Finite Sample Comparison of Alternative Estimators for Fractional Gaussian Noise*

Shuping Shi†, Jun Yu††, Chen Zhang††

† Macquarie University

††Singapore Management University

November 22, 2022

Abstract

The fractional Brownian motion (fBm) process is a continuous-time Gaussian process with its increment being the fractional Gaussian noise (fGn). It has enjoyed widespread empirical applications across many fields, from science to economics and finance. The dynamics of fBm and fGn are governed by a fractional parameter $H \in (0, 1)$. This paper first derives an analytical expression for the spectral density of fGn and investigates the accuracy of various approximation methods for the spectral density. Next, we conduct an extensive Monte Carlo study comparing the finite sample performance and computational cost of alternative estimation methods for H under the fGn specification. These methods include the log periodogram regression method, the local Whittle method, the time-domain maximum likelihood (ML) method, the Whittle ML method, and the change-of-frequency method. We implement two versions of the Whittle method, one based on the analytical expression for the spectral density and the other based on Paxson's approximation. Special attention is paid to highly anti-persistent processes with H close to zero, which are of empirical relevance to financial volatility modelling. Considering the trade-off between statistical and computational efficiency, we recommend using either the Whittle ML method based on Paxson's approximation or the time-domain ML method. We model the log realized volatility dynamics of 40 financial assets in the US market from 2012 to 2019 with fBm. Although all estimation methods suggest rough volatility, the implied degree of roughness varies substantially with the estimation methods, highlighting the importance of understanding the finite sample performance of various estimation methods.

*We would like to thank Brendan Beare, Masaaki Fukasawa, Andrew Patton, and Tetsuya Takabatake for helpful comments. Shuping Shi, Department of Economics, Macquarie University, Sydney, Australia. Email: shuping.shi@mq.edu.au. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: yujun@smu.edu.sg. Chen Zhang, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: chenzhang@smu.edu.sg. Shi acknowledges support from the Australian Research Council (Grant No. DE190100840). Yu and Zhang would like to acknowledge that this research/project is supported by the Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 2 (Award Number MOE-T2EP402A20-0002).

JEL classification: C12, C22, G01
Keywords: Fractional Brownian motion; Fractional Gaussian noise; Semiparametric method; Maximum likelihood; Whittle likelihood; Change-of-frequency; Realised volatility

1 Introduction

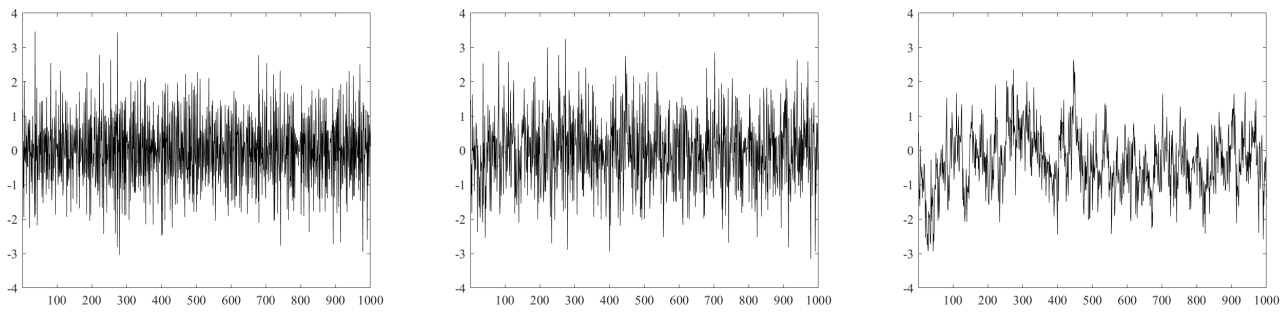
The fractional Brownian motion (fBm) is a continuous-time Gaussian process that generalises the standard Brownian motion. This process was shown useful in characterising the empirical feature discovered by Hurst (1951) in over 75 different geophysical phenomena (later known as long memory) in the 1960s (Mandelbrot, 1965; Mandelbrot and Van Ness, 1968). It has since enjoyed widespread empirical applications across many fields, including hydrologic sciences, earth science, data network traffic, economics, and finance. See Molz et al. (1997); Korvin (1992); Park and Willinger (2000); Graves et al. (2017) for reviews.

The fBm process is governed by a self-similar parameter H, also known as the fractional parameter or the Hurst exponent, ranging between zero and one (non-inclusive). It captures qualitatively different behaviour with different values of H. The standard Brownian motion corresponds to the case of $H = 0.5$. The increment of fBm is the fractional Gaussian noise (fGn), with its autocorrelation function decaying at a hyperbolical rate. Figure 1 shows different traces of fGn. The long-run variance of fGn is not summable when $H > 0.5$ and the time series is said to have a *long memory*, whereas when $H < 0.5$ the long-run variance is zero and the time series is *anti-persistent* or *rough*. The long memory fGn process behaves like a long-term trend, whereas the rough fGn generates short-term reversal. In the continuous-time literature, the fGn process is often embedded in a general continuous-time specification such as the fractional Ornstein-Uhlenbeck (fOU) process (Comte, 1996; Comte and Renault, 1996).

Figure 1: Typical sample path of fGn

(a) $H = 0.1$ (b) $H = 0.5$ (c) $H = 0.9$

.



In economics and finance, another popular process for capturing the long memory or anti-

persistent feature of empirical data is the fractionally integrated noise (fIn), governed by the memory parameter d. The two processes share similar empirical features. Geweke and Porter-Hudak (1983) show that the ratio between the spectral densities of fGn and fIn is uniformly bounded in the spectral frequency when $H = d + 0.5$. The fIn process is most often seen embedded in a stationary autoregressive moving average model, leading to an autoregressive fractionally integrated moving average (ARFIMA) model (Granger, 1980; Granger and Joyeux, 1980; Hosking, 1981).[1] The discrete-time ARFIMA model is empirically more popular than the fBm-based continuous-time processes due to its simple expression for the spectral density and the fact that we only have discrete observations in practice. See Graves et al. (2017) for an interesting survey that compares the two classes of models and their history.

Nevertheless, in recent years, fGn-based continuous-time models have received much attention in mathematical finance, financial engineering, and financial econometrics. In a seminar paper by Gatheral et al. (2018), the log volatility surface is found to be well fitted by a rough fBm process (i.e., $H \approx 0.14$). Several other studies fit the fBm or fOU models to actual realized volatility data and find that volatility is rough (i.e., $H < 0.5$) (Wang et al., 2021; Fukasawa et al., 2022; Bolko et al., 2022; Wang et al., 2022). Moreover, the rough fGn-based models are shown to provide better out-of-sample forecasting results for volatility than some popular discrete-time models such as the HAR model of Corsi (2009). See Gatheral et al. (2018) and Wang et al. (2021). The rough fGn-based models have found a wide range of financial applications, including in option pricing (Livieri et al., 2018; Bayer et al., 2016; Garnier and Sølna, 2018), variance swaps (Bayer et al., 2016), portfolio choice (Fouque and Hu, 2018), and dynamic hedging (El Euch et al., 2018). The Rough Volatility website contains more than 200 papers in this rapidly growing literature.[2] Many applications rely critically on the mathematical tractability in the adopted continuous-time framework.

This paper investigates the finite sample performance of several alternative estimation methods for H in fGn (the increment of fBm) from discrete observations. Unlike the memory parameter d in ARFIMA models where finite sample performance of various estimation methods is well understood,[3] a comprehensive study comparing the finite sample performance of estimation approaches for H, especially when H takes values in the interval $(0, 0.3)$, is yet to be conducted. The estimation methods considered include the log periodogram regression method (Geweke and Porter-Hudak, 1983, GPH), the local Whittle (LW) method (Robinson, 1995a; Künsch, 1987), the time-domain maximum likelihood (TDML) method, the Whittle (frequency domain) maximum likelihood (ML)

[1] Tanaka (2013) and Wang and Yu (2022) show that the ARFIMA process converges weakly to fOU (or fBm) when the autoregressive coefficient is local-to-unity (or unity) and the number of discrete-time observations goes to infinity. Despite their asymptotic similarity, these two classes of processes might perform differently in finite samples.

[2] See `https://sites.google.com/site/roughvol` for a selective list of papers in this growing literature.

[3] See Shi and Yu (2022) for a recent study on the topic.

method (Whittle, 1953, 1954), and the change-of-frequency method (Brouste et al., 2020, CoF).

The two semiparametric methods (GPH and LW) are most often employed to estimate d in ARFIMA and rely on a theoretical feature of its spectral density at the local-to-zero frequency. Since fGn shares a similar feature with fIn, the semiparametric methods can be readily applied to fGn. The log likelihood of TDML involves the determinant and the inverse of the variance-covariance matrix, which can be computationally intensive. Brouste et al. (2020) propose a one-step maximum likelihood (TDML-OS) method to improve the computational speed without compromising its statistical efficiency. Both the original TDML and the TDML-OS methods are implemented in the simulation study. The implementation of the Whittle ML method requires the spectral density. The spectral density of fGn involves an infinite sum that converges extremely slowly when H is close to zero and brings challenges to the computation. We consider two versions of the Whittle ML method. The first version is based on the analytical expression for the spectral density that we derive in this paper, referred to as the exact Whittle ML (EWML) method. The second version is based on Paxson's approximation of the spectral density as in Fukasawa and Takabatake (2019a), referred to as the approximate Whittle ML (AWML) method. Special attention is paid to a highly rough fGn with H close to 0.

Simulation results lead to several interesting findings. First, in terms of root mean square errors (RMSE), the four ML methods (TDML, TDML-OS, EWML, AWML) significantly outperform CoF and the two semiparametric methods, whereas CoF is found superior to the two semiparametric methods. Second, AWML provides almost identical estimates to EWML and outperforms TDML-OS. Third, TDML is more efficient than the two Whittle ML methods when H is close to zero and performs similarly to the Whittle methods when H is larger. Fourth, among the four ML methods, AWML is computationally most efficient and EWML is computationally least efficient. Specifically, the computation time of EWML, TDML, and TDML-OS is about 419, 85, and 21 times of AWML, respectively. In light of the statistical and computational efficiency, we recommend using either TDML or AWML for the estimation of H in fGn.

In the empirical application, we model the log volatility dynamics of ten exchange-traded funds (ETF), including the S&P 500 index ETF and nine industry index ETFs, and 30 Dow Jones stocks with fBm. The estimated self-similar parameters from the TDML, EWML and AWML methods are very close to each other. The volatility of the S&P 500 market index ETF is found to be the smoothest with H around 0.22, while the volatility of industry ETFs and individual stocks are found to be rougher with a lower bound of 0.1. The estimated coefficients from CoF and TDML-OS are higher than those of the ML methods, whereas estimates from the semiparametric are significantly lower. All estimation methods suggest a rough dynamic for all financial assets with $H < 0.5$.

The rest of the paper is organised as follows. Section 2 presents the model specification and

its spectral density. Section 3 discusses various estimation methods for H. Simulation results are detailed in Section 4. Section 5 presents the empirical results. Section 6 concludes. The appendix contains an additional table for the empirical application.

2 Fractional Brownian Motion

Let $B(t)$ be the standard Brownian motion. The fractional Brownian motion process, denoted by $B^H(t)$, is a moving average of past $\mathrm{d}\,B(t)$, which takes the form of

$$B^H(t) = \frac{1}{\Gamma(H+0.5)} \left\{ \int_{-\infty}^{0} \left[(t-s)^{H-0.5} - (-s)^{H-0.5} \right] \mathrm{d}\,B(s) + \int_{0}^{t} (t-s)^{H-0.5}\,\mathrm{d}\,B(s) \right\}.$$

See Mandelbrot and Van Ness (1968).[4] It is self-similar in the sense that $B^H(ct) = c^{H-0.5}B^H(t)$ with c being an arbitrary constant. In other words, the properties of fBm are preserved with respect to scaling in time. The fBm process reduces to the standard Brownian motion when $H = 0.5$.

The fBm process is Gaussian with mean zero and covariance

$$Cov\left(B^H(t), B^H(s)\right) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right), \quad \forall t, s. \tag{1}$$

The increment of fBm is fGn and denoted by y_t. Using discrete time notations, we have

$$y_t = \sigma\left(B^H(t) - B^H(t-1)\right) \text{ and } B^H(t) = B^H(t-1) + \frac{1}{\sigma}y_t.$$

The process $B^H(t)$ can be viewed as a unit root process with an fGn error. The autocovariance of y_t is, $\forall k \geq 0$,

$$\begin{aligned} Cov\left(y_t, y_{t+k}\right) &= \frac{\sigma^2}{2}\left[(k+1)^{2H} + (k-1)^{2H} - 2k^{2H}\right], \\ &\sim \sigma^2 H(2H-1)k^{2H-2} \text{ for large } k, \end{aligned} \tag{2}$$

where the approximation is based on the Taylor expansion. When $H \in (0.5, 1)$, the autocovariances of fGn are not absolutely summable and fGn has a long memory. When $H \in (0, 0.5)$, it is easy to verify that $\forall k \neq 0$, $Cov\left(y_t, y_{t+k}\right) < 0$ and $\sum_{k=-\infty}^{\infty} Cov\left(y_t, y_{t+k}\right) = 0$. Under this setting, fGn is anti-persistent and the corresponding fBm is rough.

Let λ be the spectral frequency. The spectral density of fGn is given by Sinai (1976) and takes the form of

$$f(\lambda) = 2C_H(1-\cos(\lambda)) \sum_{k=-\infty}^{\infty} |2\pi k + \lambda|^{-1-2H} \tag{3}$$

[4] This is also referred to as the Type I fBm. When s is restricted to be between 0 and t (inclusive), it becomes a Type II fBm. See, for example, Comte and Renault (1996).

for $\lambda \in (0, \pi]$, where $C_H = \sigma^2 (2\pi)^{-1} \Gamma(2H+1) \sin(\pi H)$. Note that the spectral density involves an infinite summation, which makes the computation of the spectral density challenging. Computing the spectral density is, however, essential for the implementation of Whittle ML methods. In the next few subsections, we consider various computational methods for the spectral density, including an analytical expression and three approximation methods.

Analytical Expression

According to Fukasawa and Takabatake (2019b), no closed-form expression for the spectral density of fGn is known so far. Here, we provide an analytical expression for $f(\lambda)$. Let

$$\zeta(s, q) := \sum_{j=0}^{\infty} (j+q)^{-s}$$

be Riemann's general zeta or Hurwitz's zeta function. The infinite sum in the spectral density of fGn can be unfolded and expressed using the Hurwitz zeta function as the following:

$$\begin{aligned}
\sum_{j=-\infty}^{\infty} |2\pi j + \lambda|^{-1-2H} &= \sum_{j=-\infty}^{-1} |2\pi j + \lambda|^{-1-2H} + \sum_{j=0}^{\infty} |2\pi j + \lambda|^{-1-2H} \\
&= (2\pi)^{-1-2H} \left(\sum_{j=1}^{\infty} \left| j - \frac{\lambda}{2\pi} \right|^{-1-2H} + \sum_{j=0}^{\infty} \left| j + \frac{\lambda}{2\pi} \right|^{-1-2H} \right) \\
&= (2\pi)^{-1-2H} \left(\sum_{j=0}^{\infty} \left| j + 1 - \frac{\lambda}{2\pi} \right|^{-1-2H} + \sum_{j=0}^{\infty} \left| j + \frac{\lambda}{2\pi} \right|^{-1-2H} \right) \\
&= (2\pi)^{-1-2H} \left[\zeta\left(1+2H, 1 - \frac{\lambda}{2\pi}\right) + \zeta\left(1+2H, \frac{\lambda}{2\pi}\right) \right]. \qquad (4)
\end{aligned}$$

By replacing the infinite sum in (3) with the expression in (4), we obtain the analytical expression which is displayed in Theorem 2.1 below.[5]

Theorem 2.1 *The spectral density of fGn is given by*

$$f(\lambda) = 2C_H(1 - \cos(\lambda)) (2\pi)^{-1-2H} \left[\zeta\left(1+2H, 1 - \frac{\lambda}{2\pi}\right) + \zeta\left(1+2H, \frac{\lambda}{2\pi}\right) \right], \qquad (5)$$

where $C_H = \sigma^2 (2\pi)^{-1} \Gamma(2H+1) \sin(\pi H)$ *and* $\zeta(s, q) = \sum_{j=0}^{\infty} (j+q)^{-s}$ *is the Hurwitz zeta function.*

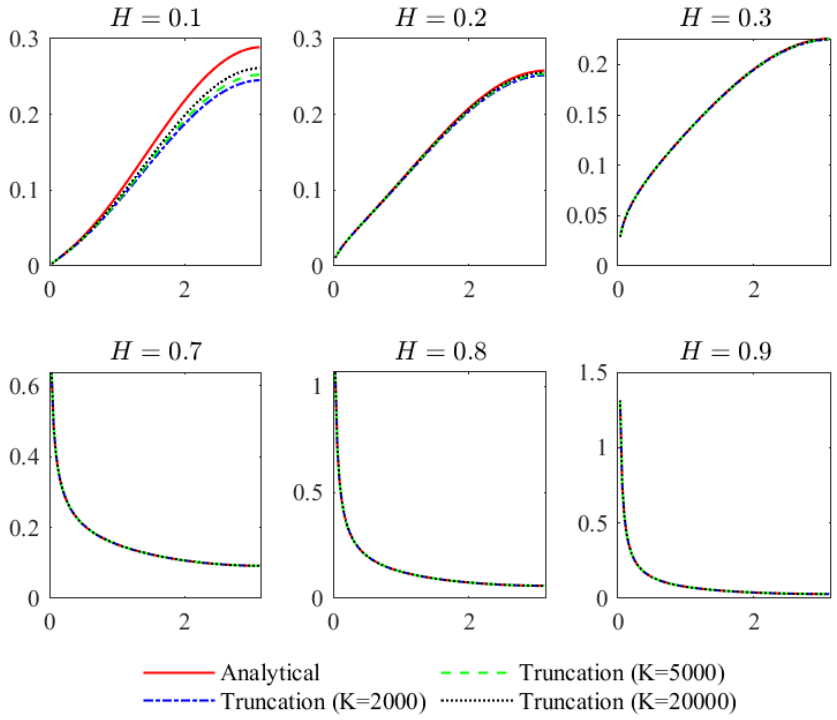
The in-built function *hurwitzZeta* in MATLAB can be employed to compute $\zeta(s, q)$. But unfortunately, the computational speed of $\zeta(s, q)$ and hence $f(\lambda)$ is quite slow which might hamper the widespread applications of the Whittle ML method. Therefore, we search for suitable approximation methods next. The analytical expression (5) is, however, crucial for evaluating the accuracy of various approximation methods.

[5] We follow the traditional wisdom that special functions and cumulative density functions of well-known distributions are regarded as analytical expressions since these functions can be computed up to machine precision and vector operations are possible using standard software.

Truncation Method

It is natural to approximate the infinite sum $\sum_{k=-\infty}^{\infty}$ in (3) by $\sum_{k=-K}^{K}$, with K being the upper bound of the sum. We refer to it as the truncation method. The choice of K depends on the convergence rate of the summand; the slower the convergence, the larger K is required to ensure approximation accuracy. When H is close to zero, the summand $|2\pi k + \lambda|^{-1-2H}$ is almost divergent and an extremely large K is needed. The calculation of the summation and hence the spectral density of fGn under this setting using the truncation method will be computationally costly. If K is not large enough, the computed spectral density might be far from its actual value.

Figure 2: The truncation-based approximation to $f(\lambda)$ with $K = 2000, 5000, 20000$ for $\lambda \in (0, \pi]$.



As an illustration, we plot in Figure 2 the true value of $f(\lambda)$ computed from the analytical expression (5) and the approximated spectral density using the truncation method with $K = 2000, 5000, 20000$ for $\lambda \in (0, \pi]$. We set $H = 0.1, 0.2, 0.3, 0.7, 0.8, 0.9$. It is clear that when $H = 0.1$ all truncation numbers considered cannot lead to a well approximated spectral density. The approximation gets poorer when H moves closer to zero and the spectral frequency λ is higher. Given the empirical relevance of small H values in fGn for log volatility modelling (Gatheral et al., 2018; Fukasawa and Takabatake, 2019a), the large approximation errors are expected to have important implications for estimation methods that rely on the global approximation to $f(\lambda)$ such as the Whittle ML method.

Paxson's Approximation Method

In light of the difficulty described above for the truncation method, we consider an alternative approximation method. The spectral density of fGn can be rewritten as

$$f(\lambda) = 2C_H(1-\cos(\lambda))\left(|\lambda|^{-\gamma_H} + \sum_{j=1}^{\infty} b(j,\lambda)\right),$$

where $\gamma_H = 2H+1$ and $b(j,\lambda) = (2\pi j+\lambda)^{-\gamma_H} + (2\pi j-\lambda)^{-\gamma_H}$. The Paxson approximation of the spectral density (Paxson, 1997; Fukasawa and Takabatake, 2019a) is

$$f(\lambda) \approx 2C_H(1-\cos(\lambda))\left\{|\lambda|^{-\gamma_H} + \sum_{j=1}^{K} b(j,\lambda) + \frac{1}{2}\left[a(K,\lambda)+a(K+1,\lambda)\right]\right\}, \tag{6}$$

where K is a pre-specified integer and

$$a(k,\lambda) = \frac{1}{4\pi H}\left[(2\pi k+\lambda)^{1-\gamma_H} + (2\pi k-\lambda)^{1-\gamma_H}\right].$$

Like the truncation method, the infinite summation $\sum_{j=1}^{\infty} b(j,\lambda)$ is replaced with a truncated quantity $\sum_{j=1}^{K} b(j,\lambda)$. The Paxson approximated spectral density, however, has an additional term, $\frac{1}{2}\left[a(K,\lambda)+a(K+1,\lambda)\right]$, which is expected to reduce the approximation error.

Figure 3: Paxson's approximation of the spectral density

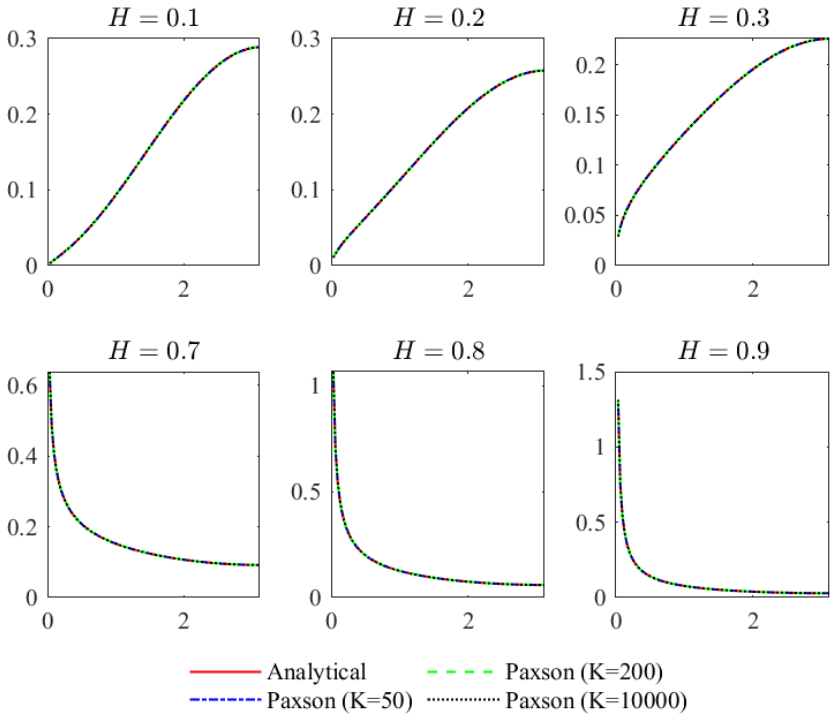


Figure 4: The ratios of the true spectral density to Paxson's approximation

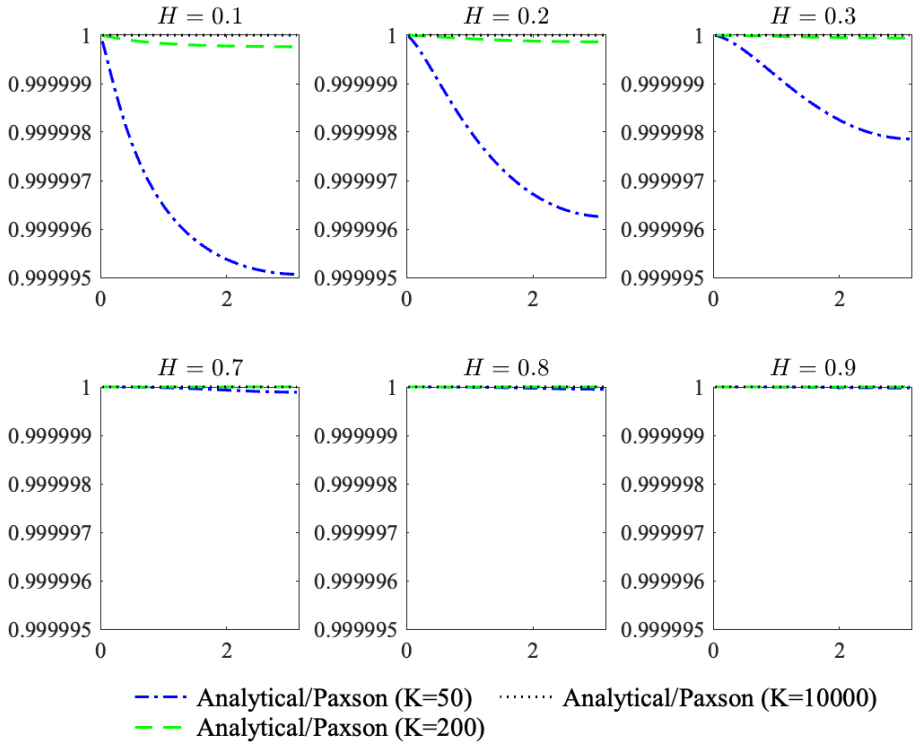


Figure 3 displays the approximated spectral density (Paxson) with $K = 50, 200, 1000$ and the analytical values of $f(\lambda)$ when $H = 0.1, 0.2, 0.3, 0.7, 0.8, 0.9$ and $\lambda \in (0, \pi]$. Figure 4 shows the ratios of the true spectral density $f(\lambda)$ to the approximated ones. Evidently, Paxson's approximation is much more accurate than the truncation-based approximation (see Figure 2), especially when $H = 0.1, 0.2$. The approximation works very well even when K is as small as 50.

Taylor-series Approximation for $\lambda \to 0$

At the near zero frequency (i.e., $\lambda \to 0$), the spectral density can be approximated by

$$f(\lambda) \sim C_H \lambda^{1-2H}, \tag{7}$$

using the Taylor expansion and L'Hopital's rule. This is compared to the spectral density of fIn(d) with $d = H - 0.5$ or ARFIMA$(0, H - 0.5, 0)$:[6]

$$\tilde{f}(\lambda) = \frac{\sigma^2}{2\pi} \left(2 - 2\cos(\lambda)\right)^{-(H-0.5)} \sim \frac{\sigma^2}{2\pi} \lambda^{1-2H}, \quad \text{as} \quad \lambda \to 0, \tag{8}$$

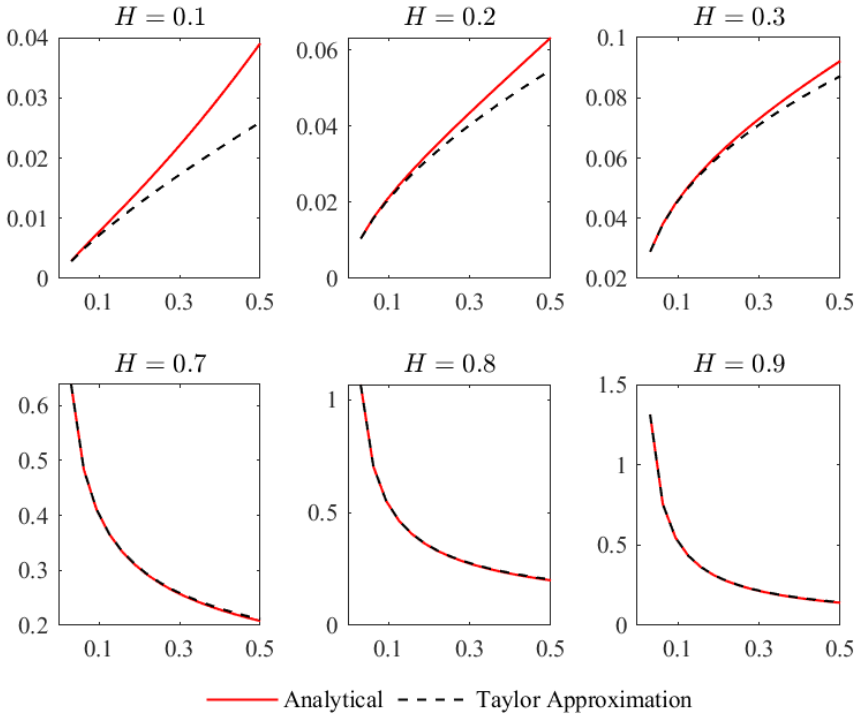
by the Taylor approximation. Both spectral densities converge to zero at a rate of $O(\lambda^{1-2H})$ as $\lambda \to 0$. Despite the similarity, the two spectral densities take very different forms and behave

[6]The fIn(d) process is defined as $u_t = (1 - L)^{-d} e_t$ where L is the lag operator, $e_t \overset{iid}{\sim} (0, \sigma^2)$ with $\sigma^2 > 0$, and $d \in (-0.5, 0.5)$.

differently when the frequency is away from zero; see a detailed discussion on this point in Section 2 of Geweke and Porter-Hudak (1983).

We now examine the accuracy of the Taylor approximation (7) of fGn at the near zero frequency. Figure 5 plots, in the black dash line, the local-to-zero-frequency approximation for $\lambda \in (0, 0.5]$ and the red solid line is the corresponding analytical spectral density. The Taylor expansion general works well when $\lambda \to 0$. The approximation accuracy, however, deteriorates as H becomes smaller. For example, when $H = 0.1$, it cannot well approximate the true spectral density even when λ is as low as 0.1. This suggests that estimation methods that rely on the local-to-zero-frequency approximation (e.g., the semiparametric methods which will be introduced in the next section) are expected to perform poorly when H is small. Once again, given the empirical relevance of small H values in the volatility modelling literature, the approximation errors are expected to have important implications for the semiparametric methods.

Figure 5: The Taylor approximation of the spectral density



3 Estimation Methods

We introduce the two semiparametric methods (GPH and LW), the time-domain and frequency domain ML methods, and the CoF approach in this section. Both semiparametric methods are based on the Taylor approximation of the spectral density, i.e., $f(\lambda) \sim C_H \lambda^{1-2H}$.

3.1 GPH method

The estimation method proposed by Geweke and Porter-Hudak (1983) is as follows. By taking log of both sides of (7), we have

$$\log f(\lambda) = \log\left(C_H\right) + (1-2H)\log(\lambda). \tag{9}$$

Simple re-organisation of (9) leads to the following regression model:

$$\log I(\lambda_j) = \log\left(C_H\right) + (1-2H)\log(\lambda) + \varepsilon_j,$$

where $\varepsilon_j = \ln I(\lambda_j) - \ln f(\lambda)$. Let $z_j = \ln(\lambda)$. The ordinary least squares estimator of the coefficient $1-2H$ is

$$\hat{\beta}_{GPH} = \frac{\sum_{j=1}^{m}(z_j - \bar{z})\ln I(\lambda_j)}{\sum_{j=1}^{m}(z_j-\bar{z})^2},$$

where $\bar{z} = m^{-1}\sum_{j=1}^{m} z_j$ and $I\left(\lambda_j\right)$ is the periodogram at the j^{th} Fourier frequency $\lambda_j = 2\pi j/T$ (with $j = 1, 2, \ldots, m$) defined as

$$I\left(\lambda_j\right) = \frac{1}{2\pi T}\left|\sum_{t=0}^{T} y_t \exp\left(-it\lambda_j\right)\right|^2. \tag{10}$$

Note that m represents the number of spectral frequencies (i.e., sample size) employed in the regression and is a tuning parameter chosen by the user. The GPH estimator of H is obtained as

$$\hat{H}_{GPH} = \frac{1}{2}\left(1 - \hat{\beta}_{GPH}\right).$$

Under the settings of $m \to \infty$ and $m/T \to 0$, the GPH estimator converges to the true value H at rate $O_p(\sqrt{m})$ and has a limiting distribution of the following:

$$\sqrt{m}\left(\hat{H}_{GPH} - H\right) \xrightarrow{d} N(0, \pi^2/6) \text{ with } \pi^2/6 = 1.645. \tag{11}$$

See Robinson (1995b). Assumption $m/T \to 0$ implies that $\max_j\{\lambda_j\} = 2\pi m/T \to 0$. That is, one should only use near zero frequencies to estimate H.

From Figure 5, the Taylor approximation does not work well when H is small and λ is not near zero. Therefore, we expect GPH to suffer from a large bias when m is chosen to be a large number in finite samples. On the other hand, if the sample size m is chosen to be a very small number, the variance of the GPH estimator will increase substantially. Andrews and Guggenberger (2003) give the asymptotic RMSE optimal choice of the tuning parameter, that is, $m = T^{0.8}$. However, in the ARFIMA literature, smaller values of m (e.g., $m = T^{\delta}$ with $\delta = 0.5, 0.6, 0.65, 0.7$) are often used to reduce biases (Agiakloglou et al., 1993; Smith et al., 1997; Nielsen and Frederiksen, 2005; Nadarajah et al., 2021).

3.2 Local Whittle method

The local Whittle method proposed by Künsch (1987) and Robinson (1995a) solves the following optimisation problem:

$$\begin{aligned}\left(\hat{C}_{LW}, \widehat{H}_{LW}\right) &= \arg\max_{C,H} \frac{1}{m}\sum_{j=1}^{m}\left[-\log f\left(\lambda_j\right) - \frac{I\left(\lambda_j\right)}{f\left(\lambda_j\right)}\right] \\ &= \arg\max_{C_H,H} \frac{1}{m}\sum_{j=1}^{m}\left[-\log C_H + (2H-1)\log\lambda_j - \frac{1}{C_H}\lambda^{2H-1}I(\lambda_j)\right], \end{aligned} \tag{12}$$

where the spectral density $f\left(\lambda_j\right)$ is replaced by its Taylor approximation. The analytical solution of LWE is

$$\begin{aligned}\hat{C}_{LW} &= \frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2H-1}I(\lambda_j), \\ \widehat{H}_{LW} &= \arg\max_{H}\left[-\log\left(\hat{C}_{LW}\right) + (2H-1)\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j\right].\end{aligned}$$

Under the settings of $m \to \infty$ and $m/T \to 0$, the local Whittle estimator is a consistent estimator of H and has the following limiting distribution (Robinson, 1995a):

$$\sqrt{m}\left(\hat{H}_{LW} - H\right) \xrightarrow{d} N(0, 1/4).$$

Clearly, the LW estimator is more efficient than the GPH estimator (0.25 versus 1.645 for the asymptotic variance). Both GPH and LW rely on the Taylor series approximation, which was shown to be inaccurate when H is close to zero and λ is not so close to zero. Therefore, like GPH, we expect LW to suffer from a large bias in finite samples when m is too large. However, if m is too small, a large variance for $\hat{H}_{LW}$ is expected. In the ARFIMA literature, a setting of $m = T^{\delta}$ with $\delta \in [0.5, 0.8]$ is often used; see for example, Nielsen and Frederiksen (2005) and (Giraitis et al., 2012, Chapter 8.7).

3.3 Time-domain ML method

Let $y = (y_1, y_2, \cdots, y_T)'$, $\theta = (H, \sigma)$, and Σ_y be the covariance matrix of y whose elements can be found from (2). Since $y \sim N(\mu, \Sigma_y)$ with $\mu = 0$, the log likelihood function of fGn is

$$\log L(\theta) \propto \frac{1}{2T}\log|\Sigma_y| + \frac{1}{2T}(y-\mu)'\Sigma_y^{-1}(y-\mu). \tag{13}$$

In the case of unknown μ, μ is replaced with a consistent estimator (e.g., the sample mean). We replace $|\Sigma_y|$ with the product of the eigenvalues of Σ_y when computing the likelihood value. The time domain ML estimator is defined as

$$\hat{\theta}_{ML} = \arg\max_{\theta} \log L(\theta).$$

The asymptotic theory of $\hat{\theta}_{ML}$ under the model specification of fGn was studied in Dahlhaus (1989) when $H \in (0.5, 1)$ and Lieberman et al. (2012) when $H \in (0, 0.5)$. Under both settings, the TDML estimator has the limiting property that

$$\sqrt{T}\left(\hat{\theta}_{ML} - \theta\right) \xrightarrow{d} N(0, \Gamma(\theta)^{-1}) \text{ with } \Gamma(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial \log f(\lambda)}{\partial \theta}\left(\frac{\partial \log f(\lambda)}{\partial \theta}\right)' d\lambda \qquad (14)$$

as $T \to \infty$. The TDML estimator is shown asymptotic efficient in the sense of Fisher in Dahlhaus (1989), while the local asymptotic normal (LAN) property of the estimator is obtained in Cohen et al. (2011).

The TDML method is time consuming as it requires computing the inverse matrix and searching for the global maximum of the objective function via a numerical method. To improve the computational speed without compromising asymptotic efficiency, Brouste et al. (2020) propose a one-step time-domain ML method. They consider an in-fill framework where the distance between two consecutive observations (denoted by Δ_n) is assumed to converge to zero and the time period T is fixed.[7] The total number of observations within the sample period is $n = T/\Delta_n$. Let $\tilde{\theta}$ be an initial estimator of θ that is not as efficient (but rate-efficient) as TDML but easy to compute (e.g., the change-of-frequency estimator introduced later). The TDML-OS method updates the initial estimator with the following scoring procedure:

$$\hat{\theta}_{TDML-OS} = \tilde{\theta} + \left(\psi^{-1}(\tilde{\theta})'\Gamma(\tilde{\theta})\psi^{-1}(\tilde{\theta})\right)^{-1} \nabla \log L(\tilde{\theta}),$$

where the rate matrix $\psi(\theta)$ is required to satisfy certain properties and can be taken as

$$\psi(\theta) = \frac{1}{\sqrt{n}}\begin{pmatrix} 1 & 0 \\ -\sigma \log \Delta_n & 1 \end{pmatrix},$$

the Fisher information matrix is

$$\Gamma(\theta) = \begin{pmatrix} \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{\partial \log f(\lambda)}{\partial H}\right)^2 d\lambda & \frac{1}{2\pi\sigma}\int_{-\pi}^{\pi}\frac{\partial \log f(\lambda)}{\partial H}d\lambda \\ \frac{1}{2\pi\sigma}\int_{-\pi}^{\pi}\frac{\partial \log f(\lambda)}{\partial H}d\lambda & 2\sigma^{-2} \end{pmatrix},$$

and $\nabla \log L(\cdot)$ is the derivative of $\log L(\cdot)$.

Under the in-fill asymptotic setting, Brouste et al. (2020) show that the TDML-OS estimator is asymptotically normal and efficient in Fisher's sense such that

$$\psi(\theta)^{-1}\left(\hat{\theta}_{TDML-OS} - \theta\right) \to_d N\left(0, \Gamma^{-1}(\theta)\right).$$

Nevertheless, because the update of the initial estimator is only conducted once, the performance of the TDML-OS estimator is expected to be dependent on the initial estimate $\tilde{\theta}$ and less efficient than the TDML estimator in finite samples.

[7] We set $\Delta_n = 1$ in the simulation study and empirical applications.

Since the TDML and TDML-OS methods do not use information about the spectral density, the approximation errors in calculating the spectral density are irrelevant, unlike the two semi-parametric methods reviewed earlier and the Whittle ML method which will be reviewed next.

3.4 Whittle ML method

To avoid computing Σ_y^{-1} and $|\Sigma_y|$, Whittle (1951) propose to employ the approximations $\Sigma_y^{-1} \approx [a_{jk}]_{j,k=1}^{T}$ and $\log|\Sigma_y| \approx T(2\pi)^{-1}\int_{-\pi}^{\pi}\log f(\lambda)\,\mathrm{d}\lambda$ where $a_{jk} \approx (2\pi)^{-2}\int_{-\pi}^{\pi} f(\lambda)^{-1} e^{i(j-k)\lambda}\,\mathrm{d}\lambda$, leading to the Whittle likelihood function:

$$\log L_W(\theta) = -\frac{1}{m}\sum_{j=1}^{m}\log f(\lambda_j) - \frac{1}{m}\sum_{j=1}^{m}\frac{I(\lambda_j)}{f(\lambda_j)}, \tag{15}$$

where $m = \lfloor T/2 \rfloor$ and $\lambda_j = 2\pi j/T$ with $j = 1, ..., m$. The Whittle estimator $\hat{\theta}_W$ is defined as

$$\hat{\theta}_W = \arg\max_{\theta} \log L_W(\theta).$$

Suppose that $f^*(\lambda)$ is a normalised spectral density that satisfies the property:

$$\int_{-\pi}^{\pi}\log f^*(\lambda)\,\mathrm{d}\lambda = 0.$$

Since $\frac{1}{m}\sum_{j=1}^{m}\log f^*(\lambda_j) \to \int_{-\pi}^{\pi}\log f^*(\lambda)\,d\lambda = 0$ as $T \to \infty$, the Whittle likelihood function can be simplified to

$$\log L_W(\theta) = -\frac{1}{m}\sum_{j=1}^{T}\frac{I(\lambda_j)}{f^*(\lambda_j)}.$$

Note that the spectral density of fIn $\tilde{f}(\lambda)$ in (8) is self-normalised since $\int_{-\pi}^{\pi}\log\left[2-2\cos(\lambda)\right]^{-(H-0.5)}\mathrm{d}\lambda = 0$ (Breidt et al., 1998). The first term in the Whittle likelihood function (15) is therefore often omitted when applying the Whittle ML method to ARFIMA (Cheung and Diebold, 1994). In contrast, the spectral density of fGn is not self-normalised and hence the first term of the Whittle likelihood function $-\frac{1}{m}\sum_{j=1}^{m}\log f(\lambda_j)$ should not be omitted when estimating H in fGn. Unreported simulations reveal that omitting it can cause significant estimation bias.

Under some regularity conditions, Fox and Taqqu (1986) and Giraitis and Surgailis (1990) show that for strongly dependent stationary Gaussian processes,

$$\sqrt{T}\left(\hat{\theta}_W - \theta\right) \xrightarrow{d} N\left(0, 4\pi W^{-1}(\theta)\right),$$

where $W(\theta) = [\omega_{jk}]_{j,k=1}^{2}$, $\omega_{jk} = \int_{-\pi}^{\pi} f^*(\lambda)\frac{\partial^2}{\partial\theta_j\partial\theta_k}\left[f^*(\lambda)\right]^{-1}d\lambda$, and $\theta = (H,\sigma)'$. Dahlhaus (1989) shows that the Whittle ML estimator is asymptotic efficient in the sense of Fisher. More recently,

the LAN property of the Whittle estimator for fGn under the high frequency setting was investigated in a few papers; see, for example, Brouste and Fukasawa (2018); Fukasawa and Takabatake (2019a).

We compute the spectral density of fGn using either the analytical expression provided in Theorem 2.1 (referred to as the exact Whittle ML or EWML) or Paxson's approximation (referred to as the approximate Whittle ML or AWML).

3.5 Change-of-frequency method

The CoF approach (Brouste et al., 2020) estimates H from the quadratic generalised variations and takes the form of

$$\hat{H}_{CoF} = \frac{1}{2}\log_2\left(\frac{\sum_{i=1}^{T-3}\left(y_{(i+2)}+y_{(i+3)}-y_{(i+1)}-y_i\right)^2}{\sum_{i=1}^{T-1}\left(y_{(i+1)}-y_i\right)^2}\right).$$

The limiting properties of $\hat{H}_{CoF}$ for fGn was studied in Brouste et al. (2020).[8] They show that $\hat{H}_{CoF}$ is rate-efficient but less efficient than the ML estimator. This finding was echoed by Wang et al. (2021) where CoF is used to estimate H in fOU. Under the model specification of fGn, the asymptotic distribution of $\hat{H}_{CoF}$ is

$$\sqrt{T}\left(\hat{H}_{CoF}-H\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}}{(2\log 2)^2}\right). \tag{16}$$

where

$$\begin{aligned}
\Sigma_{22} &= 2+4\sum_{j=1}^{\infty}\rho_j^2,\\
\Sigma_{11} &= 2+2^{2-4H}\sum_{j=1}^{\infty}\left(\rho_{j+2}+4\rho_{j+1}+6\rho_j+4\rho_{|j-1|}+\rho_{|j-2|}\right)^2,\\
\Sigma_{12} &= 2^{1-2H}\left(4(\rho_1+1)^2+2\sum_{j=0}^{\infty}\left(\rho_{j+2}+2\rho_{j+1}+\rho_j\right)^2\right),
\end{aligned}$$

with $\rho_j = \frac{1}{2(4-2^{2H})}\left(-|j+2|^{2H}+4|j+1|^{2H}-6|j|^{2H}+4|j-1|^{2H}-|j-2|^{2H}\right)$.

4 Monte Carlo Studies

Many studies have examined the finite sample performance of alternative estimation methods in the context of ARFIMA; see, for example, Agiakloglou et al. (1993); Cheung and Diebold

[8] The CoF approach was also employed by Lang and Roueff (2001) for Gaussian processes with stationary increments and by Barndorff-Nielsen et al. (2013) for Brownian semi-stationary processes when $H \in (0, 0.75)$.

(1994); Smith et al. (1997); Nielsen and Frederiksen (2005); Nadarajah et al. (2021); Shi and Yu (2022) for a somewhat incomplete list. In that literature, it is often found that (1) the Whittle ML method performs as well as the time-domain ML method; (2) the Whittle ML method is computationally much cheaper than the time-domain ML method. See, for example, Tables 1-3 in Nielsen and Frederiksen (2005) for a comparison of alternative estimation methods in estimating d in ARFIMA$(0, d, 0)$.

On the other hand, the list of studies that examine finite sample performance of alternative estimation methods for fGn is much shorter. Brouste and Fukasawa (2018) compare the performance of CoF, TDML and TDML-OS. Fukasawa and Takabatake (2019a) examine the finite sample performance of the Whittle ML, where the spectral density is approximated by Paxson's method. In this section, we provide a comprehensive study investigating the finite sample performance of alternative estimation methods which were reviewed earlier.

Let $\Sigma_y = LL'$ be the Cholesky decomposition of the variance-covariance matrix of y. To simulate fGn, we first simulate a $T \times 1$ vector of independent standard normal variates, denoted by z, then generate the data by $y = Lz$. In our Monte Carlo studies, we set $H = 0.1, 0.2, 0.3, 0.7, 0.8, 0.9$. The parameter σ is set to be one and assumed to be known. The sample size T is $2,000$, and the number of replications is $1,000$. All seven methods, including the two versions of the Whittle ML method, are employed to estimate H from the simulated data, yielding $1,000$ estimates of H for each method. For the AWML method, we set $K = 50$ in Paxson's approximation. For the EWML method, we calculate the Hurwitz zeta function $\zeta(s, q)$ using the MATLAB in-built function *hurwitzZeta*. For the two semiparametric methods, we set the bandwidth parameter $m = T^{0.6}$, which implies an upper bound of $\pi m/T = 0.15$ for the spectral frequency.

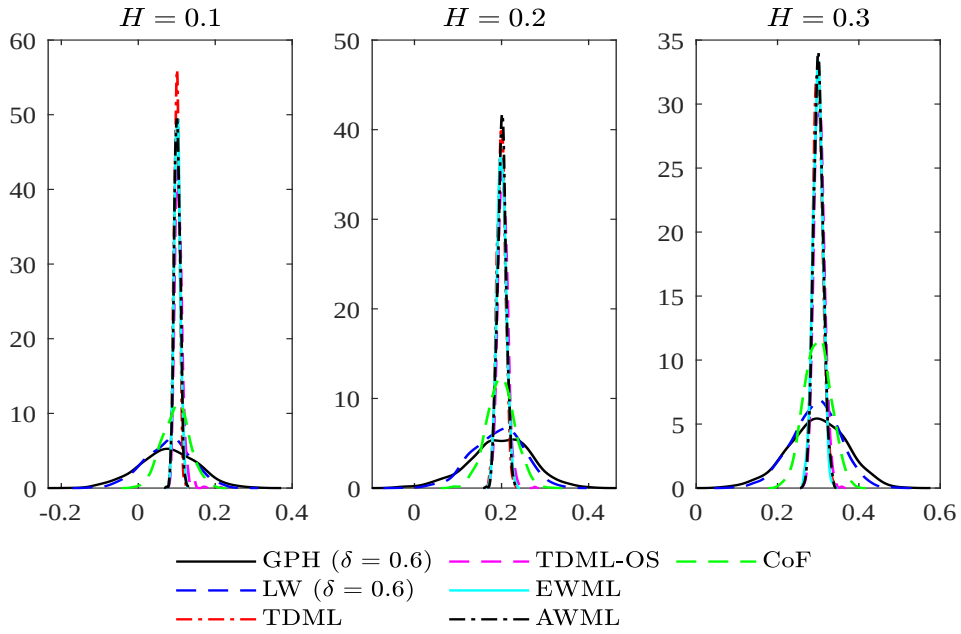
4.1 $H < 0.5$

The setting of $H < 0.5$ is highly empirically relevant for modelling financial market volatility. Figure 6 plots the kernel densities of alternative estimates of H when the true model is fGn with $H = 0.1, 0.2, 0.3$. Table 1 reports the bias, standard error (std), RMSE, and CPU time of the estimation methods.[9] Evidently, the ML methods perform much better than CoF and the two semiparametric methods. The two Whittle methods (EWML and AWML) perform equally well, suggesting that the approximation error of Paxson's method has almost no impact on the estimation outcome. The performance of TDML-OS is inferior to the Whittle methods when H is close to zero.

Between the two semiparametric methods, LW outperforms GPH, as expected. However, both semiparametric methods have a noticeable downward bias (more than 20%) when $H = 0.1$. This

[9]The CPU time is based on MacBook Pro M2 chip, 8-core CPU and 10-core GPU for one iteration and in seconds.

Figure 6: Kernel densities of estimated H



.

Table 1: Bias, standard error, RMSE, and CPU time of alternative estimators when $H = 0.1, 0.2, 0.3$. For GPH and LW, $m = T^{0.6}$. The CPU time is for one iteration and measured in seconds.

H		GPH	LW	TDML	TDML-OS	EWML	AWML	CoF
0.1	Bias	-.0235	-.0233	.0001	.0052	.0013	.0013	-.0023
	Std	.0753	.0613	.0069	.0112	.0077	.0076	.0352
	RMSE	.0789	.0656	.0069	.0123	.0079	.0077	.0353
	CPU time	.0028	.0020	7.0932	1.8447	420.0684	.0830	.0010
0.2	Bias	-.0026	-.0090	-.0009	.0031	.0003	.0007	-.0022
	Std	.0714	.0575	.0097	.0113	.0103	.0097	.0324
	RMSE	.0715	.0582	.0097	.0117	.0103	.0097	.0325
	CPU time	.0058	.0026	6.4728	1.7437	355.7278	.0712	.0017
0.3	Bias	-.0023	-.0065	-.0006	.0027	-.0007	.0006	-.0017
	Std	.0719	.0589	.0119	.0127	.0118	.0118	.0320
	RMSE	.0719	.0593	.0119	.0130	.0119	.0118	.0321
	CPU time	.0045	.0020	5.8709	1.8076	323.8706	.0647	.0010

finding is consistent with our observation from Figure 5 that when $H = 0.1$ the Taylor approximation of the spectral density is not very accurate when $\lambda \geq 0.1$, whereas the spectral frequencies used in the semiparametric methods are between zero and 0.15. The bias of GPH and LW is less

visible when $H = 0.2, 0.3$. None of the five other methods has significant bias.[10]

According to RMSE, when $H = 0.1$, TDML is the best estimation method, outperforming the two Whittle methods by approximately 14.5% and the TDML-OS method by 78%. It overwhelmingly outperforms the rest. The performance of TDML, TDML-OS, and the two Whittle ML methods are similar when $H = 0.2, 0.3$. In terms of computation time, the CoF method is the cheapest to implement, followed by the two semiparametric methods, then by AWML, TDML-OS, TDML, EWML, in this order. In particular, it takes less than one second to implement AWML, two seconds for TDML-OS, seven seconds for TDML, and around seven minutes for EWML. Although TDML-OS was introduced to improve computational efficiency of the time domain ML method, it is slower than AWML. Given the trade-off between accuracy (measured by RMSE) and computational cost (measured by the CPU time), we recommend using either TDML or AWML when H is close to zero.

4.2 $H > 0.5$

We report simulation results for the cases of $H = 0.7, 0.8, 0.9$ in this subsection. Figure 7 displays kernel densities of alternative estimates of H from 1,000 replications when $H = 0.7, 0.8, 0.9$. Summary statistics of those estimates (bias, standard error, RMSE), along with the computation time for each replication, are presented in Table 2.

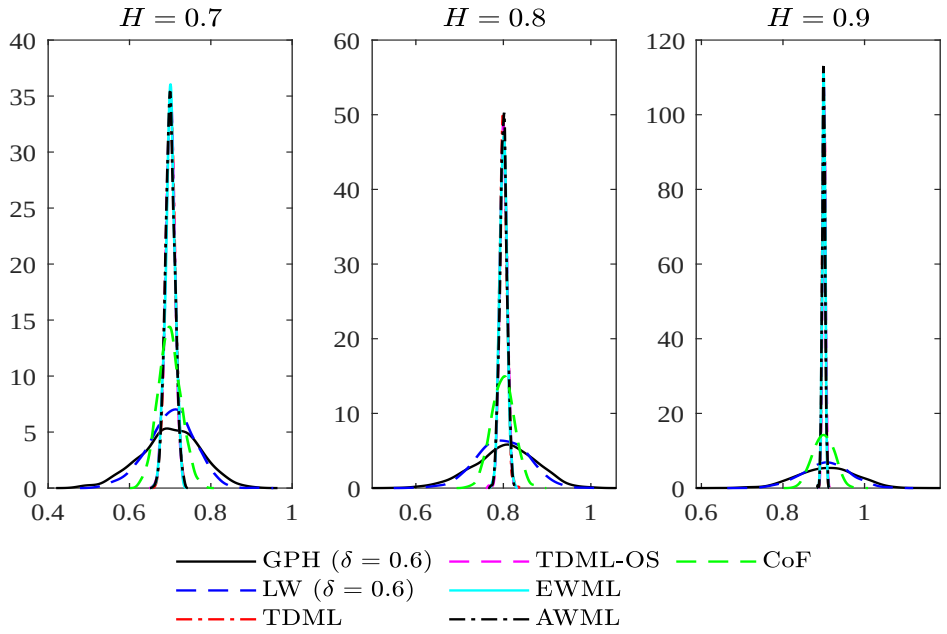
Table 2: Bias, standard error, RMSE, and CPU time of alternative estimators when $H = 0.7, 0.8, 0.9$. For GPH and LW, $m = T^{0.6}$. The CPU time is measured in seconds.

H		LPR	LW	TDML	TDML-OS	EWML	AWML	CoF
0.7	Bias	.0005	.0019	.0005	-.0002	.0001	-.0001	-.0017
	Std	.0735	.0575	.0116	.0117	.0117	.0116	.0278
	RMSE	.0735	.0575	.0116	.0117	.0117	.0116	.0279
	CPU time	.0044	.0022	6.4638	1.7886	357.6512	.0778	.0011
0.8	Bias	.0035	-.0016	-.0002	.0003	-.0002	.0001	-.0014
	Std	.0702	.0576	.0079	.0081	.0081	.0081	.0256
	RMSE	.0704	.0577	.0080	.0081	.0081	.0081	.0257
	CPU time	.0034	.0020	6.9042	1.7913	390.1940	.0831	.0010
0.9	Bias	.0069	.0046	.0001	.0006	-.0000	-.0003	-.0006
	Std	.0696	.0574	.0036	.0037	.0036	.0037	.0257
	RMSE	.0700	.0576	.0036	.0037	.0036	.0037	.0258
	CPU time	.0047	.0027	7.5326	1.7746	392.0987	.0778	.0010

The general conclusions for $H > 0.5$ are the same as those for $H = 0.2, 0.3$. Specifically, the four ML methods have similar performance and significantly outperform the CoF and semiparametric

[10]Interestingly, a slight upward bias is found in TDML-OS and is larger than that of the CoF estimator employed as the initial value.

Figure 7: Kernel densities of estimated H



.

methods in terms of RMSE. Among the four ML methods, AWML is the most computationally efficient, followed by TDML-OS, TDML and EWML, in this order. We recommend using AWML when $H > 0.5$. Since cases of $H = 0.1, 0.2$ are more empirically relevant for modelling log volatility, we focus on these two settings when conducting robustness checks in the following subsection.

4.3 Robustness Checks

We conduct a few robustness checks in this subsection. Section 4.3.1 investigates the impact of an unknown mean on the estimation accuracy of TDML. Section 4.3.2 shows the finite sample performance of the Whittle ML method when the spectral density of fGn is obtained by the truncation method. Section 4.3.3 reports the performance of the semiparametric methods with alternative settings of the tuning parameter m.

4.3.1 Time-domain ML with unknown mean

So far we assume that the population mean is zero and known. In practice, the true population mean μ is unknown. While the semiparametric, CoF, and Whittle ML methods do not require the information on μ, it is needed for the implementation of TDML.

In the context of the ARFIMA$(0, d, 0)$ model, Cheung and Diebold (1994) show that if the population mean is known, TDML is superior to the Whittle ML method. When the mean is

unknown and replaced by the sample mean (often referred to as the plug-in method), they show that the finite sample efficiency gain by TDML over the Whittle ML method is greatly reduced or even lost. The loss in efficiency occurs in the long memory case (i.e., when d close to 0.5) because the precision of the sample mean as an estimate of μ is reduced. Lieberman (2005) argues that although limiting properties of the TDML estimator remain unchanged when the plug-in strategy is employed, the mean estimator introduces a second-order bias to the estimation result.

As an alternative, one can employ the modified profile likelihood (MPL) proposed by Cox and Reid (1987), which makes the parameter of interest orthogonal to nuisance parameters by a linear transformation. An and Bloomfield (1993) and Hauser (1999) show that MPL can reduce a certain degree of estimation bias while keeping the asymptotic distribution of the estimators unchanged for ARFIMA. The MPL is specified as

$$\log L_M(y,H) = \left(\frac{1}{T}-\frac{1}{2}\right)\log|R| - \frac{1}{2}\log\left(l'R^{-1}l\right) + \frac{3-T}{2}\log\left[T^{-1}(y-\hat{\mu}l)'R^{-1}(y-\hat{\mu}l)\right], \quad (17)$$

where $R=\Sigma_y/\sigma^2$, $l=(1,\ldots,1)'$, and $\hat{\mu}=(l'R^{-1}l)^{-1}l'R^{-1}Y$.

Table 3 reports the bias, standard error, and RMSE of TDML (assuming known μ), TDML (plug-in) and MPL under the DGP of fGn. Both the plug-in method and MPL work well in the case of unknown population mean, with some subtle differences. Not surprisingly, the infeasible TDML (with true μ) has the smallest RMSE. Consistent with the finding of An and Bloomfield (1993) and Hauser (1999) for ARFIMA models, the MPL estimator has a smaller bias than the plug-in estimator. The plug-in method, however, has a smaller RMSE than MPL. This is expected, because, when $H=0.1$ the process is anti-persistent, the sample mean can estimate μ precisely.

Table 3: Bias, standard error, and RMSE of TDML, plug-in TDML and MPL.

H		TDML	TDML (plug-in)	MPL
0.1	Bias	.0001	-.0003	-.0001
	Std	.0069	.0073	.0084
	RMSE	.0069	.0073	.0084
0.2	Bias	-.0009	-.0015	.0005
	Std	.0097	.0098	.0107
	RMSE	.0097	.0100	.0107

4.3.2 Whittle ML based on truncation

For the Whittle ML method, instead of using Paxson's approximation of the spectral density as in AWML, one could use the truncation-based method (see Section 2). Table 4 reports the bias, standard error, RMSE, and CPU time of the truncation-based Whittle ML method with $K=2000, 5000, 20000$ when $H=0.1, 0.2$.

Results for EWML and AWML are duplicated here for ease of comparison. When $H = 0.1$, the truncation method leads to a noticeable upward bias under all three settings of K. Although the bias decreases as K increases, it is still around 9% when $K = 20000$. This is consistent with our observation in Figure 2 that when $H = 0.1$ the gap between the truncation-based approximation of the spectral density and the true spectral density remains sizeable even with $K = 20000$. The inaccurate approximation of the spectral density leads to biased estimates of H.

Table 4: Bias, standard error, RMSE and CPU time of EWML, AWML, and the Whittle ML method based on the truncation method with $K = 2000, 5000, 20000$. The CPU time is measured in seconds.

H		EWML	AWML	Truncation-based Whittle ML		
				$K = 2000$	$K = 5000$	$K = 20000$
0.1	Bias	.0013	.0013	.0148	.0120	.0088
	Std	.0077	.0076	.0059	.0061	.0063
	RMSE	.0079	.0077	.0160	.0135	.0109
	CPU time	649.3975	.0830	4.4394	10.9951	42.6732
0.2	Bias	.0003	.0007	.0005	.0003	.0000
	Std	.0103	.0097	.0098	.0099	.0099
	RMSE	.0103	.0097	.0098	.0099	.0099
	CPU time	487.6717	.0712	4.0965	10.2277	39.8513

Interestingly, the standard errors of the three truncation-based estimates are slightly smaller than that of the exact counterpart and AWML. However, due to the significant biases, the truncation-based methods have larger RMSEs than EWML and AWML. Moreover, as expected, the computational cost increases substantially as K increases and is much higher than that of AWML (based on Paxson's approximation).

When $H = 0.2$, the bias of the truncation-based Whittle method becomes negligible. The standard errors and the RMSEs of the three truncation methods are comparable to those of EWML and AWML. Once again, this finding is consistent with that in Figure 2. Although we have not carried out a simulation study for larger values of H, we expect the three truncation methods to perform similarly to the case of $H = 0.2$.

4.3.3 Semiparametric methods with alternative tuning parameters

In this subsection, we consider several alternative settings of the bandwidth parameter $m = T^{\delta}$ for the two semiparametric methods: $\delta = 0.5, 0.6, 0.7, 0.8$. Recall that the semiparametric methods are based on the Taylor approximation of the spectral density at the near-zero frequency (see Section 2) . The larger the approximation error, the larger the estimation bias is expected. The bandwidth parameter m regulates the upper bound of the near-zero frequencies, which are $0.07, 0.15, 0.32, 0.69$,

respectively, when $m = T^{0.5}, m = T^{0.6}, m = T^{0.7}, m = T^{0.8}$. When a larger value of m is chosen, the Taylor approximation becomes less accurate (see Figure 5) but the sample size increases. Consequently, a larger bias and a smaller variance are expected.

Table 5 reports the bias, standard deviations, and RMSE of GPH and LW under the various settings of m when $H = 0.1, 0.2$. The number in boldface corresponds to the smallest RMSE value. When $H = 0.1$, the bias of both methods rises with δ and becomes noticeable when $\delta \geq 0.6$. For the case of $H = 0.2$, the biases of the two semiparametric methods are small when $\delta = 0.5, 0.6$ but grow as δ becomes larger. These findings are consistent with our observation in Figure 5 where the Taylor approximation does not work well for $\lambda \geq 0.1$ when $H = 0.1$ and for $\lambda \geq 0.2$ when $H = 0.2$.

Table 5: Bias, standard error and RMSE of GPH and LW for alternative choices of δ when $H = 0.1, 0.2$. The number in boldface corresponds to the smallest RMSE value.

H		GPH				LW			
	δ	0.5	0.6	0.7	0.8	0.5	0.6	0.7	0.8
0.1	Bias	.0088	-.0235	-.0606	-.1177	.0020	-.0233	-.0597	-.1118
	Std	.1119	.0753	.0490	.0333	.0980	.0613	.0404	.0267
	RMSE	.1123	.0789	**.0780**	.1224	.0981	**.0656**	.0720	.1150
0.2	Bias	.0085	-.0026	-.0220	-.0542	-.0100	-.0090	-.0268	-.0551
	Std	.1117	.0714	.0470	.0314	.0933	.0575	.0372	.0249
	RMSE	.1121	.0715	**.0519**	.0626	.0939	.0582	**.0458**	.0604

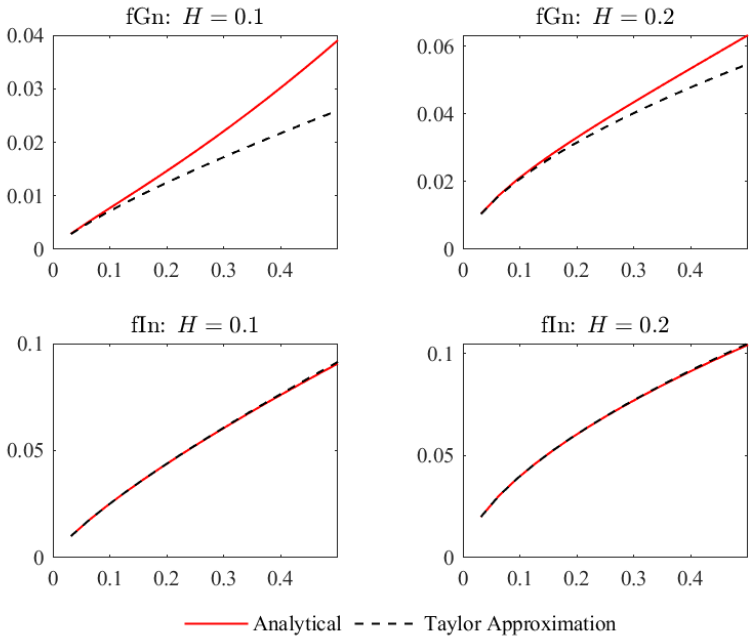
While the large bias is not surprising in light of Figure 5, it is remarkably different from what has been reported in the ARFIMA literature. For example, Nielsen and Frederiksen (2005) did not report any visible bias for the two semiparametric methods when $m = T^{0.5}, T^{0.6}$. To see why the bias is more an issue for fGn than fIn, Figure 8 compares the Taylor approximated and the true spectral densities, with the top row corresponding to fGn and the bottom row for fIn. Clearly, the Taylor approximation performs much better for fIn than for fGn, at least when $H = 0.1, 0.2$. This explains why the large bias problem has not been seen in the fIn literature.

As expected, the standard errors of both methods decline as the sample size m expands. The selection of m represents a trade-off between bias and standard error. From Table 5, the smallest RMSE value takes place when $\delta = 0.6$ or 0.7. This explains why we set $\delta = 0.6$ in our main simulation design.

5 Modelling Log Realised Volatility

Volatility modelling has been a subject of significant interest over many decades. In the more recent continuous-time finance literature, log volatilities are found to follow a rough fBm or fOU process

Figure 8: The Taylor approximation and the true spectral densities of fGn and fIn when $\lambda \in (0, 0.5]$.



using various calibration or estimation methods. Although the volatility dynamic is generally agreed to be rough (i.e., $H < 0.5$) in the continuous-time literature, the exact degree of roughness (i.e., magnitude of H) varies depending on the financial assets and estimation methods employed (Gatheral et al., 2018; Fukasawa and Takabatake, 2019a; Wang et al., 2021; Bolko et al., 2022; Bennedsen et al., 2021).

In this section, we compare estimates of H from alternative methods for log volatility using a comprehensive dataset. The log volatilities are modelled as an fBm process $\sigma B^H(t)$. While fOU is a more general model specification, the additional drift component in fOU is often found negligible (Wang et al., 2021; Bolko et al., 2022).[11] We consider seven different estimation techniques (GPH, LW, TDML, TDML-OS, EWML, AWML, and CoF) whose finite sample performance has been studied carefully in our simulations. The bandwidth parameter of the semiparametric methods m is set to be $T^{0.6}$. For TDML and TDML-OS, we assume the population mean is zero and known. However, similar results are obtained from the TDML (plug-in) method and not reported here for brevity.

The data examined are the daily realised volatility of the S&P 500 index ETF, nine industry

[11] In the context of ARFIMA$(1, d, 0)$, Li et al. (2022) raise a weak identification issue between a long memory model (with the autoregressive coefficient α close to zero and $d > 0.5$) and a rough model ($\alpha \to 1$ and $d < 0.5$). It is unclear whether the weak identification also applies to fOU. An investigation of this question involves a careful study of the spectral density of fOU, which is quite complicated and hence left for a separate paper.

index ETF, and Dow Jones 30 (DJ30) stocks[12] from 2012 to 2019, provided by Risk Lab.[13] The sample spans over a long period (8 years) and is chosen to avoid significant financial market interruptions such as the subprime mortgage crisis and the financial turbulence at the onset of Covid-19 (March 2020). The estimation techniques are applied to the first differenced log RV, which is assumed to follow an fGn process, i.e., $\sigma(B^H(t) - B^H(t-1))$. Summary statistics of the log RVs and their first differences are provided in the Appendix. Table 6 reports the estimated H from various approaches, their asymptotic standard errors, and the computation time required for the ten index EFTs. Estimation results of the DJ30 stocks are provided in Table 7. The computation time for DJ30 stocks is similar to those of the ETFs and hence omitted for brevity.

Table 6: Estimation results: S&P 500 and industry ETFs. We set $K = 50$ for AWML and $m = T^{0.6}$ for GPH and LW. CPU time is measured in seconds.

ETF	H	GPH	LW	TDML	TDML-OS	EWML	AWML	CoF
SPY	Estimate	-.026	.041	.216	.243	.217	.217	.287
	Std. error	.131	.051	.011	.011	.011	.011	.032
	CPU time	.014	.036	29.643	3.483	1653.022	.324	.003
XLB	Estimate	.043	.064	.154	.197	.155	.155	.250
	Std. error	.132	.051	.009	.011	.009	.009	.033
	CPU time	.004	.008	37.191	3.184	2013.004	.390	.001
XLE	Estimate	.115	.128	.196	.270	.197	.197	.339
	Std. error	.132	.051	.010	.011	.011	.011	.032
	CPU time	.001	.006	33.512	3.078	1986.599	.374	.000
XLF	Estimate	.040	.031	.169	.200	.171	.171	.245
	Std. error	.132	.051	.010	.011	.010	.010	.033
	CPU time	.000	.005	33.743	3.083	1826.552	.367	.000
XLI	Estimate	.027	.043	.174	.198	.175	.176	.237
	Std. error	.132	.051	.010	.011	.010	.010	.033
	CPU time	.006	.010	34.901	2.953	1795.398	.358	.002
XLK	Estimate	-.076	-.022	.182	.211	.183	.183	.258
	Std error	.132	.051	.010	.011	.010	.010	.033
	CPU time	.001	.004	33.000	2.993	1769.932	.327	.001
XLP	Estimate	-.158	-.002	.145	.151	.146	.146	.167
	Std. error	.1318	.051	.009	.011	.009	.009	.034
	CPU time	.000	.002	32.862	2.917	1923.232	.366	.000
XLU	Estimate	.120	.090	.135	.149	.142	.142	.173
	Std. error	.132	.051	.009	.011	.009	.009	.034
	CPU time	.000	.001	44.408	2.931	2396.946	.441	.000
XLV	Estimate	.013	.031	.176	.193	.177	.177	.225
	Std. error	.132	.051	.010	.011	.010	.010	.033
	CPU time	.000	.002	33.955	2.942	1757.927	.365	.000
XLY	Estimate	-.017	.048	.172	.194	.173	.173	.232
	Std. error	.132	.051	.010	.011	.010	.010	.033
	CPU time	.000	.002	30.278	2.907	1692.155	.330	.000

[12]We replace Dow Inc. (NYSE: DOW) with Exxon Mobil Co. (NYSE: XOM) as in Li et al. (2022), because DOW has a small number of observations available within the sample period .

[13]See https://dachxiu.chicagobooth.edu/#risklab.

Table 7: Estimated H and standard errors (in parentheses): DJ30 individual stocks. We set $K = 50$ for AWML and $m = T^{0.6}$ for GPH and LW.

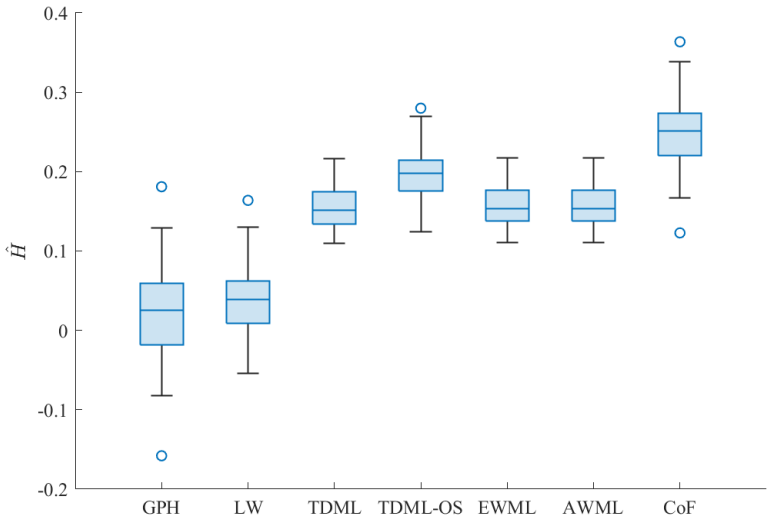
Ticker	GPH	LW	TDML	TDML-OS	EWML	AWML	CoF
AAPL	.066 (.132)	.020 (.052)	.194 (.012)	.280 (.011)	.195 (.012)	.195 (.012)	.364 (.032)
ALD	-.040 (.132)	.019 (.052)	.152 (.011)	.190 (.011)	.153 (.011)	.153 (.011)	.241 (.033)
AMGN	.071 (.132)	.030 (.052)	.144 (.010)	.173 (.011)	.145 (.010)	.145 (.010)	.216 (.034)
AXP	.006 (.132)	.039 (.052)	.133 (.010)	.179 (.011)	.134 (.010)	.134 (.010)	.235 (.033)
BA	-.021 (.132)	-.042 (.052)	.153 (.011)	.232 (.011)	.154 (.011)	.154 (.011)	.310 (.033)
BEL	-.007 (.133)	-.049 (.052)	.123 (.010)	.150 (.011)	.124 (.010)	.124 (.010)	.187 (.034)
CAT	.181 (.132)	.164 (.052)	.130 (.010)	.193 (.011)	.131 (.010)	.131 (.010)	.259 (.033)
CHV	.129 (.132)	.130 (.052)	.187 (.011)	.202 (.011)	.188 (.011)	.188 (.011)	.225 (.034)
CRM	.012 (.132)	.040 (.052)	.166 (.012)	.238 (.011)	.167 (.012)	.167 (.012)	.311 (.033)
CSCO	.043 (.132)	.060 (.052)	.165 (.011)	.218 (.011)	.167 (.011)	.167 (.011)	.275 (.033)
DIS	-.019 (.132)	-.054 (.052)	.156 (.011)	.200 (.011)	.158 (.011)	.158 (.011)	.251 (.033)
GS	.095 (.132)	.041 (.052)	.187 (.012)	.241 (.011)	.188 (.012)	.188 (.012)	.297 (.033)
HD	.033 (.132)	.043 (.052)	.148 (.010)	.195 (.011)	.150 (.011)	.150 (.011)	.250 (.033)
IBM	.069 (.132)	.096 (.052)	.148 (.011)	.220 (.011)	.150 (.011)	.150 (.011)	.291 (.033)
INTC	.023 (.132)	.047 (.052)	.167 (.012)	.208 (.011)	.168 (.012)	.168 (.012)	.256 (.033)
JNJ	.126 (.132)	.115 (.052)	.127 (.009)	.124 (.011)	.131 (.009)	.131 (.009)	.123 (.035)
JPM	.009 (.132)	.016 (.052)	.181 (.012)	.216 (.011)	.181 (.012)	.182 (.012)	.261 (.033)
KO	.022 (.132)	.044 (.052)	.140 (.010)	.171 (.011)	.142 (.010)	.142 (.010)	.211 (.034)
MCD	.052 (.132)	.077 (.052)	.122 (.010)	.210 (.011)	.126 (.010)	.127 (.010)	.288 (.033)
MMM	.088 (.132)	.075 (.052)	.151 (.011)	.198 (.011)	.153 (.011)	.153 (.011)	.257 (.033)
MRK	-.014 (.132)	.007 (.052)	.130 (.010)	.162 (.011)	.130 (.010)	.130 (.010)	.206 (.034)
MSFT	-.082 (.132)	-.003 (.052)	.190 (.012)	.237 (.011)	.191 (.012)	.191 (.012)	.292 (.033)
NIKE	.039 (.132)	.020 (.052)	.140 (.011)	.203 (.011)	.142 (.011)	.143 (.011)	.266 (.033)
PG	-.023 (.132)	.003 (.052)	.141 (.010)	.156 (.011)	.141 (.010)	.141 (.010)	.182 (.034)
SPC	-.077 (.133)	-.036 (.052)	.110 (.009)	.135 (.011)	.111 (.009)	.111 (.009)	.174 (.034)
UNH	.032 (.133)	.020 (.052)	.121 (.010)	.158 (.011)	.125 (.010)	.125 (.010)	.206 (.034)
V	.046 (.133)	.078 (.052)	.142 (.010)	.179 (.011)	.144 (.010)	.144 (.010)	.226 (.034)
WAG	-.020 (.133)	-.048 (.052)	.133 (.011)	.202 (.011)	.134 (.011)	.134 (.011)	.272 (.033)
WMT	.038 (.133)	.011 (.052)	.122 (.010)	.193 (.011)	.123 (.010)	.123 (.010)	.263 (.033)
XOM	.012 (.132)	.051 (.051)	.177 (.011)	.212 (.011)	.178 (.011)	.178 (.011)	.253 (.033)

The three outperforming ML methods (TDML, EWML, and AWML) always yield similar estimates with small standard errors, which is consistent with our simulation results. The estimated H is the largest for the S&P 500 market index ETF (approximately 0.22), implying a smoother dynamic. Volatility dynamics are believed to be driven by information flow (Clark, 1973; Tauchen and Pitts, 1983; Andersen, 1996). The market index volatility reflects the aggregated information flow, whereas the volatility dynamics of industry ETF (individual stock) are driven by both the market-wide information and industry (industry and company) specific information. It is, therefore, reasonable to see that the industry or individual stock volatilities are noisier or rougher (i.e., smaller H). The estimated H for the industry ETFs ranges between 0.13 to 0.19, with the dynamics of the utility sector ETF (XLU) being the roughest. Results for the DJ30 stocks are similar to those of the industry ETFs. The estimated H from the three ML methods falls between 0.12 and

0.2. It is worth mentioning that the computation time using the EWML (TDML) is more than $5,000$ (90) times of the AWML in some cases. Our estimation results provide further evidence supporting the rough volatility argument.

Estimation results from the two semiparametric, TDML-OS, and CoF methods are substantially different from those of the TDML, EWML, and AWML methods. Figure 9 displays summary statistics of the 39 estimated Hs (across the nine industry ETFs and 30 individual stocks) from the seven different estimation methods using box plots. One can see that the semiparametric estimates are systematically lower than those of the ML methods, with only one exception (CAT). In contrast, the TDML-OS and CoF estimates are almost always higher than those of the three outperforming ML methods (with the exception of JNJ), with the TDML-OS estimates lying between the TDML and the CoF estimates. Since those financial assets are subject to common macroeconomics and market factors, the volatilities of those assets are highly correlated. As such, it is not surprising to observe a pattern of systematic under- and over-estimation of the H parameter from the semiparametric and CoF methods, respectively.

Figure 9: Summary statistics of the estimated H across the nine industry ETF and DJ30 stocks



6 Conclusion

In this paper, we derive the analytical expression for the spectral density of the fGn process. The analytical expression allows us to examine the approximation errors introduced by alternative methods, including Paxson's method and the truncation method. It also allows us to compare the statistical performance of the Whittle ML methods based on the analytical expression and that based on the alternative approximation methods.

We compare the finite sample performance of several estimation methods for the self-similar parameter H in fGn, including two semiparametric methods, two time-domain ML methods (TDML and TDML-OS), two Whittle ML methods (EWML and AWML), and the CoF method. EWML is based on the analytical expression of the spectral density, while AWML is based on Paxson's approximation. We find that, in terms of RMSE, the four ML methods significantly outperform the CoF and two semiparametric methods. Among the four ML methods, TDML is the best performing method, while the performance of AWML and EWML are almost identical to each other and superior to TDML-OS. Specifically, when H is close to zero, TDML is more efficient than the two Whittle ML methods and TDML-OS. When H takes larger values, the two Whittle ML and TDML-OS methods yield estimates very close to TDML. However, there is a vast difference in computational cost among the four ML methods. The AWML method is computationally most efficient, followed by TDML-OS and then TDML. The EWML is very computationally intensive compared with the other three ML methods. Taking into account of the trade-off between the statistical efficiency and computational efficiency and the empirical relevance of small H, we recommend using AWML and TDML to estimate parameters in fGn.

The fBm specification is applied to the log realized volatility of 40 financial assets (including ten index ETFs and 30 Dow Jones stocks running from 2012 to 2019) and is estimated with six different methods. While all estimation methods suggest a rough volatility dynamic, the degree of roughness implied by the alternative methods varies substantially, emphasising the importance of understanding the finite sample performance of different estimation methods. Estimates of the three outperforming ML methods (TDML, EWML, and AWML) are similar for all assets and fall between 0.1 and 0.22 (with an average of 0.15). In contrast, the two semiparametric methods suggest that the self-similar parameter H is even closer to zero and could take negative values. The CoF and TDML-OS methods, on the other hand, find that the averaged values of H (across all assets considered) are 0.25 and 0.20, respectively.

The fOU specification is a generalisation to fBm and allows for short-run dynamics. Although the additional drift term in fOU is often found negligible for log RV, one might still want to employ this more general model specification. The semiparametric methods that only require information on the spectral density in the local-to-zero frequencies are, to some extent, robust to short-run dynamics; hence, the estimation results will not change. In contrast, the ML methods, which require a full specification of the likelihood or spectral density, might lead to different estimation outcomes when the drift component is nonzero. How to implement the Whittle ML estimation method for fOU and how alternative estimation methods perform under the fOU specification will be studied in a separate paper.

References

Agiakloglou, C., P. Newbold, and M. Wohar (1993). Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis 14*(3), 235–246. 11, 15

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*. 20

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance 51*(1), 169–204. 25

Andrews, D. W. and P. Guggenberger (2003). A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica 71*(2), 675–712. 11

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. 15

Bayer, C., P. Friz, and J. Gatheral (2016). Pricing under rough volatility. *Quantitative Finance 16*(6), 887–904. 3

Bennedsen, M., A. Lunde, and M. S. Pakkanen (2021). Decoupling the short-and long-term behavior of stochastic volatility. *Journal of Financial Econometrics, forthcoming*. 23

Bolko, A. E., K. Christensen, M. S. Pakkanen, and B. Veliyev (2022). A GMM approach to estimate the roughness of stochastic volatility. *Journal of Econometrics, forthcoming*. 3, 23

Breidt, F. J., N. Crato, and P. De Lima (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics 83*(1-2), 325–348. 14

Brouste, A. and M. Fukasawa (2018). Local asymptotic normality property for fractional Gaussian noise under high-frequency observations. *The Annals of Statistics 46*(5), 2045–2061. 15, 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. 4, 13, 15

Cheung, Y.-W. and F. X. Diebold (1994). On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *Journal of Econometrics 62*(2), 301–316. 14, 15, 19

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica 41*(1), 135–155. 25

Cohen, S., F. Gamboa, C. Lacaux, and J.-M. Loubes (2011). LAN property for some fractional type brownian motion. *arXiv preprint arXiv:1111.1077*. 13

Comte, F. (1996). Simulation and estimation of long memory continuous time models. *Journal of Time Series Analysis 17*(1), 19–36. 2

Comte, F. and E. Renault (1996). Long memory continuous time models. *Journal of Econometrics 73*(1), 101–149. 2, 5

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196. 3

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. 20

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 1749–1766. 13, 14

El Euch, O., M. Fukasawa, and M. Rosenbaum (2018). The microstructual foundations of leverage effect and rough volatility. *Finance and Stochastics 22*(2), 241–280. 3

Fouque, J.-P. and R. Hu (2018). Optimal portfolio under fast mean-reverting fractional stochastic environment. *SIAM Journal on Financial Mathematics 9*(2), 564–601. 3

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. 14

Fukasawa, M. and T. Takabatake (2019a). Asymptotically efficient estimators for self-similar stationary Gaussian noises under high frequency observations. *Bernoulli 25*(3), 1870–1900. 4, 7, 8, 15, 16, 23

Fukasawa, M. and T. Takabatake (2019b). Online supplement to "asymptotically efficient estimators for self-similar stationary Gaussian noises under high frequency observations.". *Bernoulli*. 6

Fukasawa, M., T. Takabatake, and R. Westphal (2022). Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. *Mathematical Finance, forthcoming*. 3

Garnier, J. and K. Sølna (2018). Option pricing under fast-varying and rough stochastic volatility. *Annals of Finance 14*(4), 489–516. 3

Gatheral, J., T. Jaisson, and M. Rosenbaum (2018). Volatility is rough. *Quantitative Finance 18*(6), 933–949. 3, 7, 23

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. 3, 10, 11

Giraitis, L., H. L. Koul, and D. Surgailis (2012). *Large sample inference for long memory processes*. Imperial College Press, London. World Scientific. 12

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. 14

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. 3

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics 14*(2), 227–238. 3

Graves, T., R. Gramacy, N. Watkins, and C. Franzke (2017). A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951–1980. *Entropy 19*(9), 437. 2, 3

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. 20

Hosking, J. R. (1981). Fractional differencing. *Biometrika 68*(1), 165–76. 3

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers 116*(1), 770–799. 2

Korvin, G. (1992). *Fractal models in the earth sciences*. Amsterdam: Elsevier. 2

Künsch, H. (1987). Statistical aspects of self-similar processes. In *Proceedings of the First Congress of the Bernoulli Society, 1987*. 3, 12

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*(3), 283–306. 15

Li, J., P. C. Phillips, S. Shi, and J. Yu (2022). Weak identification of long memory with implications for inference. *Available at SSRN*. 23, 24

Lieberman, O. (2005). On plug-in estimation of long memory models. *Econometric Theory*, 431–454. 20

Lieberman, O., R. Rosemarin, and J. Rousseau (2012). Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes. *Econometric Theory 28*(2), 457–470. 13

Livieri, G., S. Mouti, A. Pallavicini, and M. Rosenbaum (2018). Rough volatility: evidence from option prices. *IISE transactions 50*(9), 767–776. 3

Mandelbrot, B. (1965). Une classe de processus stochastiques homothetiques a soi-application a la loi climatologique de he hurst. *Comptes Rendus Hebdomadaires des Seances de l Academie des Sciences 260*(12), 3274–3277. 2

Mandelbrot, B. B. and J. W. Van Ness (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review 10*(4), 422–437. 2, 5

Molz, F., H. Liu, and J. Szulga (1997). Fractional Brownian motion and fractional Gaussian noise in subsurface hydrology: A review, presentation of fundamental properties, and extensions. *Water Resources Research 33*(10), 2273–2286. 2

Nadarajah, K., G. M. Martin, and D. Poskitt (2021). Optimal bias correction of the log-periodogram estimator of the fractional parameter: A jackknife approach. *Journal of Statistical Planning and Inference 211*, 41–79. 11, 16

Nielsen, M. Ø. and P. H. Frederiksen (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews 24*(4), 405–443. 11, 12, 16

Park, K. and W. Willinger (2000). Self-similar network traffic: An overview. *Self-Similar Network Traffic and Performance Evaluation*, 1–38. 2

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

Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics 23*(5), 1630–1661. 3, 12

Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics 23*(4), 1048–1072. 11

Shi, S. and J. Yu (2022). Volatility puzzle: Long memory or anti-persistency. *Management Science, forthcoming*. 3, 16

Sinai, Y. G. (1976). Self-similar probability distributions. *Theory of Probability & Its Applications 21*(1), 64–80. 5

Smith, J., N. Taylor, and S. Yadav (1997). Comparing the bias and misspecification in arfima models. *Journal of Time Series Analysis 18*(5), 507–527. 11, 16

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. 3

Tauchen, G. E. and M. Pitts (1983). The price variability-volume relationship on speculative markets. *Econometrica 51*(2), 485–505. 25

Wang, X., W. Xiao, and J. Yu (2021). Modeling and forecasting realized volatility with the fractional Ornstein–Uhlenbeck process. *Journal of Econometrics, forthcoming*. 3, 15, 23

Wang, X. and J. Yu (2022). Latent local-to-unit root models. *Working paper, Singapore Management University*. 3

Wang, X., J. Yu, and C. Zhang (2022). On the optimal forecast with the fractional Brownian motion. *Working paper, Singapore Management University*. 3

Whittle, P. (1951). *Hypothesis testing in time series analysis*, Volume 4. Almqvist & Wiksells boktr. 14

Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv för matematik 2*(5), 423–434. 4

Whittle, P. (1954). On stationary processes in the plane. *Biometrika 41*, 434–449. 4

Appendix

Table 8: Summary statistics of log RVs and the first differences of log RV

	log RV				first differenced log RV			
Ticker	Mean	Std.	Skew	Kurto	Mean	Std.	Skew.	Kurto.
ETFs								
SPY	-2.53	.42	.36	3.23	-.0002	.26	.18	3.32
XLB	-2.20	.35	.30	3.44	-.0003	.24	.10	3.95
XLE	-1.95	.35	.58	3.41	-.0002	.20	.08	3.28
XLF	-2.18	.34	.79	7.71	-.0003	.25	.42	14.92
XLI	-2.32	.36	.40	3.56	-.0002	.26	.16	3.60
XLK	-2.30	.39	.61	4.69	.0000	.28	.17	4.91
XLP	-2.50	.34	.67	4.64	.0000	.26	.21	4.65
XLU	-2.16	.29	.17	3.56	-.0003	.22	.39	4.74
XLV	-2.32	.37	.80	4.74	.0000	.25	.29	6.57
XLY	-2.36	.38	.57	3.76	-.0001	.27	.10	3.27
DJ30 stocks								
AAPL	-1.80	.36	.47	3.12	.0003	.25	.53	4.55
ALD	-2.02	.34	.57	3.80	-.0002	.27	.19	4.17
AMGN	-1.68	.34	.59	3.39	-.0001	.25	.19	4.56
AXP	-1.94	.31	.66	4.20	-.0005	.26	.39	5.25
BA	-1.81	.34	.88	4.95	.0000	.26	.43	5.53
BEL	-1.99	.28	.90	7.28	.0000	.25	.53	9.84
CAT	-1.71	.33	.60	3.76	-.0003	.24	.54	5.36
CHV	-1.90	.34	.83	4.05	-.0002	.21	.23	3.75
CRM	-1.50	.35	.26	3.46	-.0003	.25	.25	5.18
CSCO	-1.82	.31	.52	4.16	-.0002	.23	.22	5.16
DIS	-1.94	.31	.75	4.63	.0000	.25	.21	4.64
GS	-1.72	.30	.45	3.53	-.0003	.21	.21	3.49
HD	-1.92	.31	1.08	6.60	-.0001	.25	.60	9.77
IBM	-2.03	.30	.78	4.30	-.0003	.24	.16	4.22
INTC	-1.73	.31	.70	4.60	-.0002	.23	.29	5.24
JNJ	-2.12	.33	1.07	6.13	-.0002	.26	.59	6.92
JPM	-1.82	.32	.90	6.64	-.0004	.23	.53	14.10
KO	-2.14	.30	.71	4.61	-.0002	.23	.39	5.60
MCD	-2.12	.30	.88	4.95	-.0003	.25	.35	4.96
MMM	-2.07	.36	.58	3.72	-.0001	.26	.34	4.17
MRK	-1.92	.31	.96	5.94	-.0001	.26	.44	7.34
MSFT	-1.83	.32	.60	4.10	-.0001	.22	.17	3.92
NIKE	-1.82	.29	.90	4.64	-.0003	.24	.07	4.23
PG	-2.13	.29	.73	5.03	.0000	.23	.48	7.11
SPC	-2.07	.30	.55	4.12	-.0003	.27	.44	5.28
UNH	-1.79	.32	.71	4.84	-.0001	.26	.57	6.13
V	-1.91	.32	.70	4.64	-.0003	.25	.38	5.40
WAG	-1.71	.32	.57	4.38	-.0002	.28	.35	4.34
WMT	-2.05	.30	.90	4.92	-.0002	.26	.60	6.15
XOM	-2.01	.32	.74	4.21	-.0001	.21	.15	3.95