11-2020

Persistent and rough volatility

Xiaobin LIU

Shuping SHI

Jun YU
Singapore Management University, yujun@smu.edu.sg

SMU ECONOMICS & STATISTICS

SMU SINGAPORE MANAGEMENT UNIVERSITY

Persistent and Rough Volatility

Xiaobin Liu, Shuping Shi, Jun Yu

November 2020

Paper No. 23-2020

Persistent and Rough Volatility*

Xiaobin Liu
Zhejiang University

Shuping Shi
Macquarie University

Jun Yu
Singapore Management University

November 3, 2020

Abstract

This paper contributes to an ongoing debate on volatility dynamics. We introduce a discrete-time fractional stochastic volatility (FSV) model based on the fractional Gaussian noise. The new model has the same limit as the fractional integrated stochastic volatility (FISV) model under the in-fill asymptotic scheme. We study the theoretical properties of both models and introduce a memory signature plot for a model-free initial assessment. A simulated maximum likelihood (SML) method, which maximizes the time-domain log-likelihoods obtained by the importance sampling technique, is employed to estimate the model parameters. Simulation studies suggest that the SML method can accurately estimate both models. Our empirical analysis of several financial assets reveals that volatilities are both persistent and rough. It is persistent in the sense that the estimated autoregressive coefficients of the log volatilities are very close to unity, which explains the observed long-range dependent feature of volatilities. It is rough as the estimated Hurst (fractional) parameters of the FSV (FISV) model are significantly less than half (zero), which is consistent with the findings of the recent literature on 'rough volatility'.

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

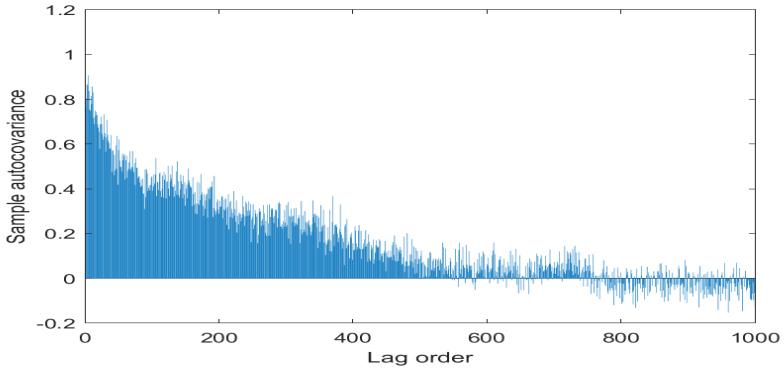
1 Introduction

Volatility is central to risk management, portfolio allocation, and the pricing of financial assets. Temporal dependence in volatility has been one of the most studied problems in financial econometrics. One prominent feature of volatility dynamic 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 over the period from January 2, 1964 to October 1, 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 (demeaned) returns of the S&P 500 index over the period from January 2, 1964 to October 1, 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

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

with L being the lag operator, and the fractional parameter $d \in (0, 0.5)$. This process has a long memory 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:

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

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 parameter in the literature[2] and assumed to be $H \in (0.5, 1)$ in Comte and Renault (1996, 1998) and Rosenbaum (2008). Under this setting of H, the fBM has a long memory in the sense that its autocovariance function decays at a hyperbolic rate and the one-sided long-run variance $\sum_{n=0}^{\infty} \mathbb{E}\left(B_1^H (B_{n+1}^H - B_n^H)\right) = \infty$.

[1] The fractional Brownian motion becomes the standard Brownian motion W_t when $H = 0.5$ because $\mathbb{E}(B_t^H B_s^H) = \min\{t, s\}$.

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

The long-range dependence of volatilities is captured by either the fractional integrated process (1) with $d \in (0, 0.5)$ or the fBM process B_t^H with $H \in (0.5, 1)$. In fact, by letting $d = H - 1/2$, one can show under the in-fill asymptotic scheme the log volatility process of the FISV model converge to h_t in the fSV model (Davydov, 1970; Tanaka, 2013). The volatility trajectory is smooth and its long-run variance diverges to infinity under those model specifications.

There is, however, a burgeoning literature documenting evidence that the Hurst parameter of the fSV model is smaller than 0.5 for volatility time series and volatility surface over the past several years. See, for example, Gatheral et al. (2018); Fukasawa et al. (2019); Wang et al. (2019); Bennedsen et al. (2017); Bolko et al. (2020). Interestingly, when $H \in (0, 0.5)$ the fBM is negatively correlated, the sample path generated by B_t^H is rough, and the long-run variance (both two-sided and one-sided) of fBM tends to a constant. In other words, contradicting the traditional view on volatility dynamics, they believe the sample path of the volatility process is rough. As a result, a new generation of stochastic volatility (SV) models, coined in Gatheral et al. (2018) as 'rough volatility models', appeared both in the literature on volatility modeling and the literature on asset pricing. A publicly accessible website contains more than 100 papers written on this subject matter.[3] Several studies document the superior performance of the rough SV model in forecasting volatility relative to long-memory volatility models (Gatheral et al., 2018; Bennedsen et al., 2017; Wang et al., 2019). Gatheral et al. (2018) show via simulations that standard tests (Andersen et al., 2001, 2003) might identify 'spurious long memory' when the data are actually generated from a rough volatility model.

The contribution of this paper is four folded. First, we consider an Euler discretized version of the fSV model, with a more general setting of the Hurst parameter. We refer to the discrete-time model as FSV. Unlike Comte and Renault (1996, 1998) and Rosenbaum (2008), we assume the Hurst parameter $H \in (0, 1)$. Properties of autocovariances and long-run variance of log volatilities are examined for various ranges of H (i.e., $H \in (0, 0.5)$, $H = 0.5$, and $H \in (0.5, 1)$) and compared with those of the FISV model with $d \in (-0.5, 0)$, $d = 0$, and $d \in (0, 0.5)$. The long-run variance of log volatilties diverges under the specfication of FSV with $H \in (0.5, 1)$ or under the FISV specification with $d \in (0, 0.5)$ and converges to a constant when $H \in (0, 0.5]$ or $d \in (-0.5, 0]$.

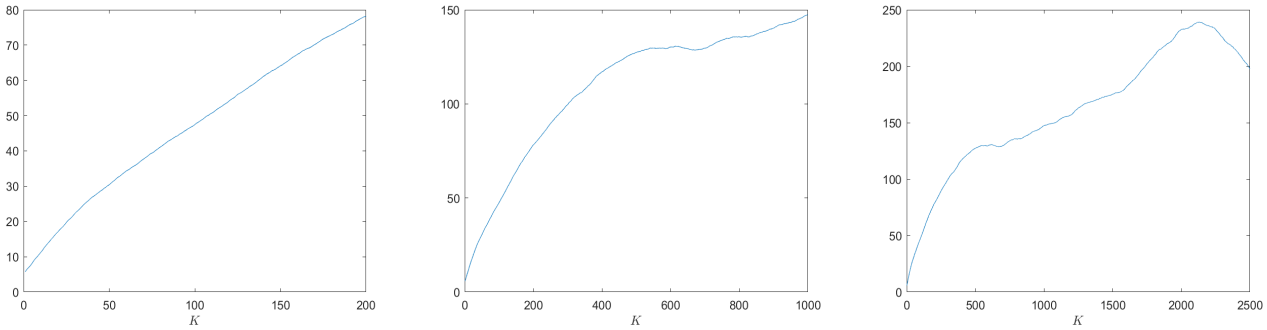
Second, we introduce a memory signature plot that provides a simple and 'model-free' assessment for the memory parameter (H or d). The memory signature plot displays the dynamic of the cumulative sums of autocovariances of log volatilities, which is expected to diverge when $H \in (0.5, 1)$ or $d \in (0, 0.5)$ and be bounded otherwise. As such, it can reveal the possible range of the memory parameter. Surprisingly, despite the considerable empirical evidence on long memory volatilities (Andersen and Bollerslev, 1997; Breidt et al., 1998; Bollerslev and Wright, 2000; Andersen et al., 2003; Bollerslev et al., 2000), we do not find divergences of the cumulative sums for all daily data series considered in our empirical studies. For example, Figure 2 shows the memory signature plots of the S&P 500 index from 1964 to 1988 with $K = 200$, $K = 1,000$, and $K = 2,500$. The y-axes are the cumulative sums of autocovariances $\sum_{k=0}^{K} \hat{\gamma}(k)$ with $\hat{\gamma}(k)$ being the sample autocovariance of log volatilities at lag order

[3] https://sites.google.com/site/roughvol/home/risks-1.

k, and the x-axes are K. As we can see, $\sum_{k=0}^{K} \hat{\gamma}(k)$ rises rapidly initially but eventually declines as K increases. This pattern is in strong agreement with those of the FSV model when $H \in (0, 0.5)$ and the FISV model when $d \in (-0.5, 0)$.

Figure 2: The memory signature plots of the S&P 500 from 1964 to 1988: the y-axes are the cumulative sums of autocovariances $\sum_{k=0}^{K} \hat{\gamma}(k)$ and the x-axes are the value of K.

(a) $K = 200$ (b) $K = 1000$ (c) $K = 2500$



Third, we propose a simulated maximum likelihood (SML) method to estimate the FSV and FISV models. The likelihood functions are computed in the time domain with an importance sampling technique, instead of the frequency domain as in Fukasawa et al. (2019) for the fSV model and Breidt et al. (1998) for the FISV model.[4] We treat the log volatility as a latent variable and construct the likelihoods from log prices. This is in contrast to Gatheral et al. (2018) and Wang et al. (2019) who assume h_t is observable. Our approach is also different from the quasi-maximum likelihood (QML) method of Fukasawa et al. (2019) and the generalized method of moment (GMM) of Bolko et al. (2020), where the spectral density is obtained from realized volatilities and the moments is obtained from integrated volatilities, respectively. Since the time-domain likelihood function is maximized, the SML method is expected to be asymptotically efficient. Also, one can use the classical asymptotic theory to make statistical inferences. Our simulation results show that the SML method can provide reasonably accurate estimations for all model parameters in both FSV and FISV.

Lastly, our estimation results reveal that volatilities are both persistent and rough. With a less stringent assumption on the memory parameter and a new estimation method, we show that the 'long-range dependence' of volatilities is due to the autoregressive coefficient that is very close to unity (a.k.a. persistent). Meanwhile, roughness is also present in the data and driven by the memory parameter $H < 0.5$ or $d < 0$. Specifically, we estimate both the FSV and FISV models with the SML method for four financial assets (S&P 500, Nikkei 225, EURO/USD, and GBP/USD) over two 25-year periods.

[4]Fukasawa et al. (2019) propose to approximate the Whittle likelihood (i.e., the likelihood function in the frequency domain) of increments of the log daily realized volatilities from the fSV model with $\gamma = 0$. Breidt et al. (1998) propose to maximize the Whittle likelihood of returns from the FSV model given by equations (16)-(17). Unfortunately, Dahlhaus (1988) shows that the Whittle estimators may perform poorly in finite samples for autoregressive processes, particularly when the characteristic root of the time series is close to unity, which is the exact situation we are facing empirically, as it will become clear later.

The first period starts from their first available observation from DataStream (which varies for each data series) and lasts for 25 years. This sample period allows us to connect with the earlier empirical literature on long memory volatilities. The second sample is on the most recent 25 years running from 1996 to 2020, aligning with the more recent research on rough volatility. The estimated autoregressive parameters of log volatilities fall between 0.99 and unity for all data series, suggesting very persistent dynamics of log volatilities. The estimated Hurst parameters of the FSV model are less than 0.5, ranging between 0.08 to 0.25, while the estimated fractional parameters of the FISV model are negative and fall between -0.27 to -0.49. Our analysis supports the recent literature on rough volatility.

Our paper is related to Bennedsen et al. (2017) where two continuous-time SV models are introduced to incorporate roughness and persistence in volatilities. The first model is based on the Cauchy process of Gneiting and Schlather (2004) and found to perform well in volatility-forecasting (Bennedsen et al., 2017). The second one is a Brownian semistationary (BSS) process originally introduced in Barndorff-Nielsen and Schmiegel (2008). In Bennedsen et al. (2017), roughness occurs at short lags of the autocorrelations of volatilities, while long memory is at longer lags. Both the Cauchy and BSS processes can generate roughness and long memory in volatilities if the kernel function is selected carefully. Unlike the fSV model that has been used to price options (Bayer et al., 2016), how to apply the Cauchy process and BSS to pricing financial assets is yet to know.

The paper is organized as follows. In Section 1, we provide a primer on the fractional Gaussian noise that is the basis for the fBM. Section 2 introduces the FSV model and derives its statistical properties, and Section 3 is on the FISV model. Section 4 discusses the SML method in the time domain. Section 5 checks the performance of the proposed SML estimator using data simulated from the FSV and FISV models. Section 6 applies the FSV and FISV models and the SML estimation method to the log prices of several financial assets, including stock markets and foreign exchange markets. Section 7 concludes the paper. The appendix collects the proof of theoretic results and additional results for the empirical and simulation studies.

2 A Primer on Fractional Gaussian Noise

Let η_t^H denote a fractional Gaussian noise and $\gamma_\eta(k) = Cov\left(\eta_t^H, \eta_s^H\right)$ with $k = |t-s|$ be the autocovariance of η_t^H. The autocovariance has the following expression,

$$\gamma_\eta(k) = \frac{1}{2}\left[(k+1)^{2H} + (k-1)^{2H} - 2k^{2H}\right]. \tag{4}$$

The variance $\gamma_\eta(0) = 1$. When $H = 1/2$, the fractional Gaussian noise becomes the standard Gaussian noise and $\gamma_\eta(k) = 0$ for all $k \neq 0$. If $H \neq 1/2$, $\gamma_\eta(k) \neq 0$ and, by the first order Taylor series expansion,

$$\gamma_\eta(k) = \frac{1}{2}k^{2H}\left[\left(1+\frac{1}{k}\right)^{2H} + \left(1-\frac{1}{k}\right)^{2H} - 2\right] \sim H(2H-1)k^{2H-2} \tag{5}$$

for large k. The autocovariance $\gamma_\eta(k)$ decays at a hyperbolic rate as k goes to infinity. See the top row of Figure 3 for an illustration.

When $H > 1/2$, the one-sided long-run variance is approximately

$$\sum_{k=0}^{\infty} \gamma_\eta(k) \sim H(2H-1) \sum_{k=0}^{\infty} k^{2H-2} \to \infty,$$

and hence, the two-sided long-run variance $\sum_{k=-\infty}^{\infty} \gamma_\eta(k) = 1 + 2\sum_{k=1}^{\infty} \gamma_\eta(k) \to \infty$.[5] When $H < 1/2$, the one-sided long-run variance of η_t^H is

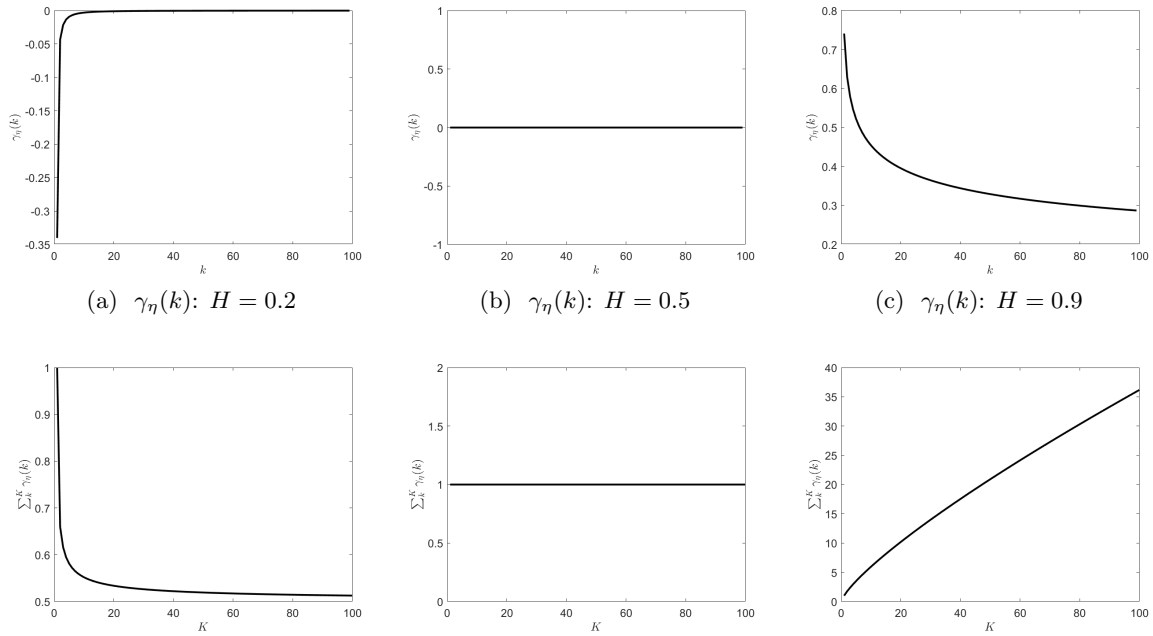
$$\sum_{k=0}^{\infty} \gamma_\eta(k) = 1 + \sum_{k=1}^{\infty} \frac{1}{2}\left[(k+1)^{2H} + (k-1)^{2H} - 2k^{2H}\right] = 0.5,$$

and the two-sided long-run variance is

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

See the bottom row of Figure 3 for an illustration of the one-sided long-run variance.

In summary, if $H > 1/2$, η_t^H is positively autocorrelated and has a long memory since $\sum_{k=-\infty}^{\infty} \gamma_\eta(k) = \infty$, whereas if $H < 1/2$, η_t^H is negatively autocorrelated and anti-persistent with $\sum_{k=-\infty}^{\infty} \gamma_\eta(k) = 0$.



(a) $\gamma_\eta(k)$: $H = 0.2$ (b) $\gamma_\eta(k)$: $H = 0.5$ (c) $\gamma_\eta(k)$: $H = 0.9$

(d) $\sum_{k=0}^{K} \gamma_\eta(k)$: $H = 0.2$ (e) $\sum_{k=0}^{K} \gamma_\eta(k)$: $H = 0.5$ (f) $\sum_{k=0}^{K} \gamma_\eta(k)$: $H = 0.9$

Figure 3: The dynamics of $\gamma_\eta(k)$ (with $k = 1, \cdots, 100$) and $\sum_{k=0}^{K} \gamma_\eta(k)$ (with $K = 0, \cdots, 100$).

[5] By the Cauchy condensation test, $\sum_{k=0}^{\infty} k^{2H-2}$ converges if and only if $\sum_{k=0}^{\infty} 2^{k(2H-1)}$ converges.

3 Fractional Stochastic Volatility Model

Assume there are $T+1$ equidistant log prices over the period $[0, N]$. The interval between observations is Δ and $N = T\Delta$. Log returns $r_{t\Delta} = y_{t\Delta} - y_{(t-1)\Delta}$ are available on grid $t\Delta$ with $t = 1, 2, \ldots, T$. Applying the Euler approximation to model (2), we have the following discrete-time fractional SV model

$$r_{t\Delta} = \sigma e^{h_{t\Delta}/2}\varepsilon_t, \tag{6}$$

$$h_{t\Delta} = \beta h_{(t-1)\Delta} + \sigma_\eta^* \eta_{t\Delta}^H, \tag{7}$$

where $\beta = 1 + \gamma\Delta$, $\varepsilon_t \overset{iid}{\sim} N(0,1)$, $\eta_{t\Delta}^H = B_{t\Delta}^H - B_{(t-1)\Delta}^H$ is a fractional Gaussian noise, and $\sigma = \sigma^*\sqrt{\Delta}$. We assume $\gamma < 0$ in subsequent analysis. Note that, by the self-similarity property of fBM,

$$Cov\left(\eta_{t\Delta}^H, \eta_{s\Delta}^H\right) = \Delta^{2H}\gamma_\eta(k) \text{ with } k = |t-s|.$$

When the time interval N is fixed and $\Delta \to 0$, the AR coefficient $\beta = 1 + \gamma N/T \to 1$. The dynamic of $h_{t\Delta}$ is local-to-unity with the local parameter γN. When Δ is fixed and $\gamma \to 0$, the process $h_{t\Delta}$ follows a fractional random walk process and its first difference $h_{t\Delta} - h_{(t-1)\Delta}$ is a fractional Gaussian noise.

3.1 Properties of h_t : when the initialization is in the infinite past

We discuss the dynamic properties of $h_{t\Delta}$ in this section. By recursive substitution, equation (7) can be rewritten as

$$h_{t\Delta} = \sigma_\eta^* \sum_{i=0}^{\infty} \beta^i \eta_{(t-i)\Delta}^H. \tag{8}$$

Assume $s \geq t$. The autocovariance between $h_{t\Delta}$ and $h_{s\Delta}$, denoted by $\gamma_h(k) \equiv Cov(h_{t\Delta}, h_{s\Delta})$ with $k = s - t$, is[6]

$$\gamma_h(k) = \sigma_\eta^{*2}\Delta^{2H} \sum_{i=k}^{\infty}\sum_{j=0}^{\infty} \beta^{i+j-k}\gamma_\eta(|i-j|). \tag{9}$$

Standard Gaussian noise: $H = 0.5$

We have the basic SV model when $H = 0.5$. The autocovariance $\gamma_\eta(k) = 0$ for all $k \neq 0$ under this setting. It follows from (9) that the autocovariance function of $h_{t\Delta}$ is

$$\gamma_h(k) = \sigma_\eta^{*2}\Delta \sum_{i=k}^{\infty} \beta^{2i-k}\gamma_\eta(0) = \sigma_\eta^{*2}\Delta \sum_{i=k}^{\infty} \beta^{2i-k} = \sigma_\eta^{*2}\Delta \frac{\beta^k}{1-\beta^2}. \tag{10}$$

This autocovariance function decreases geometrically as k increases. The variance of $h_{t\Delta}$ is $\gamma_h(0) = \sigma_\eta^{*2}\Delta/\left(1-\beta^2\right)$. The one-sided long-run variance is

$$\sum_{k=0}^{\infty} \gamma_h(k) = \frac{\sigma_\eta^{*2}\Delta}{(1-\beta^2)(1-\beta)}.$$

[6] We slightly abuse the notations as we let $\gamma_\eta(k) \equiv Cov(\eta_t, \eta_s)$ but let $\gamma_h(k) \equiv Cov(h_{t\Delta}, h_{s\Delta})$.

Fractional Gaussian noise: $H \neq 0.5$

Now, we consider the case of $H \neq 0.5$. The η_t^H process follows a fractional Gaussian noise under this setting. It has a long memory when $H > 0.5$ and is anti-persistent when $H < 0.5$. We now explore the properties of $\gamma_h(0)$ and $\gamma_h(k)$ with both $H < 0.5$ and $H > 0.5$.

Theorem 3.1 *Let* $\varsigma_{H,\beta,s} \equiv \sum_{l=s}^{\infty} \beta^l \gamma_\eta(l)$. *When* $H \neq 0.5$, *the variance is*

$$\gamma_h(0) = \frac{\sigma_\eta^{*2} \Delta^{2H}}{1-\beta^2} \left[1 + 2\varsigma_{H,\beta,1}\right] < \infty, \tag{11}$$

and the autocovariance function is

$$\gamma_h(k) = \frac{\sigma_\eta^{*2} \Delta^{2H}}{1-\beta^2} \beta^k \left[1 + \sum_{l=1}^{k} \beta^{-l} \gamma_\eta(l) + \beta^{-2k} \varsigma_{H,\beta,k+1} + \varsigma_{H,\beta,1}\right] < \infty. \tag{12}$$

The one-sided long-run variance is

$$\sum_{k=0}^{\infty} \gamma_h(k) < \infty \text{ if } H < 0.5 \text{ and } \sum_{k=0}^{\infty} \gamma_h(k) = \infty \text{ if } H > 0.5.$$

When $\beta \in (0,1)$ *and* $H > 0.5$, $\gamma_h(k) > 0$ *for any* k.

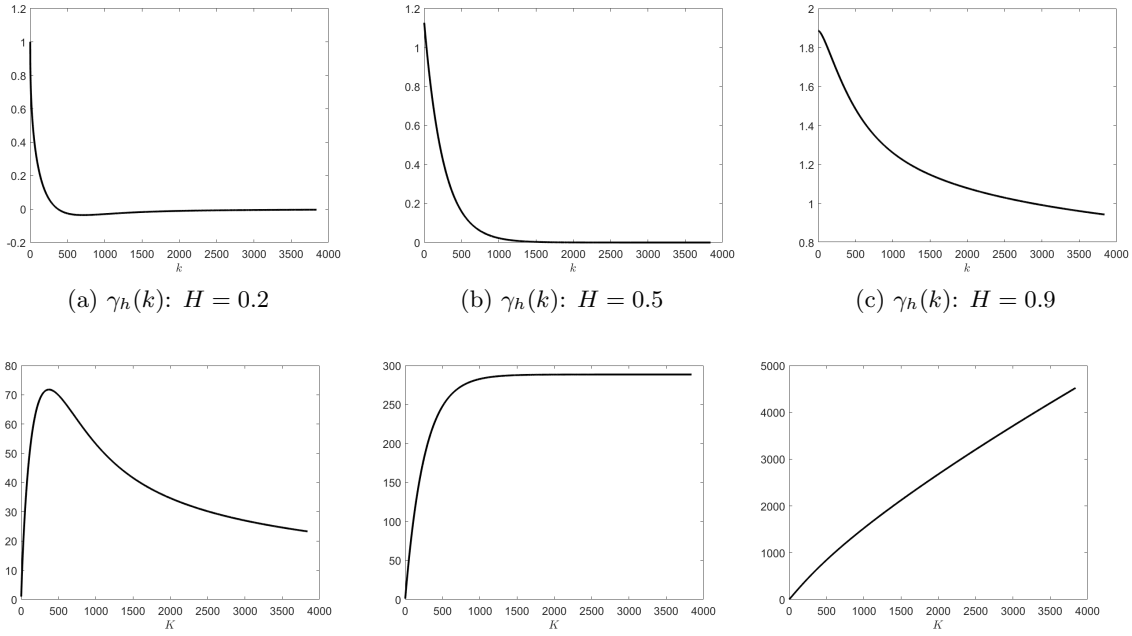
According to Theorem 3.1, the behavior of $h_{t\Delta}$ is similar to that of $\eta_{t\Delta}^H$ in the sense that the one-sided long-run variance is bounded when $H < 0.5$ and diverges to infinity when $H > 0.5$. Equations (11) and (12) provide a convenient way for the computation of $\gamma_h(k)$ and hence, the variance-covariance matrix of $h_{t\Delta}$ which is required for the SML method detailed in the next section.

As an illustration, we plot the dynamics of $\gamma_h(k)$ and $\sum_{k=0}^{K} \gamma_h(k)$ with various settings of H. We set the maximum number of l in $\zeta_{H,\beta,s}$ to $10,000$ and consider the following parameter setting: $N = 15$, $\sigma_\eta^* = 1.5$, $\gamma = -1$, and $\Delta = 1/256$, according to the empirical results. The autocovariance functions of $h_{t\Delta}$ against the lag order k are displayed in the top row of Figure 4, with $H = \{0.2, 0.5, 0.9\}$, while the cumulative sums of autocovariances are plotted in the second row. The autocovariance function decays as k increases. The rate of decaying of the autocovariance is the slowest when the process has a long memory (i.e., $H = 0.9$) and fastest when it is rough (i.e., $H = 0.2$). The cumulative sums of autocovariances converge when $H \leq 0.5$ and diverge when $H = 0.9$. When $H = 0.2$, the cumulative sums first go up and then go down. This feature generally holds for any $H \in (0, 0.5)$.

3.2 Properties of h_t : when the initialization is at period zero

Suppose the h_t process is initiated at period zero, instead of the infinite past as in (8). One can rewrite equation (7) as

$$h_{t\Delta} = \sigma_\eta^* \sum_{i=1}^{t} \beta^{t-i} \eta_{i\Delta}^H + h_0,$$



(a) $\gamma_h(k)$: $H = 0.2$ (b) $\gamma_h(k)$: $H = 0.5$ (c) $\gamma_h(k)$: $H = 0.9$

(d) $\sum_{k=0}^{K} \gamma_h(k)$: $H = 0.2$ (e) $\sum_{k=0}^{K} \gamma_h(k)$: $H = 0.5$ (f) $\sum_{k=0}^{K} \gamma_h(k)$: $H = 0.9$

Figure 4: The dynamics of $\gamma_h(k)$ (with $k = 0, \cdots, N/\Delta$) and $\sum_{k=0}^{K} \gamma_h(k)$ (with $K = 0, \cdots, N/\Delta$): $\gamma = -1$, $\sigma_\eta^* = 1.5$, $\Delta = 1/256$, and $N = 15$.

with h_0 being the initial value. The autocovariance between $h_{t\Delta}$ and $h_{s\Delta}$ is

$$Cov\left(h_{t\Delta}, h_{s\Delta}\right) = \sigma_\eta^{*2} \Delta^{2H} \sum_{i=1}^{t} \sum_{j=1}^{s} \beta^{t+s-i-j} \gamma_\eta\left(|i-j|\right). \tag{13}$$

Consider the long-span asymptotic that Δ is fixed and $N \to \infty$. The autocovariance function has the following limiting properties.

Theorem 3.2 *Under the long-span asymptotic scheme (Δ fixed, $N \to \infty$, and $\gamma < 0$), as $t \to \infty$, when $H = 0.5$,*

$$Cov\left(h_{t\Delta}, h_{s\Delta}\right) \to \sigma_\eta^{*2} \Delta \frac{\beta^k}{1-\beta^2};$$

when $H \neq 0.5$, the variance is

$$\gamma_h(0) \to \frac{\sigma_\eta^{*2} \Delta^{2H}}{1-\beta^2} \left(1 + 2\varsigma_{H,\beta,1}\right), \tag{14}$$

and the autocovariance function is

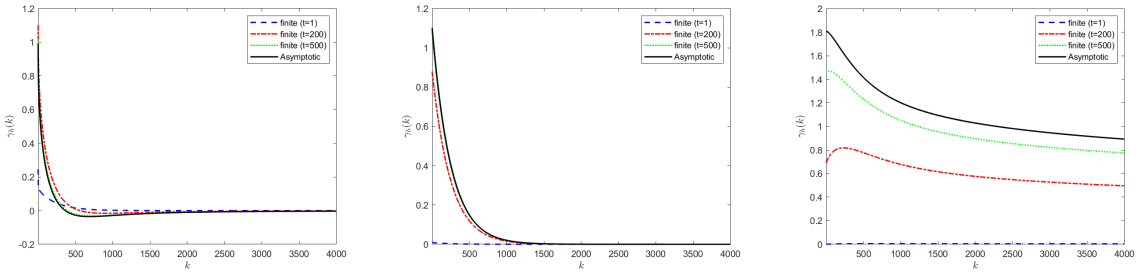
$$\gamma_h(k) \to \frac{\sigma_\eta^{*2} \Delta^{2H}}{1-\beta^2} \beta^k \left[1 + \sum_{l=1}^{k} \beta^{-l} \gamma_\eta(l) + \beta^{-2k} \varsigma_{H,\beta,k+1} + \varsigma_{H,\beta,1}\right]. \tag{15}$$

Theorem 3.2 shows that the autocovariance function (13) converges to (9) under the long-span asymptotic scheme. As such, we refer to (13) as the finite sample autocovariance, with (9) being its

asymptotic counterpart. Figure 5 shows the finite sample autocovariance function when $t = 1$, 500, 1000, along with their corresponding asymptotic function. The model parameter settings are the same as in those for Figure 4. It is obvious that the speed of convergence depends on H. The gaps between the finite sample and asymptotic autocovariance functions are wider when $H = 0.9$ than those when $H = \{0.2, 0.5\}$.

Figure 5: The convergence of the finite sample autocovariance function: $\gamma = -1$, $\sigma_\eta^* = 1.5$, and $\Delta = 1/256$.

(a) $H = 0.2$ (b) $H = 0.5$ (c) $H = 0.9$



4 Fractional Integrated Stochastic Volatility Model

The fractional integrated stochastic volatility model of Breidt et al. (1998) takes the expression

$$r_{t\Delta} = \sigma e^{h_{t\Delta}/2}\varepsilon_t,\ \varepsilon_t \overset{iid}{\sim} N(0,1), \tag{16}$$

$$h_{t\Delta} = \rho h_{(t-1)\Delta} + \sigma_u u_t, \tag{17}$$

where u_t is a stationary fractional integrated process and takes the form of (1). The model of Harvey (2007) is special case of the above model with $\rho = 0$ in (17). Here, we allow the fractional parameter d to be between -0.5 and 0.5, that is, $d \in (-0.5, 0.5)$.

The fractional difference operator $(1-L)^d$ in (1) has the expression

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

where $\Gamma(\cdot)$ is a gamma function defined by

$$\Gamma(x) := \begin{cases} \int_0^\infty t^{x-1}e^{-t}dt, & x > 0 \\ \infty, & x = 0 \\ x^{-1}\Gamma(1+x), & x < 0 \end{cases}.$$

The autocovariance function of u_t, denoted by $\gamma_u(k)$, is

$$\gamma_u(k) := \mathbb{E}(u_t u_{t-k}) = \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)}. \tag{18}$$

See, e.g., Page 468 in Brockwell and Davis (1987).

The FISV model is closely linked to the fractional stochastic volatility model (2). By letting

$$d = H - 1/2,\ \rho = e^{\gamma\Delta},\ \sigma_u^2 = \frac{e^{2\gamma\Delta}-1}{2\gamma}\sigma_\eta^{*2},$$

we have the following weak convergence. As $\Delta \to 0$,

$$\frac{\delta_H \Gamma(H+1/2)}{T^H} h_{\lfloor Ts \rfloor \Delta} \Rightarrow h_s\,, \forall\ 0 \leq s \leq 1\,, \tag{19}$$

where $\delta_H = \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}}$, $h_{\lfloor Ts \rfloor \Delta}$ is the process defined in (17) with $\lfloor . \rfloor$ denoting the integer part of the argument, and h_s is the process defined in equation (2). See Davydov (1970) and Tanaka (2013).

4.1 Properties of $h_{t\Delta}$

In analogy to the FSV model, let the autocovariance $\gamma_h(k) \equiv Cov(h_{t\Delta}, h_{s\Delta})$. It either takes the finite-sample form

$$\gamma_h(k) = \sigma_u^2 \sum_{i=1}^{t}\sum_{j=1}^{s} \rho^{t+s-i-j}\gamma_u(|i-j|)\,, \tag{20}$$

or the asymptotic form

$$\gamma_h(k) = \sigma_u^2 \sum_{i=k}^{\infty}\sum_{j=0}^{\infty} \rho^{i+j-k}\gamma_u(|i-j|)\,, \tag{21}$$

with $\gamma_u(\cdot)$ defined in (18). From Hosking (1981, Lemma 1(c)), the asymptotic autocovariance (21) has the expression

$$\gamma_h(k) = \frac{\sigma_u^2}{1-\rho^2}\gamma_u(k)\left[F(1, d+k; 1-d+k; \rho) + F(1, d-k; 1-d-k; \rho) - 1\right], \tag{22}$$

where $F(\cdot)$ being the hypergeometric function.[7] It could also take the form of Sowell (1992, eq. (8)-(9))

$$\gamma_h(k) = \frac{\sigma_u^2}{\rho(1-\rho^2)}\gamma_u(k) A(k,\rho)\,, \tag{23}$$

$$A(k,\rho) = B(k,1)\left[\rho^2 C(1-k,\rho) + C(k-1,\rho) - 1\right], \tag{24}$$

where $B(k,m) = \frac{\Gamma(1-d-k)\Gamma(d-k+m)}{\Gamma(d-k)\Gamma(1-d-k+m)}$, and $C(s,\rho) = F(d+s, 1; 1-d+s; \rho)$.

Both asymptotic expressions require the evaluation of the hypergeometric function $F(\cdot)$, which could be computational costly when the dimension of the variance-covariance matrix is large. The speed of computation, however, can be improved significantly with the approximation method proposed by Chung (1994). Specifically, we rewrite $A(k,\rho)$ in (24) as

$$A(k,\rho) = \rho\left[C(k,\rho) + C(-k,\rho) - 1\right], \tag{25}$$

[7] See, for example, Gradshteyn and Ryzhik (2014).

and compute $C(k,\rho)$ recursively using the formula

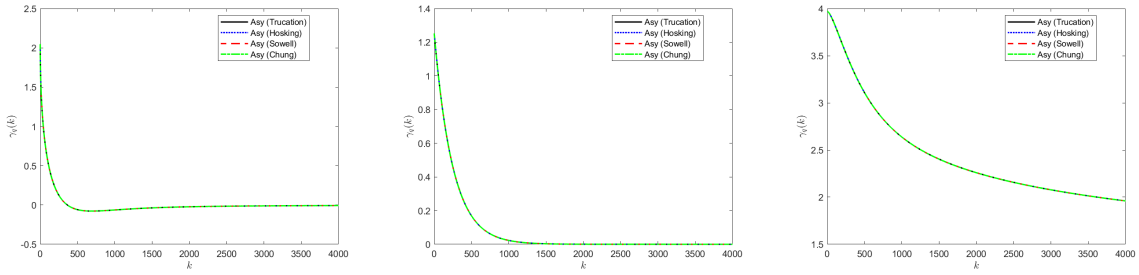
$$C(k,\rho) = C(T,\rho)\frac{\rho^{T-k}}{B(T,T-k)} + \sum_{i=1}^{T-k-1}\rho^{i}B(-k,i)+1, \tag{26}$$

$$C(-k,\rho) = C(0,\rho)\frac{\rho^{k}}{B(0,k)} + \sum_{i=1}^{k-1}\rho^{i}B(k,i)+1. \tag{27}$$

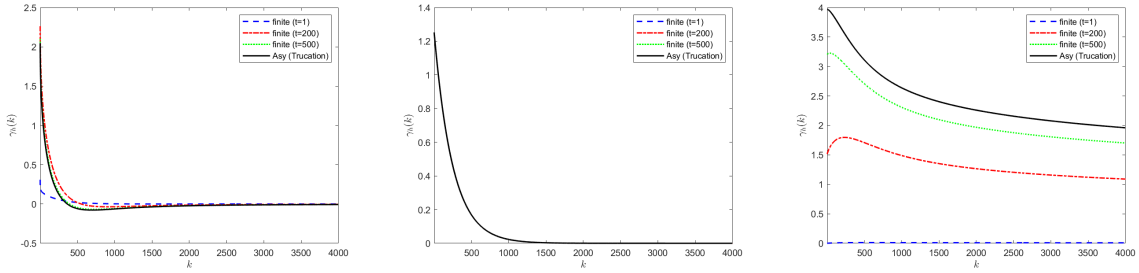
With the proposed approximation of (25)-(27), one would only need to compute the hypergeometric function twice (as opposed to T times), namely $C(T,\rho)$ and $C(0,\rho)$, and hence, is able to reduce the computational cost substantially.

Figure 6: The asymptotic and finite-sample autocovariance $\gamma_h(k)$ of FISV: $\rho = 0.996$, $\sigma_u = 0.1\Delta^d$, and $\Delta = 1/256$.

(a) Asy: $d = -0.3$ (b) Asy: $d = 0$ (c) Asy: $d = 0.4$



(d) Finite v.s. Asy: $d = -0.3$ (e) Finite v.s. Asy: $d = 0$ (f) Finite v.s. Asy: $d = 0.4$

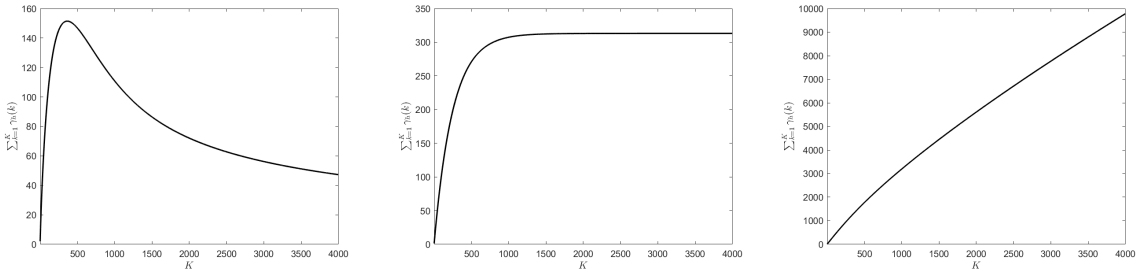


For comparison, we compute the autocovariance function $\gamma_h(k)$ using the finite sample autocovariance (20) and the four different approaches for the asymptotic autocovariance: equation (21) with a truncation of $10,000$ for i and j (labeled as 'Asy (truncation)'), equation (22) (labeled as 'Asy (Hosking)'), equations (23)-(24) (labeled as 'Asy (Sowell)'), and equations (23) and (25)-(27) (labeled as 'Asy (Chung)'). The fractional parameter takes three values $d = \{-0.3, 0, 0.4\}$. The remaining parameters are set according to our empirical results, namely $\Delta = 1/256$, $\rho = 0.996$, and $\sigma_u = 0.1\Delta^d$ taking values $\{0.528, 0.1, 0.011\}$ when $d = \{-0.3, 0, 0.4\}$. From the first row of Figure 6, there is no visual difference in the computed asymptotic covariance functions under this parameter setting. Similar to the FSV model, the convergence speed of the finite sample covariance is slow when $d = 0.4$ and faster when $d = -0.3$. See the second row of Figure 6. Figure 7 displays the cumulative sums of autocovariances

$\sum_{k=1}^{K}\gamma_h(k)$ of $h_{t\Delta}$ under the FISV specification. We can see that $\sum_{k=1}^{K}\gamma_h(k)$ diverges to infinity when $d>0$ and converges when $d\leq 0$.

Figure 7: The asymptotic cumulative sums of autocovariances $\sum_{k=1}^{K}\gamma_h(k)$ of FISV: $\rho=0.996$, $\sigma_u=0.1\Delta^d$, and $\Delta=1/256$.

(a) $\sum_{k=1}^{K}\gamma_h(k)$: $d=-0.3$ (b) $\sum_{k=1}^{K}\gamma_h(k)$: $d=0$ (c) $\sum_{k=1}^{K}\gamma_h(k)$: $d=0.4$



Finally, it is important to note that the hypergeometric function could be of extremely large magnitude when k is large and ρ is far from unity. Suppose $d=0.4$ and $k=2000$. While $F(1,d+k;1-d+k;\rho)=99.06$ when $\rho=0.99$, it takes value of 5.21×10^{34} when $\rho=0.93$. The large magnitude of $F(\cdot)$ leads to extremely large values of the autocovariance functions of Hosking (1981), Sowell (1992), and Chung (1994), as shown in Figure 8 where we set $\rho=0.93$ and $d=0.4$. The remain parameters are the same as those in Figure 6. Similar pattern is observed for the case of $d=-0.3$. Consequently, we recommend the use of 'asy (truncation)' for the asymptotic autocovariance function of the FISV model.

5 Memory Signature Plot

To help us evaluate the possible range of H or d, we introduce a memory signature plot for volatilities. Let $x_{t\Delta}=\log\left(r_{t\Delta}-\bar{r}\right)^2$ with $\bar{r}$ being the sample mean of $r_{t\Delta}$. Under the model specification of (6)-(7) and (16)-(17), we have

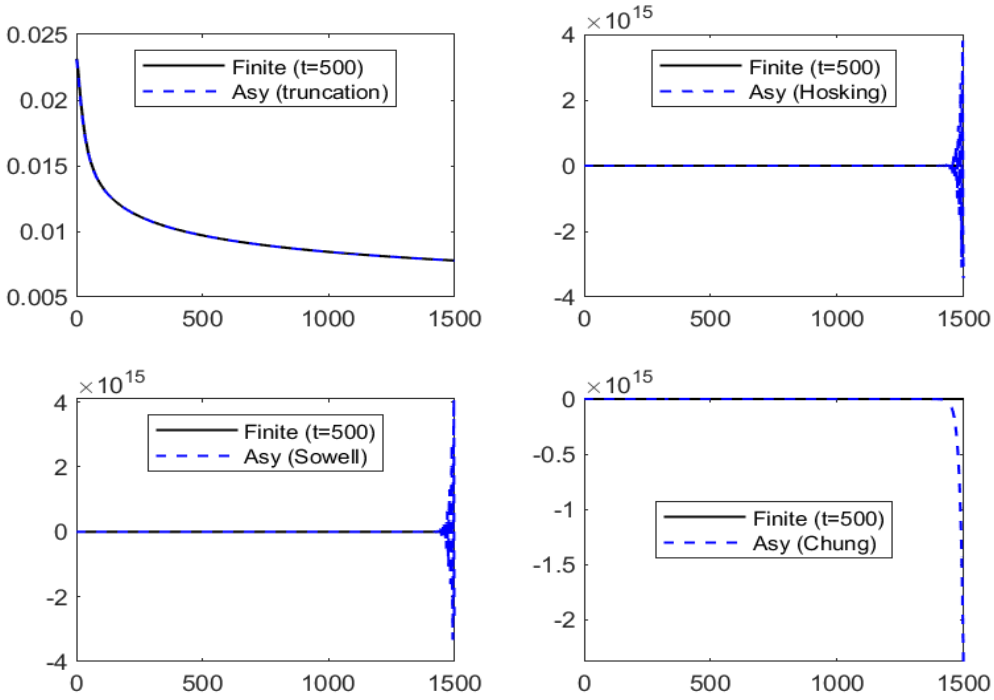
$$x_{t\Delta}=\log\left(\sigma^{*2}\Delta\right)+h_{t\Delta}+\log\varepsilon_t^2=\mu+h_{t\Delta}+\omega_t, \tag{28}$$

where $\mu=\log\left(\sigma^{*2}\Delta\right)+\mathbb{E}\left[\log\varepsilon_t^2\right]$ 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 $V\left(\log\varepsilon_t^2\right)=\pi^2/2\approx4.9$. Therefore, $x_{t\Delta}$ is the $h_{t\Delta}$ process plus an i.i.d non-Gaussian noise, with $\mathbb{E}\left(x_{t\Delta}\right)=\mu$ and

$$Var\left(x_{t\Delta}\right)=Var\left(h_{t\Delta}\right)+\pi^2/2\quad\text{and}\quad Cov\left(x_{t\Delta},x_{s\Delta}\right)=Cov\left(h_{t\Delta},h_{s\Delta}\right). \tag{29}$$

The autocovariances of $h_{t\Delta}$ are identical to those of the log squared (demeaned) returns $x_{t\Delta}$. Despite that the conditional variance $h_{t\Delta}$ is unobservable, $x_{t\Delta}$ can be easily obtained from returns. Consequently, we can compute the sample autocovariance of $h_{t\Delta}$ using $x_{t\Delta}$, which is denoted by $\hat{\gamma}(k)$ with k being the lag order. The memory signature plot shows the dynamic of $\sum_{k=0}^{K}\hat{\gamma}(k)$ as K increases, with

Figure 8: The asymptotic covariance functions of $h_{t\Delta}$: $\rho = 0.93$, $d = 0.4$, $\sigma_u = 0.1\Delta^d$, and $\Delta = 1/256$.



$\sum_{k=0}^{K} \hat{\gamma}(k)$ on the y-axis and K on the x-axis. The limit of $\sum_{k=0}^{K} \hat{\gamma}(k)$ is equivalent to the one-sided long-run variance $\sum_{k=0}^{\infty} \gamma_h(k)$ (when $T \to \infty$ followed by $K \to \infty$).

Importantly, from our previous discussions and as shown in Figure 4 and 7, the one-side long-run variance diverges to infinity when the process h_t has a long memory (i.e., $H > 0.5$ in FSV and $d > 0$ in FISV) and is bounded when $H \leq 0.5$ in FSV or $d \leq 0$ in FISV. The memory signature plot could, therefore, provide a simple assessment for the range of the memory parameters (i.e., H and d). While it is very common to plot $\hat{\gamma}(k)$ against k in the volatility literature (e.g., Ding et al. (1993); Bollerslev and Mikkelsen (1996); Breidt et al. (1998)), we argue that it is more informative to plot $\sum_{k=0}^{K} \hat{\gamma}(k)$ against K. This is because $\sum_{k=0}^{K} \hat{\gamma}(k)$ is directly related to $\sum_{k=0}^{\infty} \gamma(k)$, which has distinctive features under various specifications of the memory parameter.

6 Simulated Maximum Likelihood

Let $r = (r_{1\Delta}, r_{2\Delta}, \cdots, r_{T\Delta})'$ and $h = (h_{1\Delta}, h_{2\Delta}, \cdots, h_{T\Delta})'$. Model parameters are collected in θ, that is, $\theta = (\sigma, \beta, \sigma_\eta, H)$ for FSV and $\theta = (\sigma, \rho, \sigma_u, d)$ for FISV. Under the model specification of (6)-(7) or (16)-(17), the joint probability density function (pdf) of returns is

$$f(r|\theta) = \int f(r; h|\theta)\, dh = \int f(r|h, \theta)\, f(h|\theta)\, dh, \tag{30}$$

where the conditional density $f(r|h,\theta) = \Pi_{t=1}^{T}\phi\left(r_{t\Delta};0,\sigma^2 e^{h_{t\Delta}}\right)$ 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

$$f(h|\theta) = (2\pi)^{-T/2}\left|\Xi_\theta\right|^{-1/2}\exp\left(-\frac{1}{2}h'\Xi_\theta h\right), \tag{31}$$

with Ξ_θ being a $T\times T$ matrix whose $(t,s)^{th}$ element is given by $Cov\left(h_{t\Delta},h_{s\Delta}\right)$ for $t,s=1,2,\cdots,T$.

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

$$\frac{1}{S}\sum_{s=1}^{S} f\left(r|h^{(s)},\theta\right). \tag{32}$$

The distribution of h and the pdf $f(h|\theta)$ in (31) 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 (32) 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 $H=0.5$ or $d=0$, 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. This method has been successfully applied to the basic SV model (i.e., $H=1/2$ in the FSV model and $d=0$ in the FISV model) by Sandmann and Koopman (1998) and Yu (2011).

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,

$$f(r|\theta) = \int \frac{f(r;h|\theta)}{g(h)}dG(h) \simeq \frac{1}{S}\sum_{s=1}^{S}\frac{f\left(r;h^{(s)}|\theta\right)}{g\left(h^{(s)}\right)}. \tag{33}$$

Importantly, unlike the 'brute force' technique (32), $\left\{h^{(s)}\right\}_{s=1}^{S}$ 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

$$h_\theta^*(r) = \arg\max_h \log\left(f(r;h|\theta)\right), \tag{34}$$

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

$$\begin{aligned}\log f\left(r;h|\theta\right) &= \sum_{t=1}^{T}\log\left(\phi\left(r_t;0,\sigma^2e^{h_t}\right)\right)+\log\left(f\left(h|\theta\right)\right)\\ &= -T\log\left(2\pi\right)-\frac{T}{2}\log\left(\sigma^2\right)-\frac{1}{2}\sum_{t=1}^{T}h_{t\Delta}-\frac{1}{2}\sum_{t=1}^{T}z_{t\Delta}-\frac{1}{2}\log\left|\Xi_\theta\right|-\frac{1}{2}h'\Xi_\theta^{-1}h,\end{aligned}$$

with $z_{t\Delta}=r_{t\Delta}^2/\left(\sigma^2e^{h_{t\Delta}}\right)$. The variance-covariance matrix $\Sigma_\theta^*\left(r\right)$ is calculated as

$$\Sigma_\theta^*\left(r\right)=\left[-\frac{\partial^2\log\left(f\left(r;h|\theta\right)\right)}{\partial h\partial h'}\right]^{-1}\Bigg|_{h=h_\theta^*(r)}. \tag{35}$$

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

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

$$\hat{\theta}=\arg\max_{\theta\in\Theta}\log\left[\frac{1}{S}\sum_{s=1}^{S}\exp\left(\log\frac{f\left(r;h^{(s)}|\theta\right)}{g\left(h^{(s)}\right)}\right)\right], \tag{36}$$

where Θ is the parameter space and

$$\begin{aligned}\log\frac{f\left(r;h^{(s)}|\theta\right)}{g\left(h^{(s)}\right)} &= -\frac{T}{2}\log\left(2\pi\right)-T\log\left(\sigma^2\right)-\frac{1}{2}\sum_{t=1}^{T}h_{t\Delta}^{(s)}-\frac{1}{2}\sum_{t=1}^{T}z_{t\Delta}-\frac{1}{2}\log\left|\Xi_\theta\right|-\frac{1}{2}h^{(s)\prime}\Xi_\theta^{-1}h^{(s)}\\ &\quad+\frac{1}{2}\log\left|\Sigma_\theta^*\left(r\right)\right|+\frac{1}{2}\left(h^{(s)}-h_\theta^*\left(r\right)\right)'\Sigma_\theta^*\left(r\right)^{-1}\left(h^{(s)}-h_\theta^*\left(r\right)\right).\end{aligned} \tag{37}$$

The estimated volatility sequence $\hat{h}$ is taken as $h_{\hat{\theta}}^*\left(r\right)$, which is the modal configuration of $\log f\left(r,h|\hat{\theta}\right)$ using equation (34) with θ replaced by $\hat{\theta}$. Clearly, $\hat{h}_{t\Delta}$, the t^{th} element in $\hat{h}$, uses all the information in $\{r_{t\Delta}\}_{t=1}^{T}$ and hence, is a smoothed estimate of $h_{t\Delta}$. Correspondingly, the quantity $\exp\left(h_{t\Delta}/2\right)$, which follows a lognormal distribution, can be estimated as

$$\exp\left(\frac{h_{\hat{\theta}}^{*(t)}\left(r\right)}{2}+\frac{\Sigma_{\hat{\theta}}^{*(t,t)}(r)}{8}\right), \tag{38}$$

where $h_{\hat{\theta}}^{*(t)}\left(r\right)$ is the t^{th} element in $h_{\hat{\theta}}^*\left(r\right)$, $\Sigma_{\hat{\theta}}^{*(t,t)}(r)$ is the t^{th} diagonal element in $\Sigma_{\hat{\theta}}^*(r)$ (computed from (35) with $h_\theta^*\left(r\right)$ replaced by $h_{\hat{\theta}}^*\left(r\right)$). To obtain the filtered estimates of $h_{t\Delta}$ and $\exp\left(h_{t\Delta}/2\right)$, one can apply the Laplace approximation analogously to $\log f\left(r_\Delta,...,r_{t\Delta},h_\Delta,...,h_{t\Delta}|\hat{\theta}\right)$ instead.

6.2 Other Implementation Details

The optimization problem (34) 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

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

where

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

with $z=[z_{1\Delta},z_{2\Delta},\cdots z_{T\Delta}]$.

The distributional approximation of $h|(\theta,r)$ or the optimization of (34) is conducted independently for every given θ. To ensure the smoothness of the likelihood function (33) with respect to θ,[8] 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 $N\left(h_{\theta}^{*}\left(r\right),\Sigma_{\theta}^{*}\left(r\right)\right)$.

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

$$\begin{aligned}\log\left[\frac{1}{S}\sum_{s=1}^{S}\exp\left(\log\frac{f\left(r;h^{(s)}|\theta\right)}{g\left(h^{(s)}\right)}\right)\right] &= \log\left[\frac{1}{S}\sum_{s=1}^{S}f\left(r|h^{(s)},\theta\right)w^{(s)}\right]\\ &= \log\left(\frac{1}{S}\sum_{s=1}^{S}w^{(s)}\right)+\log\left[\sum_{s=1}^{S}f\left(r|h^{(s)},\theta\right)w^{*(s)}\right]\\ &= \log\left(\frac{1}{S}\sum_{s=1}^{S}\exp\left(A^{(s)}\right)\right)+\log\left[\sum_{s=1}^{S}\exp\left(B^{(s)}\right)w^{*(s)}\right],\end{aligned}\tag{39}$$

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

$$\begin{aligned}w^{*(s)} &= \frac{\exp\left(A^{(s)}+C_{1}\right)}{\sum_{s=1}^{S}\exp\left(A^{(s)}+C_{1}\right)},\\ \log\left(\frac{1}{S}\sum_{s=1}^{S}\exp\left(A^{(s)}\right)\right) &= -C_{1}+\log\left(\frac{1}{S}\sum_{s=1}^{S}\exp\left(A^{(s)}+C_{1}\right)\right),\\ \log\left[\sum_{s=1}^{S}\exp\left(B^{(s)}\right)w^{*(s)}\right] &= -C_{2}+\log\left[\sum_{s=1}^{S}\exp\left(B^{(s)}+C_{2}\right)w^{*(s)}\right],\end{aligned}$$

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

$$\log\left[\frac{1}{S}\sum_{s=1}^{S}\exp\left(\log\frac{f\left(r;h^{(s)}|\theta\right)}{g\left(h^{(s)}\right)}\right)\right]$$

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

$$= -C_1 - C_2 + \log\left(\frac{1}{S}\sum_{s=1}^{S}\exp\left(A^{(s)}+C_1\right)\right) + \log\left[\sum_{s=1}^{S}\exp\left(B^{(s)}+C_2\right)\frac{\exp\left(A^{(s)}+C_1\right)}{\sum_{s=1}^{S}\exp\left(A^{(s)}+C_1\right)}\right].$$

The evaluation of the likelihood function involves the variance-covariance matrix Ξ_θ, which could be of very high dimension when the sample size is large. Indeed, it will be computational costly to calculate the matrix element-by-element. The speed of computation, however, can be improved substantially with matrix-based languages. Table 1 lists the time required for computing the variance-covariance matrix in MATLAB with a standard available laptop (CPU@ 1.90GHZ) for the FSV and FISV models, which are quite fast even when $T = 7,680$.

Table 1: Time required for computing Ξ_θ (in seconds)

Computation time	Ξ_θ : FSV		Ξ_θ : FISV	
	Finite	Asy.	Finite	Asy (trunc.)
$T = 2,560$	0.81	0.16	0.80	5.94
$N = 7,680$	7.21	1.25	7.38	6.86

No asymptotic distribution has been developed for the SML estimator for general hidden non-Markov models in the literature.[9] It is a common practice to assume that the standard ML asymptotic theory remains valid for the SML estimator (e.g., Durbin and Koopman (1997)). Heuristically, let θ_0 be the true value of θ and $\widehat{\theta}$ be an interior point of the compact set Θ. Applying the first-order Taylor expansion to $\frac{1}{T}\frac{\partial \log f\left(r|\widehat{\theta}\right)}{\partial\theta}$ at θ_0 and using (36), we have, as $T \to \infty$,

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right) = \left[-\frac{\partial^2 \log f\left(r|\widetilde{\theta}\right)}{\partial\theta\partial\theta'}\right]^{-1}\sqrt{T}\frac{\partial \log f\left(r|\theta_0\right)}{\partial\theta} \xrightarrow{d} N\left(0, H^{-1}JH^{-1}\right),$$

where $H = \text{plim}\left[-\frac{\partial^2 \log f(r|\theta_0)}{\partial\theta\partial\theta'}\right]$, $J = \lim \mathbb{E}\left[\frac{\partial \log f(r|\theta_0)}{\partial\theta}\frac{\partial \log f(r|\theta_0)}{\partial\theta'}\right]$, and $\widetilde{\theta}$ is a value on the line joining $\widehat{\theta}$ and θ_0. Therefore, standard asymptotic theory for the Wald test and the likelihood ratio test are applicable to the basic SV, FSV, and FISV models.

7 Simulation Studies

In this section, we investigate the estimation accuracy of the SML method for the FSV and FISV models. We report median deviations and median absolute deviations (MAD) of each parameter (i.e., $median(\hat{\theta}_i) - \theta_i$ and $median|\hat{\theta}_i - median(\hat{\theta}_i)|$), with the biases and root mean square errors (RMSE) provided in the appendix. It is well-known that the median is less sensitive than the mean to outliers, which are likely to exist due to the failure of numerical convergence of some cases in the simulation.

7.1 Fractional Stochastic Volatility Model

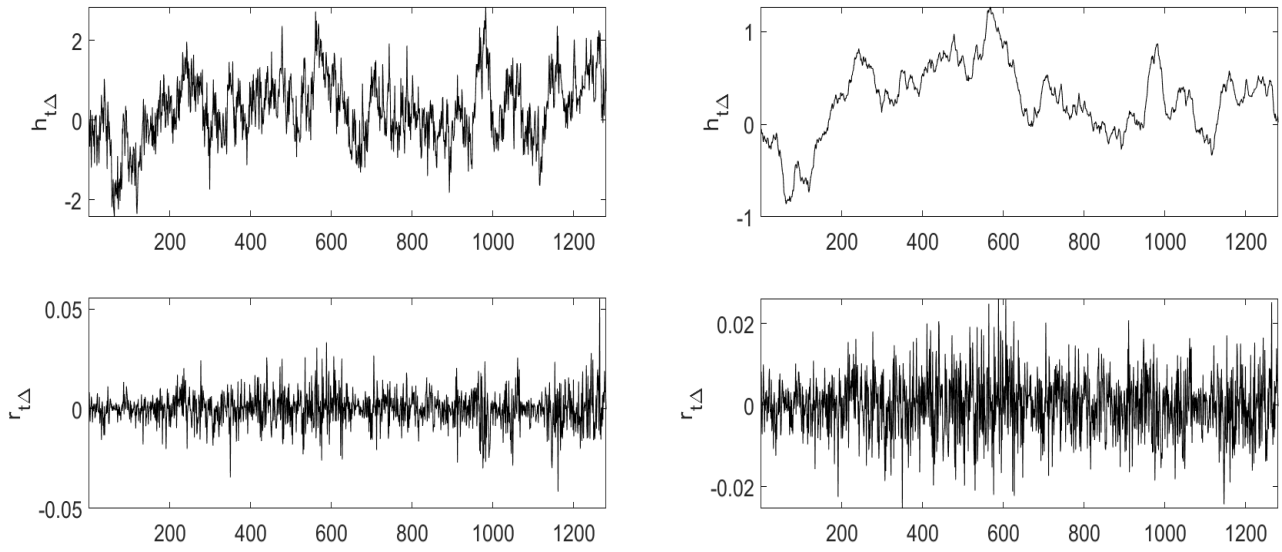
Log returns are generated from equation (6), while the latent variable $h_{t\Delta}$ is from (7), with the fractional Gaussian noise generated using fast Fourier transform (Kroese and Botev, 2015). Model parameters

[9] Douc et al. (2011) provide the consistency of the maximum likelihood estimator for general hidden Markov models, which include nonlinear and non-Gaussian state-space models as special cases.

are set according to our empirical estimation results in Section 8 as follows

$$\beta = 0.9961, \sigma^* = 0.1, \sigma_\eta^* = 1.5, \Delta = 1/256.$$

We examine the performance of the estimation method using both the asymptotic covariance and its finite sample form. Different values of Hurst parameter H are considered. Specifically, we set $H = \{0.1, 0.2, 0.5, 0.7, 0.9\}$. The time span considered are 5 and 10 years, implying 1280 and 2560 observations of $r_{t\Delta}$. Figure 9 shows two typical realized data series from this data generating process with $N = 5$. We set the Hurst parameter H to 0.2 in panel (a) and 0.7 in panel (b).



(a) FSV: $H = 0.2$ (b) FSV: $H = 0.7$

Figure 9: Typical realized trajectories of FSV: $\beta = 0.996, \sigma^* = 0.1, \sigma_\eta^* = 1.5, \Delta = 1/256, N = 5$.

Table 2: Estimated model parameters of FSV: median and median absolute deviation. The true model parameters are $\beta = 0.9961, \sigma_\eta^* = 1.5, \sigma^* = 0.1$.

	$\hat{\beta}$			$\hat{\sigma}_\eta^*$			$\hat{H}$			$\hat{\sigma}^*$		
	Median	Bias	MAD	Median	Bias	MAD	Median	Bias	MAD	Median	Bias	MAD
						$T = 1280$						
Finite sample variance-covariance matrix												
$H = 0.1$	0.9924	-0.0037	0.0035	1.5579	0.0579	0.1384	0.1292	0.0292	0.0211	0.1011	0.0011	0.0092
$H = 0.2$	0.9927	-0.0034	0.0032	1.5969	0.0969	0.1942	0.2239	0.0239	0.0395	0.0981	-0.0019	0.0124
$H = 0.5$	0.9940	-0.0021	0.0033	1.6390	0.1390	0.4501	0.5782	0.0782	0.1502	0.0972	-0.0028	0.0126
$H = 0.7$	0.9960	-0.0001	0.0021	1.4427	-0.0573	0.4540	0.7231	0.0231	0.1223	0.0962	-0.0038	0.0098
$H = 0.9$	0.9969	0.0008	0.0014	1.2470	-0.2530	0.4166	0.8990	-0.0010	0.0480	0.0993	-0.0007	0.0098
Asymptotic variance-covariance matrix												
$H = 0.1$	0.9932	-0.0029	0.0039	1.5404	0.0404	0.1331	0.1276	0.0276	0.0180	0.1024	0.0024	0.0096
$H = 0.2$	0.9927	-0.0034	0.0036	1.5767	0.0767	0.1912	0.2190	0.0190	0.0372	0.1010	0.0010	0.0123
$H = 0.5$	0.9934	-0.0027	0.0029	1.5595	0.0595	0.3644	0.5264	0.0264	0.1248	0.1002	0.0002	0.0178
$H = 0.7$	0.9953	-0.0008	0.0017	1.3097	-0.1903	0.3453	0.6206	-0.0794	0.1033	0.1036	0.0036	0.0254
$H = 0.9$	0.9967	0.0006	0.0011	0.8472	-0.6528	0.2075	0.6694	-0.2306	0.0957	0.1096	0.0096	0.0347
						$T = 2560$						
Finite sample variance-covariance matrix												
$H = 0.1$	0.9934	-0.0027	0.0022	1.4698	-0.0302	0.0763	0.1250	0.0250	0.0157	0.1002	0.0002	0.0056
$H = 0.2$	0.9946	-0.0015	0.0014	1.5227	0.0227	0.1202	0.2175	0.0175	0.0314	0.0978	-0.0022	0.0074
$H = 0.5$	0.9953	-0.0008	0.0025	1.5136	0.0136	0.3312	0.4978	-0.0022	0.0981	0.0965	-0.0035	0.0152
$H = 0.7$	0.9965	0.0004	0.0019	1.3823	-0.1177	0.3752	0.6373	-0.0627	0.1278	0.0977	-0.0023	0.0122
$H = 0.9$	0.9966	0.0005	0.0009	1.3716	-0.1284	0.2272	0.9029	0.0029	0.0269	0.0988	-0.0012	0.0094
Asymptotic variance-covariance matrix												
$H = 0.1$	0.9938	-0.0023	0.0024	1.4646	-0.0354	0.0782	0.1251	0.0251	0.0157	0.0998	-0.0002	0.0063
$H = 0.2$	0.9946	-0.0015	0.0015	1.5112	0.0112	0.1345	0.2152	0.0152	0.0297	0.0977	-0.0023	0.0078
$H = 0.5$	0.9949	-0.0011	0.0021	1.4617	-0.0383	0.2851	0.4835	-0.0165	0.0825	0.0956	-0.0044	0.0173
$H = 0.7$	0.9959	-0.0002	0.0013	1.2553	-0.2447	0.2435	0.6026	-0.0974	0.0758	0.0981	-0.0019	0.0288
$H = 0.9$	0.9972	0.0011	0.0009	0.8304	-0.6696	0.1601	0.6894	-0.2106	0.1064	0.1056	0.0056	0.0423

Table 2 reports the median and MAD of the parameter estimates when using both asymptotic covariance and its finite sample form. The number of replications for every parameter constellation is 100. The number of samples in the importance sampling is $1,000$. When H is far from unity, no matter which form of the covariance is used, the estimation method performs well even when there are only $T = 1280$ observations. The medians of all estimates are closed to the true values of parameters. The MADs of σ^* and β are stable regardless of the values of H; and the variation becomes smaller as the sample size increases.

However, when H is close to unity, the SML method performs better by using the finite sample form of the covariance. In particular, for $H = 0.9$, the median of estimates is very close to the true value if we use the finite-sample form even when the sample size is only $T = 1,280$. However, if we adopt the asymptotic form, the bias is very large. Specifically, when $T = 2,560$, the bias is -0.2106, which is 100 times larger than that of the finite sample counterpart. The reason is the divergence between the finite-sample and asymptotic covariances, illustrated in Section 3.2. In summary, when H is close to unity, the finite-sample form is more accurate than the asymptotic form.

Table B.7 reports the mean and the RMSE of estimates from all replications. We can find similar patterns of simulation results as those in Table 2. Further, it is not surprising that the bias and RMSE are larger than the median and MAD, as the median and MAD are less sensitive to outliers.

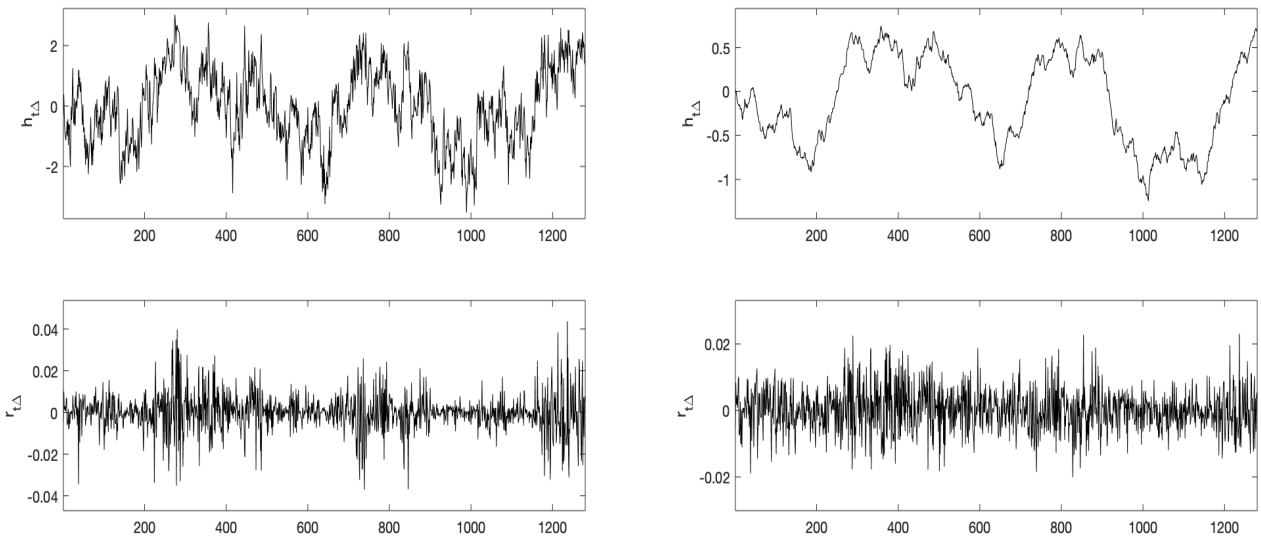
7.2 Fractional Integrated Stochastic Volatility Model

For the FISV model, the log returns are generated according to equation (16), while the latent variable $h_{t\Delta}$ is from (17), with $u_{t\Delta}$ generated using equation (1). We set the model parameters according to our empirical estimation results in Section 8, that is,

$$\rho = 0.9961, \sigma^* = 0.1, \Delta = 1/256, \sigma_u = 0.1\Delta^d.$$

The fractional parameter d is set corresponding to the H values in the previous section, that is, $d = \{-0.4, -0.3, 0, 0.2, 0.4\}$. The time span N is either 5 or 10 years (i.e., $T = 1,280$ or $2,560$). Figure 10 shows two typical realized data series from this data generating process with $N = 5$.

We estimate the model parameters using the SML method with both the asymptotic and finite sample variance-covariance matrixes. The number of replications and the number of samples in the importance sampling are the same as those for the FSV model in the previous section. Table 3 reports the median and MAD of the estimates, while Table B.8 shows the mean and RMSE.



(a) $d = -0.3$ (b) $d = 0.2$

Figure 10: Two typical realized trajectories of the fractional integrated stochastic volatility model: $\rho = 0.9961, \sigma^* = 0.1, \Delta = 1/256, N = 5$.

Table 3: Estimated model parameters of FISV: median and median absolute deviation. The true model parameters are $\rho = 0.9961, \sigma = 0.1, \sigma_u = 0.1\Delta^d$.

	$\hat{\rho}$			$\hat{\sigma}_u$			$\hat{d}$			$\hat{\sigma}$		
	Median	Bias	MAD	Median	Bias	MAD	Median	Bias	MAD	Median	Bias	MAD
						$T = 1280$						
Finite sample variance-covariance matrix												
$d = -0.4$	0.9900	-0.0061	0.0041	0.7810	-0.1379	0.0387	-0.3276	0.0724	0.0314	0.1000	-0.0000	0.0131
$d = -0.3$	0.9919	-0.0042	0.0032	0.4897	-0.0381	0.0508	-0.2601	0.0399	0.0415	0.1005	0.0005	0.0155
$d = 0$	0.9946	-0.0015	0.0019	0.0901	-0.0099	0.0099	0.0311	0.0311	0.0311	0.1000	0.0000	0.0061
$d = 0.2$	0.9956	-0.0005	0.0023	0.0347	0.0017	0.0142	0.1821	-0.0179	0.1079	0.1006	0.0006	0.0068
$d = 0.4$	0.9969	0.0008	0.0018	0.0189	0.0080	0.0089	0.2685	-0.1315	0.0888	0.1004	0.0004	0.0050
Asymptotic variance-covariance matrix												
$d = -0.4$	0.9904	-0.0057	0.0037	0.7839	-0.1350	0.0403	-0.3300	0.0700	0.0349	0.0999	-0.0001	0.0123
$d = -0.3$	0.9919	-0.0042	0.0030	0.4958	-0.0320	0.0516	-0.2671	0.0329	0.0388	0.0994	-0.0006	0.0131
$d = 0$	0.9927	-0.0034	0.0031	0.0997	-0.0003	0.0344	0.0252	0.0252	0.1010	0.1007	0.0007	0.0178
$d = 0.2$	0.9945	-0.0016	0.0019	0.0452	0.0122	0.0165	0.1339	-0.0661	0.0828	0.0994	-0.0006	0.0224
$d = 0.4$	0.9959	-0.0002	0.0015	0.0275	0.0166	0.0104	0.1918	-0.2082	0.0838	0.1001	0.0001	0.0269
						$T = 2560$						
Finite sample variance-covariance matrix												
$d = -0.4$	0.9915	-0.0046	0.0025	0.7469	-0.1721	0.0242	-0.3133	0.0867	0.0211	0.1008	0.0008	0.0097
$d = -0.3$	0.9936	-0.0025	0.0021	0.4618	-0.0660	0.0295	-0.2492	0.0508	0.0340	0.0995	-0.0005	0.0102
$d = 0$	0.9946	-0.0015	0.0016	0.0938	-0.0062	0.0063	0.0311	0.0311	0.0155	0.1000	0.0000	0.0104
$d = 0.2$	0.9956	-0.0005	0.0022	0.0349	0.0020	0.0127	0.1712	-0.0288	0.1139	0.0988	-0.0012	0.0107
$d = 0.4$	0.9966	0.0005	0.0014	0.0167	0.0058	0.0055	0.3024	-0.0976	0.0872	0.1001	0.0001	0.0089
Asymptotic variance-covariance matrix												
$d = -0.4$	0.9917	-0.0044	0.0025	0.7476	-0.1713	0.0272	-0.3147	0.0853	0.0224	0.1013	0.0013	0.0098
$d = -0.3$	0.9936	-0.0025	0.0020	0.4654	-0.0624	0.0284	-0.2485	0.0515	0.0338	0.0993	-0.0007	0.0107
$d = 0$	0.9941	-0.0020	0.0022	0.1036	0.0036	0.0331	0.0019	0.0019	0.0944	0.0984	-0.0016	0.0164
$d = 0.2$	0.9952	-0.0009	0.0016	0.0415	0.0085	0.0141	0.1315	-0.0685	0.1024	0.0982	-0.0018	0.0229
$d = 0.4$	0.9965	0.0004	0.0013	0.0231	0.0123	0.0068	0.2284	-0.1716	0.0770	0.0955	-0.0045	0.0306

The SML method can provide reasonably accurate estimates for all parameters of the FISV model. For parameter d, one can reduce the bias of the estimator by replacing the asymptotic variance-covariance matrix with its finite sample counterpart when the long memory pattern is strong. For instance, when $d = 0.4$ and $T = 2,560$, the bias of $\hat{d}$ is -0.0976 with the finite sample variance-covariance matrixes, which is 0.074 closer to the true value than the case with the asymptotic matrix. The estimation accuracy of d can be improved further by increasing the sample size. When $d = 0.4$, and with the finite sample variance-covariance matrix, the bias of $\hat{d}$ reduces from -0.1315 to -0.0976 when T increases from $1,280$ to $2,560$. There are no obvious changes in MAD under various settings.

8 Empirical Studies

In this section, we consider two stock indices and two major currency pairs. The stock indices are the S&P 500 composite index and the NIKKEI 225 index, while the currency pairs are the Eurodollar to USD (EURO/US) and the Great British Pound to USD (GBP/USD). Data are downloaded from DataStream for the maximum available sample period at the daily frequency. The starting dates of each data series are listed in Table 4. The sample terminates on October 1, 2020. We remove observations on public holidays for the two equity indices.

Figure 9 displays log returns $r_{t\Delta}$ in the left column and $x_{t\Delta}$ as defined in (28) in the right column. Table 4 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 (29).

Table 4: Summary statistics

		r_t				x_t			
	Start Date	Mean	Std. dev	Skewness	Kurtosis	Mean	Std. dev	Skewness	Kurtosis
S&P 500	2-Jan-1964	0.000	0.010	-1.03	29.53	-11.06	2.48	-1.12	5.60
NIKKEI 225	3-Apr-1950	0.000	0.012	-0.39	12.69	-10.67	2.46	-1.15	5.86
EURO/USD	2-Jan-1975	0.000	0.006	-0.06	6.56	-12.29	3.18	-1.75	6.61
GBP/USD	2-Jan-1975	-0.000	0.006	-0.27	10.31	-12.21	2.79	-1.15	4.37

For the model estimation, we consider two subsamples. Sample I starts from the beginning of the sample period and lasts for 25 years, while Sample II is from 1996 to 2020. The exact sample periods are listed in Table 5. The first subsample allows us to connect with the empirical literature on long memory which mostly published in late 1990s or early 2000s (Andersen and Bollerslev, 1997; Breidt et al., 1998; Bollerslev and Wright, 2000; Andersen et al., 2003; Bollerslev et al., 2000). The second subsample period aligns with the empirical studies of recent publications or working papers on volatility roughness (Gatheral et al., 2018; Wang et al., 2019; Fukasawa et al., 2019; Bennedsen et al., 2017; Bolko et al., 2020).

Table 5: Estimation sample periods

	S&P 500	NIKKEI 225	EURO/USD	GBP/USD
Sample I	1964-1988	1950-1974	1975-1999	1975-1999
Sample II	1996-2020	1996-2020	1996-2020	1996-2020

8.1 Memory Signature Plots

The memory signature plot offers a simple way to visualize the convergence of the cumulative sums of autocovariances $\sum_{k=0}^{K}\hat{\gamma}(k)$. Recall that the long-run variance $\sum_{k=0}^{\infty}\gamma_h(k)$ diverges to infinity when $H > 0.5$ in the FSV model or $d > 0$ in the FISV model, whereas it is bounded when $H \leq 0.5$ or $d \leq 0$. Figure 11 displays the memory signature plots (with $K = 2,500$) for the four data series over each of the two subsamples. Contradict to the traditional view that $\sum_{k=0}^{\infty}\gamma_h(k) = \infty$, we find the convergence of the cumulative sums of autocovariances for all data series over both sample periods, suggesting $d \leq 0$ or $H \leq 0.5$.

For the memory signature plot, it is crucial to set K to a large value to be consistent with the notion of 'long-run' variance. Otherwise, one might draw a false conclusion when the convergence is slow. As a case in point, suppose that one sets the value of K to be 100 or 200, as typically done in the literature when plotting $\gamma_h(k)$ (Ding et al., 1993; Breidt et al., 1998). One can see from Figure 11 that for all sample periods considered, $\sum_{k=0}^{K}\hat{\gamma}(k)$ increases rapidly with K when $K \leq 200$. Without seeing the remaining part of the graph, one would probably conclude that the cumulative sums of autocovariances diverge and that the long-run variance is unbounded, leading to the conclusion of $H > 0.5$ in the FSV model or $d > 0$ in the FISV model. This probably can been seen better from Figure 2, where we plot $\sum_{k=0}^{K}\hat{\gamma}(k)$ for Sample I of the S&P 500 index with $K = 200$, $K = 1,000$, and $K = 2,500$ on separate graphs.

8.2 Estimation Results

Next, we fit the FSV and FISV models, along with the basic SV model, to each of the return series over the two subsample periods. The log-likelihood functions are computed from the finite sample variance-covariance matrices. For the numerical optimization of (36) for the FSV and FISV models, we consider a set of $1,000$ initial values and choose the one with the largest log-likelihood value as the initial input of the optimization.[10]

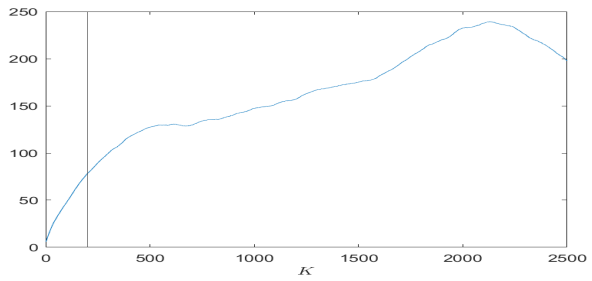
Estimation results are reported in Table 6. The top panel is for Sample I and the bottom panel is for Sample II. First, the log-likelihood values (the second last column) of the FSV and FISV models are always larger than those of the basic SV model.[11] The log-likelihood ratio statistics (last column)

[10] The $1,000$ grid points considered cover a wide range of values of the model parameters. For the FSV model, the autoregressive coefficient is from 0.925 to 0.995 with an increment of 0.01, i.e., $\beta_0 = \{0.925 : 0.01 : 0.995\}$, the Hurst parameter $H_0 = \{0.05 : 0.1, 0.95\}$, and $\sigma^*_{\eta 0} = \{0.5 : 1 : 10\}$. The grid settings for the FISV model are the same, namely $\rho_0 = \beta_0$, $d_0 = H_0 - 0.5$, and $\sigma_{u0} = \sigma^*_{\eta 0}\sqrt{\Delta}$. For both models, we set σ to be the estimate of the standard SV model.

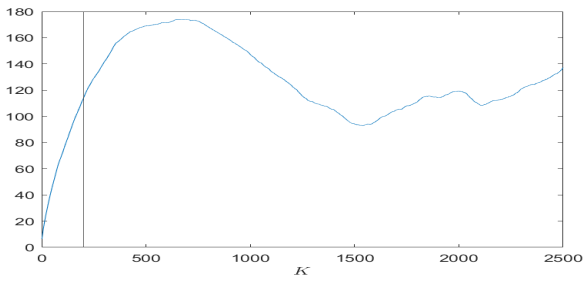
[11] The estimated coefficients in the basic SV model for the S&P 500 returns are almost identical to those provided in Sandmann and Koopman (1998) for the sample period running from 1928 to 1987.

Figure 11: The memory signature plots: the y-axis is the cumulative sums $\sum_{k=0}^{K} \hat{\gamma}(k)$ and the x-axis shows the value of K. The vertical line indicates the locations of $K = 200$.

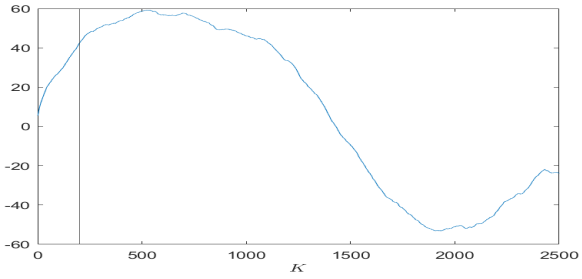
(a) S&P 500: Sample I



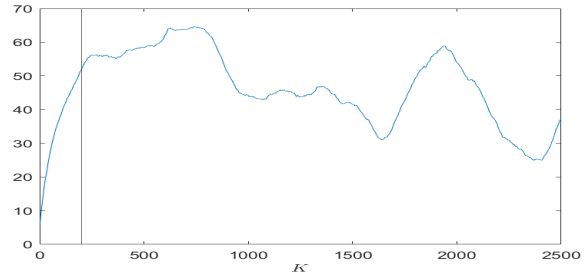
(b) S&P 500: Sample II



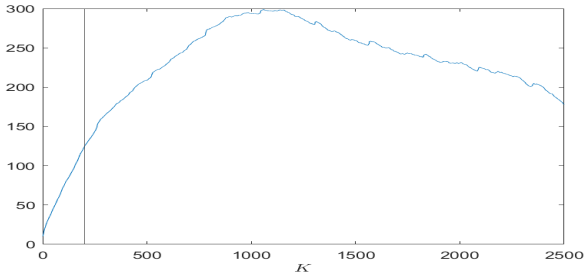
(c) NIKKEI 225: Sample I



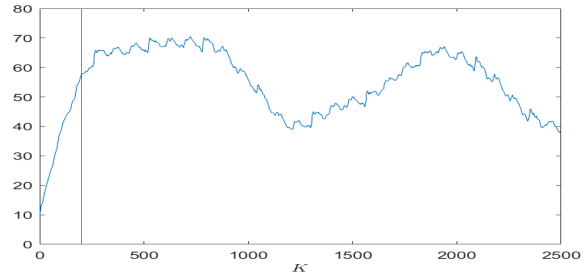
(d) NIKKEI 225: Sample II



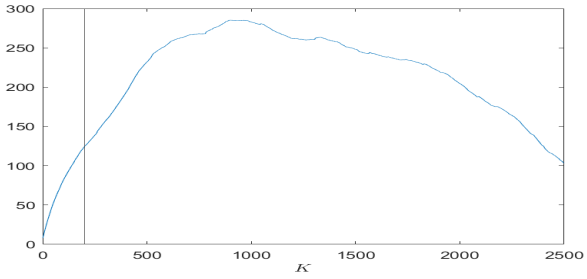
(e) EURO/USD: Sample I



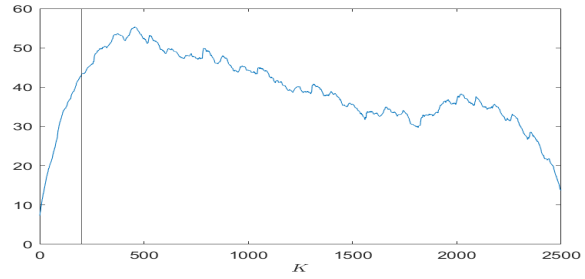
(f) EURO/USD: Sample II



(g) GBP/USD: Sample I



(h) GBP/USD: Sample II



suggest that by freeing up only the memory parameter (H or d), the FSV and FISV models improve the likelihood value, in the log scale, by a value between 8 and 176. The 1% critical value for chi-square distribution is 6.63, overwhelmingly rejecting the basic SV model in favor of the FSV or FISV model

for all data series. This means that the FSV and FISV models are far more suitable for the data series.

Second, the estimated autoregressive coefficients from both the FSV and FISV models are between 0.990 and unity, suggesting that volatilities are highly persistent.[12] Third, the estimated Hurst parameter H of the FSV model ranges between 0.12 and 0.22 for the first sample period and between 0.08 and 0.25 for the second sample period, which is smaller than 0.5 assumed in the basic SV model. Correspondingly, the estimated memory parameter of the FISV model d falls between -0.45 and -0.36 for Sample I and -0.49 and -0.27 for Sample II. Both models suggest that the volatility process is rough, confirming our observation from the memory signature plots. The estimation results imply that the long-range dependence of volatility is due to the persistence of the data series (with an autoregressive coefficient close to unity) rather than long memory with $H > 0.5$ or $d > 0$.

Our finding on the memory parameter is consistent with the recent rough volatility literature but contradicts the earlier literature on long memory. For example, Breidt et al. (1998) estimate the FISV model with a spectral likelihood estimator for the S&P 500 CRSP data over the sample period from July 1962 to December 1987. They find that the estimated d parameter is approximately 0.44, and the volatility process is less persistent with the autoregressive parameter ρ around 0.93. The long-range dependence of the volatility is driven by the fractional parameter d being greater than zero and close to 0.5. Although the data and the sample period considered in their paper is different from our S&P 500 data series, we find a similar pattern of the memory signature plot with the data provided by the authors to that of the S&P 500 series (Figure 11(a)). In other words, although their estimation results suggest $d > 0$, the memory signature plot of their data series shows otherwise.

Fourth, the log-likelihood values of FSV and FISV models are very close to each other. For Sample I, the log-likelihood value of FSV is identical to that of FISV for the S&P 500 index. It is slightly larger (smaller) than that of the FISV model for EURO/USD and GBP/USD (the Nikkei 225 index). For Sample II, the likelihood value of the FSV model is higher than that of the FISV model for all data series. Lastly, when using the asymptotic variance-covariance matrices for the likelihood functions, we obtain almost identical results (up to second or third decimal points) for all three models and all data series. This result echoes our findings in the simulations. When $H \leq 0.5$ or $d \leq 0$, there is no apparent difference between the finite sample and the asymptotic variance-covariance matrices for the SML estimation accuracy of both FSV and FISV.

The smoothed estimates of volatilities (38) from the FSV and FISV models are displayed in Figure B.13.[13] The estimated volatility trajectories are similar to those of $h_{t\Delta}$ simulated from our DGPs with $H < 0.5$ or $d < 0$ (Figure 9(a) and Figure 10(a)). The volatility of Sample II appears to be more volatile than that of Sample I for the S&P 500 index, whereas we observe the opposite for the other three data series. This observation aligns with our estimation results for σ_η and σ_u. The volatilities of the two equity indices increased to unprecedented levels during the 2008 subprime mortgage crisis period. Like

[12] Our empirical estimates of β are very close to those obtained in Table 3 of Bolko et al. (2020) based on GMM. Note that our β corresponds to $1 - \lambda$ in Bolko et al. (2020).

[13] For brevity, we do not display the estimated volatilities from the basic SV model. They appear to be much smoother than those from the FSV and FISV models.

Table 6: Estimation results for the basic SV, FSV and FISV models. Numbers in the square brackets are the 10% confidence intervals. The last column is the log-likelihood ratio test for H_0: Basic SV.

Sample I: start date + 25 years						
Basic Stochastic Volatility Model						
	β	σ_η^*	H	σ	lld.	
S&P 500	0.990 [0.986, 0.994]	1.93 [1.76, 2.10]	-	0.007 [0.006, 0.008]	21,684	
NIKKEI 225	0.944 [0.932, 0.956]	4.48 [4.04, 4.93]	-	0.0084 [0.0079, 0.0088]	20,330	
EURO-USD	0.974 [0.968, 0.980]	3.59 [3.18, 3.99]	-	0.005 [0.004, 0.005]	25,045	
GBP-USD	0.947 [0.941, 0.953]	4.88 [4.39, 5.37]	-	0.0047 [0.0044, 0.005]	24,980	
Fractional Stochastic Volatility Model						
	β	σ_η^*	H	σ	lld.	LR (H_1: FSV)
S&P 500	0.998 [0.997, 0.999]	1.05 [0.95, 1.15]	0.18 [0.13, 0.22]	0.007 [0.006, 0.008]	21,690	12***
NIKKEI 225	0.990 [0.984, 0.994]	1.68 [1.41, 1.96]	0.22 [0.17, 0.27]	0.0081 [0.0078, 0.0085]	20,334	8***
EURO-USD	0.999 [0.997, 1.000]	1.33 [1.21, 1.45]	0.13 [0.11, 0.16]	0.005 [0.004, 0.006]	25,087	84***
GBP-USD	0.9999 [0.9996, 1.0002]	1.51 [1.42, 1.60]	0.12 [0.10, 0.14]	0.0024 [0.0018,0.003]	25,068	176***
Fractional Integrated Stochastic Volatility Model						
	ρ	σ_u	d	σ	lld	LR (H_1: FISV)
S&P 500	0.998 [0.997, 1.000]	0.31 [0.25, 0.37]	-0.36 [-0.43, -0.29]	0.007 [0.006, 0.008]	21,690	12***
NIKKEI 225	0.994 [0.990, 0.997]	0.46 [0.40, 0.52]	-0.38 [-0.45, -0.32]	0.0080 [0.0077, 0.0084]	20,339	18***
EURO-USD	0.999 [0.998, 1.000]	0.50 [0.45, 0.55]	-0.44 [-0.48, -0.40]	0.0052 [0.0046 0.0059]	25,084	78***
GBP-USD	0.9999 [0.9997, 1.0002]	0.59 [0.55, 0.64]	-0.45 [-0.49, -0.42]	0.0024 [0.0018, 0.0031]	25,063	166***
Sample II: 1996 to 2020						
Basic Stochastic Volatility Model						
	β	σ_η^*	H	σ	lld.	
S&P 500	0.980 [0.975, 0.985]	3.19 [2.86, 3.52]	-	0.009 [0.008, 0.010]	20,114	
NIKKEI 225	0.974 [0.968, 0.980]	2.85 [2.62, 3.08]	-	0.0124 [0.0115, 0.0133]	17,736	
EURO-USD	0.994 [0.987, 1.000]	1.15 [0.61, 1.69]	-	0.0052 [0.0047, 0.0058]	24,568	
GBP-USD	0.983 [0.976, 0.990]	1.90 [1.55, 2.25]	-	0.0050 [0.0047, 0.0053]	24,850	
Fractional Stochastic Volatility Model						
	β	σ_η^*	H	σ	lld.	LR (H_1: FSV)
S&P 500	0.996 [0.993, 0.999]	1.63 [1.39, 1.86]	0.25 [0.19, 0.31]	0.009 [0.008, 0.010]	20,121	14***
NIKKEI 225	0.994 [0.989, 0.998]	1.40 [1.24, 1.56]	0.24 [0.18, 0.31]	0.0123 [0.0116, 0.0131]	17,742	12***
EURO-USD	0.999 [0.998, 1.000]	0.90 [0.83, 0.97]	0.08 [0.06, 0.10]	0.005 [0.004, 0.006]	24,597	58***
GBP-USD	0.998 [0.996, 0.999]	0.98 [0.90, 1.05]	0.09 [0.06, 0.11]	0.0049 [0.0047,0.0052]	24,889	78***
Fractional Integrated Stochastic Volatility Model						
	ρ	σ_u	d	σ	lld	LR (H_1: FISV)
S&P 500	0.996 [0.993, 0.999]	0.34 [0.23, 0.45]	-0.27 [-0.36, -0.17]	0.009 [0.008, 0.010]	20,120	12***
NIKKEI 225	0.994 [0.990, 0.999]	0.31 [0.24, 0.38]	-0.28 [-0.39, -0.18]	0.0124 [0.0116, 0.0131]	17,741	10***
EURO-USD	0.999 [0.998, 1.000]	0.35 [0.29, 0.41]	-0.46 [-0.52, -0.41]	0.005 [0.004, 0.006]	24,584	32***
GBP-USD	0.998 [0.997, 1.000]	0.42 [0.36, 0.48]	-0.49 [-0.55, -0.44]	0.0050 [0.0048, 0.0052]	24,882	64***

the stock markets, we observe dramatic rises in volatilities of the two exchange rates around this period. For example, the volatility of the EURO/USD rose abruptly from around 0.6 to 3.2 in 2008 according to the FSV model. Unlike the stock markets, Covid-19 has a relatively mild impact on the exchange rate volatility. There is an upsurge in the volatility of the GBP/USD when the Brexit referendum took place in June 2016.

9 Conclusions

The long-range dependence of volatilities is traditionally believed to be driven by a long memory process and modeled as either a fractionally integrated process with the memory parameter $d > 0$ or a fractional

Brownian motion process with the Hurst parameter $H > 0.5$. However, since Gatheral, Thibault, and Rosenbaum posted a working paper titled 'Volatility is rough' on SSRN in 2014, abundant empirical evidence of rough volatilities has been reported. Volatility is rough in the sense that the behavior of log volatility is well described by a fractional Brownian motion with $H < 0.5$. El Euch et al. (2018) provide a market microstructural foundation to rough volatility.

This paper contributes to this debate. We introduce a discrete-time fractional stochastic volatility model based on the fractional Gaussian noise (with the Hurst parameter H), which is asymptotically equivalent to the fractional integrated stochastic volatility model with fractional parameter $d = H - 0.5$. These two models include the standard stochastic volatility model as a special case with $H = 0.5$ for the FSV model and $d = 0$ for the FISV model. We examine the theoretical properties of these two models, allowing the memory parameter to take various values. In particular, the log volatility has a long memory and its long-run variance diverges to infinity when $H \in (0.5, 1)$ or $d \in (0, 0.5)$. In contrast, the log volatility is rough, with its long-run variance converging to a constant when $H \in (0, 0.5)$ or $d \in (-0.5, 0)$. We propose a memory signature plot, which reveals the dynamics of the cumulative sums of autocovariance of log volatilities and offers a straightforward assessment for the memory parameter range. Furthermore, we employ a simulated maximum likelihood method for the estimation of the model parameters. The (time-domain) likelihood function is evaluated with the importance sampling technique, where the Laplace approximation determines the proposal distribution. Unlike the existing rough volatility literature, we neither assume that the volatility process is directly observable nor use the estimated integrated or spot volatility to estimate the model. Instead, the estimation is based on log prices, and the volatility sequence is treated as latent variables. Our estimation method allows us to obtain both filtered and smoothed estimates of latent variables. Simulation studies show that the proposed SML method can accurately estimate both models.

We fit the proposed FSV model, the FISV model, and the standard SV model to several financial assets (S&P 500, Nikkei 225, EURO/USD, GBP/USD) over two sampling periods, each lasting about 25 years. Our empirical results suggest that log volatilities of those financial assets are persistent and rough. The estimated autoregressive coefficient of the log volatility process is very close to unity, and the estimated Hurst (fractional) parameters of the FSV (FISV) model are less than half (zero). The latter is consistent with the findings of the recent literature on rough volatility.

References

Andersen, T. G. and T. Bollerslev (1997). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Journal of Finance 52*(3), 975–1005.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association 96*(453), 42–55.

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

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. and J. Schmiegel (2008). Ambit processes; with applications to turbulence and tumour growth. In *Stochastic Analysis and Applications*, pp. 93–124. Springer.

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

Bennedsen, M., A. Lunde, and M. S. Pakkanen (2017). Decoupling the short-and long-term behavior of stochastic volatility. *arXiv preprint arXiv:1610.00332*.

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., J. Cai, and F. M. Song (2000). Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the us treasury bond market. *Journal of Empirical Finance 7*(1), 37–55.

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

Bollerslev, T. and J. H. Wright (2000). Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data. *Journal of Econometrics 98*(1), 81–106.

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 (1987). *Time Series: Theory and Methods*. Springer, New York.

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.

Dahlhaus, R. (1988). Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *The Annals of Statistics*, 808–841.

Davydov, Y. A. (1970). The invariance principle for stationary processes. *Theory of Probability & Its Applications 15*(3), 487–498.

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.

Douc, R., E. Moulines, J. Olsson, and R. van Handel (2011). Consistency of the maximum likelihood estimator for general hidden markov models. *Annals of Statistics 39*, 474–513.

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

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.

Fukasawa, M., T. Takabatake, and R. Westphal (2019). Is volatility rough? *arXiv preprint arXiv:1905.04852*.

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

Gneiting, T. and M. Schlather (2004). Stochastic models that separate fractal dimension and the hurst effect. *SIAM Review 46*(2), 269–282.

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

Gradshteyn, I. S. and I. M. Ryzhik (2014). *Table of Integrals, Series, and Products*. Academic press.

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.

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.

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.

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 (2019). Estimation and inference of fractional continuous-time model with discrete-sampled data.

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

A Appendix: Proofs

Proof of Theorem 3.1. The covariance between $h_{t\Delta}$ and $h_{s\Delta}$

$$\begin{aligned} Cov\left(h_{t\Delta}, h_{s\Delta}\right) &= \sigma_{\eta}^{*2}\Delta^{2H}\sum_{i=k}^{\infty}\sum_{j=0}^{\infty}\beta^{i+j-k}\gamma_{\eta}\left(|i-j|\right) \\ &= \sigma_{\eta}^{*2}\Delta^{2H}\beta^{-k}\left[\sum_{i=k}^{\infty}\beta^{2i}\gamma_{\eta}\left(0\right)+\sum_{i>j}\beta^{i+j}\gamma_{\eta}\left(i-j\right)+\sum_{j>i}\beta^{i+j}\gamma_{\eta}\left(j-i\right)\right]. \end{aligned}$$

The second and third terms in the squared bracket can be rewritten as

$$\begin{aligned} \sum_{i>j}\beta^{i+j}\gamma_{\eta}\left(i-j\right) &= \sum_{i=k}^{\infty}\sum_{j=0}^{i-1}\beta^{i+j}\gamma_{\eta}\left(i-j\right)=\sum_{i=k}^{\infty}\sum_{l=1}^{i}\beta^{2i-l}\gamma_{\eta}\left(l\right) \\ &= \sum_{l=1}^{k}\gamma_{\eta}\left(l\right)\sum_{i=0}^{\infty}\beta^{2(k+i)-l}+\sum_{l=k+1}^{\infty}\gamma_{\eta}\left(l\right)\sum_{i=0}^{\infty}\beta^{l+2i} \\ &= \frac{\beta^{2k}}{1-\beta^{2}}\sum_{l=1}^{k}\beta^{-l}\gamma_{\eta}\left(l\right)+\frac{1}{1-\beta^{2}}\sum_{l=k+1}^{\infty}\beta^{l}\gamma_{\eta}\left(l\right) \end{aligned}$$

and

$$\sum_{j>i}\beta^{i+j}\gamma_{\eta}\left(j-i\right)=\sum_{i=k}^{\infty}\sum_{j=i+1}^{\infty}\beta^{i+j}\gamma_{\eta}\left(j-i\right)=\sum_{i=k}^{\infty}\sum_{l=1}^{\infty}\beta^{2i+l}\gamma_{\eta}\left(l\right)=\frac{\beta^{2k}}{1-\beta^{2}}\sum_{l=1}^{\infty}\beta^{l}\gamma_{\eta}\left(l\right).$$

It follows that

$$\begin{aligned} Cov\left(h_{t\Delta}, h_{s\Delta}\right) &= \frac{\sigma_{\eta}^{*2}\Delta^{2H}}{1-\beta^{2}}\left[\beta^{k}+\beta^{k}\sum_{l=1}^{k}\beta^{-l}\gamma_{\eta}\left(l\right)+\beta^{-k}\sum_{l=k+1}^{\infty}\beta^{l}\gamma_{\eta}\left(l\right)+\beta^{k}\sum_{l=1}^{\infty}\beta^{l}\gamma_{\eta}\left(l\right)\right] \\ &= \frac{\sigma_{\eta}^{*2}\Delta^{2H}}{1-\beta^{2}}\beta^{k}\left(1+\sum_{l=1}^{k}\beta^{-l}\gamma_{\eta}\left(l\right)+\beta^{-2k}\varsigma_{H,\beta,k+1}+\varsigma_{H,\beta,1}\right) \end{aligned}$$

where $\varsigma_{H,\beta,s}\equiv\sum_{l=s}^{\infty}\beta^{l}\gamma_{\eta}\left(l\right)$. Similarly, the variance of $h_{t\Delta}$

$$\begin{aligned} \gamma_{h}\left(0\right) &= \sigma_{\eta}^{*2}\Delta^{2H}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\beta^{i+j}\gamma_{\eta}\left(|i-j|\right) \\ &= \sigma_{\eta}^{*2}\Delta^{2H}\left[\sum_{i=0}^{\infty}\beta^{2i}\gamma_{\eta}\left(0\right)+\sum_{i>j}\beta^{i+j}\gamma_{\eta}\left(i-j\right)+\sum_{j>i}\beta^{i+j}\gamma_{\eta}\left(j-i\right)\right] \\ &= \frac{\sigma_{\eta}^{*2}\Delta^{2H}}{1-\beta^{2}}\left[1+2\sum_{l=1}^{\infty}\beta^{l}\gamma_{\eta}\left(l\right)\right]=\frac{\sigma_{\eta}^{*2}\Delta^{2H}}{1-\beta^{2}}\left(1+2\varsigma_{H,\beta,1}\right) \end{aligned}$$

since

$$\sum_{i>j}\beta^{i+j}\gamma_{\eta}\left(i-j\right)=\frac{1}{1-\beta^{2}}\sum_{l=1}^{\infty}\beta^{l}\gamma_{\eta}\left(l\right)\text{ and }\sum_{j>i}\beta^{i+j}\gamma_{\eta}\left(j-i\right)=\frac{1}{1-\beta^{2}}\sum_{l=1}^{\infty}\beta^{l}\gamma_{\eta}\left(l\right).$$

We first derive properties of $\gamma_h(0)$ and $\gamma_h(k)$. By the definition of the upper incomplete Gamma function $\Gamma(s,z)=\int_z^\infty t^{s-1}e^{-t}dt$ when $s\neq 0$,

$$\int_s^\infty l^{2H-2}\beta^l dl=\left[\int_{-s\ln\beta}^\infty(-l\ln\beta)^{2H-2}e^{l\ln\beta}d(-l\ln\beta)\right](-\ln\beta)^{1-2H}=\Gamma(2H-1,-s\ln\beta)(-\ln\beta)^{1-2H}<\infty$$

as $H\neq 0.5$. This implies that both $\varsigma_{H,\beta,1}=\sum_{l=1}^\infty\beta^l\gamma_\eta(l)<\infty$ and $\varsigma_{H,\beta,k+1}=\sum_{l=k+1}^\infty\beta^l\gamma_\eta(l)<\infty$ when $H\neq 0.5$. Therefore,

$$\begin{aligned}\gamma_h(0) &= \frac{\sigma_\eta^{*2}\Delta^{2H}}{1-\beta^2}(1+2\varsigma_{H,\beta,1})<\infty\\ \gamma_h(k) &= \frac{\sigma_\eta^{*2}\Delta^{2H}}{1-\beta^2}\beta^k\left(1+\delta_{H,\beta,k}+\beta^{-2k}\varsigma_{H,\beta,k+1}+\varsigma_{H,\beta,1}\right)<\infty.\end{aligned}$$

Next, we derive properties of the one-sided long run variance, which is defined as

$$\sum_{k=0}^\infty\gamma_h(k)=\gamma_h(0)+\frac{\sigma_\eta^{*2}\Delta^{2H}}{1-\beta^2}\left[\frac{\beta}{1-\beta}+\sum_{k=1}^\infty\beta^k\sum_{l=1}^k\beta^{-l}\gamma_\eta(l)+\sum_{k=1}^\infty\beta^{-k}\varsigma_{H,\beta,k+1}+\sum_{k=1}^\infty\beta^k\varsigma_{H,\beta,1}\right].$$

(1) When $H>0.5$, $2H-1>0$, $2H-2>-1$, and

$$\beta^k\sum_{l=1}^k\beta^{-l}\gamma_\eta(l)=\beta^k\sum_{l=1}^{k-1}\beta^{-l}\gamma_\eta(l)+\gamma_\eta(k)>\gamma_\eta(k)\sim H(2H-1)k^{2H-2}>H(2H-1)k^{-1}$$

when k is large. Since $\sum_{k=1}^\infty k^{-1}=\infty$, we have $\sum_{k=1}^\infty\beta^k\sum_{l=1}^k\beta^{-l}\gamma_\eta(l)=\infty$ and hence

$$\sum_{k=0}^\infty\gamma_h(k)=\infty.$$

(2) When $H<0.5$, $2H-1<0$. We have

$$\sum_{k=1}^\infty\beta^{-k}\varsigma_{H,\beta,k+1}\sim(-\ln\beta)^{1-2H}\sum_{k=1}^\infty\beta^{-k}\Gamma(2H-1,z_k),$$

where $z_k=-(k+1)\ln\beta\to+\infty$ as $k\to\infty$. By the recurrence property of the incomplete Gamma function,

$$\Gamma(2H-1,z_k)=2(H-1)\Gamma(2H-2,z_k)+z_k^{2(H-1)}e^{-z_k}.$$

It follows that when k is large,

$$\beta^{-k}\Gamma(2H-1,z_k)\sim\beta^{-k}e^{-z_k}z_k^{2H-2}\sim(k+1)^{2H-2}.$$

By the Cauchy condensation test, $\sum_{k=1}^\infty(k+1)^{2H-2}<\infty$ when $2H-2<-1$. Therefore, $\sum_{k=1}^\infty\beta^{-k}\varsigma_{H,\beta,k+1}$ exists. Moreover,

$$\sum_{k=1}^\infty\beta^k\sum_{l=1}^k\beta^{-l}\gamma_\eta(l)=\sum_{l=1}^\infty\sum_{k=l}^\infty\gamma_\eta(l)\beta^{k-l}\sim\sum_{l=1}^\infty\sum_{k=l}^\infty l^{2H-2}\beta^{k-l}=\sum_{l=1}^\infty l^{2H-2}\sum_{k=l}^\infty\beta^{k-l}=\frac{1}{1-\beta}\sum_{l=1}^\infty l^{2H-2},$$

exists if $H<0.5$ but does not exist if $H>0.5$. Therefore, when $H<0.5$,

$$\sum_{k=0}^\infty\gamma_h(k)<\infty.$$

■

Proof of Theorem 3.2. When $H = 0.5$, the covariance (13) becomes

$$\begin{aligned} Cov\left(h_{t\Delta}, h_{s\Delta}\right) &= \sigma_\eta^{*2}\Delta\sum_{i=1}^{t}\beta^{t+s-2i}\gamma_\eta(0) = \sigma_\eta^{*2}\Delta\beta^{s-t}\sum_{i=1}^{t}\beta^{2t-2i} \\ &= \sigma_\eta^{*2}\Delta\beta^k\frac{1-\beta^{2t}}{1-\beta^2} \to \sigma_\eta^{*2}\Delta\frac{\beta^k}{1-\beta^2} \end{aligned}$$

since $\beta < 1$ and $\beta^{2t} \to 0$. When $H \neq 0.5$, (13) can be rewritten as

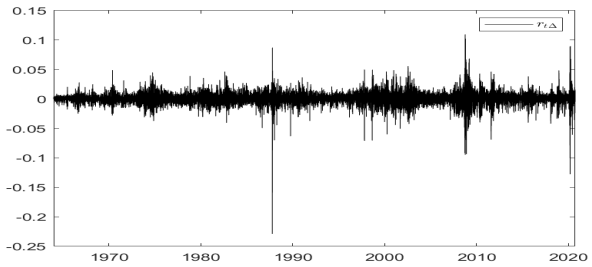
$$\begin{aligned} & Cov\left(h_{t\Delta}, h_{s\Delta}\right) \\ = \;& \sigma_\eta^{*2}\Delta^{2H}\beta^{s-t}\sum_{i=1}^{t}\sum_{j=1}^{s}\beta^{2t-i-j}\gamma_\eta\left(|i-j|\right) \\ = \;& \sigma_\eta^{*2}\Delta^{2H}\beta^k\left[\sum_{i=1}^{t}\beta^{2t-2i}\gamma_\eta\left(0\right) + \sum_{i>j}\beta^{2t-i-j}\gamma_\eta\left(i-j\right) + \sum_{j-i>k}\beta^{2t-i-j}\gamma_\eta\left(j-i\right) + \sum_{1\leq j-i\leq k}\beta^{2t-i-j}\gamma_\eta\left(j-i\right)\right] \\ = \;& \sigma_\eta^{*2}\Delta^{2H}\frac{\beta^k}{1-\beta^2}\left[1-\beta^{2t} + \sum_{l=1}^{t-1}\gamma_\eta\left(l\right)\left(\beta^l - \beta^{2t-l}\right) + \sum_{l=k+1}^{s-1}\gamma_\eta\left(l\right)\left(\beta^{-2k+l} - \beta^{2t-l}\right) + \sum_{l=1}^{k}\gamma_\eta\left(l\right)\left(\beta^{-l} - \beta^{2t-l}\right)\right] \\ \to \;& \sigma_\eta^{*2}\Delta^{2H}\frac{\beta^k}{1-\beta^2}\left[1 + \sum_{l=1}^{\infty}\beta^l\gamma_\eta\left(l\right) + \beta^{-2k}\sum_{l=k+1}^{\infty}\gamma_\eta\left(l\right)\beta^l + \sum_{l=1}^{k}\gamma_\eta\left(l\right)\beta^{-l}\right] \\ = \;& \sigma_\eta^{*2}\Delta^{2H}\frac{\beta^k}{1-\beta^2}\left[1 + \sum_{l=1}^{k}\beta^{-l}\gamma_\eta\left(l\right) + \beta^{-2k}\varsigma_{H,\beta,k+1} + \varsigma_{H,\beta,1}\right] \end{aligned}$$

since $\beta < 1$ and $\beta^{2t-l} \to 0$. ■

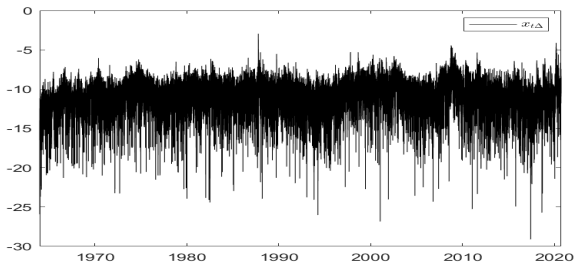
B Appendix: Figures and Tables

Figure B.12: The dynamics of $r_{t\Delta}$ and $x_{t\Delta}$ over the whole sample period

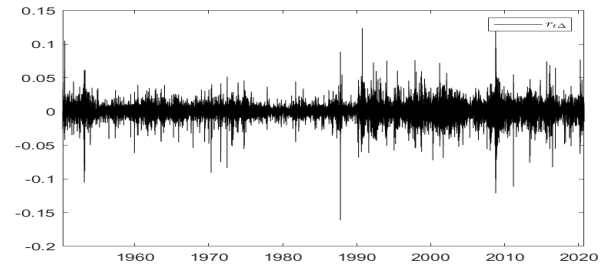
(a) S&P 500: $r_{t\Delta}$



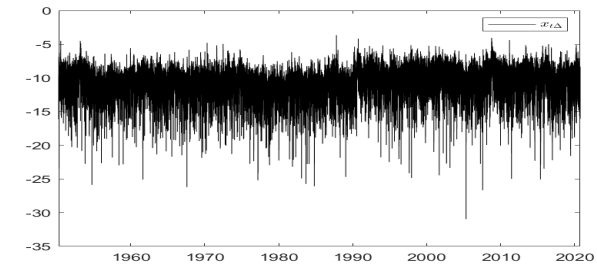
(b) S&P 500: $x_{t\Delta}$



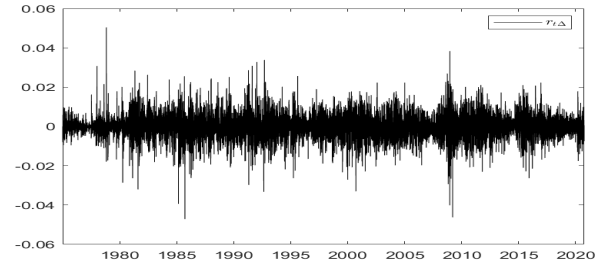
(c) NIKKEI 225: $r_{t\Delta}$



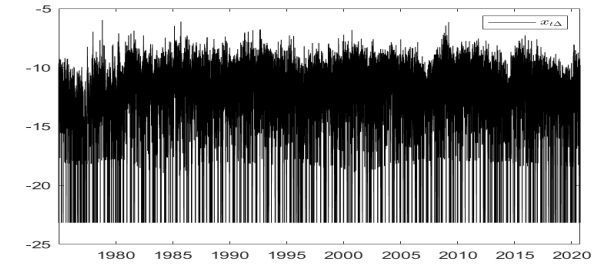
(d) NIKKEI 225: $x_{t\Delta}$



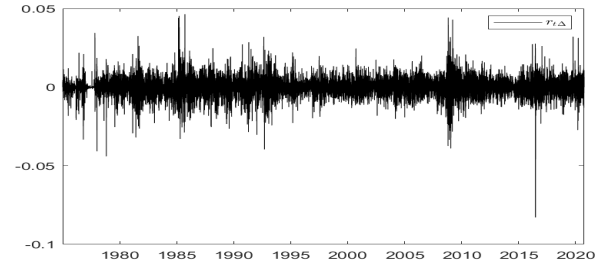
(e) EURO/USD



(f) EURO/USD



(g) GBP/USD



(h) GBP/USD

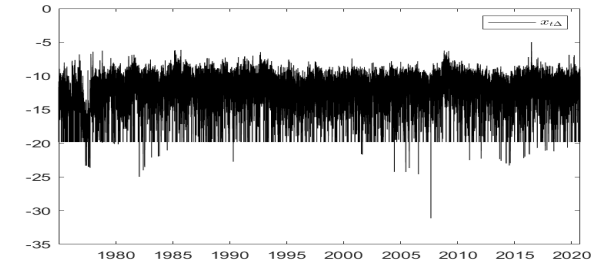
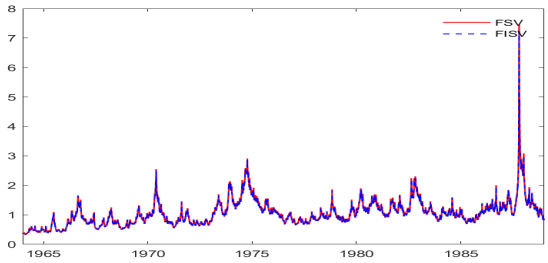
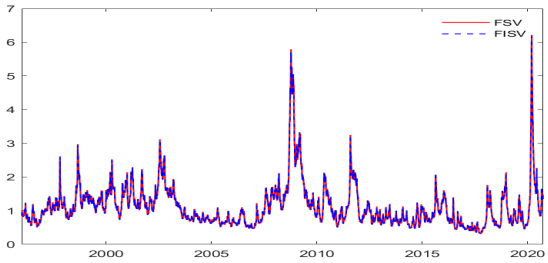


Figure B.13: The smoothed volatility (38) from the FSV and FISV models

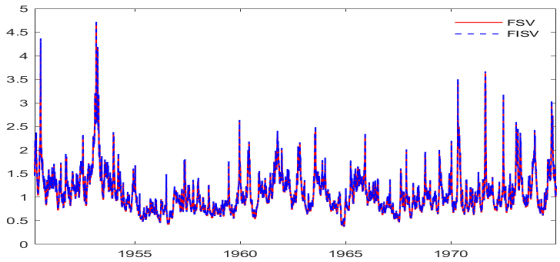
(a) S&P 500: Sample I



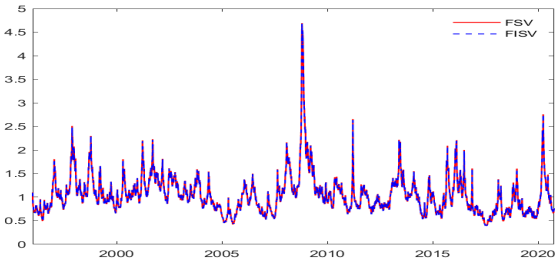
(b) S&P 500: Sample II



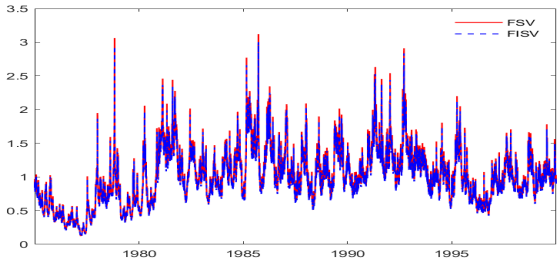
(c) NIKKEI 100: Sample I



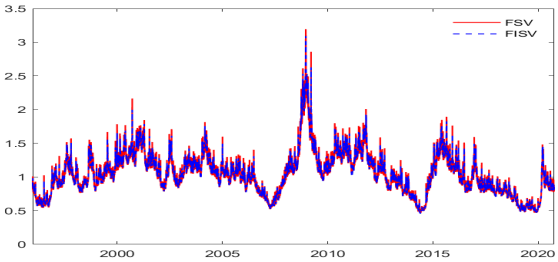
(d) NIKKEI 100: Sample II



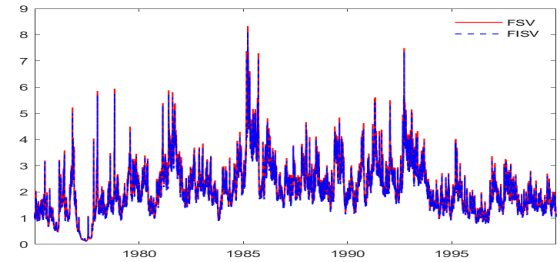
(e) EURO/USD: Sample I



(f) EURO/USD: Sample II



(g) GBP/USD: Sample I



(h) GBP/USD: Sample II

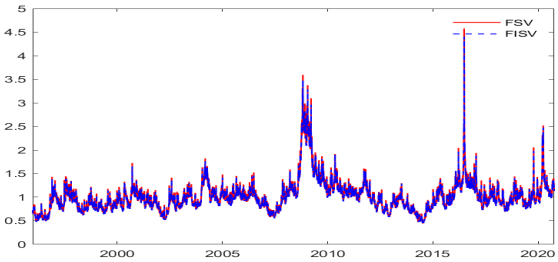


Table B.7: The mean and RMSE of estimated model parameters of FSV. The true model parameters are $\beta = 0.9961, \sigma_\eta^* = 1.5, \sigma^* = 0.1$.

	$\hat{\beta}$			σ_η^*			$\hat{H}$			$\hat{\sigma}^*$		
	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
						$T = 1280$						
Finite sample variance-covariance matrix												
$H = 0.1$	0.9895	-0.0066	0.0106	1.5601	0.0601	0.2099	0.1304	0.0304	0.0440	0.1027	0.0027	0.0143
$H = 0.2$	0.9900	-0.0061	0.0116	1.7061	0.2061	0.5124	0.2404	0.0404	0.0874	0.1015	0.0015	0.0186
$H = 0.5$	0.9923	-0.0038	0.0076	1.9659	0.4659	1.0401	0.5601	0.0601	0.1911	0.0996	-0.0004	0.0181
$H = 0.7$	0.9950	-0.0011	0.0044	1.6302	0.1302	0.8274	0.6739	-0.0261	0.1803	0.0982	-0.0018	0.0160
$H = 0.9$	0.9960	-0.0001	0.0037	1.3180	-0.1820	0.6642	0.8246	-0.0754	0.1889	0.0994	-0.0006	0.0174
Asymptotic variance-covariance matrix												
$H = 0.1$	0.9905	-0.0056	0.0100	1.5516	0.0516	0.2061	0.1293	0.0293	0.0432	0.1033	0.0033	0.0160
$H = 0.2$	0.9902	-0.0059	0.0114	1.6805	0.1805	0.4742	0.2359	0.0359	0.0833	0.1028	0.0028	0.0181
$H = 0.5$	0.9917	-0.0044	0.0076	1.7551	0.2551	0.7595	0.5167	0.0167	0.1613	0.1046	0.0046	0.0283
$H = 0.7$	0.9943	-0.0018	0.0040	1.3630	-0.1370	0.5358	0.6021	-0.0979	0.1842	0.1079	0.0079	0.0384
$H = 0.9$	0.9961	0.0000	0.0023	0.8741	-0.6259	0.7279	0.6467	-0.2533	0.3053	0.1148	0.0148	0.0516
						$T = 2560$						
Finite sample variance-covariance matrix												
$H = 0.1$	0.9926	-0.0035	0.0053	1.4798	-0.0202	0.1288	0.1260	0.0260	0.0354	0.1011	0.0011	0.0082
$H = 0.2$	0.9937	-0.0024	0.0043	1.5491	0.0491	0.1996	0.2205	0.0205	0.0481	0.0999	-0.0001	0.0113
$H = 0.5$	0.9941	-0.0020	0.0052	1.7270	0.2270	0.7939	0.4986	-0.0014	0.1597	0.0986	-0.0014	0.0176
$H = 0.7$	0.9953	-0.0008	0.0041	1.6456	0.1456	0.8633	0.6534	-0.0466	0.1791	0.0990	-0.0010	0.0161
$H = 0.9$	0.9960	-0.0000	0.0027	1.3643	-0.1357	0.5165	0.8395	-0.0605	0.1534	0.1001	0.0001	0.0134
Asymptotic variance-covariance matrix												
$H = 0.1$	0.9932	-0.0029	0.0050	1.4747	-0.0253	0.1289	0.1251	0.0251	0.0347	0.1008	0.0008	0.0091
$H = 0.2$	0.9938	-0.0023	0.0042	1.5421	0.0421	0.1984	0.2189	0.0189	0.0474	0.0999	-0.0001	0.0118
$H = 0.5$	0.9942	-0.0019	0.0040	1.5512	0.0512	0.4828	0.4689	-0.0311	0.1340	0.1012	0.0012	0.0240
$H = 0.7$	0.9955	-0.0006	0.0028	1.3140	-0.1860	0.4583	0.5867	-0.1133	0.1739	0.1056	0.0056	0.0390
$H = 0.9$	0.9968	0.0007	0.0027	0.9079	-0.5921	0.6887	0.6755	-0.2245	0.2720	0.1155	0.0155	0.0610

Table B.8: The mean and RMSE of estimated model parameters of FISV. The true model parameters are $\rho = 0.9961, \sigma = 0.1, \sigma_u = 0.1\Delta^d$.

	$\hat{\rho}$			$\hat{\sigma}_u$			$\hat{d}$			$\hat{\sigma}$		
	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
						$T = 1,280$						
Finite sample variance-covariance matrix												
$d = -0.4$	0.9872	-0.0089	0.0145	0.7799	-0.1390	0.1533	-0.3212	0.0788	0.0967	0.1032	0.0032	0.0205
$d = -0.3$	0.9909	-0.0052	0.0085	0.4845	-0.0433	0.0840	-0.2581	0.0419	0.0870	0.1024	0.0024	0.0222
$d = 0$	0.9928	-0.0033	0.0063	0.0946	-0.0054	0.0279	0.0265	0.0265	0.0544	0.1023	0.0023	0.0127
$d = 0.2$	0.9945	-0.0016	0.0049	0.0499	0.0169	0.0422	0.1492	-0.0508	0.1812	0.1003	0.0003	0.0148
$d = 0.4$	0.9960	-0.0001	0.0036	0.0296	0.0187	0.0380	0.2229	-0.1771	0.2462	0.0991	-0.0009	0.0118
Asymptotic variance-covariance matrix												
$d = -0.4$	0.9881	-0.0079	0.0114	0.7825	-0.1364	0.1505	-0.3264	0.0736	0.0898	0.1050	0.0050	0.0248
$d = -0.3$	0.9904	-0.0057	0.0095	0.4898	-0.0380	0.0817	-0.2630	0.0370	0.0832	0.1029	0.0029	0.0199
$d = 0$	0.9915	-0.0046	0.0077	0.1101	0.0101	0.0530	0.0104	0.0104	0.1457	0.1043	0.0043	0.0282
$d = 0.2$	0.9936	-0.0025	0.0046	0.0578	0.0248	0.0452	0.1006	-0.0994	0.1780	0.1053	0.0053	0.0337
$d = 0.4$	0.9952	-0.0009	0.0031	0.0387	0.0278	0.0432	0.1482	-0.2518	0.2963	0.1074	0.0074	0.0420
						$T = 2,560$						
Finite sample variance-covariance matrix												
$d = -0.4$	0.9903	-0.0058	0.0074	0.7480	-0.1710	0.1750	-0.3150	0.0850	0.0903	0.1032	0.0032	0.0137
$d = -0.3$	0.9928	-0.0033	0.0050	0.4648	-0.0630	0.0764	-0.2522	0.0478	0.0649	0.1015	0.0015	0.0160
$d = 0$	0.9937	-0.0024	0.0042	0.0971	-0.0029	0.0305	0.0265	0.0265	0.0924	0.1020	0.0020	0.0161
$d = 0.2$	0.9948	-0.0013	0.0051	0.0421	0.0091	0.0256	0.1727	-0.0273	0.1467	0.1000	0.0000	0.0140
$d = 0.4$	0.9957	-0.0004	0.0056	0.0204	0.0096	0.0167	0.2813	-0.1187	0.1755	0.0999	-0.0001	0.0141
Asymptotic variance-covariance matrix												
$d = -0.4$	0.9905	-0.0056	0.0073	0.7491	-0.1699	0.1739	-0.3170	0.0830	0.0887	0.1037	0.0037	0.0134
$d = -0.3$	0.9927	-0.0034	0.0049	0.4662	-0.0616	0.0751	-0.2540	0.0460	0.0636	0.1024	0.0024	0.0141
$d = 0$	0.9936	-0.0025	0.0045	0.1087	0.0087	0.0459	0.0044	0.0044	0.1311	0.1033	0.0033	0.0237
$d = 0.2$	0.9948	-0.0013	0.0037	0.0484	0.0154	0.0279	0.1288	-0.0712	0.1423	0.1064	0.0064	0.0343
$d = 0.4$	0.9961	-0.0000	0.0028	0.0263	0.0154	0.0216	0.2130	-0.1870	0.2181	0.1111	0.0111	0.0510