Journal of Econometrics 235 (2023) 745–778

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A GMM approach to estimate the roughness of stochastic volatility☆

Anine E. Bolko [a], Kim Christensen [a,*], Mikko S. Pakkanen [b,a], Bezirgen Veliyev [a]

[a] Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark
[b] Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 4 May 2021
Received in revised form 9 May 2022
Accepted 4 June 2022
Available online 15 September 2022

JEL classification:
C10
C50

Keywords:
Fractional Brownian motion
GMM estimation
Hurst exponent
Log-normal stochastic volatility
Realized variance
Roughness

ABSTRACT

We develop a GMM approach for estimation of log-normal stochastic volatility models driven by a fractional Brownian motion with unrestricted Hurst exponent. We show that a parameter estimator based on the integrated variance is consistent and, under stronger conditions, asymptotically normally distributed. We inspect the behavior of our procedure when integrated variance is replaced with a noisy measure of volatility calculated from discrete high-frequency data. The realized estimator contains sampling error, which skews the fractal coefficient toward "illusive roughness." We construct an analytical approach to control the impact of measurement error without introducing nuisance parameters. In a simulation study, we demonstrate convincing small sample properties of our approach based both on integrated and realized variance over the entire memory spectrum. We show the bias correction attenuates any systematic deviance in the parameter estimates. Our procedure is applied to empirical high-frequency data from numerous leading equity indexes. With our robust approach the Hurst index is estimated around 0.05, confirming roughness in stochastic volatility.

1. Introduction

Stochastic volatility (SV) models are pervasive in finance. Over the years, a variety of different models–each with its own dynamics–were developed, such as the log-normal model (e.g. Taylor, 1986), the square-root diffusion (e.g. Heston, 1993), or more complicated processes where volatility is driven by a non-Gaussian pure-jump component, e.g. Barndorff-Nielsen and Shephard (2001), Todorov and Tauchen (2011).

In this paper, we investigate the log-normal SV model, which has been extensively studied in previous work, e.g. Alizadeh et al. (2002), Christoffersen et al. (2010), Hull and White (1987). This class is a promising starting point,

☆ A previous draft of this article was circulated under the title "Roughness in spot variance? A GMM approach for estimation of fractional log-normal stochastic volatility models using realized measures". The paper was presented at the 2019 "Frontiers in Quantitative Finance" workshop at Copenhagen University, Denmark, the 2021 European Winter Meeting of the Econometric Society, the 2022 DAGStat conference in Hamburg, Germany, the 2022 Aarhus–Singapore Management University (SMU) joint workshop on Volatility, the 2022 Vienna–Copenhagen (VieCo) conference on Financial Econometrics in Copenhagen, Denmark, and in seminars at McMaster University, SMU, and University of Waterloo. We appreciate input from attendees at these venues. We are also grateful for insightful comments and suggestions made by the co-editor (Torben G. Andersen), the associate editor, and two anonymous referees at the Journal of Econometrics. In addition, we thank Mikkel Bennedsen, Christa Cuchiero, Masaaki Fukasawa, Jim Gatheral, Shin Kanaya, Jia Li, Roberto Renò, Shuping Shi and Jun Yu for feedback. Christensen was partially funded by a grant from the Independent Research Fund Denmark (DFF 1028–00030B). This work was also supported by the Center for Research in Econometric Analysis of Time Series (CREATES).

* Corresponding author.
E-mail address: kim@econ.au.dk (K. Christensen).

https://doi.org/10.1016/j.jeconom.2022.06.009

because the unconditional distribution of realized variance is close to log-normal (see e.g. Andersen et al., 2001, 2003; Christensen et al., 2019).

While there is a general acceptance that log-normal volatility offers a decent description of return variation in financial asset prices, there is no consensus on the properties of the background driving Gaussian process. In the SV setting, it is often assumed to be a standard Brownian motion. The mean-reversion and volatility-of-volatility parameters then control both the local properties of volatility and also determine its long-run persistence. There are multiple papers dealing with estimation of the parameters of the log-normal SV model, for example using method of moments- or likelihood-based approaches (e.g. Taylor, 1986; Melino and Turnbull, 1990; Duffie and Singleton, 1993; Harvey et al., 1994; Gallant et al., 1997; Fridman and Harris, 1998). In the context of generalized method of moments (GMM) estimation, Andersen and Sørensen (1996) offer further advice on how to select moment criteria and the weighting matrix in order to get good results in small samples.

When the driving process is a fractional Brownian motion, which neither is a semimartingale nor has independent increments, less is known. In this setting, part of the memory in volatility is reallocated to the fractal index or Hurst (1951) exponent. Comte and Renault (1998) propose a log-normal SV model, where the Hurst exponent is larger than one-half; the value implied by a standard Brownian motion. This induces positive serial correlation–or long-memory–in the increments of the process. Bennedsen et al. (2022), El Euch and Rosenbaum (2018), Fukasawa et al. (2022), Gatheral et al. (2018), among others, study "roughness" as captured by a fractal index smaller than one-half, rendering volatility highly erratic–or anti-persistent–at short time scales. In the fractional setting, the Hurst index is typically pre-estimated based on a semi-parametric procedure, before the other parameters are recovered. While this may yield consistent parameter estimates, it is generally inefficient and may be severely biased in finite samples.

In this paper, we extend the classical GMM procedure to joint estimation of the parameters of the log-normal SV model with a general fractal index. An attractive feature of our paper is that moment expressions are derived in convenient integral form facilitating the implementation without recourse to simulation-based approaches. As in many papers preceding this one, we appeal to the time series properties of integrated variance to construct our estimator, an idea pioneered by Bollerslev and Zhou (2002).

In practice, the integrated variance is unobserved. Realized variance, which is computed from high-frequency data, is a consistent estimator of integrated variance and often replaces it in the calculations. In previous work, this is done by showing convergence in probability or distribution of the parameter estimator in a double-asymptotic in-fill and long-span setting, such that the volatility discretization error is small enough to be ignored. In the subsequent applications, however, the volatility proxy enters directly in place of integrated variance.

Substituting the latent volatility with a proxy, however, entails a measurement error in finite samples, which obfuscates the underlying integrated variance dynamics. This can be detrimental to the estimation procedure if unaccounted for (e.g. Meddahi, 2002; Hansen and Lunde, 2014), because the moment conditions imposed on integrated variance are generally not valid for the noisy proxy. Barndorff-Nielsen and Shephard (2002) employ a state–space system and the Kalman filter to smooth out realized variance prior to maximum quasi-likelihood estimation of their SV model, see also Meddahi (2003). In this paper, we introduce a high-level assumption employing a realized measure to proxy for integrated variance following Patton (2011). We construct a bias correction that controls for the measurement error and show how to embed it analytically into the GMM setting without introducing additional nuisance parameters, as opposed to Bollerslev and Zhou (2002).

Our proposed estimator is consistent and asymptotically normal. The main asymptotic theory is long-span with time going to infinity but high-frequency data sampled at a fixed frequency. As an aside, we complement the analysis by deriving the double-asymptotic result, where the estimation error correction is immaterial.

We investigate our estimator in a simulation study, where various configurations of a fractional log-normal SV model with Hurst parameter covering the rough, standard and long-memory setting are inspected. The parameter estimates are centered closely around the true values–suggesting our procedure is approximately unbiased–and relatively accurate, once the bias correction is adopted. In an empirical application, we study an extensive selection of major equity indexes and confirm rough volatility, even after ironing out the effect of noise in the volatility proxy. In those data we consistently locate a roughness parameter around 0.05.

The rest of this paper is organized as follows. Section 2 presents the log-normal SV model and studies the properties of integrated variance within this framework. The GMM approach is introduced in Section 3. Section 4 examines the performance of our estimator in a Monte Carlo study. In Section 5, we apply the procedure to real data and compare our findings with the previous literature. We conclude in Section 6 and leave some theoretical derivations to the appendix. An online-only supplemental appendix contains further Monte Carlo results and empirical analysis.

2. The setting

We model the log-price of a financial asset, $X = (X_t)_{t\geq 0}$, as an adapted continuous-time stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. We suppose a standard arbitrage-free market, in which asset prices are of semimartingale form (e.g., Back, 1991; Delbaen and Schachermayer, 1994). We assume X can be described by an Itô process:

$$X_t = X_0 + \int_0^t \mu_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s, \qquad t \geq 0, \tag{1}$$

where X_0 is $\mathcal{F}_0$-measurable, $\mu = (\mu_t)_{t\geq 0}$ is a predictable drift process, $\sigma = (\sigma_t)_{t\geq 0}$ is a càdlàg volatility process and $W = (W_t)_{t\geq 0}$ is a standard Brownian motion.

The spot variance $\sigma^2 = (\sigma_t^2)_{t\geq 0}$ is given by:

$$\sigma_t^2 = \xi \exp\left(Y_t - \frac{1}{2}\kappa(0)\right), \qquad t \geq 0, \tag{2}$$

where $\xi \in \Xi \subset (0, \infty)$ represents the unconditional mean of the stochastic variance, while $Y = (Y_t)_{t\geq 0}$ is a mean zero stationary Gaussian process with autocovariance function (acf) $\kappa(u) = \mathrm{cov}(Y_0, Y_u) = \kappa_\phi(u)$, $u \geq 0$, parameterized by $\phi \in \Phi \subset \mathbb{R}^p$. We assume Ξ and Φ are compact, so that $\Theta = \Xi \times \Phi \subset \mathbb{R}^{p+1}$ is compact, and write $\theta = (\xi, \phi) \in \Theta$.

Note that we do not restrict volatility to a Markovian or semimartingale setting.[1] This is not a problem for absence of arbitrage and existence of an equivalent risk-neutral probability measure (although it is not unique in our setup), since volatility is not the price of a tradable asset.[2] The main limitation of (2) is that it has continuous sample paths. Our model therefore excludes jumps in volatility, which are empirically relevant (e.g., Bandi and Renò, 2016; Eraker et al., 2003; Todorov and Tauchen, 2011). We leave a theoretical development of our framework in presence of volatility jumps to future research.[3]

To maintain a streamlined exposition, we also exclude a jump component in X. The theory should at least be robust to the addition of finite-activity jumps, but then one needs to pay attention to the practical implementation.[4]

The integrated variance on day t is defined as:

$$IV_t = \int_{t-1}^{t} \sigma_s^2 \mathrm{d}s, \qquad t \in \mathbb{N}, \tag{3}$$

and holds information on the parameters of the model. Our estimation procedure exploits this by measuring integrated variance on a daily basis. The subscript t indicates the end of a day. We later substitute integrated variance with a realized measure of volatility computed from intraday high-frequency data of X.

We exploit the dynamics of integrated variance in this paper. This is in contrast to the application of spot variance in previous work, e.g. Bennedsen et al. (2022), Fukasawa et al. (2022), Gatheral et al. (2018). While spot variance is more ideal, it is associated with numerous pitfalls in practice. Firstly, spot variance estimation requires ultra high-frequency data, which may not readily be available. Even if they are, sampling at the highest frequency may induce an accumulation of microstructure noise that can distort the analysis (e.g., Hansen and Lunde, 2006). The calculation of microstructure noise-robust estimators is complicated and they suffer from poor rates of convergence (e.g., Barndorff-Nielsen et al., 2008; Jacod et al., 2009; Zhang et al., 2005). Secondly, intraday spot variance is driven by a pronounced deterministic diurnal pattern, which needs to be controlled for if the properties of the underlying stochastic process are to be uncovered (Andersen and Bollerslev, 1997, 1998b). Working at the daily frequency sidesteps this problem. Thirdly, spot variance estimators converge at a slow rate–relative to estimators of integrated variance–and, in the context of our model, often lack associated CLTs. The smoothing entailed by integrating spot variance overcomes this issue to some extent.

2.1. Moment structure of integrated variance

In this section, we derive the moment structure of integrated variance in (3) within the framework of the general log-normal SV process (1)–(2). This serves as the foundation of our GMM procedure to estimate the parameters of the model.

The starting point is the moment conditions:

$$\mathbb{E}\big[g(IV_t; \theta_0)\big] = 0, \tag{4}$$

for a measurable function g, where $\theta_0 \in \Theta$ is the true parameter vector.

[1] The log-normal distribution is invariant to (non-zero) power transformations. This implies that "volatility", which in financial economics is more often associated with the standard deviation–or the square-root of the variance–is also log-normal if the variance is (and vice versa). Hence, volatility is applied loosely here to mean either variance or standard deviation. In the text, the concrete meaning is apparent from the context and should not be the cause of any confusion.

[2] In 2004, CBOE launched derivatives on the VIX index, which is a weighted average of implied volatility from a basket of S&P 500 options, rendering volatility at least partially tradable (see, e.g., the white paper available at https://www.cboe.com/micro/vix/vixwhite.pdf).

[3] In the supplemental appendix, we simulate log-volatility as a jump–diffusion process driven by a standard Brownian motion with Hurst exponent $H = 1/2$. We implement the GMM procedure developed for the continuous sample path version of the model on it and observe that jumps in volatility do not cause any discernible spurious roughness. An explanation of this effect is offered there.

[4] The realized variance introduced in (20) is not a consistent estimator for integrated variance in the presence of price jumps. In such instances, the truncated realized variance of Mancini (2009) or the bipower variation of Barndorff-Nielsen and Shephard (2004) can be exploited. Below, we derive a bias correction for realized variance and bipower variation to account for measurement error. The correction developed for realized variance also applies to the truncated version, at least so long as the jumps are finite-activity, (see, e.g., Li et al., 2017, Proposition 1, which does not require semimartingale volatility and therefore is applicable in our setting).

The expectation in (4) can be hard to calculate for most choices of g. While spot variance is log-normal, the integrated variance is a sum of correlated log-normal random variables, and figuring out its distribution is a highly nontrivial statistical problem.[5] This also renders maximum likelihood estimation of the model complicated.

As the next theorem illuminates, the computation in (4) is manageable for raw moments, where the order of the expectation operator and volatility integral can be reversed.

Theorem 2.1. *Suppose that (1)–(2) hold. Then, the integrated variance process $(IV_t)_{t\in\mathbb{N}}$ is stationary with the following first and second-order moment structure:*

$$\begin{aligned}\mathbb{E}[IV_t] &= \xi,\\ \mathbb{E}[IV_t IV_{t+\ell}] &= \xi^2 \int_0^1 (1-y)\Big[\exp\big(\kappa(\ell+y)\big)+\exp\big(\kappa(|\ell-y|)\big)\Big]\mathrm{d}y,\end{aligned} \tag{5}$$

for $\ell \in \mathbb{N}\cup\{0\}$.

The third and fourth moment of integrated variance are:

$$\mathbb{E}[IV_t^3] = 6\xi^3 \int_0^1\int_0^x (1-x) f(x,y)\mathrm{d}y\mathrm{d}x, \tag{6}$$

$$\mathbb{E}[IV_t^4] = 24\xi^4 \int_0^1\int_0^x\int_0^y (1-x) g(x,y,z)\mathrm{d}z\mathrm{d}y\mathrm{d}x, \tag{7}$$

where

$$\begin{aligned}f(x,y) &= \exp\big(\kappa(|x-y|)+\kappa(|x|)+\kappa(|y|)\big),\\ g(x,y,z) &= \exp\big(\kappa(|x-y|)+\kappa(|x-z|)+\kappa(|y-z|)+\kappa(|x|)+\kappa(|y|)+\kappa(|z|)\big).\end{aligned} \tag{8}$$

In addition, suppose the following conditions hold:

(a) $\lim_{\ell\to\infty}\kappa(\ell)=0$,
(b) there exists an integrable function $\phi:[-1,1]\to\mathbb{R}$ *such that* $\frac{\kappa(\ell+y)}{\kappa(\ell)}\to\phi(y)$ *as* $\ell\to\infty$ *for any* $y\in[-1,1]$,
(c) $\limsup_{\ell\to\infty}\sup_{y\in[-1,1]}\left|\frac{\kappa(\ell+y)}{\kappa(\ell)}\right|<\infty$.

Then, as $\ell\to\infty$:

$$\mathbb{E}\big[(IV_t-\xi)(IV_{t+\ell}-\xi)\big] \sim \xi^2\kappa(\ell)\int_{-1}^1 (1-|y|)\phi(y)\mathrm{d}y, \tag{9}$$

where $f(\ell)\sim g(\ell)$ denotes asymptotic equivalence, i.e. $f(\ell)/g(\ell)\to 1$ as $\ell\to\infty$.

The proof of Theorem 2.1 exploits a convenient change of variables technique, which we introduce in Lemma A.1 in the appendix. This allows to express the rth moment of integrated variance as an $(r-1)$-dimensional integral.

Without proof, we conjecture a general formula for $\mathbb{E}[IV_t^r]$ by induction:

$$\mathbb{E}[IV_t^r] = r!\,\xi^r \int_0^1\int_0^{x_1}\cdots\int_0^{x_{r-2}} (1-x_1) g(x_1,\ldots,x_{r-1})\mathrm{d}x_{r-1}\ldots\mathrm{d}x_1, \tag{10}$$

where $r\geq 2$ is a positive integer, and

$$g(x_1,\ldots,x_{r-1}) = \exp\Bigg(\sum_{1\leq i<j\leq r-1}\kappa(|x_i-x_j|)+\sum_{i=1}^{r-1}\kappa(|x_i|)\Bigg). \tag{11}$$

In many realistic models, the above integrals do not possess analytic solutions and need to be approximated or solved numerically. In the fractional SV model entertained below, the acf of spot variance has a sharp incline near the origin, which gets steeper for smaller H. In that model, even the third moment can be prohibitively time-consuming to calculate, at least within the grasp of our computing powers. This makes higher-order moments unwieldy to work with in practice. Furthermore, because the distribution of integrated variance is generally heavy-tailed and highly right-skewed, such moments are also hard to estimate. As if that was not bad enough, in the noisy proxy setting the moment conditions incorporate estimation error. The updated expectations involve a convolution of the integrated variance and measurement error. This is tough to deal with even for raw moments. As a practical compromise, our estimation procedure therefore relies on low-order moments.

[5] The distribution of a sum of correlated log-normals can be approximated by the Fenton–Wilkinson method, where one replaces it with a single log-normal random variable. However, the approximation is often inaccurate.

In the last part of Theorem 2.1, condition (a) is a minimal condition that is necessary and sufficient for a stationary Gaussian process to be ergodic. This follows from the classical result of Maruyama (1949). It is evidently true for the models in this paper. Condition (b) and (c) are more technical and restrict the oscillation of $\kappa(\ell)$ as $\ell \to \infty$.

As an illustration, suppose there exists $\ell_0 > 0$ such that

$$\kappa(\ell) = \ell^{-\beta} e^{-\rho\ell} L(\ell), \quad \ell \geq \ell_0, \tag{12}$$

where $\beta \geq 0$ and $\rho \geq 0$ with $\min(\beta, \rho) > 0$, and for some slowly varying function $L : (0, \infty) \to (0, \infty)$, i.e. a function for which $\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1$ for all $c > 0$. For example, if $L(x)$ converges to a strictly positive limit as $x \to \infty$, then it is evidently slowly varying. Appealing to the uniform convergence theorem for slowly varying functions (Bingham et al., 1989, Theorem 1.5.2), condition (a)–(c) can be verified with $\phi(y) = e^{-\rho y}$.

2.2. The fractional SV model

While the theoretical results developed in this paper apply more broadly to general log-normal SV models, in our illustrations–both the simulation study and empirical application–we zoom in on the fractional SV (fSV) model, in which the volatility is assumed to be the exponential of a fractional Ornstein–Uhlenbeck (fOU) process:

$$Y_t = \nu \int_0^t e^{-\lambda(t-s)} \mathrm{d}B_s^H, \quad t \geq 0, \tag{13}$$

where $\nu, \lambda > 0$, and $B^H = (B_t^H)_{t\geq 0}$ is a fractional Brownian motion (fBm) with Hurst index $H \in (0, 1)$.[6] This is a standard log-normal SV model for $H = 1/2$. The fractional version was introduced by Comte and Renault (1998) in a long-memory setting ($H > 1/2$) and recently in a rough setting ($H < 1/2$) by Gatheral et al. (2018). Our main interest in this model is that we aim to assess the empirical level of the Hurst exponent, when all parameters are estimated jointly. Hence, in our Monte Carlo analysis we pay particular attention to the accuracy of the estimation error in H.[7]

The acf of the fOU model was derived in the following convenient form by Garnier and Sølna (2018), which–when adapted to the present parameterization of the model and assuming Y_0 is drawn from its stationary distribution–can be expressed as:

$$\kappa(\ell) = \frac{\nu^2}{2\lambda^{2H}} \left(\frac{1}{2} \int_{-\infty}^{\infty} e^{-|y|} |\lambda\ell + y|^{2H} \mathrm{d}y - |\lambda\ell|^{2H} \right), \quad \ell \geq 0. \tag{14}$$

For $H = 1/2$ this expression reduces to:

$$\kappa(\ell) = \frac{\nu^2}{2\lambda} e^{-\lambda\ell}, \tag{15}$$

which is the standard formula for the log-normal SV model.

As apparent from (15), the acf decays exponentially for $H = 1/2$. However, (14) has a hyperbolic rate of decay for other values of H, since $\kappa(\ell) = O\left(\ell^{2(H-1)}\right)$ as $\ell \to \infty$. This follows from Cheridito et al. (2003). These properties are transferred to the integrated variance process by the last result in Theorem 2.1. Moreover, the acf is integrable for $H \leq 1/2$ but not integrable for $H > 1/2$.

The variance of Y_t is:

$$\kappa(0) = \frac{\nu^2}{2\lambda^{2H}} \Gamma(1 + 2H). \tag{16}$$

The fSV model conforms to (12) with $\beta = 2(1 - H)$ and $\rho = 0$ for $H \in (0, 1/2) \cup (1/2, 1)$, by Theorem 2.3 of Cheridito et al. (2003), and with $\beta = 0$ and $\rho = \lambda$ for $H = 1/2$. We stress that L need not be given in closed form, as the proof of (12) amounts to checking that $L(\ell) \equiv \frac{\kappa(\ell)}{\ell^{-\beta} e^{-\rho\ell}}$ is slowly varying, based on the asymptotic behavior of $\kappa(\ell)$ as $\ell \to \infty$.

3. GMM estimation

In this section, for technical convenience we define all processes also for negative time indices.

[6] A fBm started at the origin ($B_0^H = 0$) with Hurst exponent $H \in (0, 1)$ is a centered Gaussian process with covariance function $E\left[B_t^H B_s^H\right] = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t - s|^{2H}\right)$. It can be constructed as a weighted infinite moving average of past increments to a standard Brownian motion following the Mandelbrot and Van Ness (1968, Definition 2.1) representation: $B_t^H = \frac{1}{\Gamma(H+1/2)} \left\{ \int_{-\infty}^0 \left[(t - s)^{H-1/2} - (-s)^{H-1/2} \right] \mathrm{d}B_s + \int_0^t (t - s)^{H-1/2} \mathrm{d}B_s \right\}$, where $\Gamma(\cdot)$ is the Gamma function. The process is self-similar of index H and has stationary but not independent increments (except for $H = 1/2$). Its sample paths are locally Hölder continuous up to order H. As readily seen, the fBm reduces to a standard Brownian motion for $H = 1/2$.

[7] In Appendix C, we provide the associated analysis for GMM estimation of an alternative log-normal SV model, namely the Brownian semistationary process.

3.1. Assumptions and examples

As described above, the spot variance $\sigma^2 = (\sigma_t^2)_{t\in\mathbb{R}}$ depends on the parameter vector $\theta = (\xi, \phi) \in \Theta$. We write $\mathbb{P}_\theta$ for the probability measure induced by θ and $\mathbb{E}_\theta$ is the corresponding expectation operator (we sometimes omit θ when there is no risk of confusion). We denote by $\mathcal{F}^\sigma$ the σ-algebra generated by σ^2 or, equivalently, Y.

We now introduce our main assumption about Y.

Assumption 1. The Gaussian process Y and its covariance function κ satisfy the following conditions:

(i) Y has continuous sample paths for any $\phi \in \Phi$,
(ii) $(u, \phi) \mapsto \kappa_\phi(u)$ is a continuous function.

Condition (i) is natural for stationary Gaussian processes, since if Y was discontinuous, its sample paths would in fact be *unbounded* almost surely (Belyaev, 1961). Condition (ii) ensures that the moments of the model are continuous with respect to θ. It is worth pointing out that neither these conditions nor the stationarity of Y say much about the long-term behavior of volatility. We return to this in Assumption 2. In the fOU process, condition (i) has been shown in Proposition 3.4 of Kaarakka and Salminen (2011), while condition (ii) can be verified by applying the dominated convergence theorem to (14).

As the main object of interest, integrated variance, is not observable in practice, it needs to be estimated. We strive for a general framework applicable to realized measures *at large*, while still remaining analytically tractable. We postulate that we observe a noisy proxy of IV_t given by

$$\widehat{IV}_t = IV_t + \varepsilon_t, \tag{17}$$

where ε_t is a random variable capturing the measurement error, which needs to adhere to a set of stylized technical conditions given in Assumption 2. Such a high-level approach to describing measurement error between a realized measure and the corresponding integrated variance is reminiscent to what Patton (2011) uses for the analysis of noisy volatility proxies in the context of forecast evaluation.

To formalize our assumptions about the process $(\varepsilon_t)_{t\in\mathbb{Z}}$, we require a filtration $\mathcal{F}_t^{\sigma,\varepsilon} = \mathcal{F}_t^\sigma \vee \mathcal{F}_t^\varepsilon$, where $\mathcal{F}_t^\varepsilon = \sigma\big(\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}\big)$, $t \in \mathbb{Z}$, is the σ-algebra generated by the errors up to time t. We also introduce a key assumption about the joint long-term behavior of $(IV_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$.

Assumption 2. The processes $(IV_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ satisfy the following conditions:

(i) $(IV_t, \varepsilon_t)_{t\in\mathbb{Z}}$ is a stationary and ergodic process under $\mathbb{P}_\theta$ for any $\theta \in \Theta$,
(ii) $\theta \mapsto c(\theta) \equiv \mathbb{E}_\theta[\varepsilon_1^2]$ is a finite-valued, continuous function on Θ,
(iii) $\mathbb{E}_\theta[\varepsilon_t \mid \mathcal{F}_{t-1}^{\sigma,\varepsilon}] = 0$ for any $t \in \mathbb{Z}$ and any $\theta \in \Theta$.

The σ^2 and $(IV_t)_{t\in\mathbb{Z}}$ processes readily inherit the stationarity and ergodicity of Y. The joint ergodicity of $(IV_t, \varepsilon_t)_{t\in\mathbb{Z}}$ is more delicate, since even if $(IV_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ are ergodic on their own and mutually independent, it does not follow that $(IV_t, \varepsilon_t)_{t\in\mathbb{Z}}$ is ergodic (see Lindgren, 2006, Exercise 5.13). But if additionally $(IV_t)_{t\in\mathbb{N}}$ or $(\varepsilon_t)_{t\in\mathbb{Z}}$ is weakly mixing, then their joint ergodicity holds (see Lindgren, 2006, Exercise 5.14). That said, in practical applications the mutual independence of $(IV_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is too strong an assumption, since the level of measurement error typically depends on the underlying level of volatility. Condition (iii) is a martingale-difference property for $(\varepsilon_t)_{t\in\mathbb{Z}}$, which implies $\mathbb{E}_\theta[\varepsilon_t] = 0$, i.e. the proxy $\widehat{IV}_t$ is unbiased. This is obviously a somewhat stylized assumption, which is not exactly satisfied by many realized measures. However, we can expect it to hold approximately and, in any case, it is crucial for the analytical tractability of the setup.[8]

We now demonstrate that a particular structural form of the error term, ε_t, conveniently accommodates concrete realized measures as proxies in an approximate sense while satisfying Assumption 2. More specifically, let

$$\varepsilon_t = h\big(Z_t, (\sigma^2_{s+t-1})_{s\in[0,1]}\big), \quad t \in \mathbb{Z}, \tag{18}$$

where Z_t, $t \in \mathbb{Z}$, are i.i.d. d-dimensional random vectors, for some $d \in \mathbb{N}$, that are independent of $\mathcal{F}^\sigma$ and $h : \mathbb{R}^d \times C([0,1]) \to \mathbb{R}$ is a continuous functional such that

$$\mathbb{E}_\theta[h(Z_1, f)] = 0, \tag{19}$$

for any $\theta \in \Theta$ and $f \in C([0,1])$. Then condition (i) in Assumption 2 can be proved using standard ergodic theory arguments, see, e.g., Lindgren (2006, Section 5.4), while condition (iii) is readily implied by (19). We can check condition (ii) on a case-by-case basis below by computing $c(\theta) = \mathbb{E}_\theta[\varepsilon_1^2]$ explicitly and invoking condition (ii) of Assumption 1 to establish its continuity in θ.

In the following examples, we construct ε_t and Z_t only for $t \in \mathbb{N}$, but we can extend them to negative indices by stationarity.

[8] Meddahi (2002) studies realized variance under a class of log-normal volatility models including drift. He finds the mean of the measurement error to be negligible at a 5-minute sampling frequency.

Example 3.1 (*Realized Variance, CLT Approximation*)**.** Suppose that we estimate the integrated variance IV_t with the realized variance (see, e.g., Andersen and Bollerslev, 1998a; Barndorff-Nielsen and Shephard, 2002):

$$RV_t^n = \sum_{i=1}^{n} \left(X_{t-1+\frac{i}{n}} - X_{t-1+\frac{i-1}{n}}\right)^2, \tag{20}$$

for any $t \in \mathbb{N}$. Under standard technical conditions, the central limit theorem (CLT)

$$\sqrt{n}(RV_t^n - IV_t) \xrightarrow[n\to\infty]{d_{\mathrm{st}}} \sqrt{2}\int_{t-1}^{t} \sigma_s^2 \mathrm{d}B_s^{\perp}, \tag{21}$$

holds jointly for all $t \in \mathbb{N}$, where $\xrightarrow{d_{\mathrm{st}}}$ denotes stable convergence in distribution and $(B_s^{\perp})_{s\geq 0}$ is a Brownian motion independent of X and σ. Note that the limiting random variables $\sqrt{2}\int_{t-1}^{t} \sigma_s^2 \mathrm{d}B_s^{\perp}$, $t \in \mathbb{N}$, are conditionally independent given $\mathcal{F}^{\sigma}$ with

$$\sqrt{2}\int_{t-1}^{t} \sigma_s^2 \mathrm{d}B_s^{\perp} \,\middle|\, \mathcal{F}^{\sigma} \sim N(0, 2IQ_t), \quad t \in \mathbb{N}, \tag{22}$$

where

$$IQ_t = \int_{t-1}^{t} \sigma_s^4 \mathrm{d}s \tag{23}$$

is the integrated quarticity. Thus,

$$Z_t = \frac{\int_{t-1}^{t} \sigma_s^2 \mathrm{d}B_s^{\perp}}{IQ_t^{1/2}} \sim N(0, 1), \quad t \in \mathbb{N}, \tag{24}$$

are both mutually independent and independent of $\mathcal{F}^{\sigma}$.

Informally, the CLT (21) says that, for any $t \in \mathbb{N}$,

$$RV_t^n \overset{d}{\approx} IV_t + \left(\frac{2}{n} IQ_t\right)^{1/2} Z_t \tag{25}$$

for large n, where "$\overset{d}{\approx}$" denotes approximate equality in distribution, as used, e.g., in Zhang et al. (2005, Section 1.2). Thus, for any $t \in \mathbb{N}$, the proxy $\widehat{IV}_t = IV_t + \varepsilon_t$ with $\varepsilon_t = \left(\frac{2}{n} IQ_t\right)^{1/2} Z_t$ approximates RV_t^n for large n. Such a proxy is analogous to what Fukasawa et al. (2022) employ in their estimation framework. We can represent the error term as ε_t in the form (18) using the continuous functional

$$h(z, f) = \left(\frac{2}{n}\int_0^1 f(s)^2 \mathrm{d}s\right)^{1/2} z, \quad z \in \mathbb{R}, \quad f \in C([0, 1]). \tag{26}$$

Then (19) holds given that $Z_1 \sim N(0, 1)$. We can compute $c(\theta)$ explicitly using Tonelli's theorem. The expression is reported in Table 1 and $\theta \mapsto c(\theta)$ is evidently continuous under Assumption 1.

Example 3.2 (*Realized Variance, No Drift or Leverage Effect*)**.** In general, the measurement error $RV_t^n - IV_t$ is analytically hard to analyze unless we resort to asymptotic approximation with $n \to \infty$ as in Example 3.1 (see also the comments below). However, in a simple specific case, we can actually work with the exact error $\varepsilon_t = RV_t^n - IV_t$, that is $\widehat{IV}_t = RV_t^n$, without losing analytical tractability.

Namely, suppose that the log-price $X = (X_t)_{t\geq 0}$ of the asset follows a drift-free Itô process

$$X_t = X_0 + \int_0^t \sigma_s \mathrm{d}W_s, \quad t \geq 0, \tag{27}$$

where $W = (W_t)_{t\geq 0}$ is a standard Brownian motion independent of $\mathcal{F}^{\sigma}$, i.e. ruling out any dependence between W and the spot variance process σ^2, stemming from the leverage effect for instance (e.g., Christie, 1982). Then, for any $t \in \mathbb{N}$,

$$\begin{aligned} RV_t^n - IV_t &= \sum_{i=1}^{n}\left(\left(\int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1} \sigma_s \mathrm{d}W_s\right)^2 - \int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1} \sigma_s^2 \mathrm{d}s\right) \\ &= \sum_{i=1}^{n} (Z_{t,i}^2 - 1)\int_{\frac{i-1}{n}}^{\frac{i}{n}} \sigma_{s+t-1}^2 \mathrm{d}s, \end{aligned} \tag{28}$$

Table 1
Formulae for $c(\theta) = \mathbb{E}_\theta[\varepsilon_1^2]$.

Proxy	Setting	Example	$c(\theta) = c(\xi, \phi)$
Realized variance	CLT	3.1	$\frac{2\xi^2}{n}\exp(\kappa_\phi(0))$
	No drift or leverage	3.2	$\frac{4\xi^2}{n}\int_0^1(1-y)\exp(\kappa_\phi(\frac{y}{n}))\mathrm{d}y$
Bipower variation	CLT	3.3	$\frac{(\frac{\pi^2}{4}+\pi-3)\xi^2}{n}\exp(\kappa_\phi(0))$

Note. In the case of Example 3.2, Theorem 2.1 to derive the expression.

where

$$Z_{t,i} = \frac{\int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1}\sigma_s \mathrm{d}W_s}{\left(\int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1}\sigma_s^2\mathrm{d}s\right)^{1/2}}, \quad t \in \mathbb{N}, \quad i = 1, \ldots, n. \tag{29}$$

Since W is independent of $\mathcal{F}^\sigma$, conditional on $\mathcal{F}^\sigma$ the random variables $Z_{t,i}$, $t \in \mathbb{N}$, $i = 1, \ldots, n$, are mutually independent and follow a standard normal distribution. Consequently, they are i.i.d. standard normal also unconditionally and independent of $\mathcal{F}^\sigma$.

Thanks to (28), we can represent the measurement error $\varepsilon_t = RV_t^n - IV_t$ in the form (18) via the functional

$$h\big((z_1, \ldots, z_n), f\big) = \sum_{i=1}^n (z_i^2 - 1)\int_{\frac{i-1}{n}}^{\frac{i}{n}} f(s)\mathrm{d}s, \quad (z_1, \ldots, z_n) \in \mathbb{R}^n, \quad f \in C([0,1]), \tag{30}$$

and i.i.d. random vectors

$$Z_t = (Z_{t,1}, \ldots, Z_{t,n}), \quad t \in \mathbb{N}, \tag{31}$$

with components given by (29), so that $d = n$. The property (19) then holds, while an integral functional representation of $c(\theta)$ is given in Table 1 and its continuity in θ follows from the dominated convergence theorem under Assumption 1.

In the presence of a leverage effect, the moments of the measurement error $\varepsilon_t = RV_t^n - IV_t$ can be analyzed using Malliavin calculus and chaos expansions, see, e.g., Peccati and Taqqu (2011). However, the resulting formulae are not in any form convenient for numerical implementation, which is why we do not pursue this approach further here.

Example 3.3 (*Bipower Variation, CLT Approximation*)**.** In the context of Example 3.1, the realized variance can be substituted with the bipower variation estimator of Barndorff-Nielsen and Shephard (2004), which is defined as:

$$BV_t^n = \frac{\pi}{2}\sum_{i=2}^n \left|X_{t-1+\frac{i}{n}} - X_{t-1+\frac{i-1}{n}}\right|\left|X_{t-1+\frac{i-1}{n}} - X_{t-1+\frac{i-2}{n}}\right|, \quad n \in \mathbb{N}, \tag{32}$$

for any $t \in \mathbb{N}$. Under standard technical conditions

$$\sqrt{n}(BV_t^n - IV_t) \xrightarrow[n\to\infty]{d_{\mathrm{st}}} \sqrt{\frac{\pi^2}{4} + \pi - 3}\int_{t-1}^t \sigma_s^2 \mathrm{d}B_s^\perp, \tag{33}$$

jointly for all $t \in \mathbb{N}$, where the structure of the limit is identical to the one in (21). BV_t^n is then approximated for large n by the proxy $\widehat{IV}_t = IV_t + \varepsilon_t$ with error term $\varepsilon_t = \left(\frac{\frac{\pi^2}{4}+\pi-3}{n} IQ_t\right)^{1/2} Z_t$, where Z_t, $t \in \mathbb{N}$, are as in Example 3.1. Retracing the arguments in Example 3.1, we can then show that ε_t can be cast in the form (29), so Assumption 2 holds.

The appealing feature of bipower variation is that it remains consistent for integrated variance in the presence of price jumps. However, in the latter setting the CLT in (33) ceases to hold and the limiting distribution of BV_t^n is not mixed Gaussian (Vetter, 2010). Even without price jumps the asymptotic theory of bipower variation generally needs to impose smoothness conditions on volatility, typically in the form of an Itô semimartingale structure or, possibly, a long-memory process. To our knowledge, roughness is ruled out or at least remains undetermined. Hence, the approximate bias correction from Example 3.3 should be applied with caution. Nevertheless, it can serve as a heuristic double check.[9]

Table 1 summarizes the derivations for the measurement error. We reiterate that Examples 3.1 and 3.3 are based on $n \to \infty$, while any implementation is always done with a finite n. In contrast, Example 3.2 is valid for all n in absence of drift and leverage. In the simulations, we compare both expressions for realized variance to gauge their impact on the estimation.

[9] In the supplemental appendix, we repeat the simulation analysis and empirical application with bipower variation. The results do not change much compared to those for realized variance reported below.

3.2. Consistency

In this section, we turn to the consistency of our GMM estimator. We take $\theta \in \Theta$, $t \in \mathbb{Z}$ and $\ell \in \mathbb{Z}$ and introduce the moment structure of the IV_t process, which is defined by:

$$g_0^{(1)}(\theta) = \mathbb{E}_\theta[IV_t], \quad g_0^{(2)}(\theta) = \mathbb{E}_\theta[IV_t^2], \quad g_\ell(\theta) = \mathbb{E}_\theta[IV_t IV_{t-\ell}], \tag{34}$$

for $\ell = 1, \dots, k$, which we collect in the column vector

$$G(\theta) = \big(g_0^{(1)}(\theta), g_0^{(2)}(\theta), g_1(\theta), \dots, g_k(\theta)\big)^\intercal. \tag{35}$$

Here, $^\intercal$ is the transpose operator. We also define

$$\begin{aligned} \mathbb{IV}_t &= \big(IV_t, IV_t^2, IV_t IV_{t-1}, \dots, IV_t IV_{t-k}\big)^\intercal, \\ \widehat{\mathbb{IV}}_t &= \big(\widehat{IV}_t, \widehat{IV}_t^2, \widehat{IV}_t \widehat{IV}_{t-1}, \dots, \widehat{IV}_t \widehat{IV}_{t-k}\big)^\intercal, \end{aligned} \tag{36}$$

which by condition (i) of Assumption 2 are stationary and ergodic processes.

By condition (ii)–(iii) of Assumption 2:

$$\begin{aligned} \mathbb{E}_\theta\big[\widehat{IV}_t\big] &= g_0^{(1)}(\theta), \\ \mathbb{E}_\theta\big[\widehat{IV}_t \widehat{IV}_{t-\ell}\big] &= \begin{cases} g_0^{(2)}(\theta) + c(\theta), & \ell = 0, \\ g_\ell(\theta), & \ell \neq 0. \end{cases} \end{aligned} \tag{37}$$

The noisy proxy changes the second moment in (37) compared to integrated variance in (5). The term $c(\theta)$ is induced by the noise. This observation was also made by Bollerslev and Zhou (2002). They account for the error-in-variables by including an additive nuisance parameter to the second-order moment condition, which is estimated independently of the structural parameters. In contrast, we incorporate the error variation directly into the model as a function of θ, which avoids the need of an extra parameter.

The first- and other second-order moments are unbiased, due to the linearity of the expectation operator and because the errors are mean zero and serially uncorrelated. In principle, we can thus avoid the negative impact of measurement errors by excluding $g_0^{(2)}(\theta)$ from the moment selection. More generally, however, it is often preferable to add the variance or absolute value to the moment conditions, because low-order moments are highly informative about the parameters of SV models (Andersen and Sørensen, 1996). To avoid any systematic deviance in the estimated values of the parameters, it is then necessary to correct the appropriate entries in the moment vector as detailed above (dealing with the measurement error is of course much more complicated for the absolute value than for the square).

We propose to compare the sample moments of $\widehat{IV}_t$ to a corrected moment function

$$G_c(\theta) = G(\theta) + \big(0, c(\theta), 0, \dots, 0\big)^\intercal. \tag{38}$$

We define a random function:

$$\widehat{m}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \widehat{\mathbb{IV}}_t - G_c(\theta), \tag{39}$$

which, in view of (37), has

$$\mathbb{E}_{\theta_0}\big[\widehat{m}_T(\theta)\big] = G_c(\theta_0) - G_c(\theta) \equiv m(\theta), \tag{40}$$

so that

$$m(\theta_0) = 0. \tag{41}$$

Our GMM estimator is then given by

$$\widehat{\theta}_T = \arg\min_\theta \ \widehat{m}_T(\theta)^\intercal \mathbb{W}_T \widehat{m}_T(\theta), \tag{42}$$

where $\mathbb{W}_T$ is a random $(k+2) \times (k+2)$ weight matrix.

We need additional conditions for the consistency of $\widehat{\theta}_T$. Firstly, we introduce a standard assumption about the limiting behavior of $\mathbb{W}_T$.

Assumption 3. $\mathbb{W}_T = A_T^\intercal A_T$ for a random $(k+2) \times (k+2)$ matrix A_T, which under $\mathbb{P}_{\theta_0}$ converges almost surely to a non-random matrix A as $T \to \infty$.

Secondly, we assume that the parameters are identifiable.

Assumption 4. $Am(\theta) = 0$ if and only if $\theta = \theta_0$.

Assumption 4 is a standard identification condition in GMM that ensures uniqueness of the solution. It is hard to check when moments are not given in algebraic form, see e.g. Barboza and Viens (2017), Newey and McFadden (1994) in the GMM setting or Corradi and Distaso (2006), Todorov (2009) in the context of estimation of SV models.

An alternative route to inspect identification is to perform a rank test on the Jacobian matrix. This amounts to verifying that $\nabla_\theta Am(\theta_0)$ has full column rank. The latter is equivalent to Assumption 4 if the moment conditions are linear in the parameters. In our setting, this is not the case (except for the mean). Hence, the rank condition can only help to identify parameters locally in a neighborhood of a solution candidate. Nevertheless, developing a formal rank test for local identification in the fSV model presents a challenge. This requires that we derive the asymptotic distribution of $T^{-1/2}\nabla_\theta A_T \widehat{m}_T(\theta)$, e.g. Wright (2003). We do not pursue the idea here but leave it to future research. Instead, in Appendix B we offer some alternative insights about identification in the fSV model based on the notion of equality of an sequence of moment conditions. As noted there, ξ and H are identified, whereas ν and λ are identified only through their ratio.

Theorem 3.4. *Suppose Assumptions 1–4 hold. As $T \to \infty$*

$$\widehat{\theta}_T \xrightarrow{a.s.} \theta_0. \tag{43}$$

In the above, our analysis assumed that the number of observations per day, n, is fixed and then relies on the noisy proxy idea. Now, following Bollerslev and Zhou (2002), Corradi and Distaso (2006), Todorov (2009), we also cover the theory of the GMM estimator in a double-asymptotic setting with $T \to \infty$ and $n \to \infty$.

To this end, we denote with V_t^n some consistent realized measure of integrated variance (e.g., realized variance, bipower variation, or truncated realized variance). For fixed $k \in \mathbb{N}$, we set

$$\mathbb{V}_t^n = \left(V_t^n, (V_t^n)^2, V_t^n V_{t-1}^n, \dots, V_t^n V_{t-k}^n\right)^\intercal, \tag{44}$$

with associated sample moments

$$\widetilde{m}_{n,T}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{V}_t^n - G(\theta), \tag{45}$$

where we employ the moments of $G(\theta)$ instead of the corrected version $G_c(\theta)$, which is of no consequence for the following result since $n \to \infty$.

Then,

$$\widetilde{\theta}_{n,T} = \arg\min_{\theta\in\Theta} \widetilde{m}_{n,T}(\theta)^\intercal \mathbb{W}_{n,T} \widetilde{m}_{n,T}(\theta), \tag{46}$$

is our GMM estimator.

In this setting, we replace Assumption 2 with the following requirement.

Assumption 5. The processes $(IV_t)_{t\in\mathbb{Z}}$ and $(V_t^n)_{t\in\mathbb{Z},n\in\mathbb{N}}$ admit the following:

(i) $(IV_t)_{t\in\mathbb{Z}}$ is a stationary and ergodic process under $\mathbb{P}_\theta$ for any $\theta \in \Theta$,
(ii) $\sup_{t\in\mathbb{Z}} \mathbb{E}_{\theta_0}[(V_t^n - IV_t)^2] \to 0$ as $n \to \infty$.

Theorem 3.5. *Suppose Assumptions 1, 3–5 hold. As $T \to \infty$ and $n \to \infty$*

$$\widetilde{\theta}_{n,T} \xrightarrow{\mathbb{P}} \theta_0. \tag{47}$$

This result is equivalent to Theorem 1 (and Corollary 1) in Todorov (2009) and Theorem 1 in Corradi and Distaso (2006).

In Appendix A.5, we show that under a boundedness condition on the drift and volatility:

$$\sup_{t\in\mathbb{Z}} \mathbb{E}[(RV_t^n - IV_t)^2] \le Cn^{-1}, \tag{48}$$

for some $C > 0$. Hence, in this setting Assumption 5 holds for RV_t^n.

3.3. Asymptotic normality

To establish asymptotic normality of our GMM estimator, for technical reasons we assume that under $\mathbb{P}_{\theta_0}$ the Gaussian process Y admits a causal moving average representation

$$Y_t = \int_{-\infty}^{t} K(t-u)\mathrm{d}B_u, \quad t \in \mathbb{R}, \tag{49}$$

for a two-sided standard Brownian motion $B = (B_t)_{t\in\mathbb{R}}$ and measurable kernel $K : (0,\infty) \to \mathbb{R}$ such that $\int_0^\infty K(u)^2\mathrm{d}u < \infty$. We can extend K to the entire real line by setting $K(u) = 0$ for $u \le 0$ when necessary. (49) is not restrictive, since a stationary Gaussian process admits such a representation under weak conditions. In particular, the moving average

structure exists if and only if Y satisfies a mild, albeit technical, condition known as *pure non-determinism*, see Karhunen (1950, Satz 5) and Dym and McKean (1976, Section 4.5). The fSV model incorporated in this paper adheres to form (49), since the fOU process has such a representation (e.g., Barndorff-Nielsen and Basse-O'Connor, 2011).

The asymptotic behavior of $K(u)$ as $u \to \infty$ governs the long-term memory of Y. To derive the asymptotic normality of our GMM estimator, we need to constrain that memory.

Assumption 6. $K(u) = O(u^{-\gamma})$ as $u \to \infty$ for some $\gamma > 1$.

Garnier and Sølna (2018) showed that the kernel in the moving average representation of the fOU process is asymptotically, as $u \to \infty$, proportional to $u^{H-3/2}$ for $H(0, 1/2) \cup (1/2, 1)$. Moreover, the definition in (13) with $H = 1/2$ implies $K(u) = \nu e^{-\lambda u} = o(u^{-\gamma})$, for all $\gamma > 1$, as $u \to \infty$. Thereby, the fSV model requires $H \leq 1/2$ to be covered by Assumption 6, allowing for rough volatility but ruling out the long-memory version.

We believe the constraint in Assumption 6 is nearly optimal in the sense that if $K(u)$ is asymptotically proportional to $u^{-\gamma}$ for $\gamma \in (0, 1)$, e.g. with the fSV model for $H > 1/2$, then asymptotic normality with a standard rate of convergence ceases to hold. In this case, we can show that the expression for the asymptotic covariance matrix in our central limit theorem of Proposition 3.6 does not converge. It is possible that a non-central limit theorem with a non-standard rate of convergence holds, as commonly encountered in the realm of long-memory processes, see, e.g., Taqqu (1975). Proving such an extension is rather non-trivial, however, and therefore beyond the scope of the present exposition.[10]

Additionally, we introduce stronger assumptions about the error process $(\varepsilon_t)_{t\in\mathbb{Z}}$. In what follows, we write $\|X\|_{L^2(\mathbb{P}_\theta)} = \mathbb{E}_\theta\big[X^2\big]^{1/2}$ for any square integrable random variable X and work with the filtrations $\mathcal{F}_t^{\widehat{\mathbb{IV}}} = \sigma\big\{\widehat{\mathbb{IV}}_t, \widehat{\mathbb{IV}}_{t-1}, \ldots\big\}$ and $\mathcal{F}_t^{B,\varepsilon} = \sigma\{\varepsilon_t, \varepsilon_{t-1}, \ldots\} \vee \sigma\{B_u : u \leq t\}$, $t \in \mathbb{Z}$.

Assumption 7. The processes B and $(\varepsilon_t)_{t\in\mathbb{Z}}$ satisfy the following conditions:

(i) $\mathbb{E}[\varepsilon_1^4] < \infty$,

(ii) $\Big\|\mathbb{E}_{\theta_0}\Big[\varepsilon_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - \mathbb{E}_{\theta_0}[\varepsilon_1^2]\Big\|_{L^2(\mathbb{P}_{\theta_0})} = O(r^{-\gamma+1/2})$ as $r \to \infty$,

(iii) B has independent increments with respect to $(\mathcal{F}_t^{B,\varepsilon})_{t\in\mathbb{Z}}$ (i.e., for any $t \in \mathbb{Z}$ the process $(B_u - B_t)_{u\geq t}$ is independent of $\mathcal{F}_t^{B,\varepsilon}$).

Condition (ii) constrains the memory in the squared measurement error. In the high-frequency setting, the measurement error usually depends on volatility (as demonstrated in Examples 3.1–3.3). So here Assumption 6 implies condition (ii), see Proposition A.6 in Appendix A.6. Condition (iii) ensures that the measurement error does not anticipate future increments of the driving Brownian motion B, which is not very restrictive.

The next result presents the CLT for the sample mean of our statistic.

Proposition 3.6. *Suppose that Assumptions 1, 2, 6 and 7 hold. Then, as $T \to \infty$, under $\mathbb{P}_{\theta_0}$,*

$$\sqrt{T}\widehat{m}_T(\theta_0) \xrightarrow{d} N\big(0, \Sigma_{\widehat{\mathbb{IV}}}\big), \tag{50}$$

where

$$\Sigma_{\widehat{\mathbb{IV}}} = \sum_{\ell=-\infty}^{\infty} \Gamma_{\widehat{\mathbb{IV}}}(\ell), \tag{51}$$

and

$$\Gamma_{\widehat{\mathbb{IV}}}(\ell) = \mathbb{E}_{\theta_0}\Big[\big(\widehat{\mathbb{IV}}_1 - G_c(\theta_0)\big)\big(\widehat{\mathbb{IV}}_{1+\ell} - G_c(\theta_0)\big)^\intercal\Big]. \tag{52}$$

The proof of Proposition 3.6 builds on Theorem 4.10 from Merlevède et al. (2019, Section 4.2), where a martingale approximation CLT is derived for an L^2-mixingale of size $-1/2$. Theorem 24.5 in Davidson (1994) states a related CLT for an L^1-mixingale of size -1. In other words, we impose a slightly stronger moment restriction to attain a better size condition. The rate improvement is important in our context, as a mixingale of size -1 does not cover the fOU with $H \in (0, 1/2)$. Moreover, L^1 integrability does not add anything here, because the process is always square integrable. That is, in our setting Merlevède et al. (2019) deliver the right foundation. Conversely, with a CLT that only requires a mixingale of size -1, we need to additionally assume that $\gamma > 3/2$ in Assumption 6. As the decay exponent of the fOU is $H - 3/2$, such a theory does not here apply for any $H > 0$.

A final assumption for the CLT of our GMM estimator is presented next. Here, we introduce the function $\mathbf{g} : \mathbb{R}^{k+2} \times \Theta \to \mathbb{R}$ via $\mathbf{g}(x, \theta) = x - G_c(\theta)$.

[10] It may be possible to extend the permissible range of H values by first-differencing the realized measure series. However, since the present estimator is consistent for any $H \in (0, 1)$ and the empirical results presented below do not contain estimates near the region $H > 1/2$, we do not pursue this extension here.

Assumption 8. It holds that

(i) θ_0 is an interior point of Θ.
(ii) $J^\intercal \mathbb{W} J$ is non-singular, where $J = \mathbb{E}_{\theta_0}\left[\nabla_\theta \mathbf{g}(\widehat{\mathbb{V}}_1, \theta_0)\right]$ and $\mathbb{W} = A^\intercal A$.
(iii) The function $\theta \mapsto \mathbf{g}(x, \theta)$ is continuously differentiable. In addition, $\mathbb{E}_{\theta_0}\left[\|\mathbf{g}(\widehat{\mathbb{V}}_1, \theta_0)\|^2\right] < \infty$ and $\mathbb{E}_{\theta_0}\left[\sup_{\theta\in\Theta} \|\nabla_\theta \mathbf{g}(\widehat{\mathbb{V}}_1, \theta)\|\right] < \infty$.

Now, we are ready to present the asymptotic distribution of $\widehat{\theta}_T$.

Theorem 3.7. *Suppose Assumptions 1–4 and 6–8 hold. As $T \to \infty$,*

$$\sqrt{T}\big(\widehat{\theta}_T - \theta_0\big) \xrightarrow{d} N\big(0, (J^\intercal \mathbb{W} J)^{-1} J^\intercal \mathbb{W} \Sigma_{\widehat{\mathbb{V}}} \mathbb{W} J (J^\intercal \mathbb{W} J)^{-1}\big). \tag{53}$$

As usual, to minimize the asymptotic variance in (53) and derive the efficient GMM estimator–for given moment conditions–we choose an optimal weight matrix as (the inverse of) a consistent estimator of $\Sigma_{\widehat{\mathbb{V}}}$. We propose a Newey and West (1987) HAC-type estimator:

$$\widehat{\Sigma}_T = \widehat{\Gamma}(0) + \sum_{\ell=1}^{T-1} w(\ell/L)\left[\widehat{\Gamma}(\ell) + \widehat{\Gamma}(\ell)^\intercal\right], \tag{54}$$

where w is a weight function, $L = o(T^{1/2})$ is the lag length, and

$$\widehat{\Gamma}(\ell) = \frac{1}{T}\sum_{t=1}^{T-l}\big(\widehat{\mathbb{V}}_t - G_c(\widehat{\theta}_T)\big)\big(\widehat{\mathbb{V}}_{t+\ell} - G_c(\widehat{\theta}_T)\big). \tag{55}$$

Following Davidson (2020), we impose a weak regularity condition on w that is fulfilled by a large class of common choices in practice, such as the Bartlett or Parzen kernel.

Assumption 9. It holds that

(i) $w(0) = 1$ and $\sup_{x\geq 0} |w(x)| < \infty$,
(ii) w is continuous at 0,
(iii) $\int_0^\infty \bar{w}(x)\mathrm{d}x < \infty$, where $\bar{w}(x) = \sup_{y\geq x} |w(y)|$.

Theorem 3.8. *Suppose Assumptions 1–4 and 6–9 hold. As $T \to \infty$,*

$$\widehat{\Sigma}_T \xrightarrow{\mathbb{P}} \Sigma_{\widehat{\mathbb{V}}}. \tag{56}$$

Setting $\mathbb{W}_T = \widehat{\Sigma}_T^{-1}$, it follows that

$$\sqrt{T}\big(\widehat{\theta}_T - \theta_0\big) \xrightarrow{d} N\Big(0, \big(J^\intercal \Sigma_{\widehat{\mathbb{V}}}^{-1} J\big)^{-1}\Big). \tag{57}$$

To make inference on the parameters we plug-in $\widehat{\Sigma}_T^{-1}$ in (57).

By standard results for quadratic forms of multivariate normal variables, the minimized objective function value times the sample size has an asymptotic chi-square distribution:

$$\mathcal{J}_{\mathrm{HS}} = T\widehat{m}_T(\widehat{\theta}_T)^\intercal \widehat{\Sigma}_T^{-1} \widehat{m}_T(\widehat{\theta}_T) \xrightarrow{d} \chi^2(k - p + 1). \tag{58}$$

where $k - p + 1$ is the number of overidentifying restrictions. This facilitates a Sargan–Hansen omnibus specification test of the model.

To finish this section, we study the CLT of our GMM estimator in the double-asymptotic setting, where $T \to \infty$ and $n \to \infty$, such that the discretization error is negligible.

Assumption 10. The processes $(IV_t)_{t\in\mathbb{Z}}$ and $(V_t^n)_{t\in\mathbb{Z},n\in\mathbb{N}}$ satisfy the following conditions:

(i) $(IV_t)_{t\in\mathbb{Z}}$ is a stationary and ergodic process under $\mathbb{P}_\theta$ for any $\theta \in \Theta$,
(ii) $\sup_{t\in\mathbb{Z}} \mathbb{E}_{\theta_0}\left[\left(\sqrt{T}(V_t^n - IV_t)\right)^2\right] \to 0$ as $T \to \infty$ and $n \to \infty$.

In this setting, we again introduce a HAC-type estimator $\widehat{\Sigma}_{n,T}$, which merely substitutes $\widehat{\mathbb{V}}_t - G_c$ with $\mathbb{V}_t^n - G$ in (55), and then we take $\mathbb{W}_{n,T} = \widehat{\Sigma}_{n,T}^{-1}$.

Theorem 3.9. *Suppose Assumptions 1, 3–4, 6, 8–10 hold. As $T \to \infty$ and $n \to \infty$,*

$$\sqrt{T}\big(\widetilde{\theta}_{n,T} - \theta_0\big) \xrightarrow{d} N\big(0, (\widetilde{J}^\intercal \Sigma_{\mathbb{IV}}^{-1} \widetilde{J})^{-1}\big), \tag{59}$$

where

$$\Sigma_{\mathbb{IV}} = \sum_{\ell=-\infty}^{\infty} \Gamma_{\mathbb{IV}}(\ell), \tag{60}$$

and

$$\Gamma_{\mathbb{IV}}(\ell) = \mathbb{E}_{\theta_0}\big[(\mathbb{IV}_1 - G(\theta_0))(\mathbb{IV}_{1+\ell} - G(\theta_0))^\intercal\big] \tag{61}$$

with $\widetilde{J} = \mathbb{E}_{\theta_0}\big[\nabla_\theta \mathbf{g}(\mathbb{IV}_1, \theta_0)\big]$.

In closing, we remark that Theorems 3.5 and 3.9 also hold for the bias-corrected estimator $\hat{\theta}_T$ provided it fulfills the additional conditions imposed by Assumption 5(ii) for consistency or Assumption 10(ii) for asymptotic normality, as $T \to \infty$ and $n \to \infty$. If that is so, the bias correction vanishes under the double-asymptotic framework. In practice, even if n is large and the bias term is small, we recommend to add the correction as a precaution, as nothing is lost by doing so.

Remark. In comparison to Fukasawa et al. (2022) (FTW), our article differs in several aspects. Firstly, FTW employ quasi-likelihood based on a Whittle approximation, while we propose GMM estimation. Secondly, FTW show consistency, but we further prove asymptotic normality of our estimator. Thirdly, many of our theoretical results apply more broadly to general log-normal SV models, whereas FTW concentrate on the fSV model. Fourthly, our approach is fully parametric, whereas FTW is a semi-parametric procedure, where the dynamic of the drift is left unspecified. Moreover, their implementation relies on the first-difference of log-realized variance to enforce mean zero. Hence, FTW estimate two parameters (volatility-of-volatility and the Hurst index), whereas we recover a four-dimensional parameter vector (including the mean and mean reversion coefficient). Fifthly, FTW also control for measurement error in the volatility proxy, but they employ a double-asymptotic CLT for log-realized variance with the length of the time interval approaching zero. In the implementation, however, they still rely on daily realized variance with 5-minute sampling frequency, thereby mixing properties of spot and integrated variance.

4. Simulation study

In the above, we developed a full-blown large sample GMM framework for estimation of the log-normal fSV model with a general Hurst index. We now review the finite sample properties of our approach. The aim is to assess the accuracy of the procedure in a realistic setup. We inspect both the infeasible setting where estimation is based on integrated variance and a feasible implementation relying on realized variance. For the latter, we gauge the performance both with and without the quarticity correction in (38).

We assume the log-price, X_t, evolves as a driftless Itô process:

$$\mathrm{d}X_t = \sigma_t \mathrm{d}W_t, \quad t \geq 0, \tag{62}$$

with initial condition $X_0 \equiv 0$. Here, σ_t is the spot volatility and W_t is a standard Brownian motion. We discretize X via an Euler scheme.

The log-variance, $Y_t = \ln(\sigma_t^2)$, is a fOU process:

$$\mathrm{d}Y_t = -\lambda(Y_t - \eta)\mathrm{d}t + \nu \mathrm{d}B_t^H, \tag{63}$$

where B_t^H is a fbM. We assume $W \perp\!\!\!\perp B^H$, so there is no leverage effect.

The SDE in (63) is solved to get a more convenient expression for Y:[11]

$$Y_t = \eta + (Y_{t-\Delta} - \eta)e^{-\lambda\Delta} + \nu \int_{t-\Delta}^{t} e^{-\lambda(t-s)} \mathrm{d}B_s^H. \tag{64}$$

The stochastic integral is approximated as $\int_{t-\Delta}^{t} e^{-\lambda(t-s)}\mathrm{d}B_s^H \simeq e^{-\lambda\Delta/2}(B_t^H - B_{t-\Delta}^H)$ meaning that increments to a discretely sampled fBm are required. These can be produced in many ways to get an exact discretization, e.g. Cholesky factorization or circulant embedding (see Asmussen and Glynn, 2007). While the former has complexity $O(x^3)$, the latter entails a markedly lower budget of $O(x\log(x))$ and is our preferred algorithm.

We draw 10,000 independent replications of this model with a path length of $T = 4{,}000$ days as a default. In each simulation, the log-variance process is started at random from its stationary distribution, $Y_0 \sim N\big(\eta, \mathrm{var}(Y_t)\big)$, where $\mathrm{var}(Y_t)$

[11] The ploy is as always to use Itô's Lemma with the integrating factor $e^{\lambda t}Y_t$. The math is a bit more involved here though, since we are dealing with a fBm, where a stochastic calculus may not exist. Nevertheless, it goes through in this particular instance, see, e.g., Cheridito et al. (2003).

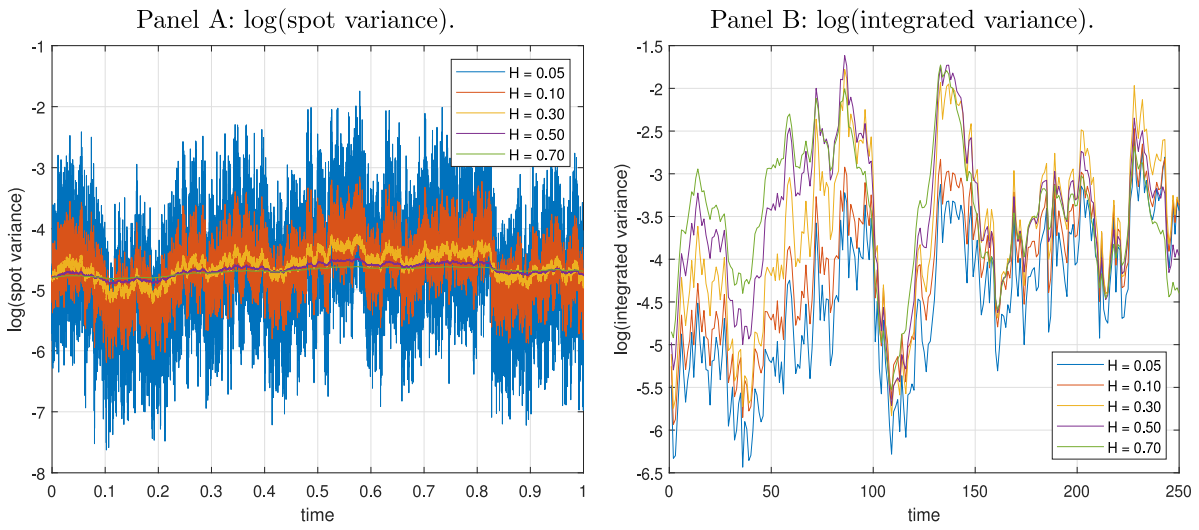

Fig. 1. Sample path of spot and integrated variance.
Note. In Panel A, we simulate a sample path of the log-spot variance for a single day as a function of H. In Panel B, we show the associated integrated variance dynamics over 250 trading days.

is given by (16). To get an almost continuous-time realization of the processes and minimize discretization bias, we partition $[t-1, t]$, for $t = 1, \dots, T$, into $N = 23{,}400$ discrete points of length $\Delta = 1/N$. In the US equity market, this roughly amounts to a 16-year sample of the stock price recorded every second in a 6.5-hour trading day.

Our procedure is inspected on several distinct sets of parameters to gauge its robustness. Throughout, we set $\eta = \ln(\xi) - 0.5\mathrm{var}(Y_t)$, where $\xi = \mathbb{E}(\sigma_t^2) = 0.0225$. This ensures the unconditional mean of the variance process is identical across settings and implies an annualized standard deviation σ_t of about 15% on average, close to the aggregate level of volatility in the empirical data analyzed in Section 5. As we are particularly attentive to estimation of the Hurst index, we choose $H = [0.05, 0.10, 0.30, 0.50, 0.70]$ as in Fukasawa et al. (2022), thus covering both the rough, standard and long-memory case. We calibrate λ and ν to minimize the distance from the model-implied autocovariance of integrated variance at lag 0, 20 and 50 to the associated sample autocovariances of realized variance of the .SPX (that is, the S&P 500) equity index, after controlling for sampling error in the noisy proxy, which are subsequently rounded off to the nearest convenient values.

The parameters are presented in Table 2. A realization of the spot and integrated variance processes from each model are plotted in Fig. 1. While the pathwise properties of volatility are notably different at a microscopic scale, they are much harder to discriminate after we integrate them up to the daily horizon.

In addition to integrated variance we also collect realized variance with $n = 78$, i.e. with 5-minute data. The advantage of this choice is that there is no concern about microstructure noise at this sampling frequency in practice. The input to the optimizer is therefore either $(IV_t)_{t=1}^T$ or $(RV_t^n)_{t=1}^T$. We restrict the description of the implementation details below to the feasible setting with realized variance.

The parameter vector is $\theta_0 = (\xi, \lambda, \nu, H)$, which we estimate via the gradient-based non-linear least squares Matlab function `lsqnonlin`. We employ the default search algorithm "trust-region-reflective" with a tolerance level of 10^{-6}.

We launch the engine at initial values determined as follows: ξ is started at the average realized variance, i.e. $\overline{RV} = T^{-1}\sum_{t=1}^T RV_t^n$. To set H and ν we exploit the auxiliary two-stage procedure proposed in Gatheral et al. (2018), which relies on the scaling law:

$$\gamma_h \equiv \mathbb{E}\big[|Y_{t+h} - Y_t|^q\big] \to K_q \nu^q |h|^{qH}, \tag{65}$$

as $h \to 0$, where $K_q = 2^{q/2}\frac{\Gamma(\frac{q+1}{2})}{\sqrt{\pi}}$ is the q'th moment of the absolute value of a standard normal random variable. This entails a log-linear relationship between γ_h and $|h|$: $\ln(\gamma_h) = \ln\left(K_q \nu^q\right) + qH\ln(|h|)$. We employ RV_t^n as a proxy for the instantaneous variance and substitute the left-hand side of (65) by the sample mean:

$$\hat{\gamma}_h = \frac{1}{T-m}\sum_{t=1}^{T-m} |\ln(RV_{t+h}^n) - \ln(RV_t^n)|^q, \tag{66}$$

for $h = 1, \dots, m$. H and ν are then estimated by OLS with $q = 2$ and $m = 6$. The results are rather robust against this configuration. At last, λ is pre-estimated such that the theoretical variance of Y_t equals the sample variance of $\ln(RV_t^n)$.

Table 2
Parameter estimation of the log-normal fSV model in simulated data.

Parameter	Value	Integrated variance		Realized variance			
					Uncorrected	Corrected estimate	
		Initial value	Estimate	Initial value	estimate	Exact	Approximate
Panel A:							
ξ	0.0225	0.0225 (0.0014)	0.0214 (0.0014)	0.0225 (0.0014)	0.0207 (0.0015)	0.0213 (0.0014)	0.0213 (0.0014)
λ	0.0050	0.0326 (0.0068)	0.0072 (0.0057)	0.0260 (0.0063)	0.0075 (0.0062)	0.0072 (0.0056)	0.0081 (0.0061)
ν	1.2500	0.4580 (0.0058)	1.3806 (0.3543)	0.5233 (0.0066)	2.6245 (0.7784)	1.3834 (0.3942)	1.0704 (0.1129)
H	0.0500	0.2661 (0.0101)	0.0460 (0.0202)	0.2208 (0.0098)	0.0168 (0.0121)	0.0466 (0.0212)	0.0646 (0.0171)
Panel B:							
ξ	0.0225	0.0225 (0.0009)	0.0216 (0.0010)	0.0225 (0.0009)	0.0211 (0.0011)	0.0216 (0.0009)	0.0216 (0.0010)
λ	0.0100	0.0381 (0.0065)	0.0117 (0.0059)	0.0286 (0.0059)	0.0084 (0.0049)	0.0117 (0.0055)	0.0124 (0.0056)
ν	0.7500	0.3683 (0.0046)	0.7767 (0.0871)	0.4390 (0.0055)	1.7914 (0.5549)	0.7750 (0.1066)	0.7275 (0.0700)
H	0.1000	0.2998 (0.0104)	0.0939 (0.0212)	0.2363 (0.0099)	0.0258 (0.0143)	0.0950 (0.0242)	0.1036 (0.0215)
Panel C:							
ξ	0.0225	0.0226 (0.0031)	0.0190 (0.0025)	0.0226 (0.0031)	0.0182 (0.0025)	0.0189 (0.0025)	0.0189 (0.0025)
λ	0.0150	0.0318 (0.0054)	0.0187 (0.0150)	0.0225 (0.0046)	0.0098 (0.0087)	0.0187 (0.0149)	0.0188 (0.0149)
ν	0.5000	0.3624 (0.0045)	0.5019 (0.0454)	0.4282 (0.0052)	0.7131 (0.1794)	0.4998 (0.0621)	0.4982 (0.0594)
H	0.3000	0.4380 (0.0108)	0.2781 (0.0490)	0.3620 (0.0108)	0.1759 (0.0530)	0.2793 (0.0589)	0.2802 (0.0583)
Panel D:							
ξ	0.0225	0.0226 (0.0037)	0.0191 (0.0031)	0.0226 (0.0037)	0.0168 (0.0028)	0.0190 (0.0031)	0.0190 (0.0031)
λ	0.0350	0.0424 (0.0053)	0.0437 (0.0264)	0.0240 (0.0040)	0.0161 (0.0163)	0.0473 (0.0357)	0.0473 (0.0357)
ν	0.3000	0.2484 (0.0032)	0.2990 (0.0236)	0.3320 (0.0039)	0.4712 (0.1947)	0.2982 (0.0332)	0.2981 (0.0331)
H	0.5000	0.5715 (0.0106)	0.4839 (0.0638)	0.4256 (0.0114)	0.2725 (0.0948)	0.4912 (0.0933)	0.4913 (0.0932)
Panel E:							
ξ	0.0225	0.0226 (0.0057)	0.0195 (0.0049)	0.0226 (0.0057)	0.0159 (0.0040)	0.0193 (0.0049)	0.0193 (0.0049)
λ	0.0700	0.0521 (0.0056)	0.0828 (0.0315)	0.0207 (0.0037)	0.0188 (0.0258)	0.0794 (0.0373)	0.0794 (0.0373)
ν	0.2000	0.1691 (0.0023)	0.2004 (0.0147)	0.2749 (0.0032)	0.3571 (0.2121)	0.2055 (0.0208)	0.2054 (0.0208)
H	0.7000	0.6777 (0.0098)	0.6803 (0.0670)	0.4254 (0.0120)	0.3079 (0.1318)	0.6580 (0.1000)	0.6580 (0.1000)

Note. We simulate 10,000 replications of a fractional Ornstein–Uhlenbeck process $dY_t = -\lambda(Y_t - \eta)dt + \nu dB_t^H$ on $[0, T]$ with $T = 4{,}000$ and a discretization step of $\Delta = 1/23{,}400$. The true model parameters $\theta_0 = (\xi, \lambda, \nu, H)$ appear in Panel A – E, where $\xi = e^{\eta + 0.5\text{var}(Y_t)}$. We estimate θ_0 with the GMM procedure developed in the main text, where the theoretical mean and autovariance (at lag 0, 1, 2, 3, 5, 20, and 50) of integrated variance is matched with the sample. The optimizer is launched with initial values from the two-stage procedure in Gatheral, Jaisson, and Rosenbaum (2018). We report the average of the initial values and the associated parameter estimates based on integrated variance (left) and realized variance (right). The latter is computed both excluding (“uncorrected”) and including (“corrected”) the bias correction in (38). The form of $c(\theta)$ is available in Table 1 and is either based on the no drift and leverage assumption (“exact”) or the CLT-based approximation (“approximate”). Standard deviation across simulations appear in parenthesis.

As shown in Table 2, the initial values display very low variation between replications, but they are often highly biased. For instance, using IV_t the starting point of H increases with the true value, but as expected it is too high on average, whereas for RV_t^n it is largely unaffected by the actual roughness of the model.

As such, there is a lot of work left to the GMM procedure. We match the sample average of RV_t^n with the mean of IV_t and the ℓ'th sample autocovariance of RV_t^n with the autocovariance structure of IV_t–available based on numerical integration of (5) together with (14)–with $\ell = [0, 1, 2, 3, 5, 20, 50]$. To motivate this choice, note that the short-term behavior of integrated variance and the impact of measurement error are described by rapid changes of autocovariances at a short time scale. Hence, we select 0, 1, 2, 3 and 5 into the set of lags. Additionally, to capture the medium- to long-term persistence of volatility, we include autocovariances at large lags. Their variation is slower, however, and it suffices to select a sparser subset, so we add lags 20 and 50.[12] In total, this yields eight moment conditions with four overidentifying restrictions.

While the above may seem arbitrary, it follows the previous literature both in terms of number of moment conditions per parameter and with its emphasis on first- and second-order moments (see, e.g., Bollerslev and Zhou, 2002; Corradi and Distaso, 2006; Todorov, 2009). The selected autocovariances–taken from an infinite set of possibilities–are also meant to maximize the efficiency of our GMM estimator, while minimizing partially redundant information in the moment conditions (e.g., Breusch et al., 1999; Hall et al., 2007). On the one hand, we perceive short-run autocovariances to be very informative about the parameters of the model. On the other hand, we expect nearby autocovariances to exhibit higher correlation. By picking a variety of mostly short lags and employing a few lags farther out as a variance reduction device, we attempt to exploit the structure of the problem as much as possible without resorting to formal econometric analysis. An alternative approach to address this problem is Carrasco and Florens (2000), who extend GMM to a continuum of moment conditions. However, the latter entails additional technical details and complicates the implementation further, and we therefore postpone it to future research.

[12] In a robustness check, and to better capture the persistence of log-variance with $H = 0.7$, we also attempted to include lag 100 and 200. However, the results did not change much. This is consistent with Andersen and Sørensen (1996), who note that estimation of SV models does not always improve by adding more information.

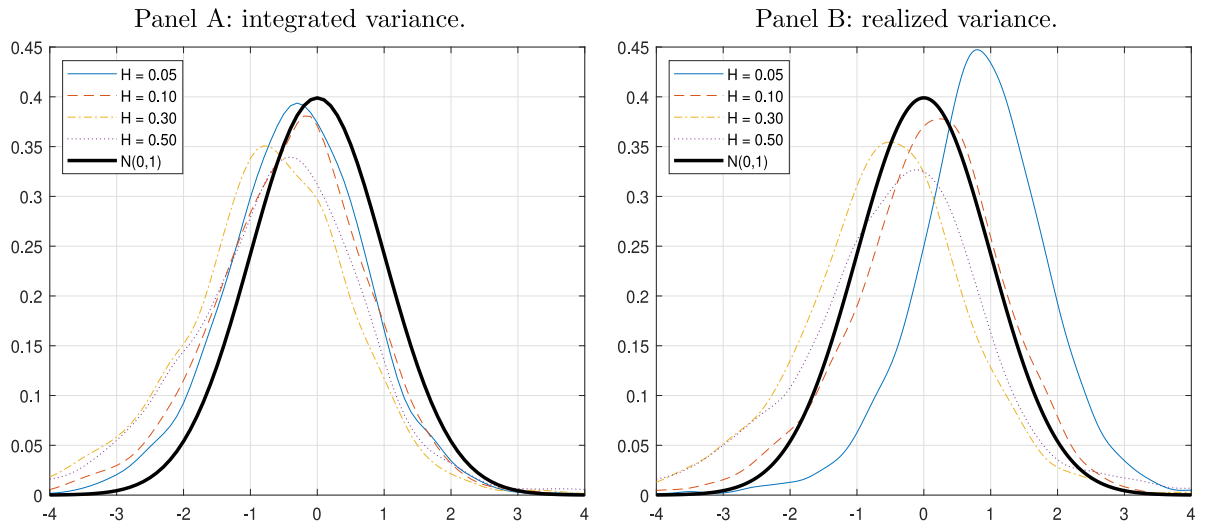

Fig. 2. Kernel density estimate of standardized H.
Note. We construct kernel smoothed densities of the statistic $(\widehat{H} - H)/\widehat{\mathrm{se}}(\widehat{H})$, where $\widehat{\mathrm{se}}(\widehat{H})$ is the estimated standard error of $\widehat{H}$. In Panel A, we show the analysis for integrated variance, whereas Panel B is for realized variance with the CLT-based bias correction. The density function of a standard normal random variable is superimposed as a reference point.

We employ iterated GMM with a maximum of three iterations. In the first step, a preliminary estimate of θ_0 is acquired by setting $\mathbb{W}_T$ equal to the identity matrix. In subsequent stages, in accordance with Theorem 3.8, the weight matrix is recalculated as in (54) with a Parzen kernel to ensure positive semi-definiteness and automatic lag selection based on Andrews (1991) using an approximating ARMA(1, 1) structure for realized variance. The benefit of iterated GMM is its invariance to the initial weighting matrix. Typically, only a single iteration is required to converge, but when volatility is really rough, an extra computation is sometimes helpful due to poor starting values.

The results are presented in Table 2. We report the mean estimate (standard error in parenthesis) both for the initial and final parameter value, where all calculations are done across replica. The left-hand side shows the outcome based on integrated variance, whereas the right-hand side is for realized variance with and without the correction in (38). As a robustness check, we gauge both the exact solution and CLT-based approximation available in Table 1. The former applies, since there is no drift nor leverage in the model.

We begin by commenting on the estimation based on integrated variance. Several interesting findings emerge. In the infeasible setting, the GMM procedure returns parameter estimates that are close to their population counterparts, thus verifying the robustness and accuracy of our approach. Across the board, the typical estimate of H tracks the true value with minor deviations. We do notice a minuscule underestimation of ξ and overestimation of λ. The drift parameters pull the acf of integrated variance in opposite directions, causing an offsetting impact on the moment matching. Nevertheless, both estimates remain within about a Monte Carlo standard error of the true value. ν is typically recovered with little bias, but for $H = 0.05$ the estimates are shifted upward and exhibit high variation. The intuition is that it is tougher to recover the volatility-of-volatility parameter with severe roughness.

Turning attention to the feasible results for realized variance, we record a significant deterioration in the estimation of H without the quarticity adjustment. As explained, realized variance is a noisy proxy for integrated variance, which induces "illusive roughness" and yields H estimates that are vastly below target, when the measurement error is unaccounted for. Also, ν increases while λ decreases to compensate for this effect. Including the bias correction rectifies this problem and leads to a huge improvement in all parameter estimates. It is assuring to see how the analysis for the exact correction aligns that integrated variance. Moreover, the distinction between the exact and approximate correction is often immaterial, but for $H \leq 0.10$ notable differences start to creep in. In particular, the latter produces smaller ν and larger H estimates. This indicates that the CLT-based correction may be somewhat inaccurate in finite samples when volatility is very erratic, which is not entirely unexpected. However, it is important to note that this amounts to *less* roughness. Overall, our simulations suggest the CLT-based approach is accurate enough even with 5-minute sampling.

To gauge the accuracy of our asymptotic theory, Fig. 2 portrays kernel smoothed densities of the statistic $(\widehat{H} - H)/\widehat{\mathrm{se}}(\widehat{H}) \xrightarrow{d} N(0, 1)$, where $\widehat{\mathrm{se}}(\widehat{H})$ is standard error estimate extracted from (54) and the Jacobi matrix provided by the optimizer. The left-hand side depicts the outcome for integrated variance, whereas the right-hand side contains the normalized realized variance estimates with the CLT-based bias correction. The exact correction is omitted, as its graphs are virtually identical to those reported for integrated variance.[13] As readily seen, the standard normal is a good

[13] $H = 0.7$ is excluded from the figure as well, because the distribution theory does not cover the long-memory model. Indeed, the approximation is markedly worse in that setting.

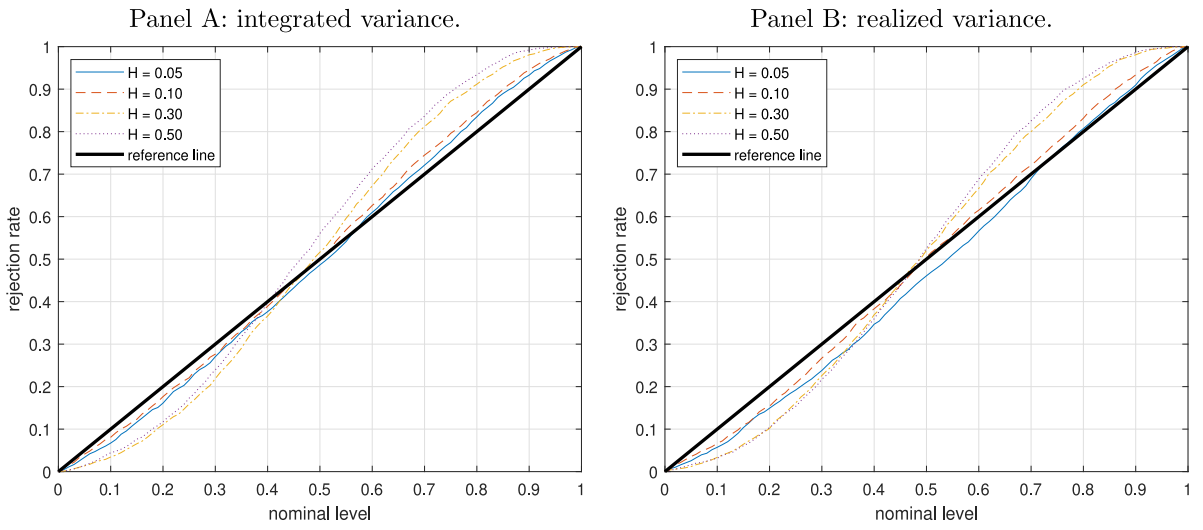

Fig. 3. Distribution of the J-test of overidentifying restrictions.
Note. We construct P-P plots to describe the variation in the J-test of overidentifying restrictions. The test statistic, $\mathcal{J}_{\text{HS}}$, is asymptotically $\chi^2(4)$-distributed under the null. In Panel A, we show the analysis for integrated variance, whereas Panel B is for realized variance with the CLT-based bias correction. The 45-degree line is superimposed as a reference point.

description of the variation in the parameter estimates of H, apart from a modest off-centering with realized variance for $H = 0.05$.

In Fig. 3, we inspect the finite sample behavior of the Sargan–Hansen J-test of overidentifying restrictions. We contrast the empirical distribution function of the test statistic, $\mathcal{J}_{\text{HS}}$, against its asymptotic distribution under the null hypothesis, which is $\chi^2(4)$. Again, we show the integrated variance in Panel A and realized variance with the CLT-based bias correction in Panel B. The impression is that the test statistic lines up fairly well with the predicted values. However, the curves are located to the right of the reference line in the lower half of diagram and vice versa in the upper half. This implies an overconcentration of mass in the center of the empirical distribution, meaning the test statistic is slightly underdispersed. This leads to a conservative test with rejection rates that are below the nominal level. At the 5% significance level, for example, the test statistic exceeds the critical value of 9.4877 about 3.6% of the times for $H = 0.05$, which drops to 0.28% for $H = 0.50$. Interestingly, the approximation is better for smaller values of H.

5. Empirical application

The log-normal fSV model is estimated from empirical high-frequency data covering a comprehensive selection of asset return series. We downloaded version 0.3 of the Oxford-Man Institute's "realized library" via: https://realized.oxford-man.ox.ac.uk/. The website tracks thirty-one leading stock indexes covering major financial markets. At the end of each trading day, the library is refreshed with information from Thomson Reuters DataScope Tick History and several nonparametric volatility estimators are calculated and appended to the database. We here employ the daily realized variance defined in (20). In line with our comments above, we decide on a 5-minute sampling frequency to suppress microstructure noise. As the trading hours of each stock exchange varies, this corresponds to n falling between 78 and 102 for most indexes, equivalent to an opening period of 6.5–8.5 h.

As it is, the database contains several data entries we suspect are erroneous. For instance, realized variance is occasionally identically equal to zero. While not impossible due to price discreteness, it is implausible for liquid securities. We therefore remove these from the sample. In addition, on February 7 2010 a realized variance corresponding to an annualized volatility of 250% is reported for the .OMXHPI equity index. We searched both Factiva and Google for relevant news articles. There is nothing immediate to suggest anything out of the extraordinary occurred in the Finnish stock market that day. However, without access to the underlying high-frequency data, it is difficult to reconcile what is causing such an abnormal deviation in realized variance. The volatility in the subsequent days are back to normal levels, so it is probably an isolated outlier. Removal of this single observation is enough to raise the first-order sample autocorrelation of the .OMXHPI realized variance series from 0.209 to 0.592.

To account for such irregularities, we further discard realized variance estimates differing more than 30 mean absolute deviations from the average realized variance calculated from 50 observations on a rolling window centered around, but excluding, the data point under investigation. This is a light filter that removes none but the most egregious data. The median number of data deleted for each index with this algorithm is two. In the US equity market (.DJI, .IXIC, .RUT and

Table 3
Parameter estimation of the log-normal fSV model in stock index data.

Code	Index	Location	Start date	Sample size	n	$\overline{RV}$	ρ_1	GMM estimate				$\mathcal{J}_{\text{HS}}$
								ξ	$\lambda \times 100$	ν	H	
.AEX	AEX	Netherlands	2000–01	4,990	102	0.028	0.743	0.024	0.026	1.800	0.032	0.726
.AORD	All Ordinaries	Australia	2000–01	4,941	72	0.011	0.575	0.008	4.873	2.345	0.035	0.128
.BFX	BEL 20	Belgium	2000–01	4,987	102	0.022	0.644	0.018	0.012	1.938	0.021	0.603
.BSESN	BSE Sensex	India	2000–01	4,852	75	0.035	0.625	0.030	0.012	2.029	0.021	0.621
.BVLG	PSI All-Share	Portugal	2012–10	1,731	96	0.011	0.630	0.010	0.027	1.893	0.015	0.764
.BVSP	Bovespa	Brazil	2000–01	4,819	84	0.038	0.718	0.032	0.103	1.852	0.025	0.847
.DJI	DJIA	USA	2000–01	4,907	78	0.027	0.677	0.021	0.010	2.008	0.021	0.773
.FCHI	CAC 40	France	2000–01	4,989	102	0.033	0.668	0.029	0.108	1.837	0.035	0.789
.FTMIB	FTSE MIB	Italy	2009–06	2,581	90	0.029	0.658	0.023	0.018	1.877	0.016	0.303
.FTSE	FTSE 100	United Kingdom	2000–01	4,933	102	0.028	0.541	0.022	0.020	2.211	0.019	0.645
.GDAXI	DAX	Germany	2000–01	4,965	102	0.041	0.702	0.037	0.021	2.009	0.027	0.922
.GSPTSE	TSX Composite	Canada	2002–05	4,314	78	0.020	0.599	0.012	0.001	2.254	0.017	0.464
.HSI	Hang Seng	Hong Kong	2000–01	4,791	78	0.024	0.677	0.017	0.001	2.102	0.008	0.322
.IBEX	IBEX 35	Spain	2000–01	4,957	102	0.035	0.657	0.033	0.172	1.894	0.030	0.878
.IXIC	Nasdaq 100	USA	2000–01	4,906	78	0.030	0.696	0.020	0.002	1.445	0.029	0.361
.KS11	KOSPI	South Korea	2000–01	4,814	72	0.030	0.763	0.024	0.003	1.643	0.027	0.661
.MXX	IPC Mexico	Mexico	2000–01	4,907	78	0.020	0.487	0.016	0.007	2.442	0.010	0.478
.N225	Nikkei 225	Japan	2000–02	4,762	72	0.025	0.683	0.023	0.183	1.847	0.029	0.564
.NSEI	Nifty 50	India	2000–01	4,849	75	0.030	0.598	0.029	0.092	2.146	0.032	0.955
.OMXC20	OMXC20	Denmark	2005–10	3,431	96	0.030	0.641	0.021	0.306	2.247	0.026	0.671
.OMXHPI	OMX Helsinki	Finland	2005–10	3,467	102	0.028	0.592	0.018	0.001	1.475	0.028	0.360
.OMXSPI	OMX Stockholm	Sweden	2005–10	3,468	102	0.024	0.562	0.015	0.001	2.432	0.012	0.487
.OSEAX	Oslo Exchange	Norway	2001–09	4,462	101	0.031	0.634	0.023	0.001	2.317	0.012	0.379
.RUT	Russel 2000	USA	2000–01	4,907	78	0.018	0.663	0.013	0.051	2.180	0.019	0.674
.SMSI	Madrid General	Spain	2005–07	3,587	101	0.031	0.648	0.029	0.218	2.008	0.030	0.591
.SPX	S&P 500	USA	2000–01	4,911	78	0.026	0.699	0.019	0.100	1.610	0.043	0.641
.SSEC	Shanghai Composite	China	2000–01	4,724	66	0.042	0.674	0.036	0.015	1.865	0.025	0.692
.SSMI	Swiss Market Index	Switzerland	2000–01	4,904	102	0.020	0.732	0.017	0.022	1.613	0.037	0.747
.STOXX50E	EURO STOXX 50	Europe	2000–01	4,989	102	0.039	0.585	0.033	0.092	2.013	0.030	0.783
Average						0.028	0.640	0.022	0.214	1.980	0.024	0.607

Note. "code" is based on the Oxford-Man Institute's naming convention. "sample size' is the number of observations for the stock index. n is the number of intraday returns at the 5-minute sampling frequency. $\overline{RV}$ is the sample average realized variance. ρ_1 is the first-order autocorrelation of realized variance. ξ is the average level of the spot variance process, λ is the speed of mean reversion (multiplied by 100), ν is the volatility-of-volatility, while H is the Hurst exponent. The cross-sectional average of each descriptive statistic is shown in the bottom row. $\mathcal{J}_{\text{HS}}$ reports the P-value of the Sargan–Hansen test of overidentifying restrictions, which is asymptotically $\chi^2(4)$-distributed.

.SPX) nothing from the financial crisis is flagged as outlying. Only June 24 2015 and August 24 2016 are deleted. Both these days correspond to infamous flash crashes that ravage volatility estimates (e.g. Christensen et al., 2022).

An overview of the remaining data is presented in Table 3. It reports the starting date of each index and the sample size. We include information up to 31 July 2019 and exclude .KSE og .STI from our investigation, as there are sizeable gaps in their data series.

The GMM estimation follows the setup from the simulation section. Since we are dealing with equity data, we can expect both drift and leverage to be present, so the approximate correction from Table 1 is employed. The right-hand side of Table 3 shows the outcome for individual stock indexes, where the bottom row presents the cross-sectional average of each parameter. Looking at the table, the results are remarkably stable across assets. The $\bar{\xi} = 0.022$ estimate corresponds to about 14.91% annualized volatility in the aggregate stock market. We do observe a slight underestimation of the mean volatility level compared to the sample average realized variance, which is consistent with the Monte Carlo analysis.

The reported Hurst exponents suggest a very rough volatility process with an average level of $\bar{H} = 0.024$. This is on par with Bayer et al. (2016) and Fukasawa et al. (2022) but slightly smaller compared to Bennedsen et al. (2022) and Gatheral et al. (2018). A possible reason for this discrepancy is that the latter employ realized variance as a proxy for spot variance. However, the former is a consistent estimator of the integrated variance, which is much smoother than instantaneous variance (see Fig. 1). This ought to bias their H estimates upwards. Our procedure does not suffer from that problem, as we directly compare realized variance to the dynamics of integrated variance in the fSV model, so the averaging "cancels out".

In Panel A of Fig. 4, we report the empirical estimates of H together with 90% confidence intervals (CIs). The latter are constructed by exponentiating the CIs associated with the log-based distribution theory, where we apply the delta rule to conclude that $\frac{\sqrt{T}(\ln(\widehat{H})-\ln(H))}{\widehat{\text{se}}(H)/\widehat{H}} \xrightarrow{d} N(0,1)$ as $T \to \infty$.[14] Here, $\widehat{\text{se}}(\widehat{H})$ is the standard error of $\widehat{H}$ extracted from the estimate of the asymptotic variance–covariance matrix in (57). Looking at the graph, we discover that H is typically estimated

[14] This approach has the advantage of enforcing non-negativity on the CI for H.

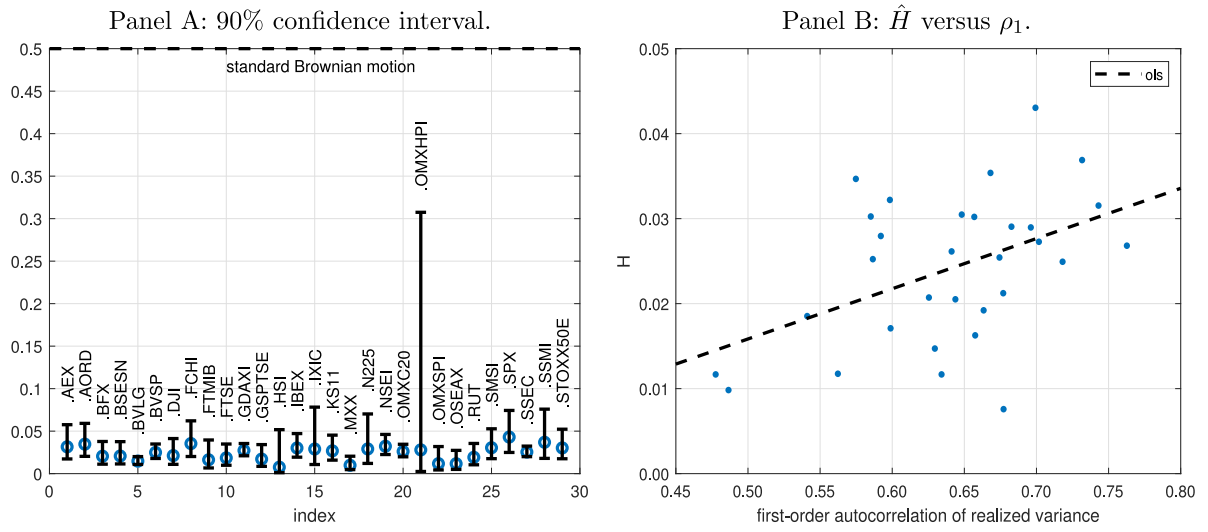

Fig. 4. Properties of the roughness estimate.
Note. In Panel A, we report 90% CIs for H. The CIs are based on the log-based distribution theory to enforce non-negativity. This also makes the CIs asymmetric. In Panel B, we show a scatter plot of $\widehat{H}$ versus ρ_1 together with the fitted regression line $\widehat{H} = \hat{a} + \hat{b}\rho_1$.

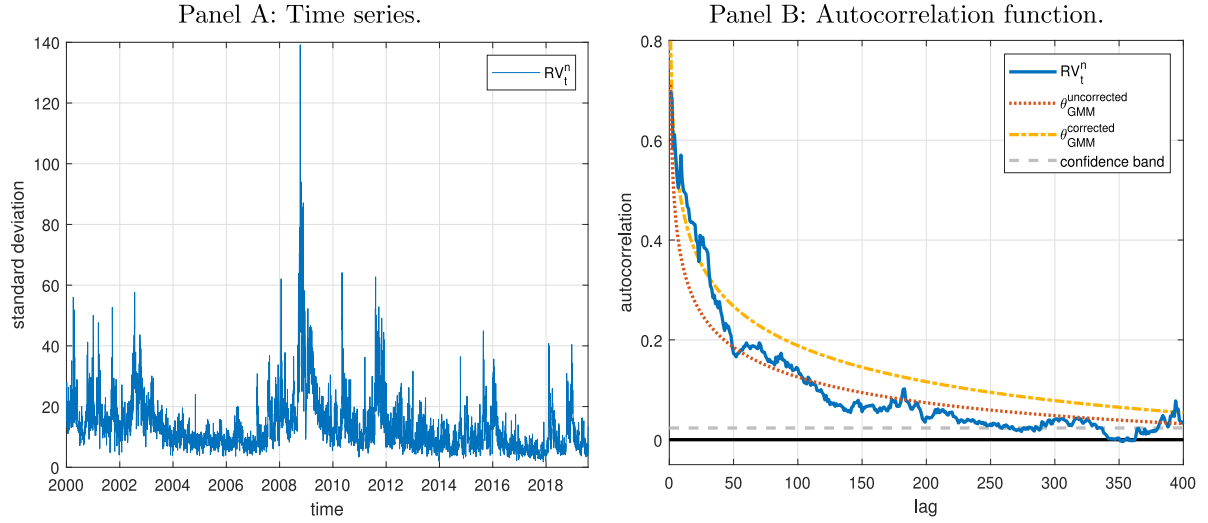

Fig. 5. Properties of .SPX realized variance.
Note. In Panel A, we plot the realized variance of .SPX converted to standard deviation per annum. In Panel B, we show the sample acf of realized variance with a 95% white noise confidence band. We compare this to the theoretical acf of the log-normal fSV model implied by the estimated parameter vector $\hat{\theta}_{GMM}$, where the latter is reported without and with the bias correction from (38).

very accurately, meaning realized variance is informative about the true level of roughness in the data. The exception is the Finnish stock market, where the range of plausible values is rather extensive. Irrespective of this, the upper bound of the CIs are far away from the level implied by a standard Brownian motion. In Panel B, we add a scatter plot of $\widehat{H}$ versus the persistence of realized variance, as measured by its first-order autocorrelation coefficient, ρ_1. This reveals a pronounced positive association between the series. A linear regression $\widehat{H} = a + b\rho_1 + \epsilon$ shows an $R^2 = 0.2120$ and a slope estimate $\hat{b} = 0.0591$ that is statistically significant with a P-value of 0.0091 (the estimated intercept $\hat{a} = -0.0137$ is highly insignificant).

To gauge the statistical fit of the model, we calculate the J-test for overidentifying restrictions. To reiterate, the test statistic $\mathcal{J}_{HS}$ has an asymptotic $\chi^2(4)$-distribution under $\mathcal{H}_0$. The results appear in the last column of Table 3. Overall, the P-values are relatively high, so the fSV process does a good job in describing the data.

As an illustration of our findings we zoom in at .SPX, which represents the S&P 500 index and is therefore related to developments in the US stock market. In Panel A of Fig. 5, we show the realized variance of .SPX (the raw estimator has

been converted to standard deviation per annum for convenience). It displays the customary volatility clustering present in most financial asset return series. In Panel B, we plot the first 400 lags of the associated acf of RV_t^n together with a Bartlett one-sided 95% white noise confidence band. It is suggestive of significant memory in integrated variance. The acf retains the impression of a hyperbolic decay, which is consistent with $H \neq 0.5$. The initial steep decline is indicative of roughness driving the serial correlation, in contrast to the much slower decay symptomatic of long-memory, where the latter entails a Hurst exponent above 0.5. As a comparison, we superimpose the model-implied acf recovered from the GMM parameter estimation, where the latter is shown both with and without the approximate bias correction in (38). The acf of the uncorrected estimator tracks the sample counterpart based on realized variance closely both at the short and long end. Meanwhile, the effect of the bias correction is to lift the acf higher, indicating a larger amount of memory in integrated variance. Note that this was to be expected, since the impact of measurement error in a time series is to attenuate the acf (e.g., Hansen and Lunde, 2014).[15]

As a robustness check, we fetched from the NYSE Trade and Quote (TAQ) database a 5-minute transaction price series for the ticker symbol SPY; an exchange-traded fund tracking the S&P 500 and equivalent to the .SPX index. The purpose is to verify the findings via our in-house data sources and to gauge their sensitivity with respect to price jumps. We pair SPY with the .SPX sample and construct realized variance, bipower variation and truncated realized variance. The latter are jump-robust estimators of integrated variance. In that ordering, the GMM procedure delivers the following H estimates: 0.033, 0.038, 0.035. As expected, bipower variation and truncated realized variance return slightly larger measures of H, but the various estimates are practically speaking identical and consistent with the .SPX estimate from Table 3. The other parameter estimates are also close. Hence, the presence and degree of roughness is not dictated by the choice of noisy proxy, nor is it caused by application of the non-robust realized variance.

In sum, our empirical results point toward a very erratic volatility process in line with–or even exceeding–previous research. As these findings are not induced by microstructure noise nor discretization error, we are bound to conclude there is roughness in variance.

6. Conclusion

We propose a GMM framework for estimation of the log-normal SV model governed by a fractional Brownian motion. Our procedure is built from the dynamic properties of integrated variance, but it employs a realized measure of volatility computed from high-frequency as a noisy proxy. We explicitly account for the inherent measurement error in the selected estimator by adjusting an appropriate moment condition. We prove consistency and asymptotic normality our estimator in a classical long-span setting. A Monte Carlo study shows our routine is capable of recovering the parameters of the model across the entire memory spectrum. We implement the approach on high-frequency data from leading equity market indexes and confirm the presence of substantial roughness in the stochastic variance process, as consistent with recent findings in the literature.

Appendix A. Proofs

A.1. Auxiliary result

To prove Theorem 2.1, we need the following auxiliary result that enables to express certain two-dimensional integrals in a one-dimensional form.

Lemma A.1. *Assume $f : [0, \infty) \to \mathbb{R}$ is a continuous function and let $k \in \mathbb{N}$. Then,*

$$\int_{k-1}^{k}\int_{0}^{1} f(|s-t|)\mathrm{d}s\mathrm{d}t = \int_{0}^{1}(1-y)\big(f(|k-1-y|)+f(k-1+y)\big)\mathrm{d}y.$$

Proof. Write

$$\int_{k-1}^{k}\int_{0}^{1} f(|s-t|)\mathrm{d}s\mathrm{d}t = \iint_{[k-1,k]\times[0,1]} f(|s-t|)\mathrm{d}s\mathrm{d}t$$

and introduce the linear (bijective) change of variables:

$$\begin{bmatrix} s \\ t \end{bmatrix} = \frac{1}{2}\begin{bmatrix} (u+v) \\ (-u+v) \end{bmatrix} = \underbrace{\frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{\equiv A}\begin{bmatrix} u \\ v \end{bmatrix} \equiv \begin{bmatrix} \varphi_1(u,v) \\ \varphi_2(u,v) \end{bmatrix} \equiv \varphi(u,v).$$

[15] We also estimated the fSV model with a driving standard Brownian motion, i.e. pre-imposing $H = 0.5$. The remaining parameter estimates were $(\hat{\xi}, \hat{\lambda}, \hat{\nu}) = (0.016, 0.101, 0.552)$, which broadly aligns with previous studies. Intuitively, to fit the sample acf of realized variance the GMM procedure has to select a larger mean-reversion parameter λ to compensate for the extra memory induced by forcing H to one-half.

Applying this to the inequalities $k-1 \le s \le k$ and $0 \le t \le 1$, we find they are equivalent to

$$v \le 2k-u, \quad v \ge 2(k-1)-u, \quad v \ge u, \quad \text{and} \quad v \le u+2.$$

Therefore, the set

$$B = \{(u,v) \in \mathbb{R}^2 : v \le 2k-u, \quad v \ge 2(k-1)-u, \quad v \ge u, \quad \text{and} \quad v \le u+2\}$$

is mapped by φ to $[k-1,k] \times [0,1]$. Note also that $B = B_1 \cup B_2$, where

$$B_1 \equiv \{(u,v) \in \mathbb{R}^2 : k-2 \le u < k-1 \quad \text{and} \quad 2(k-1)-u \le v \le u+2\}$$
$$B_2 \equiv \{(u,v) \in \mathbb{R}^2 : k-1 \le u \le k \quad \text{and} \quad u \le v \le 2k-u\}$$

are disjoint. Now, the Jacobian $(D\varphi)(u,v)$ of φ equals A for any $(u,v) \in \mathbb{R}^2$, whereby

$$|\det(D\varphi)(u,v)| = |\det(A)| = \frac{1}{2},$$

and since $s-t = \varphi_1(u,v) - \varphi_2(u,v) = \frac{1}{2}(u+v) - \frac{1}{2}(-u+v) = u$, we get by multivariate integration by substitution:

$$\begin{aligned}\iint_{[k-1,k]\times[0,1]} f(|s-t|)\mathrm{d}s\mathrm{d}t &= \iint_{\varphi(B)} f(|s-t|)\mathrm{d}s\mathrm{d}t \\ &= \iint_B f(|\varphi_1(u,v) - \varphi_2(u,v)|)|\det(D\varphi)(u,v)|\mathrm{d}u\mathrm{d}v \\ &= \frac{1}{2}\left(\iint_{B_1} f(|u|)\mathrm{d}u\mathrm{d}v + \iint_{B_2} f(|u|)\mathrm{d}u\mathrm{d}v\right).\end{aligned}$$

Firstly,

$$\begin{aligned}\iint_{B_1} f(|u|)\mathrm{d}u\mathrm{d}v = \int_{k-2}^{k-1}\left(\int_{2(k-1)-u}^{u+2} f(|u|)\mathrm{d}v\right)\mathrm{d}u &= 2\int_{k-2}^{k-1}\big(u-(k-2)\big)f(|u|)\mathrm{d}u \\ &= 2\int_0^1 (1-y)f(|k-1-y|)\mathrm{d}y,\end{aligned}$$

via the substitution $y = k-1-u$. Secondly,

$$\begin{aligned}\iint_{B_2} f(|u|)\mathrm{d}u\mathrm{d}v = \int_{k-1}^{k}\left(\int_{u}^{2k-u} f(|u|)\mathrm{d}v\right)\mathrm{d}u &= 2\int_{k-1}^{k}(k-u)f(u)\mathrm{d}u \\ &= 2\int_0^1 (1-y)f(k-1+y)\mathrm{d}y,\end{aligned}$$

by substituting $y = u-(k-1)$ and noting $u \ge k-1 \ge 0$. Thus, the asserted formula follows. ■

A.2. Proof of Theorem 2.1

To prove the first part of the theorem, we note that since the variance process $(\sigma_t^2)_{t\ge 0}$ is stationary, Fubini's theorem yields that

$$\mathbb{E}[IV_t] = \int_{t-1}^{t} \mathbb{E}\big[\sigma_s^2\big]\mathrm{d}s = \mathbb{E}\big[\sigma_0^2\big] = \xi.$$

We proceed with the second-order moments of IV_t by noting that

$$\begin{aligned}\mathbb{E}[\sigma_t^2\sigma_s^2] &= \xi^2\mathbb{E}\big[\exp\big(Y_t + Y_s - \kappa(0)\big)\big] \\ &= \xi^2\exp\big(\kappa(|t-s|)\big),\end{aligned}$$

where the last equation follows from $Y_t + Y_s \sim N(0, 2\kappa(|t-s|) + 2\kappa(0))$. We deduce that

$$\begin{aligned}\mathbb{E}[IV_1 IV_{1+\ell}] &= \int_\ell^{\ell+1}\int_0^1 \mathbb{E}\big[\sigma_s^2\sigma_t^2\big]\mathrm{d}s\mathrm{d}t \\ &= \xi^2\int_\ell^{\ell+1}\int_0^1 \exp(\kappa(|t-s|))\mathrm{d}s\mathrm{d}t \\ &= \xi^2\int_0^1 (1-y)\big[\exp(\kappa(|\ell-y|)) + \exp(\kappa(\ell+y))\big]\mathrm{d}y,\end{aligned}$$

as a consequence of Lemma A.1.

Next, we deal with the third moment of integrated variance:

$$\mathbb{E}[\sigma_t^2\sigma_s^2\sigma_u^2] = \xi^3 \exp\Big(Y_t + Y_s + Y_u - \frac{3}{2}\kappa(0)\Big),$$

where $Y_t + Y_s + Y_u$ is Gaussian with mean zero and

$$\mathrm{var}(Y_t + Y_s + Y_u) = 3\kappa(0) + 2\big(\kappa(|t-s|) + \kappa(|t-u|) + \kappa(|s-u|)\big).$$

Then, it follows that

$$\mathbb{E}[\sigma_t^2\sigma_s^2\sigma_u^2] = \xi^3 \exp\left(\kappa(|t-s|) + \kappa(|t-u|) + \kappa(|s-u|)\right),$$

from which we deduce

$$\begin{aligned}\mathbb{E}[IV_t^3] &= \xi^3 \int_0^1\int_0^1\int_0^1 \mathbb{E}[\sigma_t^2\sigma_s^2\sigma_u^2]\mathrm{d}u\mathrm{d}s\mathrm{d}t \\ &= \xi^3 \int_0^1\int_0^1\int_0^1 \exp\big(\kappa(|t-s|) + \kappa(|t-u|) + \kappa(|s-u|)\big)\mathrm{d}u\mathrm{d}s\mathrm{d}t.\end{aligned}$$

As above, we exploit the symmetric structure of the problem. We start by reducing the three-dimensional integral applying the following two-dimensional change of variables:

$$x = u - t, \quad y = s - t,$$

such that

$$\begin{aligned}\mathbb{E}[IV_t^3] &= \xi^3 \int_0^1\int_{-t}^{1-t}\int_{-t}^{1-t} f(x,y)\mathrm{d}x\mathrm{d}y\mathrm{d}t = \xi^3 \int_0^1\int_{-1}^1\int_{-1}^1 1_{\{-t\le x,y\le 1-t\}} f(x,y)\mathrm{d}x\mathrm{d}y\mathrm{d}t \\ &= \xi^3 \int_{-1}^1\int_{-1}^1 c(x,y) f(x,y)\mathrm{d}x\mathrm{d}y,\end{aligned}$$

where $f(x,y) = \exp\big(\kappa(|x-y|) + \kappa(|x|) + \kappa(|y|)\big)$ and $c(x,y) = \int_0^1 1_{\{-t\le x,y\le 1-t\}}\mathrm{d}t$. Note that $c(x,y) = c(-x,-y)$, which is seen from substituting $s = 1 - t$. Now, when $(x,y) \in [0,1]\times[0,1]$:

$$c(x,y) = 1 - \max(x,y).$$

Moreover, if $(x,y) \in [0,1]\times[-1,0]$, we conclude that $c(x,y) = \max(1-x+y, 0)$. Inserting these terms and noting that $f(x,y) = f(-x,-y) = f(y,x)$:

$$\begin{aligned}\mathbb{E}[IV_t^3] &= 2\xi^3 \int_0^1\int_0^1 (1-\max(x,y)) f(x,y)\mathrm{d}y\mathrm{d}x + 2\xi^3 \int_0^1\int_{-1}^0 \max(1-x+y, 0) f(x,y)\mathrm{d}y\mathrm{d}x \\ &= 4\xi^3 \int_0^1\int_0^x (1-x) f(x,y)\mathrm{d}y\mathrm{d}x + 2\xi^3 \int_0^1\int_{x-1}^0 (1-x+y) f(x,y)\mathrm{d}y\mathrm{d}x \\ &= 6\xi^3 \int_0^1\int_0^x (1-x) f(x,y)\mathrm{d}y\mathrm{d}x.\end{aligned}$$

The last equality holds, since both double integrals in the above expression agree in value. To show this, we substitute $z = x - y$ and reexpress $f(x,z) = f(x,x-z)$, which yields

$$\int_{x-1}^0 (1-x+y) f(x,y)\mathrm{d}y = \int_x^1 (1-z) f(x,z)\mathrm{d}z.$$

Finally, by exploiting $f(x,y) = f(y,x)$ again, we further conclude

$$\int_0^1\int_0^x (1-x) f(x,y)\mathrm{d}y\mathrm{d}x = \int_0^1\int_x^1 (1-z) f(x,z)\mathrm{d}z\mathrm{d}x.$$

To calculate the fourth moment of IV_t, we proceed as above:

$$\mathbb{E}[\sigma_t^2\sigma_s^2\sigma_u^2\sigma_v^2] = \xi^4 \exp\big(\kappa(|t-s|) + \kappa(|t-u|) + \kappa(|s-u|) + \kappa(|t-v|) + \kappa(|s-v|) + \kappa(|u-v|)\big).$$

Subsequently, to compute the four-dimensional integral we apply the change of variables:

$$x = u - t, \quad y = s - t, \quad z = v - t,$$

which, by recalling the definition of g, implies

$$\begin{aligned}\mathbb{E}[IV_t^4] &= \xi^4 \int_0^1 \int_{-t}^{1-t} \int_{-t}^{1-t} \int_{-t}^{1-t} g(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t \\ &= \xi^4 \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 d(x,y,z)g(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z,\end{aligned}$$

where $g(x,y,z) = \exp\big(\kappa(|x-y|)+\kappa(|x-z|)+\kappa(|y-z|)+\kappa(|x|)+\kappa(|y|)+\kappa(|z|)\big)$ and $d(x,y,z) = \int_0^1 1_{\{-t\le x,y,z\le 1-t\}}\mathrm{d}t$.

To compute this term, we split the cube $[-1,1]\times[-1,1]\times[-1,1]$ into eight quadrants determined by the signs of the variables. And since $g(x,y,z) = g(-x,-y,-z)$ and $d(x,y,z) = d(-x,-y,-z)$, we concentrate our efforts to four quadrants, which are handled case-by-case analogously to the derivation of the third moment.

1. $(x,y,z)\in[0,1]\times[0,1]\times[0,1]$:

Here $d(x,y,z) = 1-\max(x,y,z)$. Since g and d are invariant under permutation of its three variables, it suffices to compute the integral for $z\le y\le x$. As a result,

$$\int_0^1\int_0^1\int_0^1 d(x,y,z)g(x,y,z)\mathrm{d}z\mathrm{d}y\mathrm{d}x = 6\int_0^1\int_0^x\int_0^y (1-x)g(x,y,z)\mathrm{d}z\mathrm{d}y\mathrm{d}x.$$

2. $(x,y,z)\in[0,1]\times[0,1]\times[-1,0]$:

Assume $y\le x$. Then, $d(x,y,z) = \max(1-x+z,0)$ and

$$\begin{aligned}\int_0^1\int_0^1\int_{-1}^0 d(x,y,z)g(x,y,z)\mathrm{d}z\mathrm{d}y\mathrm{d}x &= 2\int_0^1\int_0^x\int_{-1}^0 \max(1-x+z,0)g(x,y,z)\mathrm{d}z\mathrm{d}y\mathrm{d}x \\ &= 2\int_0^1\int_0^x\int_{x-1}^0 (1-x+z)g(x,y,z)\mathrm{d}z\mathrm{d}y\mathrm{d}x.\end{aligned}$$

3. $(x,y,z)\in[0,1]\times[-1,0]\times[0,1]$:

Assume $z\le x$. Then, $d(x,y,z) = \max(1-x+y,0)$ and

$$\int_0^1\int_0^1\int_{-1}^0 d(x,y,z)g(x,y,z)\mathrm{d}y\mathrm{d}z\mathrm{d}x = 2\int_0^1\int_0^x\int_{-1}^0 \max(1-x+y,0)g(x,y,z)\mathrm{d}y\mathrm{d}z\mathrm{d}x,$$

which reduces to the integral over the second region, because $g(x,y,z) = g(x,z,y)$.

4. $(x,y,z)\in[-1,0]\times[0,1]\times[0,1]$:

Assume $z\le y$. Then, $d(x,y,z) = \max(1-y+x,0)$ and

$$\int_0^1\int_0^1\int_{-1}^0 d(x,y,z)g(x,y,z)\mathrm{d}x\mathrm{d}z\mathrm{d}y = 2\int_0^1\int_0^y\int_{-1}^0 \max(1-y+x,0)g(x,y,z)\mathrm{d}x\mathrm{d}z\mathrm{d}y,$$

which is identical to the integral over the third region due to $g(x,y,z) = g(y,x,z)$.

Next, we show that the second integral (and by implication the third and fourth) is equal to the first one. To this end, in the inner two integrals over the second region, we substitute $u = x-y$, $v = x-z$ and employ the identity $g(x,y,z) = g(x,x-z,x-y)$. This leads to

$$\int_0^x\int_{x-1}^0 (1-x+z)g(x,y,z)\mathrm{d}z\mathrm{d}y = \int_0^x\int_x^1 (1-v)g(x,u,v)\mathrm{d}v\mathrm{d}u,$$

and hence

$$\int_0^1\int_0^x\int_x^1 (1-v)g(x,u,v)\mathrm{d}v\mathrm{d}u\mathrm{d}x = \int_0^1\int_0^v\int_0^x (1-v)g(v,x,u)\mathrm{d}u\mathrm{d}x\mathrm{d}v,$$

which concludes the proof of the claimed equality of integrals. Summing up the terms, we get a factor $24 = 2\times 12$ in front of the integral.

As for the second part of Theorem 2.1, note that from condition (a) there exists $\ell_0 > 0$ such that $|\kappa(u)|\le 1$ for any $u\ge \ell_0 - 1$. Denoting $\gamma_{\ell+1,1} = \mathbb{E}[IV_tIV_{t+\ell}] - \xi^2$, we thus find that

$$\gamma_{\ell+1,1} = \xi^2\int_0^1 (1-y)\big(\exp\big(\kappa(|\ell-y|)\big) - 1 + \exp\big(\kappa(\ell+y)\big) - 1\big)\mathrm{d}y.$$

Introducing $r(x)\equiv \exp(x)-1-x$, $x\in\mathbb{R}$ allows to further write

$$\frac{\gamma_{\ell+1,1}}{\xi^2\kappa(\ell)} = \underbrace{\int_0^1 (1-y)\left(\frac{\kappa(\ell-y)}{\kappa(\ell)} + \frac{\kappa(\ell+y)}{\kappa(\ell)}\right)\mathrm{d}y}_{\equiv I_1} + \underbrace{\int_0^1 (1-y)\left(\frac{r(\kappa(\ell-y))}{\kappa(\ell)} + \frac{r(\kappa(\ell+y))}{\kappa(\ell)}\right)\mathrm{d}y}_{\equiv I_2},$$

for any $\ell\ge\ell_0$.

As $|r(x)| \leq 3x^2$, $x \in [0,1]$, it follows that

$$|I_2| \leq 3 \sup_{y\in[-1,1]} \left|\frac{\kappa(\ell+y)}{\kappa(\ell)}\right| \int_0^1 \underbrace{(1-y)(|\kappa(|\ell-y|)| + |\kappa(\ell+y)|)}_{\equiv v_\ell(y)} \mathrm{d}y,$$

where for any $y \in [0,1] : 0 \leq v_\ell(y) \leq 1$, while $v_\ell(y) \to 0$, as $\ell \to \infty$ by (a). Applying the dominated convergence theorem and condition (c) implies that:

$$\limsup_{\ell\to\infty} |I_2| \leq 3 \limsup_{\ell\to\infty} \sup_{\bar{y}\in[-1,1]} \left|\frac{\kappa(\ell+\bar{y})}{\kappa(\ell)}\right| \lim_{\ell\to\infty} \int_0^1 v_\ell(y)\mathrm{d}y = 0.$$

Finally, for $y \in [0,1]$ the integrand $u_\ell(y) \equiv (1-y)\left(\frac{\kappa(\ell-y)}{\kappa(\ell)} + \frac{\kappa(\ell+y)}{\kappa(\ell)}\right)$ in I_1 is bounded uniformly in ℓ by some constant from condition (c), while

$$\lim_{\ell\to\infty} u_\ell(y) = (1-y)(\phi(-y)+\phi(y)), \quad y \in [0,1],$$

by condition (b). Thus, by dominated convergence

$$\lim_{\ell\to\infty} I_1 = \int_0^1 (1-y)(\phi(-y)+\phi(y))\mathrm{d}y = \int_{-1}^1 (1-|y|)\phi(y)\mathrm{d}y,$$

which concludes the proof. ■

A.3. Proof of Theorem 3.4

We apply Theorem 2.1 of Hansen (1982), the sufficient conditions of which are implied by our Assumptions 1–4. It remains to verify $G_c(\theta)$ is continuous in θ, which also renders the random function $\widehat{m}_T(\theta)$ continuous in θ. Next, note that the moduli of continuity of $\widehat{m}_T(\theta)$ and $G_c(\theta)$ coincide, so $G_c(\theta)$ being continuous readily implies the so-called first moment continuity of $\widehat{m}_T(\theta)$, see Definition 2.1 in Hansen (1982).

To establish continuity of $G_c(\theta)$ in $\theta = (\xi, \phi)$, note that $c(\theta)$ is continuous by Assumption 2, whereby it suffices to prove the continuity of $G(\theta)$. The first component of $G(\theta)$ is $g_0^{(1)}(\theta) = \xi$, which is evidently continuous, while the remaining components are given in integral form in Theorem 2.1. Their continuity is then a consequence of the dominated convergence theorem, given condition (ii) of Assumption 1. ■

A.4. Proof of Theorem 3.5

We introduce the notation:

$$\widetilde{Q}_{n,T}(\theta) = \widetilde{m}_{n,T}(\theta)^\intercal \mathbb{W}_T \widetilde{m}_{n,T}(\theta),$$
$$Q(\theta) = m(\theta)^\intercal \mathbb{W} m(\theta).$$

where $m(\theta) = G(\theta_0) - G(\theta)$ and $\mathbb{W} = A^\intercal A$. The claim then follows under the conditions of Theorem 2.1 of Newey and McFadden (1994):

(i) $Q(\theta)$ is uniquely minimized at θ_0,
(ii) Θ is compact,
(iii) $\theta \to Q(\theta)$ is continuous, and
(iv) $\sup_{\theta\in\Theta} |\widetilde{Q}_{n,T}(\theta) - Q(\theta)| \xrightarrow{\mathbb{P}} 0$.

We note condition (i) is implied by Assumption 4, since for $\theta \neq \theta_0$:

$$Q(\theta) = (Am(\theta))'Am(\theta) > 0 = Q(\theta_0).$$

Condition (ii) is immediate. We already showed condition (iii) in the proof of Theorem 3.4. Now, we pass to the last condition (iv). In view of the Cauchy–Schwarz inequality,

$$\begin{aligned}
\left|\widetilde{Q}_{n,T}(\theta) - Q(\theta)\right| &\leq \left|(\widetilde{m}_{n,T}(\theta) - m(\theta))^\intercal \mathbb{W}_T (\widetilde{m}_{n,T}(\theta) - m(\theta))\right| + \left|m(\theta)^\intercal (\mathbb{W}_T + \mathbb{W}_T^\intercal)(\widetilde{m}_{n,T}(\theta) - m(\theta))\right| \\
&\quad + \left|m(\theta)^\intercal (\mathbb{W}_T - \mathbb{W}) m(\theta)\right| \\
&\leq \|\widetilde{m}_{n,T}(\theta) - m(\theta)\|^2 \|\mathbb{W}_T\| + 2\|m(\theta)\| \|\mathbb{W}_T\| \|\widetilde{m}_{n,T}(\theta) - m(\theta)\| \\
&\quad + \|m(\theta)\|^2 \|\mathbb{W}_T - \mathbb{W}\|.
\end{aligned}$$

Then, in view of Assumption 3, it suffices to prove that

$$\sup_{\theta\in\Theta} \|\widetilde{m}_{n,T} - m(\theta)\| \xrightarrow{\mathbb{P}} 0, \quad \text{as } T \to \infty \text{ and } n \to \infty.$$

Let $m_T(\theta) = T^{-1}\sum_{t=1}^{T}\big[\mathbb{V}_t - G(\theta)\big]$. Since the convergence $\sup_{\theta\in\Theta}\|m_T(\theta) - m(\theta)\| \xrightarrow{\mathbb{P}} 0$ was already covered by the proof of Theorem 3.4 (setting $\varepsilon_t = c(\theta) = 0$), it remains to show

$$\sup_{\theta\in\Theta}\|\widetilde{m}_{n,T}(\theta) - m_T(\theta)\| \xrightarrow{\mathbb{P}} 0, \quad \text{as } T\to\infty \text{ and } n\to\infty.$$

To this end, we observe that

$$\begin{aligned}
\|\widetilde{m}_{n,T}(\theta) - m_T(\theta)\| &\le \frac{1}{T}\sum_{t=1}^{T}\|\mathbb{V}_t^n - \mathbb{V}_t\| \\
&\le \frac{1}{T}\sum_{t=1}^{T}\left[|V_t^n - IV_t| + \sum_{j=0}^{k}|V_t^nV_{t-j}^n - IV_tIV_{t-j}|\right] \\
&\le \frac{1}{T}\sum_{t=1}^{T}|V_t^n - IV_t|(1 + |V_t^n| + IV_t) \\
&+ \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{k}|V_t^n - IV_t||V_{t-j}^n| + IV_t|V_{t-j}^n - IV_{t-j}|.
\end{aligned}$$

From Assumption 5 and the Cauchy–Schwarz inequality, we deduce that:

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\|\widetilde{m}_{n,T}(\theta) - m_T(\theta)\|\right] \to 0, \quad \text{as } T\to\infty \text{ and } n\to\infty,$$

which was to be shown. ■

A.5. Verifying Assumption 5 for realized variance

Suppose that $\sup_{s\in\mathbb{R}}\mathbb{E}[\mu_s^4] + \sup_{s\in\mathbb{R}}\mathbb{E}[\sigma_s^4] < \infty$. Then, there exists a constant C such that

$$\sup_{t\in\mathbb{Z}}\mathbb{E}\big[(RV_t^n - IV_t)^2\big] \le \frac{C}{n}.$$

To see this, we apply Itô's Lemma to get

$$\big(X_{t-1+\frac{i}{n}} - X_{t-1+\frac{i-1}{n}}\big)^2 = 2\int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}}(X_s - X_{t-1+\frac{i-1}{n}})\mathrm{d}X_s + \int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}}\sigma_s^2\mathrm{d}s.$$

Consequently,

$$\begin{aligned}
RV_t^n - IV_t &= 2\sum_{i=1}^{n}\int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}}(X_s - X_{t-1+\frac{i-1}{n}})\mathrm{d}X_s \\
&= 2\sum_{i=1}^{n}\int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}}(X_s - X_{t-1+\frac{i-1}{n}})\mu_s\mathrm{d}s + 2\sum_{i=1}^{n}\int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}}(X_s - X_{t-1+\frac{i-1}{n}})\sigma_s\mathrm{d}W_s.
\end{aligned}$$

In turn, this combined with Cauchy–Schwarz and Jensen's inequality leads to

$$\begin{aligned}
\mathbb{E}\big[(RV_t - IV_t)^2\big] &\le 4\sum_{i=1}^{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}}\mathbb{E}[(X_s - X_{\frac{i-1}{n}})^2\mu_s^2]\mathrm{d}s + 4\sum_{i=1}^{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}}\mathbb{E}[(X_s - X_{\frac{i-1}{n}})^2\sigma_s^2]\mathrm{d}s \\
&\le \frac{C}{n},
\end{aligned}$$

where in the last inequality we exploited $\sup_{s\ge0}\mathbb{E}[\mu_s^4] + \sup_{s\ge0}\mathbb{E}[\sigma_s^4] < \infty$ along with Burkholder's inequality: $\sup_{s\in\left[\frac{i-1}{n},\frac{i}{n}\right]}\mathbb{E}\big[(X_s - X_{\frac{i-1}{n}})^4\big] \le \frac{C}{n^2}$. ■

A.6. Proof of Proposition 3.6

The proof relies on a martingale approximation central limit theorem for stationary and ergodic processes (Merlevède et al., 2019, Section 4.2), and it requires some preparation. First, we state and prove a couple of generic, elementary lemmas.

Lemma A.2. *Suppose X is a random variable such that $\mathbb{E}[X^2] < \infty$ and let $\mathcal{F}$ and $\mathcal{G}$ be σ-algebras such that $\mathcal{F} \subset \mathcal{G}$. Then,*

$$\|\mathbb{E}[X \mid \mathcal{F}]\|_{L^2(\mathbb{P})} \leq \|\mathbb{E}[X \mid \mathcal{G}]\|_{L^2(\mathbb{P})}.$$

Proof. Since $\mathcal{F} \subset \mathcal{G}$, we get by the tower property of conditional expectations,

$$\|\mathbb{E}[X \mid \mathcal{F}]\|^2_{L^2(\mathbb{P})} = \mathbb{E}\big[\mathbb{E}[X \mid \mathcal{F}]^2\big] = \mathbb{E}\big[\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{F}]^2\big].$$

Applying Jensen's inequality for conditional expectations,

$$\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{F}\big]^2 \leq \mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}]^2 \mid \mathcal{F}\big].$$

Hence,

$$\mathbb{E}\big[\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]|\mathcal{F}]^2\big] \leq \mathbb{E}\big[\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]^2|\mathcal{F}]\big] = \mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}]^2\big] = \|\mathbb{E}[X \mid \mathcal{G}]\|^2_{L^2(\mathbb{P})}. \quad \square$$

Lemma A.3. *Suppose that $X \sim N(\mu, \lambda^2)$ for some $\mu \in \mathbb{R}$ and $\lambda > 0$. Then,*

$$\mathbb{E}\big[(e^X - 1)^2\big] \leq \big(e^{\mu+\lambda^2} + 1\big)^2\big(8|\mu| + 6\lambda^2\big).$$

Proof. Note that

$$\mathbb{E}\big[(e^X - 1)^2\big] = e^{2(\mu+\lambda^2)} - 2e^{\mu+\frac{1}{2}\lambda^2} + 1 \leq e^{2(\mu+\lambda^2)} + 2e^{\mu+\lambda^2} + 1 = (e^{\mu+\lambda^2} + 1)^2,$$

while

$$e^{2(\mu+\lambda^2)} - 2e^{\mu+\frac{1}{2}\lambda^2} + 1 = e^{2(\mu+\lambda^2)} - 1 + 2(1 - e^{\mu+\frac{1}{2}\lambda^2}) \leq 8|\mu| + 6\lambda^2 \leq \underbrace{(e^{\mu+\lambda^2} + 1)^2}_{\geq 1}(8|\mu| + 6\lambda^2),$$

for $|\mu|+\lambda^2 < \frac{1}{2}$ due to the elementary inequality $|e^x - 1| \leq 2|x|$, for $|x| \leq 1$. However, if $|\mu|+\lambda^2 \geq \frac{1}{2}$, then $8|\mu|+6\lambda^2 \geq 1$, so the inequality holds also unconditionally. ■

Secondly, for convenience we formulate here a multivariate version of the martingale approximation central limit theorem, as a corollary of the results from Section 4.2 in Merlevède et al. (2019). However, we state the result using the concept of a mixingale.

Definition 1. A square-integrable (possibly multivariate) stochastic process $(\zeta_t)_{t\in\mathbb{Z}}$, which is adapted to a filtration $(\mathcal{G}_t)_{t\in\mathbb{Z}}$, is called an *$L^2$-mixingale* of size $-\varphi_0 \leq 0$ with respect to $(\mathcal{G}_t)_{t\in\mathbb{Z}}$ if there exists $\varphi > \varphi_0$ such that for any $t \in \mathbb{Z}$,

$$\|\mathbb{E}[\zeta_{t+r} \mid \mathcal{G}_t] - \mathbb{E}[\zeta_t]\|_{L^2(\mathbb{P})} = O(r^{-\varphi}), \quad \text{as } r \to \infty. \tag{67}$$

Notably, if $(\zeta_t)_{t\in\mathbb{Z}}$ is stationary and $(\mathcal{G}_t)_{t\in\mathbb{Z}}$ is the natural filtration, i.e. $\mathcal{G}_t = \mathcal{F}^\zeta_t \equiv \sigma\{\zeta_t, \zeta_{t-1}, \ldots\}$ for any $t \in \mathbb{Z}$, then condition (67) reduces to

$$\|\mathbb{E}[\zeta_r \mid \mathcal{G}_0] - \mathbb{E}[\zeta_0]\|_{L^2(\mathbb{P})} = O(r^{-\varphi}), \quad \text{as } r \to \infty. \tag{68}$$

The next lemma states the relevant mixingale CLT.

Lemma A.4. *Suppose $(\zeta_t)_{t\in\mathbb{Z}}$ is a multivariate, stationary and ergodic process, which is additionally an L^2-mixingale of size $-\frac{1}{2}$ in reference to $(\mathcal{F}^\zeta_t)_{t\in\mathbb{Z}}$. Then, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\big(\zeta_t - \mathbb{E}[\zeta_0]\big) \xrightarrow{d} N(0, \Sigma_\zeta), \tag{69}$$

where $\Sigma_\zeta \equiv \sum_{\ell=-\infty}^{\infty} \mathbb{E}\big[(\zeta_1 - \mathbb{E}[\zeta_0])(\zeta_{1+\ell} - \mathbb{E}[\zeta_0])^\intercal\big]$.

Proof. By the Cramér–Wold device, we can reduce the multivariate convergence (69) to a univariate one by linear projection. To this end, assume that $(\zeta_t)_{t\in\mathbb{Z}}$ is d-dimensional for some $d \in \mathbb{N}$ and let $a \in \mathbb{R}^d$ be arbitrary. Defining $\xi_t \equiv a^\intercal(\zeta_t - \mathbb{E}[\zeta_0])$, $t \in \mathbb{Z}$, we deduce that $\mathcal{F}^\xi_t = \sigma\{\xi_t, \xi_{t-1}, \ldots\} \subset \mathcal{F}^\zeta_t$ for any $t \in \mathbb{Z}$, whereby it follows that

$$\big\|\mathbb{E}[\xi_r \mid \mathcal{F}^\xi_0]\big\|_{L^2(\mathbb{P})} \leq \|a\|_{\mathbb{R}^d}\big\|\mathbb{E}[\zeta_r \mid \mathcal{F}^\zeta_0] - \mathbb{E}[\zeta_0]\big\|_{L^2(\mathbb{P})} = O(r^{-\varphi}),$$

as $r \to \infty$ for some $\varphi > 1/2$. Thus,

$$\sum_{r=1}^{\infty} r^{-1/2}\big\|\mathbb{E}[\xi_r \mid \mathcal{F}^\xi_0]\big\|_{L^2(\mathbb{P})} < \infty,$$

so that, by Theorem 4.10 and Remark 4.13 of Merlevède et al. (2019),

$$a^\intercal \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\zeta_t - \mathbb{E}[\zeta_0]) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_t \xrightarrow{d} N\left(0, \sum_{\ell=-\infty}^{\infty} \mathbb{E}[\xi_1 \xi_{1+\ell}]\right)$$

as $T \to \infty$. It remains to note that

$$\sum_{\ell=-\infty}^{\infty} \mathbb{E}[\xi_1 \xi_{1+\ell}] = \sum_{\ell=-\infty}^{\infty} \mathbb{E}[a^\intercal(\zeta_1 - \mathbb{E}[\zeta_0])(\zeta_{1+\ell} - \mathbb{E}[\zeta_0])^\intercal a] = a^\intercal \Sigma_\zeta a,$$

and the claim follows. ■

To prove Proposition 3.6, building on Lemma A.4, we still need the following technical lemma that estimates the memory of integrated variance.

Lemma A.5. *Suppose that Assumptions 1 and 6 hold. Moreover, suppose that $(\mathcal{F}_t)_{t\in\mathbb{R}}$ is a filtration such that B is adapted and has independent increments with respect to it (cf. condition (iii) in Assumption 7). Then, for any $p > 0$, $0 \le s \le t$ and $0 \le s' \le t'$,*

(i) $\left\| \mathbb{E}_{\theta_0}\left[\int_{s+r}^{t+r} \sigma_u^p \mathrm{d}u \mid \mathcal{F}_0\right] - \mathbb{E}_{\theta_0}\left[\int_s^t \sigma_u^p \mathrm{d}u\right] \right\|_{L^2(\mathbb{P}_{\theta_0})} = O(r^{-\gamma+1/2})$,

(ii) $\left\| \mathbb{E}_{\theta_0}\left[\int_{s+r}^{t+r} \sigma_u^p \mathrm{d}u \int_{s'+r}^{t'+r} \sigma_{u'}^p \mathrm{d}u' \mid \mathcal{F}_0\right] - \mathbb{E}_{\theta_0}\left[\int_s^t \sigma_u^p \mathrm{d}u \int_{s'}^{t'} \sigma_{u'}^p \mathrm{d}u'\right] \right\|_{L^2(\mathbb{P}_{\theta_0})} = O(r^{-\gamma+1/2})$,

as $r \to \infty$.

Proof. We only prove (ii) as the proof of (i) is analogous. In explicit terms, Assumption 6 says that there exist constants $u_0 \ge 0$ and $c > 0$ such that

$$|K(u)| \le c u^{-\gamma}, \quad u \ge u_0. \tag{70}$$

Without loss of generality, assume $r \ge u_0$ from now on. By Tonelli's theorem,

$$\begin{aligned} &\mathbb{E}_{\theta_0}\left[\int_{s+r}^{t+r} \sigma_u^p \mathrm{d}u \int_{s'+r}^{t'+r} \sigma_{u'}^p \mathrm{d}u' \mid \mathcal{F}_0\right] - \mathbb{E}_{\theta_0}\left[\int_s^t \sigma_u^p \mathrm{d}u \int_{s'}^{t'} \sigma_{u'}^p \mathrm{d}u'\right] \\ &\qquad = \int_s^t \int_{s'}^{t'} \left(\mathbb{E}_{\theta_0}[\sigma_{u+r}^p \sigma_{u'+r}^p \mid \mathcal{F}_0] - \mathbb{E}_{\theta_0}[\sigma_u^p \sigma_{u'}^p]\right) \mathrm{d}u \mathrm{d}u', \end{aligned} \tag{71}$$

where

$$\sigma_v^p \sigma_{v'}^p = \xi^p e^{-\frac{p}{2}\kappa(0)} \exp\left(\int_{-\infty}^{\infty} K^+(v, v', \tau) \mathrm{d}B_\tau\right)$$

with $K^+(v, v', \tau) = \frac{p}{2}\big[K(v-\tau) + K(v'-\tau)\big]$ for any $v, v' \ge 0$ (recall we set $K(v) = 0$ for any $v \le 0$). Subsequently,

$$\mathbb{E}_{\theta_0}[\sigma_u^p \sigma_{u'}^p] = \xi^p e^{-\frac{p}{2}\kappa(0)} \exp\left(\frac{1}{2}\int_{-\infty}^{\infty} K^+(u, u', \tau)^2 \mathrm{d}\tau\right),$$

while, by the assumed properties of the Brownian motion B,

$$\begin{aligned} &\mathbb{E}_{\theta_0}[\sigma_{u+r}^p \sigma_{u'+r}^p \mid \mathcal{F}_0] \\ &\quad = \xi^p e^{-\frac{p}{2}\kappa(0)} \exp\left(\int_{-\infty}^{0} K^+(u+r, u'+r, \tau) \mathrm{d}B_\tau\right) \mathbb{E}\left[\exp\left(\int_0^{\infty} K^+(u+r, u'+r, \tau) \mathrm{d}B_\tau\right)\right] \\ &\quad = \xi^p e^{-\frac{p}{2}\kappa(0)} \exp\left(\int_{-\infty}^{0} K^+(u, u', \tau - r) \mathrm{d}B_\tau + \frac{1}{2}\int_0^{\infty} K^+(u, u', \tau - r)^2 \mathrm{d}\tau\right), \end{aligned}$$

using the property $K^+(u+r, u'+r, \tau) = K^+(u, u', \tau - r)$.

Therefore,

$$\mathbb{E}_{\theta_0}[\sigma_{u+r}^p \sigma_{u'+r}^p \mid \mathcal{F}_0] - \mathbb{E}_{\theta_0}[\sigma_u^p \sigma_{u'}^p] = \xi^2 e^{-\frac{p}{2}\kappa(0)} \exp\left(\frac{1}{2}\int_{-\infty}^{\infty} K^+(u, u', \tau)^2 \mathrm{d}\tau\right)\left(\exp(Y_r^{u,u'}) - 1\right), \tag{72}$$

with

$$Y_r^{u,u'} = \int_{-\infty}^{0} K^+(u, u', \tau - r) \mathrm{d}B_\tau - \frac{1}{2}\overline{K}^{u,u'}(r) \sim N\left(-\frac{1}{2}\overline{K}^{u,u'}(r),\ \overline{K}^{u,u'}(r)\right) \tag{73}$$

and $\overline{K}^{u,u'}(r) = \int_{-\infty}^{-r} K^+(u,u',\tau)^2 \mathrm{d}\tau = \int_{-\infty}^{0} K^+(u,u',\tau-r)^2 \mathrm{d}\tau$. Applying Tonelli's theorem and Jensen's inequality to (71), we conclude that:

$$\mathbb{E}_{\theta_0}\left[\left(\mathbb{E}_{\theta_0}\left[\int_{s+r}^{t+r}\sigma_u^p \mathrm{d}u \int_{s'+r}^{t'+r}\sigma_{u'}^p \mathrm{d}u' \mid \mathcal{F}_0\right] - \mathbb{E}_{\theta_0}\left[\int_s^t \sigma_u^p \mathrm{d}u \int_{s'}^{t'}\sigma_{u'}^p \mathrm{d}u'\right]\right)^2\right]$$
$$\leq (t-s)(t'-s')\int_s^t\int_{s'}^{t'} \mathbb{E}_{\theta_0}\left[(\mathbb{E}_{\theta_0}[\sigma_{u+r}^p\sigma_{u'+r}^p \mid \mathcal{F}_0] - \mathbb{E}_{\theta_0}[\sigma_u^p\sigma_{u'}^p])^2\right]\mathrm{d}u\mathrm{d}u',$$

where, by (72)–(73) and Lemma A.3,

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left[(\mathbb{E}_{\theta_0}[\sigma_{u+r}^p\sigma_{u'+r}^p \mid \mathcal{F}_0] - \mathbb{E}_{\theta_0}[\sigma_u^p\sigma_{u'}^p])^2\right] \\
&= \xi^{2p}e^{-p\kappa(0)}\exp\left(\int_{-\infty}^{\infty} K^+(u,u',\tau)^2\mathrm{d}\tau\right)\mathbb{E}_{\theta_0}\left[(\exp(Y_r^{u,u'})-1)^2\right] \\
&\leq 14\xi^{2p}e^{-p\kappa(0)}\exp\left(\int_{-\infty}^{\infty} K^+(u,u',\tau)^2\mathrm{d}\tau\right)\left(e^{\frac{3}{2}\overline{K}^{u,u'}(r)}+1\right)^2\overline{K}^{u,u'}(r) \\
&\leq 14\xi^{2p}e^{\left(\frac{p^2}{2}-p\right)\kappa(0)}\left(e^{\frac{3p^2}{4}\kappa(0)}+1\right)^2\overline{K}^{u,u'}(r),
\end{aligned}$$

after observing that

$$\overline{K}^{u,u'}(r) \leq \int_{-\infty}^{\infty} K^+(u,u',\tau)^2\mathrm{d}\tau \leq \frac{p^2}{2}\int_0^{\infty} K(\tau)^2\mathrm{d}\tau = \frac{p^2}{2}\kappa(0).$$

Moreover, if $\tau \leq -r$ then $-\tau \geq r \geq u_0$, whereby $u-\tau \geq u_0$ and $u'-\tau \geq u_0$ since $u \geq s \geq 0$ and $u' \geq s' \geq 0$. Thus, by (70),

$$\begin{aligned}
\overline{K}^{u,u'}(r) &= \frac{p^2}{4}\int_{-\infty}^{-r}\left(K(u-\tau)+K(u'-\tau)\right)^2\mathrm{d}\tau \\
&\leq \frac{c^2p^2}{2}\int_{-\infty}^{-r}\left((u-\tau)^{-2\gamma}+(u'-\tau)^{-2\gamma}\right)\mathrm{d}\tau \\
&\leq c^2p^2\int_r^{\infty}\tau^{-2\gamma}\mathrm{d}\tau = \frac{c^2p^2}{1-2\gamma}r^{-2\gamma+1}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\left\|\mathbb{E}_{\theta_0}\left[\int_{s+r}^{t+r}\sigma_u^p\mathrm{d}u\int_{s'+r}^{t'+r}\sigma_{u'}^p\mathrm{d}u' \mid \mathcal{F}_0\right] - \mathbb{E}_{\theta_0}\left[\int_s^t\sigma_u^p\mathrm{d}u\int_{s'}^{t'}\sigma_{u'}^p\mathrm{d}u'\right]\right\|_{L^2(\mathbb{P}_{\theta_0})} \\
&\qquad \leq (t-s)(t'-s')\left(\frac{14}{1-2\gamma}\right)^{1/2}\xi^p e^{\left(\frac{p^2}{4}-\frac{p}{2}\right)\kappa(0)}\left(e^{\frac{3p^2}{4}\kappa(0)}+1\right)cpr^{-\gamma+1/2} \\
&\qquad = O(r^{-\gamma+1/2}),
\end{aligned}$$

as $r \to \infty$, which concludes the proof of (ii). ■

Proof of Proposition 3.6. Thanks to Lemma A.4, it suffices to show that the stationary process $(\widehat{\mathbb{V}}_t)_{t\in\mathbb{Z}}$ is an L^2-mixingale of size $-\frac{1}{2}$ with respect to $(\mathcal{F}_t^{\widehat{\mathbb{V}}})_{t\in\mathbb{Z}}$.

Let $r \geq 1$. First, we consider:

$$\mathbb{E}_{\theta_0}\left[\widehat{IV}_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}}\right] - g_0^{(1)}(\theta_0) = \mathbb{E}_{\theta_0}\left[IV_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}}\right] - \mathbb{E}_{\theta_0}[IV_1] + \mathbb{E}_{\theta_0}\left[\varepsilon_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}}\right],$$

where

$$\mathbb{E}_{\theta_0}\left[\varepsilon_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}}\right] = \mathbb{E}_{\theta_0}\left[\mathbb{E}_{\theta_0}[\varepsilon_r \mid \mathcal{F}_{r-1}^{\sigma,\varepsilon}] \mid \mathcal{F}_0^{\widehat{\mathbb{V}}}\right] = 0$$

by Assumption 2 and the tower property of conditional expectations, which is applicable since $\mathcal{F}_0^{\widehat{\mathbb{V}}} \subset \mathcal{F}_{r-1}^{\sigma,\varepsilon}$. Therefore,

$$\begin{aligned}
\left\|\mathbb{E}_{\theta_0}\left[\widehat{IV}_r\middle|\mathcal{F}_0^{\widehat{\mathbb{V}}}\right] - g_0^{(1)}(\theta_0)\right\|_{L^2(\mathbb{P}_{\theta_0})} &= \left\|\mathbb{E}_{\theta_0}\left[IV_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}}\right] - \mathbb{E}_{\theta_0}[IV_1]\right\|_{L^2(\mathbb{P}_{\theta_0})} \\
&\leq \left\|\mathbb{E}_{\theta_0}\left[IV_r \mid \mathcal{F}_0^{W,\varepsilon}\right] - \mathbb{E}_{\theta_0}[IV_1]\right\|_{L^2(\mathbb{P}_{\theta_0})} \\
&= O(r^{-\gamma+1/2}), \quad r \to \infty,
\end{aligned} \tag{74}$$

which follows by Lemma A.2, again since $\mathcal{F}_0^{\widehat{\mathbb{V}}} \subset \mathcal{F}_0^{W,\varepsilon}$, and Lemma A.5(i).

Secondly,

$$\mathbb{E}_{\theta_0}\Big[\widehat{IV}_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - g_0^{(2)}(\theta_0) - c(\theta_0) = \mathbb{E}_{\theta_0}\Big[IV_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - \mathbb{E}_{\theta_0}[IV_1^2] + 2\mathbb{E}_{\theta_0}\Big[\varepsilon_r IV_r \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] + \mathbb{E}_{\theta_0}\Big[\varepsilon_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - \mathbb{E}_{\theta_0}[\varepsilon_1^2],$$

where

$$\mathbb{E}_{\theta_0}\Big[\varepsilon_r IV_r \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] = \mathbb{E}_{\theta_0}\Big[\mathbb{E}_{\theta_0}\big[\varepsilon_r \mid \mathcal{F}_{r-1}^{\sigma,\varepsilon}\big] IV_r \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] = 0,$$

by the tower property, since $\mathcal{F}_0^{\widehat{\mathbb{IV}}} \subset \mathcal{F}_{r-1}^{\sigma,\varepsilon}$, and Assumption 2. By virtue of condition 7 in Assumption 7 and Minkowski's inequality:

$$\begin{aligned}\Big\|\mathbb{E}_{\theta_0}\Big[\widehat{IV}_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - g_0^{(2)}(\theta_0) - c(\theta_0)\Big\|_{L^2(\mathbb{P}_{\theta_0})} &= \big\|\mathbb{E}_{\theta_0}\Big[IV_1^2 \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - \mathbb{E}_{\theta_0}[IV_1^2]\big\|_{L^2(\mathbb{P}_{\theta_0})} + O(r^{-\gamma+1/2}) \\ &\le \big\|\mathbb{E}_{\theta_0}\Big[IV_r^2 \mid \mathcal{F}_0^{W,\varepsilon}\Big] - \mathbb{E}_{\theta_0}[IV_1^2]\big\|_{L^2(\mathbb{P}_{\theta_0})} + O(r^{-\gamma+1/2}) \\ &= O(r^{-\gamma+1/2}),\end{aligned} \tag{75}$$

as $r \to \infty$, by Lemmas A.2 and A.5(ii).

Lastly, for $\ell = 1, \dots, k$ (assuming $r > k$ without loss of generality):

$$\begin{aligned}\mathbb{E}_{\theta_0}\Big[\widehat{IV}_r \widehat{IV}_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - g_\ell(\theta_0) &= \mathbb{E}_{\theta_0}\Big[IV_r IV_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - \mathbb{E}_{\theta_0}[IV_1 IV_{1-\ell}] \\ &\quad + \mathbb{E}_{\theta_0}\Big[\varepsilon_r IV_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] + \mathbb{E}_{\theta_0}\Big[IV_r \varepsilon_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] + \mathbb{E}_{\theta_0}\Big[\varepsilon_r \varepsilon_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big],\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}_{\theta_0}\Big[\varepsilon_r IV_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] &= \mathbb{E}_{\theta_0}\Big[\mathbb{E}_{\theta_0}\big[\varepsilon_r \mid \mathcal{F}_{r-1}^{\sigma,\varepsilon}\big] IV_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] = 0, \\ \mathbb{E}_{\theta_0}\Big[IV_r \varepsilon_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] &= \mathbb{E}_{\theta_0}\Big[IV_r \mathbb{E}_{\theta_0}\big[\varepsilon_{r-\ell} \mid \mathcal{F}_{r-\ell-1}^{\sigma,\varepsilon}\big] \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] = 0, \\ \mathbb{E}_{\theta_0}\Big[\varepsilon_r \varepsilon_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] &= \mathbb{E}_{\theta_0}\Big[\mathbb{E}_{\theta_0}\big[\varepsilon_r \mid \mathcal{F}_{r-1}^{\sigma,\varepsilon}\big] \varepsilon_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] = 0,\end{aligned}$$

from tower property, because $\mathcal{F}_0^{\widehat{\mathbb{IV}}} \subset \mathcal{F}_{r-\ell-1}^{\sigma,\varepsilon} \subset \mathcal{F}_r^{\sigma,\varepsilon}$, and Assumption 2. Thus, applying yet again Lemma A.2, we get

$$\begin{aligned}\Big\|\mathbb{E}_{\theta_0}\Big[\widehat{IV}_r \widehat{IV}_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - g_\ell(\theta_0)\Big\|_{L^2(\mathbb{P}_{\theta_0})} &= \big\|\mathbb{E}_{\theta_0}\Big[IV_r IV_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - \mathbb{E}_{\theta_0}[IV_1 IV_{1-\ell}]\big\|_{L^2(\mathbb{P}_{\theta_0})} \\ &\le \big\|\mathbb{E}_{\theta_0}\Big[IV_r IV_{r-\ell} \mid \mathcal{F}_0^{W,\varepsilon}\Big] - \mathbb{E}_{\theta_0}[IV_1 IV_{1-\ell}]\big\|_{L^2(\mathbb{P}_{\theta_0})} \\ &= O(r^{-\gamma+1/2}),\end{aligned} \tag{76}$$

as $r \to \infty$, due to Lemma A.5(ii).

Combining (74)–(76), we deduce that

$$\Big\|\mathbb{E}_{\theta_0}\Big[\widehat{\mathbb{IV}}_r \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - \mathbb{E}_{\theta_0}\Big[\widehat{\mathbb{IV}}_r\Big]\Big\|_{L^2(\mathbb{P}_{\theta_0})} = \Big\|\mathbb{E}_{\theta_0}\Big[\widehat{\mathbb{IV}}_r \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - G_c(\theta_0)\Big\|_{L^2(\mathbb{P}_{\theta_0})} = O(r^{-\gamma+1/2}),$$

as $r \to \infty$. Since $\gamma > 1$, this implies that $(\widehat{\mathbb{IV}}_t)_{t\in\mathbb{Z}}$ is an L^2-mixingale of size $-\frac{1}{2}$. ■

Proposition A.6. *Suppose that Assumptions 1 and 6 hold. Then condition* (ii) *of Assumption 7 applies in Examples 3.1–3.3.*

Proof. The error terms in Examples 3.1 and 3.3 differ only by scaling factor, so it suffices to look at the former. Then,

$$\varepsilon_t = \left(\frac{2}{n}\int_{t-1}^{t} \sigma_u^4 \mathrm{d}u\right)^{1/2} Z_t, \quad t \in \mathbb{Z},$$

so that, using the filtration $\mathcal{F}_t^{B,Z} = \sigma\{Z_t, Z_{t-1}, \dots\} \vee \sigma\{B_u : u \le t\}$, $t \in \mathbb{Z}$:

$$\mathbb{E}_{\theta_0}\Big[\varepsilon_r^2 \mid \mathcal{F}_0^{B,Z}\Big] - \mathbb{E}_{\theta_0}[\varepsilon_1^2] = \frac{2}{n}\left(\mathbb{E}_{\theta_0}\left[\int_{r-1}^{r} \sigma_u^4 \mathrm{d}u \mid \mathcal{F}_0^{B,Z}\right] - \mathbb{E}_{\theta_0}\left[\int_0^1 \sigma_u^4 \mathrm{d}u\right]\right), \quad r \ge 1,$$

since $\int_{r-1}^{r} \sigma_u^4 \mathrm{d}u$ and Z_r are conditionally independent on $\mathcal{F}_0^{B,Z}$. We can then apply Lemma A.5(i) and A.2 to show the conjecture of this part.

In Example 3.2,

$$\varepsilon_t = \sum_{i=1}^{n} (Z_{t,i}^2 - 1) \int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}} \sigma_u^2 \mathrm{d}u, \quad t \in \mathbb{Z},$$

whereby for any $r \geq 1$,

$$\mathbb{E}_{\theta_0}\Big[\varepsilon_r^2 \mid \mathcal{F}_0^{B,Z}\Big] - \mathbb{E}_{\theta_0}[\varepsilon_1^2] = 2\sum_{i=1}^{n}\left(\mathbb{E}_{\theta_0}\left[\left(\int_{r-1+\frac{i-1}{n}}^{r-1-\frac{i}{n}} \sigma_u^2 \mathrm{d}u\right)^2 \mid \mathcal{F}_0^{B,Z}\right] - \mathbb{E}_{\theta_0}\left[\left(\int_{\frac{i-1}{n}}^{\frac{i}{n}} \sigma_u^2 \mathrm{d}u\right)^2\right]\right).$$

Applying Minkowski's inequality and Lemmas A.5(i) and A.2 concludes the proof. ■

A.7. Proof of Theorem 3.7

We set $Q_T = \widehat{m}_T(\theta)^\intercal \mathbb{W}_T \widehat{m}_T(\theta)$ and note that Q_T attains its minimum value at $\widehat{\theta}_T$. Combining this with the mean value theorem yields that

$$0 = \nabla_\theta Q_T(\widehat{\theta}_T) = \nabla_\theta Q_T(\theta_0) + \nabla^2_{\theta\theta} Q_T(\bar{\theta}_T)(\widehat{\theta}_T - \theta_0),$$

where $\bar{\theta}_T$ lies between $\widehat{\theta}_T$ and θ_0. Now,

$$\nabla_\theta Q_T(\theta) = 2\nabla_\theta \widehat{m}_T(\theta)^\intercal \mathbb{W}_T \widehat{m}_T(\theta),$$
$$\nabla^2_{\theta\theta} Q_T(\theta) = 2\nabla_{\theta\theta} \widehat{m}_T(\theta)^\intercal \mathbb{W}_T \widehat{m}_T(\theta) + 2\nabla_{\theta\theta} \widehat{m}_T(\theta)^\intercal \mathbb{W}_T \nabla_\theta \widehat{m}_T(\theta)$$

This leads to:

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) = \big(\nabla^2_{\theta\theta} Q_T(\bar{\theta}_T)\big)^{-1} 2\nabla_\theta \widehat{m}_T(\theta_0)^\intercal \mathbb{W}_T \sqrt{T}\widehat{m}_T(\theta_0). \tag{77}$$

Invoking the assumptions of the theorem, it follows that $\bar{\theta}_T \xrightarrow{\mathbb{P}} \theta_0$ as $T \to \infty$. In addition, and recalling Proposition 3.6, we deduce that as $T \to \infty$:

$$\sqrt{T}\widehat{m}_T(\theta_0) \xrightarrow{d} N\big(0, \Sigma_{\widehat{\mathbb{V}}}\big),$$
$$\nabla_\theta \widehat{m}_T(\theta_0) \xrightarrow{\mathbb{P}} J,$$
$$\nabla^2_{\theta\theta} Q_T(\bar{\theta}_T) \xrightarrow{\mathbb{P}} J^\intercal \mathbb{W} J,$$

where the last part uses that $\widehat{m}_T(\theta_0) \xrightarrow{\mathbb{P}} 0$. Then, Slutsky's theorem finishes the proof. ■

A.8. Proof of Theorem 3.8

By the Cramér–Wold device, we can reduce the proof to a univariate setting. Then, denoting $\Sigma_T = \Gamma(0) + 2\sum_{\ell=1}^{T-1}(1 - \ell/T)\Gamma(\ell)$, it is enough to show that $\widehat{\Sigma}_T - \Sigma_T \xrightarrow{\mathbb{P}} 0$. This assertion follows from Theorem 2.1 of Davidson (2020). To sketch the main ideas, we write

$$\widehat{\Sigma}_T/\Sigma_T - 1 = A_1 + A_2,$$

where

$$A_1 = \frac{1}{\Sigma_T}\left[\widehat{\Gamma}(0) - \Gamma(0) + 2\frac{1}{T}\sum_{\ell=1}^{T-1} w(\ell/L)\sum_{t=1}^{T-\ell}\big[\big(\widehat{\mathbb{V}}_t - G_c\big)\big(\widehat{\mathbb{V}}_{t+\ell} - G_c\big) - \Gamma(\ell)\big]\right],$$
$$A_2 = \frac{1}{\Sigma_T} 2\sum_{\ell=1}^{T-1}\big(w(\ell/L) - 1\big)(1 - \ell/T)\Gamma(\ell).$$

Note that the term A_2 is deterministic. Moreover, it follows from the assumptions imposed on w that $A_2 = o(1)$.

To deal with the term A_1, we first note that based on Lemma A.5 and the proof of Proposition 3.6, for each $\ell \geq 0$, the sequence $\{\big(\widehat{\mathbb{V}}_t - G_c\big)\big(\widehat{\mathbb{V}}_{t+\ell} - G_c\big) - \Gamma(\ell)\}$, is an L^2-mixingale of size $-\frac{1}{2}$. Then, Corollary 16.10 in Davidson (1994) implies that

$$\mathbb{E}\left[\left(\sum_{t=1}^{T-\ell}\Big(\big(\widehat{\mathbb{V}}_t - G_c\big)\big(\widehat{\mathbb{V}}_{t+\ell} - G_c\big) - \Gamma(\ell)\Big)\right)^2\right] \leq CT.$$

Along with the convergence

$$\frac{1}{L}\sum_{\ell=0}^{T-1}|w(\ell/L)| \longrightarrow \int_0^\infty |w(x)|\mathrm{d}x,$$

we deduce that $\mathbb{E}[|A_1|] = O(LT^{-1/2}) = o(1)$ based on the assumption on L. ■

A.9. Proof of Theorem 3.9

First, we provide a result for a general weight matrix as in (the proof of) Theorem 3.7. Proceeding as in the proof there, and denoting $\widetilde{Q}_{n,T} = \widetilde{m}_{n,T}(\theta)^{\intercal}\mathbb{W}_{n,T}\widetilde{m}_{n,T}(\theta)$, we find that

$$\sqrt{T}(\widetilde{\theta}_{n,T} - \theta_0) = \left(\nabla^2_{\theta\theta}\widetilde{Q}_{n,T}(\check{\theta}_{n,T})\right)^{-1} 2\nabla_\theta \widetilde{m}_{n,T}(\theta_0)^{\intercal}\mathbb{W}_{n,T}\sqrt{T}\widetilde{m}_{n,T}(\theta_0).$$

Then, as $T \to \infty$ and $n \to \infty$,

$$\begin{aligned}
\sqrt{T}\widetilde{m}_{n,T}(\theta_0) &\xrightarrow{d} N(0, \Sigma_{\mathbb{IV}}),\\
\nabla_\theta \widetilde{m}_{n,T}(\theta_0) &\xrightarrow{\mathbb{P}} \widetilde{J},\\
\nabla^2_{\theta\theta}\widetilde{Q}_{n,T}(\check{\theta}_{n,T}) &\xrightarrow{\mathbb{P}} \widetilde{J}^{\intercal}\mathbb{W}\widetilde{J},
\end{aligned}$$

To wrap this up, we again exploit Slutsky's theorem.

To reach the final conclusion involving the estimator $\widehat{\Sigma}_{n,T}$, it is enough to further show the convergence $\widehat{\Sigma}_{n,T} \xrightarrow{\mathbb{P}} \Sigma_{\mathbb{IV}}$ as $T \to \infty$ and $n \to \infty$. As in the proof of Theorem 3.8, this follows by noting that the error $V^n_t - IV_t$ is negligible in front of Assumption 10. ■

Appendix B. Identification

Assumption 4 is difficult to check, as it involves solving a system of nonlinear equations, which cannot be expressed in closed-form. In lieu of this, it is worthwhile to look briefly at a related question: Does equality of *all* first and second-order moments, i.e.

$$g^{(1)}_0(\theta_1) = g^{(1)}_0(\theta_2) \quad \text{and} \quad g_\ell(\theta_1) = g_\ell(\theta_2), \quad \text{for all } \ell \in \mathbb{N} \cup \{0\}, \tag{78}$$

for $\theta_1, \theta_2 \in \Theta$ hold if and only if $\theta_1 = \theta_2$; the "if" direction being trivial. While an affirmative answer falls short of settling the original question, it does provide circumstantial evidence to suggest that there is no immediate issue with identification, so long as sufficiently many moments are included in the GMM estimation.

Note that since $g^{(1)}_0(\theta) = \xi$, the parameter ξ is always identifiable.

In the fSV model, where the parameter vector is $\theta = (\xi, \lambda, \nu, H)$, we only get a partial identification result. We first observe that, on the one hand, (78) implies

$$\lim_{\ell\to\infty} \frac{\mathbb{E}_{\theta_1}\left[(IV_t - \xi_1)(IV_{t+l} - \xi_1)\right]}{\mathbb{E}_{\theta_2}\left[(IV_t - \xi_2)(IV_{t+l} - \xi_2)\right]} = \lim_{\ell\to\infty} \frac{g_\ell(\theta_1) - g^{(2)}_0(\theta_1)^2}{g_\ell(\theta_2) - g^{(2)}_0(\theta_2)^2} = 1.$$

On the other hand, when $H_1 \neq 0.5 \neq H_2$, Theorem 2.1 and Eq. (8) in Garnier and Sølna (2018) imply that

$$\lim_{\ell\to\infty} \frac{\mathbb{E}_{\theta_1}\left[(IV_t - \xi_1)(IV_{t+l} - \xi_1)\right]}{\mathbb{E}_{\theta_2}\left[(IV_t - \xi_2)(IV_{t+l} - \xi_2)\right]} = \lim_{\ell\to\infty} \frac{\frac{\xi_1^2 H_1(2H_1+1)\nu_1^2}{\lambda_1^2}\ell^{2H_1-2}}{\frac{\xi_2^2 H_2(2H_2+1)\nu_2^2}{\lambda_2^2}\ell^{2H_2-2}} = \lim_{\ell\to\infty} \underbrace{\frac{\frac{\xi_1^2 H_1(2H_1+1)\nu_1^2}{\lambda_1^2}}{\frac{\xi_2^2 H_2(2H_2+1)\nu_2^2}{\lambda_2^2}}}_{>0}\ell^{2(H_1-H_2)},$$

where the limiting value is strictly positive and finite only if $H_1 = H_2$.[16] In that case, the limit is further equal to one only if

$$\frac{\nu_1}{\lambda_1} = \frac{\nu_2}{\lambda_2},$$

since $\xi_1 = \xi_2$ by (78).

We conjecture that a closer analysis of $g^{(2)}_0(\theta)$ and $g_\ell(\theta)$, $\ell \in \mathbb{N}$, may help to uniquely identify also λ and ν, but as it does not appear completely trivial, we do not pursue this avenue further.

Appendix C. The gamma-$\mathcal{BSS}$ process

In this section, we assume that Y is a Brownian semistationary ($\mathcal{BSS}$) SV process, i.e. a Gaussian process constructed with a serial correlation that is locally equivalent to a fBm, whereas the long-range dependence structure can differ a lot. We derive the expressions required to estimate the parameters of a restricted version of this model via our GMM approach.

[16] If either $H_1 = 0.5$ or $H_2 = 0.5$ then either the numerator or denominator decays exponentially, so the limit of the ratio cannot equal one.

The $\mathcal{BSS}$ process is defined as:

$$Y_t = \nu \int_{-\infty}^{t} h(t-s)\mathrm{d}B_s, \quad t \geq 0, \tag{79}$$

where $\nu > 0$ and $h : (0, \infty) \to \mathbb{R}$ is a kernel (subject to some regularity conditions).

A convenient choice is the gamma kernel $h(x) = x^{\alpha}e^{-\lambda x}$ with $\alpha > -1/2$ and $\lambda > 0$ resulting in the Gamma-$\mathcal{BSS}$. This model has local properties that are equivalent to the fSV model, and while not formally long-memory it does allow for substantial persistence in the time series. Its covariance structure was derived in Bennedsen et al. (2022), which we include in (80) for self-containedness. The autocovariance function of integrated variance in (81) is new.

Lemma C.1. *If Y follows the Gamma-$\mathcal{BSS}$ process, $\kappa(\ell)$ has the form:*

$$\begin{aligned} \kappa(0) &= \frac{\nu^2}{(2\lambda)^{2\alpha+1}}\Gamma(2\alpha+1), \\ \kappa(\ell) &= \frac{\nu^2\Gamma(\alpha+1)}{\sqrt{\pi}}\left(\frac{\ell}{2\lambda}\right)^{\alpha+\frac{1}{2}} K_{\alpha+1/2}(\lambda\ell), \quad \ell > 0, \end{aligned} \tag{80}$$

where $K_a(x)$ is the Bessel function of the second kind. In addition, as $\ell \to \infty$, it follows that

$$\mathbb{E}\big[(IV_t - \xi)(IV_{t+\ell} - \xi)\big] \sim \frac{\nu^2\xi^2\Gamma(\alpha+1)(\exp(\lambda)-1)^2}{2^{\alpha+1}\lambda^{\alpha+2}}\ell^{\alpha}\exp(-\lambda(\ell+1)). \tag{81}$$

Proof. Recall that $\kappa(\ell) = \mathrm{cov}(Y_t, Y_{t+\ell})$, where for the Gamma-$\mathcal{BSS}$ process:

$$\kappa(\ell) = \nu^2 \int_0^{\infty} h(x)h(x+\ell)\mathrm{d}x.$$

Inserting the Gamma kernel, $h(x) = x^{\alpha}e^{-\lambda x}$, we deduce the identity:

$$\begin{aligned} \kappa(0) &= \nu^2 \int_0^{\infty} x^{2\alpha}\exp(-2\lambda x)\mathrm{d}x \\ &= \nu^2(2\lambda)^{-2\alpha-1}\int_0^{\infty} z^{2\alpha}\exp(-2z)\mathrm{d}z \\ &= \nu^2(2\lambda)^{-2\alpha-1}\Gamma(2\alpha+1). \end{aligned}$$

Now, for each $\ell > 0$,

$$\begin{aligned} \kappa(\ell) &= \nu^2\exp(-\lambda\ell)\int_0^{\infty} x^{\alpha}(x+\ell)^{\alpha}\exp(-2\lambda x)\mathrm{d}x \\ &= \frac{\nu^2\Gamma(\alpha+1)}{\sqrt{\pi}}\left(\frac{\ell}{2\lambda}\right)^{\alpha+\frac{1}{2}} K_{\alpha+1/2}(\lambda\ell), \end{aligned}$$

where the last equality follows from Gradshteyn and Ryzhik (2007), formula 3.383 (8).

As $\kappa(\ell)$ adheres to (12) with $\beta = \alpha$ and $\rho = \lambda$ by Remark 4.4 in Bennedsen et al. (2022), Theorem 2.1 applies with

$$\int_{-1}^{1}(1-|y|)\phi(y)\mathrm{d}y = \int_{-1}^{1}(1-|y|)\exp(-\lambda y)\mathrm{d}y = \frac{\exp(-\lambda)(\exp(\lambda)-1)^2}{\lambda^2},$$

so it follows that

$$\gamma_{\ell+1,1} \sim F(\ell; \alpha, \lambda, \nu, \xi), \quad \ell \to \infty,$$

where

$$F(\ell; \alpha, \lambda, \nu, \xi) \equiv \underbrace{\frac{\nu^2\xi^2\Gamma(\alpha+1)(\exp(\lambda)-1)^2}{2^{\alpha+1}\lambda^{\alpha+2}}}_{>0}\ell^{\alpha}\exp(-\lambda(\ell+1))$$

for $\ell > 0$, $\alpha > -\frac{1}{2}$, $\lambda > 0$, $\nu > 0$ and $\xi > 0$. ■

The Gamma-$\mathcal{BSS}$ process achieves Assumption 6 in its entire parameter space. Condition *(i)* has been established for general $\mathcal{BSS}$ processes in Bennedsen et al. (2017, Proposition 2.2). Condition *(ii)* follows by noting that the modified Bessel function of the second kind, K_{α}, that appears in its covariance function is continuous. Moreover, the process is already expressed in the form of (49).

There is a more complete result about identification for the Gamma-$\mathcal{BSS}$ process.

Proposition C.2. *In the Gamma-$\mathcal{BSS}$ process, the moment equality (78) holds if and only if $\theta_1 = (\xi_1, \nu_1, \lambda_1, \alpha_1)$ equals $\theta_2 = (\xi_2, \nu_2, \lambda_2, \alpha_2)$.*

Proof. Assume (78), then by Lemma C.1:

$$\begin{aligned}\lim_{\ell\to\infty}\frac{\mathbb{E}_{\theta_1}\big[(IV_t-\xi_1)(IV_{t+l}-\xi_1)\big]}{\mathbb{E}_{\theta_2}\big[(IV_t-\xi_2)(IV_{t+l}-\xi_2)\big]} &= \lim_{\ell\to\infty}\frac{\frac{\nu_1^2\xi_1^2\Gamma(\alpha_1+1)(\exp(\lambda_1)-1)^2}{2^{\alpha_1+1}\lambda_1^{\alpha_1+2}}\ell^{\alpha_1}\exp(-\lambda_1(\ell+1))}{\frac{\nu_2^2\xi_2^2\Gamma(\alpha_2+1)(\exp(\lambda_2)-1)^2}{2^{\alpha_2+1}\lambda_2^{\alpha_2+2}}\ell^{\alpha_2}\exp(-\lambda_2(\ell+1))}\\ &= \lim_{\ell\to\infty}\underbrace{\frac{\frac{\nu_1^2\xi_1^2\Gamma(\alpha_1+1)(\exp(\lambda_1)-1)^2}{2^{\alpha_1+1}}}{\frac{\nu_2^2\xi_2^2\Gamma(\alpha_2+1)(\exp(\lambda_2)-1)^2}{2^{\alpha_2+1}}}}_{>0}\ell^{\alpha_1-\alpha_2}\exp(-(\lambda_1-\lambda_2)\ell).\end{aligned} \tag{82}$$

In order for this limit to be strictly positive and finite, it is necessary that $\lambda_1 = \lambda_2$. In this vein, we also deduce that $\alpha_1 = \alpha_2$ is necessary. Finally, since we already know that $\xi_1 = \xi_2$ from the equality of the first moment, it remains to note that the limit in (82) can only be equal to one if $\nu_1 = \nu_2$.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2022.06.009.

References

Alizadeh, S., Brandt, M.W., Diebold, F.X., 2002. Range-based estimation of stochastic volatility models. J. Finance 57 (3), 1047–1092.
Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. J. Empir. Financ. 4 (2), 115–158.
Andersen, T.G., Bollerslev, T., 1998a. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. Int. Econ. Rev. 39 (4), 885–905.
Andersen, T.G., Bollerslev, T., 1998b. Deutsche Mark-Dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. J. Finance 53 (1), 219–265.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. J. Financ. Econ. 61 (1), 43–76.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71 (2), 579–625.
Andersen, T.G., Sørensen, B.E., 1996. GMM estimation of a stochastic volatility model: A Monte Carlo study. J. Bus. Econom. Statist. 14 (3), 328–352.
Andrews, D.W.K., 1991. Heteroscedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.
Asmussen, S., Glynn, P.W., 2007. Stochastic Simulation: Algorithms and Analysis, first ed. Springer, Berlin.
Back, K., 1991. Asset prices for general processes. J. Math. Econom. 20 (4), 371–395.
Bandi, F.M., Renò, R., 2016. Price and volatility co-jumps. J. Financ. Econ. 119 (1), 107–146.
Barboza, L.A., Viens, F.G., 2017. Parameter estimation of Gaussian stationary processes using the generalized method of moments. Electron. J. Stat. 11 (1), 401–439.
Barndorff-Nielsen, O.E., Basse-O'Connor, A., 2011. Quasi Ornstein–Uhlenbeck processes. Bernoulli 17 (3), 916–941.
Barndorff-Nielsen, O.E., Hansen, P.R., Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. Econometrica 76 (6), 1481–1536.
Barndorff-Nielsen, O.E., Shephard, N., 2001. Non-Gaussian Orstein-Uhlenbeck-based models and some of their uses in financial economics. J. R. Stat. Soc. Ser. B Stat. Methodol. 63 (2), 167–241.
Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. J. R. Stat. Soc. Ser. B Stat. Methodol. 64 (2), 253–280.
Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. J. Financ. Econom. 2 (1), 1–48.
Bayer, C., Friz, P., Gatheral, J., 2016. Pricing under rough volatility. Quant. Finance 16 (6), 887–904.
Belyaev, Y.K., 1961. Continuity and Hölder's conditions for sample functions of stationary Gaussian processes. In: Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability. Vol. II. University of California Press, Berkeley, pp. 23–33.
Bennedsen, M., Lunde, A., Pakkanen, M.S., 2017. Hybrid scheme for Brownian semistationary processes. Finance Stoch. 21 (4), 931–965.
Bennedsen, M., Lunde, A., Pakkanen, M.S., 2022. Decoupling the short- and long-term behavior of stochastic volatility. J. Financ. Econom. (Forthcoming).
Bingham, N.H., Goldie, C.M., Teugels, J.L., 1989. Regular Variation, first ed. Cambridge University Press, Cambridge.
Bollerslev, T., Zhou, H., 2002. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. J. Econometrics 109 (1), 33–65.
Breusch, T., Qian, H., Wyhowski, D., 1999. Redundancy of moment conditions. J. Econometrics 91 (1), 89–111.
Carrasco, M., Florens, J.-P., 2000. Generalization of GMM to a continuum of moment conditions. Econom. Theory 16 (6), 797–834.
Cheridito, P., Kawaguchi, H., Maejima, M., 2003. Fractional Ornstein–Uhlenbeck processes. Electron. J. Probab. 8 (3), 1–14.
Christensen, K., Oomen, R.C.A., Renò, R., 2022. The drift burst hypothesis. J. Econometrics 227 (2), 461–497.
Christensen, K., Thyrsgaard, M., Veliyev, B., 2019. The realized empirical distribution function of stochastic variance with application to goodness-of-fit testing. J. Econometrics 212 (2), 556–583.
Christie, A.A., 1982. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. J. Financial Econ. 10 (4), 407–432.
Christoffersen, P.F., Jacobs, K., Mimouni, K., 2010. Volatility dynamics for the S & P500: Evidence from realized volatility, daily returns, and option prices. Rev. Financial Stud. 23 (8), 3141–3189.
Comte, F., Renault, E., 1998. Long memory in continuous-time stochastic volatility models. Math. Finance 8 (4), 291–323.
Corradi, V., Distaso, W., 2006. Semi-parametric comparison of stochastic volatility models using realized measures. Rev. Econom. Stud. 73 (3), 635–667.
Davidson, J., 1994. Stochastic Limit Theory, first ed. Oxford University Press, Oxford.

Davidson, J., 2020. A new consistency proof for HAC variance estimators. Econom. Lett. 186 (1), 108811.
Delbaen, F., Schachermayer, W., 1994. A general version of the fundamental theorem of asset pricing. Math. Ann. 300 (1), 463–520.
Duffie, D., Singleton, K.J., 1993. Simulated moments estimation of markov models of asset prices. Econometrica 61 (4), 929–952.
Dym, H., McKean, H.P., 1976. Gaussian Processes, Function Theory, and the Inverse Spectral Problem, first edn. Academic Press, Massachusetts.
El Euch, O., Rosenbaum, M., 2018. The characteristic function of rough Heston models. Math. Finance 29 (1), 3–38.
Eraker, B., Johannes, M., Polson, N., 2003. The impact of jumps in volatility and returns. J. Finance 58 (3), 1269–1300.
Fridman, M., Harris, L., 1998. A maximum likelihood approach for non-Gaussian stochastic volatility models. J. Bus. Econom. Statist. 16 (3), 284–291.
Fukasawa, M., Takabatake, T., Westphal, R., 2022. Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. Math. Finance (Forthcoming).
Gallant, A.R., Hsieh, D.A., Tauchen, G.E., 1997. Estimation of stochastic volatility with diagnostics. J. Econometrics 81 (1), 159–192.
Garnier, J., Sølna, K., 2018. Option pricing under fast-varying and rough stochastic volatility. Ann. Finance 14 (4), 489–516.
Gatheral, J., Jaisson, T., Rosenbaum, M., 2018. Volatility is rough. Quant. Finance 18 (6), 933–949.
Gradshteyn, I.S., Ryzhik, I.M., 2007. Table of Integrals, Series, and Products, seventh ed. Academic Press, Massachusetts.
Hall, A.R., Inoue, A., Jana, K., Shin, C., 2007. Information in generalized method of moments estimation and entropy-based moment selection. J. Econometrics 138 (2), 488–512.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.
Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. J. Bus. Econom. Statist. 24 (2), 127–161.
Hansen, P.R., Lunde, A., 2014. Estimating the persistence and the autocorrelation function of a time series that is measured with error. Econom. Theory 30 (1), 60–93.
Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Rev. Financ. Stud. 61 (2), 247–264.
Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. 6 (2), 327–343.
Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. J. Finance 42 (2), 281–300.
Hurst, H.E., 1951. Long-term storage capacity of reservoirs. Trans. Am. Soc. Civ. Eng. 116 (1), 770–799.
Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: The pre-averaging approach. Stochastic Process. Appl. 119 (7), 2249–2276.
Kaarakka, T., Salminen, P., 2011. On fractional Ornstein–Uhlenbeck processes. Commun. Stoch. Anal. 5 (1), 121–133.
Karhunen, K., 1950. Über die struktur stationärer zufälliger Funktionen. Ark. Mat. 1 (2), 141–160.
Li, J., Todorov, V., Tauchen, G., 2017. Adaptive estimation of continuous-time regression models using high-frequency data. J. Econometrics 200 (1), 36–47.
Lindgren, G., 2006. Lectures on Stationary Stochastic Processes, Lecture Notes. Lund University.
Mancini, C., 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. Scand. J. Stat. 36 (2), 270–296.
Mandelbrot, B.B., Van Ness, J.W., 1968. Fractional brownian motions, fractional noises and applications. SIAM Rev. 10 (4), 422–437.
Maruyama, G., 1949. The harmonic analysis of stationary stochastic processes. In: Memoirs of the Faculty of Science. Vol. 4. (1), Kyūshū University, pp. 45–106.
Meddahi, N., 2002. A theoretical comparison between integrated and realized volatility. J. Appl. Econometrics 17 (5), 479–508.
Meddahi, N., 2003. ARMA representation of integrated and realized variances. Econom. J. 6 (2), 335–356.
Melino, A., Turnbull, S.M., 1990. Pricing foreign currency options with stochastic volatility. J. Econometrics 45 (1–2), 239–265.
Merlevède, F., Peligrad, M., Utev, S., 2019. Functional Gaussian Approximation for Dependent Structures, first ed. Oxford University Press, Oxford.
Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis. In: Engle, R.F., McFadden, D. (Eds.), Handbook of Econometrics: Vol. IV. North-Holland, Amsterdam, pp. 2112–2245.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.
Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. J. Econometrics 160 (1), 246–256.
Peccati, G., Taqqu, M.S., 2011. Wiener Chaos: Moments, Cumulants and Diagrams, first ed. Springer, Berlin.
Taqqu, M.S., 1975. Weak convergence to fractional Brownian motion and to the Rosenblatt process. Z. Wahrscheinlichkeitstheor. Verwandte Geb. 31 (4), 287–302.
Taylor, S.J., 1986. Modelling Financial Time Series, first ed. John Wiley and Sons, Chichester.
Todorov, V., 2009. Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. J. Econometrics 148 (2), 131–148.
Todorov, V., Tauchen, G., 2011. Volatility jumps. J. Bus. Econom. Statist. 29 (3), 356–371.
Vetter, M., 2010. Limit theorems for bipower variation of semimartingales. Stochastic Process. Appl. 120 (1), 22–38.
Wright, J.H., 2003. Detecting lack of identification in GMM. Econom. Theory 19 (2), 322–330.
Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100 (472), 1394–1411.