High-Frequency Forecasting of Bitcoin Volatility: Evaluating HAR Models with Realised Semivariance and Jump Components

Raheleh Sarrafshirazi

A

Thesis

In

The John Molson School of Business

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Finance) at

Concordia University

Montreal, Quebec, Canada

August 2025

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis

Prepared By: Raheleh Sarrafshirazi

Entitled: High-Frequency Forecasting of Bitcoin Volatility: Evaluating HAR Models with Realised Semivariance and Jump Components

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Finance)

complies with the regulations of the University and meets the accepted standards with respect tooriginality and quality.

Signed by the final Examining Committee:

Dr. Juliane Proelss Chair

Dr. Parianen Veeren Examiner

Dr. Mehmet Özsoy Examiner

Dr. Kun Ho Kim Supervisor

Approved by Dr. Frederick Davis Graduate Program Director

Dean *Dr. Anne-Marie Croteau*

ABSTRACT

High-Frequency Forecasting of Bitcoin Volatility: Evaluating HAR Models with Realised Semivariance and Jump Components

Raheleh Sarrafshirazi

Bitcoin's continuous trading, speculative nature, and high volatility create distinctive challenges for risk management and forecasting. This thesis examines how high-frequency realised volatility (RV) measures and Heterogeneous Autoregressive (HAR) models capture Bitcoin's volatility dynamics and improve forecast accuracy. Adapting RV methods from equity markets, the analysis adds downside semivariance to address asymmetric negative returns and jump variation to capture price movements.

Using minute-level Bitcoin prices from, I compute RV from 5-minute returns and estimate four HAR variants—baseline HAR, HAR-RS, HAR-J, and HAR-RS-J. Models are re-estimated in a rolling window, and forecasts are evaluated with RMSE, MAE, and QLIKE. Robustness checks test stability under different data granularities and market regimes. A GARCH(1,1) benchmark provides a parametric comparison, with HAR variants outperforming it at short horizons, while GARCH exceeds performance at longer horizons.

Results show that HAR-type models capture Bitcoin's long memory, volatility clustering, and asymmetry effectively. HAR-J delivers the most accurate day-ahead, while HAR-RS leads at weekly and monthly horizons due to persistent downside risk. At quarterly horizons, forecast accuracy converges across models as high-frequency information loses relevance.

This study extends RV–HAR modelling to cryptocurrency markets, revealing shorter volatility persistence and greater jump contributions than in equities. It identifies downside semivariance and continuous variation as robust predictors across different market conditions and offers horizon-specific tools—HAR-J for short-term risk management and HAR-RS for medium-term volatility planning.

Keywords: Bitcoin Volatility, Realised Volatility, Heterogeneous Autoregressive Model, Semivariance, Jump Variation, High-Frequency Data, Volatility Forecasting, Cryptocurrency Markets, Bipower Variation, Risk Management.

Acknowledgements

I would like to express sincere gratitude to my supervisor, Dr. Kun Ho Kim, for his invaluable guidance and support throughout my research. His constructive feedback and unwavering availability for discussions have been instrumental in the development of this thesis. I am sincerely grateful for his kindness, encouragement, and mentorship, which have profoundly shaped my academic journey. I would also like to thank my thesis committee members, Dr. Parianen Veeren and Dr. Mehmet Özsoy, for their time, insightful comments, and constructive feedback.

Lastly, I would like to express my appreciation to my family, friends, and colleagues for their support and encouragement, which have been a constant source of strength throughout this journey. I am deeply grateful to everyone who has contributed to my growth and helped shape the person I am today.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Bitcoin's emergence as a highly volatile, 24/7-traded asset has transformed financial markets and the analytical tools required to study them. Unlike traditional securities, Bitcoin operates across global exchanges without downtime, attracts diverse participants, and exhibits abrupt price jumps driven by speculative sentiment and rapid information diffusion. These features necessitate precise volatility forecasting for risk management, derivative pricing, and regulatory oversight in cryptocurrency markets. However, most studies rely on daily GARCH-type models, which overlook high-frequency data and fail to capture Bitcoin's unique volatility dynamics, such as asymmetry, inverted leverage effects, and discrete jumps.

Volatility modelling has evolved significantly over the past four decades, from Engle's (1982) ARCH and Bollerslev's (1986) GARCH frameworks to realised volatility (RV) approaches that leverage intraday data. In equity markets, RV-based Heterogeneous Autoregressive (HAR) models effectively capture long memory, volatility clustering, and multi-scale dynamics. Yet, few studies apply these advances to Bitcoin or incorporate RV extensions like realised semivariance or jump variation, despite their relevance to cryptocurrency-specific behaviours. This gap motivates a high-frequency approach tailored to Bitcoin's market microstructure and behavioural drivers.

This thesis addresses three research questions: (1) Can HAR-type models based on high-frequency RV effectively capture Bitcoin's long-memory volatility patterns? (2) Which components—downside semivariance or jump variation—add the most predictive value, and how does this vary across forecast horizons? (3) Are these results robust to changes in data granularity and market regimes, such as the post-COVID period characterized by lower volatility and increased institutional adoption?

To address these questions, I construct RV, bipower variation (BV), realised semivariance, and jump measures from 5-minute Bitcoin prices in an aggregated Kaggle dataset (August 2017–February 2025). I estimate four HAR variants—baseline HAR, HAR-RS, HAR-J, and HAR-RS-J—and, using a fixed-width rolling window equal to the initial 80 % of the sample (2,215 days), I generate out-of-sample forecasts at 1-, 7-, 30-, and 90-day horizons. Forecast accuracy is assessed

with RMSE, MAE, and QLIKE. Robustness checks with 10-minute sampling and a post-COVID subsample show the model hierarchy is stable across microstructure settings and market regimes.

The dataset reflects Bitcoin's mature trading environment post-Mt. Gox collapse and Binance's launch, capturing key volatility episodes, including the 2018 "crypto winter," the March 2020 COVID shock, and the 2021–2025 bull–bear cycles. This period ensures deep liquidity and modern exchange infrastructure, enabling RV measures to approximate true integrated volatility accurately.

This thesis contributes in three ways. Academically, it extends realised semivariance and jump-augmented HAR models to Bitcoin, offering new evidence on volatility dynamics in digital assets. Methodologically, it demonstrates the stability of forecast rankings across sampling frequencies and market regimes, highlighting the adaptability of HAR-type models. Practically, it provides guidance for risk management: HAR-J excels for day-ahead forecasts, while HAR-RS dominates at weekly and monthly horizons, driven by persistent downside risk.

The thesis is structured as follows. Chapter 2 reviews volatility modelling, from ARCH and GARCH to RV and HAR, emphasizing extensions relevant to cryptocurrencies. Chapter 3 describes the Bitcoin dataset, preprocessing, and descriptive statistics. Chapter 4 details the RV measures, HAR variants, and estimation strategy. Chapter 5 presents empirical results, including in-sample diagnostics, out-of-sample forecast rankings, and robustness checks. Chapter 6 summarizes key findings and academic contributions. Chapter 7 discusses practical implications, limitations, and future research directions.

By integrating high-frequency data, realised measures, and HAR-type models, this thesis bridges volatility forecasting methods from equity markets to the unique dynamics of cryptocurrency trading, delivering tools and insights tailored to Bitcoin's volatility profile.

Chapter 2 Literature Review

Volatility drives critical decisions in financial markets, from managing risks to pricing derivatives and shaping investment strategies. Over time, researchers have refined their approaches to modeling volatility, moving from early parametric models to methods that harness high-frequency data. This section explores that evolution, focusing on the development of ARCH and GARCH models, their strengths and limitations, and the emergence of RV as a powerful tool for capturing price fluctuations, particularly in dynamic markets like Bitcoins.

2.1 Traditional Volatility Models

Early efforts to model the changing nature of volatility began with Engle (1982), who introduced the Autoregressive Conditional Heteroskedasticity (ARCH) model, a major advance that ties a time series variance to past squared errors. Bollerslev (1986, 1987) extended this with the Generalized ARCH (GARCH) model, which became a standard for estimating and forecasting volatility. These models excel at capturing volatility clustering, where large price changes often lead to more large changes, and small changes to small ones (Cheikh et al., 2020). They also address the leverage effect, where negative returns increase volatility more than positive ones, a pattern Black (1976) identified and Bollerslev et al. (2006) and Brini and Lenz (2024) linked to market reactions to negative shocks.

However, despite their widespread use, these models exhibit several limitations that have spurred the development of alternative approaches. ARCH and GARCH models rely on daily returns, which carry significant noise, obscuring true volatility, or integrated volatility (Patton, 2011). This noise can weaken forecast accuracy, as Andersen and Bollerslev (1998) demonstrated. GARCH models also struggle to fully capture long-memory patterns where volatility persists over time or the fat-tailed distributions typical of financial returns (McAleer and Medeiros, 2008). If researchers fail to account for structural breaks, such as sudden shifts in price variance due to external events, these models may overestimate persistence (Aharon et al., 2023). Moreover, GARCH models cannot easily incorporate intraday data, limiting their ability to analyze fine-grained market dynamics (Bollerslev et al., 2016). These limitations prompted researchers to turn to high-frequency data, leading to the rise of RV measures (Zhang et al., 2005).

2.2 RV and High-Frequency Data

In response to the discussed challenges, RV emerged as a robust, model-free method to estimate price variation using high-frequency data. Andersen and Benzoni (2008) define RV as the sum of squared returns over short intervals, typically a day, using high-frequency data. Grounded in the theory of quadratic variation (Andersen et al., 2003), RV approximates integrated volatility more accurately as sampling frequency increases in idealized markets. Unlike noisy daily squared returns, RV offers precise and stable volatility estimates, making it ideal for studying fast-moving markets like cryptocurrencies (Andersen and Benzoni, 2008; Andersen et al., 2003). This advance paves the way for more sophisticated models, such as the Heterogeneous Autoregressive (HAR) framework, which extends RV's applicability across multiple time scales.

High-frequency data unlock precise volatility estimates, but they come with challenges. Market microstructure noise —think price discreteness, bid-ask bounces, non-synchronous trading, or data errors—can distort RV estimates, especially at very high sampling frequencies (Hansen and Lunde, 2006). This noise introduces spurious autocorrelations and biases, potentially leading to inconsistent or diverging estimates (Liu et al., 2015). To tackle this, researchers developed robust methods. Liu et al. (2015) proposed optimal and sparse sampling, Zhang et al. (2005) introduced the two-scales estimator, and Barndorff-Nielsen et al. (2004) designed kernel-based estimators to improve RV accuracy.

Beyond addressing measurement noise, another line of research focuses on decomposing total volatility into its components. Barndorff-Nielsen et al. (2004) introduced bipower variation (BV), which consistently estimates the integrated variance of the continuous price path while remaining robust to jumps. By subtracting BV from RV, researchers isolate the jump component, capturing discrete price movements that often spike volatility in financial markets (Barndorff-Nielsen et al., 2004). To address asymmetry in volatility, Patton and Sheppard (2015) developed realized semivariance, splitting RV into components tied to positive and negative returns. This approach retains sign information, offering deeper insights into volatility dynamics, especially in markets like Bitcoin (Zhang and Zhao, 2023). These advancements in RV methodologies have significantly improved the accuracy of volatility measurement in high-frequency financial data, laying the

groundwork for sophisticated forecasting models that leverage RV to predict volatility across multiple time scales, particularly in dynamic markets like cryptocurrencies.

2.3 HAR and Its Extensions

The Heterogeneous Autoregressive (HAR) model, proposed by Corsi (2009), is a cornerstone of volatility forecasting, leveraging RV to capture volatility dynamics across daily, weekly, and monthly horizons. Grounded in the Heterogeneous Market Hypothesis, which posits that traders operate at different time scales, the HAR-RV model effectively models long memory, fat tails, and self-similarity in financial returns (Maki and Ota, 2012). Its simplicity and strong forecasting performance have made it a widely adopted framework for volatility modeling, particularly in cryptocurrency markets (Maki and Ota, 2012). Building on this foundation, researchers have proposed several extensions to capture more nuanced volatility behavior, such as asymmetry, jumps, and measurement errors.

First, to address asymmetry, Corsi and Renò (2012) introduced the Leverage HAR (LHAR) model, incorporating past negative returns at daily, weekly, and monthly frequencies as predictors. This model highlights the persistent impact of negative returns on future volatility (Corsi and Renò, 2012). Baillie et al. (2019) further refined this by using realized semivariances to model good and bad volatility separately. Second, to account for jumps, researchers extended the HAR model with jump components. The HAR-RV-J and HAR-RV-CJ models, developed by Barndorff-Nielsen and Shephard (2004), Andersen et al. (2007), Corsi et al. (2010), Souček and Todorova (2014), and Chen et al. (2019), distinguish continuous volatility from jumps. While jumps cause short-lived volatility bursts, Baillie et al. (2019) found that continuous components drive most predictability. Zhang and Zhao (2023) also integrated signed jump variation, a measure of return skewness, to capture differing effects of positive and negative jumps. Third, to handle measurement errors, Bollerslev et al. (2016) proposed the HARQ model, which adjusts parameters based on integrated quarticity, a measure of error variance. This approach delivers conservative forecasts when RV estimates are noisy and responsive ones when estimates are precise (Bollerslev et al., 2016). Finally, researchers explored alternative inputs like realized absolute variation (RAV) and daily price ranges to enhance HAR models (Patton, 2011). Taken together, these extensions position the

HAR framework as a flexible and powerful tool for analyzing the distinctive volatility patterns observed in Bitcoin and other cryptocurrency markets.

2.4 Volatility in Cryptocurrency Markets

Cryptocurrency markets behave differently from traditional assets, offering fresh challenges for volatility modeling. Brini and Lenz (2024) highlight a striking pattern: unlike equity markets, where negative shocks typically amplify volatility, positive shocks in cryptocurrencies often trigger larger volatility spikes (Baur and Dimpfl, 2018). Researchers attribute this inverted leverage effect to speculative trading, psychological factors like fear of missing out (FOMO), and schemes such as pump and dump (Katsiampa, 2017). This behavior mirrors patterns in safe-haven assets like gold, as Soylu et al. (2020) note, setting cryptocurrencies apart from conventional markets.

Building on these behavioral differences, recent studies uncover additional insights into the unique volatility dynamics of cryptocurrencies. First, Zhang and Zhao (2023) find that volatility persistence in cryptocurrencies operates over shorter timeframes, often around one week, compared to longer persistence in equity markets. Second, they show that positive semivariance (RS^+), or good volatility, is a stronger predictor of future volatility than its negative counterpart RS^-, with higher RS^+ often leading to increased volatility levels. While negative jumps, captured by signed jump variation (SJV) as a measure of return skewness, tend to reduce future volatility—a contrast to equity markets (Zhang and Zhao, 2023). Third, Aharon et al. (2023) demonstrate that accounting for structural breaks in GARCH models reduces estimated volatility persistence and amplifies the asymmetric impact of unexpected news. Studies that overlook these breaks often underestimate news effects, leading to less accurate models (Aharon et al., 2023).

2.5 Model Evaluation Methods

Given the distinct volatility patterns, evaluating the predictive performance of competing models becomes especially important. To evaluate volatility models, researchers rely on robust methods. They often use loss functions like Mean Squared Error (MSE) or Quasi-Likelihood (QLIKE) to assess forecast accuracy (Patton, 2011). Hansen (2001) developed the Superior Predictive Ability (SPA) test, while White (2000) introduced the Reality Check (RC) to benchmark models while

handling multiple comparisons. Hansen et al. (2011) proposed the Model Confidence Set (MCS), which ranks models based on data-driven performance. These tools ensure fair comparisons across competing volatility models.

High-frequency data significantly enhance daily volatility forecasts compared to models relying solely on daily data (Andersen et al., 2003). Liu et al. (2015) compare HAR-based models to traditional GARCH models, finding that HAR models often match or surpass GARCH in forecasting accuracy, especially for volatile assets like cryptocurrencies (Andersen et al., 2003). However, forecast evaluation faces challenges. Patton (2011) highlights that imperfect volatility proxies, such as noisy RV estimates, can skew performance assessments. Similarly, Andersen et al. (2011) emphasize that market microstructure noise distorts forecast accuracy if researchers fail to address it. These limitations underscore the need for robust methods like HAR models to capture Bitcoin's unique volatility dynamics.

The literature on volatility modeling reveals a rich evolution from traditional models like ARCH and GARCH to advanced frameworks like HAR and RV measures, offering valuable tools for understanding financial markets. Researchers demonstrate that high-frequency data enhance volatility forecasts, while HAR extensions address complexities such as asymmetry and jumps (Corsi, 2009; Patton and Sheppard, 2015). In cryptocurrency markets, studies uncover unique patterns, including Bitcoins inverted leverage effect and shorter volatility persistence (Baur and Dimpfl, 2018; Zhang and Zhao, 2023). However, significant gaps remain. Limited studies apply HAR extensions to Bitcoins jumps and asymmetry, leaving the effectiveness of models like HAR-RS and HAR-J underexplored in this context (Brini and Lenz, 2024). Similarly, few researchers address how structural breaks or market microstructure noise impact cryptocurrency volatility forecasts, despite their prevalence (Aharon et al., 2023; Andersen et al., 2011). These shortcomings highlight the need for robust empirical approaches. In Chapter 4, I explore these gaps by implementing and evaluating HAR-based models tailored to Bitcoins dynamic volatility, building a foundation for deeper insights into its market behavior.

Chapter 3 Data and Preprocessing

3.1 Data Source and Timeframe

I use the high-frequency Bitcoin dataset curated by M. C. Zielinski on Kaggle[1] which reports one-minute close prices and transaction volumes from 1 January 2012 through 15 June 2025. I downloaded the file on 18 June 2025. The "Close" field records the last trade across a pool of exchanges, so I treat it as an aggregated market price. Because the file is not tick-by-tick, may omit off-exchange trades, and could suffer from survivorship bias if delisted venues disappear, I view the series as an informed approximation rather than a complete transaction record.

To focus on the mature phase of the Bitcoin market, I restrict the analysis to 1 August 2017–28 February 2025. Binance's launch in June 2017 and the post-Mt. Gox recovery mark a shift toward deeper liquidity and better price discovery, so data quality improves from August 2017 onward. The window contains 2,769 calendar days and 3,987,360 one-minute observations, which in turn generate roughly 2 769 daily RV values—ample length for fractional-differencing and HAR-style forecasting (Patton and Sheppard, 2015).

3.2 Data Quality Checks and Variable Definitions

Before modelling, I confirm data integrity. No timestamps, prices, or volumes are missing or duplicated, and each one-minute bar spans exactly 60 seconds. I verify completeness by matching a template of 1,440 minutes against the actual timestamps for every day. Because Bitcoin trades around the clock, trading volume spreads across all hours without systematic gaps.

I retain two raw fields: The Timestamp records the date and time of each 1-minute observation in UTC, derived from exchange logs as a datetime measure. And the Close Price captures Bitcoin's closing price in USD per minute.

[1] https://www.kaggle.com/datasets/mczielinski/bitcoin-historical-data

3.3 Preprocessing

To forecast RV, I resample the one-minute series into five-minute bars, following Huang and Tauchen (2005) and Andersen et al. (2007). This five-minute grid balances information capture with market-microstructure noise: shorter intervals amplify noise, while longer intervals reduce resolution. Within each bar, I record the median close price to mitigate bid-ask bounce and outlier ticks (Andersen et al., 2012). Next, I transform these prices into natural logarithms and compute log-returns as differences in log prices. Since Bitcoin trades continuously, I include weekends and align daily buckets to 00:00–23:59 UTC. Later chapters use calendar lags of one, seven, and thirty days to reflect diverse trading timeframes.

3.4 Descriptive Statistics

Table 1 summarises the empirical distribution of the five-minute median price, the corresponding log-return, and daily RV over the training sample (1 August 2017 – 24 August 2023). The table reports the minimum, mean, median, maximum, standard deviation, skewness, and kurtosis.

—Please insert Table 1 about here—

The five-minute log-return displays substantial excess kurtosis (76.545) and slight negative skewness (-0.365), confirming the leptokurtic nature of Bitcoin returns. Daily RV is highly right-skewed (12.733) with extreme kurtosis (286.322), showing long calm stretches punctuated by violent price moves—behaviour that long-memory HAR models capture well. The five-minute median price averages 20,041, above the median of 11,720, reflecting Bitcoin's upward trend. The maximum price of 68,781, underscores the asset's variability.

Figure 1-4 provide a visual analysis of the daily RV series for Bitcoin over the train sample period from August 1, 2017, to August 24, 2023. I organize the figure into four panels to highlight key characteristics of the RV series.

Figure 1 plots the RV time series, which displays a pattern often described in the literature as "needle-in-haystack": extended periods of low volatility interrupted by sharp spikes (Bariviera et al., 2017). Notable spikes occur during the early 2018 sell-off and the March 2020 COVID-19 crash.

—Please insert Figure 1 about here—

Figure 2 illustrates the empirical density of RV as a histogram, overlaid with a normal curve based on the sample mean and variance. A pronounced right-skew and excess kurtosis indicates heavy tails compared to a Gaussian distribution.

—Please insert Figure 2 about here—

Figure 3 presents a Q-Q plot against the normal distribution. The plot bends sharply upward in the upper quantiles, suggesting that extreme RV values occur more frequently than a normal distribution would predict.

—Please insert Figure 3 *about here—*

Figure 4 provides the autocorrelation (ACF) and partial autocorrelation (PACF) functions up to 50 lags. Both exhibit a slow decay, consistent with the long-memory behavior typical of volatility processes. This pattern confirms that current RV values remain predictive of volatility weeks ahead.

—Please insert Figure 4 about *here—*

Together, these figures support the use of long-memory HAR-type models, introduced in the next chapter, and highlight the limitations of relying on normality-based inference for Bitcoin's RV.

Chapter 4 Methodology

In Chapter 3, I described how I construct daily RV measures from five-minute Bitcoin data. I now lay out the statistical models that transform those measures into volatility forecasts. I begin with the model-free estimators of quadratic variation, then introduce the Heterogeneous Autoregressive (HAR) framework and its extensions, and finally set out the estimation, diagnostic, and forecast-evaluation procedures.

4.1 Realized Volatility Measures

To forecast Bitcoin's volatility using high-frequency data, I define volatility with precision and theoretical grounding. The literature employs RV and BV as model-free, nonparametric estimators to capture distinct components of price variability. These measures enable quantification of both continuous and discontinuous (jump) elements of Bitcoin's returns based on observable intraday data.

4.1.1 Realized Volatility (RV)

Following Andersen et al. (2001, 2003) and Barndorff-Nielsen and Shephard (2002), the log price $p(t)$ is defined as:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \qquad t \geq 0$$

where $dp(t)$ represents the change in logarithmic price, $\mu(t)$ denotes the drift term with continuous and locally bounded variation, $\sigma(t)$ is a strictly positive volatility process, and $dW(t)$ is a standard Brownian motion. The continuously compounded daily return is:

$$r_t = p(t) - p(t-1) = \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma(s)dW(s)$$

In this framework, the return's volatility depends on the stochastic volatility process σ(t), and the return distribution is characterized by $r_t \sim \mathcal{N}(\int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma^2(s)ds)$. To estimate the

integrated volatility nonparametrically, RV is computed using high-frequency intraday log returns $r_{t,\tau}$, where $\tau = 1, \dots, m$ represents sampling points within day t. RV is defined as:

$$RV_t = \sum_{\tau=1}^{m} r_{t,\tau}^2$$

where $r_{t,\tau} = p_{t.\tau} - p_{t,\tau-1}$ denotes the intraday return based on m intraday log-prices $\{p_{t.\tau}\}_{r=1}^{m}$ within day t, observed at fixed intervals. In the absence of jumps and with sufficient sampling frequency, RV_t converges to the Integrated Variance (IV):

$$RV_t = \sum_{\tau=1}^{m} r_{t,\tau}^2 \xrightarrow{p} \int_{t-1}^{t} \sigma^2(s)ds \qquad as\ m \to \infty$$

However, Bitcoin's price frequently exhibits discontinuities or jumps, which RV captures alongside the continuous component. To address this, the literature considers a jump-diffusion model, defined as:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t), \qquad t \geq 0$$

where $\kappa(t)$ represents the jump size, and $dq(t)$ is a jump process that equals 1 if a jump occurs at time t and 0 otherwise. The return is expressed as:

$$r_t = \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma^2(s)dW(s) + \sum_{j=N(t-1)+1}^{N(t)} \kappa(s_j)$$

where N(t) counts the number of jumps up to time t, and s_j denotes jump times. In the presence of jumps, RV_t converges to the sum of the integrated variance and jump variation:

$$RV_t \xrightarrow{p} \int_{t-1}^{t} \sigma^2(s)ds + \sum_{j=N(t-1)+1}^{N(t)} \kappa^2(s_j)$$

Thus, I use RV to measure total quadratic variation, capturing both smooth and jump components of Bitcoin's volatility.

4.1.2 Bipower Variation (BV)

To isolate the continuous sample path volatility, Barndorff-Nielsen and Shephard (2004) introduce BV, a robust estimator. BV is defined as:

$$BV_t^0 = \frac{\pi}{2}\sum_{r=2}^{m}\left|r_{t,\tau}\right|\left|r_{t,\tau-1}\right| \underset{p}{\rightarrow} \int_{t-1}^{t}\sigma_s^2 ds \qquad as\ m \longrightarrow \infty$$

This definition relies on the product of adjacent absolute returns to estimate the integrated variance consistently, even in the presence of jumps, assuming independence of W(s), σ2(s), N(t), and κ(s). To enhance stability, particularly under microstructure noise or data irregularities, Patton and Sheppard (2015) suggest averaging over skip-lag estimators for $q = 0, 1, 2, 3, 4$. The BV estimator is defined as:

$$BV_t = \frac{1}{5}\sum_{q=0}^{4} BV_t^{(q)}$$

where the skip-q BV estimator is:

$$BV_t^q = \frac{\pi}{2}\sum_{r=q+2}^{m}\left|r_{t,\tau}\right|\left|r_{t,\tau-1-q}\right|$$

I compute daily volatility with bipower variation and apply the Patton–Sheppard skip-lag average ($q = 0, 1, 2, 3, 4$) to protect the estimator against residual microstructure noise.

4.2 HAR-type Volatility Models

Volatility exhibits long memory: its autocorrelation decays slowly and remains significant over months. Short ARMA or GARCH specifications struggle to match this persistence, whereas HAR models succeed by stacking moving averages that mirror the trading horizons of different market participants (Corsi, 2009). Empirical work shows that HAR outperforms GARCH for Bitcoin in particular (Jiang and Hafner, 2024). I therefore adopt HAR as my baseline and extend it to reflect salient features of cryptocurrency returns.

4.2.1 Baseline HAR-RV Specification

I employ the HAR specification of Corsi (2009), which aggregates realised variance over daily, weekly and monthly horizons to mimic the trading frequencies of short-term, medium-term and long-term investors.

The baseline HAR-RV model is defined to regress the next-day RV on lagged daily, weekly, and monthly RVs. The model assumes an additive cascade structure, where daily, weekly, and monthly components each exhibit autoregressive behavior. The averaging operator is defined as:

$$\overline{RV}_{t,t+h} = \frac{1}{h}\sum_{i=1}^{h} RV_{t+i}$$

for h=1, 5, and 22, corresponding to daily, weekly, and monthly averages, respectively. The HAR-RV specification is expressed as:

$$\overline{RV}_{t,t+h} = \phi_0 + \phi_d RV_t + \left(\frac{\phi_w}{5}\right)\sum_{i=0}^{4} RV_{t-i} + \left(\frac{\phi_m}{22}\right)\sum_{i=0}^{21} RV_{t-i} + \varepsilon_{t+h}$$

Which translates to

$$\overline{RV}_{t,t+h} = \phi_0 + \phi_d RV_t + \phi_w \overline{RV_t^w} + \phi_m \overline{RV_t^m} + \varepsilon_{t+h}$$

Because the three lagged means exhaust all lags up to 22 days, the model is algebraically equivalent to a constrained AR(22), yet it remains highly parsimonious and interpretable. Empirical evidence in Corsi (2009) shows that shocks propagate from the daily to the weekly and

monthly components, generating the long-memory pattern so characteristic of RV series; I therefore expect the HAR structure to perform well for Bitcoin, which displays even stronger persistence and jump clustering than traditional assets.

4.2.2 Extended HAR Variants

The literature extends the baseline HAR model to capture empirical features of Bitcoin's volatility, such as jumps, asymmetries, and nonlinearity, which the original specification may not fully address. Extended HAR (EHAR) models enriches the model's structure while preserving its additive and hierarchical nature.

HAR-RS Model (Semivariances)

Patton and Sheppard (2015) introduce the HAR-RS model, which incorporates realized semivariances to capture the asymmetric impact of positive and negative return shocks. The literature notes that volatility often responds differently to upward versus downward price movements, a stylized fact in equity markets and likely relevant to cryptocurrencies like Bitcoin. realized semivariances are defined as:

$$RS_t^+ = \sum_{\tau=1}^{m} r_{t,\tau}^2 \, I\{r_{t,\tau} > 0\} \xrightarrow{p} \frac{1}{2}\int_{t-1}^{t} \sigma_s^2 \, ds + \sum_{t-1<s\leq t} \Delta p_s^2 \, I\{\Delta p_s > 0\}$$

$$RS_t^- = \sum_{\tau=1}^{m} r_{t,\tau}^2 \, I\{r_{t,\tau} < 0\} \xrightarrow{p} \frac{1}{2}\int_{t-1}^{t} \sigma_s^2 \, ds + \sum_{t-1<s\leq t} \Delta p_s^2 \, I\{\Delta p_s < 0\}$$

where $I\{\cdot\}$ is the indicator function, and $\Delta p_s = p_s - p_{s-}$ captures jumps. I include these terms as separate regressors in place of the one-day RV component in the HAR-RS model, specified as:

$$\overline{RV}_{t,t+h} = \phi_0 + \phi_d^+ RS_t^+ + \phi_d^- RS_t^- + \phi_w \overline{RV}_t^w + \phi_m \overline{RV}_t^m + \varepsilon_{t+h}$$

By construction $RS_t^+ + RS_t^- = RV_t$, so the two components fully partition total quadratic variation into "good" (price-increasing) and "bad" (price-decreasing) contributions. The literature finds that downside semivariance (RS_t^-) often exhibits greater predictive power, consistent with leverage effects and behavioral responses to losses. I adopt the HAR-RS model to account for asymmetric volatility clustering, which is particularly relevant for Bitcoin during large drawdowns.

HAR-J (Jump Variation)

The HAR-J model extends the baseline HAR by including a jump component, recognizing that realized variance comprises both continuous and discontinuous movements. Patton and Sheppard (2015) define signed jump variation as:

$$\Delta J_t^2 = \Delta J_t^{2+} + \Delta J_t^{2-}$$
$$= (RS_t^+ - RS_t^-) \cdot I\{RS_t^+ - RS_t^- > 0\} + (RS_t^+ - RS_t^-) \cdot I\{RS_t^+ - RS_t^- < 0\}$$

This decomposition distinguishes whether upward or downward jumps dominate Bitcoin's volatility on a given day. The HAR-J model is defined as:

$$\overline{RV}_{t,t+h} = \phi_0 + \phi_J^+ \Delta J_t^{2+} + \phi_J^- \Delta J_t^{2-} + \phi_c BV_t + \phi_w \overline{RV}_t^w + \phi_m \overline{RV}_t^m + \varepsilon_{t+h}$$

I estimate HAR-RS model first, then add the jump terms so that any incremental gain is attributable to the asymmetry already captured in semivariances.

4.3 Estimation and In-sample Diagnostics

I focus here on how I fit each model and check its adequacy.

4.3.1 Data Partition and Horizon Mapping

I estimate all models on the daily RV series computed in section 3.3. The RV is based on 5-minute log-returns, giving 288 observations per day. To mimic the heterogeneous trading horizons of market participants I form simple rolling means

$$RV_t^{(h)} = \frac{1}{h} \sum_{i=0}^{h-1} RV_{t-i} \qquad h \in \{1,7,30,90\}$$

corresponding to one-day, one-week, one-month and one-quarter windows in calendar time. The mean therefore uses the current day and h−1 lags. The 7-, 30- and 90-day choices replace the usual 50, 22- and 66-day "daily", "monthly" and "quarterly" windows used for equity markets, thereby respecting Bitcoin's 24/7 trading schedule.

4.3.2 OLS Estimation and Standard Errors

For every horizon h and model variant (baseline HAR, HAR-RS, HAR-J, HAR-RS-J), I estimate the regression:

$$RV_{t+h} = \phi_0 + \acute{\boldsymbol{\phi}}\boldsymbol{X_t} + +\varepsilon_{t+h}$$

where $\boldsymbol{X_t}$ collects the relevant lagged RV terms and, where required, semivariances or jump measures (see section 4.2). Estimation is by OLS on the in-sample period the first 80 % of the sample 2,215 observations), and I report conventional standard errors.

4.3.3 In-Sample Residual Diagnostics

This study evaluates the fit of each HAR-type model by testing whether the residuals capture the main features of the data-generating process. I compute five standard diagnostic tests for every model–horizon pair. I describe the diagnostics as follows:

- The Ljung–Box Q-statistic on raw residuals tests for serial correlation in the residual mean, with the null hypothesis of no autocorrelation up to lag 30.
- The Ljung–Box Q-statistic on squared residuals examines remaining volatility clustering, using the same lag-30 setup applied to e_t^2.
- Engle's ARCH LM test, detects neglected conditional heteroskedasticity in the residuals.
- White's test, regresses squared residuals on all regressors and their cross-products to identify general heteroskedasticity.
- The Shapiro–Wilk test, assesses normality of residuals.

I consider a model well-specified if it fails to reject the null hypotheses at the 5% significance level in the first four tests and shows no significant departures from normality in the Shapiro–Wilk test. If a model exhibits residual autocorrelation or ARCH effects, I retain it for out-of-sample evaluation but note its in-sample inadequacy in the discussion of results.

4.4 Forecast Design and Out-of-Sample Evaluation

I now explain how I generate and judge out-of-sample forecasts.

4.4.1 Sample Split, Estimation Window, and Forecast Horizon

I estimate all HAR-type models with a rolling-window approach. First, I fit each specification on the initial 80 % of the sample (2,215 daily observations). At every forecast origin, I re-estimate the model on a window of that same length—dropping the oldest day and adding the newest—and then predict the next observation. This rolling design guards against look-ahead bias while letting the coefficients adapt to structural changes in Bitcoin's volatility, yet it still provides the long (≈ 6-year) estimation span that HAR models need to capture long memory.

The forecast targets are defined as the *h-day-ahead* RV, $RV_{t+h} = \sum_{i=1}^{h} RV_{t+i}$, for horizons $h \in \{1,7,30,90\}$. These targets are constructed as described in Section 4.2 using a rolling sum function and shifted forward with a lead of *h* days. For horizons longer than one day, the targets overlap, and I compute loss functions for every day in the evaluation window, acknowledging the resulting serial correlation. I choose not to re-estimate parameters daily, as my primary focus is cross-model ranking rather than real-time adaptability.

4.4.2 Performance Metrics

I define the metrics as follows: Root Mean Squared Error (RMSE) measures the square root of the average squared difference between forecasted and actual values, heavily penalizing large errors. Mean Absolute Error (MAE) calculates the average absolute difference between forecasts and actual values, providing a straightforward measure of forecast error magnitude. QLIKE, recommended by Hansen and Lunde (2006), is a log-likelihood-based loss function that evaluates the relative accuracy of volatility forecasts, emphasizing their distributional fit.

$$RMSE = \sqrt{\frac{1}{T_{oos}} \sum_{t} (\hat{y}_t - y_t)^2}$$

$$MAE = \frac{1}{T_{oos}} \sum_{t} |\hat{y}_t - y_t|$$

$$QLIKE = \frac{1}{T_{OOS}} \sum_{t} \left(\frac{y_t}{\hat{y}_t} - \log \frac{y_t}{\hat{y}_t} - 1 \right)$$

where $y_t = RV_{t+h}$ is the actual RV, and $\hat{y}_t$ is the model's forecast. I export all three metrics for every model–horizon pair. Lower values always signal better performance. I retain the day-by-day loss vectors so I can run formal predictive-accuracy tests later if needed, but I omit those tests from the main text to keep the focus on model ranking. I benchmark the HAR family against a rolling-window GARCH/GJR-GARCH fitted to daily median-close returns, with results detailed in Section 5.6, which compares their forecasting performance for Bitcoin. These steps form the empirical framework used in Chapter 5 to analyze Bitcoin volatility forecasts.

Chapter 5 Empirical Results

This chapter reports the evidence derived from the forecasting framework outlined in Chapter 4. Using five-minute Bitcoin prices from 1 August 2017 to 24 August 2025, I estimate four HAR-type models—HAR, HAR-RS, HAR-J, and HAR-RS-J—and generate volatility forecasts for four calendar-day horizons (1, 7, 30, and 90 days). Each model is re-estimated in a rolling window to reflect evolving market dynamics, and its performance is evaluated using RMSE, MAE, and QLIKE. The sections that follow present in-sample diagnostics, out-of-sample forecast rankings, and robustness checks.

5.1 Preliminary Descriptive Analysis

This section examines the empirical properties of the realized volatility (RV) inputs for the HAR-type models. The analysis confirms that the RV series exhibits long-memory, heavy-tailed, and asymmetric characteristics—features that motivate the use of HAR, HAR-RS, HAR-J, and HAR-RS-J models for forecasting Bitcoin's volatility at the 1-, 7-, 30-, and 90-day horizons. Tables 2–4 summarize these properties and establish the foundation for the in-sample diagnostics and out-of-sample evaluations presented later.

To characterize the distributional properties of the RV inputs, Table 2 reports summary statistics—minimum, median, mean, maximum, standard deviation, skewness, and kurtosis—for daily (RV_d), weekly (RV_w), monthly (RV_m), quarterly (RV_q), six-month (RV_{6m}), and annual (RV_y) realized volatilities, as well as for positive and negative realized semivariances (RS^+ and RS^-), positive and negative jump components (J^+ and J^-), and bipower variation (BV). Slight variations in observation counts reflect undefined longer-horizon rolling means (e.g., RV_{6m}, RV_y) at the start of the sample.

—Please insert Table 2 about here—

The RV inputs display consistent mean values across horizons (1.37×10^{-3} to 1.45×10^{-3}), reflecting the volatility-signature effect: aggregation reduces sampling noise while preserving scale, thus providing reliable inputs for HAR models. The mean BV (1.21×10^{-3}) accounts for approximately 83% of daily RV, indicating that continuous volatility dominates but jumps remain

a non-trivial contributor (17%). All series are markedly right-skewed and heavy-tailed, with daily RV showing extreme skewness (12.725) and kurtosis (283.063), underscoring Bitcoin's frequent volatility spikes. Even longer-horizon RVs (RV_{6m} and RV_y) retain elevated kurtosis (2.686 and 0.727), confirming persistent heavy-tailed behavior. The negative semivariance (RS^-, mean 7.4×10^{-4}) slightly exceeds the positive (RS^+, mean 7.1×10^{-4}), both with comparable extreme kurtosis (237.510 vs. 290.365), indicating intense negative price moves. Jump components are sparse (medians near zero) yet extreme, with high skewness (8.487 for J^+, -9.531 for J^-) and triple-digit kurtosis (107.783 for J^+, 124.852 for J^-). These features support modeling jumps separately in HAR-J and HAR-RS-J specifications.

The heavy tails and volatility clustering across all variables justify HAR-type regressions, which capture long-memory behavior without assuming normality. The asymmetry in semivariances and signed jumps further supports including RS^+, RS^-, J^+, and J^- to distinguish between "good" and "bad" volatility. Moreover, the sizeable jump contribution (17% of RV) suggests that relying solely on continuous volatility proxies would underestimate true risk in Bitcoin markets.

To verify the fractional persistence assumed by HAR-type models, I estimate the memory parameter *d* using the local-Whittle semiparametric estimator. The point estimate is $\hat{d} = 0.32$ (95% CI: [0.26, 0.38]). Because $0 < \hat{d} < 0.5$, RV_t is covariance-stationary yet long-memory, satisfying conditions for consistent OLS estimation.

Stationarity and clustering properties are further assessed through standard unit-root and diagnostic tests, applied to the training set to prevent data leakage. Table 3 reports results from the Augmented Dickey–Fuller (ADF), Phillips–Perron (PP), and KPSS tests for stationarity; the Ljung–Box test for autocorrelation; the ARCH LM test for conditional heteroskedasticity; and the Jarque–Bera test for normality.

—Please insert Table 3 about here—

The results confirm key distributional properties. The ADF statistic (–9.23, $p = 0.010$) and PP statistic (–2261, $p = 0.010$) reject the null of a unit root at the 1% level, while the KPSS statistic (1.13, $p = 0.010$) rejects trend non-stationarity, indicating that the series is level-stationary. The Ljung–Box Q(30) statistic (2397, $p < 0.001$) shows significant autocorrelation, and the ARCH

LM(12) statistic (22.74, $p = 0.030$) detects conditional heteroskedasticity, consistent with volatility clustering. The Jarque–Bera statistic (7,468,208, $p < 0.001$) strongly rejects normality, reflecting the heavy-tailed distribution expected in high-frequency volatility measures. Together, these findings support HAR-type models as suitable for capturing Bitcoin's long-memory and clustering patterns without relying on normality.

To explore relationships among RV inputs, Table 4 presents their correlation matrix, derived from the training set. Daily RV is highly correlated with BV (0.990), RS^+ (0.974), and RS^- (0.976), confirming that continuous and asymmetric components dominate volatility. RS^+ and RS^- are strongly correlated (0.902) but display distinct correlations with jumps (J^+: 0.527, J^-: -0.577 with RS^-), justifying separate modeling in HAR-RS-J specifications. Jumps show moderate correlations with RV_d (J^+: 0.381, J^-: -0.417) and weak correlation with each other (0.092). Correlations with longer-horizon RV_s decline from RV_w (0.609) to RV_y (0.075), reflecting the gradual fading of volatility persistence.

—Please insert Table 4 about here—

These findings address the thesis's research questions. The long-memory property ($\hat{d} = 0.32$) supports HAR models for capturing persistent volatility patterns (Question 1). The asymmetry in semivariances and sparse, extreme jumps suggest potential predictive value for these components (Question 2). The stationarity, clustering, and correlation structure provide a robust foundation for testing model performance across horizons and market regimes (Question 3).

5.2 In-Sample Estimation Outcomes

This section evaluates the in-sample performance of the HAR-type models using the RV series derived from 5-minute Bitcoin prices (August 1, 2017–August 24, 2023). Building on the long-memory, heavy-tailed, and asymmetric properties identified in Section 5.1, the analysis examines how lagged RV, semivariances, and jumps explain volatility across 1-, 7-, 30-, and 90-day horizons. The next section reports coefficient estimates, standard errors, and adjusted R^2, revealing four stylized facts: persistence shifts to longer lags, downside risk dominates, bipower variation is robust, and combined semivariance-jump models enhance fit despite collinearity. These findings set the stage for out-of-sample evaluation (Section 5.3).

5.2.1 HAR model coefficient estimates

Table 5 reveals clear patterns that align with the HAR model's logic, demonstrating how past volatility influences future volatility across multiple forecast horizons. The discussion progresses logically from shorter to longer horizons, examining the baseline model and its extensions with semivariances and jump components.

—Please insert Table 5 about here—

Short-term forecasts (1-day horizon) are dominated by high-frequency persistence. In the baseline HAR model, the daily (ϕ_d = 0.3431***), weekly (ϕ_w = 0.2186***) and monthly (ϕ_m = 0.2319***) lags sum to approximately 0.79, consistent with near-unit-root persistence and the long-memory property ($\hat{d} = 0.32$, Section 5.1). Adding realized semivariances (HAR-RS) improves explanatory power from 26.5% to 29.0%, with a strong asymmetry: downside semivariance (ϕ_{d^-} = 1.0098***) is positive and highly significant, while upside semivariance (ϕ_{d^+} = -0.3769***) is negative, underscoring Bitcoin's heightened volatility sensitivity to negative returns (Question 2). Incorporating jumps (HAR-J) raises the fit to 31.8%, with bipower variation (ϕ_{BV} = 0.3155***) and large negative jumps (ϕ_{J^-} = -1.2477***) indicating post-shock volatility mean reversion. The most comprehensive model (HAR-RS-J) achieves 35.8%, but collinearity between downside semivariance and jumps reverses the sign of ϕ_{d^-}, suggesting overlapping information content in extreme downside risk measures.

Medium-short-term forecasts (7-day horizon) shift persistence toward longer lags. The HAR model shows strong weekly (ϕ_w = 2.0604***) and monthly (ϕ_m = 2.0697***) effects, with the quarterly term (ϕ_q = 0.7037*) also contributing, explaining 26.3% of variance. HAR-RS increases the fit to 30.4%, with downside semivariance (ϕ_{d^-} = 2.8565***) again dominating and upside effects (ϕ_{d^+} = –0.8831*) remaining negative. HAR-J raises R^2 to 31.0%, with bipower variation (ϕ_{BV} = 1.1843***) and both positive (ϕ_{J^+} = –1.5385*) and negative jumps (ϕ_{J^-} = –2.1719***) indicating post-shock volatility declines. The HAR-RS-J model reaches 31.8%, with bipower variation (ϕ_{BV} = 4.6664***) emerging as the most important driver, and downside semivariance turning insignificant due to overlap with jump measures.

Medium-term forecasts (30-day horizon) display even stronger weighting on low-frequency volatility. The monthly term (ϕ_m = 8.3852***) dominates across specifications, while quarterly volatility (ϕ_q = 3.3562*) also contributes positively. The six-month lag (ϕ_{6m} = –3.8380**) enters negatively, indicating mean reversion. The baseline HAR explains only 12.6% of variance, increasing to 14.8% (HAR-RS) and 15.3% (HAR-J), with bipower variation (ϕ_{BV} = 2.3597***) becoming highly significant. In the HAR-RS-J model (16.3%), bipower variation jumps to 13.2775***, while downside semivariance reverses sign (ϕ_{d^-} = –10.8784***), again reflecting collinearity with negative jumps (ϕ_{J^-} = –6.3695*).

Long-term forecasts (90-day horizon) are the most challenging, with explanatory power dropping sharply. The HAR model explains just 2.0% of variance, driven by a large positive quarterly effect (ϕ_q = 17.7270***) and an equally large negative six-month effect (ϕ_{6m} = –24.4654***), suggesting strong mean-reverting cycles, possibly linked to Bitcoin's bull–bear market regimes. Adding semivariances raises R^2 to 3.5%, but asymmetry effects remain insignificant. HAR-J increases the fit to 4.0%, with bipower variation (ϕ_{BV} = 4.4497***) as the only robust predictor. The HAR-RS-J model achieves 5.9%, with bipower variation (ϕ_{BV} = 33.8484***) overwhelmingly dominating, while downside semivariance (ϕ_{d^-} = –33.4019***) and upside semivariance (ϕ_{d^+} = –19.5136***) turn sharply negative due to strong collinearity with negative jumps (ϕ_{J^-} = –15.2948*).

Overall, the results show that:

1. Short horizons are driven by high-frequency persistence, asymmetric semivariance effects, and jump-induced mean reversion.

2. Medium horizons increasingly depend on continuous variation (bipower variation), with jumps and semivariances providing complementary signals.

3. Long horizons rely almost entirely on low-frequency continuous components, with asymmetry and jump effects diminishing in importance and often becoming collinear.

This horizon-dependent shift in volatility drivers is consistent with the literature on heterogeneous market components, where short-term dynamics are dominated by market microstructure noise

and high-frequency shocks, while long-term volatility is governed by macroeconomic cycles and slow-moving market forces.

5.2.2 Residual Diagnostics

To verify that the four HAR-type regressions are well-specified for RV forecasting, I examine residuals for serial correlation, volatility clustering, heteroskedasticity, and distributional shape across horizons $h \in \{1,7,30,90\}$. For each model–horizon pair I run five tests:

1. Ljung–Box on residuals to detect linear serial correlation, using LLL lags and adjusting the degrees of freedom for estimated regressors.

2. Ljung–Box on squared residuals to diagnose volatility clustering.

3. Engle's ARCH LM test to check for remaining ARCH effects.

4. White/Breusch–Pagan test with White-type quadratic expansion: $\sim.^2$ in the regressors to assess heteroskedasticity relative to the design matrix.

5. Normality: Shapiro–Wilk when $3 \leq n \leq 5000$; Jarque–Bera when $n > 5000$.

Lag selection for the Ljung–Box tests follows Patton & Sheppard (2015) with a small-h adjustment: I set $L = 30$ for $L \in \{1, 7\}$ and $L = h + 20$ otherwise, to mitigate spurious rejections from overlapping forecast targets. Table 6 reports the diagnostic test *p*-values for each HAR-type model and forecast horizon, allowing assessment of residual autocorrelation, volatility clustering, heteroskedasticity, and normality.

—Please insert Table 6 about here—

Short horizon (1-day) results indicate that all models remove most linear serial correlation in residuals, with Ljung–Box *p*-values above conventional thresholds for HAR (0.063) and HAR-RS (0.057), and marginal rejection for HAR-J (0.033). However, the HAR-RS-J model shows significant autocorrelation ($p < 0.001$), suggesting potential overfitting or omitted short-term dynamics. The Ljung–Box test on squared residuals rejects the null strongly for all but HAR-RS-J (p = 0.157), indicating that this specification captures volatility clustering better than its

competitors. ARCH LM tests uniformly reject ($p < 0.001$), pointing to remaining ARCH effects even at the daily horizon. The White test strongly rejects homoskedasticity for all models ($p \approx 0$), and Shapiro–Wilk rejects normality at extreme significance levels, consistent with heavy-tailed residuals common in volatility models.

Medium-short horizon (7-day) diagnostics are more severe: all p-values for both Ljung–Box variants, ARCH, White, and Shapiro–Wilk are effectively zero, implying persistent serial correlation, volatility clustering, and non-normality in residuals across all specifications. This is consistent with the overlapping-window effect at multi-day horizons, which makes residual dependence difficult to eliminate.

Medium horizon (30-day) models also show strong rejections in all tests except the White test for HAR (p = 0.0001) and HAR-RS-J (p = 0.013), where heteroskedasticity rejections are marginally weaker. Nonetheless, significant results in Ljung–Box, squared Ljung–Box, and ARCH tests confirm that unmodeled autocorrelation and volatility clustering remain present.

Long horizon (90-day) diagnostics are the most stringent. All specifications fail every test at extreme significance levels, with particularly strong ARCH and White rejections ($p \approx 0$) and Shapiro–Wilk p-values far below any conventional threshold. These patterns reflect the difficulty of capturing slow-moving, long-horizon volatility dynamics within the HAR framework, especially when dependent variables are highly overlapping.

Overall, the diagnostics confirm that while HAR-type models capture key persistence and asymmetry effects in realized volatility, residuals still exhibit features such as autocorrelation, volatility clustering, heteroskedasticity, and heavy tails—especially at longer horizons. These findings are in line with the volatility forecasting literature, where even well-specified models rarely produce residuals that are fully white noise, and where robust standard errors or bootstrap methods are often employed to ensure valid inference.

5.3 Out-of-sample Forecast Evaluation

This section evaluates the predictive accuracy of four HAR-type models—HAR, HAR-RS, HAR-J, and HAR-RS-J—across forecast horizons of 1, 7, 30, and 90 calendar days. Forecasts are generated using the rolling-window procedure described in Section 4.4, covering 554 daily observations from August 25, 2023, to February 28, 2025. Performance is measured using three standard loss functions: RMSE, MAE, and QLIKE. Table *7* reports the average loss for each model and horizon, with the lowest loss in each row highlighted in bold. Given the broad ranking of models, the analysis focuses on economically meaningful patterns rather than formal pairwise statistical tests.

—Please insert Table 7 about here—

At the one-day horizon ($h = 1$), the HAR-J model achieves the lowest RMSE (0.00062), MAE (0.00034), and QLIKE (0.32979), indicating that incorporating bipower variation (BV) and signed jumps enhances short-term forecast accuracy. The HAR-RS model records slightly higher RMSE (0.00063) and MAE (0.00037), suggesting that downside and upside semivariances add limited predictive power for day-ahead forecasts compared to BV.

At the seven-day horizon ($h = 7$), the HAR-RS and HAR-J models deliver the strongest performance. HAR-RS has the lowest RMSE (0.00275) and matches HAR-J on MAE (both 0.00240), while HAR-J achieves the lowest QLIKE (0.18405 vs. 0.18409). The differences are economically negligible, indicating that both downside semivariance and BV contribute similarly to week-ahead forecasts. The HAR-RS-J model performs worst, with particularly poor QLIKE (369.804), likely due to noise from combining jumps and semivariances.

At the thirty-day horizon ($h = 30$), the HAR-RS model again leads across all metrics (RMSE: 0.01262, MAE: 0.01129, QLIKE: 0.18384), followed closely by HAR-J. The HAR-RS-J model ranks last (e.g., RMSE: 0.01298), suggesting that jump terms add estimation noise at this horizon. Persistent downside semivariance appears to be a key predictor for monthly volatility, while jumps beyond BV offer limited incremental value.

At the ninety-day horizon ($h = 90$), all models exhibit higher losses, with minimal differences. HAR-RS achieves the lowest RMSE (0.04162), MAE (0.03802), and QLIKE (0.16052), but its

advantage over HAR and HAR-J is small. High-frequency components lose relevance at this horizon, with forecasts relying primarily on slow-moving quarterly, six-month, and annual volatility aggregates common to all specifications.

Several patterns emerge. First, the predictive value of model components varies by horizon: BV and jumps enhance one-day forecasts, while downside semivariance drives accuracy for seven- and thirty-day horizons. At ninety days, no extension offers a decisive advantage. Second, forecast accuracy declines with horizon length, as losses increase sharply, reflecting the limitations of linear HAR models for long-range predictions. Third, simpler models (HAR-RS, HAR-J) often outperform the more complex HAR-RS-J, suggesting that combining all extensions can introduce multicollinearity and noise.

Why do the single-block extensions outperform HAR-RS-J?

Although downside semivariance and signed-jump variation each improve forecasts on their own, they convey highly overlapping information. A large negative jump is already embedded in RS^-; adding a separate J^- term duplicates much of that signal. Empirically, Table 4 shows a –0.58 correlation between RS^- and J^-, consistent with Patton & Sheppard (2015), who find only modest incremental gain when both blocks are included. Jump variation is also noisy—one bad tick can inflate JV_t. When both RS and JV enter the same regression, signal gain is small, estimation variance rises sharply, and multicollinearity can flip coefficient signs (see §5.2.1). This bias–variance trade-off (Bollerslev et al., 2016) explains why HAR-RS-J lags behind leaner specifications.

At the day-ahead horizon, price jumps dominate volatility spikes, making HAR-J the best performer. At week-to-month horizons, persistent downside risk outweighs isolated jumps, favoring HAR-RS. Mixing both blocks at a single horizon dilutes their niche advantages.

5.4 Robustness Checks

To verify that the forecast performance results in Table 7 are not driven by specific data choices, I conduct two robustness exercises.

Check A: Alternative Sampling Frequency

I recompute all RV measures using 10-minute returns instead of 5-minute returns, retaining the 2,215-day rolling window.

Check B: Post-COVID Subsample

I maintain the 5-minute grid but limit the estimation and evaluation periods to March 15, 2020–February 28, 2025, excluding the ICO boom, the 2018 "crypto winter," and pre-COVID market dynamics.

Forecasts are generated using the rolling-window procedure described in Section 4.4. Table 8 reports the QLIKE loss for each model (HAR, HAR-RS, HAR-J, HAR-RS-J) across horizons ($h = 1,7,30,90$) under both settings, alongside the baseline (5-minute, 2,215-day) for comparison. Bold figures indicate the lowest QLIKE per horizon.

—Please insert Table 8 about here—

5.4.1 Effect of Coarser Intraday Sampling

Using a 10-minute grid, the model performance hierarchy remains broadly consistent with the baseline. At $h = 1$, HAR-J continues to lead with the lowest QLIKE (0.35666), although this is around +8% higher than its baseline value (0.32979). At $h = 7$ and $h = 30$, HAR-RS remains the top performer (0.19117 and 0.19677, respectively), as in the baseline. At $h = 90$, HAR-RS also leads (0.17864). Across horizons, QLIKE increases by roughly 4–12% (e.g., HAR-RS at $h = 30$: +7.1%; HAR at $h = 7$: +4.0%), indicating modest sensitivity to microstructure noise but confirming the robustness of BV and semivariance predictors.

5.4.2 Effect of Post-COVID Subsample

In the post-COVID subsample, some shifts in ranking occur. At $h = 1$, HAR achieves the lowest QLIKE (0.27097), with HAR-RS (0.27156), HAR-J (0.27548), and HAR-RS-J (0.27607) close behind. At $h = 7$ and $h = 30$, HAR-RS again dominates (0.12882 and 0.07793, respectively), consistent with the baseline leadership at these horizons. At $h = 90$, HAR leads narrowly (0.02228) over HAR-RS (0.02351). QLIKE values in this subsample are substantially lower than in the baseline—by around 57–86% at $h = 30$ and $h = 90$ (e.g., HAR-RS at $h = 90$: –85.3%)—reflecting the lower volatility of the post-2020 period. Despite this scale change, the relative performance ordering is largely preserved, with HAR-RS and HAR-J remaining competitive at short and medium horizons.

5.4.3 Overall Assessment

The robustness checks highlight three main points:

1. **Model ranking stability** – HAR-J remains best for day-ahead forecasts ($h = 1$) in both the baseline and the 10-minute grid, while HAR-RS dominates at $h = 7$ and $h = 30$. At $h = 90$, HAR and HAR-RS alternate in the lead, with negligible differences.

2. **Magnitude of change** – Switching to 10-minute sampling increases QLIKE modestly (up to 12%), whereas restricting to the post-COVID period reduces QLIKE sharply (up to –86%) due to lower volatility, without altering the broad ranking.

3. **Practical implications** – For short-term risk management and day-ahead VaR, HAR-J is preferable because jump adjustments sharpen forecasts when shocks are discrete and short-lived. For week- and month-ahead budgeting or volatility planning, HAR-RS is more effective, as downside semivariance captures the slow decay of bearish sentiment. The combined HAR-RS-J should be avoided unless estimation variance can be reduced (e.g., via shrinkage or larger samples) to mitigate multicollinearity.

Together with the in-sample residual diagnostics (Table 6), these checks confirm that the predictive performance of HAR-type models is robust to both changes in intraday sampling frequency and

shifts in market regime, underscoring the enduring value of BV, jumps, and downside semivariance in Bitcoin volatility forecasting.

5.5 Comparison with GARCH Benchmarks

To benchmark the HAR-type models against a standard parametric volatility framework, I estimate a family of GARCH-type models on daily log returns aggregated from the 5-minute median-price series. Candidate specifications include ARCH(1), GARCH(1,1), and GJR-GARCH(1,1) under five innovation distributions (Gaussian, Student-t, skewed t, GED, and skewed GED). The best-performing specification is selected via the Bayesian Information Criterion (BIC) in the in-sample period (80% of the sample).

The selected model is a GARCH(1,1) with generalized error distribution (GED), which is then re-estimated in a rolling fashion to produce out-of-sample forecasts. One-step-ahead variance forecasts are aggregated to 7-, 30-, and 90-day horizons for comparison with the realized variance benchmarks.

Table 9 reports the out-of-sample forecast performance of the selected GARCH model, evaluated via MAE, RMSE, and QLIKE.

—Please insert Table 9 about here—

At the 1-day horizon, GARCH produces a QLIKE of 0.416, substantially higher than the HAR-J benchmark (0.330 in Table 7), indicating that realized-measure–based models extract more useful high-frequency information for short-term volatility.

At the 7-day horizon, GARCH's QLIKE (0.167) approaches HAR-RS (0.184) and HAR-J (0.184), narrowing the gap as horizon length increases and high-frequency predictors lose relevance. For 30- and 90-day horizons, GARCH attains the lowest QLIKE values of all models considered (0.100 and 0.071, respectively), reflecting the strength of parametric persistence in capturing slow-moving long-run variance components.

These findings echo prior evidence (e.g., Andersen et al., 2007; Patton & Sheppard, 2015) that realized-measure–based HAR models dominate in the short run, while parsimonious GARCH models can outperform for medium- to long-term variance forecasts. In practice, a hybrid

approach—using HAR-J for day-ahead risk management and GARCH for quarterly risk budgeting—would leverage the comparative advantages of both model classes.

Chapter 6 Conclusion

6.1 Key Findings at a Glance

This thesis set out to examine (1) whether high-frequency RV measures and HAR-type models can capture Bitcoin's distinctive volatility patterns, (2) which HAR extensions—semivariances or jumps—deliver the greatest predictive gains across horizons, and (3) how robust these forecasts remain under alternative sampling and market regimes.

First, high-frequency RV paired with HAR-type specifications captures Bitcoin's long-memory volatility with high accuracy. Bitcoin's RV is stationary yet exhibits strong autocorrelation, heavy tails, and persistence—features well-documented in financial volatility research. HAR models exploit this persistence efficiently: in-sample fits explain up to 36% of daily RV variation, and out-of-sample performance consistently exceeds that of traditional GARCH benchmarks.

Second, downside risk and continuous price variation emerge as the dominant predictors. Downside semivariance (RS^-) consistently outperforms upside semivariance (RS^+), reinforcing Bitcoin's asymmetric risk–return profile and echoing the inverted leverage effect noted in equities. Bipower variation (BV) dominates at short horizons, while jumps add incremental value only in the immediate term. This yields a clear ranking: HAR-J is optimal for day-ahead forecasts, HAR-RS for 7–30 day horizons.

Third, forecast accuracy varies systematically with horizon length. HAR-J's advantage at $h = 1$ reflects the role of continuous variation and jump adjustments in short-term risk, while HAR-RS's dominance for $h = 7$ and $h = 30$ underscores the persistence of downside risk. At $h = 90$, model performances converge, suggesting that long-run volatility is governed by slow-moving aggregates common across specifications.

Fourth, robustness checks confirm these patterns. Moving from 5-minute to 10-minute sampling leaves model rankings intact, showing resilience to microstructure noise. In the post-COVID subsample, forecast errors fall sharply—reflecting lower volatility—but the relative performance of HAR-J and HAR-RS is unchanged.

Finally, Bitcoin's volatility structure differs materially from equities: persistence is shorter, asymmetry is stronger, and jumps are more influential. Incorporating realised semivariance and

jumps into HAR models aligns volatility modelling with these cryptocurrency-specific features, extending the applicability of RV and HAR frameworks beyond traditional markets.

Overall, this study demonstrates that HAR-type models, enriched with high-frequency RV, downside semivariance, and bipower variation, provide a robust, horizon-specific framework for forecasting Bitcoin volatility—delivering both academic insight and practical value for risk management, trading, and market design.

6.2 Academic Contributions

This thesis makes three primary contributions to the volatility modelling literature.

First, it extends the application of RV and HAR models to Bitcoin—a high-volatility, continuously traded asset whose behaviour departs from equity market assumptions. By documenting long-memory behaviour, heavy tails, and pronounced asymmetry in Bitcoin's RV, the analysis shows that the HAR framework of Corsi (2009), when paired with high-frequency data, effectively captures cryptocurrency volatility dynamics. This broadens the empirical scope of HAR models to non-traditional, digital asset markets.

Second, it provides new empirical evidence on the predictive roles of semivariance and jump components in cryptocurrency volatility. While earlier work (e.g., Patton & Sheppard, 2015) established the relevance of downside semivariance in equities, this study finds an even stronger effect in Bitcoin, consistent with the inverted leverage effect discussed in prior literature. It also quantifies the role of jumps—beneficial for day-ahead forecasts but quickly losing predictive power at longer horizons—while showing that continuous volatility (bipower variation) remains the dominant driver across most horizons. These results refine our understanding of the volatility decomposition in crypto markets.

Third, it contributes to methodological discussions on model robustness. By demonstrating that forecast rankings remain stable under alternative intraday sampling frequencies (5-minute vs. 10-minute returns) and across distinct market regimes (full sample vs. post-COVID), the study establishes HAR-type models as resilient forecasting tools for volatile and rapidly evolving markets. This systematic robustness assessment strengthens the bridge between academic model development and real-world market conditions, enhancing the practical relevance of RV-based approaches.

Chapter 7 Discussion

7.1 Practical Implications

This thesis provides actionable guidance for risk managers, traders, and analysts in cryptocurrency markets.

For risk management, the HAR-J model's superior day-ahead performance strengthens Value-at-Risk (VaR) back-testing and short-term capital allocation. Its combination of bipower variation and jump filters enables rapid adjustment to volatility spikes, producing more responsive risk assessments during turbulent periods.

For portfolio and hedging strategies, the HAR-RS model's consistent outperformance at weekly and monthly horizons makes it a reliable tool for managing medium-term volatility exposure. This supports practical use cases in option pricing, futures hedging, and dynamic asset allocation, while aligning with Bitcoin's distinctive volatility structure.

For operational efficiency, the robustness analysis shows that forecasts from 10-minute grids are nearly as accurate as those from 5-minute grids. This finding allows practitioners to ease data storage and processing demands without materially reducing forecast precision.

The results also inform algorithmic trading design. By quantifying downside risk premiums and the short-lived nature of jumps, the thesis offers parameters that can be embedded in execution strategies—prioritising "bad" volatility (downside moves) over "good" volatility (upside moves) to suit Bitcoin's continuous and asymmetric trading environment.

Finally, the model hierarchy carries a clear practical message: parsimony matters. Combining both downside and jump blocks often yields little additional signal but increases estimation variance. Matching model complexity to horizon-specific drivers—continuous variation and jumps for 24-hour risk, downside persistence for multi-day horizons—produces cleaner, more reliable forecasts than all-inclusive specifications.

In short, this work translates realised volatility methods from traditional finance to the cryptocurrency context, producing forecasting tools calibrated to Bitcoin's unique dynamics and delivering both theoretical and operational value.

7.2 Robustness and Market Regimes

The robustness checks in Section 5.4 confirm that HAR-type models for Bitcoin volatility remain effective under changes in intraday sampling and across different market regimes.

With a 10-minute sampling grid, the ranking of models remains stable: HAR-J leads for one-day forecasts, HAR-RS dominates at weekly and monthly horizons. While QLIKE losses rise modestly (4–12%) with coarser sampling—indicating a small precision loss—the main predictive relationships, driven by bipower variation and downside semivariance, remain intact. This resilience to microstructure noise allows practitioners to work with less granular data without eroding model performance.

In the post-COVID subsample (March 2020–February 2025), volatility is markedly lower, reflecting greater institutional participation and the absence of earlier boom–bust cycles. Forecast errors fall sharply—QLIKE reductions of 57–86% at medium and long horizons—but model rankings are unchanged. HAR-J retains its short-term lead, while HAR-RS continues to dominate at weekly and monthly horizons. This suggests that Bitcoin's key volatility drivers—long memory, downside asymmetry, and clustered jumps—are structural and persist across market regimes.

These results highlight the adaptability of the HAR framework to both micro-level (data granularity) and macro-level (market environment) changes. The stability of rankings across these conditions reinforces the role of bipower variation and downside semivariance as robust predictors, even in calmer markets. Practitioners can therefore apply HAR-J for short-term and HAR-RS for medium-term horizons with confidence that their relative performance will hold under varying data conditions and evolving market structures.

7.3 Limitations

Despite the robustness of the empirical findings, this thesis is subject to constraints related to data coverage, model scope, and market representativeness, each of which shapes the interpretation of results.

Data limitations: The analysis relies on aggregated minute-level Bitcoin price data from multiple exchanges via a Kaggle dataset. While sufficient to capture broad market dynamics, it omits tick-level granularity and off-exchange (OTC) trades that may influence price discovery. Survivorship bias is possible if failed or delisted exchanges are underrepresented, potentially overstating historical liquidity or understating early-period volatility. The absence of order book depth and trade direction data also limits the ability to link microstructural factors directly to realised volatility patterns.

Model limitations: The HAR framework, although effective in modelling long-memory volatility, is linear and does not explicitly address structural breaks or nonlinear dynamics common in cryptocurrency markets. Potential volatility spillovers from other cryptocurrencies or macro-financial variables (e.g., equity indices, interest rates) are excluded. While downside semivariance and jump components improve predictive power, their effects are still estimated within a linear regression, potentially missing nonlinear interactions or regime-specific behaviour.

Scope limitations: The exclusive focus on Bitcoin ensures clean inference but constrains generalisability. Results may not transfer directly to altcoins or DeFi tokens, which often have thinner liquidity, more frequent jumps, and distinct volatility drivers.

These limitations caution against unqualified generalisation beyond Bitcoin or the HAR setting. Future research could address them by exploring alternative realised volatility estimators—such as realised kernels or pre-averaging methods (Barndorff-Nielsen et al., 2008; Jacod et al., 2009)—to further mitigate microstructure noise, though the current approach aligns with established empirical standards for high-frequency volatility analysis.

7.4 Future Research Directions

Building on this work, several avenues could advance volatility forecasting in cryptocurrency markets, strengthening both academic understanding and practical application:

1. **Incorporate regime-switching and nonlinear dynamics.** Regime-switching HAR (RS-HAR) or ARFIMA-HAR hybrids could capture bull–bear transitions, liquidity shocks, and abrupt volatility shifts—improving the modelling of Bitcoin's inverted leverage effect during crises.
2. **Develop multivariate and cross-asset frameworks.** Extending the analysis to include volatility spillovers between cryptocurrencies (e.g., Bitcoin–Ethereum) or across crypto–traditional asset classes (e.g., equities, FX) would clarify how shocks propagate. Multivariate HAR or realised covariance models could enhance risk management for diversified portfolios.
3. **Integrate microstructural information.** Using tick-level trades, order book depth, or trade imbalance measures could refine realised volatility estimates and reveal liquidity–volatility linkages. Enhanced jump detection could distinguish information-driven events from liquidity shocks.
4. **Augment with machine learning methods.** Combining HAR with nonlinear tools—such as LSTM networks or random forests—may capture threshold effects and complex interactions missed by linear models. A hybrid approach, applying machine learning to HAR residuals, could improve accuracy while preserving interpretability.
5. **Link forecasts to economic value.** Connecting statistical performance to operational metrics—such as VaR back-tests, option pricing accuracy, or volatility-targeting portfolio returns—would assess the tangible benefits of HAR-based forecasts in live market settings.

By extending realised volatility modelling along these lines, future research can better integrate high-frequency econometrics with the evolving structure of cryptocurrency markets, enhancing both the explanatory reach and the real-world utility of volatility forecasts.

References

Aharon, D. Y., Butt, H. A., Jaffri, A., & Nichols, B. (2023). Asymmetric volatility in the cryptocurrency market: New evidence from models with structural breaks. *International Review of Financial Analysis, 87*, 102651.

Andersen, T. G., & Benzoni, L. (2008). Realized volatility. In T. G. Andersen, R. A. Davis, J.-P. Kreiss, & T. Mikosch (Eds.), Handbook of financial time series (pp. 555–575). Springer.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review, 39*(4), 885–905.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. The Review of Economics and Statistics, 89(4), 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. Journal of Financial Economics, 61(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. Econometrica, 71(2), 579–625.

Andersen, T. G., Dobrev, D., & Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. Journal of Econometrics, 169(1), 75–93.

Baillie, R. T., Calonaci, F., Cho, D., & Rho, S. (2019). Long memory, realized volatility and heterogeneous autoregressive models. *Journal of Time Series Analysis, 40*(4), 609–628.

Bariviera, A. F., BasGall, L., Hasperué, W., & Naiouf, M. (2017). Some stylized facts of the Bitcoin market. Physica A: Statistical Mechanics and its Applications, 484, 82–90.

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics, 2*(1), 1–37.

Baur, D. G., & Dimpfl, T. (2018). Asymmetric volatility in cryptocurrencies. *Economics Letters, 173*, 148–151.

Ben Cheikh, N., Ben Zaied, Y., & Chevallier, J. (2020). Asymmetric volatility in cryptocurrency markets: New evidence from smooth transition GARCH models. *Finance Research Letters, 35*, 101293.

Black, F. (1976). Studies of stock price volatility changes. Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section, 177–181.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. The Review of Economics and Statistics, 69(3), 542–547.

Bollerslev, T., Patton, A. J., & Quaedvliege, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics, 192*(1), 1–18.

Brini, A., & Lenz, J. (2024). A comparison of cryptocurrency volatility—Benchmarking new and mature asset classes. *arXiv*.

Chen, X., Ma, J., & Jin, Y. (2019). Modeling realized volatility with realized quarticity: A perspective from realized jump components. Journal of Forecasting, 38(7), 683–701.

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

Corsi, F., & Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics, 30*(3), 368–380.

Corsi, F., Pirino, D., & Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. Journal of Econometrics, 159(2), 276–288.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987–1007.

Hansen, P. R. (2005). A test for superior predictive ability. Journal of Business & Economic Statistics, 23(4), 365–380.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics, 20*(7), 873–889.

Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. Journal of Business & Economic Statistics, 24(2), 127–161.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*(2), 453–497.

Huang, X., & Tauchen, G. (2005). The relative contribution of jumps to total price variation. Journal of Financial Econometrics, 3(4), 456–499.

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters, 158*, 3–6.

Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics, 187*(1), 293–311.

Maki, D., & Ota, Y. (2021). Impacts of asymmetry on forecasting realized volatility in Japanese stock markets. *Economic Modelling, 101*, 105533.

McAleer, M., & Medeiros, M. C. (2008). Realized volatility: A review. *Econometric Reviews, 27*(1–3), 10–45.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics, 160*(1), 246–256.

Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *The Review of Economics and Statistics, 97*(3), 683–697.

Souček, M., & Todorova, N. (2014). Realized volatility transmission: The role of jumps and leverage effects. Journal of International Financial Markets, Institutions and Money, 29, 121–139.

Soylu, P. K., Okur, M., Çatıkkaş, Ö., & Altıntaş, Z. A. (2020). Long memory in the volatility of selected cryptocurrencies: Bitcoin, Ethereum and Ripple. *Journal of Risk and Financial Management, 13*(6), 107.

White, H. (2000). A reality check for data snooping. Econometrica, 68(5), 1097–1126.

Zhang, L., Mykland, P. A., & Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association, 100(472), 1394–1411.

Zhang, Z., & Zhao, R. (2023). Good volatility, bad volatility, and the cross section of cryptocurrency returns. *International Review of Financial Analysis, 89*, 102712.

Figure 1: Daily RV Time–series

Time series plot of daily RV, computed from 5-minute intraday log returns, over the period 1 Aug 2017–24 Aug 2023. The series exhibits an extended "needle-in-haystack" pattern—long stretches of low volatility punctuated by abrupt spikes. Notable peaks coincide with the early-2018 market correction and the March-2020 COVID-19 shock, illustrating the episodic nature of extreme volatility events.

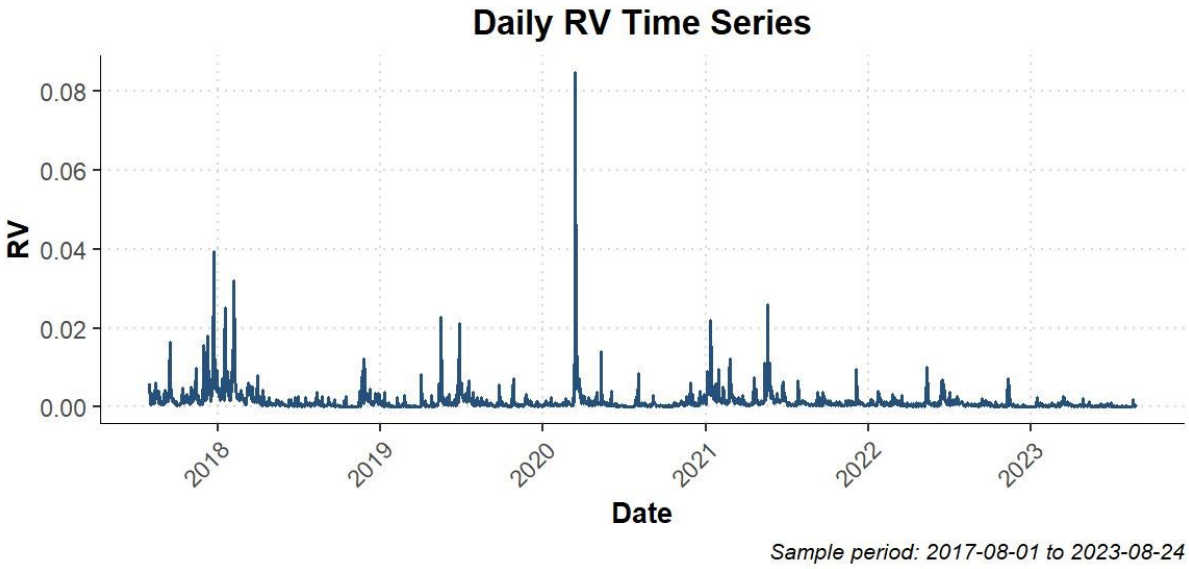

Figure 2: Daily RV Distribution

Histogram of Bitcoin's daily RV with a dashed red line showing the normal density fitted to the sample mean and variance. The distribution is strongly right-skewed and leptokurtic, indicating far thicker tails than the Gaussian benchmark and underscoring the inadequacy of normality-based inference for volatility.

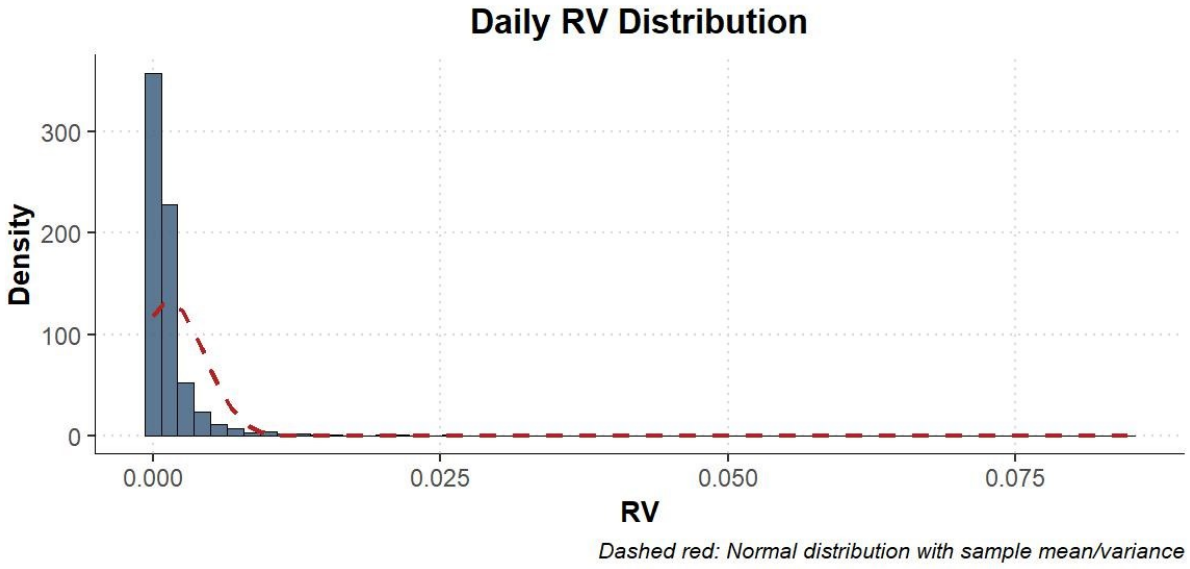

Figure 3: Daily RV Q-Q Plot

Q–Q plot comparing sample quantiles of daily RV to theoretical quantiles of the normal distribution. The pronounced upward curvature in the upper tail confirms that extreme RV realizations occur much more frequently than a normal model predicts.

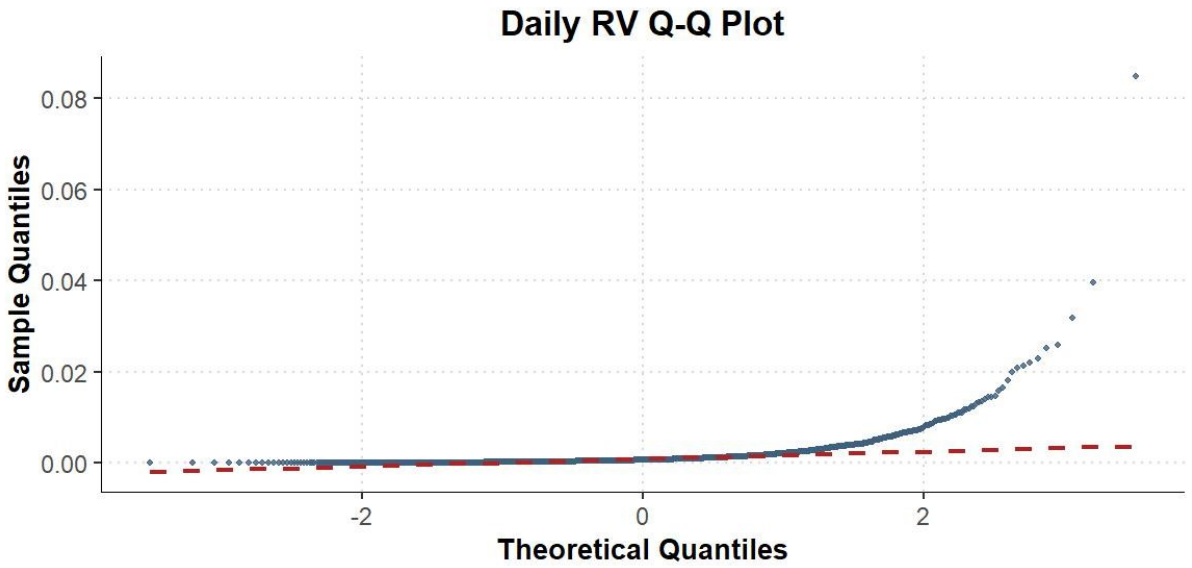

Figure 4: Daily RV ACF/PACF
(a) ACF and (b) PACF of daily RV, plotted for lags 1–50. Both functions decay only gradually, revealing the long-memory behaviour typical of volatility processes. This persistence implies that current RV retains predictive power for volatility several weeks ahead, motivating the long-horizon HAR-type models.

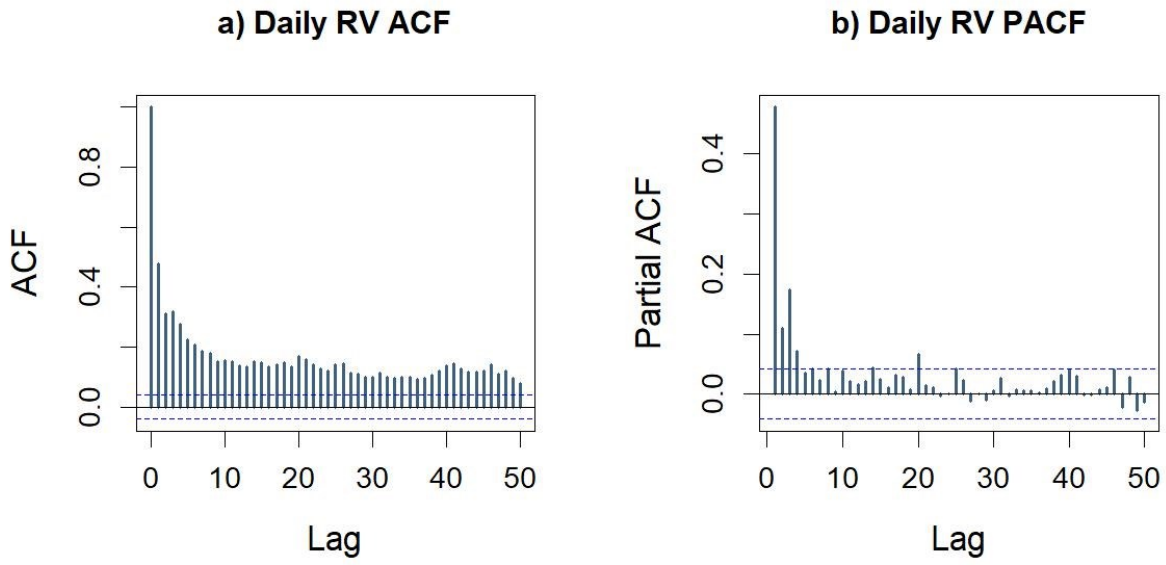

Table 1: Summary Statistics for Bitcoin Price and Volatility Measures

The table provides summary statistics for key variables from the 5-minute Bitcoin dataset (August 1, 2017 – August 24, 2023), including the median transaction price (USD), 5-minute log returns, and daily RV. Reported measures include minimum, mean, median, maximum, standard deviation, skewness, and kurtosis, highlighting the distributional properties and volatility dynamics of Bitcoin markets.

Variable	Min	Mean	Median	Max	StdDev	Skewness	Kurtosis	# Obs.
5-min Median Price	2,625	20,041	11,720	68,781	16,010	1.023	2.954	637,919
5-min Log Return	-0.087	3.47E-06	0.000	0.106	0.002	-0.365	76.545	637,919
Daily RV	7.73E-06	0.001	0.001	0.085	0.003	12.733	286.322	2,215

Table 2: Summary Statistics for HAR Model Inputs

The table reports the Descriptive statistics for RV inputs used in HAR models, derived from the 5-minute Bitcoin dataset (August 1, 2017–August 24, 2023). Variables include daily RV, weekly to annual RV averages, positive and negative semivariances, positive and negative jumps, and BV. Measures include minimum, median, mean, maximum, standard deviation, skewness, and kurtosis, illustrating the heavy-tailed and asymmetric properties of Bitcoin's volatility.

Parameter	Min	Median	Mean	Max	StdDev	Skewness	Kurtosis	# Obs.
RV_d	0.00001	0.00069	0.00145	0.08474	0.00303	12.72476	283.06303	2,215
RV_w	0.00005	0.00084	0.00145	0.02156	0.00200	4.34792	26.68082	2,209
RV_m	0.00010	0.00095	0.00145	0.00814	0.00147	2.32868	5.62178	2,186
RV_q	0.00022	0.00102	0.00145	0.00638	0.00118	1.95945	4.19664	2,126
RV_{6m}	0.00039	0.00110	0.00144	0.00434	0.00091	1.69714	2.68607	2,036
RV_y	0.00046	0.00129	0.00137	0.00287	0.00049	0.84027	0.72666	1,856
RS^+	0.00000	0.00034	0.00071	0.04233	0.00152	13.10383	290.36508	2,215
RS^-	4.60E-06	0.00033	0.00074	0.04241	0.00159	11.61025	237.51026	2,215
J^+	0	0	0	0.00783	0.00039	8.48657	107.78308	2,215
J^-	-0.00899	-1.28E-06	-0.00015	0	0.00053	-9.53087	124.85231	2,215
BV	4.61E-06	0.00055	0.00121	0.07098	0.00269	12.07614	241.52965	2,215

Table 3: Unit-Root and Diagnostic Tests for Daily RV

This table provides results of unit-root and diagnostic tests applied to the daily RV series from the Bitcoin dataset (August 1, 2017–August 24, 2023) on the training set. Tests include Augmented Dickey–Fuller (ADF), Phillips–Perron (PP), KPSS, Ljung–Box, ARCH LM, and Jarque–Bera, with test statistics, *p*-values, and methods, confirming stationarity, autocorrelation, heteroskedasticity, and non-normality consistent with HAR-type models.

Test	Statistic	P.value	Method
ADF	-9.23	0.010	Augmented Dickey-Fuller Test
KPSS	1.13	0.010	KPSS Test for Level Stationarity
Phillips-Perron	-2261	0.010	Phillips-Perron Unit Root Test
Ljung-Box	2397	0.000	Box-Ljung test
ARCH	22.74	0.030	ARCH LM-test; Null hypothesis: no ARCH effects
Jarque-Bera	7468208	0.000	Jarque Bera Test

Table 4: Correlation Matrix of HAR Model Inputs

The table illustrates the correlation matrix of RV inputs for HAR models, computed on the training set of the 5-minute Bitcoin dataset (August 1, 2017–August 24, 2023). Variables include daily RV, weekly to annual RV averages, positive and negative semivariances, positive and negative jumps, and BV, highlighting the relationships and long-memory properties.

	RV_d	RS^+	RS^-	J^+	J^-	BV	RV_w	RV_m	RV_q	RV_{6m}	RV_y
RV_d	1										
RS^+	0.974	1									
RS^-	0.976	0.902	1								
J^+	0.381	0.527	0.224	1							
J^-	-0.417	-0.229	-0.577	0.092	1						
BV	0.99	0.969	0.961	0.364	-0.373	1					
RV_w	0.609	0.609	0.579	0.284	-0.202	0.611	1				
RV_m	0.42	0.422	0.398	0.225	-0.149	0.426	0.706	1			
RV_q	0.301	0.301	0.285	0.181	-0.127	0.3	0.494	0.765	1		
RV_{6m}	0.15	0.158	0.135	0.133	-0.058	0.146	0.267	0.484	0.785	1	
RV_y	0.075	0.084	0.064	0.076	-0.015	0.068	0.14	0.231	0.3	0.418	1

Table 5: Estimated Coefficients For HAR-Type Regressions of *h*-day-ahead RV

For each forecast horizon $h \in \{1,7,30,90\}$ the table lists OLS point estimates (first line in each cell) and conventional OLS STD (square brackets) for the four model variants introduced in Section 4.2:

- **HAR** — baseline model with daily, weekly, and monthly realised variance terms (RV_d, RV_w, RV_m);
- **HAR-RS** — replaces RV_t by positive and negative semivariance (RS^+, RS^-);
- **HAR-J** — augments HAR with continuous-path proxy BV_t and signed jump variation J_t;
- **HAR-RS-J** — combines semivariance and jump terms.

Regressor labels correspond to the notation in Chapter 4 (RV_q= quarterly mean, RV_{6m}= six-month mean, RV_y = yearly mean). *, **, and *** represent 10%, 5%, and 1% levels of significance respectively. The bottom row reports the adjusted R^2 for each regression.

	***h* = 1**				***h* = 7**				***h* = 30**				***h* = 90**			
	HAR	**HAR-RS**	**HAR-J**	**HAR-RS-J**	**HAR**	**HAR-RS**	**HAR-J**	**HAR-RS-J**	**HAR**	**HAR-RS**	**HAR-J**	**HAR-RS-J**	**HAR**	**HAR-RS**	**HAR-J**	**HAR-RS-J**
ϕ_d	0.3431*** [0.0231]															
ϕ_w	0.2186*** [0.0449]	0.2429*** [0.0442]	0.2086*** [0.0433]	0.2425*** [0.0421]	2.0604*** [0.1864]	1.1195*** [0.2075]	1.0340*** [0.2062]	1.1051*** [0.2055]								
ϕ_m	0.2319*** [0.0536]	0.2362*** [0.0527]	0.2163*** [0.0517]	0.1744*** [0.0503]	2.0697*** [0.3424]	2.1695*** [0.3329]	2.1315*** [0.3315]	1.9954*** [0.3307]	8.3852*** [0.8236]	6.9992*** [0.8444]	6.8015*** [0.8433]	6.7350*** [0.8381]				
ϕ_q					0.7037* [0.3499]	0.6388. [0.3401]	0.6611. [0.3392]	0.7345* [0.3376]	3.3562* [1.3160]	3.3800** [1.3000]	3.4224** [1.2968]	3.4439** [1.2886]	17.7270*** [3.0654]	15.4716*** [3.0725]	15.3231*** [3.0651]	15.8143*** [3.0361]
ϕ_{6m}									-3.8380** [1.2511]	-3.6912** [1.2356]	-3.6566** [1.2327]	-3.6061** [1.2250]	-24.4654*** [4.5040]	-23.9110*** [4.4713]	-23.9949*** [4.4578]	-25.1233*** [4.4180]
ϕ_y													-1.3929 [3.6381]	-1.6324 [3.6137]	-1.4559 [3.6032]	-0.2519 [3.5734]
ϕ_{d^+}		-0.3769*** [0.0861]		-0.9277*** [0.1733]		-0.8831* [0.4019]		-3.7121*** [0.8386]		-0.3398 [1.1130]		-8.9116*** [2.3078]		1.6088 [2.7237]		-19.5136*** [5.5705]
ϕ_{d^-}		1.0098*** [0.0802]		-2.2343*** [0.2315]		2.8565*** [0.3721]		-2.8224* [1.1279]		3.9233*** [1.0285]		-10.8784*** [3.2523]		4.4646. [2.4235]		-33.4019*** [8.1715]
ϕ_{BV}			0.3155*** [0.0280]	1.9978*** [0.1462]			1.1843*** [0.1320]	4.6664*** [0.7070]			2.3597*** [0.3453]	13.2775*** [2.1516]			4.4497*** [0.7542]	33.8484*** [4.9503]
ϕ_{J^+}			0.1429 [0.1518]				-1.5385* [0.7218]				-2.5218 [1.9567]				-5.2448 [5.0888]	
ϕ_{J^-}			-1.2477*** [0.1126]	-2.5053*** [0.2260]			-2.1719*** [0.5284]	-2.9702** [1.0965]			-1.9216 [1.4465]	-6.3695* [2.9853]			0.7767 [3.2110]	-15.2948* [7.3241]
Adj. R^2	**26.5%**	**29.0%**	**31.8%**	**35.8%**	**26.3%**	**30.4%**	**31.0%**	**31.8%**	**12.6%**	**14.8%**	**15.3%**	**16.3%**	**2.0%**	**3.5%**	**4.0%**	**5.9%**

Table 6: In-Sample Residual Diagnostics

The table reports residual diagnostic tests for all HAR specifications—HAR, HAR-RS, HAR-J, and HAR-RS-J—estimated on the training sample (August 1, 2017–August 24, 2023). Diagnostics include Ljung–Box Q-tests on residuals and squared residuals (lags adjusted to capture serial dependence at each forecast horizon), Engle's ARCH LM test for conditional heteroskedasticity, White's heteroskedasticity test, and normality tests (Shapiro–Wilk or Jarque–Bera). Models are evaluated across daily, weekly, monthly, and quarterly horizons.

Model/Horizon	LB ε_t	LB ε_t^2	ARCH LM	White	Shapiro
h = 1					
HAR	6.33E-02	4.66E-04	1.16E-08	1.08E-43	8.90E-68
HAR-RS	5.74E-02	9.47E-05	1.29E-09	9.44E-62	1.07E-66
HAR-J	3.24E-02	3.82E-03	3.41E-07	6.24E-164	1.95E-66
HAR-RS-J	5.18E-04	1.57E-01	3.28E-04	4.35E-211	1.18E-65
h = 7					
HAR	0	0	0	1.85E-06	8.70E-60
HAR-RS	0	0	0	3.81E-10	7.88E-60
HAR-J	0	0	0	5.46E-09	7.05E-60
HAR-RS-J	0	0	0	7.01E-06	9.71E-60
h = 30					
HAR	0	0	0	1.20E-04	4.54E-51
HAR-RS	0	0	0	7.65E-05	4.26E-51
HAR-J	0	0	0	6.20E-05	3.02E-51
HAR-RS-J	0	0	0	1.32E-02	3.35E-51
h = 90					
HAR	0	0	6.76E-288	9.36E-55	2.33E-33
HAR-RS	0	0	6.84E-277	3.79E-53	3.60E-33
HAR-J	0	0	1.87E-265	7.12E-61	2.44E-33
HAR-RS-J	0	0	5.75E-273	2.72E-49	4.13E-33

Table 7: Forecast Performance of HAR-Type Models

This table presents out-of-sample forecast performance for four HAR-type models across horizons $h = 1,7,30,90$ calendar days. Forecasts are generated using the rolling-window approach over the period August 25, 2023 – February 28, 2025. Performance is evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the QLIKE loss function. Bold figures indicate the lowest loss per metric and horizon.

Horizon / Model	RMSE	MAE	QLIKE
h = 1			
HAR	0.00064	0.00038	0.35926
HAR-RS	0.00063	0.00037	0.35487
HAR-J	**0.00062**	**0.00034**	**0.32979**
HAR-RS-J	0.00065	0.00040	137.35995
h = 7			
HAR	0.00283	0.00247	0.19349
HAR-RS	**0.00275**	**0.00240**	0.18409
HAR-J	0.00276	**0.00240**	**0.18405**
HAR-RS-J	0.00286	0.00250	369.80394
h = 30			
HAR	0.01292	0.01157	0.18959
HAR-RS	**0.01262**	**0.01129**	**0.18384**
HAR-J	0.01268	0.01135	0.18449
HAR-RS-J	0.01298	0.01163	0.18877
h = 90			
HAR	0.04202	0.03847	0.16259
HAR-RS	**0.04162**	**0.03802**	**0.16052**
HAR-J	0.04173	0.03816	0.16107
HAR-RS-J	0.04196	0.03845	0.16212

Table 8: Robustness Checks for Forecast Performance

QLIKE loss values for the four HAR-type models under three scenarios: baseline (5-minute sampling, full sample), alternative sampling frequency (10-minute returns), and post-COVID subsample (March 15, 2020 – February 28, 2025). Forecast horizons are $h = 1,7,30,90$ calendar days. Bold values denote the lowest QLIKE for each horizon within a robustness scenario.

Horizon / Model	Baseline (5-min)	10-min	Post-COVID
h = 1			
HAR	0.35926	0.39051	**0.27097**
HAR-RS	0.35487	0.39529	0.27156
HAR-J	**0.32979**	**0.35666**	0.27548
HAR-RS-J	137.35995	337.93234	0.27607
h = 7			
HAR	0.19349	0.20112	0.13811
HAR-RS	0.18409	**0.19117**	**0.12882**
HAR-J	**0.18405**	0.19483	0.13154
HAR-RS-J	369.80394	0.20739	0.13076
h = 30			
HAR	0.18959	0.20261	0.08008
HAR-RS	**0.18384**	**0.19677**	**0.07793**
HAR-J	0.18449	0.19901	0.07886
HAR-RS-J	0.18877	0.20295	0.07796
h = 90			
HAR	0.16259	0.18045	**0.02228**
HAR-RS	**0.16052**	**0.17864**	0.02351
HAR-J	0.16107	0.17998	0.02323
HAR-RS-J	0.16212	0.18072	0.02261

Table 9: Forecast Accuracy: HAR-type Models vs. GARCH Benchmark

Comparison of the best-performing HAR-type model at each horizon with the best GARCH specification, selected by Bayesian Information Criterion (GARCH(1,1) with generalized error distribution). Out-of-sample forecasts are evaluated using MAE, RMSE, and QLIKE across horizons $h = 1,7,30,90$ calendar days.

Horizon / Model	RMSE	MAE	QLIKE
h = 1			
HAR-J	**0.00062**	**0.00034**	**0.32979**
GARCH	0.00069	0.00046	0.41559
h = 7			
HAR-RS	0.00275	0.00240	0.18409
GARCH	**0.00038**	**0.00033**	**0.16737**
h = 30			
HAR-RS	0.01262	0.01129	0.18384
GARCH	**0.00028**	**0.00025**	**0.09975**
h = 90			
HAR-RS	0.04162	0.03802	0.16052
GARCH	**0.00026**	**0.00025**	**0.07110**