

Realized Volatility Forecasting: Continuous versus Discrete Time Models*

Shuping Shi[†], Jun Yu^{††}, Chen Zhang^{†††}

[†] Macquarie University

^{††} University of Macau

^{†††} Sun Yat-sen University

October 21, 2025

Abstract

Forecasting realized volatility (RV) is central to financial econometrics, with important implications for risk management, asset allocation, and derivative pricing. Motivated by the ongoing debate on volatility modeling, this paper provides a comprehensive empirical comparison of many alternative models. We evaluate leading continuous time models estimated using state-of-the-art methods from the rough volatility literature, together with both standard long-memory autoregressive fractionally integrated moving average (ARFIMA) models and their rough-volatility extensions, as well as several variants of the heterogeneous autoregressive (HAR) model and their logarithmic counterparts. The models are applied to a large panel of equities and cryptocurrencies, with performance assessed using both statistical and economic criteria. Our results show that for equities, continuous time models consistently outperform discrete time alternatives across all evaluation criteria and forecasting horizons. The fractional Brownian motion model for log RV performs best at short horizons, while the fractional Ornstein Uhlenbeck model for log RV dominates in the long run. For cryptocurrencies, a mild divergence emerges between economic and statistical performance: based on realized utility, the quarticity-augmented heterogeneous autoregressive (HARQ) model for RV leads in the short term and the Brownian semistationary models prevail at longer horizons, whereas the HAR-type models for log RV deliver superior statistical accuracy.

JEL classification: C12, C22, G01

Keywords: Realized volatility, Continuous-time models, Discrete-time models, forecasting, economic utility

*Shuping Shi, Department of Economics, Macquarie University, Sydney, Australia. Email: shuping.shi@mq.edu.au. Jun Yu, Department of Finance and Business Economics, Faculty of Business Administration, University of Macau, Avenida da Universidade, Taipa, Macao, China. Email: junyu@um.edu.mo. Chen Zhang, Lingnan College, Sun Yat-sen University. Email: chen.zhangyw@gmail.com. Yu would like to acknowledge the financial support from the University of Macau Development Fund.

1 Introduction

Volatility is a fundamental concept in financial economics. Since the seminal work of Engle (1982), extensive efforts have been devoted to modeling and forecasting the dynamics of financial market volatility (Poon and Granger, 2003; Hansen and Lunde, 2005). Early studies relied on daily returns to estimate volatility, but it is now well recognized that daily returns provide noisy and inefficient measures. Andersen and Bollerslev (1998) demonstrate the advantages of using intraday returns to measure, model, and forecast daily volatility. The realized volatility (RV), defined as the sum of squared intraday returns, offers a much more accurate estimate of daily volatility than the squared daily return. Consequently, the literature on modeling and forecasting realized volatility has expanded rapidly.

Forecasting realized volatility is a central and extensively explored topic in financial econometrics, with direct implications for risk management, portfolio allocation, and derivative pricing. Empirical studies have documented that realized volatility exhibits long memory (Bollerslev et al., 2000; Andersen et al., 2001, 2003), characterized by slowly decaying autocorrelations and a persistent response to past shocks. To capture this feature, most existing studies rely on discrete time frameworks, particularly long-memory models like Autoregressive Fractionally Integrated Moving Average (ARFIMA) models (Granger, 1980; Granger and Joyeux, 1980) with $d > 0$ where d is the memory parameter and Heterogeneous Autoregressive (HAR) variants (Corsi, 2009), due to their simplicity and empirical success (Andersen et al., 2003; Corsi and Renò, 2012; Bollerslev et al., 2016; Patton and Sheppard, 2015).

More recently, a growing body of literature has provided evidence supporting the rough volatility paradigm in log RV (see, e.g., Gatheral et al., 2018; Bolko et al., 2022; Wang et al., 2023; Chong and Todorov, 2025). This line of research, grounded in continuous time modeling, suggests that log volatility follows trajectories similar to those generated by a fractional Brownian motion (fBm) or a fractional OU (fOU) process with the Hurst parameter (denoted by H) below 0.5, whose Hölder continuity is less than that of a standard Brownian motion, thereby giving the name of roughness. This line of results seems to be contradicting with that of long-memory ARFIMA models in the lens of weak convergence.¹ While the debate over whether volatility is fundamentally rough or exhibits long memory remains unresolved (Shi and Yu, 2023; Li et al., 2025), this paper shifts the focus away from theoretical distinctions toward practical considerations: namely, which class of models (discrete or continuous) delivers better out-of-sample forecasts of realized volatility.

Despite their theoretical appeal and widespread applications in the rough volatility literature and the option pricing literature (see, e.g., Sundaresan, 2000; Gatheral et al., 2018), rough continuous time models have been less

¹As shown in Tanaka (2013) and Wang and Yu (2023), under an infill scheme, ARFIMA(1, d , 0) weakly converges to fOU(H) with $H = d + 0.5$, implying that $d > 0$ and $H < 0.5$ cannot co-exist.

commonly applied to forecast RV, with some notable exceptions. [Wang et al. \(2023\)](#) and [Wang et al. \(2024\)](#) examine the predictive performance of continuous time models such as fBm and fOU process. In particular, [Wang et al. \(2023\)](#) compares the fOU model to fBm, ARFIMA and HAR models in forecasting realized volatility. However, the scope of that analysis is limited. It considers only a narrow set of models, some of which are estimated by inefficient methods.

Moreover, the weak identification problem in ARFIMA models was recently reported in ([Shi and Yu, 2023](#); [Li et al., 2025](#)), suggesting that it is difficult to distinguish a stationary near unit root ARFIMA(1, d , 0) model with $d \in (-0.5, 0)$ from a nonstationary near zero root ARFIMA(1, d , 0) model with $d \in (0.5, 1)$. Hence, it is important to examine the forecasting performance of discrete-time models with long-memory and rough features, as well as those exhibiting stationary and nonstationary properties. Since the a stationary near unit root ARFIMA(1, d , 0) model weakly converges to a rough fOU process, such an extension is natural given the strong evidence of roughness in the continuous time literature.

We also evaluate several variants of the heterogeneous autoregressive (HAR) model. Since the seminal paper by [Corsi \(2009\)](#), numerous interesting HAR variants have been proposed (see, e.g., [Patton, 2011](#); [Patton and Sheppard, 2015](#); [Laurent et al., 2024](#)). However, empirical comparisons are often benchmarked at the traditional HAR or log HAR models (see, e.g. [Clements and Preve, 2021](#)), leaving it unclear how far the literature has progressed relative to the best HAR-type model. Finally, much of the existing forecasting literature relies on suboptimal estimation methods. For instance, [Andersen et al. \(2003\)](#) estimate ARFIMA models using semiparametric techniques that yield inefficient estimators, while [Wang et al. \(2023\)](#) estimate the fOU process via the method of moments, which is also inefficient. It is therefore essential to reassess model performance when all competing frameworks are estimated using efficient and comparable methods.

In this paper, we examine four major continuous time models: fBm, fOU, and two Brownian semi-stationary (BSS) processes ([Bennedsen et al., 2022](#)). These are compared against seven discrete time models, including stationary ARFIMA, non-stationary ARFIMA, RV-HAR, RV-HARQ (quarticity-augmented HAR), RV-HAR-SV (semivariance-based HAR), and the log-transformed counterparts of the HAR-type models. These models are illustrated below and classified into two categories: continuous time models, as developed in the rough volatility literature, and all remaining models, which are labeled as discrete time models. To ensure a fair and robust comparison, we apply *state-of-the-art* estimation techniques to all models. Forecasting performance is assessed using both statistical and economic criteria. Statistically, we employ the model confidence set (MCS, [Hansen et al., 2011](#)) test to identify superior models. Economically, we evaluate realized utility from the perspective of a risk parity investor as in [Bollerslev et al. \(2018\)](#).

The analysis spans a broad range of financial assets (including 11 major index ETFs, 30 Dow Jones Industrial

Average stocks, and 30 cryptocurrencies), covering the period from 2010 to 2024 for equities and from 2017 to 2024 for cryptocurrencies. This diverse and extensive dataset enables a robust evaluation across asset classes and market conditions.

Our findings reveal several important patterns across asset classes. For equities, continuous time models consistently outperform all discrete time alternatives, achieving higher realized utility and lower forecast losses across horizons. The fBm model for log RV estimated via the approximate Whittle maximum likelihood method performs best at short and medium horizons, while the mean-reverting fOU models for log RV dominate in the long run. For cryptocurrencies, the HARQ model for RV delivers the highest short-term utility, whereas the continuous time BSS models for log RV prevail at medium and long horizons. Interestingly, while the continuous time models for log RV yield greater economic gains at these horizons, the HAR-type models for log RV exhibit superior statistical accuracy across all forecast horizons, highlighting a subtle yet important distinction between economic and statistical forecasting performance.

Five key innovations distinguish our study from prior studies. First, we consider a substantially broader set of both continuous time and discrete time models, enabling a more comprehensive comparison of alternative models for RV. In this sense, our study can be viewed as an extension of [Poon and Granger \(2003\)](#) from the GARCH literature to the realized volatility literature. Second, we employ advanced estimation techniques to improve estimation accuracy. Third, in the context of the ARFIMA literature, we explicitly address the identification problem by considering the long-memory ARFIMA model and the ARFIMA model that is rough and stationary. Fourth, we complement conventional statistical evaluation with an economic utility-based criterion, offering a more practical and investor-relevant perspective on forecasting performance. Fifth, we evaluate model performance across a wide-ranging empirical dataset that includes both traditional financial assets and non-traditional ones such as cryptocurrencies, enhancing the generalizability of our findings.

The remainder of the paper is structured as follows. Section 2 introduces the continuous time models considered in our analysis, while Section 3 presents the discrete time models. Section 4 describes the forecasting methods and evaluation criteria. The empirical results are reported in Section 5, where we compare model performance across asset classes and forecast horizons. Section 6 concludes.

2 Continuous time Models for Volatility Dynamics

This section introduces four major continuous time models including fBm, fOU, and BSS, all used to model log RV. For each model, we outline its defining features and present the corresponding estimation methods.

2.1 Fractional Brownian motion process

The fBm process $\{y_t : t \in \mathbb{R}\}$ is

$$y_t = \sigma B^H(t) \text{ with } H \in (0, 1). \quad (1)$$

It is a continuous time Gaussian process with zero mean and the autocovariance function of the following:

$$\text{Cov}(y_t, y_s) = \frac{1}{2} \sigma^2 \left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right), \quad \forall t, s \in (-\infty, +\infty), \quad (2)$$

where $\sigma > 0$ is a constant and $H \in (0, 1)$ is called the Hurst parameter. When $H = 0.5$, $B^H(r) = B(r)$ becomes a standard Brownian motion.

Suppose that y_t is sampled at grids $\Delta, 2\Delta, \dots, T\Delta$, where Δ is the sampling interval. Let $x_{i\Delta} = y_{i\Delta} - y_{(i-1)\Delta}$ denote the increment of fBm, where $i = 1, 2, \dots, T$. The variable $x_{i\Delta}$ is known in the literature as the fractional Gaussian noise (fGn). Both fBm and fGn are Gaussian processes. While fBm is nonstationary, fGn is stationary. The autocovariance of $x_{i\Delta}$ is given by

$$\begin{aligned} \text{Cov}(x_{i\Delta}, x_{(i-j)\Delta}) &= \frac{\sigma^2}{2} \Delta^{2H} [(j+1)^{2H} + (j-1)^{2H} - 2j^{2H}], \text{ for any } j \geq 0 \\ &\sim \sigma^2 \Delta^{2H} H(2H-1) j^{2H-2} \text{ for large } j, \end{aligned} \quad (3)$$

where $\sim$ denotes asymptotic equivalence. The near-zero-frequency behavior of $\text{Cov}(x_{i\Delta}, x_{(i-j)\Delta})$ for fGn is the same as that for ARFIMA(0, d , 0) when $d = H - 0.5 \in (-0.5, 0.5)$. Hence, fBm y_t is expected to have similar low frequency behavior as ARFIMA(0, $d + 1$, 0).

Equation (3) shows that when $H \in (0.5, 1)$, $x_{i\Delta}$ has positive serial dependence, and the autocovariances of $x_{i\Delta}$ are not absolutely summable. As a result, $x_{i\Delta}$ has long memory when $H > 0.5$. In contrast, if $H \in (0, 0.5)$, it can be proved that, $x_{i\Delta}$ has negative autocovariances and the long variance is zero, i.e.,

$$\sum_{j=-\infty}^{+\infty} \text{Cov}(x_{i\Delta}, x_{(i-j)\Delta}) = 0.$$

In this case, $x_{i\Delta}$ is anti-persistent. The fGn generates short-term reversal, and hence, the corresponding fBm has sample paths that are rougher than those of the standard Brownian motion. When H gets closer to 0, the sample path becomes rougher.

Estimation Methods

The fBm process has two unknown parameters, H and σ^2 . [Gatheral et al. \(2018\)](#) and [Bennedsen et al. \(2022\)](#) use ordinary least squares and nonlinear least squares, respectively, to estimate the fractional parameter. Meanwhile, [Lang and Roueff \(2001\)](#), [Barndorff-Nielsen et al. \(2013\)](#), and [Brouste et al. \(2020\)](#) propose using the change-of-frequency (CoF) estimators. Although these estimators are not efficient since they only utilize partial information from the process, fBm models with these estimators still yield more accurate forecasts of RV compared to the HAR model of [Corsi \(2009\)](#), as shown by [Gatheral et al. \(2018\)](#), [Bennedsen et al. \(2022\)](#), and [Wang et al. \(2024\)](#).

Parameters of the fBm process can be estimated from fGn. [Fukasawa and Takabatake \(2019\)](#) propose an approximate Whittle maximum likelihood (AWML) method that utilizes an approximated spectral density, demonstrating asymptotic efficiency in the Fisher sense when ignoring approximation errors. [Shi et al. \(2025\)](#) develop a computationally feasible expression for the spectral density, resulting in the exact Whittle maximum likelihood (EWML) method. Since the autocovariances are available in closed-form and fGn is a Gaussian process, a time-domain maximum likelihood (TDML) method can also be implemented. Although the time-domain likelihood and the Whittle likelihood are asymptotically equivalent, they may differ in finite sample. The difference can be further exacerbated by approximation errors in the spectral density. Their simulation study reveals that the TDML method yields the most accurate estimates. AWML is comparable to the EWML but slightly less accurate. In terms of computational cost, the ranking of methods is AWML, then TDML, and finally EWML. We now introduce the TDML and AWML methods for fGn.

TDML Let $X = (x_\Delta, x_{2\Delta}, \dots, x_{T\Delta})'$ and $\varphi = (H, \sigma)$. Under the model specification, $X \sim N(0, \sigma^2 \Sigma_X)$, where the elements of $\sigma^2 \Sigma_X$ is from (3). The log likelihood function of fGn is,

$$l(\varphi; X) = -T \ln \sigma - \frac{1}{2} \ln |\Sigma_X| - \frac{1}{2\sigma^2} X' \Sigma_X^{-1} X. \quad (4)$$

Since Σ_X only depends on H , the parameter σ^2 is estimated by

$$\hat{\sigma}_{ML}^2(H) = \frac{1}{T} X' \Sigma_X^{-1} X.$$

The TDML estimator $\hat{H}_{ML}$ is then computed as

$$\hat{H}_{ML} = \arg \max_{\varphi} l(H, \hat{\sigma}_{ML}^2(H); X).$$

See Appendix A.1 for implementation details.

AWML Let λ be the spectral frequency and $f_x(\lambda)$ be the spectral density of the fGn process. The spectral density $f(\lambda)$ is given by [Sinai \(1976\)](#) and takes the form of the following:

$$f_x^\Delta(\lambda; \varphi) = \frac{\sigma^2}{\pi} \Delta^{2H} C_H (1 - \cos \lambda) \sum_{k=-\infty}^{\infty} |2\pi k + \lambda|^{-1-2H} := \sigma^2 g_X^\Delta(\lambda; H)$$

for $\lambda \in (0, \pi]$, where $C_H = \Gamma(2H + 1) \sin(\pi H)$ with $\Gamma(\cdot)$ being the Gamma function.

The Whittle log likelihood function for parameters in fGn is

$$l_W(H; X) = -\frac{1}{2\pi} \int_0^\pi \left(\ln \sigma^2 g_X^\Delta(\lambda; H) + \frac{I(\lambda)}{\sigma^2 g_X^\Delta(\lambda; H)} \right) d\lambda, \quad (5)$$

where $I(\lambda)$ is the periodogram is defined as

$$I(\lambda) = \frac{1}{2\pi T} \left| \sum_{s=1}^T x_{s\Delta} \exp(-is\lambda) \right|^2. \quad (6)$$

Since σ^2 is functionally independent of H , these two parameters can be estimated separately. The profiled log likelihood of the Whittle method is given by

$$\tilde{l}_W(H; X) = -\ln \left(\frac{1}{\pi} \int_0^\pi \frac{I(\lambda)}{g_X^\Delta(\lambda; H)} d\lambda \right) - \frac{1}{\pi} \int_0^\pi \ln g_X^\Delta(\lambda; H) d\lambda. \quad (7)$$

and $\sigma^2 = \frac{1}{\pi} \int_0^\pi \frac{I(\lambda)}{g_X^\Delta(\lambda; H)} d\lambda$. Note that the spectral density diverges to infinity as $\lambda \rightarrow 0$ when $H > 1/2$. To address this singularity issue, we divide the integrand in the objective function (7) into two segments: one ranging from 0 to ε , and the other from ε to 2π , as in [Fukasawa et al. \(2022\)](#) and [Shi et al. \(2024\)](#). The log likelihood function can be further reformulated as the follows. We first consider the estimation of H .

Lemma 2.1 *The Whittle log likelihood function for estimating H in fGn is given by $l_W^+(H; X) = \tilde{l}_W(H; X)$ when $H \leq 1/2$; when $H > 1/2$,*

$$l_W^+(H; X) = -\left\{ \log \left[\frac{1}{\pi} \left(\int_\varepsilon^\pi \frac{I_n(\lambda)}{g_X^\Delta(\lambda; H)} d\lambda + B^\ddagger(H, \varepsilon) \right) \right] + \frac{1}{\pi} \left(\int_\varepsilon^\pi \ln g_X^\Delta(\lambda; H) d\lambda + B^\dagger(H, \varepsilon) \right) \right\}$$

where

$$B^\dagger(H, \varepsilon) = \ln \left(\frac{1}{2\pi} C_H \Delta^{2H} \right) \varepsilon + (1 - 2H) (\varepsilon \ln \varepsilon - \varepsilon),$$

$$B^\ddagger(H, \varepsilon) = \delta_H(0, \varepsilon) \hat{\gamma}_n(0) + \sum_{\tau=1}^{n-1} \delta_H(\tau, \varepsilon) \hat{\gamma}_n(\tau),$$

with $\delta_H(0, \varepsilon) \approx \frac{\varepsilon^{2H}}{2HC_H\Delta^{2H}}$ and $\delta_H(\tau, \varepsilon) \approx \frac{1}{C_H\Delta^{2H}} \sum_{j=0}^J \frac{(-1)^j \tau^{2j}}{(2j)!} \frac{\varepsilon^{2(j+H)}}{2(j+H)}$ for $\tau > 1$.²

Moreover, the spectral density $g_X^\Delta(\lambda; H)$ involves infinite summation which brings challenges computationally. A Paxson approximation of the spectral density is considered by Paxson (1997); Fukasawa and Takabatake (2019), which approximates $g_X^\Delta(\lambda; H)$ with the average of its upper and lower bounds. The Paxson approximated spectral density is denoted by $\tilde{g}_x^\Delta(\lambda; H)$ and given by

$$\tilde{g}_x^\Delta(\lambda; H) = \frac{1}{\pi} C_H (1 - \cos(\lambda)) \Delta^{2H} \left\{ |\lambda|^{-\gamma_H} + \sum_{j=1}^K b(j, \lambda) + \frac{1}{2} [a(K, \lambda) + a(K+1, \lambda)] \right\}, \quad (8)$$

where K is a pre-specified integer, $\gamma_H = 2H + 1$,

$$\begin{aligned} a(k, \lambda) &= \frac{1}{4\pi H} [(2\pi k + \lambda)^{1-\gamma_H} + (2\pi k - \lambda)^{1-\gamma_H}], \\ b(j, \lambda) &= (2\pi j + \lambda)^{-\gamma_H} + (2\pi j - \lambda)^{-\gamma_H}. \end{aligned}$$

The AWML estimator $\hat{H}_{AW}$ is computed as

$$\hat{H}_{AW} = \arg \max_{\phi} l_{AW}^+(H; X),$$

where $l_{AW}^+(H; X)$ is the objective function of the approximate Whittle method, where we replace $g_X^\Delta(\lambda; H)$ in $l_W^+(H; X)$ with $\tilde{g}_X^\Delta(\lambda; H)$ and set $K = 50$. The volatility parameter σ^2 is estimated by

$$\hat{\sigma}_{AW}^2 = \frac{1}{\pi} \left[\int_{\varepsilon}^{\pi} \frac{I_n(\lambda)}{g_X^\Delta(\lambda; \hat{H}_{AW})} d\lambda + B^\ddagger(\hat{H}_{AW}, \varepsilon) \right].$$

2.2 Fractional Ornstein-Uhlenbeck process

The fOU process $\{y_t : t \in \mathbb{R}\}$ is given by

$$dy_t = \kappa(\mu - y_t) dt + \sigma dB_t^H \text{ with } y_0 = O_p(1),$$

where $\kappa, \sigma > 0$, μ is a constant, and B_s^H is a fBm process with $H \in (0, 1)$. The fOU process has a unique path-wise solution:

$$y_t = e^{-\kappa t} y_0 + (1 - e^{-\kappa t}) \mu + \int_0^t \sigma e^{-\kappa(t-u)} dB_u^H,$$

²We set $J = 100$ and $\varepsilon = 2\pi/n$ in practice.

where $E(y_t) = \mu$ and $Var(y_t) = \sigma^2 \kappa^{-2H} H \Gamma(2H)$. The fOU process reduces to the traditional OU process when $H = 0.5$ and to a fractional Brownian motion process when $\kappa = 0$.

Suppose that $\{y_{i\Delta}\}_{i=1}^T$ are sampled from the fOU process. The autocovariance of $y_{i\Delta}$ is (Garnier and Sølna, 2018)

$$Cov(y_{i\Delta}, y_{(i-j)\Delta}) = \frac{\sigma^2}{2\kappa^{2H}} \left(\frac{1}{2} \int_{-\infty}^{\infty} e^{-|s|} |\kappa j\Delta + s|^{2H} ds - |\kappa j\Delta|^{2H} \right). \quad (9)$$

Cheridito et al. (2003) show that for any $\kappa > 0$, autocovariances of fOU behave like that of fGn when the lag length j is large. When $H \in (0.5, 1)$, the autocovariances are not absolutely summable. Unlike fBm that is always nonstationary, the fOU process is stationary as long as $\kappa > 0$. Of course, when κ is positive and close to zero, fOU should have similar properties to fBm in finite sample.

Estimation Method

The fOU model includes two additional parameters, κ and μ , compared to the fBm process. Wang et al. (2023) propose a two-stage estimation method: first, they estimate H and σ from second-order differences, followed by estimating μ using the sample mean and κ based on the previously estimated parameters. When modelling log RV, Wang et al. (2023) find that the fOU model provides more accurate forecasts for RV than both ARFIMA and HAR models, as well as fBm. However, the comparison between fOU and fBm may be biased, as Wang et al. (2023) employed sub-optimal forecasting formulas for both models, as noted by Wang et al. (2024). Additionally, the estimation methods used for both fOU and ARFIMA may not be optimal.

For the fOU model, Bennedsen et al. (2022) propose a Maximum Composite Likelihood (MCL) method that outperforms the CoF method in simulation studies. However, both approaches use only partial information from the process, limiting their efficiency. Shi et al. (2024) introduce an AWML method for estimating κ , σ , and H in fOU, yielding more accurate estimates than the MCL and CoF methods. To estimate μ , Shi et al. (2024) use the sample mean. Additionally, Wang et al. (2024) propose the TDML method to estimate all four parameters in fOU, showing improved estimation accuracy. We now introduce the TDML and AWML methods.

TDML Let $\mathbf{y} = (y_{1\Delta}, y_{2\Delta}, \dots, y_{T\Delta})'$ and $\theta = (\mu, \sigma, \kappa, H)$. Under the fOU process, $\mathbf{y} \sim N(0, \sigma^2 \Sigma_{\mathbf{y}})$. The log likelihood function of fOU can be written as

$$l(\theta; \mathbf{y}) \propto -T \ln \sigma - \frac{1}{2} \ln |\Sigma_{\mathbf{y}}| - \frac{1}{2\sigma^2} (\mathbf{y} - \mu \mathbf{1})^\top \Sigma_{\mathbf{y}}^{-1} (\mathbf{y} - \mu \mathbf{1}). \quad (10)$$

One important task is to compute the variance-covariance matrix $\sigma^2 \Sigma_{\mathbf{y}}$ using (9) which involves an integral.

Wang et al. (2024) provide an analytical expression for elements in $\sigma^2 \Sigma_{\mathbf{y}}$:

$$\text{Cov}(y_{i\Delta}, y_{(i-j)\Delta}) = \frac{\sigma^2}{2\kappa^{2H}} \left[\cosh(\kappa j \Delta) \Gamma(2H+1) - (\kappa j \Delta)^{2H} {}_1F_2 \left(1; H + \frac{1}{2}, H+1; \frac{(\kappa j \Delta)^2}{4} \right) \right], \quad (11)$$

where $\cosh(x) = [\exp(x) + \exp(-x)]/2$ is the hyperbolic cosine function and ${}_1F_2(\cdot; \cdot; \cdot)$ denotes the generalized hypergeometric function as

$${}_1F_2 \left(1; H + \frac{1}{2}, H+1; \frac{(\kappa j \Delta)^2}{4} \right) = \sum_{n=0}^{\infty} \frac{\Gamma(H+1/2) \Gamma(H+1)}{\Gamma(H+1/2+n) \Gamma(H+1+n)} \left(\frac{\kappa j \Delta}{2} \right)^{2n}. \quad (12)$$

Note that the elements in $\Sigma_{\mathbf{y}}$ depend on κ and H only. We can profile the log likelihood by

$$\mu(\kappa, H) = \frac{\mathbf{1}^\top \Sigma_{\mathbf{y}}^{-1} \mathbf{y}}{\mathbf{1}^\top \Sigma_{\mathbf{y}}^{-1} \mathbf{1}}, \quad (13)$$

which leads to

$$\sigma^2(\kappa, H) = \frac{1}{T} \left[\mathbf{y}^\top \Sigma_{\mathbf{y}}^{-1} \mathbf{y} - \frac{(\mathbf{1}^\top \Sigma_{\mathbf{y}}^{-1} \mathbf{y})^2}{\mathbf{1}^\top \Sigma_{\mathbf{y}}^{-1} \mathbf{1}} \right]. \quad (14)$$

Substituting (13) and (14) into (10) yields the following profile log likelihood function

$$l(\kappa, H; \mathbf{y}) \propto -T \ln \sigma(\kappa, H) - \frac{1}{2} \ln |\Sigma_{\mathbf{y}}|. \quad (15)$$

Therefore, the TDML estimators of κ and H are obtained as

$$(\hat{\kappa}_{ML}, \hat{H}_{ML}) = \arg \max_{\kappa, H} l(\kappa, H; \mathbf{y}). \quad (16)$$

Consequently, using (13) and (14), the TDML estimators of μ and σ are

$$\hat{\mu}_{ML} = \mu(\hat{\kappa}_{ML}, \hat{H}_{ML}) \text{ and } \hat{\sigma}_{ML} = \sigma(\hat{\kappa}_{ML}, \hat{H}_{ML}). \quad (17)$$

AWML Let $\beta = (H, \kappa, \sigma)$. The discrete time spectral density of the fOU process is provided by Hult (2003) and takes the form of

$$f_{\mathbf{y}}^{\Delta}(\lambda; \beta) = \frac{\sigma^2}{2\pi} C(H) \Delta^{2H} \sum_{k=-\infty}^{\infty} \frac{|\lambda + 2\pi k|^{1-2H}}{(\kappa \Delta)^2 + (\lambda + 2\pi k)^2} \text{ for } \lambda \in (0, 2\pi).$$

The spectral density of fOU involves infinite sum and hence is computationally challenging. [Shi et al. \(2024\)](#) provides a modified Paxson approximation of the spectral density, given by

$$\begin{aligned} \tilde{f}_y^\Delta(\lambda; \beta) = & \frac{\sigma^2}{2\pi} C(H) \Delta^{2H} \left[\sum_{k=1}^K Q_{1,k}(\lambda) + \sum_{k=1}^K Q_{2,k}(\lambda) + \frac{\lambda^{1-2H}}{(\Delta\kappa)^2 + \lambda^2} \right. \\ & \left. + \frac{1}{2}a(K, H, \lambda) + \frac{1}{2}b(K, H, \lambda) + c(K, H, \lambda) \right], \end{aligned} \quad (18)$$

where K is a pre-specified integer,

$$\begin{aligned} Q_{1,k}(\lambda) &= \frac{(2\pi k - \lambda)^{1-2H}}{(\Delta\kappa)^2 + (2\pi k - \lambda)^2}, Q_{2,k}(\lambda) = \frac{(\lambda + 2\pi k)^{1-2H}}{(\Delta\kappa)^2 + (\lambda + 2\pi k)^2}, \\ a(K, H, \lambda) &\equiv \frac{1}{4\pi} [2\pi(K+1) - \lambda]^{-2H} \left\{ \frac{1}{H} - \frac{(\Delta\kappa)^2}{(1+H)[2\pi(K+1) - \lambda]^2} \right\}, \\ b(K, H, \lambda) &\equiv \frac{1}{4\pi} [2\pi(K+1) + \lambda]^{-2H} \left\{ \frac{1}{H} - \frac{(\Delta\kappa)^2}{(1+H)[2\pi(K+1) + \lambda]^2} \right\}, \\ c(K, H, \lambda) &\equiv \frac{1}{8\pi H} (2\pi K - \lambda)^{-2H} + \frac{1}{8\pi H} (2\pi K + \lambda)^{-2H}. \end{aligned}$$

We set $K = 200$ in the application.

The approximated Whittle log likelihood function is given by

$$l_{AW}(\beta; \mathbf{y}) = -\frac{1}{2\pi} \int_0^\pi \left(\ln \tilde{f}_y^\Delta(\lambda) + \frac{I(\lambda)}{\tilde{f}_y^\Delta(\lambda)} \right) d\lambda, \quad (19)$$

where $I(\lambda)$ denotes the periodogram at frequency λ computed as

$$I(\lambda_j) = \frac{1}{2\pi T} \left| \sum_{s=1}^T (y_{s\Delta} - \hat{\mu}) \exp(-is\lambda_j) \right|^2, \quad (20)$$

where $\hat{\mu}$ is obtained as the sample mean. The spectral density diverges to infinity at the near zero frequencies (i.e., $\lambda \rightarrow 0$) when $H > 1/2$. To address the singularity problem, and given that σ is functionally independent of κ and H , the objective function can be reformulated as the following:

$$\begin{aligned} l_{AW}^+(\kappa, H; \mathbf{y}) = & - \left\{ \log \left[\frac{1}{\pi} \int_0^\pi \frac{I_n(\lambda)}{\tilde{g}_y^\Delta(\lambda; H, \kappa)} d\lambda \right] + \frac{1}{\pi} \int_0^\pi \ln \tilde{g}_y^\Delta(\lambda; H, \kappa) d\lambda \right\}, \text{ if } H \leq 1/2, \\ l_{AW}^+(\kappa, H; \mathbf{y}) = & - \left\{ \log \left[\frac{1}{\pi} \left(\int_\varepsilon^\pi \frac{I_n(\lambda)}{\tilde{g}_y^\Delta(\lambda; H, \kappa)} d\lambda + B^\ddagger(\kappa, H, \varepsilon) \right) \right] + \frac{1}{\pi} \left(\int_\varepsilon^\pi \ln \tilde{g}_y^\Delta(\lambda; H, \kappa) d\lambda + B^\dagger(\kappa, H, \varepsilon) \right) \right\}, \text{ if } H > 1/2, \end{aligned}$$

where

$$\tilde{g}_y^\Delta(\lambda; H, \kappa) = \tilde{f}_y^\Delta(\lambda; H, \kappa) / \sigma^2,$$

$$B^\dagger(\kappa, H, \varepsilon) = \ln \left(\frac{C_H \Delta^{2H}}{2\pi(\Delta\kappa)^2} \right) \varepsilon + \varepsilon(\ln \varepsilon - 1)(1 - 2H),$$

$$B^\ddagger(\kappa, H, \varepsilon) = \delta^*(0, \varepsilon) \hat{\gamma}_n(0) + \sum_{\tau=1}^{n-1} \delta^*(\tau, \varepsilon) \hat{\gamma}_n(\tau),$$

with $\hat{\gamma}_n(\tau) = n^{-1} \sum_{j=1}^{n-\tau} y_{j\Delta} y_{j-\tau\Delta}$, $\delta^*(0, \varepsilon) = \frac{\Delta^{2-2H} \kappa^2 \varepsilon^{2H}}{2HC_H}$,

$$\delta^*(\tau, \varepsilon) = \frac{\Delta^{2-2H} \kappa^2}{C_H} \sum_{j=0}^J \frac{(-1)^j \tau^{2j}}{(2j)!} \frac{\varepsilon^{2(j+H)}}{2(j+H)},$$

and J being a large number. See [Shi et al. \(2024\)](#) for details.

The AWML estimator is computed as

$$(\hat{\kappa}_{AW}, \hat{H}_{AW}) = \arg \max_{\kappa, H} l_{AW}^+(\kappa, H; \mathbf{y}).$$

The estimator is shown to be efficient in Fisher's sense ([Shi et al., 2024](#)). To estimate μ , [Shi et al. \(2024\)](#) use the sample mean of $y_{t\Delta}$.³ The volatility parameter $\hat{\sigma}_{AW}^2$ is estimated by

$$\hat{\sigma}_{AW}^2 = \frac{1}{\pi} \int_{\varepsilon}^{\pi} \frac{I_n(\lambda)}{g_y^\Delta(\lambda; \hat{H}_{AW})} d\lambda \quad \text{if } H < 1/2,$$

and

$$\hat{\sigma}_{AW}^2 = \frac{1}{\pi} \left(\int_{\varepsilon}^{\pi} \frac{I_n(\lambda)}{g_y^\Delta(\lambda; \hat{H}_{AW})} d\lambda + B^\ddagger(\hat{H}_{AW}, \varepsilon) \right) \quad \text{if } H \geq 1/2.$$

2.3 Brownian Semi-Stationary process

The Brownian Semi-Stationary (BSS) process was introduced to model volatility by [Bennedsen et al. \(2022\)](#). The process is defined as the Gaussian moving averages:

$$y_t = \mu + \sigma \int_{-\infty}^t g(t-s) dB(s),$$

³We utilize a continuous time specification of the Whittle log likelihood function. An unreported analysis indicates that when μ is known, the estimator obtained by maximizing the discrete Whittle log likelihood function performs slightly worse than the one from the continuous version. However, when μ is unknown and estimated using the sample mean, the performance of both estimators is nearly identical.

where $B(s)$ is Brownian motion on $\mathbb{R}$ and $g : (0, \infty) \rightarrow \mathbb{R}$ is a square-integrable kernel function. [Bennedsen et al. \(2022\)](#) give two specific examples: Power-BSS process and Gamma-BSS process.

The kernel function of the power-BSS process takes the form of the following:

$$g(v) = v^\alpha (1+v)^{-\alpha-\gamma}, \quad t > s,$$

with $\alpha \in (-\frac{1}{2}, \frac{1}{2})$ and $\gamma > \frac{1}{2}$. The parameter α controls the roughness of the process, while γ controls the long memory of the process. Therefore, it enables the decoupling of roughness and long memory, offering a more general framework than the fOU and ARFIMA models. See Proposition 1.2 of [Bennedsen et al. \(2022\)](#) for explanations. The autocovariance of $y_{i\Delta}$ is given by

$$\text{Cov}(y_{i\Delta}, y_{(i-j)\Delta}) = \sigma^2 \int_{-\infty}^{i\Delta} g(i\Delta - s)g(i\Delta - s + j\Delta)ds, \quad \text{for } j > 0, \quad (21)$$

$$\text{Var}(y_{i\Delta}) = \sigma^2 B(2\alpha + 1, 2\gamma - 1), \quad (22)$$

where $B(x_1, x_2) = \int_0^1 t^{x_1-1} (1-t)^{x_2-1} dt$ is the beta-function.

The kernel function of the Gamma-BSS process takes the form of

$$g(v) = v^{\tilde{\alpha}} e^{-\tilde{\gamma}v}, \quad t > s,$$

with $\tilde{\alpha} \in (-\frac{1}{2}, \frac{1}{2})$ and $\tilde{\gamma} > 0$. The parameter $\tilde{\alpha}$ controls the roughness of the process and $\tilde{\gamma}$ controls the persistence of the process. See Proposition 1.3 of [Bennedsen et al. \(2022\)](#) for explanations. The autocovariance is given by

$$\text{Cov}(y_{i\Delta}, y_{(i-j)\Delta}) = \sigma^2 \frac{\Gamma(\tilde{\alpha} + 1)}{\sqrt{\pi}} \left(\frac{j\Delta}{2\tilde{\gamma}} \right)^{\tilde{\alpha}+1/2} K_{\tilde{\alpha}+1/2}(\tilde{\gamma}j\Delta), \quad \text{for } j > 0, \quad (23)$$

where $K_\nu(\cdot)$ is the modified Bessel function of the third kind with index ν (see, for example [Gradshteyn and Ryzhik, 2014](#), Section 8.4). The variance of $y_{i\Delta}$ is given by

$$\text{Var}(y_{i\Delta}) = \sigma^2 (2\tilde{\gamma})^{-2\tilde{\alpha}-1} \Gamma(2\tilde{\alpha} + 1). \quad (24)$$

Let $\delta = (\alpha, \gamma)$ for the power-BSS process and $\delta = (\tilde{\alpha}, \tilde{\gamma})$ for the Gamma-BSS process. Denote the autocorrelation function (ACF) of y_t implied by the model as $\rho(h; \delta) := \text{Corr}(y_{i\Delta}, y_{i\Delta-h})$. The BSS process is considered rough if its ACF satisfies the following asymptotic relationship:

$$1 - \rho(h; \delta) \sim c|h|^{2\alpha+1}, \quad |h| \rightarrow 0, \quad (25)$$

where $c > 0$ is a constant, and the roughness parameter $\alpha \in (-\frac{1}{2}, \frac{1}{2})$. The process has long memory if:

$$\rho(h; \delta) \sim c|h|^{-\beta}, \quad |h| \rightarrow \infty,$$

for some $\beta \in (0, 1)$. Both the power-BSS and Gamma-BSS processes satisfy these two conditions. For further details, see [Bennedsen et al. \(2017\)](#).

Estimation Methods

[Barndorff-Nielsen et al. \(2013\)](#) examine the CoF estimator for Brownian semi-stationary (BSS) processes when $H \in (0, 1/2) \cup (1/2, 3/4)$. [Corcuera et al. \(2013\)](#) introduce a modified CoF estimator suitable for BSS processes with $H \in (3/4, 1)$. [Bennedsen et al. \(2022\)](#) employ a two-stage procedure (including nonlinear least squares (NLS) and the method of moments (MM)) to estimate parameters, concluding that BSS models outperform ARFIMA and HAR models and are strong competitors to the fBm model. However, the comparison between the BSS model and the fBm may be flawed due to the use of sub-optimal forecasting formulas for the fBm model. In an earlier version, [Bennedsen et al. \(2017\)](#) apply the MCL method to BSS models for modeling realized volatilities, with simulation results indicating that MCL for the BSS process outperforms the CoF method.

Method of Moments The mean parameter μ is estimated using the sample mean. For the remaining model parameters, [Bennedsen et al. \(2022\)](#) propose a two-step procedure. First, the roughness parameter α or $\tilde{\alpha}$ is estimated semi-parametrically based on the scaling relationship in (25). Then, the remaining parameters are estimated using a parametric method of moments.

Let $(\delta_1 = \alpha, \delta_2 = \gamma)$ for the power-BSS process and $(\delta_1 = \tilde{\alpha}, \delta_2 = \tilde{\gamma})$ for the Gamma-BSS process. From (25), the second-order variogram, denoted by $\gamma_2(\cdot)$, satisfies:

$$\gamma_2(h) := E \left[(y_{i\Delta} - y_{i\Delta-h})^2 \right] \sim c|h|^{2\delta_1+1}, \quad |h| \rightarrow 0.$$

This motivates OLS regression of the form:

$$\ln \hat{\gamma}_2(j\Delta) = a_0 + a_1 \ln(j\Delta) + \varepsilon_j, \quad j = 1, 2, \dots, m$$

where $m \in \mathbb{N}$ is a bandwidth parameter, ε_j is an error term, and $\hat{\gamma}_2(j\Delta) = \frac{1}{n-j} \sum_{i=1}^{n-j} (y_{i\Delta} - y_{(i-j)\Delta})^2$ is the empirical variogram corresponding to $\gamma_2(h)$. The OLS estimate of δ_1 is then given by $\hat{\delta}_1 = (\hat{a}_1 - 1)/2$, where $\hat{a}_1$ is the OLS estimator of a_1 from the regression. Following [Bennedsen et al. \(2022\)](#), the bandwidth parameter is set to $m = 6$.

The second step estimates the long memory parameter δ_2 based on the parametric ACF. The empirically estimated ACF, $y_{i\Delta}$, is fitted to the model implied parametric ACF $\rho(h; \delta)$, using the first-step estimate of α . The estimator of δ_2 is defined as:

$$\hat{\delta}_2 = \arg \min_{\delta_2} \sum_{k=1}^K \left[\hat{\rho}(j\Delta) - \rho(j\Delta; \hat{\delta}_1, \delta_2) \right]^2,$$

where $K = \lfloor T^{1/3} \rfloor$ and $\hat{\delta}_1$ is the first-step estimator of α . The parameter σ is estimated as the sample standard deviation of $y_{i\Delta}$ divided by $\rho(j\Delta; \hat{\delta}_1, \hat{\delta}_2)$. [Bennedsen et al. \(2022\)](#) show the asymptotic consistency of $\hat{\delta}_1$ and $\hat{\delta}_2$.

MCL Despite its asymptotic efficiency, the TDML method can be computationally intensive, which makes it unsuitable for large dimensional datasets. [Bennedsen et al. \(2024\)](#) proposes the MCL method, which aims to reduce computational cost while retaining some of the favorable theoretical properties of MLE. Let $\theta_{BSS} = (\mu, \sigma, \delta)$. The log likelihood function of the MCL is given by:

$$l^c(\theta_{BSS}; \mathbf{y}) = \sum_{j=1}^K \sum_{i=1}^{T-j} \ln \omega(y_{(i+j)\Delta}, y_{i\Delta}; \theta_{BSS}),$$

where $\omega(y_{(i+j)\Delta}, y_{i\Delta}; \theta_{BSS})$ is the pairwise joint probability density function (pdf) of $(y_{(i+j)\Delta}, y_{i\Delta})$. Since $(y_{(i+j)\Delta}, y_{i\Delta})$ follows a bivariate normal distribution, its log-density function is given by:

$$\ln \omega(y_{(i+j)\Delta}, y_{i\Delta}; \theta_{BSS}) = -\frac{1}{2} \ln \det(\Sigma_{\mathbf{z}}) - \log(2\pi) - \frac{1}{2} \mathbf{z}^\top \Sigma_{\mathbf{z}}^{-1} \mathbf{z},$$

where $\mathbf{z} = (y_{(i+j)\Delta} - \mu, y_{i\Delta} - \mu)^\top$ and $\Sigma_{\mathbf{z}}$ is the covariance matrix of $\mathbf{z}$. We set $K = 5$ and replace μ by the sample mean of the observations in applications. The computation of $\Sigma_{\mathbf{z}}$ is based on the variance and covariance formulas provided in (21) -(24). The indefinite integral in the covariance of the power-BSS is again evaluated by numerical integral using *quadgk* in Matlab. The MCL estimator is defined as

$$\hat{\theta}_{MCL}^{BSS} = \arg \max_{\theta_{BSS}} l^c(\theta_{BSS}; \mathbf{y}).$$

3 Discrete time Models

We consider two types of discrete time models: the ARFIMA(1, d , 0) model and HAR-type models. In contrast to continuous time models, where $t \in \mathbb{R}^+$ represents continuous time and $\{i\Delta\}_{i=1}^T$ indicates discrete observations, we use t to denote discrete time points for our discrete time models, i.e., $t = 1, 2, \dots, T$.

3.1 ARFIMA(1, d , 0)

For notation simplicity, we refer to the ARFIMA(1, d , 0) model as FAR(α, d) subsequently. The FAR(α, d) model is specified as

$$(1 - \alpha L)(y_t - \mu) = \sigma(1 - L)^{-d} \varepsilon_t \text{ with } |\alpha| < 1, \quad (26)$$

where L is the lag operator, d is the memory parameter, and $\varepsilon_t \sim_{iid} N(0, 1)$. It reduces to a standard autoregressive process when $d = 0$.

When $d \in (-1/2, 1/2)$, the process is stationary (Bloomfield, 1985). Let $u_t = (1 - L)^{-d} \varepsilon_t$ be the fractionally integrated process. Let $\gamma_u(j) = \text{Cov}(u_t, u_{t-j})$ be the j th order autocovariance of u_t . According to Hosking (1981), the autocovariance function of u_t is

$$\gamma_u(j) = \frac{(-1)^j \Gamma(1 - 2d)}{\Gamma(j - d + 1) \Gamma(1 - j - d)}. \quad (27)$$

The long run variance covariance $\sum_{j=-\infty}^{\infty} \gamma_u(j) \rightarrow \infty$ when $d \in (0, 1/2)$ and $\sum_{j=-\infty}^{\infty} \gamma_u(j) = 0$ when $d \in (-1/2, 0)$. Therefore, u_t has a long memory if $d \in (0, 1/2)$ and is anti-persistent if $d \in (-1/2, 0)$. The covariance function for the stationary process can be written as (Brockwell and Davis, 2009):

$$\gamma_y(k) = \sum_{s=-\infty}^{\infty} \tilde{\gamma}(s) \gamma_u(k - s), \quad (28)$$

where $\tilde{\gamma}(s)$ is the autocovariance of the pure AR component, i.e., $\tilde{\gamma}(s) = \sigma^2 \alpha^s / (1 - \alpha^2)$ for $s > 0$.⁴

When $d \in (1/2, 3/2)$ in FAR(α, d), y_t is nonstationary. Multiplying both sides of (26) by $1 - L$, we obtain

$$(1 - L)(1 - \alpha L)(y_t - \mu) = \sigma(1 - L)^{-(d-1)} \varepsilon_t \text{ with } |\alpha| < 1,$$

which leads to

$$(1 - \alpha L)(y_t - y_{t-1}) = \sigma(1 - L)^{-d^*} \varepsilon_t \text{ with } |\alpha| < 1 \text{ and } d^* = d - 1 \in (-0.5, 0.5). \quad (29)$$

Hence, FAR(α, d) with $d \in (0.5, 1.5)$ for y_t is equivalent to FAR(α, d^*) with $d^* \in (-0.5, 0.5)$ for $y_t - y_{t-1}$. In this case, although y_t is nonstationary, $y_t - y_{t-1}$ is stationary. Interestingly, μ is superfluous in the latter model.

⁴In practice, the summation is truncated at value S . We follow the rule recommended by Shi and Yu (2023) for the selection of S to ensure accurate estimation of the variance-covariance matrix particularly for processes with an autoregressive root close to unity. Specifically, for $\alpha \leq 0.9$, $S = 200$ is sufficient, but larger values are needed as α increases: $S = 300$ for $0.9 < \alpha \leq 0.95$, $S = 1,700$ for $0.95 < \alpha \leq 0.99$, $S = 3,000$ for $0.99 < \alpha \leq 0.995$, and $S = 7,000$ for $0.995 < \alpha < 1$.

Model Specifications and Estimation

Several estimation techniques have been proposed, including semi-parametric methods, the TDML method, and the EWML methods. Semi-parametric approaches, such as the local Whittle method (Künsch, 1987; Robinson, 1995a) and the log periodogram method (Geweke and Porter-Hudak, 1983; Robinson, 1995b), leverage features of the spectral density at low frequencies, which is robust to short-run dynamics of the data series asymptotically. The autoregressive coefficient α can be estimated from pre-filtered data based on the estimated d , and μ can be estimated by the sample mean. Shi and Yu (2023) point out that semi-parametric methods struggle to differentiate between persistence arising from the autoregressive coefficient and that from the fractional parameter when the autoregressive coefficient is close to unity.

The parametric ML method generally performs well when α is far away from unity and zero. However, if α is near unity or zero, there is a weak identification problem documented in Shi and Yu (2023) and Li et al. (2025). The weak identification problem arises because the spectral density of this model is asymptotically indistinguishable in two key regions: (1) as $\alpha \rightarrow 1$ with $d = d^* \in (-1/2, 0)$ (which is a stationary model, representing rough dynamics), and (2) as $\alpha \rightarrow 0$ with $d = 1 + d^* \in (1/2, 1)$ (which is a nonstationary model, indicating long memory). The autoregressive coefficient $\alpha \rightarrow 1$ and the fractional parameter $d > 0$ both play a similar role in capturing the slow decay pattern of ACF.

We consider three specifications of $\text{FAR}(\alpha, d)$. First, we assume the process is stationary and exhibits long memory, with $d \in [0, 0.5)$, as in the traditional long memory literature (see, e.g., Granger and Joyeux, 1980; Granger, 1980; Hosking, 1981), and estimate it using the modified profile likelihood (MPL) method that will be reviewed later (labeled *ARFIMA-LM-S (MPL)*). Second, we maintain the stationarity assumption but allow the process to exhibit rough behavior, with $d \in (-0.5, 0)$, also estimated using MPL (labeled *ARFIMA-R-S (MPL)*). Third, we consider a long memory process that may be nonstationary, with $d \in (0, 1)$, and estimate it using the Whittle maximum likelihood method that will be reviewed later (labeled *ARFIMA-LM-NS (Whittle)*).

MPL Let $\theta_{\text{FAR}} = (\mu, \sigma, \delta^\dagger)$ with $\delta^\dagger = (\alpha, d)$. Under the model specification, $\mathbf{y} - \mu\mathbf{1}$ follows a normal distribution with mean zero and variance-covariance matrix $\sigma^2 \Sigma_{\mathbf{y}}$. The objective function of the ML estimator is as in (10). Elements of the variance covariance matrix $\sigma^2 \Sigma_{\mathbf{y}}$ is computed from (28), where the summand is truncated at S . In case of unknown μ , one could use the plug-in method, which substitutes μ by a consistent estimator of the mean (e.g., the sample mean). Although the method provides a $\sqrt{T}$ consistent and asymptotically normal estimator, it is contaminated by an additional second-order negative bias (Lieberman, 2005) due to the need of estimating μ .

The MPL estimator of Cox and Reid (1987) is obtained by modifying the standard profile likelihood function to address the second-order bias. The modification introduces an adjustment term derived from the information

matrix of the nuisance parameters, which penalizes the likelihood heavier when the plug-in estimator of these parameters is noisier. This adjustment reduces the second-order bias while preserving asymptotic efficiency (An and Bloomfield, 1993; Hauser, 1999). The resulting modified profile likelihood function is

$$l_M(\delta^\dagger; \mathbf{y}) = \left(\frac{1}{T} - \frac{1}{2} \right) \ln |\Sigma_{\mathbf{y}}| - \frac{1}{2} \ln \left(\mathbf{1}^\top \Sigma_{\mathbf{y}}^{-1} \mathbf{1} \right) + \frac{3-T}{2} \ln \left[T^{-1} (\mathbf{y} - \hat{\mu})^\top \Sigma_{\mathbf{y}}^{-1} (\mathbf{y} - \hat{\mu}) \right], \quad (30)$$

where $\hat{\mu} = \frac{\mathbf{1}^\top \Sigma_{\mathbf{y}}^{-1} \mathbf{y}}{\mathbf{1}^\top \Sigma_{\mathbf{y}}^{-1} \mathbf{1}}$. The MPL estimator is denoted by $\hat{\delta}_{MPL}^\dagger$ and defined as

$$\hat{\delta}_{MPL}^\dagger = \arg \max_{\delta^\dagger} l_M(\delta^\dagger; \mathbf{y}).$$

The parameter μ is estimated by

$$\hat{\mu}_{MPL} = \frac{\mathbf{1}^\top \hat{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y}}{\mathbf{1}^\top \hat{\Sigma}_{\mathbf{y}}^{-1} \mathbf{1}} \quad \text{with} \quad \hat{\Sigma}_{\mathbf{y}} = \Sigma_{\mathbf{y}}(\hat{\delta}_{MPL}^\dagger),$$

while the variance σ^2 is estimated as

$$\hat{\sigma}_{MPL}^2 = \frac{1}{T} (\mathbf{y} - \hat{\mu}_{MPL} \mathbf{1})^\top \Sigma_{\mathbf{y}}^{-1} (\mathbf{y} - \hat{\mu}_{MPL} \mathbf{1}).$$

MPL requires that the process be stationary, ensuring the variance-covariance matrix remains time-invariant.

Whittle MLE Let $\delta^\dagger = (\alpha, d)$. The spectral density of $\text{FAR}(\alpha, d)$ is given by

$$f_y(\lambda; \delta^\dagger, \sigma) = \frac{\sigma^2}{2\pi} \frac{(2 - 2\cos(\lambda))^{-d}}{1 - 2\alpha \cos(\lambda) + \alpha^2} \quad \text{for } -\pi \leq \lambda \leq \pi. \quad (31)$$

When the data series is nonstationary (i.e., $d > 0.5$), $f_y(\lambda; \delta^\dagger, \sigma)$ is not integrable over the interval $[-\pi, \pi]$. However, it can be interpreted as the limit of the expected sample periodogram (Hurvich and Ray, 1995; Velasco and Robinson, 2000). Consequently, it is referred to as a ‘pseudo spectral density’. The standard Whittle ML method is shown to yield $\sqrt{T}$ -consistent, asymptotically normal, and efficient parameter estimates when $d \in (0, 1/2)$ (Hannan, 1973; Fox and Taqqu, 1986; Giraitis and Surgailis, 1990). Velasco and Robinson (2000) proposes using the Whittle MLE with tapering to account for potential nonstationarity, i.e., $d \in [1/2, 1)$ in the data series.

The parameter μ is estimated by the sample mean $\hat{\mu}$. Define the tapered periodogram by

$$I^p(\lambda_j) = \frac{1}{2\pi \sum_{t=1}^T h_t^2} \left| \sum_{t=1}^T h_t (y_t - \hat{\mu}) \exp(-it\lambda_j) \right|^2,$$

where h_t is the data taper series. Let $f_y(\lambda; \delta^\dagger, \sigma) = \frac{\sigma^2}{2\pi} \eta_y(\lambda; \delta^\dagger)$. The profiled Whittle objective function is given by:

$$l_W^p(\delta^\dagger; \mathbf{y}) = \frac{2\pi p}{T} \sum_{j \in J(p)} \frac{I^p(\lambda_j)}{\eta_y^*(\lambda_j; \delta^\dagger)}. \quad (32)$$

where $J(p) = \{p, 2p, \dots, T-p\}$ (assuming that T/p is an integer) and $\eta_y^*(\lambda_j; \delta^\dagger) = \eta_y(\lambda; \delta^\dagger) / Z$ with $Z := \exp\left(\frac{p}{T} \sum_{j \in J(p)} \ln \eta_y(\lambda_j; \delta^\dagger)\right)$. Note that $\eta_y^*(\lambda_j; \delta^\dagger)$ is the normalized spectral density which satisfies $\int_{-\pi}^{\pi} \ln \eta_y^*(\lambda; \delta^\dagger) d\lambda = 0$. See the appendix for the derivation of the profiled Whittle objective function. The tapered Whittle estimator is

$$\hat{\delta}_W^{\dagger p} = \arg \min_{\theta} l_W^p(\delta^\dagger; \mathbf{y}) \text{ and } (\hat{\sigma}_W^p)^2 = \frac{2\pi p}{T} \sum_{j \in J(p)} \frac{I^p(\lambda_j)}{\eta_y(\lambda_j; \hat{\delta}_W^{\dagger p})}.$$

Under certain regularity conditions, the tapered Whittle estimator is shown to have asymptotic normality for $d \in (-0.5, 1)$:

$$\sqrt{m} \left(\hat{\delta}_W^{\dagger p} - \delta_0^\dagger \right) \rightarrow N(0, 4\pi p \Phi_p \Sigma_0^{-1}),$$

where $\delta_0^\dagger$ is the true model parameter,

$$\Sigma_0 = \int_{-\pi}^{\pi} \left\{ \frac{\partial}{\partial \delta^\dagger} \ln \eta(\lambda; \delta_0^\dagger) \right\} \left\{ \frac{\partial}{\partial \delta^\dagger'} \ln \eta(\lambda; \delta_0^\dagger) \right\} d\lambda,$$

and $\Phi_p = \lim_{n \rightarrow \infty} \sum_{k \in J(p)} \phi^2(\lambda_k)$ is the taper variance inflator, with $\phi(\lambda) = (\sum_{t=1}^n h_t^2)^{-1} (\sum_{t=1}^n h_t^2 \cos(t\lambda))$. The estimation method includes the standard Whittle method as a special case with $h_t = 1$ and $p = 1$, leading to a taper inflator of $\Phi_p = 1$ and an asymptotic variance of $4\pi \Sigma_0^{-1}$. In our implementation, the taper series and h_t is assume to be the Kolmogorov weights ([Žurbenko, 1979](#)) with order $p = 3$, which takes the form of the full cosine bell $h_t = 0.5 \left(1 - \cos \frac{2\pi t}{n}\right)$. The parameter μ is estimated by its sample mean.

3.2 HAR-type models

HAR-type models have gained significant attention in financial econometrics for their ability to capture the dynamics of volatility and also for their simplicity in estimation and forecasting. This section introduces three models in this category: the HAR model [Corsi \(2009\)](#), the HARQ model of [Bollerslev et al. \(2016\)](#), and the semi-variance HAR (HAR-SV) model of [Patton and Sheppard \(2015\)](#).

The HAR model proposed by [Corsi \(2009\)](#) for RV serves as a widely used benchmark for forecasting. This model effectively captures the long memory dynamics of volatility through a straightforward linear specification,

incorporating lagged daily, weekly, and monthly average RV as regressors. Specifically,

$$RV_{t+h} = \beta_0 + \beta_1 RV_t + \beta_2 RV_{t|t-4} + \beta_3 RV_{t|t-21} + \varepsilon_{t+h}, \quad (33)$$

where $RV_{t-j|t-k} = \frac{1}{k-j+1} \sum_{i=j}^k RV_{t-i}$ with $j \leq k$ and ε_t is a disturbance term. The quantity $RV_{t|t-4}$ and $RV_{t|t-21}$ represent the weekly and monthly average RV, respectively. The HAR-RV specification can be viewed as a constrained version of an AR(21) model and its parameters can be estimated using the OLS method.

To take the measure errors of RV into consideration, [Bollerslev et al. \(2016\)](#) proposes to directly adjust the daily coefficient based on the magnitude of its measurement error. The HARQ model proposed by [Bollerslev et al. \(2016\)](#) is specified as follows:

$$RV_{t+h} = \beta_0 + \left(\beta_1 + \beta_{1Q} \sqrt{RQ_t} \right) RV_t + \beta_2 RV_{t|t-4} + \beta_3 RV_{t|t-21} + \varepsilon_{t+h}, \quad (34)$$

where RQ is the realized quardicity which serves as a proxy for the measure error of RV. The measurement error variance for the weekly and monthly (normalized) realized volatility is much smaller, so these coefficients are kept constant. The realized quarticity is calculated from intraday observations as: $RQ = \frac{M}{3} \sum_{i=1}^M r_{t,i}^4$, where M is the number of observations within the day, and $r_{t,i}$ represents the i th return on day t . The HARQ is found to significant improvements in the accuracy of the resulting forecasts compared to the forecasts from other HAR-type models ([Bollerslev et al., 2016](#))

The HAR-SV model extends the traditional HAR framework by incorporating realized semivariances to capture the asymmetric impact of positive and negative returns on future volatility. The realized semivariances are defined as

$$RS_t^+ = \sum_{i=1}^n r_{i,t}^2 \mathbb{I}\{r_{i,t} > 0\}, \quad RS_t^- = \sum_{i=1}^n r_{i,t}^2 \mathbb{I}\{r_{i,t} < 0\},$$

where $r_{i,t}$ denotes the intraday return at time i on day t with n being the number of observations within the day, and $\mathbb{I}\{\cdot\}$ is the indicator function. The forecasting equation for the h -step ahead realized variance is specified as

$$RV_{t+h} = \beta_0 + \beta_1^+ RS_t^+ + \beta_1^- RS_t^- + \beta_2 RV_{t|t-4} + \beta_3 RV_{t|t-21} + \varepsilon_{t+h}.$$

To address non-normality for the distribution of RV, we apply the HAR, HARQ, and HAR-SV models to the log RV. That is, we replace RV with log RV, and substitute $\log RSV^+$ and $\log RSV^-$ for RSV^+ and RSV^- , respectively, in each HAR-type model. We refer to the original specifications as RV-HAR, RV-HARQ, and RV-HAR-SV, and denote their log counterparts as logRV-HAR, logRV-HARQ, and logRV-HAR-SV.

4 Forecasting Method

We begin by discussing forecasting methods for continuous time models. In the seminal paper, [Gatheral et al. \(2018\)](#) forecast future realized volatility (RV) by discretizing an infinite-past forecasting formula derived by [Nuzman and Poor \(2000\)](#). Several subsequent studies ([Gatheral et al., 2018](#); [Bennedsen et al., 2017](#); [Wang et al., 2023](#)) have followed this approach when applying the fractional Brownian motion (fBm) model to RV forecasting. Theoretically, the infinite-past forecasting formula is optimal when a continuous, infinite record is available. However, when applied to empirical data, it fails to yield optimal forecasts due to the discrete and finite nature of observed samples. With discrete time observations and assuming the variable follows a Gaussian process, the optimal forecast of the target variable (in term of minimizing root mean squared error (RMSE)) is given by its conditional expectation.⁵ The conditional expectation approach for forecast is applicable to both the continuous time models and the FAR(α, d) model.

4.1 Optimal Forecast

Let z_t be a generic process of interest. If y_t is a stationary process (e.g., fOU, BSS, or FAR model with $d \in (0, 1/2)$), then

$$z_t = y_t - \mu.$$

Conversely, if y_t is nonstationary (e.g., fBm or FAR with $d \in [1/2, 1)$),

$$z_t = y_t - y_{t-1}.$$

The h -step-ahead optimal prediction of z_T is achieved at the conditional mean since z_t is Gaussian. Let $\mathbf{z} = (z_1, \dots, z_T)'$. The optimal forecast $\hat{z}_{T+h} = \mathbb{E}(z_{T+h}|\mathbf{z})$, which is a linear combination of the past observations, i.e.,

$$\hat{z}_{T+h} = \phi^{(h)'} \mathbf{z} \text{ with } \phi^{(h)} = \Sigma_{\mathbf{z}}^{-1} \Gamma_h,$$

where $\phi^{(h)}$ is the optimal weight assigned to each past observations with $\Gamma_h = (\gamma_z(T+h-1), \dots, \gamma_z(h))'$, $\gamma_z(j) = \text{Cov}(z_{i\Delta}, z_{(i-j)\Delta})$, and $\Sigma_{\mathbf{z}}$ being the $T \times T$ dimension variance-covariance matrix of $\mathbf{z}$. The theoretical mean squared prediction errors of z_{T+h} is given by

$$\mathbb{E} \left[(z_{T+h} - \hat{z}_{T+h})^2 \right] = \gamma_z(0) + \Gamma_h' \Sigma_{\mathbf{z}}^{-1} \Gamma_h.$$

⁵[Wang et al. \(2024\)](#) assess the efficiency loss of the infinite-past forecasting approach against the conditional expectation method using RMSE. They find that [Gatheral et al. \(2018\)](#)'s method underestimates the fBm model's performance by 5% for both simulated data and daily realized volatility.

For stationary processes,

$$\hat{y}_{(T+h)\Delta} = \hat{\mu} + \hat{z}_{(T+h)\Delta},$$

whereas for nonstationary processes,

$$\hat{y}_{T+h} = y_T + \sum_{j=1}^h \hat{z}_{T+j}.$$

Consequently, for nonstationary processes, the mean squared prediction error of $y_{(T+h)\Delta}$ is the following:

$$\mathbb{E} \left[(y_{T+h} - \hat{y}_{T+h})^2 \right] = \sum_{j=1}^h \mathbb{E} \left[(z_{T+j} - \hat{z}_{T+j})^2 \right] = \mathbf{1}_h' \Sigma_z \mathbf{1}_h - \left(\sum_{j=1}^h \Gamma_j' \right) \Sigma_z^{-1} \left(\sum_{j=1}^h \Gamma_j \right).$$

When y_t represents log RV, which is assumed to follow a Gaussian distribution, RV follows a lognormal distribution. Therefore, the h -step-ahead forecast for the RV is given by

$$\widehat{RV}_{T+h} = \exp \left(\hat{y}_{T+h} + \frac{1}{2} \widehat{CV}_{T+h} \right), \quad (35)$$

where $\widehat{CV}_{T+h} = \gamma_{z,0} + \Gamma_h' \Sigma_z^{-1} \Gamma_h$ for stationary processes and $\widehat{CV}_{T+h} = \mathbf{1}_h' \Sigma_z \mathbf{1}_h - \left(\sum_{j=1}^h \Gamma_j' \right) \Sigma_z^{-1} \left(\sum_{j=1}^h \Gamma_j \right)$ for nonstationary processes.

For the HAR-type models, the h -step-ahead forecast of RV is simply the fitted value of the regression model with RV_{t+h} as the dependent variable. For the logRV models, we apply the same transformation as in equation (35) to convert log forecasts back to the RV scale. The HAR and HARQ model can sometimes generate implausibly large or small forecasts. To mitigate this issue, [Bollerslev et al. \(2016\)](#) implement an ‘insanity filter’ for all forecasts. Specifically, any forecast that exceeds the maximum or falls below the minimum of the dependent variable observed during the estimation period is replaced with the sample average from that period. We adopt this approach in our analysis.

4.2 Performance Evaluation

We perform a rolling forecasting exercise for RV using a five-year window, resulting in $T_0 = 251 \times 5 = 1,255$ data points. Three forecasting horizons (h) are considered: $h = 1$, $h = 5$, and $h = 21$.

Statistical Measures The loss functions used in the analysis include squared forecast error (SFE) and QLIKE given by

$$SFE : L_{t,h}^S = \left(\widehat{RV}_{t+h} - RV_{t+h} \right)^2,$$

$$QLIKE : L_{t,h}^Q = \ln(\widehat{RV}_{t+h}) + RV_{t+h}/\widehat{RV}_{t+h},$$

where $t = T_0, \dots, T - h$. [Patton \(2011\)](#) shows that unlike absolute forecast errors, SFE and QLIKE yield inferences that are invariant to the choice of units.

We use the model confidence set (MCS) approach by [Hansen et al. \(2011\)](#) to assess the statistical differences between competing models. The MCS provides a confidence set containing the best models with a probability greater than or equal to a specified level (e.g., 25%). It also assigns a p-value to each individual model, allowing for a comprehensive evaluation of their statistical significance.

Suppose we have a set of competing models indexed by $\mathbb{M}_0 = \{i = 1, \dots, M\}$. The loss function can be either squared error, absolute error, or QLIKE. We calculate the relative performance as $d_{ij,t} = L_{i,t} - L_{j,t}$ for all $i, j \in \mathbb{M}_0$. The MCS procedure involves an iterative process to identify the best model set. For iteration s , the null and alternative hypotheses are as follows:

$$H_{0,\mathbb{M}_s} : \mathbb{E}(d_{ij,t}) = 0 \text{ for all } i, j \in \mathbb{M}_s \subset \mathbb{M}_0,$$

and the alternative

$$H_{A,\mathbb{M}_s} : \mathbb{E}(d_{ij,t}) \neq 0 \text{ for some } i, j \in \mathbb{M}_s.$$

We perform a model equivalence test using the $T_{\max,M}$ statistic with a bootstrapped implementation (block length of 20 and 5,000 replications), as recommended by [Hansen et al. \(2011\)](#). This statistic measures the maximum absolute difference between the empirical distribution functions of the two models, offering insights into their overall dissimilarity. The block bootstrapping procedure is implemented to account for potential serial correlation and heteroskedasticity in the loss differences. If $H_{0,\mathbb{M}_s}$ is not rejected, the best model confidence set is $\mathbb{M}_s$. Otherwise, we apply an elimination rule to remove models from $\mathbb{M}_s$ according to the guidelines specified in ([Hansen et al., 2011](#)) and repeat the test.

Let $P_{H_{0,\mathbb{M}_s}}$ denote the p-value associated with the null hypothesis $H_{0,\mathbb{M}_s}$, and let $e_{\mathbb{M}_s}$ represent the model eliminated from the set $\mathbb{M}_s$ when $H_{0,\mathbb{M}_s}$ is rejected. The MCS p-value for model $e_{\mathbb{M}_s}$ is defined as

$$\hat{p}_{e_{\mathbb{M}_s}} = \max_{k \leq s} P_{H_{0,\mathbb{M}_k}},$$

where $\mathbb{M}_1 \supset \mathbb{M}_2 \supset \dots \supset \mathbb{M}_s$.

Economic Measures To assess the economic value of various forecasting models, we adopt a risk parity trading strategy following [Bollerslev et al. \(2018\)](#). In this framework, a risk parity investor dynamically allocates wealth

between a risky asset and a risk-free asset. Given a fraction ω_t of wealth W_t invested in the risky asset with return r_t , and the remainder in the risk-free asset with return r_f , the investor's wealth evolves as:

$$W_{t+1} = W_t(1 + r_f) + W_t\omega_t r_{t+1}^e,$$

where $r_{t+1}^e = r_{t+1} - r_f$ represents the excess return. The investor's conditional expected utility is modeled as:

$$\mathbb{E}_t(U(W_{t+1})) = \mathbb{E}_t(W_{t+1}) - \frac{\gamma^A}{2} \mathbb{V}_t(W_{t+1}),$$

where $\gamma^A = -u''/u'$ denotes the absolute risk aversion of the investor. Given that the risk parity investor maintains a constant Sharpe ratio $SR = \mathbb{E}_t(r_{t+1}^e)/\sqrt{\mathbb{V}_t(r_{t+1}^e)}$, the expected utility function simplifies to:

$$U(\omega_t) = W_t \left[1 + r_f + \omega_t SR \sqrt{\mathbb{V}_t(r_{t+1}^e)} - \frac{\gamma}{2} \omega_t^2 \mathbb{V}_t(r_{t+1}^e) \right] \quad (36)$$

where $\gamma = \gamma^A W_t$ represents relative risk aversion. The investor optimally selects ω_t^* to maximize expected utility, yielding:

$$\omega_t^* = \frac{SR/\gamma}{\sqrt{\mathbb{V}_t(r_{t+1}^e)}} \approx \frac{SR/\gamma}{\sqrt{RV_{t+1}}},$$

where realized volatility (RV_{t+1}) is used as a proxy for the conditional variance of excess returns. The investment decision adjusts based on volatility predictions: if the predicted volatility exceeds the risk tolerance threshold (SR/γ), the investor reduces exposure to the risky asset ($\omega_t^* < 1$); otherwise, the investor takes leveraged positions ($\omega_t^* > 1$).

The key objective is to forecast RV_{t+1} , for which we employ both continuous and discrete time models. Competing forecasting models are denoted by $\mathcal{M}_0$ and the forecast from model $m \in \mathcal{M}_0$ is represented as $\widehat{RV}_{t+1}^m$. The optimal portfolio weight based on model m is given by:

$$\omega_t^{*m} = \frac{SR/\gamma}{\sqrt{\widehat{RV}_{t+1}^m}}. \quad (37)$$

Substituting this into the utility function, we derive the per unit of wealth utility (UoW) for each model at period $t + 1$:

$$UoW_{t+1}^m = \frac{SR^2}{\gamma} \left(\frac{\sqrt{RV_{t+1}}}{\sqrt{\widehat{RV}_{t+1}^m}} - \frac{1}{2} \frac{RV_{t+1}}{\widehat{RV}_{t+1}^m} \right),$$

where the constant term $1 + r_f$ is omitted for direct comparison across models. The cumulative realized utility is

then computed as:

$$UoW^m = \sum_{t=T_0}^{T-1} UoW_{t+1}^m.$$

Following [Bollerslev et al. \(2018\)](#), we assume an annualized Sharpe ratio of $SR = 0.4$ and a risk aversion parameter of $\gamma = 2$. In the ideal scenario where $\widehat{RV}_{t+1}^m = RV_{t+1}$, the realized utility UoW^m reaches 4%. Thus, the forecasting performance is evaluated based on how closely the realized utility approaches this benchmark, with higher accuracy indicating superior predictive ability.

5 Empirical Findings

The data consist of daily realized volatilities (RVs) for the Nasdaq Composite Index ETF, the S&P 500 Index ETF, nine industry index ETFs, the 30 Dow Jones Industrial Average stocks, and 30 cryptocurrencies. Equity index ETFs and individual stock prices are sourced from Refinitiv Tick History, while cryptocurrency prices are obtained from Binance. All data are sampled at 5-minute intervals. For the ETFs, we exclude observations outside standard trading hours (9:30 a.m. to 4:00 p.m.) and remove any trading days with less than 50% of the expected 5-minute observations. Individual stock prices are adjusted for splits. The sample period for equity index ETFs and individual stocks spans from January 4, 2010, to December 30, 2024 (with some exceptions). Summary statistics are reported in [Tables 2](#). In contrast, cryptocurrencies trade continuously, 24 hours a day. The sample periods vary by currency, depending on their launch dates, which range from 2017 to 2023. The end date for all cryptocurrency series is also set to December 30, 2024. See [Table 3](#) for details. Note that a five-year rolling window is used in the forecasting exercise.⁶ Accordingly, several assets with data histories shorter than five years – namely, HON, SHIBI, ICP, LDO, QNT, APE, and ARB – are excluded from the analysis. Models considered are listed in [Table 1](#).

[Figures 1](#) present the forecasting performance comparison of all candidate models for equities based on UoW ratios, mean square forecast error (MSFE) ratios, and QLIKE ratios relative to the chosen benchmark model, namely, fOU (AWML). The benchmark is selected primarily for clarity of presentation. Each panel reports model performance relative to the benchmark, where values above one for the UoW ratio indicate superior utility compared to fOU (AWML), while values below one for MSFE and below minus one for QLIKE ratios indicate higher forecast accuracy.⁷ The dots represent ratios for all assets, and the black cross (red diamond) denotes the mean (median) ratio across all assets.

The results reveal distinct differences in forecasting performance across model classes. Overall, for equities, the discrete time specifications, namely the ARFIMA, HAR-type, and log-HAR families, tend to underperform

⁶Robustness checks using a two-year rolling window yield similar results.

⁷The QLIKE ratio is defined as the ratio of a given model's QLIKE value to the absolute value of the QLIKE of the benchmark model.

Table 1: Model Abbreviations and Full Names

Abbreviation	Full Name
ARFIMA-LM-S (MPL)	ARFIMA: $\alpha \in (-1, 1)$ & $d \in (0, 0.5)$ estimated by MPL
ARFIMA-R-S (MPL)	ARFIMA: $\alpha \in (-1, 1)$ & $d \in (-0.5, 0)$ estimated by MPL
ARFIMA-LM-NS (Whittle)	ARFIMA : $\alpha \in (-1, 1)$ and $d \in (0, 1)$ estimated by Whittle method
RV-HAR	HAR model using raw RV
RV-HARQ	HARQ model using raw RV
RV-HAR-SV	HAR-SV model using raw RV
logRV-HAR	HAR model using log-transformed RV
logRV-HARQ	HARQ model using log-transformed RV
logRV-HAR-SV	HAR-SV model using log-transformed RV
fOU (MLE)	fOU estimated by MLE
fOU (AWLE)	fOU estimated by AWMLE
fBm (MLE)	fBm estimated by MLE
fBm (AWLE)	fBm estimated by AWMLE
BSSP (MM)	BSSP estimated by method of moments
BSSP (MCL)	BSSP, estimated by MCL
BSSG (MM)	BSSG estimated by method of moments
BSSG (MCL)	BSSG estimated by MCL

relative to the continuous time counterparts across all evaluation criteria and horizons. They tend to produce lower UoW values, larger values for MSFE and QLIKE for all three horizons. Moreover, the dispersion of the dots in Figure 1 shows that discrete time models exhibit larger cross-sectional variability and occasional extreme outliers, reflecting unstable predictive accuracy across assets. Within this group, the logRV-HAR and logRV-HAR-SV variants stand out as the most competitive, offering relatively higher realized utilities and lower MSFE and QLIKE ratios, especially at short horizons. However, their median and mean performances remain below those of the continuous time benchmarks.⁸

As the differences among alternative continuous time models are not easy to see in Figure 1, in Figure 2, we consider only continuous time models. In this figure, finer contrasts emerge across horizons. At shorter and medium horizons ($h = 1$ and $h = 5$), the fBm model estimated via the maximum approximate Whittle likelihood method, i.e., fBm (AWML), delivers consistently higher mean UoW ratio and lower mean forecast losses, reflecting its strength in capturing short-run persistence and roughness in volatility. However, the variability is large in all cases. The fOU model follows closely in terms of mean ratios with much smaller variability. By contrast, as the forecast horizon lengthens ($h = 21$), the fOU models (estimated by both AWML and MLE) become dominant, achieving superior mean realized utility, the lowest mean MSFE and QLIKE ratios, and the smallest variability. This pattern suggests that the mean reverting structure of the fOU models better captures long-horizon reversion and memory decay in volatility dynamics, yielding more accurate long-term forecasts. Judged by the mean performance and

⁸Although the average realized utilities of the HAR-type models are close to one when $h = 21$, the dispersion remains large, and their statistical accuracy measures remain inferior to the continuous time benchmarks.

variability in all three horizons, it is reasonable to claim that fOU for log RV is an overall choice to predict RV for equities.

For the 26 valid cryptocurrencies, the benchmark model is the logRV HAR specification. Figure 3 reveals that the HARQ model achieves the highest realized utility at the short horizon ($h = 1$), suggesting that incorporating realized quarticity enhances short run risk and return trade offs. At medium and long horizons ($h = 5$ and $h = 21$), the continuous time BSS models, including both BSSP and BSSG, dominate, reflecting their ability to capture persistent volatility and long memory dynamics.

In contrast, the statistical measures (MSFE and QLIKE ratios) convey a different message. The logHAR type models (including logRV-HAR, logRV-HARQ, and logRV-HAR-SV) consistently outperform other specifications in terms of forecast accuracy, with similar performance across horizons and followed closely by the BSS models. This contrast indicates that while the continuous time models deliver higher realized utility and thus stronger economic gains, the log-HAR type models achieve greater statistical precision, revealing a subtle distinction between economic and statistical forecasting performance. In Appendix C, we also report the results of the model confidence set tests (detailed in Section 4.2) based on both SFE and QLIKE. Similar conclusions are obtained for both equities and cryptocurrencies.

In light of the weak identification problem associated with the ARFIMA model Shi and Yu (2023); Li et al. (2025), our model specification mitigates this concern by constraining the parameter space to plausible ranges. Interestingly, we find that, within the ARFIMA family, the rough specification yields the best performance for equities, whereas the long-memory variants perform better for cryptocurrencies.

6 Conclusion

This paper provides a comprehensive evaluation of alternative models to forecast RV, comparing a broad set of continuous time and discrete time approaches across traditional equities and cryptocurrencies. By applying state-of-the-art estimation techniques and assessing performance using both statistical loss functions and economic utility, we offer a robust and practical perspective on model effectiveness.

The comparative analysis reveals a clear hierarchy of forecasting performance across assets and evaluation criteria. For equities, continuous time models clearly dominate, with the fBm (AWML) model for log RV performing best at short and medium horizons and the fOU (AWML and MLE) models for log RV taking the lead at the long horizon. These models deliver both higher realized utility and lower forecast losses, reflecting their strength in capturing volatility persistence and mean-reverting dynamics. Among the discrete time specifications, the logRV-HAR and logRV-HAR-SV variants perform best but remain below the continuous time benchmarks.

Figure 1: Forecasting performance comparison for equities: realized utility (UoW) ratios, SE ratios, and QLIKE ratios relative to fOU (AWML) for forecast horizons $h = 1, 5, 21$. A higher UoW ratio indicates better performance, whereas lower SE and QLIKE ratios indicate better accuracy. The rolling window size is five years.



Figure 2: Forecasting performance comparison for equities: realized utility (UoW) ratios, SE ratios, and QLIKE ratios relative to fOU (AWML) for forecast horizons $h = 1, 5, 21$. A higher UoW ratio indicates better performance, whereas lower SE and QLIKE ratios indicate better accuracy. The rolling window size is five years.

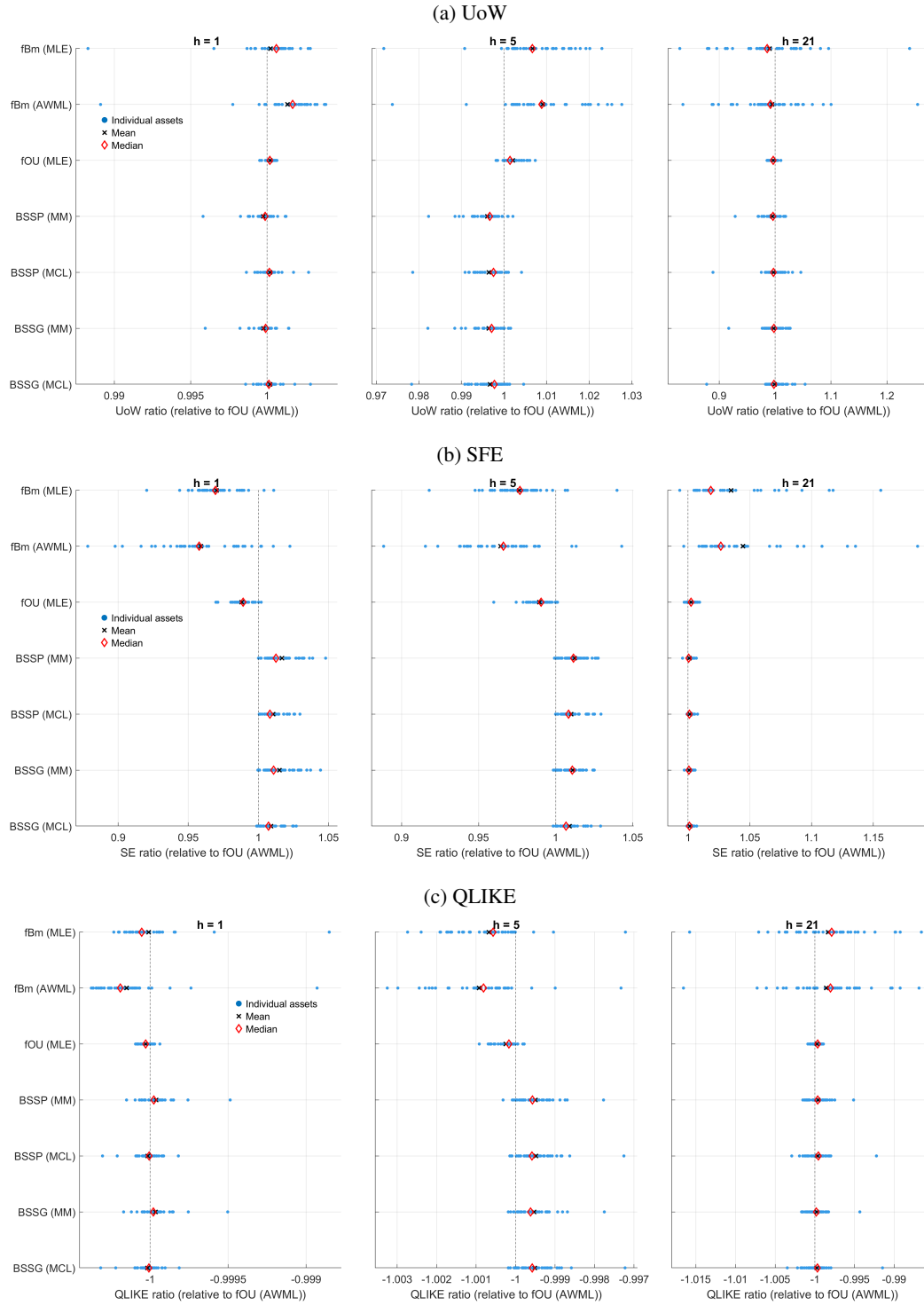
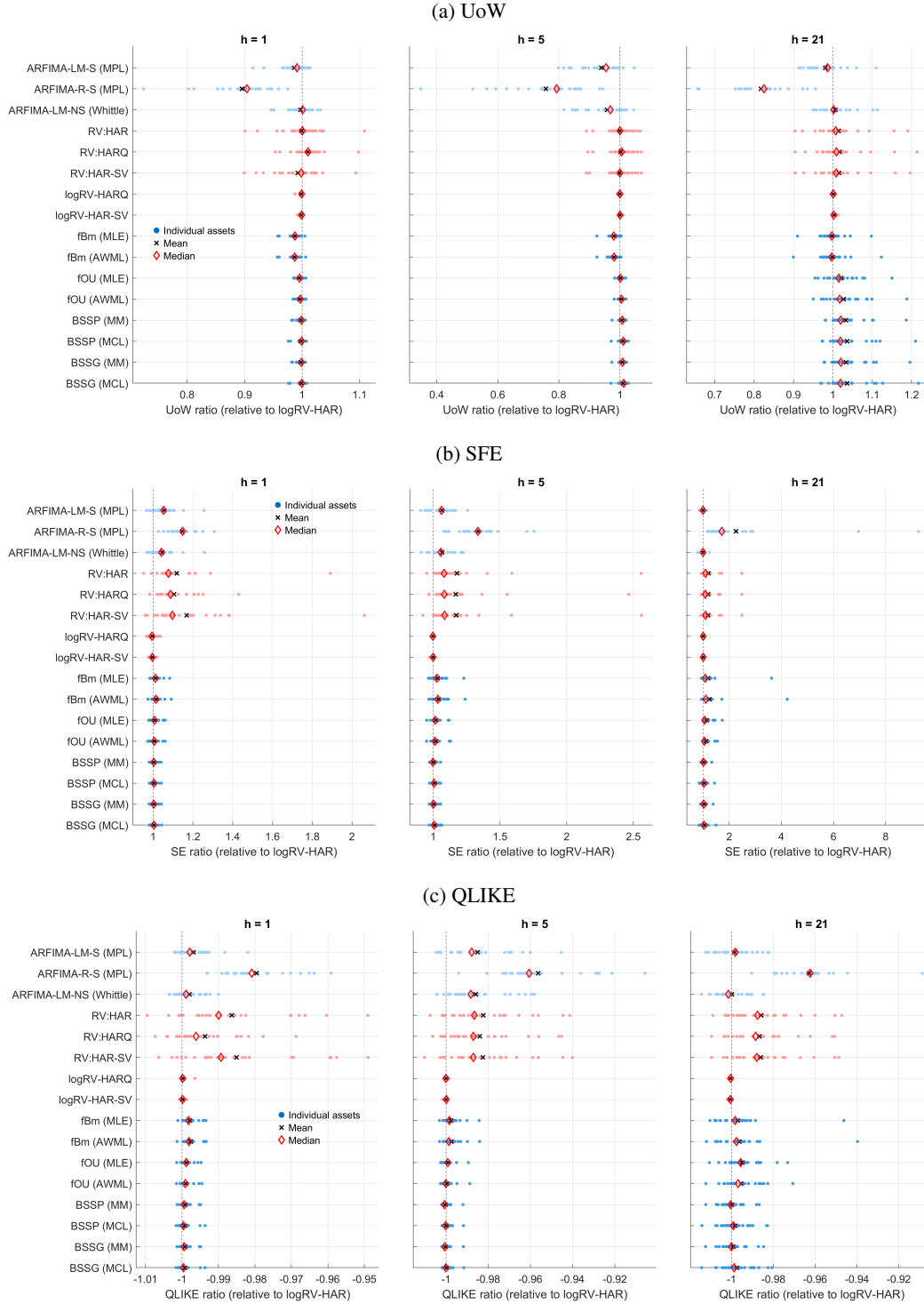


Figure 3: Forecasting performance comparison for cryptocurrencies: realized utility (UoW) ratios, SE ratios, and QLIKE ratios relative to fOU (AWML) for forecast horizons $h = 1, 5, 21$. A higher UoW ratio indicates better performance, whereas lower SE and QLIKE ratios indicate better accuracy. The rolling window size is five years.



For cryptocurrencies, the HARQ model for RV achieves the highest realized utility at the short horizon, while the continuous time BSS models (BSSP and BSSG) for log RV dominate at medium and long horizons. In contrast, the statistical measures (MSFE and QLIKE) favor the log-HAR-type models, which exhibit the highest forecast accuracy and closely match the BSS models’ performance. This divergence underscores the importance of evaluating both economic and statistical criteria: while continuous time models yield greater economic gains, log-HAR-type specifications remain statistically more precise.

Our study contributes to the literature by expanding the modeling scope, improving estimation accuracy, integrating economic relevance into model evaluation, and leveraging a rich dataset that spans multiple asset classes and market regimes.

References

- An, S. and P. Bloomfield (1993). Cox and Reid’s modification in regression models with correlated errors. Department of Statistics, North Carolina State University, Raleigh. 18
- Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39(4), 885–905. 2
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96(453), 42–55. 2
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. Econometrica 71(2), 579–625. 2, 3
- Barndorff-Nielsen, O. E., J. M. Corcuera, and M. Podolskij (2013). Limit theorems for functionals of higher order differences of brownian semi-stationary processes. In Prokhorov and contemporary probability theory, pp. 69–96. Springer. 6, 14
- Bennedsen, M., K. Christensen, and P. Christensen (2022). Likelihood-based estimation of rough stochastic volatility models. Working paper. 9
- Bennedsen, M., K. Christensen, and P. Christensen (2024). Composite likelihood estimation of stationary Gaussian processes with a view toward stochastic volatility. arXiv preprint arXiv:2403.12653. 15
- Bennedsen, M., A. Lunde, and M. S. Pakkanen (2017). Decoupling the short- and long-term behavior of stochastic volatility. Technical report, Department of Economics and Business Economics, Aarhus University. 14, 21
- Bennedsen, M., A. Lunde, and M. S. Pakkanen (2022). Decoupling the short-and long-term behavior of stochastic volatility. Journal of Financial Econometrics. 3, 6, 12, 13, 14, 15
- Bloomfield, P. (1985). On series representations for linear predictors. The Annals of Probability, 226–233. 16

- Bolko, A. E., K. Christensen, M. S. Pakkanen, and B. Veliev (2022). A GMM approach to estimate the roughness of stochastic volatility. Journal of Econometrics, forthcoming. 2
- Bollerslev, T., J. Cai, and F. M. Song (2000). Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the us treasury bond market. Journal of Empirical Finance 7(1), 37–55. 2
- Bollerslev, T., B. Hood, J. Huss, and L. H. Pedersen (2018). Risk everywhere: Modeling and managing volatility. The Review of Financial Studies 31(7), 2729–2773. 3, 23, 25
- Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2016). Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics 192(1), 1–18. 2, 19, 20, 22
- Brockwell, P. J. and R. A. Davis (2009). Time series: theory and methods. Springer Science & Business Media. 16
- Brouste, A., M. Soltane, and I. Votsi (2020). One-step estimation for the fractional Gaussian noise at high-frequency. ESAIM: Probability and Statistics 24, 827–841. 6
- Cheridito, P., H. Kawaguchi, and M. Maejima (2003). Fractional Ornstein-Uhlenbeck processes. Electronic Journal of Probability 8, 1–14. 9
- Chong, C. H. and V. Todorov (2025). A nonparametric test for rough volatility. Journal of the American Statistical Association (just-accepted), 1–23. 2
- Clements, A. and D. P. Preve (2021). A practical guide to harnessing the har volatility model. Journal of Banking & Finance 133, 106285. 3
- Corcuera, J. M., E. Hedevar, M. S. Pakkanen, and M. Podolskij (2013). Asymptotic theory for brownian semi-stationary processes with application to turbulence. Stochastic Processes and their Applications 123(7), 2552–2574. A Special Issue on the Occasion of the 2013 International Year of Statistics. 14
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7(2), 174–196. 2, 3, 6, 19
- Corsi, F. and R. Renò (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. 2
- Cox, D. R. and N. Reid (1987). Parameter orthogonality and approximate conditional inference. Journal of the Royal Statistical Society: Series B (Methodological) 49(1), 1–18. 17
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50(4), 987–1007. 2
- Fox, R. and M. S. Taqqu (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. The Annals of Statistics, 517–532. 18

- Fukasawa, M. and T. Takabatake (2019). Asymptotically efficient estimators for self-similar stationary Gaussian noises under high frequency observations. Bernoulli 25(3), 1870–1900. 6, 8
- Fukasawa, M., T. Takabatake, and R. Westphal (2022). Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. Mathematical Finance. 7, 37
- Garnier, J. and K. Sølna (2018). Option pricing under fast-varying and rough stochastic volatility. Annals of Finance 14(4), 489–516. 9
- Gatheral, J., T. Jaisson, and M. Rosenbaum (2018). Volatility is rough. Quantitative Finance 18(6), 933–949. 2, 6, 21
- Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. Journal of Time Series Analysis 4(4), 221–238. 17
- Giraitis, L. and D. Surgailis (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle’s estimate. Probability Theory and Related Fields 86(1), 87–104. 18
- Gradshteyn, I. S. and I. M. Ryzhik (2014). Table of Integrals, Series, and Products. Academic press. 13
- Granger, C. W. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis 1(1), 15–29. 2, 17
- Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14(2), 227–238. 2, 17
- Hannan, E. J. (1973). The asymptotic theory of linear time-series models. Journal of Applied Probability, 130–145. 18
- Hansen, P. R. and A. Lunde (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? Journal of Applied Econometrics 20(7), 873–889. 2
- Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. Econometrica 79(2), 453–497. 3, 23
- Haslett, J. and A. E. Raftery (1989). Space-time modelling with long-memory dependence: Assessing ireland’s wind power resource. Journal of the Royal Statistical Society. Series C (Applied Statistics) 38(1), 1–50. 36
- Hauser, M. A. (1999). Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study. Journal of Statistical Planning and Inference 80(1-2), 229–255. 18
- Hosking, J. R. (1981). Fractional differencing. Biometrika 68(1), 165–76. 16, 17
- Hult, H. (2003). Topics on fractional Brownian motion and regular variation for stochastic processes. Ph. D. thesis, Matematik. 10
- Hurvich, C. M. and B. K. Ray (1995). Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. Journal of Time Series Analysis 16(1), 17–41. 18

- Künsch, H. (1987). Statistical aspects of self-similar processes. In Proceedings of the First Congress of the Bernoulli Society, 1987. 17
- Lang, G. and F. Roueff (2001). Semi-parametric estimation of the hölder exponent of a stationary gaussian process with minimax rates. Statistical Inference for Stochastic Processes 4, 283–306. 6
- Laurent, S., R. Renò, and S. Shi (2024). Realized drift. Journal of Econometrics, 105813. 3
- Li, J., P. C. B. Phillips, S. Shi, and J. Yu (2025). Weak identification of long memory with implications for volatility modeling. The Review of Financial Studies 138. 2, 3, 17, 27
- Lieberman, O. (2005). On plug-in estimation of long memory models. Econometric Theory, 431–454. 17
- Nuzman, C. J. and H. V. Poor (2000). Linear estimation of self-similar processes via lamperti’s transformation. Journal of Applied Probability 37(2), 429–452. 21
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160(1), 246–256. 3, 23
- Patton, A. J. and K. Sheppard (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. Review of Economics and Statistics 97(3), 683–697. 2, 3, 19
- Paxson, V. (1997). Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic. ACM SIGCOMM Computer Communication Review 27(5), 5–18. 8
- Poon, S.-H. and C. W. Granger (2003, June). Forecasting volatility in financial markets: A review. Journal of Economic Literature 41(2), 478–539. 2, 4
- Ramsey, F. L. (1974). Characterization of the partial autocorrelation function. The Annals of Statistics 2(6), 1296–1301. 36
- Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. The Annals of Statistics 23(5), 1630–1661. 17
- Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. The Annals of Statistics 23(4), 1048–1072. 17
- Shi, S., Y. Jun, , and C. Zhang (2025). Fractional Gaussian noise: Spectral density and estimation methods. 46. 6
- Shi, S., Y. Jun, and C. Zhang (2024). On the spectral density of fractional Ornstein-Uhlenbeck processes. Journal of Econometrics 245(05872). 7, 9, 11, 12, 37
- Shi, S. and J. Yu (2023). Volatility puzzle: Long memory or anti-persistency. Management Science 69(7), 3861–3883. 2, 3, 16, 17, 27
- Sinai, Y. G. (1976). Self-similar probability distributions. Theory of Probability & Its Applications 21(1), 64–80. 7

- Sundaresan, S. M. (2000). Continuous-time methods in finance: A review and an assessment. The Journal of Finance 55(4), 1569–1622. [2](#)
- Tanaka, K. (2013). Distributions of the maximum likelihood and minimum contrast estimators associated with the fractional Ornstein–Uhlenbeck process. Statistical Inference for Stochastic Processes 16(3), 173–192. [2](#)
- Velasco, C. and P. M. Robinson (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. Journal of the American Statistical Association 95(452), 1229–1243. [18](#)
- Wang, X., W. Xiao, and J. Yu (2023). Modeling and forecasting realized volatility with the fractional ornstein–uhlenbeck process. Journal of Econometrics 232(2), 389–415. [2](#), [3](#), [9](#), [21](#)
- Wang, X., W. Xiao, J. Yu, and C. Zhang (2024). Maximum likelihood estimation of fractional ornstein-uhlenbeck process with discretely sampled data. Working paper. [6](#), [9](#), [10](#)
- Wang, X. and J. Yu (2023). Latent local-to-unity models. Econometric Reviews 42(7), 586–611. [2](#)
- Wang, X., J. Yu, and C. Zhang (2024). On the optimal forecast with the fractional Brownian motion. Quantitative Finance. [3](#), [9](#), [21](#)
- Žurbenko, I. G. (1979). On the efficiency of estimates of a spectral density. Scandinavian Journal of Statistics, 49–56. [19](#)

A Mathematical Annex

A.1 Implementation Details: TDML for fBm

The MLE method is known for its computational intensity, primarily due to the necessity of calculating the inverse matrix of Σ_X . To enhance computational efficiency without sacrificing asymptotic performance, we employ the recursive approach introduced by [Haslett and Raftery \(1989\)](#) for the computation of the log likelihood function. Since, conditional on $X^{t-1} = (x_\Delta, x_{2\Delta}, \dots, x_{T-1\Delta})'$, $x_{t\Delta}$ has a normal distribution. The log likelihood function can be rewritten as

$$\tilde{l}(\varphi; X) = -\frac{1}{2} \sum_{t=1}^T \log v_{t\Delta} - \frac{1}{2} \sum_{t=1}^T \frac{(x_{t\Delta} - \eta_{t\Delta})^2}{v_{t\Delta}}, \quad (38)$$

where $\eta_{t\Delta}$ is the conditional mean of $x_{t\Delta}$ and $v_{t\Delta}$ is the conditional variance. Let $\gamma_x(k)$ be the autocovariance of $x_{t\Delta}$ given in (3) and ϕ_{tj} be the partial linear regression coefficients. Using results in [Ramsey \(1974, Theorem 2\)](#) for Gaussian processes, we have

$$\eta_{t\Delta} = \sum_{j=1}^{t-1} \phi_{tj} x_{(t-j)\Delta} \quad \text{and} \quad v_{t\Delta} = \gamma_x(0) \Pi_{j=1}^{t-1} (1 - \phi_{jj}^2).$$

The log likelihood function is computed from the algorithm below.

Algorithm 1 Computation of the TDML Log-Likelihood

- 1: Compute autocovariances $\gamma_x(k)$ for $k = 0, \dots, T-1$.
- 2: Use the Durbin–Levinson recursion (with $\sigma^2 = 1$) to obtain:
 - Partial regression coefficients ϕ_{tj} ,
 - Scaled variances $\bar{v}_{t\Delta}$,
 - Conditional means $\eta_{t\Delta} = \sum_{j=1}^{t-1} \phi_{tj} x_{(t-j)\Delta}$.
- 3: Estimate σ^2 by $\hat{\sigma}_{ML}^2 = T^{-1} \sum_{t=1}^T (x_{t\Delta} - \eta_{t\Delta})^2 / \bar{v}_{t\Delta}$.
- 4: Form $\hat{v}_{t\Delta} = \bar{v}_{t\Delta} \hat{\sigma}_{ML}^2$.
- 5: Evaluate the log-likelihood

$$\tilde{l}(\varphi; X) = -\frac{1}{2} \sum_{t=1}^T \log \hat{v}_{t\Delta} - \frac{1}{2} \sum_{t=1}^T \frac{(x_{t\Delta} - \eta_{t\Delta})^2}{\hat{v}_{t\Delta}}.$$

A.2 Proof of Lemma 2.1

By definition,

$$g_X^\Lambda(\lambda; H) = \frac{1}{\pi} \Delta^{2H} C_H (1 - \cos \lambda) \sum_{k=-\infty}^{\infty} |2\pi k + \lambda|^{-1-2H}$$

$$\approx \frac{1}{2\pi} \Delta^{2H} C_H \lambda^{1-2H} (1 + o(1)) \quad \text{when } \lambda \rightarrow 0 \text{ and } H > 1/2, \quad (39)$$

since $1 - \cos \lambda \sim \frac{1}{2} \lambda^2$ for small λ . The derivation below follows closely of [Fukasawa et al. \(2022\)](#) and [Shi et al. \(2024\)](#). First, note that

$$\begin{aligned} B^\dagger(H, \varepsilon) &= \int_0^\varepsilon \ln \left(\frac{1}{2\pi} C_H \Delta^{2H} \lambda^{1-2H} \right) d\lambda \\ &= \ln \left(\frac{1}{2\pi} C_H \Delta^{2H} \right) \varepsilon + \int_0^\varepsilon \ln(\lambda^{1-2H}) d\lambda, \\ &= \ln \left(\frac{1}{2\pi} C_H \Delta^{2H} \right) \varepsilon + (1-2H)(\varepsilon \ln \varepsilon - \varepsilon). \end{aligned}$$

since for small λ , $1 - \cos \lambda \sim \frac{1}{2} \lambda^2$.

Second, note that the periodogram can be expressed as $I_n(\lambda) = \frac{1}{2\pi} \sum_{\tau=-(n-1)}^{n-1} \hat{\gamma}_n(\tau) e^{-i\lambda\tau}$, where $e^{-i\lambda\tau} = \cos(\lambda\tau) - i\sin(\lambda\tau)$. Since the periodogram is real-valued, only the real part, $\cos(\lambda\tau)$, contributes to the integral below. Therefore,

$$B^\dagger(H, \varepsilon) := \int_0^\varepsilon \frac{I_n(\lambda)}{g_X^\Delta(\lambda; H)} d\lambda = \delta_H(0, \varepsilon) \hat{\gamma}_n(0) + 2 \sum_{\tau=1}^{n-1} \delta_H(\tau, \varepsilon) \hat{\gamma}_n(\tau),$$

where

$$\delta_H(\tau, \varepsilon) := \frac{1}{2\pi} \int_0^\varepsilon \frac{\cos(\tau\lambda)}{g_X^\Delta(\lambda; H)} d\lambda \approx \frac{1}{C_H \Delta^{2H}} \int_0^\varepsilon \frac{\cos(\tau\lambda)}{\lambda^{1-2H}} d\lambda.$$

It follows that

$$\delta_H(0, \varepsilon) \approx \frac{1}{C_H \Delta^{2H}} \int_0^\varepsilon \lambda^{2H-1} d\lambda = \frac{\varepsilon^{2H}}{2H C_H \Delta^{2H}}$$

and for $\tau \geq 1$,

$$\delta_H(\tau, \varepsilon) \approx \frac{1}{C_H \Delta^{2H}} \sum_{j=0}^{\infty} \frac{(-1)^j \tau^{2j}}{(2j)!} \int_0^\varepsilon \lambda^{2j-1+2H} d\lambda \approx \frac{1}{C_H \Delta^{2H}} \sum_{j=0}^J \frac{(-1)^j \tau^{2j}}{(2j)!} \frac{\varepsilon^{2(j+H)}}{2(j+H)},$$

using the Maclaurin series expansion for cosine, where J is chosen to be a large number to ensure accuracy.

A.3 Profiled Whittle Likelihood Function

Step 1: Rewrite the objective function Substituting $f_X^\Delta(\lambda; \psi) = \sigma^2 g_X^\Delta(\lambda; H)$, we get:

$$l_W(H; x) = -\frac{1}{2\pi} \int_0^\pi \left(\ln \sigma^2 + \ln g_X^\Delta(\lambda; H) + \frac{I(\lambda)}{\sigma^2 g_X^\Delta(\lambda; H)} \right) d\lambda.$$

Split the terms:

$$l_W(H; x) = -\frac{1}{2\pi} \int_0^\pi \ln \sigma^2 d\lambda - \frac{1}{2\pi} \int_0^\pi \ln g_X^\Delta(\lambda; H) d\lambda - \frac{1}{2\pi} \int_0^\pi \frac{I(\lambda)}{\sigma^2 g_X^\Delta(\lambda; H)} d\lambda.$$

Step 2: Profile out σ^2 . Let $Q(H) = \frac{1}{2\pi} \int_0^\pi \frac{I(\lambda)}{g_X^\Delta(\lambda; H)} d\lambda$. The objective function

$$l_W(H; x) = -\frac{1}{2} \ln \sigma^2 - \frac{1}{\sigma^2} Q(H) - \frac{1}{2\pi} \int_0^\pi \ln g_X^\Delta(\lambda; H) d\lambda.$$

The derivative with respect to σ^2 is:

$$-\frac{1}{2\sigma^2} + \frac{Q(H)}{\sigma^4} = 0 \Rightarrow \sigma^2 = 2Q(H).$$

Step 3: Substitute $\sigma^2 = 2Q(H)$ into the objective function. The profiled objective function is:

$$\tilde{l}_W(H; x) = -\frac{1}{\pi} \int_0^\pi \ln g_X^\Delta(\lambda; H) d\lambda - \ln \left(\frac{1}{\pi} \int_0^\pi \frac{I(\lambda)}{g_X^\Delta(\lambda; H)} d\lambda \right).$$

A.4 Tapered Whittle Estimator

The Whittle objective function is as follows:

$$\begin{aligned} l_W^p(\delta^\dagger; \mathbf{y}) &= \frac{p}{T} \sum_{j \in J(p)} \ln f_{\mathbf{y}}(\lambda_j; \delta^\dagger, \sigma) + \frac{p}{T} \sum_{j \in J(p)} \frac{I^p(\lambda_j)}{f_{\mathbf{y}}(\lambda_j; \delta^\dagger, \sigma)}, \\ &\simeq \ln \sigma^2 - \ln(2\pi) + \frac{p}{T} \sum_{j \in J(p)} \ln \eta_{\mathbf{y}}(\lambda_j; \delta^\dagger) + \frac{p}{T} \frac{2\pi}{\sigma^2} \sum_{j \in J(p)} \frac{I^p(\lambda_j)}{\eta_{\mathbf{y}}(\lambda_j; \delta^\dagger)}, \end{aligned}$$

where $J(p) = \{p, 2p, \dots, T-p\}$ (assuming that T/p is an integer). Take the derivative with respect to σ^2 ,

$$\frac{1}{\sigma^2} - \frac{p}{T} \frac{2\pi}{\sigma^4} \sum_{j \in J(p)} \frac{I^p(\lambda_j)}{\eta_{\mathbf{y}}(\lambda_j; \delta^\dagger)} = 0 \Rightarrow \sigma^2 = \frac{2\pi p}{T} \sum_{j \in J(p)} \frac{I^p(\lambda_j)}{\eta_{\mathbf{y}}(\lambda_j; \delta^\dagger)}.$$

Substitute it back into the original objective function,

$$\begin{aligned} l_W^p(\delta^\dagger; \mathbf{y}) &= \ln \sigma^2 - \ln(2\pi) + \frac{p}{T} \sum_{j \in J(p)} \ln \eta_{\mathbf{y}}(\lambda_j; \delta^\dagger) + \frac{p}{T} \frac{2\pi}{\sigma^2} \sum_{j \in J(p)} \frac{I^p(\lambda_j)}{\eta_{\mathbf{y}}(\lambda_j; \delta^\dagger)}, \\ &= \ln \left(\frac{2\pi p}{T} \sum_{j \in J(p)} \frac{I^p(\lambda_j)}{\eta_{\mathbf{y}}(\lambda_j; \delta^\dagger)} \right) + \frac{p}{T} \sum_{j \in J(p)} \ln \eta_{\mathbf{y}}(\lambda_j; \delta^\dagger) - \ln(2\pi) + 1, \end{aligned}$$

$$\propto \ln \left\{ \frac{2\pi p}{T} \sum_{j \in J(p)} \frac{I^p(\lambda_j)}{\eta_{\mathbf{y}}(\lambda_j; \delta^\dagger)} \exp \left(\frac{p}{T} \sum_{j \in J(p)} \ln \eta_{\mathbf{y}}(\lambda_j; \delta^\dagger) \right) \right\}.$$

Let $\eta_{\mathbf{y}}^*(\lambda_j; \delta^\dagger) = \eta_{\mathbf{y}}(\lambda_j; \delta^\dagger) / Z$ with $Z \equiv \exp \left(\frac{p}{T} \sum_{j \in J(p)} \ln \eta_{\mathbf{y}}(\lambda_j; \delta^\dagger) \right)$. The tapered Whittle estimator can be rewritten as the following:

$$\hat{\delta}_W^\dagger = \arg \min_{\delta^\dagger} \ln \frac{2\pi p}{T} \sum_{j \in J(p)} \frac{I^p(\lambda_j)}{\eta_{\mathbf{y}}^*(\lambda_j; \delta^\dagger)},$$

and

$$(\hat{\sigma}_W^p)^2 = \frac{2\pi p}{T} \sum_{j \in J(p)} \frac{I^p(\lambda_j)}{\eta_{\mathbf{y}}(\lambda_j; \delta^\dagger)}.$$

B Summary Statistics

Table 2: Summary Statistics: log RV

Name (RIC)	Start date EFTs	Nob	Mean	Std.	Skew.	Kurto.
Nasdaq composite index ETF (QQQ)	23-Mar-2011	3232	-9.84	0.96	0.48	3.67
S&P 500 market ETF (SPY)	04-Jan-2010	3487	-10.38	1.02	0.40	3.69
Industry ETF: Material (XLB)	04-Jan-2010	3648	-9.69	0.88	0.68	4.27
Industry ETF: Energy (XLE)	04-Jan-2010	3648	-9.15	0.90	0.66	4.10
Industry ETF: Financial (XLF)	04-Jan-2010	3648	-9.65	0.87	1.01	6.12
Industry ETF: Industrial (XLI)	04-Jan-2010	3648	-9.90	0.92	0.85	5.68
Industry ETF: Technology (XLK)	04-Jan-2010	3648	-9.76	0.95	0.56	3.77
Industry ETF: Consumer staples (XLP)	04-Jan-2010	3648	-10.38	0.80	1.17	6.48
Industry ETF: Utilities (XLU)	04-Jan-2010	3648	-9.74	0.73	1.15	7.41
Industry ETF: Health care (XLV)	04-Jan-2010	3648	-10.08	0.84	1.11	6.75
Industry ETF: consumer discretionary (XLY)	04-Jan-2010	3648	-9.84	1.00	0.58	3.89
Dow Jones 30						
Apple Inc (AAPL)	04-Jan-2010	3628	-9.09	0.86	0.50	3.80
Honeywell International Inc (ALD or HON)	11-May-2021	886	-9.29	0.62	0.48	3.15
Amgen Inc (AMGN)	04-Jan-2010	3628	-8.97	0.73	0.93	5.35
American Express Co (AEXP or AXP)	04-Jan-2010	3647	-9.16	0.86	0.74	4.62
Boeing Co (BA)	04-Jan-2010	3647	-8.82	0.96	0.74	4.53
Verizon Communications Inc (BEL or VZ)	04-Jan-2010	3647	-9.55	0.69	0.95	6.69
Caterpillar Inc (CAT)	04-Jan-2010	3647	-8.87	0.77	0.54	4.00
Chevron Corp (CHV or CVX)	04-Jan-2010	3647	-9.19	0.85	0.91	5.29
Salesforce.Com Inc (CRM)	04-Jan-2010	3647	-8.50	0.81	0.46	3.93
Cisco Systems Inc (CSCO)	04-Jan-2010	3628	-9.18	0.74	0.81	5.14
Walt Disney Co (DIS)	04-Jan-2010	3647	-9.21	0.84	0.76	4.76
Goldman Sachs Group Inc (GS)	04-Jan-2010	3647	-8.91	0.76	0.81	5.10
Home Depot Inc (HD)	04-Jan-2010	3647	-9.26	0.77	0.99	6.25
International Business Machines Corp (IBM)	04-Jan-2010	3647	-9.54	0.76	0.99	5.73
Intel Corps (INTC)	04-Jan-2010	3628	-8.79	0.80	0.58	4.20
Johnson & Johnson (JNJ)	04-Jan-2010	3647	-9.80	0.75	1.12	6.61
JPMorgan Chase & Co (JPM)	04-Jan-2010	3647	-9.09	0.82	0.76	4.87
Coca-Cola Co (KO)	04-Jan-2010	3647	-9.83	0.73	1.17	7.12
McDonald's Corp (MCD)	04-Jan-2010	3647	-9.75	0.75	1.34	8.16
3M Co (MMM)	04-Jan-2010	3647	-9.45	0.85	0.61	4.51
Merck & Co Inc (MRK)	04-Jan-2010	3647	-9.37	0.71	0.91	5.68
Microsoft Corp (MSFT)	04-Jan-2010	3628	-9.13	0.79	0.64	4.42
Nike Inc (NIKE)	04-Jan-2010	3647	-9.12	0.73	1.06	6.20
Procter & Gamble Co (PG)	04-Jan-2010	3647	-9.80	0.74	1.27	8.15
Travelers Companies Inc (SPC or TRV)	04-Jan-2010	3646	-9.47	0.79	0.80	5.22
UnitedHealth Group Inc (UNH)	04-Jan-2010	3646	-9.07	0.78	0.85	5.20
Visa Inc (V)	04-Jan-2010	3647	-9.27	0.83	0.88	5.23
Walgreens Boots Alliance Inc (WAG or WBA)	02-Jan-2015	2412	-8.69	0.83	0.60	4.32
Walmart Inc (WMT)	04-Jan-2010	3647	-9.66	0.72	1.17	6.64
Exxon Mobile Co (XMO)	04-Jan-2010	3647	-9.27	0.89	0.77	4.77

Note: Since Dow Inc. (NYSE: DOW) is listed on NYSE only since 2019, its sample size is substantially shorter than all the other stocks. For this reason we replace it with Exxon Mobil Co. (NYSE: XOM), which belonged to the Dow Jones index until August 31, 2020.

Table 3: Summary Statistics: log RV

Name (RIC)	Start date	Nob	Mean	Std.	Skew.	Kurto.
Cryptocurrency						
Bitcoin (BTC)	18-Aug-2017	2571	-5.50	1.18	0.24	3.63
Ethereum (ETH)	18-Aug-2017	2571	-5.04	1.12	0.23	3.72
Binance Coin (BNB)	07-Nov-2017	2494	-4.99	1.20	0.55	3.89
XRP (XRP)	05-May-2018	2328	-4.93	1.13	0.63	3.68
Cardano (ADA)	18-Apr-2018	2344	-4.64	0.97	0.37	3.54
Dogecoin (DOGE)	06-Jul-2019	1923	-4.59	1.15	0.78	4.82
Polygon (MATIC)	27-Apr-2019	1870	-4.27	1.16	0.41	3.52
Solana (SOL)	12-Aug-2020	1541	-4.25	1.00	0.48	3.57
Polkadot (DOT)	19-Aug-2020	1534	-4.62	1.07	0.21	3.49
Litecoin (LTC)	14-Dec-2017	2459	-4.75	1.02	0.38	3.60
Tron (TRX)	12-Jun-2018	2291	-5.36	1.34	0.14	2.64
Shiba Inu (SHIB)	11-May-2021	1284	-4.44	1.08	0.85	4.04
Avalanche (AVAX)	23-Sep-2020	1500	-4.31	1.06	0.24	3.13
Chainlink (LINK)	17-Jan-2019	2084	-4.38	0.97	0.37	3.71
Cosmos (ATOM)	01-May-2019	1985	-4.41	1.03	0.10	3.27
Uniswap (UNI)	18-Sep-2020	1505	-4.43	0.99	0.22	3.55
Monero (XMR)	16-Mar-2019	1705	-4.84	0.91	0.46	4.01
Ethereum Classic (ETC)	13-Jun-2018	2290	-4.70	1.06	0.43	3.49
Internet Computer (ICP)	12-May-2021	1283	-4.27	0.94	0.40	3.72
Stellar (XLM)	01-Jun-2018	2302	-4.79	1.01	0.78	3.78
Bitcoin Cash (BCH)	29-Nov-2019	1784	-4.80	0.97	0.64	3.86
Filecoin (FIL)	16-Oct-2020	1478	-4.51	0.97	0.42	3.92
Hedera (HBAR)	01-Oct-2019	1840	-4.18	0.96	0.87	3.85
Lido DAO (LDO)	10-May-2022	934	-4.12	1.01	0.55	3.81
Arbitrum (ARB)	25-Mar-2023	626	-4.79	0.88	0.21	3.67
Near Protocol (NEAR)	15-Oct-2020	1479	-4.12	0.97	0.12	3.27
VeChain (VET)	26-Jul-2018	2251	-4.38	1.01	0.35	3.56
ApeCoin (APE)	18-Mar-2022	985	-4.39	0.90	0.59	4.16
Quant (QNT)	30-Jul-2021	1206	-4.53	1.01	0.53	3.12
Algorand (ALGO)	22-Jun-2019	1936	-4.28	1.01	0.41	3.13

C Model Confidence Set Tests

Figure 4 presents the forecasting performance comparison of all candidate models for equities based on each model's MCS p-value. The top (bottom) panel plots the MCS p-values based on MSFE (QLIKE). The dots represent p-values for all assets, and the black cross (red diamond) denotes the mean (median) p-value across all assets. The higher the p-value, the better the model performs. The overall results are the same as those based on Figure 1. That is, in general, while the fBm (AWML) model for log RV performing best at short and medium horizons, the fOU (AWML and MLE) models for log RV outperform at the long horizon. Taking into account of both the mean p-value and variability of the p-values, we would like to recommend to use the fOU model for log RV to forecast RV for equity.

Figure 5 presents the forecasting performance comparison of all candidate models for cryptocurrencies based on each model's MCS p-value. A similar conclusion to that from Figure 3 can be drawn from Figure 5. That is,

the three log-HAR-type models exhibit the highest forecast accuracy at the short and medium horizons. In the long horizon, however, the ARFIMA-LM-S model outperforms other models.

Figure 4: Forecasting performance comparison for equities: p-values of the model confidence set tests based on FSE and QLIKE. The rolling window size is five years.

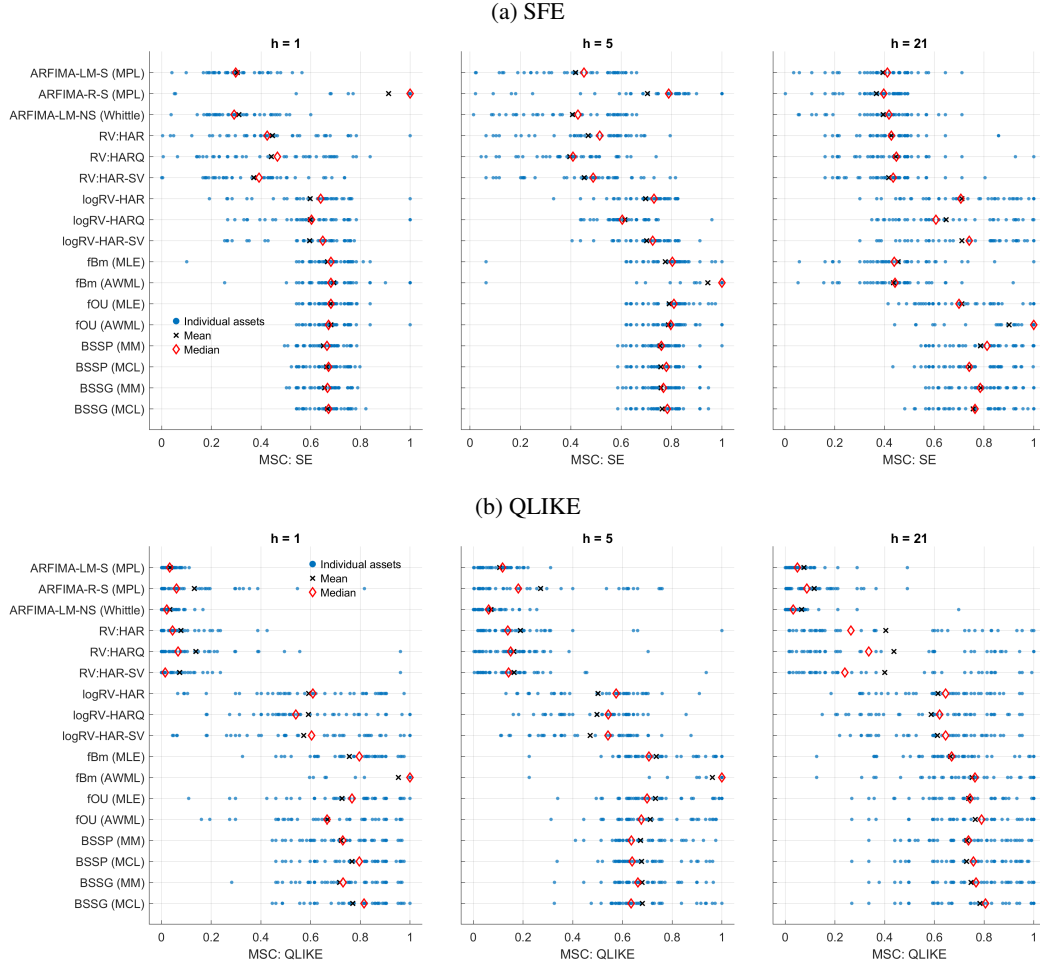


Figure 5: Forecasting performance comparison for cryptocurrencies: p-values of the model confidence set tests based on FSE and QLIKE. The rolling window size is five years.

