Roughness in spot variance? A GMM approach for estimation of fractional log-normal stochastic volatility models using realized measures*

Anine E. Bolko† Kim Christensen† Mikko S. Pakkanen‡,†
Bezirgen Veliyev†

October 12, 2020

Abstract

In this paper, we develop a generalized method of moments approach for joint estimation of the parameters of a fractional log-normal stochastic volatility model. We show that with an arbitrary Hurst exponent an estimator based on integrated variance is consistent. Moreover, under stronger conditions we also derive a central limit theorem. These results stand even when integrated variance is replaced with a realized measure of volatility calculated from discrete high-frequency data. However, in practice a realized estimator contains sampling error, the effect of which is to skew the fractal coefficient toward "roughness." We construct an analytical approach to control this error. 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 that the bias correction attenuates any systematic deviance in the estimated parameters. 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 integrated variance.

JEL Classification: C10; C50.

Keywords: GMM estimation; realized variance; rough volatility; stochastic volatility.

---

*This paper was presented at the 2019 "Frontiers in Quantitative Finance" workshop at Copenhagen University. We appreciate comments from attendees at the venue. We are also grateful for insightful suggestions by Shin Kanaya. The authors were partially funded by a grant from the Danish Council for Independent Research (DFF – 4182-00050). Our work was also supported by CREATES, Center for Research in Econometric Analysis of Time Series. Please address correspondence to: aeb@econ.au.dk.

†CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark.

‡Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.

1 Introduction

Stochastic volatility (SV) models are pervasive in finance. Over the years, a variety of different models—each with its own volatility 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, Brandt, and Diebold (2002). This class is a promising starting point, because the unconditional distribution of realized variance is close to log-normal (see e.g. Andersen, Bollerslev, Diebold, and Ebens, 2001; Andersen, Bollerslev, Diebold, and Labys, 2003). However, while there is general agreement 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 a standard setting, it is assumed to be a Brownian motion. The mean-reversion and volatility-of-volatility parameters of the model then control both the local properties of volatility and also determine its longer-run persistence. In this instance, there are multiple papers dealing with estimation of the parameters of the log-normal SV model, for example using the method of moments- or likelihood-based approaches (e.g. Taylor, 1986; Melino and Turnbull, 1990; Duffie and Singleton, 1993; Harvey, Ruiz, and Shephard, 1994; Gallant, Hsieh, and Tauchen, 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 does not have independent increments, less is known. In this non-Markovian setting, part of the memory in volatility is reallocated to the background driving process via an additional parameter, the fractal index or Hurst exponent (after Hurst, 1951). Comte and Renault (1998) propose such a version of the log-normal SV model, where the Hurst exponent is larger than one-half—as implied by a standard Brownian motion—thus inducing positive serial correlation in the increments of the process. Bennedsen (2016); Euch and Rosenbaum (2018); Gatheral, Jaisson, and Rosenbaum (2018), among others, study roughness in volatility captured by a fractal index smaller than one-half, rendering volatility anti-persistent and highly erratic at short time scales. The typical estimation approach in the fractional setting is a semi-parametric two-stage procedure, where the Hurst index is pre-estimated, before the other parameters are recovered. While this procedure may yield consistent parameter estimates, it is generally inefficient and may be severely biased in finite samples.

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

Distaso (2006); Todorov (2009).

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

Substitution of the latent volatility with a proxy, however, entails a measurement error. This obfuscates the underlying integrated variance dynamics, which can be detrimental to the estimation procedure if unaccounted for (e.g. Meddahi, 2002; Hansen and Lunde, 2014). In particular, the imposed moment conditions for integrated variance are not valid for the 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 construct an analytic bias correction that controls for the measurement error. Following Patton (2011) we introduce a high-level assumption employing a generic realized measure to proxy for integrated variance. We show in the GMM setting that the additional sampling variation is captured and corrected by adding the measurement error variance to the second moment of integrated variance. Our main asymptotic theory is therefore 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 correction is immaterial.

We investigate our estimator in a simulation study, where various configurations of a fractional log-normal stochastic volatility model with different Hurst parameters covering the rough and long-memory setting are inspected. We note that our procedure is both unbiased and relatively accurate for the unknown parameters, once the above bias correction is adopted. In an empirical application, we study an extensive selection of major equity indexes and confirm roughness in the volatility process, even after smoothing out the effect of noise in the volatility proxy. In those data we consistently locate a roughness parameter around 0.05, in line with the findings of recent work by Fukasawa, Takabatake, and Westphal (2019).

The rest of this paper is organized as follows. Section 2 presents the general log-normal SV model and studies the properties of integrated variance within this framework. The GMM procedure is introduced in Section 3, along with theoretical results of the estimation. Section 4 examines the performance of our estimator in a Monte Carlo study. In Section 5, we apply the estimation 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.

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 arbitragefree 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 $\xi \in \Xi \subset (0, \infty)$ is a scale parameter, representing the unconditional mean of the stochastic variance, while $Y = (Y_t)_{t\geq 0}$ is a zero-mean stationary Gaussian process with covariance function $\kappa(u) = \mathrm{cov}(Y_0, Y_u) = \kappa_\phi(u)$, $u \geq 0$, 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 the model to a Markovian volatility process nor a 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 the volatility itself is not the price of a tradable asset.[2]

To maintain a streamlined exposition, we 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.[3]

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 daily basis and 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.

Note that we exploit the dynamics of integrated variance in this paper. This follows previous work of Fukasawa, Takabatake, and Westphal (2019) on rough volatility, but is in contrast to the

[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. The meaning should be apparent from the context and not cause much confusion in this particular setting.

[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 for an explanation of VIX products).

[3] In particular, the realized variance introduced below is not a consistent estimator for integrated variance in the presence of jumps. In such instances, the bipower variation of Barndorff-Nielsen and Shephard (2004) can be exploited. Below, we derive a bias correction term for bipower variation, which is practical for this purpose.

application of spot variance in, e.g., Bennedsen, Lunde, and Pakkanen (2017a); Gatheral, Jaisson, and Rosenbaum (2018). While spot variance is more ideal, it is associated with numerous pitfalls in practice. First, 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, Hansen, Lunde, and Shephard, 2008; Jacod, Li, Mykland, Podolskij, and Vetter, 2009; Zhang, Mykland, and Aït-Sahalia, 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.

2.1 Properties of integrated variance

In this section, we derive some basic properties of integrated variance in the framework of the general log-normal SV model (1) – (2). This serves as the foundation for our GMM approach to estimate the parameters.

Henceforth, we denote asymptotic equivalence with $f(\ell) \sim g(\ell)$ meaning that $f(\ell)/g(\ell) \to 1$ as $\ell \to \infty$.

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:*

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

for $\ell \in \mathbb{N} \cup \{0\}$*. 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* $\dfrac{\kappa(\ell+y)}{\kappa(\ell)} \to \phi(y)$ *as* $\ell \to \infty$ *for any* $y \in [-1,1]$,

(c) $\limsup_{\ell\to\infty} \sup_{y\in[-1,1]} \left|\dfrac{\kappa(\ell+y)}{\kappa(\ell)}\right| < \infty$.

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

$$\mathbb{E}\big[(IV_t - \xi)(IV_{t+\ell} - \xi)\big] \sim \xi^2 \kappa(\ell) \int_{-1}^1 (1-|y|)\phi(y)\mathrm{d}y. \tag{4}$$

The integral describing the second-order moments of integrated variance depends on Y. In many realistic models, the integral does not possess an analytic solution and has to be either approximated or solved numerically. Moreover, while moments of higher order can be expressed in this way as well, increasing the order of integration by one makes the resulting expressions unwieldy to work with in practice. Hence, our estimation procedure relies on low-order moments.

As an illustration of Theorem 2.1, suppose there exists $\ell_0 > 0$ such that

$$\kappa(\ell) = \ell^{-\beta} e^{-\rho \ell} L(\ell), \quad \ell \geq \ell_0, \tag{5}$$

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 $\lim_{x \to \infty} \dfrac{L(cx)}{L(x)} = 1$ 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, Goldie, and Teugels, 1989, Theorem 1.5.2), condition (a)–(c) of the theorem can be verified with $\phi(y) = e^{-\rho y}$.

2.1.1 Examples

Example 2.2 (Fractional SV (fSV) model) *In a fSV model, the volatility process is driven by a fractional Ornstein-Uhlenbeck (fOU) process:*

$$Y_t = \nu \int_{-\infty}^{t} e^{-\lambda(t-s)} \mathrm{d}B_s^H, \quad t \geq 0, \tag{6}$$

where $\nu, \lambda > 0$, and $B^H = (B_t^H)_{t \geq 0}$ is a fractional Brownian motion (fBm) with Hurst index, $H \in (0,1)$.[4] *This model reduces to a standard log-normal SV model for $H = 0.5$. 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, Jaisson, and Rosenbaum (2018).*

Below, we describe the covariance structure of the fSV model.

Lemma 2.3 *If Y follows the fSV model, we deduce that:*

$$\begin{aligned} \kappa(0) &= \frac{\nu^2}{2\lambda^{2H}} \Gamma(1 + 2H), \\ \kappa(\ell) &= \kappa(0) \cosh(\lambda \ell) - \frac{\nu^2 |\ell|^{2H}}{2} {}_1F_2\left(1; H + \frac{1}{2}, H + 1; \frac{\lambda^2 \ell^2}{4}\right), \quad \ell \geq 0, \end{aligned} \tag{7}$$

where ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x)$ is the generalized hypergeometric function with p parameters of type 1 and q parameters of type 2.

[4] A fBm started at the origin ($B_0^H = 0$) with Hurst exponent $H \in (0,1)$ can be constructed as a weighted infinite moving average of past increments to a standard Brownian motion following the representation in Mandelbrot and Van Ness (1968, Definition 2.1): $B_t^H = \dfrac{1}{\Gamma(H + 1/2)} \left\{ \int_{-\infty}^{0} \left[(t-s)^{H-1/2} - (-s)^{H-1/2} \right] \mathrm{d}B_s + \int_0^t (t-s)^{H-1/2} \mathrm{d}B_s \right\}$, where $\Gamma(\cdot)$ is the Gamma function. This is also termed a fractional (or Weyl) integral of B. As readily seen, the fBm reduces to a standard Brownian motion for $H = 1/2$.

With $H = 1/2$, the above covariance function reduces to

$$\kappa(\ell) = \frac{\nu^2}{2\lambda} e^{-\lambda \ell}, \tag{8}$$

which is the standard formula the log-normal SV model with an exponentially decaying autocovariance function.

Remark 2.4 *In the fSV model, the second-order moment structure of integrated variance can be approximated by the following expression for $\ell \geq 1$:*

$$\begin{aligned}
\mathbb{E}\left[IV_t^2\right] &\approx \xi^2 \exp(\kappa(0))\Big(1 - \kappa(0) + \frac{2\kappa(0)}{\lambda^2}\left(\cosh(\lambda) - 1\right) - c_1 F_2\big(1; H + \frac{3}{2}, H + 2; \frac{\lambda^2}{4}\big)\Big), \\
\mathbb{E}\left[IV_t IV_{t+\ell}\right] &\approx \xi^2 \exp(\kappa(\ell))\Big(1 - \kappa(l) + \frac{2\kappa(0)}{\lambda^2}\cosh(\lambda\ell)\left(\cosh(\lambda) - 1\right) \\
&\quad - \xi^2 \exp(\kappa(\ell))\frac{c}{2}(\ell+1)^{2H+2}{}_1F_2\left(1; H + \frac{3}{2}, H + 2; \frac{\lambda^2(\ell+1)^2}{4}\right) \\
&\quad - \xi^2 \exp(\kappa(\ell))\frac{c}{2}(\ell-1)^{2H+2}{}_1F_2\left(1; H + \frac{3}{2}, H + 2; \frac{\lambda^2(\ell-1)^2}{4}\right) \\
&\quad + \xi^2 \exp(\kappa(\ell)) c\ell^{2H+2}{}_1F_2\left(1; H + \frac{3}{2}, H + 2; \frac{\lambda^2\ell^2}{4}\right),
\end{aligned} \tag{9}$$

where $c = \dfrac{\nu^2}{(2H+1)(2H+2)}$.

Example 2.5 (Brownian semistationary ($\mathcal{BSS}$) SV model) *In this model, Y is a $\mathcal{BSS}$ process, i.e. a Gaussian process constructed with a serial correlation that is locally equivalent to a fBm, whereas the global dependence structure can differ a lot:*

$$Y_t = \nu \int_{-\infty}^{t} h(t-s)\mathrm{d}B_s, \quad t \geq 0, \tag{10}$$

where $\nu > 0$ and $h : (0, \infty) \to \mathbb{R}$ is a kernel function (subject to suitable regularity conditions). A popular choice is the gamma kernel $h(x) = x^\alpha e^{-\lambda x}$ with $\alpha > -1/2$ and $\lambda > 0$ (Gamma-$\mathcal{BSS}$). This model has local properties that are quite similar to the fSV model, and while not formally long-memory it does allow for substantial persistence in the process.

The fSV model conforms to (5) with $\beta = 2(1 - H)$ and $\rho = 0$ for $H \in (0, 0.5) \cup (0.5, 1)$, by Theorem 2.3 of Cheridito, Kawaguchi, and Maejima (2003), and with $\beta = 0$ and $\rho = \lambda$ for $H = 0.5$. The Gamma-$\mathcal{BSS}$ model does so with $\beta = \alpha$ and $\rho = \lambda$ by Remark 4.4 in Bennedsen, Lunde, and Pakkanen (2017a). We stress that L needs not be given in closed form, as the proof of (5) 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 $\ell \to \infty$.

The precise covariance structure of the Gamma-$\mathcal{BSS}$ model was derived in Bennedsen, Lunde, and Pakkanen (2017a) and given by the following result.

Lemma 2.6 *If Y follows the Gamma-$\mathcal{BSS}$ process of Example 2.5, $\kappa(\ell)$ has the form:*

$$\begin{aligned}\kappa(0) &= \frac{\nu^2}{(2\lambda)^{2\alpha+1}}\Gamma(2\alpha+1),\\ \kappa(\ell) &= \frac{\nu^2\Gamma(\alpha+1)}{\sqrt{\pi}}\left(\frac{\ell}{2\lambda}\right)^{\alpha+\frac{1}{2}} K_{\alpha+1/2}(\lambda\ell), \quad \ell>0,\end{aligned} \tag{11}$$

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

$$\mathbb{E}\big[(IV_t-\xi)(IV_{t+\ell}-\xi)\big] \sim \frac{\nu^2\xi^2\Gamma(\alpha+1)(\exp(\lambda)-1)^2}{2^{\alpha+1}\lambda^{\alpha+2}}\ell^\alpha\exp(-\lambda(\ell+1)). \tag{12}$$

3 GMM Estimation

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

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 $\theta=(\xi,\phi)\in\Theta$. The true value is denoted by $\theta_0\in\Theta$ and is fixed. We write $\mathbb{P}_\theta$ for the probability measure induced by θ and $\mathbb{E}_\theta$ is the corresponding expectation operator. Additionally, 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 by a classical result of Belyaev (1961). Condition (ii) is crucial below in ensuring 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 under this model. We return to this in Assumption 2 below.

Example 3.1 *In the context of the fOU process used in the fSV model, condition* (i) *has been shown, e.g., in Proposition 3.4 of Kaarakka and Salminen (2011), while condition* (ii) *follows from the continuity of the hyperbolic cosine and the hypergeometric function ${}_1F_2$ that appears in the covariance function of the process.*

A general result establishing condition (i) *for $\mathcal{BSS}$ processes is derived in Bennedsen, Lunde, and Pakkanen (2017b, Proposition 2.2). This also covers the Gamma-$\mathcal{BSS}$ process. Condition* (ii) *follows for the Gamma-$\mathcal{BSS}$ process by noting that the modified Bessel function of the second kind, K_α, that appear in its covariance function is continuous.*

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

$$\widehat{IV}_t = IV_t + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{13}$$

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. We remark that such a high-level approach to describing measurement error between a realized measure and the corresponding integrated variance is similar in spirit 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 $\theta \in \Theta$.*

Regarding condition (i), we remark that the Gaussian process Y is ergodic by the classical result of Maruyama (1949) provided

$$\kappa_\phi(u) \to 0, \tag{14}$$

as $u \to \infty$, which is evidently true for all models considered in this paper. The processes σ^2 and $(IV_t)_{t\in\mathbb{Z}}$ readily inherit the stationarity and ergodicity of Y. The joint ergodicity of $(IV_t, \varepsilon_t)_{t\in\mathbb{Z}}$ is a more delicate matter, since even if $(IV_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ are ergodic on their own right 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 (Lindgren, 2006, see 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.[5]

[5] Meddahi (2002, Section 4) studies the properties of the measurement error of realized variance under a class of log-normal volatility models including drift. He finds that the mean of the measurement error is negligible at 5-minute sampling frequency.

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

$$\varepsilon_t = h\big(Z_t, (\sigma^2_{s+t-1})_{s\in[0,1]}\big), \quad t \in \mathbb{Z}, \tag{15}$$

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

$$\mathbb{E}_\theta[h(Z_1, f)] = 0, \tag{16}$$

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 (16). 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.

Example 3.2 (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):*

$$RV_t^n = \sum_{i=1}^{n} \big(X_{t-1+\frac{i}{n}} - X_{t-1+\frac{i-1}{n}}\big)^2, \tag{17}$$

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

$$\sqrt{n}(RV_t^n - IV_t) \xrightarrow[n\to\infty]{d_{\mathrm{st}}} \sqrt{2}\int_{t-1}^{t} \sigma_s^2 \mathrm{d}B_s^\perp, \tag{18}$$

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

$$\sqrt{2}\int_{t-1}^{t} \sigma_s^2 \mathrm{d}B_s^\perp \,\bigg|\, \mathcal{F}^\sigma \sim N(0, 2IQ_t), \quad t \in \mathbb{N}, \tag{19}$$

where

$$IQ_t = \int_{t-1}^{t} \sigma_s^4 \mathrm{d}s \tag{20}$$

is the integrated quarticity. Thus the random variables

$$Z_t = \frac{\int_{t-1}^{t} \sigma_s^2 \mathrm{d}B_s^\perp}{IQ_t^{1/2}} \sim N(0,1), \quad t \in \mathbb{N}, \tag{21}$$

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

Informally, the CLT (18) *says that, for any* $t \in \mathbb{N}$,

$$RV_t^n \overset{d}{\approx} IV_t + \left(\frac{2}{n} IQ_t\right)^{1/2} Z_t \tag{22}$$

for large n*, where "*$\overset{d}{\approx}$*" denotes approximate equality in distribution, as used, e.g., in Zhang, Mykland, and Aït-Sahalia (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, Takabatake, and Westphal (2019) employ in their estimation framework. We can represent the error term as* ε_t *in the form* (15) *using the continuous functional*

$$h(z, f) = \left(\frac{2}{n} \int_0^1 f(s)^2 \mathrm{d}s\right)^{1/2} z, \quad z \in \mathbb{R}, \quad f \in C([0,1]). \tag{23}$$

Then (16) *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 3.3 (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 3.2 (see also Remark 3.4 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*

$$X_t = X_0 + \int_0^t \sigma_s \mathrm{d}W_s, \quad t \geq 0, \tag{24}$$

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* $t \in \mathbb{N}$,

$$\begin{aligned} RV_t^n - IV_t &= \sum_{i=1}^n \left(\left(\int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1} \sigma_s \mathrm{d}W_s \right)^2 - \int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1} \sigma_s^2 \mathrm{d}s \right) \\ &= \sum_{i=1}^n (Z_{t,i}^2 - 1) \int_{\frac{i-1}{n}}^{\frac{i}{n}} \sigma_{s+t-1}^2 \mathrm{d}s, \end{aligned} \tag{25}$$

where

$$Z_{t,i} = \frac{\int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1} \sigma_s \mathrm{d}W_s}{\left(\int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1} \sigma_s^2 \mathrm{d}s\right)^{1/2}}, \quad t \in \mathbb{N}, \quad i = 1, \ldots, n. \tag{26}$$

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 (25), we can represent the measurement error $\varepsilon_t = RV_t^n - IV_t$ in the form (15) via the functional

$$h\big((z_1,\dots,z_n), f\big) = \sum_{i=1}^{n}(z_i^2 - 1)\int_{\frac{i-1}{n}}^{\frac{i}{n}} f(s)\mathrm{d}s, \quad (z_1,\dots,z_n)\in\mathbb{R}^n, \quad f\in C([0,1]), \tag{27}$$

and i.i.d. random vectors

$$Z_t = (Z_{t,1},\dots,Z_{t,n}), \quad t\in\mathbb{N}, \tag{28}$$

with components given by (26), so that $d = n$. The property (16) 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.

Remark 3.4 *The moments of the measurement error $\varepsilon_t = RV_t^n - IV_t$ can be analyzed under the leverage effect using Malliavin calculus and chaos expansions, see, e.g., Peccati and Taqqu (2011). However, the resulting formulae are not very convenient for numerical use, which is why we do not pursue this approach here.*

Example 3.5 (Bipower variation, CLT approximation) *In the context of Example 3.2, the realized variance can be substituted with the bipower variation estimator of Barndorff-Nielsen and Shephard (2004), which is defined as:*

$$BV_t^n = \frac{\pi}{2}\sum_{i=2}^{n}\big|X_{t-1+\frac{i}{n}} - X_{t-1+\frac{i-1}{n}}\big|\big|X_{t-1+\frac{i-1}{n}} - X_{t-1+\frac{i-2}{n}}\big|, \quad n\in\mathbb{N}, \tag{29}$$

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

$$\sqrt{n}(BV_t^n - IV_t) \xrightarrow[n\to\infty]{d_{\mathrm{st}}} \sqrt{\frac{\pi^2}{4}+\pi-3}\int_{t-1}^{t}\sigma_s^2\mathrm{d}B_s^{\perp}, \tag{30}$$

jointly for all $t\in\mathbb{N}$, where the structure of the limit is identical to the one in (18). BV_t^n is then approximated for large n by the proxy $\widehat{IV}_t = IV_t + \varepsilon_t$ with error term $\varepsilon_t = \Big(\frac{\frac{\pi^2}{4}+\pi-3}{n}IQ_t\Big)^{1/2}Z_t$, where Z_t, $t\in\mathbb{N}$, are as in Example 3.2. Retracing the arguments in Example 3.2, we can then show that ε_t can be cast in the form (26) so Assumption 2 holds.

3.2 Consistency

Turning to the consistency result for our GMM estimator, we introduce the moment structure of the IV_t process, which is defined by:

$$g_0^{(1)}(\theta) = \mathbb{E}[IV_t(\theta)], \quad g_0^{(2)}(\theta) = \mathbb{E}[IV_t^2(\theta)], \quad g_\ell(\theta) = \mathbb{E}[IV_t(\theta)IV_{t-\ell}(\theta)], \quad \ell\in\mathbb{Z}, \quad \theta\in\Theta, \tag{31}$$

for a fixed $k\in\mathbb{N}$, which we collect in the vector

$$G(\theta) = \big(g_0^{(1)}(\theta), g_0^{(2)}(\theta), g_1(\theta),\dots,g_k(\theta)\big), \quad \theta\in\Theta. \tag{32}$$

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

Proxy	Setting	$c(\theta) = c(\xi, \phi)$
Realized variance	CLT approximation (Example 3.2)	$\frac{2\xi^2}{n} \exp\big(\kappa_\phi(0)\big)$
	No drift or leverage (Example 3.3)	$\frac{4\xi^2}{n} \int_0^1 (1-y) \exp\big(\kappa_\phi(\frac{y}{n})\big) \mathrm{d}y$
Bipower variation	CLT approximation (Example 3.5)	$\frac{\left(\frac{\pi^2}{4}+\pi-3\right)\xi^2}{n} \exp\big(\kappa_\phi(0)\big)$

Note. In the case of Example 3.3, we use Theorem 2.1 to derive the expression.

We also define

$$\begin{aligned} \mathbb{IV}_t &= \big(IV_t, IV_t^2, IV_t IV_{t-1}, \dots, IV_t IV_{t-k}\big), \quad t \in \mathbb{Z}, \\ \widehat{\mathbb{IV}}_t &= \big(\widehat{IV}_t, \widehat{IV}_t^2, \widehat{IV}_t \widehat{IV}_{t-1}, \dots, \widehat{IV}_t \widehat{IV}_{t-k}\big), \quad t \in \mathbb{Z}, \end{aligned} \tag{33}$$

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

By conditions (ii)-(iii) of Assumption 2, we find that for any $\theta \in \Theta$, $t \in \mathbb{Z}$ and $\ell \in \mathbb{Z}$:

$$\begin{aligned} \mathbb{E}_\theta\big[\widehat{IV}_t\big] &= g_0^{(1)}(\theta), \\ \mathbb{E}_\theta\big[\widehat{IV}_t \widehat{IV}_{t-\ell}\big] &= \begin{cases} g_0^{(2)}(\theta) + c(\theta), & \ell = 0, \\ g_\ell(\theta), & \ell \neq 0. \end{cases} \end{aligned} \tag{34}$$

The expressions in (34) show that application of a noisy proxy $\widehat{IV}_t$ leads to biased estimation of a single moment: the variance of integrated variance, $g_0^{(2)}(\theta)$. The other moments are unbiased, 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 selected second-order moments. 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 therefore propose to compare the sample moments of $\widehat{IV}_t$ to a corrected moment function

$$G_c(\theta) = G(\theta) + \big(0, c(\theta), 0, \dots, 0\big), \quad \theta \in \Theta, \tag{35}$$

to ensure an unbiased and consistent GMM estimator. We define a random function:

$$\widehat{m}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \widehat{\mathbb{IV}}_t - G_c(\theta), \quad \theta \in \Theta, \tag{36}$$

which, in view of (34), has

$$\mathbb{E}_{\theta_0}\big[\widehat{m}_T(\theta)\big] = G_c(\theta_0) - G_c(\theta) \tag{37}$$

so that

$$\mathbb{E}_{\theta_0}\left[\widehat{m}_T(\theta_0)\right] = 0. \tag{38}$$

Our GMM estimator is then defined as:

$$\widehat{\theta}_T = \arg\min_{\theta \in \Theta} \widehat{m}_T(\theta)' \mathbb{W}_T \widehat{m}_T(\theta), \tag{39}$$

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 the weight matrix.

Assumption 3 *$\mathbb{W}_T = A_T' A_T$ 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 $T \to \infty$.*

Note that any weight matrix $\mathbb{W}_T$, which is a continuous function of the sample statistic

$$\frac{1}{T} \sum_{t=1}^{T} w\big(\widehat{\mathbb{IV}}_t\big), \tag{40}$$

where $w : \mathbb{R}^{k+2} \to \mathbb{R}^{d'}$ is a measurable function such that $\mathbb{E}_{\theta_0}\Big[\big\|w\big(\widehat{\mathbb{IV}}_{k+1}\big)\big\|_{\mathbb{R}^{d'}}\Big] < \infty$, for some $d' \in \mathbb{N}$, fulfills the convergence criterion in Assumption 3 by the stationarity and ergodicity of $\big(\widehat{\mathbb{IV}}_t\big)_{t \in \mathbb{Z}}$.

Secondly, we assume that the parameter θ is identifiable.

Assumption 4 *$A(G_c(\theta) - G_c(\theta_0)) = 0$ if and only if $\theta = \theta_0$, where A is the limiting matrix in Assumption 3.*

Assumption 4 is an identifying condition, which is equivalent to $A\mathbb{E}_{\theta_0}[\widehat{m}_T(\theta)] = 0$ if and only if $\theta = \theta_0$. It is difficult to check this in practice, because there are no closed-form expressions for the moments of our model, i.e. for the components of $G_c(\theta)$.

Theorem 3.6 *Suppose Assumptions 1 – 4 hold. As $T \to \infty$*

$$\widehat{\theta}_T \xrightarrow{a.s.} \theta_0. \tag{41}$$

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, e.g., 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 the pre-averaging estimator). For fixed $k \in \mathbb{N}$, we denote

$$\mathbb{V}_t^n = \big(V_t^n, (V_t^n)^2, V_t^n V_{t-1}^n, \ldots, V_t^n V_{t-k}^n\big), \quad t \in \mathbb{Z}, \tag{42}$$

with associated sample moments

$$\widetilde{m}_{n,T}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{V}_t^n - G(\theta), \tag{43}$$

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,

$$\widetilde{\theta}_{n,T} = \arg\min_{\theta\in\Theta} \widetilde{m}_{n,T}(\theta)' \mathbb{W}_T \widetilde{m}_{n,T}(\theta), \tag{44}$$

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) $\sup_{t\in\mathbb{Z}} \mathbb{E}[(V_t^n - IV_t)^2] \to 0$ as $n \to \infty$.

Theorem 3.7 *Suppose Assumptions 1 and 3 – 5 hold. As $T \to \infty$ and $n \to \infty$*

$$\widetilde{\theta}_{n,T} \xrightarrow{\mathbb{P}} \theta_0. \tag{45}$$

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

Remark 3.8 *In Appendix A.8, we show that under mild assumptions*

$$\sup_{t\in\mathbb{Z}} \mathbb{E}[(RV_t^n - IV_t)^2] \leq Cn^{-1}, \tag{46}$$

for some $C > 0$, hence 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

$$Y_t = \int_{-\infty}^{t} K(t-u)\mathrm{d}B_u, \quad t \in \mathbb{R}, \tag{47}$$

for a two-sided standard Brownian motion $B = (B_t)_{t\in\mathbb{R}}$ and measurable kernel function $K : (0,\infty) \to \mathbb{R}$ 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 $u \leq 0$ when necessary. (47) is not very 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 somewhat technical, condition known as *pure non-determinism*, see Karhunen (1950, Satz 5) and Dym and McKean (1976, Section 4.5). The

SV models incorporated in this paper adhere to this form. The $\mathcal{BSS}$ model is already expressed in this way, while the fractional Ornstein-Uhlenbeck process also has such a representation (e.g., Barndorff-Nielsen and Basse-O'Connor, 2011).

In the above, it is the asymptotic behavior of $K(u)$ as $u \to \infty$ that governs the long-term memory of Y. To derive the asymptotic normality of our GMM estimator, we need to constrain that memory, which we do by the following:

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

The Gamma-$\mathcal{BSS}$ model achieves Assumption 6 in the entire parameter space. Moreover, Garnier and Sølna (2018) showed that the kernel $K(u)$ in the moving average representation of the fOU process is asymptotically, as $u \to \infty$, proportional to $u^{H-3/2}$. Thereby, the fSV model requires the restriction $H < 1/2$ to be covered by Assumption 6, allowing for rough volatility to be included 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 ceases to hold. In this case, we can show that the expression for the asymptotic covariance matrix in our central limit theorem (Proposition 3.9) does not converge. It is possible that a non-central limit theorem with non-standard scaling holds, a common phenomenon in the realm of long-memory processes, see, e.g., Taqqu (1975). Proving such an extension is rather non-trivial, however, and beyond the scope of the present paper.

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

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 exemplified in Example 3.2, 3.3, and 3.5 above). So here Assumption 6 implies condition (ii), see Proposition A.6 in Appendix A.9. Condition (iii) ensures that the measurement error does not anticipate future increments of the driving Brownian motion W, which is not very restrictive anyway. It is evidently true in the above examples.

The next result presents the central limit theorem for the sample mean of our statistic.

Proposition 3.9 *Suppose that Assumptions 1, 2, 6, and 7 hold. Then, as* $T \to \infty$*, under* $\mathbb{P}_{\theta_0}$*,*

$$T^{1/2}\widehat{m}_T(\theta_0) \xrightarrow{d} N\big(0, \Sigma_{\widehat{\mathbb{IV}}}\big), \tag{48}$$

where $\Sigma_{\widehat{\mathbb{IV}}} = \sum_{\ell=-\infty}^{\infty} \Gamma_{\widehat{\mathbb{IV}}}(\ell)$ *with* $\Gamma_{\widehat{\mathbb{IV}}}(\ell) = \mathbb{E}_{\theta_0}\big[(\widehat{\mathbb{IV}}_1 - G_c(\theta_0))(\widehat{\mathbb{IV}}_{1+\ell} - G_c(\theta_0))'\big]$*,* $\ell \in \mathbb{Z}$*.*

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 $\mathbf{g}(x, \theta) = x - G_c(\theta)$.

Assumption 8 *We have*

(i) θ_0 *is an interior point of* Θ*.*

(ii) $G'\mathbb{W}G$ *is non-singular, where* $G = \mathbb{E}\Big[\nabla_\theta \mathbf{g}(\widehat{\mathbb{IV}}_1, \theta_0)\Big]$ *and* $\mathbb{W} = A'A$*.*

(iii) The function $\theta \mapsto \mathbf{g}(x, \theta)$ *is continuously differentiable. In addition,* $\mathbb{E}[\|\mathbf{g}(\widehat{\mathbb{IV}}_1, \theta_0)\|^2] < \infty$ *and* $\mathbb{E}[\sup_{\theta\in\Theta} \|\nabla_\theta \mathbf{g}(\widehat{\mathbb{IV}}_1, \theta)\|] < \infty$*.*

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

Theorem 3.10 *Suppose Assumptions 1 – 8 hold. As* $T \to \infty$*,*

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \xrightarrow{d} N\big(0, (G'\mathbb{W}G)^{-1}G'\mathbb{W}\Sigma_{\widehat{\mathbb{IV}}}\mathbb{W}G(G'\mathbb{W}G)^{-1}\big). \tag{49}$$

To finish this section, we also 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 9 *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) $T \sup_{t\in\mathbb{Z}} \mathbb{E}[(V_t^n - IV_t)^2] \to 0$ *as* $n \to \infty, T \to \infty$*.*

Theorem 3.11 *Suppose Assumptions 1, 3 – 9 hold. As* $T \to \infty$ *and* $n \to \infty$*,*

$$\sqrt{T}(\widetilde{\theta}_{n,T} - \theta_0) \xrightarrow{d} N\big(0, (\widetilde{G}'\mathbb{W}\widetilde{G})^{-1}\widetilde{G}'\mathbb{W}\Sigma_{\widehat{\mathbb{IV}}}\mathbb{W}\widetilde{G}(\widetilde{G}'\mathbb{W}\widetilde{G})^{-1}\big), \tag{50}$$

where $\Sigma_{\mathbb{IV}} = \sum_{\ell=-\infty}^{\infty} \Gamma_{\mathbb{IV}}(\ell)$ *with* $\Gamma_{\mathbb{IV}}(\ell) = \mathbb{E}_{\theta_0}\big[(\mathbb{IV}_1 - G_c(\theta_0))(\mathbb{IV}_{1+\ell} - G_c(\theta_0))'\big]$*,* $\ell \in \mathbb{Z}$*, and* $\widetilde{G} = \mathbb{E}[\nabla_\theta \mathbf{g}(\mathbb{IV}_1, \theta_0)]$*.*

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 (35).

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

$$\mathrm{d}X_t = \sigma_t \mathrm{d}W_t, \quad t \geq 0, \tag{51}$$

with initial condition $X_0 \equiv 0$. Here, σ_t is the spot volatility and W_t is a standard Brownian motion.

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

$$\mathrm{d}Y_t = -\lambda(Y_t - \eta)\mathrm{d}t + \nu \mathrm{d}B_t^H, \tag{52}$$

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

We draw 500 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 $\mathrm{var}(Y_t) = \dfrac{\nu^2}{2\lambda^{2H}}\Gamma(1+2H)$. To get an almost continuous-time realization of the processes and minimize the 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. We then discretize X via an Euler scheme.

The SDE in (52) is solved to get a more convenient expression for Y:

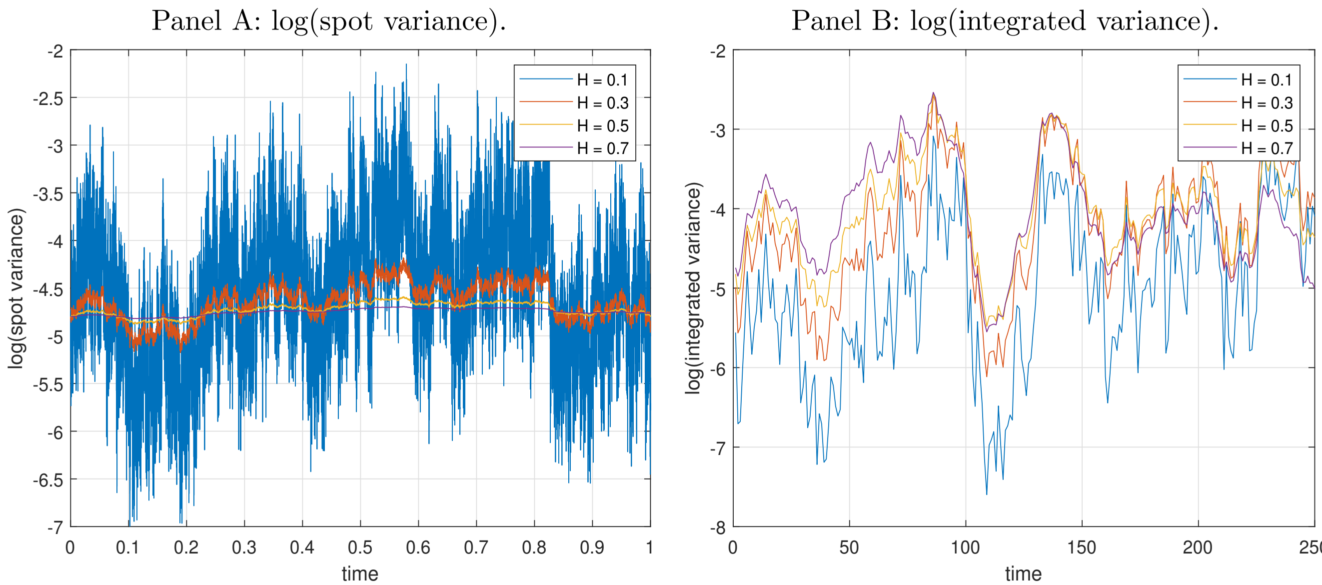
$$Y_t = \eta + (Y_{t-\Delta} - \eta)e^{-\lambda\Delta} + \nu \int_{t-\Delta}^{t} e^{-\lambda(t-s)}\mathrm{d}B_s^H.^{6} \tag{53}$$

The stochastic integral is approximated as $\int_{t-\Delta}^{t} e^{-\lambda(t-s)}\mathrm{d}B_s^H \simeq e^{-\lambda\Delta/2}\int_{t-\Delta}^{t}\mathrm{d}B_s^H$ 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 $O([TN]^3)$, the latter entails a markedly lower budget of $O(TN\log(TN))$ and is our preferred algorithm.

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 $\xi = E(\sigma_t^2) = 0.0225$. This ensures that the unconditional mean of the variance process is identical across settings and implies that the annualized standard deviation σ_t is about 15% on average, close to the aggregate level of volatility in the

[6] 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 fractional Brownian motion, where a stochastic calculus may not exist. Nevertheless, it goes through in this particular instance, see, e.g., Cheridito, Kawaguchi, and Maejima (2003).

Figure 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.

data analyzed in Section 5. As we are particularly attentive to estimation of the Hurst index, we choose $H = [0.1, 0.3, 0.5, 0.7]$ as in Fukasawa, Takabatake, and Westphal (2019), thus covering both the rough, standard and long-memory case. We calibrate λ and ν by equating $\text{std}(IV_t) = 0.05$, which is on par with the corresponding attribute of the bipower variation of the .SPX equity (that is, the S&P 500 index) in the empirical part. This effectively locks in ν/λ^H for each H, and as an identifying restriction we match the lag 100 autocovariance of integrated variance to the sample autocovariance of .SPX bipower variation.

The parameters are presented in Table 2. A realization of the spot and integrated variance processes from each model are plotted in Figure 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 unknown parameter vector is $\theta_0 = (\xi, \lambda, \nu, H)$, which we estimate with the non-gradient-based Nealder-Mead algorithm available via `fminsearch` in MatLab. That function does not accept boundary conditions, so we reparameterize the model by log-transforming ν and λ, while H is bounded by the logistic function. We launch the engine at initial values determined as follows: ξ is started at the unconditional mean of 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, Jaisson, and Rosenbaum (2018), which relies on the scaling law:

$$\gamma_h \equiv \mathbb{E}\big[|Y_{t+h} - Y_t|^q\big] \to K_q \nu^q |h|^{qH}, \tag{54}$$

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 $|h|$: $\ln(\gamma_h) = \ln(K_q\nu^q) + qH\ln(|h|)$. We employ RV_t^n as a proxy for the instantaneous variance and substitute the left-hand side of (54) by the sample mean:

$$\hat{\gamma}_h = \frac{1}{T-m}\sum_{t=1}^{T-m} |\ln(RV_{t+h}^n) - \ln(RV_t^n)|^q, \tag{55}$$

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 $\ln(RV_t^n)$.

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 still a lot of work left for the GMM procedure. We match the ℓ'th sample autocovariance of RV_t^n with the approximate second-order moment structure of IV_t—available in closed-form from Remark 2.4—with $\ell = [0, 1, 2, 5, 10, 20, 50]$.[7] The weight matrix, $\mathbb{W}_T$, is a data-driven Newey and West (1987) HAC-type estimator computed with a Bartlett kernel and a lag length equal to $[T^{1/3}]$.

The results are presented in Table 2. We report the mean parameter estimate (standard error in parenthesis) both for the initial and final value, where all calculations are done across simulations. 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 (35). Several interesting findings emerge. Firstly, the setting with integrated variance leads to parameter estimates that are generally close to their population counterparts across models, thus verifying the robustness and accuracy of our procedure. Secondly, we record a significant deterioration in the estimation of H for realized variance without the quarticity adjustment. As explained, realized variance is a noisy proxy for integrated variance and this translates to additional roughness in the $(RV_t^n)_{t=1}^T$ process. In that case, we further note ν increases, while λ decreases in order to "compensate" for the spurious loss of memory, even as H is decreasing. Including the bias correction leads to a huge improvement in the estimates that are surprisingly accurate also for small H. While a marginal downward bias is retained, the typical estimate is nevertheless within about a standard error of the true value and very much in line with the infeasible results recovered using the benchmark integrated variance.

[7] 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.

Table 2: Parameter estimation of the fSV model in simulated data.

parameter	value	integrated variance		realized variance		
		initial value	sim. average	initial value	sim. average no correction	sim. average correction
Panel A:						
ξ	0.0225	0.0268 (0.0246)	0.0251 (0.0222)	0.0268 (0.0244)	0.0249 (0.0213)	0.0251 (0.0213)
λ	0.0020	0.0189 (0.0066)	0.0029 (0.0029)	0.0155 (0.0059)	0.0025 (0.0104)	0.0034 (0.0036)
ν	1.2080	0.5963 (0.0076)	1.3686 (0.3147)	0.6499 (0.0084)	2.1987 (1.1309)	1.1734 (0.1588)
H	0.1000	0.3005 (0.0102)	0.0817 (0.0440)	0.2680 (0.0098)	0.0492 (0.0394)	0.0969 (0.0330)
Panel B:						
ξ	0.0225	0.0222 (0.0067)	0.0212 (0.0061)	0.0223 (0.0067)	0.0206 (0.0058)	0.0211 (0.0061)
λ	0.0062	0.0187 (0.0041)	0.0074 (0.0043)	0.0107 (0.0029)	0.0035 (0.0028)	0.0075 (0.0048)
ν	0.4395	0.3183 (0.0040)	0.4541 (0.0485)	0.3907 (0.0048)	0.6729 (0.1183)	0.4505 (0.0640)
H	0.3000	0.4420 (0.0108)	0.2706 (0.0474)	0.3493 (0.0110)	0.1708 (0.0461)	0.2726 (0.0540)
Panel C:						
ξ	0.0225	0.0229 (0.0081)	0.0220 (0.0075)	0.0229 (0.0081)	0.0201 (0.0069)	0.0218 (0.0074)
λ	0.0126	0.0205 (0.0038)	0.0159 (0.0085)	0.0051 (0.0015)	0.0032 (0.0034)	0.0168 (0.0124)
ν	0.2116	0.1761 (0.0022)	0.2168 (0.0294)	0.2814 (0.0033)	0.4665 (0.0948)	0.2178 (0.0369)
H	0.5000	0.5867 (0.0103)	0.4800 (0.0709)	0.3594 (0.0114)	0.2166 (0.0754)	0.4813 (0.0895)
Panel D:						
ξ	0.0225	0.0223 (0.0101)	0.0213 (0.0097)	0.0223 (0.0101)	0.0189 (0.0085)	0.0212 (0.0096)
λ	0.0300	0.0300 (0.0041)	0.0381 (0.0203)	0.0046 (0.0013)	0.0041 (0.0060)	0.0355 (0.0205)
ν	0.1455	0.1268 (0.0017)	0.1614 (0.0258)	0.2503 (0.0029)	0.4244 (0.1470)	0.1694 (0.0339)
H	0.7000	0.7123 (0.0096)	0.6577 (0.0912)	0.3683 (0.0120)	0.2356 (0.1074)	0.6321 (0.0977)

Note. We simulate 500 replications of a fractional Ornstein-Uhlenbeck process $\mathrm{d}Y_t = -\lambda(Y_t - \eta)\mathrm{d}t + \nu \mathrm{d}B_t^H$ 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 – D, where $\xi = e^{\eta + 0.5\mathrm{var}(Y_t)}$. We estimate θ_0 with the GMM procedure developed in the main text, where the theoretical mean and autovariance (at lag 0, 1, 2, 5, 10, 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 parameters estimates, based on integrated variance (left) and realized variance (right). The latter is computed both without and with the correction in (35). Standard deviation across simulations appear in parenthesis.

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 non-parametric volatility estimators are calculated and appended to the database. We here employ the daily bipower variation defined in (29). This is a consistent and jump-robust measure of integrated variance with $\text{avar}\big[\sqrt{n}(BV_t^n - IV_t)\big] = 2.6\mathbb{E}\big[\int_{t-1}^{t} \sigma_s^4 \mathrm{d}s\big]$. In line with our comments above, a 5-minute sampling frequency corresponding to $n = 78$ has been set to suppress microstructure noise.

An overview of the 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 sizable gaps in their data series.

The GMM estimation of the model parameters follows the setup from the simulation section. 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. The $\bar{\xi} = 0.0214$ estimate corresponds to about 14.49% annualized volatility in the aggregate stock market.

The reported Hurst exponents suggest a very rough volatility process with an average level of $\bar{H} = 0.035$. This is on par with Bayer, Friz, and Gatheral (2016) and Fukasawa, Takabatake, and Westphal (2019) but slightly smaller compared to Bennedsen, Lunde, and Pakkanen (2017a) and Gatheral, Jaisson, and Rosenbaum (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 Figure 1). This ought to bias their H estimates upwards. Our procedure does not suffer from that problem here, as we directly compare bipower variation to the dynamics of integrated variance in the fSV model, so the averaging "cancels out."

Looking at the table, the results are remarkably stable across assets. We do observe a minor deviation for the Finnish .OMXHPI index. On manual inspection of the data, we found that its sample autocorrelation function (acf) differs notably from the other stock indexes. This can possibly be explained by the fact that its evolution has for a large portion of our sample been dominated by a single stock, following the rise and fall of Nokia.

To gauge the statistical fit of the model, we calculate a $\mathcal{J}$-test for overidentifying restrictions. The test statistic has an asymptotic χ^2-distribution with four degrees of freedom 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.

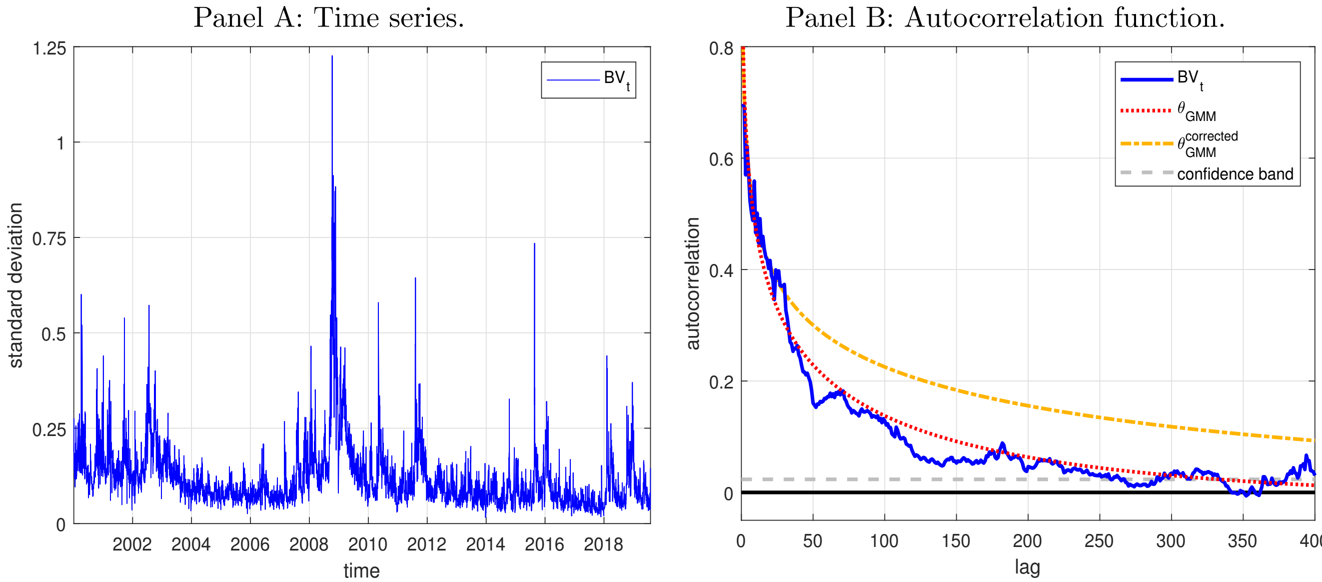
As an illustration of our findings we zoom in at .SPX, which represents the S&P 500 index and is therefore related to developments in the US stock market. In Panel A of Figure 2, we show the bipower variation of .SPX (the raw estimator has been converted to standard deviation per

Table 3: Parameter estimation of the log-normal fSV model from stock index return data.

					parameter estimate				
code	index	location	starting date	sample size	ξ	$\lambda \times 100$	ν	H	$\mathcal{J}$-test
AEX	AEX	Netherlands	2000–01	4,992	0.025	0.046	1.545	0.046	0.929
AORD	All Ordinaries	Australia	2000–01	4,942	0.010	0.031	2.064	0.025	0.984
BFX	BEL 20	Belgium	2000–01	4,989	0.019	0.119	1.708	0.039	0.934
BSESN	BSE Sensex	India	2000–01	4,858	0.031	0.020	1.954	0.026	0.699
BVLG	PSI All-Share	Portugal	2012–10	1,731	0.010	0.019	1.898	0.015	0.944
BVSP	Bovespa	Brazil	2000–01	4,822	0.031	0.118	1.789	0.030	0.907
DJI	Dow Jones Industrial Average	USA	2000–01	4,909	0.018	0.043	1.986	0.032	0.745
FCHI	CAC 40	France	2000–01	4,991	0.030	0.075	1.796	0.037	0.989
FTMIB	FTSE MIB	Italy	2009–06	2,582	0.025	0.019	1.989	0.016	0.895
FTSE	FTSE 100	United Kingdom	2000–01	4,937	0.020	0.033	2.482	0.021	0.885
GDAXI	DAX	Germany	2000–01	4,966	0.034	0.052	1.679	0.041	0.906
.GSPTSE	TSX Composite	Canada	2002–05	4,316	0.011	0.019	1.828	0.042	0.821
.HSI	Hang Seng	Hong Kong	2000–01	4,794	0.021	0.041	1.875	0.030	0.968
IBEX	IBEX 35	Spain	2000–01	4,959	0.031	0.060	1.904	0.026	0.941
IXIC	Nasdaq 100	USA	2000–01	4,907	0.028	0.027	1.901	0.033	0.973
KS11	KOSPI	South Korea	2000–01	4,818	0.027	0.064	1.442	0.053	0.676
MXX	IPC Mexico	Mexico	2000–01	4,912	0.015	0.029	2.215	0.017	0.871
N225	Nikkei 225	Japan	2000–02	4,765	0.021	0.100	1.910	0.031	0.820
NSEI	Nifty 50	India	2000–01	4,855	0.024	0.004	2.078	0.020	0.757
OMXC20	OMXC20	Denmark	2005–10	3,438	0.021	0.090	1.223	0.065	0.881
OMXHPI	OMX Helsinki	Finland	2005–10	3,470	0.021	0.147	0.511	0.186	0.872
OMXSPI	OMX Stockholm	Sweden	2005–10	3,470	0.017	0.018	1.836	0.030	0.789
OSEAX	Oslo Exchange	Norway	2001–09	4,465	0.019	0.025	1.826	0.032	0.867
RUT	Russel 2000	USA	2000–01	4,907	0.014	0.072	1.980	0.035	0.748
SMSI	Madrid General	Spain	2005–07	3,589	0.023	0.140	2.005	0.029	0.962
SPX	S&P 500	USA	2000–01	4,913	0.019	0.055	1.579	0.048	0.779
SSEC	Shanghai Composite	China	2000–01	4,726	0.036	0.009	1.833	0.025	0.777
SSMI	Swiss Market Index	Switzerland	2000–01	4,906	0.018	0.046	2.146	0.027	0.898
STOXX50E	EURO STOXX 50	Europe	2000–01	4,991	0.031	0.060	1.792	0.036	0.954
Average					0.021	0.051	1.702	0.035	

Note. "code" is based on the Oxford-Man Institute's naming convention. 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 the estimated parameter vector is reported in the bottow row. $\mathcal{J}$-test is the P-value of the Sargan-Hansen test of overidentifying restrictions, which is asymptotically χ^2-distributed with four degress of freedom.

Figure 2: Properties of .SPX bipower variation.



Note. In Panel A, we plot the bipower variation of .SPX converted to standard deviation per annum. In Panel B, we show the sample acf of bipower variation 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 $\hat{\theta}_{\text{GMM}}$, where the latter is reported with and without bias correction from (35).

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 BV_t^n together with a Bartlett one-sided 95% white noise confidence band. The slow decay is consistent with significant memory in integrated variance. 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 bias correction in (35). The acf of the uncorrected estimator tracks the sample counterpart based on bipower variation 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). Meanwhile, it retains the impression of an exponential decline, which suggests that short-memory components are driving the serial correlation, in contrast to a hyperbolic decay symptomatic of true long-memory, where the latter entails a Hurst exponent above 0.5.[8]

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 integrated variance.

[8]We also estimated the fSV model with a driving standard Brownian motion, i.e. pre-imposing $H = 0.5$. The remaining parameters were $(\hat{\xi}, \hat{\lambda}, \hat{\nu}) = (0.016, 0.612, 1.048)$, which broadly aligns with previous studies, e.g., Tegnér and Poulsen (2018). Intuitively, to fit the sample acf of bipower variation 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.

6 Conclusion

We propose a GMM framework for estimation of the log-normal stochastic volatility model governed by a general fractional Brownian motion. Our procedure is built from the dynamic properties of integrated variance in this model, but it employs a generic 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 proposed routine is capable of recovering the parameters of the model across the entire memory spectrum. We implement the approach on a vast array of 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.

In future work, we envision our theoretical results can be extended to other classes of fSV models. The Heston (1993) model, for example, has been studied in both rough and long-memory form (e.g., Comte, Coutin, and Renault, 2012; Guennoun, Jacquier, Roome, and Shi, 2018). With its affine structure and appeal to option pricing (Duffie, Pan, and Singleton, 2000; Euch and Rosenbaum, 2018), it should be interesting to adapt the GMM estimation routine outlined in this paper to the generalized fractional version of that model. We leave that for another endeavour.

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

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

and introduce the linear (bijective) change of variables:

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

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:

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

Firstly,

$$\iint\limits_{B_1} f(|u|)\mathrm{d}u\mathrm{d}v = \int_{k-2}^{k-1}\left(\int_{2(k-1)-u}^{u+2} f(|u|)\mathrm{d}v\right)\mathrm{d}u = 2\int_{k-2}^{k-1}\big(u-(k-2)\big)f(|u|)\mathrm{d}u$$
$$= 2\int_0^1 (1-y)f(|k-1-y|)\mathrm{d}y,$$

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

$$\iint\limits_{B_2} f(|u|)\mathrm{d}u\mathrm{d}v = \int_{k-1}^{k}\left(\int_{u}^{2k-u} f(|u|)\mathrm{d}v\right)\mathrm{d}u = 2\int_{k-1}^{k}(k-u)f(u)\mathrm{d}u$$
$$= 2\int_0^1 (1-y)f(k-1+y)\mathrm{d}y,$$

by substituting $y = u-(k-1)$ and noting $u \geq k-1 \geq 0$. 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 $(\sigma_t^2)_{t\geq 0}$ 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

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

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

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

as a consequence of Lemma A.1.

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

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

for any $\ell \geq \ell_0$.

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 (2.1). 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 $u_\ell(y) \equiv (1-y)\big(\frac{\kappa(\ell-y)}{\kappa(\ell)} + \frac{\kappa(\ell+y)}{\kappa(\ell)}\big)$ 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 Lemma 2.3

In view of Remark 2.4 in Cheridito, Kawaguchi, and Maejima (2003), for any $\ell \geq 0$

$$\kappa(\ell) = \frac{\nu^2\Gamma(2H+1)\sin(\pi H)}{2\pi} \int_{-\infty}^{\infty} e^{i\ell z} \frac{|z|^{1-2H}}{\lambda^2 + z^2}\mathrm{d}z.$$

Appealing to Euler's formula $e^{ix} = \cos(x) + i\sin(x)$ and the cosine (sine) function being even (odd), we further conclude that:

$$\begin{aligned}\kappa(\ell) &= \frac{\nu^2\Gamma(2H+1)\sin(\pi H)}{\pi} \int_0^\infty \cos(\ell z)\frac{z^{1-2H}}{\lambda^2+z^2}\mathrm{d}z \\ &= \frac{\nu^2\Gamma(2H+1)\sin(\pi H)}{\pi} \int_0^\infty \cos(\ell\lambda x)\frac{\lambda^{1-2H}x^{1-2H}}{\lambda^2(1+x^2)}\lambda\mathrm{d}x \\ &= \frac{\nu^2\Gamma(2H+1)\sin(\pi H)}{\lambda^{2H}\pi} \int_0^\infty \cos(\ell\lambda x)\frac{x^{1-2H}}{1+x^2}\mathrm{d}x.\end{aligned}$$

To compute the variance term ($\ell = 0$), this expression is evaluated using Gradshteyn and Ryzhik (2007), formula 8.380 (3), 8.384 (1) and 8.334 (3):

$$\begin{aligned}
\kappa(0) &= \frac{\nu^2\Gamma(2H+1)\sin(\pi H)}{\lambda^{2H}\pi}\int_0^\infty \frac{x^{1-2H}}{1+x^2}\mathrm{d}x \\
&= \frac{\nu^2\Gamma(2H+1)\sin(\pi H)}{2\lambda^{2H}\pi}B(1-H,H) \\
&= \frac{\nu^2}{2\lambda^{2H}}\Gamma(2H+1),
\end{aligned}$$

where $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\mathrm{d}t$ is the Beta function.

Now, we proceed to compute the covariance, i.e. $\ell > 0$. In view of Gradshteyn and Ryzhik (2007), formula 3.773 (4) [with $\mu = 1$ and $\nu = 1/2 - H$], this reduces to:

$$\begin{aligned}
\int_0^\infty \cos(\ell\lambda x)\frac{x^{1-2H}}{1+x^2}\mathrm{d}x &= \frac{1}{2}B(1-H,H)\,{}_1F_2\left(1-H;1-H,\frac{1}{2};\frac{\ell^2\lambda^2}{4}\right) \\
&\quad + \frac{\sqrt{\pi}(\ell\lambda)^{2H}}{4^{H+1/2}}\frac{\Gamma(-H)}{\Gamma(H+\frac{1}{2})}\,{}_1F_2\left(1;H+\frac{1}{2},H+1;\frac{\ell^2\lambda^2}{4}\right) \\
&= \frac{\pi\cosh(\ell\lambda)}{2\sin(\pi H)} - \frac{\sqrt{\pi}(\ell\lambda)^{2H}}{2^{2H+1}H}\frac{\Gamma(1-H)}{\Gamma(H+\frac{1}{2})}\,{}_1F_2\left(1;H+\frac{1}{2},H+1;\frac{\ell^2\lambda^2}{4}\right) \\
&= \frac{\pi\cosh(\ell\lambda)}{2\sin(\pi H)} - \frac{(\ell\lambda)^{2H}}{2\Gamma(2H+1)}\frac{\pi}{\sin(\pi H)}\,{}_1F_2\left(1;H+\frac{1}{2},H+1;\frac{\ell^2\lambda^2}{4}\right),
\end{aligned}$$

where we used the identity $\Gamma(z)\Gamma(1-z) = \pi/\sin(\pi z)$ and the duplication formula $\Gamma(z)\Gamma(z+1/2) = 2^{1-2z}\sqrt{\pi}\Gamma(2z)$. This yields

$$\begin{aligned}
\kappa(\ell) &= \frac{\nu^2\Gamma(2H+1)\sin(\pi H)}{\lambda^{2H}\pi}\int_0^\infty \cos(\ell\lambda x)\frac{x^{1-2H}}{1+x^2}\mathrm{d}x \\
&= \frac{\nu^2\Gamma(2H+1)\cosh(\ell\lambda)}{2\lambda^{2H}} - \frac{\nu^2\ell^{2H}}{2}\,{}_1F_2\left(1;H+\frac{1}{2},H+1;\frac{\ell^2\lambda^2}{4}\right),
\end{aligned}$$

and we are done. ■

A.4 Derivation of Remark 2.4

We start with the variance. By a standard linear Taylor approximation $\exp(x) \approx 1 + x$ and setting $x = \kappa(y) - \kappa(0)$, it follows that

$$\begin{aligned}
\mathbb{E}[IV_t^2] &= 2\xi^2\int_0^1 (1-y)\exp(\kappa(y))\mathrm{d}y \\
&\approx 2\xi^2\exp(\kappa(0))\int_0^1 (1-y)(1+\kappa(y)-\kappa(0))\mathrm{d}y \\
&= \xi^2\exp(\kappa(0))\Big(1-\kappa(0)+2\kappa(0)\int_0^1 (1-y)\cosh(\lambda y)\mathrm{d}y\Big)
\end{aligned}$$

$$- \xi^2 \exp(\kappa(0))\nu^2 \int_0^1 (1-y)y^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 y^2}{4}\right) \mathrm{d}y.$$

The first integral is calculated as

$$\int_0^1 (1-y)\cosh(\lambda y)\,\mathrm{d}y = \frac{\cosh(\lambda)-1}{\lambda^2}.$$

To handle the second integral, the substitution $x = y^2$ is done, and hereafter we apply Gradshteyn and Ryzhik (2007), formula 7.512 (11) twice [with $\mu = 1, \nu = H+1$ and then $\mu = 1, \nu = H + \frac{1}{2}$],

$$\begin{aligned}
&\int_0^1 (1-y)y^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 y^2}{4}\right) \mathrm{d}y \\
&\qquad = \frac{1}{2}\int_0^1 (x^{H-1/2} - x^H) {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x}{4}\right) \mathrm{d}x \\
&\qquad = \frac{\Gamma(H+\frac{1}{2})}{2\Gamma(H+\frac{3}{2})} {}_1F_2\left(1; H+\frac{3}{2}, H+1; \frac{\lambda^2}{4}\right) \\
&\qquad - \frac{\Gamma(H+1)}{2\Gamma(H+2)} {}_1F_2\left(1; H+\frac{1}{2}, H+2; \frac{\lambda^2 x}{4}\right)\Bigg) \\
&\qquad = \frac{1}{2} \frac{(H+1) {}_1F_2\left(1; H+\frac{3}{2}, H+1; \frac{\lambda^2 x}{4}\right) - (H+\frac{1}{2}) {}_1F_2\left(1; H+\frac{1}{2}, H+2; \frac{\lambda^2}{4}\right)}{(H+\frac{1}{2})(H+1)} \\
&\qquad = \frac{{}_1F_2\left(1; H+\frac{3}{2}, H+2; \frac{\lambda^2 x}{4}\right)}{(2H+1)(2H+2)},
\end{aligned}$$

where the property $\Gamma(z+1) = z\Gamma(z)$ is applied repeatedly along with the definition of the generalized hypergeometric function. At last, setting $c = \dfrac{\nu^2}{(2H+1)(2H+2)}$ completes the derivation of the variance term.

For each $\ell \geq 1$, we approximate the covariance as:

$$\begin{aligned}
\mathbb{E}[IV_t IV_{t+\ell}] &\approx \xi^2 \exp(\kappa(\ell)) \int_0^1 (1-y)\left(2 + \kappa(\ell+y) - \kappa(\ell) + \kappa(\ell-y) - \kappa(\ell)\right) \mathrm{d}y \\
&= \xi^2 \exp(\kappa(\ell)) \left(1 - \kappa(\ell) + \int_0^1 (1-y)\big(\kappa(\ell+y) + \kappa(\ell-y)\big)\mathrm{d}y\right) \\
&= \xi^2 \exp(\kappa(\ell)) \left(1 - \kappa(l) + \kappa(0) \int_0^1 (1-y)\big(\cosh(\ell+y) + \cosh(\ell-y)\big)\mathrm{d}y\right) \\
&- \xi^2 \exp(\kappa(\ell))\nu^2/2 \int_0^1 (1-y)(\ell+y)^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell+y)^2}{4}\right) \mathrm{d}y \\
&- \xi^2 \exp(\kappa(\ell))\nu^2/2 \int_0^1 (1-y)(\ell-y)^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell-y)^2}{4}\right) \mathrm{d}y.
\end{aligned}$$

We calculate each of the above integrals in turn. First,

$$I_1 = \int_0^1 (1-y)\cosh\big(\lambda(l+y)\big)\mathrm{d}y + \int_0^1 (1-y)\cosh\big(\lambda(l-y)\big)\mathrm{d}y$$

$$
\begin{aligned}
&= \frac{\cosh\big(\lambda(l+1)\big) - \cosh(\lambda l) - \lambda\sinh(\lambda l)}{\lambda^2} + \frac{\cosh\big(\lambda(l-1)\big) - \cosh(\lambda l) + \lambda\sinh(\lambda l)}{\lambda^2} \\
&= \frac{\cosh\big(\lambda(l+1)\big) - 2\cosh(\lambda l) + \cosh\big(\lambda(l-1)\big)}{\lambda^2} \\
&= \frac{2}{\lambda^2}\cosh(\lambda l)(\cosh(\lambda) - 1).
\end{aligned}
$$

Secondly,

$$
\begin{aligned}
I_2 &= \int_0^1 (1-y)(\ell+y)^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell+y)^2}{4}\right)\mathrm{d}y \\
&= \int_\ell^{\ell+1} (1+\ell-x)x^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right)\mathrm{d}x \\
&= \int_0^{\ell+1} (1+\ell-x)x^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right)\mathrm{d}x \\
&- \int_0^{\ell} (1+\ell-x)x^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right)\mathrm{d}x.
\end{aligned}
$$

Thirdly, we express the last integral as

$$
\begin{aligned}
I_3 &= \int_0^1 (1-y)(\ell-y)^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell-y)^2}{4}\right)\mathrm{d}y \\
&= \int_{\ell-1}^{\ell} (1-\ell+x)x^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right)\mathrm{d}x \\
&= \int_0^{\ell} (1-\ell+x)x^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right)\mathrm{d}x \\
&- \int_0^{\ell-1} (1-\ell+x)x^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right)\mathrm{d}x.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
I_2 + I_3 &= \int_0^{\ell+1} (1+\ell-x)x^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right)\mathrm{d}x \\
&- 2\int_0^{\ell} (\ell-x)x^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right)\mathrm{d}x \\
&+ \int_0^{\ell-1} (\ell-1-x)x^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right)\mathrm{d}x \\
&= (\ell+1)^{2H+2}\int_0^1 (1-z)z^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell+1)^2 z^2}{4}\right)\mathrm{d}z \\
&- 2\ell^{2H+2}\int_0^1 (1-z)z^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2\ell^2 z^2}{4}\right)\mathrm{d}z \\
&+ (\ell-1)^{2H+2}\int_0^1 (1-z)z^{2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell-1)^2 z^2}{4}\right)\mathrm{d}z.
\end{aligned}
$$

At this stage, we proceed as for the variance term and apply Gradshteyn and Ryzhik (2007), formula 7.512 (11). This leads to the evaluation

$$
\begin{aligned}
I_2 + I_3 = & \frac{(\ell+1)^{2H+2}}{(2H+1)(2H+2)} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell+1)^2}{4}\right) \\
& - \frac{2\ell^{2H+2}}{(2H+1)(2H+2)} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2\ell^2}{4}\right) \\
& + \frac{(\ell-1)^{2H+2}}{(2H+1)(2H+2)} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell-1)^2}{4}\right).
\end{aligned}
$$

Plugging the terms into the original expression results in the claimed approximation. ■

A.5 Proof of Lemma 2.6

Recall that $\kappa(\ell) = \text{cov}(Y_t, Y_{t+\ell})$, 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 (5), 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

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

where

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

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

A.6 Proof of Theorem 3.6

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 Hansen (1982, Definition 2.2).

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.7 Proof of Theorem 3.7

We introduce the notation:

$$\widetilde{Q}_{n,T}(\theta) = \widetilde{m}_{n,T}(\theta)'\mathbb{W}_T\widetilde{m}_{n,T}(\theta),$$
$$Q(\theta) = m(\theta)'\mathbb{W}m(\theta).$$

where $m(\theta) = G(\theta_0) - G(\theta)$ and $\mathbb{W} = A'A$. 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 that 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 3.6. Now, we pass to the last condition (iv). In view of the Cauchy-Schwarz inequality,

$$\begin{aligned}\left|\widetilde{Q}_{n,T}(\theta) - Q(\theta)\right| \leq{}& \left|(\widetilde{m}_{n,T}(\theta) - m(\theta))'\mathbb{W}_T(\widetilde{m}_{n,T}(\theta) - m(\theta))\right| + \left|m(\theta)'(\mathbb{W}_T + \mathbb{W}_T')(\widetilde{m}_{n,T}(\theta) - m(\theta))\right| \\ &+ \left|m(\theta)'(\mathbb{W}_T - \mathbb{W})m(\theta)\right| \\ \leq{}& ||\widetilde{m}_{n,T}(\theta) - m(\theta)||^2\, ||\mathbb{W}_T|| + 2\, ||m(\theta)||\, ||\mathbb{W}_T||\, ||\widetilde{m}_{n,T}(\theta) - m(\theta)|| \\ &+ ||m(\theta)||^2\, ||\mathbb{W}_T - \mathbb{W}||.\end{aligned}$$

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 $m_T(\theta) = T^{-1}\sum_{t=1}^{T}\left[\mathbb{IV}_t - G(\theta)\right]$. Since the convergence $\sup_{\theta\in\Theta} ||m_T(\theta) - m(\theta)|| \xrightarrow{\mathbb{P}} 0$ was already covered by the proof of Theorem 3.6 (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

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

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.8 Derivation of Remark 3.8

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

$$\sup_{t\in\mathbb{Z}} \mathbb{E}\left[(RV_t^n - IV_t)^2\right] \leq Cn^{-1}.$$

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

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

Consequently,

$$RV_t^n - IV_t = 2\sum_{i=1} \int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}} (X_s - X_{t-1+\frac{i-1}{n}})\mathrm{d}X_s$$

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

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

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

where in the last inequality we exploited $\sup_{s\ge 0}\mathbb{E}[\mu_s^4] + \sup_{s\ge 0}\mathbb{E}[\sigma_s^4] < \infty$ along with Burkholder's inequality: $\sup_{s\in\left[\frac{i-1}{n},\frac{i}{n}\right]}\mathbb{E}[(X_s - X_{\frac{i-1}{n}})^4] \le Cn^{-2}$. ∎

A.9 Proof of Proposition 3.9

The proof relies on a martingale approximation central limit theorem of Peligrad and Utev (2006) and 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,*

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

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

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

Applying Jensen's inequality for conditional expectations,

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

Hence,

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

∎

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

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

Proof. Note that

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

while

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

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

Secondly, we exploit these results to prove the following two technical lemmas that estimate the memory of integrated variance and its noisy proxy.

Lemma A.4 *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) $\left\|\mathbb{E}_{\theta_0}\left[\int_{s+r}^{t+r}\sigma_u^p \mathrm{d}u\int_{s'+r}^{t'+r}\sigma_{u'}^p \mathrm{d}u' \mid \mathcal{F}_0\right]-\mathbb{E}_{\theta_0}\left[\int_s^t \sigma_u^p \mathrm{d}u\int_{s'}^{t'}\sigma_{u'}^p du'\right]\right\|_{L^2(\mathbb{P}_{\theta_0})}=O(r^{-\gamma+1/2})$,

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

$$|K(u)|\le cu^{-\gamma},\quad u\ge u_0. \tag{56}$$

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

$$\begin{aligned}\mathbb{E}_{\theta_0}\left[\int_{s+r}^{t+r}\sigma_u^p \mathrm{d}u\int_{s'+r}^{t'+r}\sigma_{u'}^p \mathrm{d}u' \mid \mathcal{F}_0\right]&-\mathbb{E}_{\theta_0}\left[\int_s^t \sigma_u^p \mathrm{d}u\int_{s'}^{t'}\sigma_{u'}^p \mathrm{d}u'\right]\\ &=\int_s^t\int_{s'}^{t'}\left(\mathbb{E}_{\theta_0}[\sigma_{u+r}^p\sigma_{u'+r}^p\mid\mathcal{F}_0]-\mathbb{E}_{\theta_0}[\sigma_u^p\sigma_{u'}^p]\right)\mathrm{d}u\mathrm{d}u',\end{aligned} \tag{57}$$

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,

$$\mathbb{E}_{\theta_0}[\sigma_{u+r}^p\sigma_{u'+r}^p\mid\mathcal{F}_0]-\mathbb{E}_{\theta_0}[\sigma_u^p\sigma_{u'}^p]=\xi^2 e^{-\frac{p}{2}\kappa(0)}\exp\left(\frac{1}{2}\int_{-\infty}^{\infty}K^+(u,u',\tau)^2\mathrm{d}\tau\right)\big(\exp(Y_r^{u,u'})-1\big), \tag{58}$$

with

$$Y_r^{u,u'} = \int_{-\infty}^0 K^+(u,u',\tau - r)\mathrm{d}B_\tau - \frac{1}{2}\overline{K}^{u,u'}(r) \sim N\Big(-\frac{1}{2}\overline{K}^{u,u'}(r),\, \overline{K}^{u,u'}(r)\Big) \tag{59}$$

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 (57), we conclude that:

$$\begin{aligned}\mathbb{E}_{\theta_0}\bigg[\bigg(\mathbb{E}_{\theta_0}\bigg[\int_{s+r}^{t+r}\sigma_u^p\mathrm{d}u\int_{s'+r}^{t'+r}\sigma_{u'}^p\mathrm{d}u' \mid \mathcal{F}_0\bigg] - \mathbb{E}_{\theta_0}\bigg[\int_s^t\sigma_u^p\mathrm{d}u\int_{s'}^{t'}\sigma_{u'}^p\mathrm{d}u'\bigg]\bigg)^2\bigg]& \\ \le (t-s)(t'-s')\int_s^t\int_{s'}^{t'}\mathbb{E}_{\theta_0}\big[(\mathbb{E}_{\theta_0}[\sigma_{u+r}^p\sigma_{u'+r}^p \mid \mathcal{F}_0] - \mathbb{E}_{\theta_0}[\sigma_u^p\sigma_{u'}^p])^2\big]\mathrm{d}u\mathrm{d}u',&\end{aligned}$$

where, by (58) – (59) and Lemma A.3,

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

after observing that

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

Moreover, if $\tau \le -r$ then $-\tau \ge r \ge u_0$, whereby $u - \tau \ge u_0$ and $u' - \tau \ge u_0$ since $u \ge s \ge 0$ and $u' \ge s' \ge 0$. Thus, by (56),

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

Consequently,

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

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

Lemma A.5 *If Assumptions 1 – 2 and 6 – 7 hold, then*

$$\sum_{r=1}^{\infty} r^{-1/2} \big\| \mathbb{E}_{\theta_0}\Big[\widehat{\mathbb{IV}}_r \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - G_c(\theta_0) \big\|_{L^2(\mathbb{P}_{\theta_0})} < \infty.$$

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

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

where

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

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

$$\begin{aligned}
\big\| \mathbb{E}_{\theta_0}\Big[\widehat{IV}_r | \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - g_0^{(1)}(\theta_0) \big\|_{L^2(\mathbb{P}_{\theta_0})} &= \| \mathbb{E}_{\theta_0}\Big[IV_r \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - \mathbb{E}_{\theta_0}[IV_1] \|_{L^2(\mathbb{P}_{\theta_0})} \\
&\leq \| \mathbb{E}_{\theta_0}\Big[IV_r \mid \mathcal{F}_0^{W,\varepsilon}\Big] - \mathbb{E}_{\theta_0}[IV_1] \|_{L^2(\mathbb{P}_{\theta_0})} \qquad (60) \\
&= O(r^{-\gamma+1/2}), \quad r \to \infty,
\end{aligned}$$

which follows by Lemma A.2, again since $\mathcal{F}_0^{\widehat{\mathbb{IV}}} \subset \mathcal{F}_0^{W,\varepsilon}$, and Lemma A.4(i).

Secondly,

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

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:

$$\begin{aligned}
\big\| \mathbb{E}_{\theta_0}\Big[\widehat{IV}_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - g_0^{(2)}(\theta_0) - c(\theta_0) \big\|_{L^2(\mathbb{P}_{\theta_0})} &= \| \mathbb{E}_{\theta_0}\Big[IV_1^2 \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - \mathbb{E}_{\theta_0}[IV_1^2] \|_{L^2(\mathbb{P}_{\theta_0})} + O(r^{-\gamma+1/2}) \\
&\leq \| \mathbb{E}_{\theta_0}\Big[IV_r^2 \mid \mathcal{F}_0^{W,\varepsilon}\Big] - \mathbb{E}_{\theta_0}[IV_1^2] \|_{L^2(\mathbb{P}_{\theta_0})} + O(r^{-\gamma+1/2}) \\
&= O(r^{-\gamma+1/2}),
\end{aligned} \qquad (61)$$

as $r \to \infty$, by Lemma A.2 and A.4(ii).

Lastly, for $\ell = 1, \ldots, k$ (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

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

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

$$\begin{aligned}\big\|\mathbb{E}_{\theta_0}\Big[\widehat{IV}_r\widehat{IV}_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - g_\ell(\theta_0)\big\|_{L^2(\mathbb{P}_{\theta_0})} &= \|\mathbb{E}_{\theta_0}\Big[IV_rIV_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\Big] - \mathbb{E}_{\theta_0}[IV_1IV_{1-\ell}]\|_{L^2(\mathbb{P}_{\theta_0})} \\ &\leq \|\mathbb{E}_{\theta_0}\Big[IV_rIV_{r-\ell} \mid \mathcal{F}_0^{W,\varepsilon}\Big] - \mathbb{E}_{\theta_0}[IV_1IV_{1-\ell}]\|_{L^2(\mathbb{P}_{\theta_0})} \\ &= O(r^{-\gamma+1/2}),\end{aligned} \tag{62}$$

as $r \to \infty$, due to Lemma A.4(ii).

Combining (60) – (62), we deduce that

$$r^{-1/2}\big\|\mathbb{E}_{\theta_0}\Big[\widehat{\mathbb{IV}}_r \mid \mathcal{F}_1^{\widehat{\mathbb{IV}}}\Big] - G_c(\theta_0)\big\|_{L^2(\mathbb{P}_{\theta_0})} = r^{-1/2}O(r^{-\gamma+1/2}) = O(r^{-\gamma}),$$

as $r \to \infty$, whereby the result follows since $\gamma > 1$. ■

Proof of Proposition 3.9. By the Cramér–Wold device, we can reduce the multivariate convergence in (48) to a univariate problem. To this end, fix arbitrary $a \in \mathbb{R}^{k+2}$ and set $S_T = \sum_{t=1}^T \xi_t$, where $\xi_t = a'\big(\widehat{\mathbb{IV}}_t - G_c(\theta_0)\big)$, $t \in \mathbb{Z}$, so that

$$T^{-1/2}S_T = T^{1/2}a'\widehat{m}_T(\theta_0).$$

Note that $(\xi_t)_{t\in\mathbb{Z}}$ inherits the stationarity and ergodicity of $\big(\widehat{\mathbb{IV}}_t\big)_{t\in\mathbb{Z}}$. Moreover, $\mathbb{E}_{\theta_0}[\xi_1] = 0$ and $\mathbb{E}_{\theta_0}[\xi_1^2] < \infty$ under the present assumptions. As the natural filtration $\mathcal{F}_t^\xi = \sigma\{\xi_t, \xi_{t-1}, \ldots\}$, $t \in \mathbb{Z}$, of $(\xi_t)_{t\in\mathbb{Z}}$ has $\mathcal{F}_t^\xi \subset \mathcal{F}_t^{\widehat{\mathbb{IV}}}$ for any $t \in \mathbb{Z}$, then by Lemmas A.2 and A.5,

$$\sum_{r=1}^\infty r^{-1/2}\|\mathbb{E}[\xi_r \mid \mathcal{F}_0^\xi]\|_{L^2(\mathbb{P}_{\theta_0})} \leq \|a\|_{\mathbb{R}^{k+2}}\sum_{r=1}^\infty r^{-1/2}\big\|\mathbb{E}\big[\widehat{\mathbb{IV}}_r\big|\mathcal{F}_0^{\widehat{\mathbb{IV}}}\big] - G_c(\theta_0)\big\|_{L^2(\mathbb{P}_{\theta_0})} < \infty.$$

Appealing to Theorem 1 and Corollary 2 of Peligrad and Utev (2006),

$$T^{-1/2}S_T \xrightarrow[T\to\infty]{\mathcal{L}} N\bigg(0, \sum_{\ell=-\infty}^\infty \mathbb{E}_{\theta_0}[\xi_1\xi_{1+\ell}]\bigg),$$

with long-run variance

$$\sum_{\ell=-\infty}^\infty \mathbb{E}_{\theta_0}[\xi_1\xi_{1+\ell}] = \sum_{\ell=-\infty}^\infty \mathbb{E}_{\theta_0}\big[a'(\widehat{\mathbb{IV}}_1 - G_c(\theta_0))(\widehat{\mathbb{IV}}_{1+\ell} - G_c(\theta_0))'a\big] = a'\Sigma_{\mathbb{IV}}a,$$

and the proposition is verified. ■

Proposition A.6 *Suppose that Assumptions 1 and 6 hold. Then condition* (ii) *of Assumption 7 applies in Examples 3.2, 3.3, and 3.5.*

Proof. The error terms in Examples 3.2 and 3.5 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 $\mathcal{F}_t^{B,Z} = \sigma\{Z_t, Z_{t-1}, \dots\} \vee \sigma\{B_u : u \leq t\}$, $t \in \mathbb{Z}$:

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

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.4(i) and A.2 to show the conjecture of this part.

In Example 3.3,

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

whereby for any $r \geq 1$,

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

Applying Minkowski's inequality and Lemmas A.4(i) and A.2 concludes the proof. ■

A.10 Proof of Theorem 3.10

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

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

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

$$\begin{aligned}
\nabla_\theta Q_T(\theta) &= 2\nabla_\theta \widehat{m}_T(\theta)' \mathbb{W}_T \widehat{m}_T(\theta), \\
\nabla_{\theta\theta}^2 Q_T(\theta) &= 2\nabla_{\theta\theta} \widehat{m}_T(\theta)' \mathbb{W}_T \widehat{m}_T(\theta) + 2\nabla_{\theta\theta} \widehat{m}_T(\theta)' \mathbb{W}_T \nabla_\theta \widehat{m}_T(\theta)
\end{aligned}$$

This leads to:

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) = \big(\nabla_{\theta\theta}^2 Q_T(\bar{\theta}_T)\big)^{-1} 2\nabla_\theta \widehat{m}_T(\theta_0)' \mathbb{W}_T \sqrt{T} \widehat{m}_T(\theta_0). \tag{63}$$

Invoking the assumptions of the theorem, it follows that $\bar{\theta}_T \xrightarrow{\mathbb{P}} \theta$ as $T \to \infty$. In addition, and recalling Proposition 3.9, we deduce that as $T \to \infty$:

$$\begin{aligned}
\sqrt{T}\widehat{m}_T(\theta_0) &\xrightarrow{d} N\big(0, \Sigma_{\widehat{\mathbb{IV}}}\big), \\
\nabla_\theta \widehat{m}_T(\theta_0) &\xrightarrow{\mathbb{P}} G, \\
\nabla_{\theta\theta}^2 Q_T(\bar{\theta}_T) &\xrightarrow{\mathbb{P}} G'WG,
\end{aligned}$$

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

A.11 Proof of Theorem 3.11

Proceeding as above, and denoting $\widetilde{Q}_{n,T} = \widetilde{m}_{n,T}(\theta)'\mathbb{W}_T\widetilde{m}_{n,T}(\theta)$, we find that

$$\sqrt{T}(\widetilde{\theta}_{n,T} - \theta_0) = \left(\nabla^2_{\theta\theta}\widetilde{Q}_{n,T}(\check{\theta}_{n,T})\right)^{-1} 2\nabla_\theta \widetilde{m}_{n,T}(\theta_0)'\mathbb{W}_T\sqrt{T}\widetilde{m}_{n,T}(\theta_0).$$

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

$$\begin{aligned}
\sqrt{T}\widetilde{m}_{n,T}(\theta_0) &\xrightarrow{d} N(0, \Sigma_{\mathbb{IV}}), \\
\nabla_\theta \widetilde{m}_{n,T}(\theta_0) &\xrightarrow{\mathbb{P}} \widetilde{G}, \\
\nabla^2_{\theta\theta}\widetilde{Q}_{n,T}(\check{\theta}_{n,T}) &\xrightarrow{\mathbb{P}} \widetilde{G}'W\widetilde{G},
\end{aligned}$$

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

References

Alizadeh, S., M. W. Brandt, and F. X. Diebold, 2002, "Range-based estimation of stochastic volatility models," *Journal of Finance*, 57(3), 1047–1092.

Andersen, T. G., and T. Bollerslev, 1997, "Intraday periodicity and volatility persistence in financial markets," *Journal of Empirical Finance*, 4(2), 115–158.

——— , 1998a, "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *International Economic Review*, 39(4), 885–905.

——— , 1998b, "Deutsche Mark-Dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies," *Journal of Finance*, 53(1), 219–265.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61(1), 43–76.

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

Andersen, T. G., and B. E. Sørensen, 1996, "GMM estimation of a stochastic volatility model: A Monte Carlo study," *Journal of Business and Economic Statistics*, 14(3), 328–352.

Asmussen, S., and P. W. Glynn, 2007, *Stochastic Simulation: Algorithms and Analysis*. Springer-Verlag, 1st edn.

Back, K., 1991, "Asset prices for general processes," *Journal of Mathematical Economics*, 20(4), 371–395.

Barndorff-Nielsen, O. E., and A. Basse-O'Connor, 2011, "Quasi Ornstein–Uhlenbeck processes," *Bernoulli*, 17(3), 916–941.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 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., and N. Shephard, 2001, "Non-Gaussian Orstein-Uhlenbeck-based models and some of their uses in financial economics," *Journal of the Royal Statistical Society: Series B*, 63(2), 167–241.

——— , 2002, "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B*, 64(2), 253–280.

——— , 2004, "Power and bipower variation with stochastic volatility and jumps," *Journal of Financial Econometrics*, 2(1), 1–48.

Bayer, C., P. Friz, and J. Gatheral, 2016, "Pricing under rough volatility," *Quantitative 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, Volume II*. University of California Press, Berkeley, pp. 23–33.

Bennedsen, M., 2016, "Semiparametric inference on the fractal index of Gaussian and conditionally Gaussian time series data," Working paper, Aarhus University.

Bennedsen, M., A. Lunde, and M. S. Pakkanen, 2017a, "Decoupling the short- and long-term behavior of stochastic volatility," Working paper, Aarhus University.

Bennedsen, M., A. Lunde, and M. S. Pakkanen, 2017b, "Hybrid scheme for Brownian semistationary processes," *Finance and Stochastics*, 21(4), 931–965.

Bingham, N. H., C. M. Goldie, and J. L. Teugels, 1989, *Regular Variation*. Cambridge University Press, 1st edn.

Bollerslev, T., and H. Zhou, 2002, "Estimating stochastic volatility diffusion using conditional moments of integrated volatility," *Journal of Econometrics*, 109(1), 33–65.

Cheridito, P., H. Kawaguchi, and M. Maejima, 2003, "Fractional Ornstein-Uhlenbeck processes," *Electronic Journal of Probability*, 8(3), 1–14.

Christie, A. A., 1982, "The stochastic behavior of common stock variances: Value, leverage and interest rate effects," *Journal of Financial Economics*, 10(4), 407–432.

Comte, F., L. Coutin, and E. Renault, 2012, "Affine fractional stochastic volatility models," *Annals of Finance*, 8(3), 337–378.

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

Corradi, V., and W. Distaso, 2006, "Semi-parametric comparison of stochastic volatility models using realized measures," *Review of Economic Studies*, 73(3), 635–667.

Delbaen, F., and W. Schachermayer, 1994, "A general version of the fundamental theorem of asset pricing," *Mathematische Annalen*, 300(1), 463–520.

Duffie, D., J. Pan, and K. J. Singleton, 2000, "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, 68(6), 1343–1376.

Duffie, D., and K. J. Singleton, 1993, "Simulated moments estimation of markov models of asset prices," *Econometrica*, 61(4), 929–952.

Dym, H., and H. P. McKean, 1976, *Gaussian processes, function theory, and the inverse spectral problem*. Academic Press, New York.

Euch, O. E., and M. Rosenbaum, 2018, "The characteristic function of rough Heston models," *Mathematical Finance*, 29(1), 3–38.

Fridman, M., and L. Harris, 1998, "A maximum likelihood approach for non-Gaussian stochastic volatility models," *Journal of Business and Economic Statistics*, 16(3), 284–291.

Fukasawa, M., T. Takabatake, and R. Westphal, 2019, "Is volatility rough?," Working paper, Osaka University.

Gallant, A. R., D. A. Hsieh, and G. E. Tauchen, 1997, "Estimation of stochastic volatility with diagnostics," *Journal of Econometrics*, 81(1), 159–192.

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

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

Gradshteyn, I. S., and I. M. Ryzhik, 2007, *Table of integrals, series, and products*. Academic Press, 7th edn.

Guennoun, H., A. Jacquier, P. Roome, and F. Shi, 2018, "Asymptotic behavior of the fractional Heston model," *SIAM Journal of Financial Mathematics*, 9(3), 1017–1045.

Hansen, L. P., 1982, "Large sample properties of generalized method of moments estimators," *Econometrica*, 50(4), 1029–1054.

Hansen, P. R., and A. Lunde, 2006, "Realized variance and market microstructure noise," *Journal of Business and Economic Statistics*, 24(2), 127–161.

——— , 2014, "Estimating the persistence and the autocorrelation function of a time series that is measured with error," *Econometric Theory*, 30(1), 60–93.

Harvey, A. C., E. Ruiz, and N. Shephard, 1994, "Multivariate stochastic variance models," *Review of Financial Studies*, 61(2), 247–264.

Heston, S. L., 1993, "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of Financial Studies*, 6(2), 327–343.

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

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter, 2009, "Microstructure noise in the continuous case: The pre-averaging approach," *Stochastic Processes and their Applications*, 119(7), 2249–2276.

Kaarakka, T., and P. Salminen, 2011, "On fractional Ornstein-Uhlenbeck processes," *Communications on Stochastic Analysis*, 5(1), 121–133.

Karhunen, K., 1950, "Über die Struktur stationärer zufälliger Funktionen," *Arkiv för Matematik*, 1, 141–160.

Lindgren, G., 2006, "Lectures on Stationary Stochastic Processes," Lectures notes, Lund University, available at: `http://www.maths.lth.se/matstat/staff/georg/Publications/lecture2006.pdf`.

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

Maruyama, G., 1949, "The harmonic analysis of stationary stochastic processes," *Memoirs of the Faculty of Science, Kyūsyū University, Series A*, 4, 45–106.

Meddahi, N., 2002, "A theoretical comparison between integrated and realized volatility," *Journal of Applied Econometrics*, 17(5), 479–508.

——— , 2003, "ARMA representation of integrated and realized variances," *Econometrics Journal*, 6(2), 335–356.

Melino, A., and S. M. Turnbull, 1990, "Pricing foreign currency options with stochastic volatility," *Journal of Econometrics*, 45(1-2), 239–265.

Newey, W. K., and D. McFadden, 1994, "Large sample estimation and hypothesis," *Handbook of Econometrics, IV, Edited by R. Engle and D. McFadden*, pp. 2112–2245.

Newey, W. K., and K. D. West, 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," *Journal of Econometrics*, 160(1), 246–256.

Peccati, G., and M. S. Taqqu, 2011, *Wiener chaos: moments, cumulants and diagrams*. Springer, Milan.

Peligrad, M., and S. Utev, 2006, "Central limit theorem for stationary linear processes," *Annals of Probability*, 34(4), 1608–1622.

Taqqu, M. S., 1975, "Weak convergence to fractional Brownian motion and to the Rosenblatt process," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 31, 287–302.

Taylor, S. J., 1986, *Modelling Financial Time Series*. John Wiley & Sons, 1 edn.

Tegnér, M., and R. Poulsen, 2018, "Volatility is log-normal – but not for the reason you think," *Risks*, 6(2), –.

Todorov, V., 2009, "Estimation of continuous-time stochastic volatility models with jumps using high-frequency data," *Journal of Econometrics*, 148(2), 131–148.

Todorov, V., and G. Tauchen, 2011, "Volatility jumps," *Journal of Business and Economic Statistics*, 29(3), 356–371.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia, 2005, "A tale of two time scales: determining integrated volatility with noisy high-frequency data," *Journal of the American Statistical Association*, 100(472), 1394–1411.