Fractional Stochastic Volatility Model*

Shuping Shi
Macquarie University

Xiaobin Liu
Zhejiang University

Jun Yu
Singapore Management University

October 11, 2021

Abstract

This paper introduces a discrete-time fractional stochastic volatility model (FSV) based on fractional Gaussian noise. The new model includes the standard stochastic volatility model as a special case and has the same limit as the fractional integrated stochastic volatility (FISV) model. A simulated maximum likelihood method, which maximizes the time-domain log-likelihood function calculated by the importance sampling technique, and a frequency-domain quasi maximum likelihood method (or Whittle) are employed to estimate the model parameters. Simulation studies suggest that, while both estimation methods can accurately estimate the model, the simulated maximum likelihood method outperforms the Whittle method. As an illustration, we fit the FSV and FISV models with the proposed estimation techniques to the S&P 500 composite index over a sample period spanning 45 years. Our results reveal that the volatilities of the data series are persistent and rough.

JEL classification: C15, C22, C32
Keywords: Fractional Brownian motion; stochastic volatility; long memory; variance-covariance matrix; spectral density; rough volatility

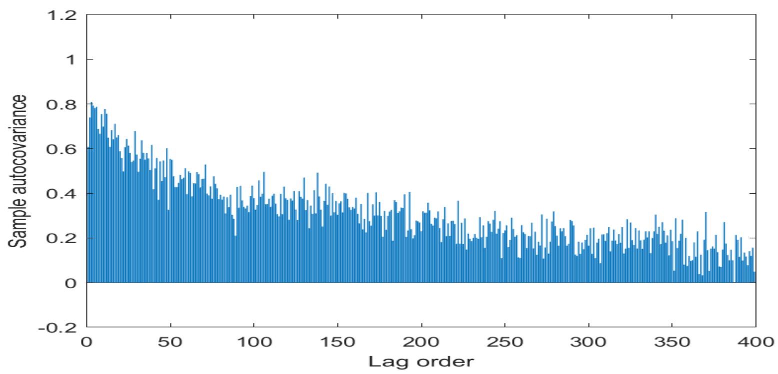
1 Introduction

Temporal dependence in volatility has been one of the most studied problems in financial econometrics. One prominent feature of volatility dynamics is its slowly decaying autocovariance function (Ding et al., 1993). As illustrated in Figure 1, the sample autocovariance of the daily log squared returns of the S&P 500 index (from 1975 to 2020) remains non-negligible even at a very large lag order. This feature of volatility is often referred to as 'long-range dependence'.

Motivated by this empirical feature, many long-memory volatility models have been put forward. In the discrete-time framework, we have, for example, the fractional integrated generalized autoregressive

*Liu acknowledges the financial support from the National Natural Science Foundation of China (No.72003171). Shi acknowledges research support from the Australian Research Council under project No. DE190100840. Yu acknowledges the support from the Lee Foundation. Xiaobin Liu, School of Economics, Zhejiang University, Hangzhou, 310058, China. Email: liuxiaobin@zju.edu.cn. Shuping Shi, Department of Economics, Macquarie University; E-mail: shuping.shi@mq.edu.au. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Rd, Singapore 178903. Email: yujun@smu.edu.sg.

Figure 1: The sample autocovariance of the daily log squared returns of the S&P composite 500 index over the period from January 3, 1975 to September 30, 2020



conditional heteroskedastic (FIGARCH) model (Baillie et al., 1996; Bollerslev and Mikkelsen, 1996) and the fractional integrated stochastic volatility (FISV) models (Breidt et al., 1998; Harvey, 2007; Hurvich and Soulier, 2009). In both models, the long-range dependent feature of volatilities is captured by a fractional integrated process (Granger and Joyeux, 1980) which takes the form of

$$(1-L)^d u_t = e_t \text{ with } e_t \overset{iid}{\sim} N\left(0,1\right), \tag{1}$$

where L is the lag operator and d is the fractional parameter. This process has a long memory when $d \in (0, 0.5)$ in the sense that its autocovariances are all positive and decay at a hyperbolic rate.

Similar developments were observed in the continuous-time volatility literature, enabling more accurate pricing of derivative securities (Comte and Renault, 1996, 1998). For example, the continuous-time fractional stochastic volatility (fSV) model considered in Comte and Renault (1998) takes the following expression:

$$\begin{array}{rcl} dy_t & = & \sigma^* e^{h_t/2} dW_t \\ dh_t & = & \gamma h_t dt + \sigma_h^* dB_t^H \end{array}, \tag{2}$$

where y_t is the log price of an asset at period t, h_t is the log volatility of dy_t, W_t is a standard Brownian motion, and B_t^H is a fractional Brownian motion (fBM).[1] The model considered in Rosenbaum (2008) is similar to (2) but with a more general drift function for dh_t. The B_t^H process is a zero mean Gaussian with an autocovariance function of

$$\mathbb{E}(B_t^H B_s^H) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right). \tag{3}$$

See, e.g., Mandelbrot and Van Ness (1968). The parameter H is known as the Hurst or memory

[1] The fractional Brownian motion becomes the standard Brownian motion W_t when $H = 0.5$.

parameter in the literature[2]. When $H \in (0.5, 1)$ as in Comte and Renault (1996, 1998) and Rosenbaum (2008), the fBM has a long memory in the sense that its autocovariance function decays at a hyperbolic rate (i.e., smooth decaying) and the one-sided long-run variance $\sum_{n=0}^{\infty} \mathbb{E}\left(B_1^H(B_{n+1}^H - B_n^H)\right) = \infty$. When $H \in (0, 0.5)$, the increment of fBM is negatively correlated, the sample path generated by B_t^H is rough. Consequently, the sample path of the volatility process is rough (as opposed to smooth).

Different techniques have been proposed recently to estimate the Hurst parameter H in model (2). For example, assuming $\gamma = 0$, Fukasawa et al. (2021) propose a quasi-maximum likelihood method based on an approximate spectral density of daily realized volatilities (obtained from intra-day prices). Bolko et al. (2020) consider the generalized method of moment (GMM) using moments of daily realized volatilities. Both papers find evidence of roughness (i.e., $H < 0.5$) in the volatility process.

In this paper, motivated by model (2) and studies by Fukasawa et al. (2021) and Bolko et al. (2020), we consider the Euler discretized version of the fSV model and propose two estimation techniques for the model. This discrete-time model is referred to as FSV. We assume $H \in (0, 1)$, allowing for both smoothed and rough volatilities. Unlike Fukasawa et al. (2021) where $\gamma = 0$ is imposed, we assume $\gamma < 0$. The FSV model includes the standard discrete-time log-normal stochastic volatility (SV) model as a special case with $H = 0.5$ and shares the same limit as the FISV model when $H = d + 0.5$ under the in-fill asymptotic scheme. We discuss the autocovariance function and spectral density of the log volatility h_t under the new model setting.

For the model estimation, we propose a time-domain simulated maximum likelihood (SML) and a frequency-domain quasi maximum likelihood (a.k.a., Whittle) method. The log-likelihoods are constructed from the daily log returns, with the log volatilities being latent.[3] The Whittle method maximizes a spectral log-likelihood function of the model and has been employed by Breidt et al. (1998) for estimating the FISV model. The SML method computes the time-domain likelihood function via the importance sampling technique. The SML method has been successfully applied to the basic SV model by Sandmann and Koopman (1998) and Yu (2011) and more recently to an SV model with a general leverage effect by Catania (2021). One can easily use the classical asymptotic theory to make statistical inferences and also obtain filtered or smoothed estimates of volatilities.

Our simulation results show that both SML and Whittle methods can provide reasonably accurate

[2]It is sometimes called the self-similarity parameter because $B_{at}^H \overset{d}{=} a^H B_t^H$ for any $a \in \mathbb{R}^+$.

[3]Our approach is different from recent studies that are based on daily realized volatilities, e.g., Fukasawa et al. (2021) and Bolko et al. (2020). See the conclusion section for further discussions on this issue.

estimation for all model parameters in FSV. The SML method outperforms the Whittle method in terms of standard errors and root mean square errors, whereas the Whittle estimates have slightly smaller biases for some parameters. Additionally, our simulations reveal that the semiparametric local polynomial Whittle with noise (LPWN) method proposed by Frederiksen et al. (2012) does not work when the fractional process is 'contaminated' by a highly persistent short-run dynamic. The estimated Hurst parameter from LPWN is significantly biased.

We fit the FSV model to the daily S&P 500 composite index from January 1975 to September 2020, spanning over 45 years and consisting 11,519 observations. The estimation results from both the SML and Whittle methods suggest that the persistency of the volatilities is captured by a near-unity autoregressive coefficient. Meanwhile, roughness is also present in the data as the estimated Hurst parameter H is smaller than 0.5. With the same estimation techniques, we obtain similar results from the FISV model for the data series. Our findings are consistent with the recent literature on rough volatility. See, for example, Gatheral et al. (2018); Fukasawa et al. (2021); Wang et al. (2021b); Bennedsen et al. (2021); Bolko et al. (2020).

The paper is organized as follows. Section 2 introduces the FSV model and derives its statistical properties. In Section 3, we discuss the SML method in the time domain and the Whittle method in the frequency domain. Section 4 checks the finite-sample performance of the SML, Whittle, and LPWN methods using data simulated from the FSV model. Section 5 employs the SML and Whittle methods to estimate the SV, FSV, and FISV models for the S&P 500 composite index. Section 6 concludes the paper. The appendix collects implementation details of the SML method and estimation details of the FISV model.

2 Fractional Stochastic Volatility Model

Suppose that log returns r_t are available on grids t with $t = 1, 2, \ldots, T$. Consider the following discrete-time fractional SV model

$$r_t = \sigma e^{h_t/2} \varepsilon_t, \tag{4}$$

$$h_t = \beta h_{t-1} + \sigma_h \eta_t^H, \tag{5}$$

where $\beta \in (-1, 1)$, $\varepsilon_t \overset{iid}{\sim} N(0, 1)$ and $\eta_t^H = B_t^H - B_{t-1}^H$ is a fractional Gaussian noise with $H \in (0, 1)$. The fractional SV model reduces to the basic SV model when $H = 0.5$ as in this case $\eta_t^H \overset{iid}{\sim} N(0, 1)$.

2.1 Autocovariance function of h_t

Let $k = |t - s|$. The autocovariance between h_t and h_s is denoted by $Cov\left(h_t, h_s\right) := \gamma_h\left(k\right)$ and has the form of

$$\gamma_h\left(k\right) = \sum_{j=-\infty}^{\infty} \tilde{\gamma}\left(j\right) \gamma_\eta\left(k - j\right), \tag{6}$$

where $\tilde{\gamma}\left(j\right) = \sigma_h^2 \beta^j / \left(1 - \beta^2\right)$ is the autocovariance of the pure AR component and $\gamma_\eta\left(k\right) := Cov\left(\eta_t^H, \eta_s^H\right)$ is the autocovariance of η_t^H. See, for example, Brockwell and Davis (2009).

The autocovariance of the fractional Gaussian noise η_t^H has the following expression:

$$\gamma_\eta\left(k\right) = \frac{1}{2}\left[\left(k+1\right)^{2H} + \left(k-1\right)^{2H} - 2k^{2H}\right]. \tag{7}$$

The variance $\gamma_\eta\left(0\right) = 1$. When $H = 0.5$, the fractional Gaussian noise becomes the standard Gaussian noise and $\gamma_\eta\left(k\right) = 0$ for all $k \neq 0$. If $H \neq 0.5$, $\gamma_\eta(k) \neq 0$ and, by the first order Taylor series expansion,

$$\gamma_\eta\left(k\right) = \frac{1}{2}k^{2H}\left[\left(1+\frac{1}{k}\right)^{2H} + \left(1-\frac{1}{k}\right)^{2H} - 2\right] \sim H(2H-1)k^{2H-2} \tag{8}$$

for large k. The autocovariance $\gamma_\eta\left(k\right)$ decays at a hyperbolic rate as k goes to infinity. The two-sided long-run variance of η_t^H is approximately

$$\sum_{k=-\infty}^{\infty} \gamma_\eta\left(k\right) \sim H\left(2H-1\right)\sum_{k=-\infty}^{\infty} k^{2H-2} = \infty$$

when $H > 0.5$, and

$$\sum_{k=-\infty}^{\infty} \gamma_\eta\left(k\right) = 1 + 2\sum_{k=1}^{\infty} \gamma_\eta\left(k\right) = 0$$

when $H < 0.5$.

In summary, if $H > 0.5$, η_t^H is positively autocorrelated and has a long memory since $\sum_{k=-\infty}^{\infty} \gamma_\eta\left(k\right) = \infty$, whereas if $H < 0.5$, η_t^H is negatively autocorrelated with $\sum_{k=-\infty}^{\infty} \gamma_\eta\left(k\right) = 0$ (a.k.a. anti-persistent).

2.2 Spectral density of h_t

The spectral density of h_t at frequency λ, denoted by $f_h(\lambda)$, has the following expression:

$$f_h\left(\lambda\right) = \left|A\left(\lambda\right)\right|^2 f_\eta\left(\lambda\right) \text{ with } A\left(\lambda\right) = \sum_{s=0}^{\infty} \tilde{\gamma}(s)e^{is\lambda} \text{ and } -\pi \leq \lambda \leq \pi, \tag{9}$$

where $f_\eta\left(\lambda\right)$ is the spectral density of the fractional Gaussian noise η_t^H. See Priestley (1981, p226) and also Beran (1994, p61). Under the model specification, we have

$$\left|A\left(\lambda\right)\right|^2 = \frac{\sigma_h^2}{1 - 2\beta\cos\left(\lambda\right) + \beta^2}$$

and

$$f_\eta(\lambda) = 2C_H(1-\cos\lambda)\sum_{j=-\infty}^{\infty}|\lambda+2\pi j|^{-1-2H}$$

with $C_H = (2\pi)^{-1}\Gamma(2H+1)\sin(\pi H)$, from Beran (1994, Proposition 2.1). If η_t^H is the standard Gaussian noise, $f_\eta(\lambda) = 1/(2\pi)$.

3 Estimation Methods

In this section, we introduce the time-domain simulated maximum likelihood method and the frequency-domain quasi maximum likelihood method for the fractional stochastic volatility model defined by (4)-(5).

3.1 Time-domain Simulated Maximum Likelihood

Let $r = (r_1, r_2, \cdots, r_T)'$ and $h = (h_1, h_2, \cdots, h_T)'$. Model parameters are collected in θ, that is, $\theta = (\sigma, \beta, \sigma_h, H)$. Under the model specification of (4)-(5), the joint probability density function (pdf) of returns is

$$f(r|\theta) = \int f(r;h|\theta)\,dh = \int f(r|h,\theta)\,f(h|\theta)\,dh, \tag{10}$$

where the conditional density $f(r|h,\theta) = \Pi_{t=1}^{T}\phi\left(r_t; 0, \sigma^2 e^{h_t}\right)$ with $\phi\left(\cdot; 0, \sigma^2\right)$ being the pdf of a normal distribution with mean zero and variance σ^2 and the pdf of h

$$f(h|\theta) = (2\pi)^{-T/2}|\Xi_\theta|^{-1/2}\exp\left(-\frac{1}{2}h'\Xi_\theta h\right), \tag{11}$$

with Ξ_θ being a $T \times T$ matrix whose $(t,s)^{th}$ element is given by $Cov(h_t, h_s)$ for $t, s = 1, 2, \cdots, T$. We compute Ξ_θ using formula (6) as in Bertelli and Caporin (2002). The summand is truncated at m, which takes a larger value when β is close to unity.[4]

3.1.1 Likelihood Evaluation

The exact likelihood $f(r|\theta)$ involves a T-dimensional integral which makes it extremely difficult to evaluate. A natural alternative way of evaluating the likelihood function is by Monte Carlo simulations. One can draw $h^{(s)}$ from the multivariate normal distribution $N(0, \Xi_\theta)$ with $s = 1, \cdots, S$ and

[4]Specifically, we set $m = 1000$ for $\beta < 0.9$, $m = 2000$ for $0.9 \le \beta < 0.99$, $m = 4000$ for $0.99 \le \beta < 0.995$, and $m = 7000$ for $\beta \ge 0.995$. For formula (6) to be applicable, $|\beta|$ must be strictly smaller than one as in our model.

approximate the likelihood function via the following 'brute force' Monte Carlo method,

$$\frac{1}{S}\sum_{s=1}^{S} f\left(r|h^{(s)},\theta\right). \tag{12}$$

The distribution of h and the pdf $f(h|\theta)$ in (11) are obtained directly from the statistical assumption of the model. This importance sampler, however, ignores the crucial information brought by the data r regarding the latent variable h. As such, the approximation in (12) is extremely inefficient and requires an enormous N to gain a reasonable accurate approximation of $f(r|\theta)$ (Liesenfeld and Richard, 2003).

To improve the estimation efficiency, we employ the importance sampling technique to approximate the log-likelihood function, in the spirit of Shephard and Pitt (1997) and Durbin and Koopman (1997) for non-Gaussian and nonlinear state space models. Although our model is not a state-space model as h does not have the Markovian property unless $H = 0.5$, the idea of Shephard and Pitt (1997) and Durbin and Koopman (1997) is general enough to be applicable to our models. Maximizing the log-likelihood leads to the SML estimators of the parameters.

The idea of the importance-sampling-based approximation of the log-likelihood is to sample $h^{(s)}$ from an alternative multivariate normal distribution with mean $h_\theta^*(r)$ and variance-covariance matrix $\Sigma_\theta^*(r)$, denoted by $N\left(h_\theta^*(r), \Sigma_\theta^*(r)\right)$. Let $g(\cdot)$ and $G(\cdot)$ be the pdf and cdf of $N\left(h_\theta^*(r), \Sigma_\theta^*(r)\right)$, respectively. The pdf $f(r|\theta)$ can be approximated by the sample average of $\left\{f\left(r; h^{(s)}|\theta\right)/g\left(h^{(s)}\right)\right\}_{s=1}^{S}$, that is,

$$f(r|\theta) = \int \frac{f(r;h|\theta)}{g(h)} dG(h) \approx \frac{1}{S}\sum_{s=1}^{S}\frac{f\left(r;h^{(s)}|\theta\right)}{g\left(h^{(s)}\right)}. \tag{13}$$

Importantly, unlike the 'brute force' technique (12), $\left\{h^{(s)}\right\}_{s=1}^{S}$ are drawn from a proposal distribution $N\left(h_\theta^*(r), \Sigma_\theta^*(r)\right)$ that is obtained by the Laplace approximation. In detail, we compute $h_\theta^*(r)$ as the modal configuration of $\log\left(f(r;h|\theta)\right)$ such that

$$h_\theta^*(r) = \arg\max_h \log\left(f(r;h|\theta)\right), \tag{14}$$

where $\log\left(f(r;h|\theta)\right)$ has the form

$$\log f(r;h|\theta) = \sum_{t=1}^{T}\log\left(\phi\left(r_t;0,\sigma^2 e^{h_t}\right)\right) + \log\left(f(h|\theta)\right) \tag{15}$$

$$= -T\log(2\pi) - \frac{T}{2}\log\left(\sigma^2\right) - \frac{1}{2}\sum_{t=1}^{T}h_t - \frac{1}{2}\sum_{t=1}^{T}z_t - \frac{1}{2}\log|\Xi_\theta| - \frac{1}{2}h'\Xi_\theta^{-1}h, \tag{16}$$

with $z_t = r_t^2 / \left(\sigma^2 e^{h_t}\right)$. The variance-covariance matrix $\Sigma_\theta^*(r)$ is calculated as

$$\Sigma_\theta^*(r) = \left[-\frac{\partial^2 \log\left(f\left(r; h|\theta\right)\right)}{\partial h \partial h'} \right]^{-1} \Bigg|_{h = h_\theta^*(r)} . \tag{17}$$

Clearly, the proposal distribution $N\left(h_\theta^*(r), \Sigma_\theta^*(r)\right)$ depends on the data r.

The SML estimator of the FSV model is denoted by $\hat{\theta}$ and defined as

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \log \left[\frac{1}{S} \sum_{s=1}^{S} \exp\left(\log \frac{f\left(r; h^{(s)}|\theta\right)}{g\left(h^{(s)}\right)} \right) \right], \tag{18}$$

where Θ is the parameter space and

$$\begin{aligned} \log \frac{f\left(r; h^{(s)}|\theta\right)}{g\left(h^{(s)}\right)} &= -\frac{T}{2}\log(2\pi) - T\log\left(\sigma^2\right) - \frac{1}{2}\sum_{t=1}^{T} h_t^{(s)} - \frac{1}{2}\sum_{t=1}^{T} z_t - \frac{1}{2}\log|\Xi_\theta| - \frac{1}{2} h^{(s)\prime}\Xi_\theta^{-1} h^{(s)} \\ &\quad + \frac{1}{2}\log|\Sigma_\theta^*(r)| + \frac{1}{2}\left(h^{(s)} - h_\theta^*(r)\right)' \Sigma_\theta^*(r)^{-1} \left(h^{(s)} - h_\theta^*(r)\right). \end{aligned} \tag{19}$$

3.1.2 Smoothed and Filtered Volatilities

The estimated volatility sequence $\hat{h}$ is taken as $h_{\hat{\theta}}^*(r)$, which is the modal configuration of $\log f\left(r, h|\hat{\theta}\right)$ using equation (14) with θ replaced by $\hat{\theta}$.[5] Clearly, $\hat{h}_t$, the t^{th} element in $\hat{h}$, uses all the information in $\{r_t\}_{t=1}^T$ and hence, is a smoothed estimate of h_t. Correspondingly, the quantity $\exp(h_t/2)$, which follows a log-normal distribution, can be estimated as

$$\exp\left(\frac{h_{\hat{\theta}}^{*(t)}(r)}{2} + \frac{\Sigma_{\hat{\theta}}^{*(t,t)}(r)}{8} \right), \tag{20}$$

where $h_{\hat{\theta}}^{*(t)}(r)$ is the t^{th} element in $h_{\hat{\theta}}^*(r)$, $\Sigma_{\hat{\theta}}^{*(t,t)}(r)$ is the t^{th} diagonal element in $\Sigma_{\hat{\theta}}^*(r)$ (computed from (17) with $h_\theta^*(r)$ replaced by $h_{\hat{\theta}}^*(r)$).

To obtain the filtered estimates of h_t and $\exp(h_t/2)$, one can apply the Laplace approximation analogously to $\log f\left(r_1, ..., r_t, h_1, ..., h_t|\hat{\theta}\right)$ instead.

3.2 Frequency-domain Quasi Maximum Likelihood Method

Let $x_t = \log r_t^2$ be the log squared return at period t. Under the model specification we have

$$x_t = \log\left(\sigma^2\right) + h_t + \log\varepsilon_t^2 = \mu + h_t + \omega_t, \tag{21}$$

where $\mu = \log\left(\sigma^2\right) + \mathbb{E}\left[\log\varepsilon_t^2\right]$ and $\omega_t = \log\varepsilon_t^2 - \mathbb{E}\left[\log\varepsilon_t^2\right]$. Since $\varepsilon_t \sim N(0,1)$, $\log\varepsilon_t^2$ is a log $\chi^2_{(1)}$ distribution with $\mathbb{E}\left[\log\varepsilon_t^2\right] = -1.27$ and $\mathbb{V}\left[\log\varepsilon_t^2\right] = \mathbb{V}[\omega_t] = \pi^2/2 \approx 4.9$.

[5] Since h_t is normally distributed, the mode is the same as the expected value.

The Whittle estimator minimizes the negative of the log spectral likelihood function

$$L_n(\theta) = \frac{\pi}{n/2} \sum_{k=1}^{n/2} \left\{ \log f_x(\lambda_k) + \frac{I_n(\lambda_k)}{f_x(\lambda_k)} \right\}, \tag{22}$$

where $\lambda_k = 2\pi k/n$, $f_x(\lambda)$ is the spectral density of x_t at frequency λ_k, and $I_n(\lambda_k)$ is the k^{th} normalized periodogram ordinate specified as

$$I_n(\lambda_k) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} x_t \exp(-it\lambda_k) \right|^2 = \frac{1}{2\pi n} \left[\left(\sum_{t=1}^{n} x_t \cos(\lambda_k t) \right)^2 + \left(\sum_{t=1}^{n} x_t \sin(\lambda_k t) \right)^2 \right].$$

The spectral density of x_t takes the form of

$$f_x(\lambda_k) = f_h(\lambda_k) + \frac{\sigma_\omega^2}{2\pi},$$

where $f_h(\lambda_k)$ is given in (9) and $\sigma_\omega^2 = \mathbb{V}[\omega_t]$ is the unconditional volatility of ω_t. When ε_t is assumed to be standard normal as in our case, $\sigma_\omega^2 = \pi^2/2$ and hence need not be estimated. That is, only three parameters (i.e., H, β, σ_h) are estimated by the Whittle method.

Although the spectral log-likelihood takes the variance of ω_t into consideration, it ignores the distributional assumption of the noise (i.e., $\log \chi^2_{(1)}$ under our setting). In other words, the expression of the spectral likelihood function remains the same if $\omega_t \sim N(0, \pi^2/2)$. That is why we refer to the method as the frequency-domain quasi maximum likelihood method. In contrast, the SML method utilizes the full information for the estimation. As such, SML is expected to be more efficient than Whittle. Moreover, the filtered and smoothed estimates of volatilities can be obtained as the by-product of SML, while an additional step is needed when the Whittle method is used for parameter estimation. A significant advantage of the Whittle estimation method over SML is that it does not require the computation of the variance-covariance matrix, and hence, is less computationally costly.

4 Simulation Studies

In this section, we investigate the estimation accuracy of the proposed estimation methods for the FSV model in terms of bias, standard error, and root mean square error (RMSE). Log returns are generated from equation (4), while the latent variable h_t is from (5). The fractional Gaussian noises are generated using fast Fourier transforms (Kroese and Botev, 2015).

We consider two sets of model parameters. The first one corresponds to our empirical estimation results in Section 5 for the S&P 500 composite index (from 1975 to 2020):

$$\text{Set I: } H = 0.176, \beta = 0.998, \sigma_h = 0.464, \sigma = 0.008.$$

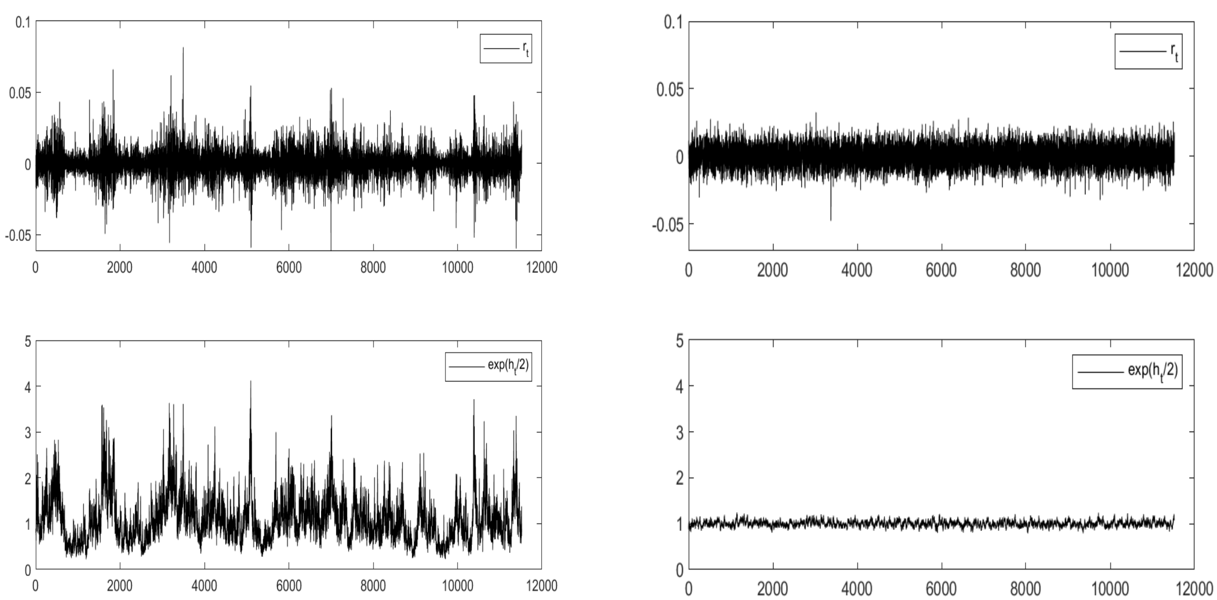
The second set of model parameters are those obtained by Breidt et al. (1998) from the S&P 500 CRSP data over the sample period from July 1962 to December 1987 for the FISV model (23)-(25) with the Whittle method.[6] The parameter values are:

$$\text{Set II: } H = 0.944, \beta = 0.932, \sigma_h = 0.0564, \sigma = 0.008.$$

The sample size T is set to be $11,520$ as in our empirical application. Figure 2 displays one typical realized sample path of the data generating process with the two different sets of model parameters. It is obvious from the graphs that the sample path based on our empirical results is more empirically realistic.

Figure 2: One typical sample path of the FSV model with the two different sets of model parameters

(a) Set I (b) Set II



The number of replications for the simulation study is 100. The choice of a small number of replications is due to the long computational time required by the SML method. Although it is relatively

[6] In particular, Breidt et al. (1998) report 0.444 and 0.932 for d and β that jointly determine the dynamics of volatilities, and 0.00318 for σ_h^2. The parameter σ, which only affects the level of volatilities, is specified by us.

fast to compute the variance-covariance matrix with the split method (Bertelli and Caporin, 2002), the sample size considered is very large, and the estimation involves numerical optimization, which evaluates the likelihood function repeatedly until a global maximum is reached. We set the initial values for the optimizations to be the true values. For SML, the number of samples in the importance sampling is $1,000$.

Table 1 reports the bias, standard error, and RMSE of the estimated model parameters using the SML and Whittle methods under both parameter settings. Let us examine the performance of the two estimation methods under the more empirically realistic parameter setting, that is, Set I. First, it can be seen that both methods can estimate all parameters accurately. The bias is always very small, although the biases obtained from the Whittle method are slightly smaller than those of SML. Secondly, the SML method outperforms the Whittle method in terms of standard error, especially for H and σ_h. The relative inefficiency of Whittle over SML is 43% and 22% for H and σ_h, respectively. This suggests that ignoring the distributional information in $\log \chi^2_{(1)}$ can result in severe efficiency loss.[7] Third, not surprisingly, the RMSEs, which summarize the trade-off between biases and variances, are smaller for SML than for Whittle for H and σ_h. To visualize the estimation biases and variations, we show in 3-D scatter plots (Panel (a) and Panel (b) of Figure 3) the estimated model parameters of the SML and Whittle methods from the 100 replications. The red circle indicates the location of true model parameters. It is obvious that the estimates by SML are more concentrated than those by Whittle.

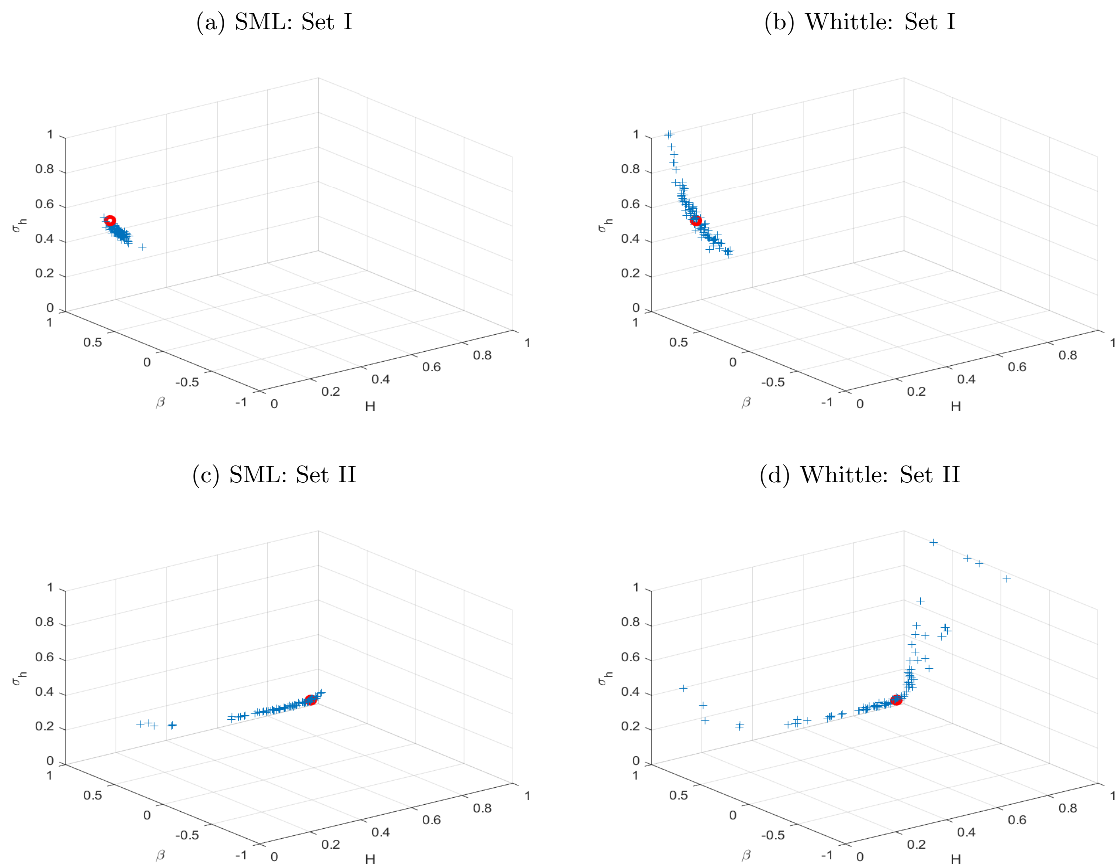
Table 1: Biases, standard errors and RMSEs of the estimated model parameters

	Bias			Std. Err.			RMSE		
	H	β	σ_h	H	β	σ_h	H	β	σ_h
Parameter Set I									
SML	0.034	-0.0008	-0.080	0.023	0.001	0.034	0.041	0.001	0.086
Whittle	0.001	-0.0002	0.033	0.054	0.001	0.151	0.054	0.001	0.154
LPWN									
$(R_{short}, R_{noise}) = (1, 0)$	0.931			0.048			0.933		
$(R_{short}, R_{noise}) = (1, 1)$	0.931			0.048			0.932		
Parameter Set II									
SML	-0.099	0.0123	-0.002	0.146	0.026	0.019	0.176	0.029	0.019
Whittle	-0.088	-0.065	0.133	0.194	0.241	0.277	0.212	0.249	0.306
LPWN									
$(R_{short}, R_{noise}) = (1, 0)$	0.091			0.062			0.110		
$(R_{short}, R_{noise}) = (1, 1)$	0.093			0.063			0.112		

[7] Kim et al. (1998) examines the difference between the $\log \chi^2_{(1)}$ and normal distributions and shows that replacing the $\log \chi^2_{(1)}$ distribution with a normal distribution can lead to poor finite sample performance.

For the second parameter setting, SML outperforms Whittle in terms of both bias and RMSE. The only exception is that the bias of H is slightly smaller when using the Whittle method. However, the estimation accuracy of both methods is not as high as that for Set I. The improvement of SML over Whittle can also be seen from Panels (c) and (d) of Figure 3.

Figure 3: Estimated model parameters of the FSV model: SML and Whittle. The red circle indicates the location of the true model parameters.



Other than using SML and Whittle to estimate the FSV model, one can also implement the semi-parametric local polynomial Whittle with noise (LPWN) method of Frederiksen et al. (2012). The LPWN method was proposed to estimate the memory parameter of a perturbed long memory process. The perturbations can be short-run dynamics and/or shocks with short-memory persistency. Under our model specification, x_t is the perturbed data series. Specifically, h_t perturbs the fractional process η_t^H with a stationary AR(1) component whose persistent level is determined by β and x_t perturbs the fractional process with the stationary AR(1) component and an i.i.d. log $\chi^2_{(1)}$ noise. When $H > 0.5$ and the short-run dynamics are far away from unit root nonstationarity, Frederiksen et al. (2012) find that LPWN performs well in finite samples. However, the applicability of LPWN for perturbed fractional processes with $H < 0.5$ and β being very close to unity is yet to be investigated. We apply LPWN to the same simulated data sets. The two tuning parameters (R_{short}, R_{noise}) of LPWN are set to be either $(1, 0)$ or $(1, 1)$ as in Frederiksen et al. (2012).

Evidently, from Table 1, the LPWN method cannot accurately estimate H under the first parameter setting. We think this inaccuracy is due to the fact that β is very close to one, although LPWN works well under the second parameter setting when $\beta = 0.932$.[8] See Shi and Yu (2021) for explanations on why local Whittle methods cannot separate persistent short-run dynamics from fractional processes. This inaccuracy is one of the reasons why more efficient estimation methods (i.e., SML and Whittle methods) were introduced for the FSV model.

5 Empirical Studies

The FSV model, along with the basic SV and the FISV models, are employed to study the S&P 500 composite index from January 3, 1975 to September 30, 2020 at the daily frequency. The FISV model is closely related to the FSV model and has been widely studied in the literature. It has a specification of the following:

$$r_t = \sigma e^{h_t/2} \varepsilon_t, \tag{23}$$

$$h_t = \beta h_{t-1} + \sigma_h u_t, \tag{24}$$

$$u_t = (1 - L)^{-d} e_t \text{ with } e_t \overset{iid}{\sim} N(0, 1). \tag{25}$$

The fractional difference operator is defined as

$$(1 - L)^{-d} = \sum_{j=0}^{\infty} \phi_j L^j \text{ with } \phi_j = \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)}, \tag{26}$$

where $\Gamma(\cdot)$ is the gamma function and $d \in (-0.5, 0.5)$. The fractional integrated process u_t defined in (26) is a 'Type I' $I(d)$ process in the sense of Marinucci and Robinson (1999) and Robinson (2005).[9] The SML method can be applied analogously to estimate the FISV model. The Whittle method has

[8] One difference between our DGPs and the DGPs considered in Frederiksen et al. (2012) is that implicitly we perturb the signal with an i.i.d. log $\chi^2_{(1)}$ noise, instead of an i.i.d. Gaussian noise. Our unreported simulation shows that the non-Gaussian noise does not lead to a larger bias in LPWN (although increasing its variance).

[9] The 'Type II' fractional integrated process u_t^* is associated with a truncated error noise e_t^*, where $e_t^* = e_t$ if $t \geq 1$ and 0 if $t < 1$.

been employed by Breidt et al. (1998) to estimate the FISV model. See Appendix B for the estimation details.

Under the in-fill asymptotic scheme (i.e., as the sampling interval goes to zero), Tanaka (2013) shows that FISV converges weakly to fSV defined by (2) with $H = d + 0.5$. Since FSV is the Euler discretization of the fSV model, when the sampling interval goes to zero, FSV also converges weakly to fSV. As such, the FISV and FSV processes are asymptotically equivalent under the in-fill scheme subject to a normalization factor, and the fractional parameter d in FISV is linked to the Hurst parameter H of the FSV model in the form of $H = d + 0.5$. See Wang et al. (2021a, Remark 3) for more details. However, as remarked in Wang et al. (2021a), there is no reason to believe the two models perform the same in finite samples.

The data are obtained from DataStream, containing 11,520 daily observations within the sample period. Figure 4 displays the dynamics of the log returns r_t and $x_t = \log(r_t - \bar{r})^2$ with $\bar{r}$ being the sample mean. Table 2 provides the summary statistics of the data series. One can see that for all data series considered, both r_t and x_t are left skewed and leptokurtic. The standard deviations of x_t are greater than $\sqrt{4.9} \approx 2.21$, consistent with our model.

Table 2: Summary statistics: the S&P 500 composite index

	Mean	Std. dev	Skew.	Kurto.
r_t	0.0003	0.011	-1.12	29.28
x_t	-10.96	2.49	-1.15	5.85

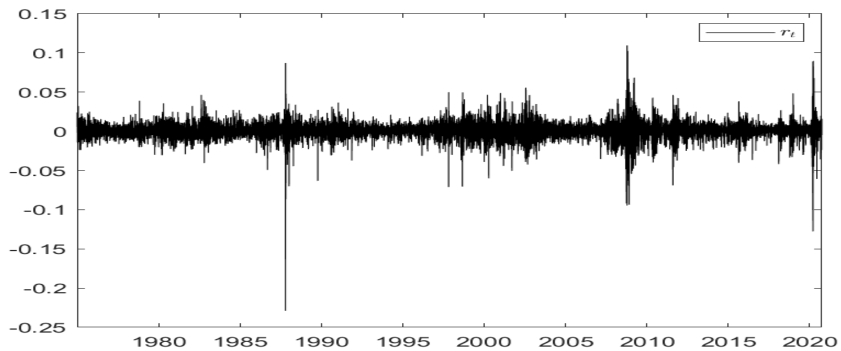
For the numerical optimization of SML, we consider a set of $2,500$ initial values and choose the one with the largest log-likelihood value as the initial input of the optimization. The $2,500$ grid points considered cover a wide range of values of the model parameters. The autoregressive coefficient is from -0.205 to 0.995 with an increment of 0.05, that is, $\beta_0 \in \{-0.205 : 0.05 : 0.995\}$, the Hurst parameter $H_0 \in \{0.001 : 0.1 : 0.999\}$, and $\sigma_{h0} \in \{0.05 : 0.05 : 0.5\}$. We set σ_0 to be the estimate of the standard SV model. Notice that we use a subscript 0 to denote the initial values. The optimization is done with the constraint maximum likelihood function *fmincon* in MATLAB, with the upper bound being one for all parameters and the lower bound being -1 for β and zero for the other three parameters.

Estimation results, including the point estimates and 90% confidence intervals of all four parameters, are reported in Table 3.[10] First, the log-likelihood values (the second last column) of the FSV and FISV

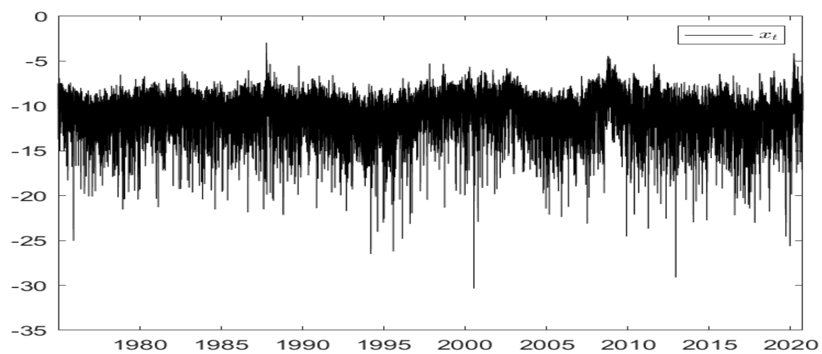
[10]When obtaining the 90% confidence intervals, we use the standard asymptotic distribution for the SML estimator.

Figure 4: The S&P 500 composite index from January 3, 1975 to September 30, 2020

(a) Return r_t



(b) Log squared return x_t



models are much larger than that of the basic SV model.[11] The log-likelihood ratio statistics (last column) suggest that by freeing up only the memory parameter, the FSV and FISV models improve the likelihood value substantially. The 1% critical value for $\chi^2_{(1)}$ is 6.63, overwhelmingly rejecting the basic SV model in favor of the FSV or FISV model. This means that the FSV and FISV models are far more suitable for the data series. The log-likelihood values of the FSV and FISV models are very close to each other, with log-likelihood value of the FSV model being higher. Interestingly, the log-likelihood value for FSV is larger than that for FISV.

However, we note that the asymptotic normal distribution may poorly approximate the finite sample distribution of β if β is very close to one and the time span of the data is small or moderately large.

[11]The estimated coefficients of the basic SV model are almost identical to those provided in Sandmann and Koopman (1998) for the sample period running from 1928 to 1987.

Table 3: Estimation results from the SML method for the basic SV, FSV and FISV models. Numbers in the square brackets are the 90% confidence intervals. The last column is the log-likelihood ratio test for H_0: Basic SV and H_1 : FSV or FISV.

SML						
	H	β	σ_h	σ	lld	LR stat.
SV	-	0.982 [0.978,0.986]	0.166 [0.150,0.181]	0.008 [0.008,0.009]	38139	-
FSV	0.176 [0.144,0.207]	0.998 [0.996,0.999]	0.464 [0.402,0.526]	0.008 [0.008,0.009]	38168	59
FISV	0.142 [0.099,0.186]	0.998 [0.997,0.999]	0.360 [0.320,0.399]	0.008 [0.008,0.009]	38164	51

Second, the estimated autoregressive coefficient from both the FSV and FISV models is 0.998, suggesting that volatilities are highly persistent. In fact, our empirical estimates of β are very close to those obtained in Table 3 of Bolko et al. (2020) based on the generalized method of moment method, although we use the daily log returns and Bolko et al. (2020) use the daily realized volatilities.[12]

Third, the Hurst parameter H is estimated to be 0.176 and 0.142 for the FSV and FISV models, respectively. The 90% confidence intervals do not cover 0.5, a value that is assumed in the basic SV model. Consequently, both FSV and FISV models suggest that the log volatility process is rough. The point estimates are very close to what Gatheral et al. (2018) and Wang et al. (2021b) found from the volatility surface and the log realized volatility. They are slightly larger than the estimates obtained by Fukasawa et al. (2021) and Bolko et al. (2020) for log spot volatilities. We expect the estimated H in FSV to be larger if $\beta = 1$ is imposed. This assumption is equivalent to have $\gamma = 0$ in the fSV model, as in Fukasawa et al. (2021).

Table 4 presents estimation results for our data series with the Whittle method and the noise-robust local Whittle method. For the Whittle method, we again employ the grid searching method for the initial values of the optimization. We allow the parameter σ_ω^2 to be either fixed at the value $\pi^2/2$ (i.e., under the Gaussian assumption of ε_t) or unknown (i.e., ε_t is allowed to have a non-Gaussian distribution). Similar to those from SML, the estimated fractional parameters in FSV and FISV under both assumptions suggest rough and persistent volatility. However, the estimated fractional parameters from the Whittle method (0.292 and 0.346) are larger than those obtained from SML (0.176 and 0.142) when we allow the distribution of ε_t to be non-Gaussian and smaller (0.082 and 0.096) when assuming $\sigma_\omega^2 = \pi^2/2$. The estimated σ_ω^2 are slightly larger than 4.9. As expected, the LPWN method cannot separate a highly persistent autoregressive coefficient from the memory parameter and leads to a large

[12]Note that our β corresponds to $1 - \lambda$ in Bolko et al. (2020).

value for H, suggesting long memory and nonstationary volatility.

Table 4: Empirical Results from the Whittle and LPWN methods

	H	β	σ_h	H	β	σ_h	σ_ω^2
Whittle method	$\sigma_\omega^2 = \pi^2/2$			*Unknown* σ_ω^2			
SV	-	0.990	0.126	-	0.992	0.113	5.419
FSV	0.082	0.999	0.895	0.292	0.997	0.249	5.371
FISV	0.096	0.999	0.403	0.346	0.996	0.180	5.390
LPWN							
$(R_{short}, R_{noise}) = (1, 0)$	1.093						
$(R_{short}, R_{noise}) = (1, 1)$	1.093						

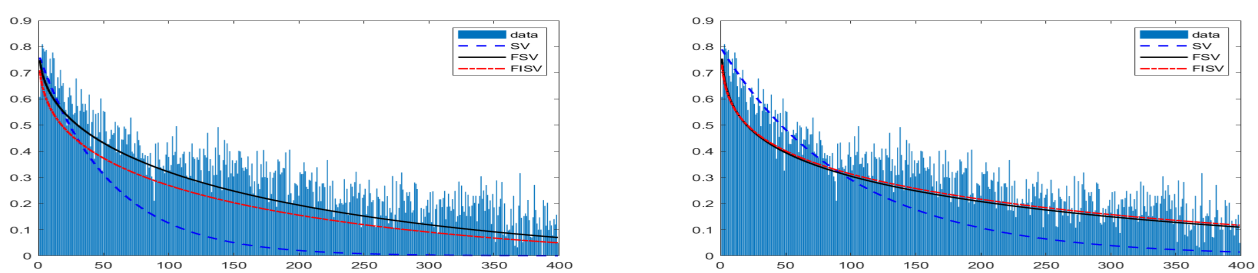
Interestingly, the Whittle estimates of FISV are remarkably different from those reported in Breidt et al. (1998) for the S&P 500 CRSP data from 1962 to 1987 where σ_ω^2 is assumed to be unknown. They find that the estimated H parameter is approximately 0.944 (or $d = 0.444$), and the autoregressive parameter β is around 0.932. The differences in the estimates may be due to the different datasets used in estimation.

The sample autocovariance of x_t, along with the theoretical implied autocovariance of the three models, are displayed in Figure 5. Overall, the autocovariance functions of the FSV and FISV models are near each other (reassuring of the estimation accuracy) and provide a much better fit for the autocovariance than the basic SV model for the data series. The three estimation methods lead to a slightly different fit of the ACF.

Figure 5: The sample autocovariance of x_t and the theoretical implied autocovariance

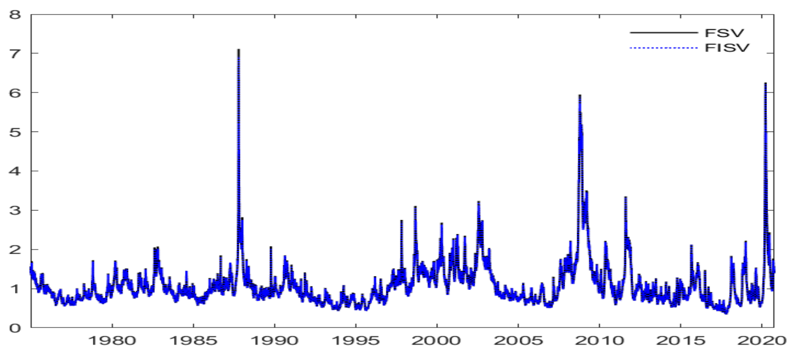
(a) SML

(b) Whittle (known σ_ω^2)



The smoothed estimates of volatilities (20) from the FSV and FISV models with SML are presented in Figure 6. The two volatility estimates are almost identical. The volatilities of the equity index increased to unprecedented levels in the 1987 stock market crash, during the 2008 subprime mortgage crisis period, and again at the onset of Covid-19 in March 2020.

Figure 6: The smoothed volatility estimates obtained from the FSV and FISV models: SML



6 Conclusions

This paper introduces a discrete-time fractional stochastic volatility model based on the fractional Gaussian noise, whose dynamic is governed by a Hurst parameter H. It includes the standard stochastic volatility model as a special case with $H = 0.5$ and is asymptotically equivalent to the fractional integrated stochastic volatility model with $H = d + 0.5$. We discuss its statistical properties, allowing the Hurst parameter to take values between 0 and 1.

We propose to estimate the model with a simulated maximum likelihood and a frequency domain quasi maximum likelihood method. For the SML method, the (time-domain) likelihood function is evaluated with the importance sampling technique, where the Laplace approximation determines the proposal distribution. The estimation method allows us to obtain both filtered and smoothed estimates of latent variables. Simulation studies show that the proposed SML and Whittle (frequency domain ML) methods can accurately estimate the FSV model. The SML method performs better than the Whittle method in terms of root mean square errors, whereas the Whittle method provides estimates with smaller biases for H.

We apply the proposed FSV model, along with the standard SV model and the FISV model, to the S&P 500 composite index over a long sampling period, spanning over 45 years. Our empirical results suggest that the log volatility of the S&P 500 index is persistent and rough. The estimated autoregressive coefficient of the log volatility process is very close to unity, and the estimated Hurst parameter of the FSV model is less than half. The latter is consistent with the findings of the recent literature on rough volatility.

The SML method is general enough to deal with more flexible model specifications. Specifically, suppose that ϵ_t in the return equation (4) or (23) is assumed to follow another parametric distribution (such as a student t distribution) to generate extra kurtosis in the distribution of r_t. One could simply modify the probability density function $f(r|h;\theta)$ in the SML algorithm to accommodate this change.

The FSV model could potentially be applied to daily log realized volatilities. Let RV_t and IV_t be the realize volatility and integrated volatility on day t. The realized volatility is computed from intraday returns and defined as the summation of the log squared returns within the day. RV_t is a consistent measure of the integrated variance $IV_t = \sigma^{*2} \int_{t-1}^{t} \exp(h_s)ds$ such that

$$RV_t = \sigma^{*2} \int_{t-1}^{t} \exp(h_s)ds + \nu_t, \tag{27}$$

where $\nu_t \overset{a}{\sim} N(0, 2IQ_t/M)$ with IQ_t being the integrated quarticity of day t and M being the number of intra-day returns within day t. See Barndorff-Nielsen and Shephard (2002). RV_t is recognized as an improved estimate of IV_t compared with the daily squared return.

Under the continuous-time specification of

$$dh_t = \gamma h_t dt + \sigma_h^* dB_t^H,$$

we have the following unique path-wise solution:

$$h_t = e^{-\gamma t} h_0 + \sigma_h^* \int_0^t e^{-\gamma(t-s)} dB_s^H, \tag{28}$$

where h_0 is the initial value. Hence,

$$h_t = e^{-\gamma} h_{t-1} + \sigma_h^* \int_{t-1}^{t} e^{-\gamma(t-s)} dB_s^H. \tag{29}$$

According to Theorem 8 of Bergstrom (1984), the time aggregation over a day leads to the following discrete-time model:

$$\int_{t-1}^{t} h_r dr = e^{-\gamma} \int_{t-2}^{t-1} h_r dr + \eta_t, \tag{30}$$

where η_t is a noise term that has an MA(1) structure.

Equations (27) and (30) form a discrete-time nonlinear state-space model that can be potentially estimated by the SML method proposed in this paper. We leave the study of the estimation approach based on realized volatility to a future study.

References

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 74*(1), 3–30.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(2), 253–280.

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

Beran, J. (1994). *Statistics for long-memory processes*. Chapman & Hall/CRC.

Bergstrom, A. R. (1984). Continuous time stochastic models and issues of aggregation over time. In Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics* (1 ed.), Volume 2. Elsevier.

Bertelli, S. and M. Caporin (2002). A note on calculating autocovariances of long-memory processes. *Journal of Time Series Analysis 23*(5), 503–508.

Bolko, A. E., K. Christensen, M. S. Pakkanen, and B. Veliyev (2020). Roughness in spot variance? a gmm approach for estimation of fractional log-normal stochastic volatility models using realized measures. *arXiv preprint arXiv:2010.04610*.

Bollerslev, T. and H. O. Mikkelsen (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics 73*(1), 151–184.

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.

Brockwell, P. J. and R. A. Davis (2009). *Time series: theory and methods*. Springer Science & Business Media.

Catania, L. (2021). A stochastic volatility model with a general leverage specification. *Journal of Business & Economic Statistics*.

Chung, C.-F. (1994). A note on calculating the autocovariances of the fractionally integrated arma models. *Economics Letters 45*(3), 293–297.

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

Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance 8*(4), 291–323.

Ding, Z., C. W. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*(1), 83–106.

Durbin, J. and S. J. Koopman (1997). Monte carlo maximum likelihood estimation of non-gaussian state space model. *Biometrika 84*(3), 669–684.

Frederiksen, P., F. S. Nielsen, and M. Ø. Nielsen (2012). Local polynomial whittle estimation of perturbed fractional processes. *Journal of Econometrics 167*(2), 426–447.

Fukasawa, M., T. Takabatake, and R. Westphal (2021). Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. *Mathematical Finance, forthcoming*.

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

Gouriéroux, C. and A. Monfort (1997). *Simulation-based Econometric Methods*. Oxford University Press, UK.

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.

Harvey, A. C. (2007). Long memory in stochastic volatility. In *Forecasting Volatility in the Financial Markets*, pp. 351–363. Elsevier.

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

Hurvich, C. M. and P. Soulier (2009). Stochastic volatility models with long memory. In *Handbook of Financial Time Series*, pp. 345–354. Springer.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies 65*(3), 361–393.

Kroese, D. P. and Z. I. Botev (2015). Spatial process simulation. In *Stochastic Geometry, Spatial Statistics and Random Fields*, pp. 369–404. Springer.

Liesenfeld, R. and J.-F. Richard (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance 10*(4), 505–531.

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

Marinucci, D. and P. M. Robinson (1999). Alternative forms of fractional brownian motion. *Journal of statistical planning and inference 80*(1-2), 111–122.

Priestley, M. B. (1981). *Spectral analysis and time series: probability and mathematical statistics*. Number ISBN 012564922-3. New York Academic Press.

Robinson, P. M. (2005). The distance between rival nonstationary fractional processes. *Journal of Econometrics 128*(2), 283–300.

Rosenbaum, M. (2008). Estimation of the volatility persistence in a discretely observed diffusion model. *Stochastic Processes and their Applications 118*(8), 1434–1462.

Sandmann, G. and S. J. Koopman (1998). Estimation of stochastic volatility models via monte carlo maximum likelihood. *Journal of Econometrics 87*(2), 271–301.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika 84*(3), 653–667.

Shi, S. and J. Yu (2021). Different strokes for different folks: long memory and roughness.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics 53*(1-3), 165–188.

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.

Wang, X., W. Xiao, and J. Yu (2021a). Asymptotic properties of least squares estimator in local to unity processes with fractional Gaussian noises. *Advances in Econometrics, forthcoming*.

Wang, X., W. Xiao, and J. Yu (2021b). Estimation and inference of the fractional continuous-time model with discrete-sampled data. *Journal of Econometrics, forthcoming*.

Yu, J. (2011). Simulation-based estimation methods for financial time series models. In *Handbook of Computational Finance*, pp. 427–465. Springer.

A Implementation Details

The optimization problem (14) can be solved numerically with the Newton-Raphson's method. Specifically, we start from an initial proposal $h_t^{(0)}$ and iterate recursively with the formula

$$h_t^{(k+1)} = h_t^{(k)} - \left[\left.\frac{\partial^2 \log f(r; h|\theta)}{\partial h \partial h'}\right|_{h_t = h_t^{(k)}} \right]^{-1} \left[\left.\frac{\partial \log f(r; h|\theta)}{\partial h}\right|_{h_t = h_t^{(k)}} \right],$$

where

$$\frac{\partial \log f(r; h|\theta)}{\partial h} = -\frac{1}{2} + \frac{1}{2} z - h' \Xi_\theta^{-1} \quad \text{and} \quad \frac{\partial^2 \log f(r; h|\theta)}{\partial h \partial h'} = -\frac{1}{2} diag(z) - \Xi_\theta^{-1}$$

with $z = [z_1, z_2, \cdots z_T]$.

The distributional approximation of $h|(\theta, r)$ or the optimization of (14) is conducted independently for every given θ. To ensure the smoothness of the likelihood function (13) with respect to θ,[13] it is essential that all importance sampling sequences $h^{(i)}$ are obtained as transformations of a common sequence of random draws. This is the so-called Common Random Numbers' technique. For our application, we use a fixed random seed to draw a random sequence of dimension $T \times N$ from the standard normal distribution, which is then transformed to have the distribution of $N(h_\theta^*(r), \Sigma_\theta^*(r))$.

Furthermore, to prevent overflow of the likelihood value, we apply some simple re-scaling techniques. Let $w^{(s)} = f(h^{(s)}|\theta) / g(h^{(s)})$. The log-likelihood function can be rewritten as

$$\begin{aligned}
\log \left[\frac{1}{S} \sum_{s=1}^{S} \exp \left(\log \frac{f(r; h^{(s)}|\theta)}{g(h^{(s)})} \right) \right] &= \log \left[\frac{1}{S} \sum_{s=1}^{S} f\left(r|h^{(s)}, \theta\right) w^{(s)} \right] \\
&= \log \left(\frac{1}{S} \sum_{s=1}^{S} w^{(s)} \right) + \log \left[\sum_{s=1}^{S} f\left(r|h^{(s)}, \theta\right) w^{*(s)} \right] \\
&= \log \left(\frac{1}{S} \sum_{s=1}^{S} \exp\left(A^{(s)}\right) \right) + \log \left[\sum_{s=1}^{S} \exp\left(B^{(s)}\right) w^{*(s)} \right], \qquad (31)
\end{aligned}$$

where $A^{(s)} = \log w^{(s)} = \log f(h^{(s)}|\theta) - \log g(h^{(s)})$, $B^{(s)} = \log f(r|h^{(s)}, \theta)$, and

$$w^{*(s)} = \frac{w^{(s)}}{\sum_{s=1}^{S} w^{(s)}} = \frac{\exp\left(A^{(s)}\right)}{\sum_{s=1}^{S} \exp\left(A^{(s)}\right)}.$$

The computation of the likelihood involves exponential functions of $A^{(s)}$ and $B^{(s)}$, which could potentially result in a numeric value that is outside of the range of computer precision and hence compromise the reliability of the program. To avoid such an overflow condition, we rescale $A^{(s)}$ and $B^{(s)}$ such that

$$\begin{aligned}
w^{*(s)} &= \frac{\exp\left(A^{(s)} + C_1\right)}{\sum_{s=1}^{S} \exp\left(A^{(s)} + C_1\right)}, \\
\log \left(\frac{1}{S} \sum_{s=1}^{S} \exp\left(A^{(s)}\right) \right) &= -C_1 + \log \left(\frac{1}{S} \sum_{s=1}^{S} \exp\left(A^{(s)} + C_1\right) \right), \\
\log \left[\sum_{s=1}^{S} \exp\left(B^{(s)}\right) w^{*(s)} \right] &= -C_2 + \log \left[\sum_{s=1}^{S} \exp\left(B^{(s)} + C_2\right) w^{*(s)} \right],
\end{aligned}$$

[13] Smoothness is essential for the numerical convergence of an optimization algorithm. See, for example, Gouriéroux and Monfort (1997).

where $C_1 = -\max_s \left\{A^{(s)}\right\} + 1$ and $C_2 = -\max_s \left\{B^{(s)}\right\} + 1$. It follows that the log-likelihood function

$$\begin{aligned} & \log\left[\frac{1}{S}\sum_{s=1}^{S}\exp\left(\log\frac{f\left(r;h^{(s)}|\theta\right)}{g\left(h^{(s)}\right)}\right)\right] \\ = \quad & -C_1 - C_2 + \log\left(\frac{1}{S}\sum_{s=1}^{S}\exp\left(A^{(s)}+C_1\right)\right) + \log\left[\sum_{s=1}^{S}\exp\left(B^{(s)}+C_2\right)\frac{\exp\left(A^{(s)}+C_1\right)}{\sum_{s=1}^{S}\exp\left(A^{(s)}+C_1\right)}\right]. \end{aligned}$$

B Estimation of the FISV model

The estimation of the FISV model with SML requires the variance-covariance matrix of h_t. The autocovariance of h_t be can again computed from

$$\gamma_h(k) = \sum_{s=-\infty}^{\infty} \tilde{\gamma}(s)\,\gamma_u(k-s), \tag{32}$$

with $\tilde{\gamma}(s) = \sigma_h^2 \beta^s / \left(1-\beta^2\right)$ and

$$\gamma_u(k) = \frac{(-1)^k\,\Gamma(1-2d)}{\Gamma(k-d+1)\,\Gamma(1-k-d)}. \tag{33}$$

Alternative expressions of the auto-covariance function are provided by Hosking (1981, Lemma 1), Sowell (1992), and Chung (1994). All three forms involves the hypergeometric function, which are more computational intensive than the form we employed here.

The spectral density of x_t under the model specification of (23)-(25), which is required for the Whittle method, is

$$f_x(\lambda) = \frac{\sigma_u^2}{2\pi}\frac{\left(\sqrt{2-2\cos(\lambda)}\right)^{-2d}}{1-2\beta\cos(\lambda)+\beta^2} + \frac{\sigma_\omega^2}{2\pi}.$$

It follows that the Whittle log likelihood function

$$L_n(\theta) = -\pi\log 2\pi - d\frac{2\pi}{n}\sum_{k=1}^{n/2}\log\left(2-2\cos(\lambda_k)\right) + \frac{2\pi}{n}\sum_{k=1}^{n/2}\log\psi_\theta(\lambda_k) \tag{34}$$

$$+\frac{4\pi^2}{n}\sum_{k=1}^{n/2}\frac{\left(2-2\cos(\lambda)\right)^d I_n(\lambda_k)}{\psi_\theta(\lambda_k)}. \tag{35}$$

where $\psi_\theta(\lambda_k) = \sigma_h^2/\left[1-2\beta\cos(\lambda)+\beta^2\right] + \sigma_\omega^2\left(2-2\cos(\lambda_k)\right)^d$. Breidt et al. (1998) argue that the term $-d\frac{2\pi}{n}\sum_{k=1}^{n/2}\log\left(2-2\cos(\lambda_k)\right)$ in the log likelihood is negligible, and hence, was omitted in their empirical application. In our paper, we employ the original log likelihood function for the purpose of more accurate evaluations of the log-likelihood.