. The ALSM-HAR model is used in one-step-ahead realized volatility forecasting for eight of the largest cryptocurrencies traded in the market. Predictive performance is evaluated in an out-of-sample framework and benchmarked against the traditional HAR model and its main extensions. Model comparisons are conducted in terms of forecast accuracy, supported by statistical testing procedures. The main contribution of the paper is the introduction of a fuzzy level set-based extension of the HAR framework to forecast realized volatility. By explicitly modeling uncertainty and ambiguity in the volatility dynamics of cryptocurrencies—where prices and risk are strongly influenced by investor sentiment and attention—the ALSM-HAR approach gives a novel flexible and interpretable alternative to standard HAR models and machine learning-based methods.

The paper proceeds as follows. Section II gives an overview of the fundamental concepts of realized volatility modeling, the HAR model and its extensions. Section III details the ALSM-HAR model. Computational experiments are reported in Section IV. Finally, Section V summarizes the main findings and outlines potential directions for future research.

Data and code availability: the intraday cryptocurrency data used in this study were obtained from a commercial data source (Binance) and are subject to usage restrictions; therefore, they cannot be publicly shared. All codes implementing the proposed ALSM-HAR model, as well as the benchmark models, will be made publicly available upon acceptance using a synthetic dataset.

II. Realized volatility modeling

The HAR model is the benchmark used in the study. HAR models realized volatility as a linear function of the daily, weekly, and monthly components, capturing heterogeneous investor horizons [5]. Realized variance is computed from intraday data as [4]

$$RV_t = \sum_{j=1}^{M} r_{t,j}^2, \tag{1}$$

where $r_{t,j}$ denotes the j-th intraday log-return on day t, computed at an intraday sampling frequency with M observations per day. Weekly and monthly realized variance components are found as

$$RV_t^{(w)} = \frac{1}{7}\sum_{i=1}^{7} RV_{t-i}, \qquad RV_t^{(m)} = \frac{1}{30}\sum_{i=1}^{30} RV_{t-i}. \tag{2}$$

For continuously traded assets such as cryptocurrencies, 7- and 30-day horizons are particularly appropriate. The baseline HAR model is as follows [5]

$$RV_t = \alpha + \beta_d RV_{t-1} + \beta_w RV_{t-1}^{(w)} + \beta_m RV_{t-1}^{(m)} + \varepsilon_t, \tag{3}$$

where α is a constant, β_d, β_w, and β_m are slope coefficients, and ε_t is a zero-mean error term.

Several extensions of the HAR framework have been developed in the literature. They are summarized next.

The HAR-CJ model decomposes realized volatility into continuous and jump components [16]. Let J_t denote the jump variation and define the continuous component as $C_t = RV_t - J_t$. Weekly and monthly continuous components are computed as

$$C_t^{(w)} = \frac{1}{7}\sum_{i=1}^{7} C_{t-i}, \qquad C_t^{(m)} = \frac{1}{30}\sum_{i=1}^{30} C_{t-i}. \tag{4}$$

The HAR-CJ model is the following

$$\begin{aligned} RV_t = \alpha + \beta_d C_{t-1} + \beta_w C_{t-1}^{(w)} + \beta_m C_{t-1}^{(m)} \\ + \gamma J_{t-1} + \varepsilon_t, \end{aligned} \tag{5}$$

where γ measures the marginal effect of jumps on future realized volatility.

To mitigate the effects of market microstructure noise on jump identification, the HAR-TCJ model employs threshold-based detection [7]. Let TJ_t be the threshold jump component and let the threshold continuous component be $TC_t = RV_t - TJ_t$. The weekly and monthly threshold continuous components are

$$TC_t^{(w)} = \frac{1}{7}\sum_{i=1}^{7} TC_{t-i}, \qquad TC_t^{(m)} = \frac{1}{30}\sum_{i=1}^{30} TC_{t-i}. \tag{6}$$

The HAR-TCJ model is given by

$$\begin{aligned} RV_t = \alpha + \beta_d TC_{t-1} + \beta_w TC_{t-1}^{(w)} + \beta_m TC_{t-1}^{(m)} \\ + \gamma TJ_{t-1} + \varepsilon_t. \end{aligned} \tag{7}$$

The LHAR-TCJ model further incorporates a leverage term to capture asymmetric volatility responses to negative returns [8]. Let r_{t-1} be the daily close-to-close log-return and define the negative return component as $r_{t-1}^- = \min(r_{t-1}, 0)$. The LHAR-TCJ model is

$$\begin{aligned} RV_t = \alpha + \beta_d TC_{t-1} + \beta_w TC_{t-1}^{(w)} + \beta_m TC_{t-1}^{(m)} \\ + \gamma TJ_{t-1} + \delta r_{t-1}^- + \varepsilon_t, \end{aligned} \tag{8}$$

where δ captures the leverage (asymmetry) effect.

Finally, the HAR-SJ model accounts for asymmetric jump effects by incorporating signed jump measures derived from realized semivariance [9]. Let BPV_t be the bipower variation proxy for the continuous component of volatility, and let SJ_t^+ and SJ_t^- be the positive and negative signed jump measures, respectively. The HAR-SJ model is as follows

$$\begin{aligned} RV_t = \alpha + \beta_w RV_{t-7} + \beta_m RV_{t-30} + \theta\, \mathrm{BPV}_t \\ + \phi^+ \exp(SJ_t^+) + \phi^- \exp(SJ_t^-) + \varepsilon_t, \end{aligned} \tag{9}$$

where θ measures the effect of the continuous component proxy, while ϕ^+ and ϕ^- capture the effects of positive and negative signed jumps on future realized volatility.

The HAR-based models (3), (5), (7), (8), and (9) are the main benchmarks in the current realized volatility forecasting literature. They provide a comprehensive basis for comparison with the ALSM-HAR model proposed in this paper.

III. Adaptive Level Set Fuzzy HAR Modeling

The ability to adapt to dynamic, nonstationary environments is a fundamental requirement in economics, engineering, and complex systems. In financial markets, volatility dynamics is subject to regime changes, structural breaks, and shifts driven by investor behavior and external information. In these markets, forecasting models must be able to learn from nonstationary data streams and update their parameters sequentially to track the changes. These challenges are central to machine learning and computational intelligence research, where several adaptive modeling strategies have been developed [17]–[19].

This paper adopts a knowledge-based adaptive fuzzy modeling perspective and extends the Adaptive Level Set Fuzzy Modeling (ALSM) framework of [14] to the framework of Heterogeneous Autoregressive (HAR) setting for realized volatility forecasting. The resulting ALSM-HAR model preserves the input structure of the HAR model, namely, the daily, weekly, and monthly realized volatility components, while replacing the linear regression mapping with an adaptive fuzzy inference mechanism based on level sets. This formulation maintains the economic interpretability of the HAR framework and introduces nonlinear, adaptive, and uncertainty-aware dynamic modeling.

The ALSM-HAR model is based on fuzzy rules of the form

$$\mathcal{R}_i: \text{ if } x \text{ is } \mathcal{A}_i \text{ then } y \text{ is } \mathcal{B}_i, \tag{10}$$

where $i = 1, 2, \ldots, N$, $x \in \mathbb{R}^p$ denotes the input vector, $\mathcal{A}_i$ is an antecedent fuzzy set with membership function $\mathcal{A}_i(x) : \mathbb{R}^p \to [0,1]$, and $\mathcal{B}_i$ is a convex consequent fuzzy set defined on the output space $\mathbb{R}$ with membership function $\mathcal{B}_i(y) : \mathbb{R} \to [0,1]$.

In volatility forecasting, the input x_t is the vector made of the HAR regressors,

$$x_t = \big(RV_{t-1},\ RV_{t-1}^{(w)},\ RV_{t-1}^{(m)}\big),$$

where RV_{t-1} denotes daily realized variance, and $RV_{t-1}^{(w)}$ and $RV_{t-1}^{(m)}$ denote the weekly and monthly realized variance components, respectively.

Given an input $x = x_t$, the activation level of rule i is computed as

$$\tau_i = \mathcal{A}_i(x), \tag{11}$$

For each activation level τ_i, the corresponding τ_i-level set of the consequent fuzzy set $\mathcal{B}_i$ is

$$\mathcal{B}_{\tau_i} = \{y \mid \tau_i \le \mathcal{B}_i(y)\} = [y_{il}, y_{iu}], \tag{12}$$

where y_{il} and y_{iu} denote the lower and upper bounds of the interval. The midpoint of this interval is

$$m_i = \frac{y_{il} + y_{iu}}{2}. \tag{13}$$

The output of the fuzzy model is obtained as a normalized weighted average of the midpoints,

$$\hat{y} = \frac{\sum_{i=1}^{N} \tau_i m_i}{\sum_{i=1}^{N} \tau_i}, \tag{14}$$

with $\tau_i > 0$ for at least one rule. The midpoint is adopted due to its simplicity and its optimality under the max-norm approximation criterion.

In the adaptive level set formulation, the explicit construction of the consequent fuzzy sets $\mathcal{B}_i$ is avoided by directly estimating an output function $\mathcal{F}_i : [0,1] \to \mathbb{R}$ from activation levels to output values. In the simplest case is to assume the output function affine,

$$\mathcal{F}_i(\tau_i) = v_i \tau_i + w_i, \tag{15}$$

where v_i and w_i are rule-specific coefficients to be estimated from data.

The ALSM-HAR model processes input data sequentially and uses a recursive correntropy-based estimation procedure [14], [20]. Let $\mathcal{D} = \{(x^k, y^k)\}_{k=1}^{K}$ denote a data set, where $x^k \in \mathbb{R}^p$ is the input and $y^k \in \mathbb{R}$ is the corresponding realized volatility observation at step k. Let the parameter vector $\mathbf{u}^k$ be the one assembled by the coefficients of the output functions $\mathbf{u}^k = [v_1^k, w_1^k, \ldots, v_N^k, w_N^k]^\top$.

Initialize the parameter vector as follows $\mathbf{u}^0 = \mathbf{0}$ and a gain matrix as $\mathbf{P}^0 = \alpha \mathbf{I}$, where $\mathbf{I}$ is the identity matrix and $\alpha > 0$ is a large constant [21].

For each input and observation data pair (x^k, y^k), the algorithm proceeds as follows:

1) Compute the prediction error
$$e^k = y^k - \mathbf{d}^k \mathbf{u}^{k-1}.$$
2) Compute the correntropy-based weight
$$\psi^k = \frac{1}{\sqrt{2\pi}\zeta^3} \exp\!\left(-\frac{|e^k|^2}{2\zeta^2}\right), \tag{16}$$
where $\zeta > 0$ is the kernel width parameter.
3) Compute the activation levels $\tau_i^k = \mathcal{A}_i(x^k)$ for $i = 1, \ldots, N$.
4) Define the regression vector
$$\mathbf{d}^k = \big[(\tau_1^k)^2/s^k,\ \tau_1^k/s^k,\ \ldots,\ (\tau_N^k)^2/s^k,\ \tau_N^k/s^k\big],$$
$$s^k = \sum_{i=1}^{N} \tau_i^k.$$
5) Update the gain matrix
$$\mathbf{P}^k = \frac{1}{\lambda}\left(\mathbf{P}^{k-1} - \frac{\psi^k \mathbf{P}^{k-1}(\mathbf{d}^k)^T \mathbf{d}^k \mathbf{P}^{k-1}}{\lambda + \psi^k \mathbf{d}^k \mathbf{P}^{k-1}(\mathbf{d}^k)^T}\right), \tag{17}$$
where $\lambda \in (0,1]$ is a forgetting factor.
6) Update the parameter vector
$$\mathbf{u}^k = \mathbf{u}^{k-1} + \psi^k \mathbf{P}^k (\mathbf{d}^k)^T \big(y^k - \mathbf{d}^k \mathbf{u}^{k-1}\big). \tag{18}$$
7) Compute the model output
$$\hat{y}^k = \mathbf{d}^k \mathbf{u}^k. \tag{19}$$

The recursive update in (17) and (18) corresponds to the maximization of the correntropy-based objective function

$$\max_{\mathbf{u}^k} J^k = \sum_{i=1}^{k} \lambda^{k-i} \frac{1}{\sqrt{2\pi}\zeta} \exp\!\left(-\frac{|y^i - \mathbf{d}^i \mathbf{u}^{i-1}|^2}{2\zeta^2}\right), \tag{20}$$

which enhances robustness to outliers and improves the tracking of time-varying dynamics compared to classical recursive least squares [20], [21].

By embedding the HAR input structure into the ALSM framework, the resulting ALSM-HAR model adds interpretability, adaptivity, and nonlinear forecasting, providing a novel flexible and uncertainty-aware approach to realized volatility forecasting.

IV. Computational experiments

A. Data and parameters of the models

Empirical analysis is conducted using intraday price data for eight major cryptocurrencies: Cardano (ADA), Binance Coin (BNB), Bitcoin (BTC), Dogecoin (DOGE), Ethereum (ETH), TRON (TRX), Stellar (XLM), and Ripple (XRP). All intraday data were obtained from Binance[1], one of the largest cryptocurrency exchanges in terms of trading volume and liquidity. Based on these high-frequency prices, realized variance and all remaining variables used in the empirical analysis were constructed.

The sample periods differ between the cryptocurrencies due to the availability of data. The initial and final dates, as well as the total number of daily observations for each series, are reported in Table I. Intraday prices sampled at five-minute intervals are used to compute daily realized variance based on closing prices, following the definition in (1). The resulting dataset spans multiple volatility regimes and distinct market phases observed in cryptocurrency markets.

The dataset was split into in-sample (for model development) and out-of-sample (for forecast evaluation) periods using a 60/40 split. Parameters of the econometric models are estimated using the ordinary least squares. The ALSM-HAR model requires selecting the model structure (number of fuzzy rules) and the coefficients of the output functions. Gaussian membership functions are employed, with centers and dispersions given by Gustafson–Kessel clustering [22]. The number of rules is found via in-sample simulations. The scenario that maximizes the predictive accuracy is the one selected. Except for DOGE, in which only four rules are needed, all the remaining cryptocurrencies use ALSM-HAR models with six rules. The coefficients of the output functions of the ALSM-HAR models are estimated assuming $\zeta = 1$.

B. Performance metrics

Model performance is evaluated using the Quasi-Likelihood (QLIKE) and Mean Squared Error (MSE) loss functions, following [23]. Both metrics are computed in an out-of-sample forecasting framework using one-step-ahead predictions, which represent the standard horizon in portfolio management and volatility-based risk measurement.

Let RV_t denote the realized variance observed at time t, and let $\widehat{RV}_{t|t-1}$ denote the corresponding one-step-ahead forecast generated at time $t-1$. The QLIKE loss function is defined as

$$\text{QLIKE} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{RV_t}{\widehat{RV}_{t|t-1}} - \log\left(\frac{RV_t}{\widehat{RV}_{t|t-1}}\right) - 1\right), \quad (21)$$

while the MSE is given by

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}\left(RV_t - \widehat{RV}_{t|t-1}\right)^2. \quad (22)$$

QLIKE is a loss function specifically designed for volatility and variance forecasts and is robust to measurement errors in realized variance, whereas MSE provides a complementary measure of average squared forecast deviations. Lower values of both metrics indicate superior predictive performance.

C. Results

Table II reports the out-of-sample MSE values. To facilitate interpretation, all reported loss values are expressed relative to the benchmark HAR model. Relative values below one indicate that a given model outperforms the benchmark, whereas values above one imply superior performance of the benchmark relative to the competing model. Consistent with [11], all HAR-type models are estimated using rolling windows of two years (730 daily observations), balancing parameter stability and adaptability to evolving market conditions. Model parameters are re-estimated on a daily basis as the estimation window advances through time. The results show that the proposed ALSM-HAR model consistently outperforms the benchmark across all eight cryptocurrencies, with relative MSE values ranging from 0.64 to 0.78. Bold values indicate the best performance among the competing models. For instance, the relative MSE reaches 0.64 for BNB and 0.70 for BTC, corresponding to reductions of approximately 36% and 30% in squared forecast errors relative to the HAR model. These improvements are economically meaningful and indicate that the adaptive fuzzy specification is able to substantially reduce large forecast deviations at the daily horizon.

In contrast, most HAR-based extensions exhibit mixed performance. While some specifications marginally outperform the benchmark for specific assets—for example, HAR-CJ for BNB (0.88) and DOGE (0.77)—their relative MSE values are frequently close to or above one. Other extensions, such as HAR-SJ and HAR-TCJ, display substantially worse performance for several assets, with relative MSE values exceeding 1.4 for BTC and ETH and reaching 2.86 for DOGE in the case of HAR-SJ. Overall, the MSE results clearly indicate the dominance of the ALSM-HAR model in terms of point forecast accuracy at the one-day horizon.

Table III presents the corresponding results based on the QLIKE loss function, which is specifically designed for evaluating variance forecasts. Although the ALSM-HAR model again outperforms the benchmark for several assets—such as ADA (0.73), BTC (0.79), ETH (0.69), XLM (0.66), and XRP (0.79)—the gains are less uniform than those observed for MSE. For BNB and TRX, the relative QLIKE values of

[1]Source: https://www.binance.com.

TABLE I
SAMPLE START AND END DATES AND NUMBER OF DAILY OBSERVATIONS FOR EACH CRYPTOCURRENCY ANALYZED IN THE STUDY.

	ADA	BNB	BTC	DOGE	ETH	TRX	XLM	XRP
Initial date	17/04/2018	30/11/2017	30/11/2017	05/07/2019	30/11/2017	11/06/2018	31/05/2018	04/05/2018
Final date	03/12/2024	02/12/2024	02/12/2024	02/12/2024	02/12/2024	03/12/2024	03/12/2024	03/12/2024
# observations	2,422	2,559	2,559	1,977	2,559	2,367	2,378	2,405

TABLE II
OUT-OF-SAMPLE RELATIVE MSE FOR THE ONE-DAY-AHEAD REALIZED VOLATILITY FORECASTING.

Model	ADA	BNB	BTC	DOGE	ETH	TRX	XLM	XRP
ALSM-HAR	**0.7844**	**0.6433**	**0.6955**	**0.7420**	**0.7438**	**0.7641**	**0.7768**	**0.7068**
HAR-CJ	1.0076	0.8773	0.9719	0.7712	0.9861	1.0312	1.0289	0.9785
HAR-SJ	1.0434	1.0867	1.1827	2.8629	1.0354	1.2175	1.1708	1.1152
HAR-TCJ	1.1525	1.5361	1.4475	0.7784	1.5978	1.1524	1.0280	1.1229
LHAR-TCJ	1.0846	1.4653	1.3887	0.7772	1.4767	1.1450	1.0524	1.1390

TABLE III
OUT-OF-SAMPLE RELATIVE QLIKE FOR THE ONE-DAY-AHEAD REALIZED VOLATILITY FORECASTING.

Model	ADA	BNB	BTC	DOGE	ETH	TRX	XLM	XRP
ALSM-HAR	**0.7284**	0.9720	**0.7877**	0.7534	**0.6915**	1.1100	**0.6624**	**0.7879**
HAR-CJ	0.8702	**0.7889**	0.8406	0.7926	0.8194	0.7994	0.8715	0.8955
HAR-SJ	0.8946	0.8602	0.9303	0.7873	0.9182	0.7658	0.8816	0.8841
HAR-TCJ	1.0041	0.8656	1.6137	0.7867	0.8361	**0.6558**	0.8735	0.8966
LHAR-TCJ	0.9787	0.8844	1.5697	**0.7507**	0.8360	0.8708	0.8628	0.8818

the ALSM-HAR model are close to or slightly above one, indicating performance comparable to or marginally worse than the HAR benchmark. In these cases, alternative HAR-based extensions, such as HAR-CJ for BNB (0.79) and HAR-TCJ for TRX (0.66), deliver the lowest QLIKE values.

These results suggest that, while the ALSM-HAR model is particularly effective at reducing large forecast errors—as reflected by the MSE—it does not uniformly dominate simpler HAR-based specifications under a loss function that penalizes relative forecast errors. This divergence between MSE and QLIKE highlights that the benefits of the fuzzy-based approach on the daily horizon are mainly driven by improvements in average squared accuracy rather than by systematic gains in all aspects of variance forecast performance.

In addition to forecast accuracy measures, the competing models are also statistically evaluated using the Model Confidence Set (MCS) procedure of [24], which identifies the subset of models that are statistically indistinguishable from the best-performing specification based on the MSE loss function. The MCS results reported in Table IV summarize the statistical significance of the MSE-based comparisons. Each column corresponds to a specific cryptocurrency and reports whether a given model belongs to the MCS for that asset, based on MSE losses. For instance, considering the BTC column, the entry "Yes" for the ALSM-HAR model indicates that its forecasting performance for Bitcoin is statistically indistinguishable from the best-performing model at the chosen confidence level. In contrast, the entries "No" for the remaining models imply that their predictive accuracy for BTC is statistically inferior and that they are excluded from the confidence set. Therefore, the BTC column shows that ALSM-HAR is the only model that remains statistically competitive for Bitcoin under the MCS procedure, highlighting its robustness relative to alternative HAR-based specifications.

The ALSM-HAR model belongs to the MCS for all eight cryptocurrencies, yielding a MCS frequency of 100% (see Table IV). In contrast, the closest competitors—HAR, HAR-TCJ and LHAR-TCJ—are included in the confidence set for seven assets each, corresponding to an MCS frequency of 87.5%. Other HAR extensions exhibit even lower inclusion frequencies. These findings reinforce the conclusion that the superior MSE performance of the ALSM-HAR model is statistically robust and is not driven by isolated assets.

The ALSM-HAR model delivers substantial reductions in squared forecast errors and dominates the MSE-based MCS analysis, but its relative gains under QLIKE are more asset-dependent. To illustrate the performance of the fuzzy model, Figure 1 shows the realized variance of Bitcoin and the forecasts from the HAR and ALSM-HAR models, highlighting the ability of the fuzzy approach to adapt to the time-varying dynamics nature of volatility.

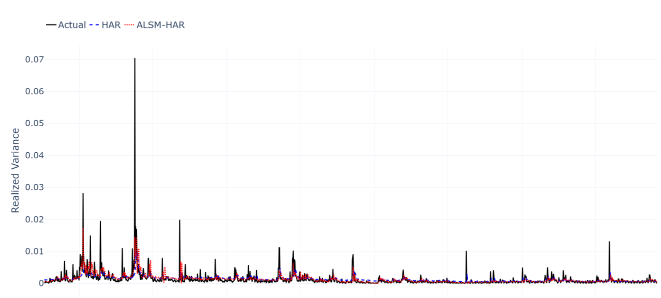

Fig. 1. Actual and predicted Bitcoin realized variance using the HAR and ALSM-HAR models.

TABLE IV
MODEL CONFIDENCE SET (MCS) RESULTS BASED ON MSE FOR ONE-DAY-AHEAD REALIZED VARIANCE FORECASTING AT THE 75% CONFIDENCE LEVEL.

Model	ADA	BNB	BTC	DOGE	ETH	TRX	XLM	XRP	Total MCS	Frequency
ALSM-HAR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8	100.0%
HAR-CJ	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	7	87.5%
HAR-SJ	No	Yes	No	No	Yes	No	Yes	Yes	4	50.0%
HAR	Yes	Yes	No	No	Yes	Yes	Yes	Yes	6	75.0%
HAR-TCJ	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	7	87.5%
LHAR-TCJ	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	7	87.5%

Note: "Yes" ("No") indicates inclusion (exclusion) of the model in the MCS for a given cryptocurrency, based on MSE losses. "Frequency" denotes the share of assets for which the model belongs to the confidence set.

V. CONCLUSION

This paper has introduced ALSM-HAR, an adaptive level set fuzzy extension of the Heterogeneous Autoregressive (HAR) framework for realized volatility forecasting. The ALSM-HAR framework endows the interpretability and structural simplicity of HAR models with the flexibility of adaptive fuzzy inference, allowing the model to capture nonlinear, regime-dependent, and uncertainty-driven dynamics. Empirical results for eight cryptocurrencies show that ALSM-HAR delivers statistically superior forecasting performance at the one-day-ahead horizon, consistently outperforming standard HAR-based benchmarks and remaining in the set of statistically superior models across assets. Overall, the findings support adaptive level set fuzzy modeling as a flexible and effective extension of the HAR framework for realized volatility forecasting. Future research shall extend the analysis to other asset classes, investigate additional HAR specifications within the ALSM framework, evaluate longer forecasting horizons and machine learning algorithms, and assess the economic value of the proposed forecasts in applications such as portfolio management and the computation of downside risk measures, including Value-at-Risk.

ACKNOWLEDGMENT

The authors thank the Brazilian National Council for Scientific and Technological Development (CNPq) for grants 301584/2025-7 and 304263/2024-9, and the Ripple Impact Fund, a donor-advised fund of the Silicon Valley Community Foundation, for grant 2018-196450 (5855), provided as part of the University Blockchain Research Initiative (UBRI).

REFERENCES

[1] N. Hu, X. Yin, and Y. Yao, "A novel HAR-type realized volatility forecasting model using graph neural network," *International Review of Financial Analysis*, vol. 98, p. 103881, 2025.

[2] R. F. Engle, "Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.

[3] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[4] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys, "Modeling and forecasting realized volatility," *Econometrica*, vol. 71, no. 2, pp. 579–625, 2003.

[5] F. Corsi, "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, vol. 7, no. 2, pp. 174–196, 2009.

[6] R. R. Branco, A. Rubesam, and M. Zevallos, "Forecasting realized volatility: Does anything beat linear models?" *Journal of Empirical Finance*, vol. 78, p. 101524, 2024.

[7] F. Corsi, D. Pirino, and R. Renò, "Threshold bipower variation and the impact of jumps on volatility forecasting," *Journal of Econometrics*, vol. 159, no. 2, pp. 276–288, 2010.

[8] F. Corsi and R. Renò, "Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling," *Journal of Business & Economic Statistics*, vol. 30, no. 3, pp. 368–380, 2012.

[9] A. J. Patton and K. Sheppard, "Good volatility, bad volatility: Signed jumps and the persistence of volatility," *The Review of Economics and Statistics*, vol. 97, no. 3, pp. 683–697, 2015.

[10] G. Dudek, P. Fiszeder, P. Kobus, and W. Orzeszko, "Forecasting cryptocurrencies volatility using statistical and machine learning methods: A comparative study," *Applied Soft Computing*, vol. 151, p. 111132, 2024.

[11] J. Chassot and F. Audrino, "HARd to beat: The overlooked impact of rolling windows in the era of machine learning," *International Journal of Forecasting*, 2025.

[12] T. Lan and M. Frömmel, "Risk factors in cryptocurrency pricing," *International Review of Financial Analysis*, vol. 105, p. 104389, 2025.

[13] Z. Zhang and R. Zhao, "Good volatility, bad volatility, and the cross section of cryptocurrency returns," *International Review of Financial Analysis*, vol. 89, p. 102712, 2023.

[14] L. Maciel, R. Ballini, and F. Gomide, "Adaptive level set fuzzy modeling," in *NAFIPS International Conference on Fuzzy Systems, Soft Computing, and Explainable AI, South Padre Island, TX, USA.*, 2024.

[15] R. Yager, "An alternative procedure for the calculation of fuzzy logic controller values," *Journal of the Japanese Soc. for Fuzzy Theory and Systems*, vol. 3, pp. 736–746, 1991.

[16] T. G. Andersen, T. Bollerslev, and F. X. Diebold, "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility," *The Review of Economics and Statistics*, vol. 89, no. 4, pp. 701–720, 2007.

[17] P. Angelov and X. Gu, *Empirical Approach to Machine Learning*. Springer, 2019.

[18] X. Gu, J. Han, Q. Shen, and P. Angelov, "Autonomous learning for fuzzy systems: A review," *Artificial Intelligence Review*, vol. 56, pp. 7549–7595, 2023.

[19] J. Junior, J. Mendes, F. Souza, and C. Premebida, "Survey on deep fuzzy systems in regression applications: A view on interpretability," *International Journal of Fuzzy Systems*, vol. 25, pp. 2568–2589, 2023.

[20] H. Rong, Z. Yang, and P. Wong, "Robust and noise-insensitive recursive maximum correntropy-based evolving fuzzy system," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 9, pp. 2277–2284, 2020.

[21] L. Jung, *System Identification: Theory for the User*. Upper Saddle River, Pearson, 1999.

[22] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *Proceedings of the IEEE Control and Decision Conference, San Diego*, 1979, pp. 761–766.

[23] Y. Ding, D. Kambouroudis, and D. G. McMillan, "Forecasting realised volatility: Does the LASSO approach outperform HAR?" *Journal of International Financial Markets, Institutions and Money*, vol. 74, p. 101386, 2021.

[24] P. R. Hansen, A. Lunde, and J. M. Nason, "The model confidence set," *Econometrica*, vol. 79, no. 2, pp. 453–497, 2011.