A Regime-Switching Approach to Bitcoin Volatility Forecasting

Cheung Yu Jeng Ethan, A0239399H

National University of Singapore

DSE4101: Capstone in Data Science and Economics

27 March, 2025

Abstract

Cryptocurrency markets, particularly Bitcoin, are characterized by extreme volatility and complex market dynamics. This study investigates the potential of Markov Regime-Switching (MRS) models to improve Bitcoin volatility forecasting by comparing single-regime and two-regime approaches using Heterogeneous Auto-Regressive (HAR) and Random Forest (RF) models. By distinguishing between high and low volatility states, we demonstrate that regime-switching methodologies can provide more nuanced and potentially more accurate volatility predictions. Our results reveal statistically significant improvements in forecasting accuracy during high volatility periods, with the two-regime HAR model outperforming single-regime baselines. Furthermore, through comprehensive model interpretability analyses, we elucidate the feature contributions of the forecasting models, hoping to enhance our understanding of the underlying drivers of Bitcoin volatility. These findings have important implications for risk management, trading strategies, and cryptocurrency market analysis.

Keywords: Cryptocurrency Volatility, Bitcoin, Markov Regime-Switching, HAR Model, Random Forest, Diebold-Mariano Test

Introduction

1.1 The Development and Volatility of Bitcoin

Since its inception in 2009, Bitcoin has spearheaded the evolution of the cryptocurrency market, introducing a decentralized digital currency paradigm that challenges traditional financial systems. Initially perceived as a niche technological experiment, Bitcoin has matured into a globally recognized financial asset, often referred to as "digital gold" due to its fixed supply and decentralized nature. This transformation has been marked by significant milestones, including its adoption as legal tender by El Salvador as early as 2021 (Hall, 2025), underscoring its growing legitimacy on the world stage. Interestingly, by early 2025, El Salvador reversed this decision, due to challenges in adoption and pressure from international financial institutions.

The cryptocurrency market has experienced exponential growth to a market capitalisation of over $2.7T USD in March 2025, up 1,460% from 5 years prior (TradingView, n.d.). Bitcoin's dominance within this market is evident, consistently accounting for a more than 40% of the total market capitalization. Its appeal as an investment vehicle and store of value has been bolstered by its decentralized framework, scarcity, and the increasing interest from institutional investors.

In recent developments, the United States has signalled a strategic shift towards embracing Bitcoin, with initiatives such as the establishment of a Strategic Bitcoin Reserve aimed at diversifying national assets and stabilizing the dollar's value during economic stress (The White House, 2025). This move reflects a broader trend of integrating Bitcoin into national reserves, highlighting its perceived value as a hedge against inflation and economic uncertainty.

Despite its growing acceptance, Bitcoin remains highly volatile compared to traditional assets like gold. While gold has long been regarded as a stable store of value, Bitcoin's price fluctuations are more pronounced, influenced by factors such as regulatory developments and security concerns, technological advancements, and broader macroeconomic factors (Lyócsa et al., 2020). This volatility poses challenges for investors and underscores the need for sophisticated forecasting models to navigate the complex dynamics of the cryptocurrency market, in order to develop effective risk management strategies and enhance the stability of cryptocurrency investments.

1.2 Research Background

This research builds on a previous group project that focused on forecasting Ethereum’s volatility, where it was found that Heterogeneous Autoregressive (HAR) models and Random Forest (RF) approaches outperformed traditional GARCH and the deep learning Long-Short-Term Memory (LSTM) neural network. The strong performance of these methods suggests that capturing longer-term volatility components (as with HAR) and exploiting non-linear relationships (as with RF) can lead to more accurate cryptocurrency volatility forecasts.

In addition, regime-switching frameworks have gained traction in recent years. A two-regime approach, as advocated by Ardia et al. (2018), can capture regime-dependent dynamics such as the stark contrast between high and low volatility periods often observed in Bitcoin markets. Leveraging these insights, the present study designs and implements a Markov-switching (two-regime) model that incorporates HAR and RF in each regime. By differentiating between volatile and tranquil market states, we aim to not only enhance forecasting accuracy but also improve model interpretability, offering clearer insights into the drivers of Bitcoin's price fluctuations.

1.3 Research Objectives

This paper aims to address several key questions relevant to both academics and practitioners in cryptocurrency markets:

1. **Regime Impact on Forecasting**: Does incorporating distinct volatility regimes (i.e., high vs. low) substantially improve Bitcoin volatility forecasts over standard single-regime approaches?
2. **Comparative Model Performance**: How do single-regime and two-regime models—specifically HAR and RF—perform across different market conditions, and which approach most effectively captures Bitcoin's pronounced volatility shifts?
3. **Risk Management Implications and Interpretability**: What are the implications of adopting a regime-switching approach for risk management in cryptocurrency markets, and how can interpretability methods (e.g., feature importance in RF, coefficient analysis in HAR) elucidate the underlying drivers of volatility?

By answering these questions, this study hopes to contribute to a deeper understanding of Bitcoin's complex market dynamics and offer robust forecasting tools that can be employed in risk management, trading strategies, and broader cryptocurrency market analysis.

Literature Review

2.1 Volatility Forecasting in Cryptocurrencies

The inherently volatile and nascent nature of cryptocurrency markets, particularly Bitcoin, has prompted significant academic interest in forecasting methodologies capable of capturing their complex dynamics. Initially, traditional econometric frameworks, such as Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, gained popularity due to their ability to account for volatility clustering and conditional variance. Studies have consistently demonstrated the effectiveness of GARCH and its variations in modelling the short-term volatility dependencies characteristic of Bitcoin prices (Katsiampa, 2017)

However, given the non-linear and evolving nature of cryptocurrency price movements, researchers have increasingly turned toward advanced machine learning approaches. Among these, the Heterogeneous Autoregressive (HAR) model and Random Forest (RF) methods have emerged as particularly promising. The HAR model, initially proposed by Corsi (2008), addresses volatility forecasting by incorporating multiple time horizons—daily, weekly, and monthly—capturing both short- and long-term volatility patterns simultaneously. Random Forests (Breiman, 2001), as ensemble learning methods, aggregate numerous decision trees to improve predictive robustness and reduce overfitting risks. They provide considerable flexibility in handling non-linear relationships and complex interactions between predictive features, qualities particularly valuable in cryptocurrency market analyses.

Empirical studies have demonstrated that both HAR and RF consistently outperforms both linear econometric models and several deep learning frameworks in forecasting cryptocurrency volatility (Dudek et al., 2022). In our preceding group project focused on Ethereum volatility forecasting, the HAR and RF models also outperformed traditional GARCH and sophisticated LSTM neural networks, including hybrid approaches such as GARCH-LSTM (Amirshahi & Lahmiri, 2023). These results underscore the efficacy of HAR and RF in addressing the nuanced volatility characteristics of cryptocurrencies. However, given the market's dynamic regime shifts, even these advanced models may fall short when applied in isolation, motivating further exploration into regime-switching frameworks.

2.2 Single-Regime vs. Multi-Regime Approaches

Single-regime models, such as standard HAR and RF, often face limitations in capturing the heterogeneous nature of cryptocurrency markets. To address this, Markov-Switching models have been proposed as a more sophisticated approach, allowing for structural changes in the underlying data-generating process. Ardia et al. (2018) showed the presence of regime changes in the GARCH volatility dynamics of Bitcoin log–returns, and that a two regime specification for high and low volatility states led to the best fit in-sample.

(Tan et al., 2021) introduced a time-varying transition probability Markov-switching GARCH (TV-MSGARCH) model that incorporates Bitcoin's daily trading volume and daily realized volatility as exogenous variables, effectively capturing the dynamic volatility patterns inherent in Bitcoin markets.

Furthermore, Ma et al. (2020) applied a Markov regime-switching mixed data sampling (MRS-MIDAS) model to forecast Bitcoin's realized volatility, finding statistically significant improvements over traditional models. This approach underscores the advantage of multi-regime models in adapting to the evolving market conditions characteristic of cryptocurrencies.

2.3 Research Gap

While regime-switching approaches have been explored in traditional financial markets and extended to applications in Bitcoin, the effectiveness of training machine learning models under different volatility regimes has not. This study is the first to address this gap by providing a comprehensive comparison of single-regime and two-regime volatility forecasting models.

Methodology

3.1.1 Data Collection

For this study, data was collected from three main sources to facilitate comprehensive Bitcoin volatility forecasting. High-frequency Bitcoin trading data at one-minute resolution from April 1, 2013, to October 8, 2023, was obtained from Kaggle (Zielak, 2025). Additionally, monthly macroeconomic indicators from the Federal Reserve Economic Data Monthly Database (FRED-MD, n.d.), encompassing key U.S. economic variables such as industrial production, inflation rates, interest rates, and employment figures, were included. Lastly, market sentiment was captured using the Crypto Fear and Greed Index, also sourced from Kaggle (Coetzee, 2023). This index integrates several indicators—market volatility, momentum, social media sentiment, Bitcoin dominance, and Google search trends—to quantify investor sentiment, providing a nuanced view of market emotions influencing cryptocurrency volatility.

3.1.2 Data Pre-Processing

To prepare the dataset for volatility forecasting, the following steps were undertaken:

Feature Selection: Utilizing the Least Absolute Shrinkage and Selection Operator (LASSO), six key predictors were identified: 'high', 'volume', 'UEMP15T26', 'UEMP27OV', 'PERMITW', and 'TB3SMFFM'. The latter four are macroeconomic indicators sourced from the Federal Reserve Economic Data (FRED) database:

- **UEMP15T26:** Represents the number of individuals unemployed for 15 to 26 weeks, indicating medium-term unemployment trends.
- **UEMP27OV:** Measures long-term unemployment (27 weeks or more), serving as a critical indicator of prolonged joblessness and economic downturns.
- **PERMITW:** Tracks building permits issued for new private housing units in the Western U.S., acting as a leading indicator of housing market activity.
- **TB3SMFFM:** Calculates the spread between the 3-month Treasury Bill rate and the Federal Funds Rate, reflecting the monetary policy stance and liquidity conditions.

Realized Variance Calculation: Daily realized variance (RV) was computed as a proxy for true volatility by summing squared minute-level log returns. To account for missing data points, the sum was normalized by $\frac{1440}{N}$, where N is the number of observations for the day.

Target Variable Transformation: The natural logarithm of the next day's realized variance, $\ln(\sigma_{t+1}^2)$ was used as the target variable. To capture temporal dependencies, lagged features were constructed, including one-day lags and weekly and monthly averages of the predictors.

3.2 Regime Classification

We modelled the volatility process as a Hidden Markov Model (HMM) with two states, hypothesizing that the Bitcoin market alternates between a "High Volatility" regime and a "Low Volatility" regime. The model was fitted on selected features – specifically the logarithms of daily, weekly, and monthly realized volatility, along with trading volume (Tan et al., 2021) – after applying a standard scaling procedure.

The HMM parameters (initial state distribution, transition probabilities, state-dependent means, and covariances) were estimated using the Expectation-Maximization algorithm. The HMM was trained on the entire dataset to obtain a smoothed state sequence, i.e., a sequence of state estimates that leverage both past and future data.

In our analysis, we experimented with two approaches for inferring state probabilities:

- **Forward Filtering:** This method computes the state probability at each time t based solely on the observations up to time t, i.e. $p(s_t \mid x_1, \ldots, x_t)$ Although ideal for real-time forecasting applications where only past information is available, forward filtering tends to yield more volatile state estimates due to its inherent lack of future information.

- **Smoothed Probabilities:** In contrast, smoothed probabilities are obtained by considering the entire observation sequence $x_1, \ldots, x_T$ when estimating the state at time t. This retrospective approach generally produces more stable and accurate regime classifications.

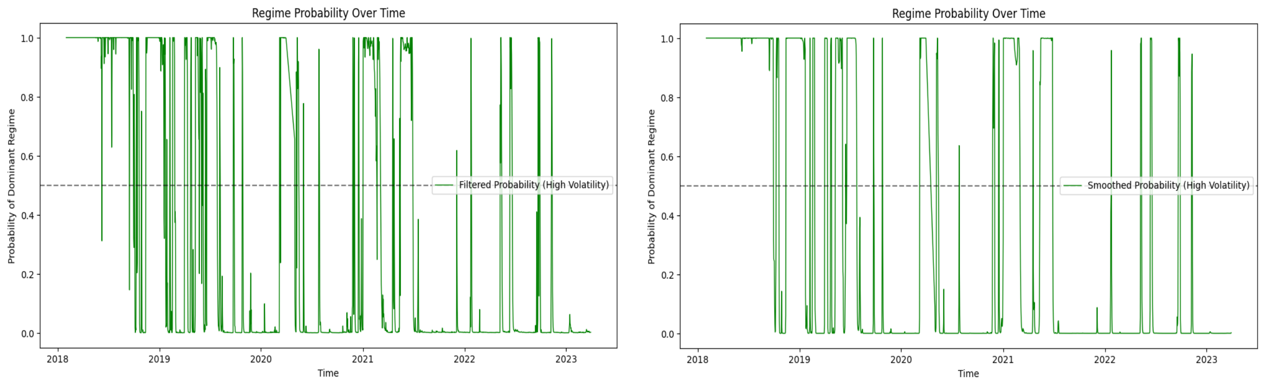

Figure 1: Forward Filtered Regime Probabilities (left) are less stable than Smoothed Regime Probabilities (Right)

While the forward filtering approach was appealing for real world applications, the stratified train-test split used later in our analysis (which does not strictly adhere to chronological order) rendered the need to address look-ahead bias less critical. Empirical comparisons showed that the smoothed probabilities resulted in more consistent regime delineations. Therefore, for subsequent model training and evaluation, we adopted the smoothed probabilities as our regime labels, shown below in Figure 2.

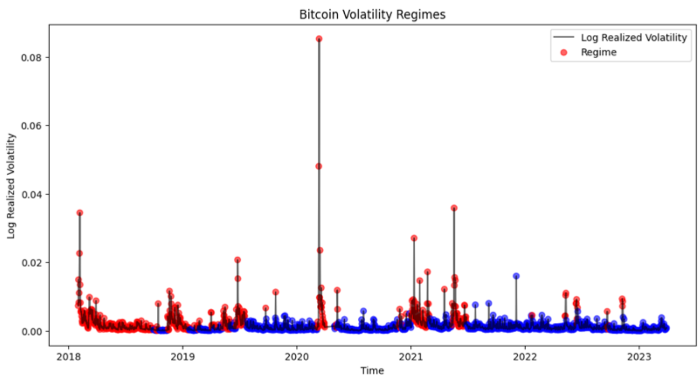

Figure 2: Realised Volatility with Classified Volatility Regimes

3.3 Stratified Train-Test Split

Given that the distribution of volatility regimes is not uniform over time—with periods of sustained high volatility often clustering—the conventional chronological split risked underrepresenting one of the regimes in either the training or testing datasets. To mitigate this, we implemented a stratified split approach, where the entire dataset was first classified into regimes using the HMM, and then the data were partitioned such that both the training and test sets contained representative samples of high and low volatility periods. All data points within each regime were combined and then split chronologically

(70% training and 30% testing). The train-test split for the one-regime baseline models was implemented by splitting the entire dataset chronologically.

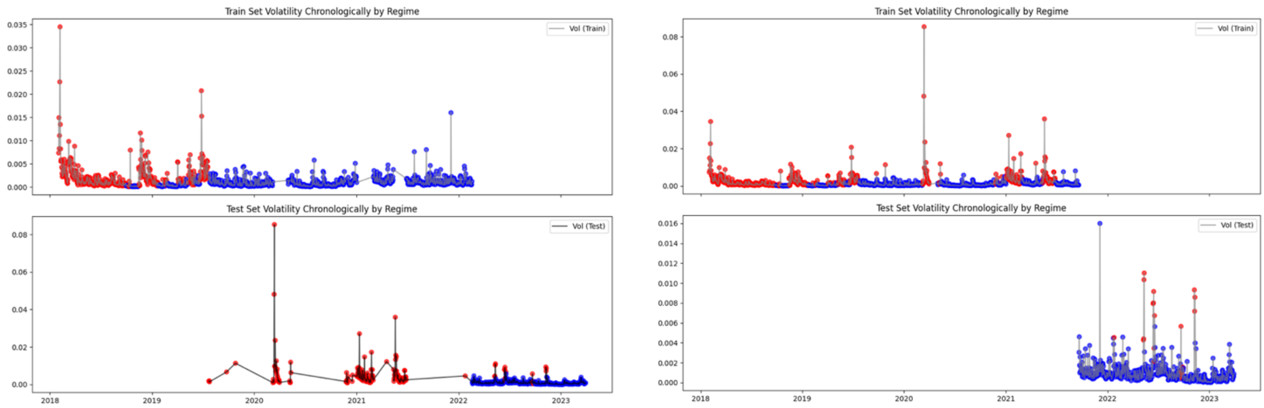

Figure 3: Train-test splits for two-regime models (left), and one-regime models (right)

3.4 Model Development

In this section, we detail the forecasting models used to predict Bitcoin's next-day volatility, namely the Heterogeneous AutoRegressive (HAR) model and the Random Forest (RF) model. Both models are applied within a regime-switching framework, where separate models are trained for periods of "High Volatility" and "Low Volatility." We also construct a one-regime baseline by training a single HAR and RF model on the entire training dataset.

Heterogeneous AutoRegressive (HAR) Model is designed to capture the persistence and long memory in volatility by incorporating information from multiple time horizons. Specifically, the HAR model for predicting next-day volatility is formulated as follows:

$$\begin{aligned} \text{target} = \beta_0 + \beta_1 \, &\text{ln\_RV\_d,t} \\ &+ \beta_2 \, \text{ln\_RV\_w,t} + \beta_3 \, \text{ln\_RV\_m,t} + \epsilon_t, \end{aligned}$$

where:

$\text{target} = \ln\left(\sigma_{t+1}^2\right)$ is the log-realized volatility for the next day,

$\text{ln\_RV\_d,t} = \ln\left(\sigma_t^2\right)$ is the daily volatility,

$\text{ln\_RV\_w,t} = \ln\left(\frac{1}{7}\sum_{i=0}^{6}\sigma_{t-i}^2\right)$ is the weekly average volatility,

$\text{ln\_RV\_m,t} = \ln\left(\frac{1}{30}\sum_{i=0}^{29}\sigma_{t-i}^2\right)$ is the monthly average volatility, and ϵ_t is the error term.

The HAR model is estimated using Ordinary Least Squares (OLS) regression. This linear framework is particularly appealing because of its simplicity, interpretability, and demonstrated forecasting performance in capturing volatility dynamics.

Random Forest (RF) Model is an ensemble, non-parametric method that constructs multiple decision trees during training and outputs the average prediction of the individual trees. This method is well-suited to capture nonlinear relationships and interactions among predictors that might be missed by linear models like HAR.

For our RF model, we use a feature set that includes not only the same volatility measures used in the HAR model but also the additional predictors determined from the above feature selection. Formally, the RF model is an approximation to the unknown function f(·) such that:

$$\begin{aligned}\text{target} = f\Big(&\text{ln\_RV\_d,t},\ \text{ln\_RV\_w,t},\ \text{ln\_RV\_m,t},\ \text{high},\\ &\text{volume},\ \text{UEMP15T26},\ \text{UEMP27OV},\ \text{PERMITW},\\ &\text{TB3SMFFM},\ \text{Fear\_Greed\_Value}\Big)\ +\ \epsilon_t.\end{aligned}$$

Unlike the HAR model, the RF does not assume any specific parametric form for f(·). Instead, it learns complex patterns and interactions from the data by aggregating the outputs of many decision trees, each built on bootstrapped samples of the data.

Regime-Specific Modelling

In the regime-switching framework, we train separate HAR and RF models for each volatility regime (High and Low). The process is as follows:

1. For the **one-regime baseline**, a single HAR and RF model are trained using the entire training set (chronological).
2. For the **two-regime models**, separate HAR and RF models are trained on data subsets (stratified) corresponding to each regime.

3.5 Evaluation Metrics

We consider four key performance metrics: **Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), R^2 (coefficient of determination), and Quasi Likelihood (Q-LIKE) loss function**. Each metric provides unique insights into the accuracy and effectiveness of the models.

Root Mean Squared Error (RMSE) measures the absolute difference between predicted and actual RV, with higher penalties for larger errors. This makes it a crucial metric for volatility forecasting, where extreme fluctuations must be captured accurately.

Mean Absolute Percentage Error (MAPE) represents relative error, indicating how much the predictions deviate from actual values in percentage terms. However, MAPE may be less reliable in evaluating the extreme nature of volatility.

R^2 (coefficient of determination) measures how well a model explains the variance in realised volatility. A higher R^2 suggests that the model provides better explanatory power.

Quasi-Likelihood (Q-LIKE) assesses how well a model fits the probabilistic structure of volatility. It is an asymmetric measure that penalises the underestimation of volatility, and therefore, it is especially relevant in financial risk assessment applications that rely on conservative estimates. Hence, under-estimation incurs higher costs. The Q-LIKE loss function is computed using the formula:

$$Q\text{-Like} = \mathbb{E}\left[\frac{y_t}{\hat{y}_t} - \ln\left(\frac{y_t}{\hat{y}_t}\right) - 1\right],$$

where y_t represents actual values and $\hat{y}_t$ denotes predicted values.

Statistical significance of the differences in performance of our competing models was assessed using the Diebold-Mariano (DM) test (Diebold and Mariano, 2002). We used the *dm_test* package created by the GitHub user *johntwk*.

Empirical Findings And Discussion

4.1 Model Performance

Common Dates Test Set Performance: Evaluating model performance on common test dates allows for direct comparison across all models. The results (Table 1, see Appendix) demonstrate that, during high volatility periods, the Two-Regime HAR model achieves the lowest RMSE (0.003346), outperforming the One-Regime HAR (0.003371), One-Regime RF (0.003721), and Two-Regime RF (0.003406) models. Additionally, the Two-Regime HAR model shows superior R^2 (0.0313) and a favourable QLIKE value (0.3356), further supporting its predictive capability during periods of market turbulence.

In low volatility periods, the Two-Regime HAR again performs best, achieving the lowest RMSE (0.000869) and highest R^2 (0.1376). It also maintains the lowest QLIKE score (0.3297), indicating its robust accuracy across different market conditions. Conversely, the RF models consistently exhibit weaker performance, particularly highlighted by significantly higher MAPE values in both volatility regimes, such as 127.70% for One-Regime RF in low volatility.

Overall, the Two-Regime HAR model provides the most reliable forecasts, consistently yielding the lowest RMSE (0.001148), highest R^2 (0.3520), and a competitive QLIKE score (0.3300) relative to other models.

Individual Model Performance on Original Test Sets: Performance on the original, regime-specific test sets further confirms the efficacy of distinguishing volatility regimes. Within the one-regime models, the HAR variant outperforms its RF counterpart across all volatility conditions, most notably in low volatility periods with significantly lower MAPE (61.47% vs. 108.49%) and better R^2 (0.0815 vs. 0.0501).

Two-regime models exhibit similar patterns, with HAR consistently outperforming RF. Specifically, the Two-Regime HAR achieves notably better results in high volatility periods (RMSE = 0.007944, R^2=0.0897, QLIKE = 0.5782) compared to the Two-Regime RF (RMSE = 0.008050, R^2=0.0653, QLIKE = 0.7126). These results reinforce the value of utilizing separate models to handle distinct market states effectively.

4.3 Statistical Significance

To validate whether observed differences in predictive accuracy were statistically significant, Diebold-Mariano tests were conducted (Table 2, see Appendix). In overall evaluations, significant statistical differences were found between One-Regime HAR and RF (DM = -2.5835, p=0.0101), favouring HAR, and similarly between One-Regime RF and Two-Regime HAR (DM = 2.6217, p=0.0091), favouring Two-Regime HAR. Furthermore, significant outperformance of the Two-Regime RF over the One-Regime RF model was observed (DM = 2.4890, p=0.0132), indicating the incremental benefit of distinguishing regimes.

In high volatility conditions, both One-Regime and Two-Regime HAR significantly outperform their RF counterparts (DM = -2.5612, p=0.0182 and DM = 2.7839, p=0.0111, respectively). Similarly, Two-

Regime RF outperforms One-Regime RF significantly (DM = 2.6493, p=0.0150). These findings strongly validate the advantage of regime differentiation, especially during volatile market phases.

Conversely, no significant differences are found among model performances in low volatility conditions, suggesting that the forecasting benefit of regime differentiation may be less pronounced or harder to detect in stable market periods.

4.4 Model interpretations

This section provides insights into the factors driving performance differences among volatility forecasting models, particularly focusing on the effectiveness of two-regime models in capturing regime-specific dynamics.

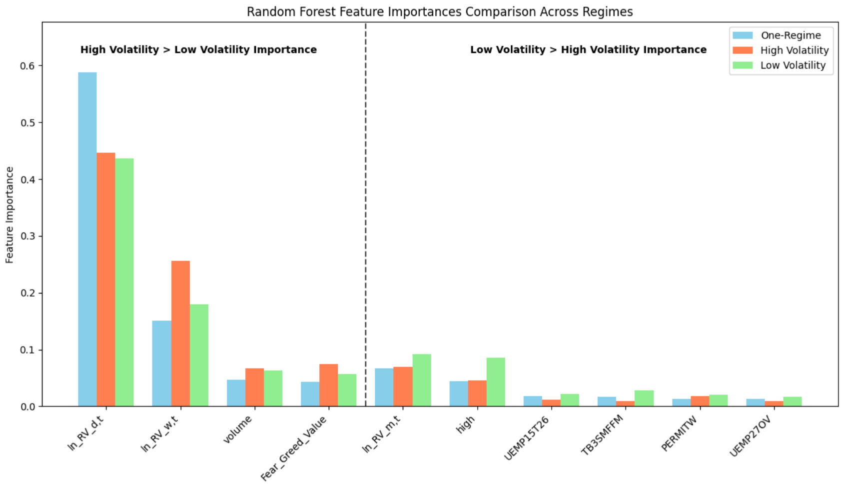

Figure 4: RF Feature Importance, features where high volatility model has higher weights than the low volatility model, is grouped on the left

The significant improvement of two-regime models over single-regime models, especially during high volatility periods, can be attributed primarily to their ability to adaptively prioritize features according to market conditions. From the feature importance analysis of the Random Forest models (see Figure 4 below and Table 4 in Appendix), we observe a clear distinction between high and low volatility regimes. Specifically, the Random Forest models assigned higher importance to short-term, responsive indicators such as daily (ln_RV_d,t) and weekly (ln_RV_w,t) realized volatility, as well as market sentiment proxies like trading volume and the Fear and Greed index during high volatility states. This aligns with the intuitive understanding that market participants react more swiftly to recent information and sentiment shifts when volatility surges.

In contrast, during periods of low volatility, the Random Forest models gave relatively greater weight to long-term indicators, including monthly realized volatility (ln_RV_m,t) and macroeconomic variables (e.g., unemployment rates, interest rates, and building permits), all of which are slowermoving, monthly-frequency data. These indicators better reflect stable, long-term economic trends rather than short-term fluctuations.

The patterns observed in the HAR model coefficients (Figure 5 and Table 3, see Appendix) further corroborate this finding. The HAR model for high volatility periods had a notably larger coefficient for daily volatility compared to its low volatility counterpart. Conversely, weekly and monthly volatility had smaller coefficients in the high volatility regime. Such differences indicate that the HAR models also dynamically adjust their emphasis on volatility horizons based on the volatility regime.

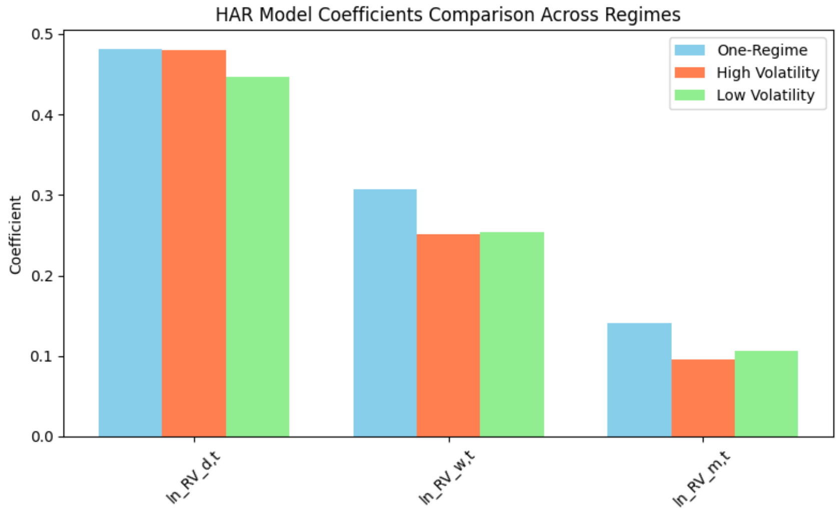

Figure 5: HAR Model Coefficients for Daily RV, Weekly and Monthly Averages of RV

This regime-dependent specialization significantly contributes to the superior predictive capability of two-regime models in high volatility environments, as verified by the statistically significant Diebold-Mariano test results (Table 2). Conversely, the lack of statistical significance during low volatility periods indicates that single-regime models are sufficiently capturing these more stable dynamics, thus reducing the relative advantage of regime-switching.

In summary, the interpretability analysis demonstrates that two-regime models provide tailored forecasts by leveraging short-term, responsive indicators during volatile periods and stable, longer-term indicators during calm periods. This dynamic adaptation enhances the model's ability to accurately forecast volatility, especially during times of heightened market uncertainty.

Conclusion

5.1 Key Findings

This study enhances our understanding of Bitcoin volatility forecasting by systematically evaluating the efficacy of regime-switching approaches compared to traditional single-regime frameworks. Our comprehensive analysis demonstrates that a two-regime Markov-switching model utilizing Heterogeneous Auto-Regressive (HAR) forecasting methods significantly improves forecasting accuracy, particularly during periods of heightened market volatility. Through rigorous performance evaluations and statistical testing using Diebold-Mariano tests, we found strong empirical support for the superiority of two-regime models, especially under high-volatility market conditions.

The interpretability analysis further revealed distinct behavioural patterns of model predictors across volatility regimes. During high volatility periods, the Random Forest and HAR models consistently emphasized short-term responsive features such as daily volatility, weekly volatility, trading volume, and market sentiment indicators. Conversely, during calmer periods, models relied more heavily on longer-term indicators, including monthly volatility and macroeconomic variables. These findings illustrate that the advantage of regime-switching models primarily stems from their adaptive ability to adjust feature importance dynamically according to market conditions.

5.2 Implications

The results of this research offer several significant implications across practical and academic domains:

Risk Management: By capturing volatility dynamics more accurately, regime-switching models can substantially enhance risk metrics such as Value-at-Risk and Expected Shortfall (Taamouti, 2009). Financial institutions and investors can leverage these improved forecasts for more informed portfolio allocation and risk mitigation strategies.

Trading Strategies: Traders can utilize improved volatility predictions to optimize their market entry, exit, and hedging decisions. Regime-dependent forecasting can help identify opportunities and manage risks more effectively, particularly in volatile trading environments.

Academic Research: This research contributes methodologically to volatility forecasting literature by validating the effectiveness of regime-switching frameworks in the cryptocurrency market context. It provides a foundation for future investigations into advanced modelling approaches and further exploration of volatility dynamics in emerging asset classes.

5.3 Limitations

While our results demonstrate significant advancements, the study is subject to several limitations:

Limited Data Range: Our analysis covers a fixed historical period from 2013 to 2023. This may not fully capture evolving market dynamics, regulatory changes, and structural breaks occurring beyond this timeframe.

Look-Ahead Bias: The offline classification of volatility regimes via Hidden Markov Models may introduce a look-ahead bias, potentially overestimating the real-world predictive performance of the models.

Noise from Stratified Train-Test Split: The training data splices disjointed ends of separate periods of market regimes, which could introduce noise into model training and evaluation, affecting the accuracy and robustness of regime-specific predictions.

5.4 Future Work

Addressing the limitations identified, future research can extend our findings through several promising directions:

Enhanced Sentiment and Macroeconomic Inputs: Future studies could incorporate more sophisticated market sentiment indicators. The effects of macroeconomic features could be enhanced by using Mixed Data Sampling (MIDAS) models, to better model the influence of monthly/quarterly frequency data on daily Bitcoin volatility (Ghysels et al., 2019).

Real-time Regime Classification: Developing real-time or online regime classification methodologies would enhance practical applicability and eliminate potential look-ahead biases, leading to more reliable forecasting and operational decision-making frameworks.

In summary, the study underscores the practical utility and theoretical significance of regime-switching approaches to cryptocurrency volatility forecasting. Continued exploration in this area holds great potential for enhancing financial decision-making and advancing academic understanding of cryptocurrency market dynamics.

References

Amirshahi, B., & Lahmiri, S. (2023). Hybrid deep learning and GARCH-family models for forecasting volatility of cryptocurrencies. *Machine Learning With Applications*, *12*, 100465. https://doi.org/10.1016/j.mlwa.2023.100465

Ardia, D., Bluteau, K., & Rüede, M. (2018). Regime changes in Bitcoin GARCH volatility dynamics. *Finance Research Letters*, *29*, 266–271. https://doi.org/10.1016/j.frl.2018.08.009

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Coetzee, L. (2023). *Bitcoin - fear and greed* [Dataset]. https://www.kaggle.com/code/lindleylawrence/bitcoin-fear-and-greed/

Corsi, F. (2008). A simple approximate Long-Memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196. https://doi.org/10.1093/jjfinec/nbp001

Crypto Dominance Chart — BTC Dominance — TradingView. (n.d.). TradingView. Retrieved March 28, 2025, from https://www.tradingview.com/markets/cryptocurrencies/dominance/

Dudek, G., Fiszeder, P., Kobus, P., & Orzeszko, W. (2023). Forecasting cryptocurrencies volatility using statistical and machine learning methods: A comparative study. *Applied Soft Computing*, *151*, 111132. https://doi.org/10.1016/j.asoc.2023.111132

Federal Reserve Bank of St. Louis. (n.d.). *FRED-MD: A monthly database for macroeconomic research* [Data set]. Federal Reserve Bank of St. Louis. Retrieved April 4, 2025, from https://files.stlouisfed.org/files/htdocs/fred-md/monthly/current.csv

Ghysels, E., Kvedaras, V., & Zemlys-Balevičius, V. (2019). Mixed data sampling (MIDAS) regression models. In *Handbook of statistics* (pp. 117–153). https://doi.org/10.1016/bs.host.2019.01.005

Hall, I. (2025, March 11). *El Salvador agrees with IMF to 'confine' Bitcoin use*. Global Government Fintech. https://www.globalgovernmentfintech.com/el-salvador-imf-agreement-confine-bitcoin-use/

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters*, *158*, 3–6. https://doi.org/10.1016/j.econlet.2017.06.023

Lyócsa, Š., Molnár, P., Plíhal, T., & Širaňová, M. (2020). Impact of macroeconomic news, regulation and hacking exchange markets on the volatility of bitcoin. *Journal of Economic Dynamics and Control*, *119*, 103980. https://doi.org/10.1016/j.jedc.2020.103980

Ma, F., Liang, C., Ma, Y., & Wahab, M. (2020). Cryptocurrency volatility forecasting: A Markov regime-switching MIDAS approach. *Journal of Forecasting*, *39*(8), 1277–1290. https://doi.org/10.1002/for.2691

Mariano, F. X. D. a. R. S. (2002). Comparing predictive accuracy. Journal of Business & Economic Statistics, 20(1), 134–144. https://www.jstor.org/stable/1392155

Taamouti, A. (2009). Analytical Value-at-Risk and Expected Shortfall under regime-switching. *ideas.repec.org*. https://ideas.repec.org/a/eee/finlet/v6y2009i3p138-151.html

Tan, C., Koh, Y., Ng, K., & Ng, K. (2021). Dynamic volatility modelling of Bitcoin using time-varying transition probability Markov-switching GARCH model. *The North American Journal of Economics and Finance*, *56*, 101377. https://doi.org/10.1016/j.najef.2021.101377

The White House. (2025, March 18). *Establishment of the Strategic Bitcoin Reserve and United States Digital Asset Stockpile*. https://www.whitehouse.gov/presidential-actions/2025/03/establishment-of-the-strategic-bitcoin-reserve-and-united-states-digital-asset-stockpile/

Zielak. (2025). *Bitcoin Historical Data* [Dataset]. https://www.kaggle.com/datasets/mczielinski/bitcoin-historical-data

Appendix

Detailed Results Tables

Table 1: Model Performance Metrics

Model	Regime	Performance Metrics				Count
		RMSE	MAPE (%)	R^2	Q-LIKE	
Common Dates Test Set Predictions[1]						
One-Regime HAR	High Volatility	0.003371	49.90	0.0163	0.3470	22
One-Regime RF	High Volatility	0.003721	55.04	0.1982	0.5218	22
Two-Regime HAR	High Volatility	0.003346	50.49	0.0313	0.3356	22
Two-Regime RF	High Volatility	0.003406	58.44	0.0041	0.3681	22
One-Regime HAR	Low Volatility	0.000870	68.89	0.1350	0.3251	385
One-Regime RF	Low Volatility	0.000887	127.70	0.1003	0.3298	385
Two-Regime HAR	Low Volatility	0.000869	66.81	0.1376	0.3297	385
Two-Regime RF	Low Volatility	0.000874	106.14	0.1264	0.3476	385
One-Regime HAR	Overall	0.001153	67.86	0.3463	0.3263	407
One-Regime RF	Overall	0.001222	123.77	0.2664	0.3402	407
Two-Regime HAR	Overall	0.001148	65.92	0.3520	0.3300	407
Two-Regime RF	Overall	0.001162	103.56	0.3366	0.3487	407
Individual Model Performance on Original Test Sets						
One-Regime Models[2]						
One-Regime HAR	High Volatility	0.003371	49.90	0.0163	0.3470	22
One-Regime RF	High Volatility	0.003721	55.04	0.1982	0.5218	22
One-Regime HAR	Low Volatility	0.001074	61.47	0.0815	0.3013	535
One-Regime RF	Low Volatility	0.001093	108.49	0.0501	0.3091	535
One-Regime HAR	Overall	0.001248	61.02	0.2460	0.3031	557
One-Regime RF	Overall	0.001301	106.38	0.1802	0.3175	557
Two-Regime Models						
Two-Regime HAR	High Volatility	0.007944	44.12	0.0897	0.5782	172
Two-Regime RF	High Volatility	0.008050	53.04	0.0653	0.7126	172
Two-Regime HAR	Low Volatility	0.000869	66.81	0.1376	0.3297	385
Two-Regime RF	Low Volatility	0.000874	106.14	0.1264	0.3476	385
Two-Regime HAR	Overall	0.004473	59.80	0.2343	0.4064	557
Two-Regime RF	Overall	0.004532	89.74	0.2141	0.4603	557

[1] Common dates determined by the intersection of one-regime and two-regime test sets.

[2] Models were not differentiated by regime during training; regime classification performed post-hoc using a Hidden Markov Model.

Table 2: Diebold-Mariano Test Results Across Volatility Regimes

Models Compared	Volatility Regime	DM Statistic	P-Value	Conclusion
Overall Performance				
OneRegHAR vs OneRegRF	Overall	-2.5835	0.0101	OneRegHAR outperforms
OneRegHAR vs TwoRegHAR	Overall	1.4386	0.1510	No significant difference
OneRegHAR vs TwoRegRF	Overall	-0.3600	0.7190	No significant difference
OneRegRF vs TwoRegHAR	Overall	2.6217	0.0091	TwoRegHAR outperforms
OneRegRF vs TwoRegRF	Overall	2.489	0.0132	TwoRegRF outperforms
TwoRegHAR vs TwoRegRF	Overall	-0.5820	0.5609	No significant difference
High Volatility Regime				
OneRegHAR vs OneRegRF	High Vol	-2.5612	0.0182	OneRegHAR outperforms
OneRegHAR vs TwoRegHAR	High Vol	2.7839	0.0111	TwoRegHAR outperforms
OneRegHAR vs TwoRegRF	High Vol	-0.2450	0.1806	No significant difference
OneRegRF vs TwoRegHAR	High Vol	2.6562	0.0148	TwoRegHAR outperforms
OneRegRF vs TwoRegRF	High Vol	2.6493	0.0150	TwoRegRF outperforms
TwoRegHAR vs TwoRegRF	High Vol	-0.4309	0.6709	No significant difference
Low Volatility Regime				
OneRegHAR vs OneRegRF	Low Vol	-1.1798	0.2388	No significant difference
OneRegHAR vs TwoRegHAR	Low Vol	0.3104	0.7564	No significant difference
OneRegHAR vs TwoRegRF	Low Vol	-0.3366	0.7366	No significant difference
OneRegRF vs TwoRegHAR	Low Vol	1.1161	0.2651	No significant difference
OneRegRF vs TwoRegRF	Low Vol	0.8352	0.4041	No significant difference
TwoRegHAR vs TwoRegRF	Low Vol	-0.4816	0.6304	No significant difference

Table 3: HAR Model Coefficients Comparison Across Regimes

Predictor	**One-Regime**		**Two-Regime High**		**Two-Regime Low**	
	Coefficient	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**
const	-0.6024	0.0002	-7.0526	0.0000	-7.1480	0.0000
ln_RV_d,t	0.4808	0.0000	0.4799	0.0000	0.4471	0.0000
ln_RV_w,t	0.3076	0.0000	0.2517	0.0006	0.2542	0.0000
ln_RV_m,t	0.1406	0.0007	0.0953	0.0393	0.1068	0.0103

Table 4: Random Forest Feature Importances Comparison Across Regimes

Predictor	**One-Regime**	**Two-Regime High**	**Two-Regime Low**
ln_RV_d,t	0.5883	0.4461	0.4366
ln_RV_w,t	0.1507	0.2560	0.1791
ln_RV_m,t	0.0670	0.0697	0.0923
volume	0.0462	0.0671	0.0629
high	0.0445	0.0457	0.0852
Fear_Greed_Value	0.0424	0.0741	0.0567
UEMP15T26	0.0178	0.01178	0.0218
TB3SMFFM	0.0163	0.0098	0.0282
PERMITW	0.0134	0.0179	0.0201
UEMP27OV	0.0132	0.0087	0.0171