The realized empirical distribution function of stochastic variance with application to goodness-of-fit testing[*]

Kim Christensen[†‡]   Martin Thyrsgaard[†]   Bezirgen Veliyev[†‡]

April 2019

Abstract

We propose a nonparametric estimator of the empirical distribution function (EDF) of the latent spot variance of the log-price of a financial asset. We show that over a fixed time span our realized EDF (or REDF)—inferred from noisy high-frequency data—is consistent as the mesh of the observation grid goes to zero. In a double-asymptotic framework, with time also increasing to infinity, the REDF converges to the cumulative distribution function of volatility, if it exists. We exploit these results to construct some new goodness-of-fit tests for stochastic volatility models. In a Monte Carlo study, the REDF is found to be accurate over the entire support of volatility. This leads to goodness-of-fit tests that are both correctly sized and relatively powerful against common alternatives. In an empirical application, we recover the REDF from stock market high-frequency data. We inspect the goodness-of-fit of several two-parameter marginal distributions that are inherent in standard stochastic volatility models. The inverse Gaussian offers the best overall description of random equity variation, but the fit is less than perfect. This suggests an extra parameter (as available in, e.g., the generalized inverse Gaussian) is required to model stochastic variance.

JEL Classification: C10; C50.

Keywords: Empirical processes; goodness-of-fit; high-frequency data; microstructure noise; pre-averaging; realized variance; stochastic volatility.

---

1 Introduction

Stochastic volatility is a central concept in financial economics with numerous implications for capital allocation, asset- and derivatives pricing, or risk management. It is therefore essential to select a model for stochastic volatility that is able to reproduce the main features observed in data from financial markets, such as the implied volatility surface, as closely as possible. There is, however, no consensus about which model is "best," as the extensive list of papers on stochastic volatility suggests (the literature is too big to enumerate, but a partial and necessarily incomplete set of proposed specifications can be found in Andersen, Benzoni, and Lund, 2002; Barndorff-Nielsen and Shephard, 2002; Chernov, Gallant, Ghysels, and Tauchen, 2003; Christoffersen, Jacobs, and Mimouni, 2010; Comte and Renault, 1998; Gatheral, Jaisson, and Rosenbaum, 2018; Heston, 1993; Hull and White, 1987, and the references therein).

The goodness-of-fit of a stochastic volatility model can be assessed in several ways. If the parameters are estimated with a method of moment-based estimator, a standard diagnostic check is available via a test of overidentifying restrictions (see also, e.g., Gallant, Hsieh, and Tauchen, 1997). This approach can be highly inefficient, however, as it depends heavily on the selection of moment conditions and the associated estimation of the weighting matrix, which may produce significant size distortions in finite samples (Andersen and Sørensen, 1996).

Another idea is to compare the model-implied distribution of volatility with a nonparametric empirical measure of it (e.g., Aït-Sahalia, 1996). In a continuous-time setting, for instance, a model is often formulated via a stochastic differential equation for the variance process (or a transformation thereof, such as the square root or natural logarithm). This, in turn, implies that the marginal distribution of spot volatility (at least in the stationary case) belongs to a particular class of distributions. So if the underlying volatility was observed, a specification test based on the distance—with respect to a suitable norm—between the empirical distribution function (EDF) and the marginal distribution imposed by the model would be feasible. However, as spot volatility is not directly observed, this approach is not immediately applicable.

The recent access to financial high-frequency data has alleviated these concerns, as it enables the computation of an error-free measure of realized volatility, allowing for a more direct evaluation of the goodness-of-fit of stochastic volatility models. Inspired by the above, a common approach is to compare (conditional) moments of the integrated variance of a parametric model with a nonparametric estimator hereof (see, e.g., Bollerslev and Zhou, 2002; Corradi and Distaso, 2006; Dette and Podolskij, 2008; Dette, Podolskij, and Vetter, 2006; Todorov, 2009; Todorov, Tauchen,

and Grynkiv, 2011; Vetter and Dette, 2012). Zu (2015) proposes a test based on a de-convolution kernel density estimator of the distribution of the integrated variance, Lin, Lee, and Guo (2013, 2016) resort to a characteristic function approach, while Bull (2017) proposes a wavelet-based test. As pointed out by Todorov and Tauchen (2012), however, the mapping between the probability distribution of the spot and integrated variance is in general not one-to-one. Thus, goodness-of-fit tests based on the latter suffer from lack of power against some alternatives, due to the smoothing entailed by integrating spot volatility over a discrete time interval.

In this article, we therefore construct goodness-of-fit tests for stochastic volatility models that are based on a realized EDF (REDF hereafter) of spot volatility. We build on Li, Todorov, and Tauchen (2013, 2016), who show that inference about the EDF can be based on the volatility occupation measure (e.g., Geman and Horowitz, 1980). In their framework, the latent spot variance at any point in time is retrieved by employing a Foster and Nelson (1996) rolling window-type estimator computed on small non-overlapping blocks of high-frequency data. In contrast, we employ overlapping blocks, which is theoretically more efficient, but it also induces further technical problems in the part of the proofs, where a uniform bound on the estimation errors of spot variance is derived.

In addition, we allow the asset price to be recorded at the tick-by-tick level, so that the instantaneous variance can be recovered as accurately as possible. This is important if the whole distribution, including the tails, is of interest and not only measures of central tendency. At that sampling frequency, however, the data are distorted by microstructure noise (see, e.g. Hansen and Lunde, 2006), and this complicates the issue. The derived problem of estimating the integrated—or cumulative—variance in the presence of noise has received a lot of attention in the literature (see, e.g. Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008; Jacod, Li, Mykland, Podolskij, and Vetter, 2009; Zhang, Mykland, and Aït-Sahalia, 2005). We here adapt the pre-averaging approach of Jacod, Li, Mykland, Podolskij, and Vetter (2009) and Podolskij and Vetter (2009a,b) to design a consistent spot volatility estimator from contaminated high-frequency data (Zu and Boswijk, 2014, suggest an alternative procedure). The REDF is then constructed from the recovered sample path of spot variance.

We show that under weak regularity conditions the REDF is a consistent noise- and jump-robust estimator of the EDF over any fixed time interval in the infill limit, thus extending Li, Todorov, and Tauchen (2013, 2016) to a setting with overlapping blocks and microstructure noise. This is merely a stepping stone in our context, however, as the main goal is to do goodness-of-fit tests for stochastic volatility models. In the second part, we therefore prove a functional CLT for the

2

REDF in a long-span asymptotic framework by letting time increase to infinity. This amounts to establish convergence of what van der Vaart and Wellner (2007) call an empirical process indexed by an estimated function, but here time is continuous and the observations are not i.i.d.

The theoretical results are refined into two new goodness-of-fit tests for measuring the discrepancy between the marginal density implied by a candidate stochastic volatility model and the nonparametric gauge at the EDF represented by the REDF. The first is reminiscent to a Kolmogorov-Smirnov statistic for an observed process, while the second is based on a weighted L^2 norm. Both are trivial to compute once the REDF has been constructed, but there are a number of subtleties that render the asymptotic distribution hard to evaluate, making it difficult to set critical regions and determine p-values. This is to some extent related to the complexities encountered in a classical setup, when the parameters of the model under the null are estimated. We resolve the issue via a parametric bootstrap (based on, e.g., Bull, 2017, and described in Appendix B), which leads to fast evaluation of the t-statistic.

The paper is organized as follows. In Section 2, we cover the setting, assumptions, and introduce the EDF of volatility. In Section 3, the properties of the REDF are analysed. The asymptotic theory for our goodness-of-fit tests is also developed here. Section 4 is devoted to a Monte Carlo study of the REDF and an assessment of the proposed goodness-of-fit tests. An empirical application is conducted in Section 5, where we recover the REDF from stock market high-frequency data and test goodness-of-fit of several marginal distributions that are induced by mainstream stochastic volatility models. We conclude in Section 6. The proofs appear in Appendix A.

2   The setting

Let $X = (X_t)_{t \geq 0}$ denote the efficient log-price process of a financial asset, which is defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P} \right)$. As consistent with no-arbitrage (e.g., Delbaen and Schachermayer, 1994), we assume X is an Itô semimartingale:

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \quad \text{for} \quad t \in [0, T], \tag{2.1}$$

where X_0 is $\mathcal{F}_0$-measurable, $(b_t)_{t \geq 0}$ is a locally bounded, predictable drift, $(\sigma_t)_{t \geq 0}$ is an adapted, càdlàg volatility, $(W_t)_{t \geq 0}$ a Brownian motion, and $(J_t)_{t \geq 0}$ is a jump process:

$$J_t = \int_0^t \int_{\mathbb{R}} \delta(s, z) 1_{\{|\delta(s,z)| \leq 1\}} (\mu - \nu)(d(s, z)) + \int_0^t \int_{\mathbb{R}} \delta(s, z) 1_{\{|\delta(s,z)| > 1\}} \mu(d(s, z)), \tag{2.2}$$

where μ is a Poisson random measure on $(\mathbb{R}_+, \mathbb{R})$, $\nu(d(s, z)) = ds \otimes \lambda(dz)$ is a compensator, λ is a σ-finite measure on $\mathbb{R}$, and $\delta : \Omega \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ is a predictable function. Furthermore, we assume

that there exists a function Γ such that $|\delta(\omega, s, z)| \leq \Gamma(z)$ for all (ω, s, z) and

$$\int_{\mathbb{R}} (1 \wedge \Gamma(z))^r \lambda(dz) < \infty, \tag{2.3}$$

for some $r \in [0, 2]$.

The constant r in (2.3) controls the activity level of the price jumps and is an upper bound on the Blumenthal-Getoor index. When $r = 0$, the jumps are of finite activity, whereas $r > 0$ corresponds to the infinite-activity setting.

We collect assumptions about the drift, volatility, and jumps of X below.

Assumption 1 *We assume that for each $p \geq 1$ and $t \geq 0$, $\mathbb{E}\big[|b_t|^p\big] + \mathbb{E}\big[|\sigma_t|^p\big] + \int_{\mathbb{R}} (1 \vee \Gamma(z))^p \lambda(dz) \leq C$ for some constant C. In addition, σ fulfils at least one of the following conditions:*

(i) There exists $H \in (0, 1)$ such that for each $p \geq 1, s \geq 0, u > 0$, and a constant C (dependent on H and p):

$$\mathbb{E}\left[\left(\sup_{s \leq t \leq s+u} |\sigma_t - \sigma_s|\right)^p\right] \leq Cu^{pH}. \tag{2.4}$$

(ii) σ has the representation:

$$\sigma_t = \sigma_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \tilde{\sigma}_s' dB_s + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s, z)(\mu - \nu)(d(s, z)), \tag{2.5}$$

where σ_0 is $\mathcal{F}_0$-measurable, $\mathbb{E}\big[|\tilde{b}_t|^p\big] + \mathbb{E}\big[|\tilde{\sigma}_t|^p\big] + \mathbb{E}\big[|\tilde{\sigma}_t'|^p\big] \leq C$ for some constant C and for each $p \geq 1$ and $t \geq 0$, while B is a Brownian motion independent of W, $\tilde{\delta}$ is a predictable function, and $|\tilde{\delta}(\omega, t, z)| \wedge 1 \leq \tilde{\Gamma}(z)$ for all (ω, t, z) and some $\tilde{\Gamma} : \mathbb{R} \to \mathbb{R}$ such that $\int_{\mathbb{R}} \tilde{\Gamma}(z)^2 \lambda(dz) < \infty$.

This places some moment conditions on b_t, σ_t, and the large jumps. While it is general and encompasses many volatility models applied in practice, it may feel a bit restrictive. The assumption can be relaxed, but this entails that a constant $\iota > 0$ (restricting the rate at which $T \to \infty$ and $\Delta_n \to 0$, cf. Section 3) has to be bounded from below, while it can be made arbitrarily small within the above setup.

Assumption 1(i) implies σ is continuous. This means it can be approximated by a locally constant process over short time intervals. The condition covers both "rough" and long-memory volatility processes (e.g., Comte and Renault, 1998; Gatheral, Jaisson, and Rosenbaum, 2018) with Hurst exponent different from one-half.

The conditions in Assumption 1(ii), which says σ is an Itô semimartingale, are common in the literature. It allows volatility to exhibit fairly unrestricted jump dynamics (at the expense of

excluding fractional Brownian motion as in 1(i)) and also enables a leverage effect, as it does not restrict the correlation structure between the increments of X and σ.

Next, we define the EDF of volatility, which in the continuous-time setting is given by the random function:

$$F_T(x) = \frac{1}{T} \int_0^T 1_{\{V_t \leq x\}} dt, \tag{2.6}$$

where $V_t \equiv \sigma_t^2$.[1]

Assumption 2 F_T *is a.s. continuous.*

This condition is fulfilled by most stochastic volatility models. It does not impose the existence of a density nor does it require σ itself to be continuous.

The right-hand side of (2.6) can be written $T^{-1} \int_0^T g(V_t)dt$, where $g(v) = 1_{\{v \leq x\}}$.[2] Then, if F_T is absolutely continuous with respect to the Lebesgue measure—which is stronger than Assumption 2—and g is a bounded or non-negative Borel function:

$$\frac{1}{T} \int_0^T g(V_t)dt = \frac{1}{T} \int_{\mathbb{R}_+} g(x) dF_T(x) = \frac{1}{T} \int_{\mathbb{R}_+} g(x) f_T(x) dx, \tag{2.7}$$

where f_T is the density of F_T with respect to the Lebesgue measure. The EDF therefore encapsulates all the information about $(V_t)_{t \geq 0}$ available in $[0, T]$ and can be viewed as a pathwise version of the distribution function of volatility.

2.1   A discrete and noisy high-frequency record of X

The main difficulty is that F_T is latent, because σ is not observable. The aim of this paper is therefore to construct a consistent estimator of F_T, while making the above minimal assumptions about X.

Of course, in an ideal world with no markets frictions and continuous trading—i.e. if the entire trajectory of X is available—we can recover the volatility process perfectly and also the exact time and size of jumps in X and σ (in constrast to the drift, which cannot be consistently estimated

---

[1]In probability theory, $F_T(x) = \int_0^T 1_{\{Y_t \leq x\}} dt$ is called the occupation—or local—time of the stochastic process Y (e.g., Geman and Horowitz, 1980). This convention was adopted by Li, Todorov, and Tauchen (2013, 2016), who applied it to high-frequency volatility estimation. We normalize F_T by T here, as we are heading toward a setting with stationary volatility and an asymptotic theory with $T \to \infty$.

[2]Much work in the high-frequency literature studies integrals of the form $\int_0^T g(V_t)dt$, where g is smooth (e.g., Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard, 2006; Jacod and Rosenbaum, 2013). Here, in contrast, g is discontinuous, which makes the theory a lot more inaccessible.

in finite time, not even if it is constant, see, e.g., Merton, 1980; Foster and Nelson, 1996). This setup is not realistic, however. In practice, we operate with discrete high-frequency data, which we assume are available at times $i\Delta_n$, for $i = 0, 1, \ldots, n$, where Δ_n is the time gap between consecutive observations and $n = \lfloor T/\Delta_n \rfloor$ is the sample size (equidistant sampling is enforced here, but it can be weakened). In a near-ideal world, the database then constitutes a high-frequency record of X, i.e. $(X_{i\Delta_n})_{i=0}^n$, and the estimator proposed by Li, Todorov, and Tauchen (2013, 2016) can be applied without further ado.

The microstructure of financial markets adds measurement error, however, which implies that we do not observe X directly (e.g., due to bid-ask bounce or price discreteness). Instead, we record a contaminated log-price Z, which we assume is related to X as follows:

$$Z_{i\Delta_n} = X_{i\Delta_n} + U_{i\Delta_n}, \quad \text{for} \quad i = 0, 1, \ldots, nT. \tag{2.8}$$

Assumption 3 *We have $U_{i\Delta_n} = \omega_{i\Delta_n}\varepsilon_{i\Delta_n}$, where $(\omega_t)_{t\geq0}$ is a càdlàg stochastic process that is adapted to $\mathcal{F}$, and $(\varepsilon_{i\Delta_n})_{i\geq0}$ is i.i.d. with mean zero, variance one, and $\mathbb{E}(|\varepsilon_{i\Delta_n}|^p) < \infty$ for every $p > 0$ and $i \geq 0$. In addition, $(\varepsilon_{i\Delta_n})_{i\geq0}$ is independent of $\mathcal{F}$.*

This setting is comparable to Assumption (K) in Jacod, Li, Mykland, Podolskij, and Vetter (2009) or Assumption (4) in the Appendix of Li, Todorov, and Tauchen (2017) (see, e.g., Bibinger and Winkelmann, 2018, for a related approach). The noise is permitted to be both heteroscedastic and serially dependent via $(\omega_t)_{t\geq0}$, so the model is broad enough to replicate realistic structures in the microstructure of financial markets (e.g., Diebold and Strasser, 2013; Hansen and Lunde, 2006). The moment condition is not standard, however, but it is merely made for technical convenience. In the proofs, we only require moments of $\varepsilon_{i\Delta_n}$ up to some order (for instance, in Lemma 2.1 the 4th moment should exist).

Later, we impose a stronger smoothness assumption on $(\omega_t)_{t\geq0}$:

Assumption 4 *Assumption 3 is satisfied. In addition, for each $p \geq 1$, $s \geq 0$, and $u > 0$:*

$$\mathbb{E}\left[\left(\sup_{s\leq t\leq s+u} |\omega_t - \omega_s|\right)^p\right] \leq Cu^p. \tag{2.9}$$

and $\mathbb{E}\left[|\omega_t|^p\right] \leq C$, for some constant C.

We define the noisy log-return:

$$\Delta_i^n Z = Z_{i\Delta_n} - Z_{(i-1)\Delta_n}, \quad \text{for} \quad i = 1, \ldots, nT. \tag{2.10}$$

To deal with the noise, we pre-average the return series (see, e.g., Jacod, Li, Mykland, Podolskij, and Vetter, 2009; Podolskij and Vetter, 2009a,b):

$$\bar{Z}_i = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) \Delta_{i+j}^n Z, \quad \text{for} \quad i = 0, \ldots, Tn - k_n + 1, \tag{2.11}$$

where k_n is a positive integer, while g is a kernel.

Any $g : [0,1] \mapsto \mathbb{R}$ with g continuous and piecewise C^1 with Lipschitz derivative g', such that $g(0) = g(1) = 0$ and $\int_0^1 g(x)^2 dx > 0$ is permitted. We follow the default choice in the literature by setting $g(x) = \min(x, 1-x)$.

The selection of k_n entails a trade-off. The intuition is that while pre-averaging lessens the noise, it also smooths out the underlying volatility of X. This can render it hard to construct estimates of spot volatility that fit the tails of the distribution if a too wide pre-averaging window is applied. As shown by Jacod, Li, Mykland, Podolskij, and Vetter (2009), an optimal k_n is achieved via:

$$k_n = \frac{\theta}{\sqrt{\Delta_n}} + o(\Delta_n^{-1/4}), \tag{2.12}$$

where $\theta > 0$ is a tuning parameter, which controls the balance struck between the above forces in small samples. As consistent with prior work, we base our analysis on $\theta = 1/3$ and $k_n = [\theta/\sqrt{\Delta_n}]$.

2.2 The realized EDF

We now exploit the pre-averaged high-frequency data to form local estimates of V_t. The estimator we propose is both jump- and noise-robust and can therefore be plugged into a "realized" version of the EDF (or REDF, as defined in (2.18)). It is computed on small blocks of pre-averaged returns. We denote the number of increments in a block by a sequence of positive integers h_n with $h_n \Delta_n \to 0$ and $h_n/k_n \to \infty$, as $\Delta_n \to 0$, so the time span of the block is decreasing, but there is an increasing amount of data within it.

Then, we set:

$$\tilde{V}_{i\Delta_n} = \frac{1}{h_n\sqrt{\Delta_n}} \sum_{m=0}^{h_n-1} \left(\bar{Z}_{i+m}\right)^2 \mathbf{1}_{\{|\bar{Z}_{i+m}| \leq v_n\}}, \quad \text{for} \quad i = 0, 1, \ldots, Tn - h_n - k_n + 1, \tag{2.13}$$

where $v_n = \alpha \Delta_n^{\bar{\omega}}$, for some $\bar{\omega} \in (0, 1/4)$ and $\alpha > 0$.

As the notation suggests, $\tilde{V}_{i\Delta_n}$ is an estimator of the spot variance at time $i\Delta_n$, i.e. $V_{i\Delta_n}$. It is convenient to extend this definition to $t \in [0, T]$. We do this by setting $\tilde{V}_t = \tilde{V}_{i\Delta_n}$, for $t \in [i\Delta_n, (i+1)\Delta_n)$, while for $t > (Tn - h_n - k_n + 1)\Delta_n$: $\tilde{V}_t = \tilde{V}_{(Tn-h_n-k_n+1)\Delta_n}$, thus holding the final spot variance estimate fixed up to time T.

Finally, we introduce the following constants, which depend on the weight function:

$$\psi_1 = \int_0^1 g'(s)^2 ds \quad \text{and} \quad \psi_2 = \int_0^1 g(s)^2 ds.^3 \tag{2.14}$$

With this notation in hand, we have the following result for our preliminary estimator of the spot variance.

Lemma 2.1 *Suppose that Assumption 3 holds and $r \in [0, 2]$. For each $t \in [0, T]$, as $\Delta_n \to 0$,*

$$\tilde{V}_t \xrightarrow{\mathbb{P}} \theta \psi_2 V_t + \frac{1}{\theta} \psi_1 \omega_t^2. \tag{2.15}$$

As an aside, we should point out it is also possible to show that $\tilde{V}_{i\Delta_n}$ retains its consistency even in absence of truncation. However, in order to derive rates of convergence, which play an important role in the subsequent theory, truncation is required when X is discontinuous. We therefore do not pursue this idea further.[4]

To strip out the residual noise variation, we need an estimator of ω_t^2. There are several options around (e.g., Bandi and Russell, 2006; Hansen and Lunde, 2006; Oomen, 2006) and each one has its own merits and disadvantages (see, e.g., Gatheral and Oomen, 2010, for a comparison). In this paper, we adopt the following:

$$\hat{\omega}_{i\Delta_n}^2 = -\frac{1}{h_n - 1} \sum_{m=1}^{h_n - 1} \Delta_{i+m}^n Z \Delta_{i+m+1}^n Z, \tag{2.16}$$

which is a consistent estimator of $\omega_{i\Delta_n}^2$ (as demonstrated in the proof of Theorem 3.1). If we define $\hat{\omega}_t^2 = \hat{\omega}_{i\Delta_n}^2$ for $t \in [i\Delta_n, (i+1)\Delta_n)$, it then follows that

$$\hat{V}_t \equiv \frac{1}{\theta \psi_2} \tilde{V}_t - \frac{\psi_1}{\theta^2 \psi_2} \hat{\omega}_t^2 \xrightarrow{\mathbb{P}} V_t. \tag{2.17}$$

Subtracting the bias implies $\hat{V}_t$ can be negative in finite samples. Nevertheless, not a single point estimate fell below zero neither in our simulations nor empirical work.

We then construct the REDF:

$$F_{n,T}(x) = \frac{1}{T} \int_0^T 1_{\{\hat{V}_t \le x\}} dt, \tag{2.18}$$

---

[3] In the proofs, after freezing volatility locally, $\psi_1^n = k_n \sum_{j=1}^{k_n} \left(g(\frac{j}{k_n}) - g(\frac{j-1}{k_n}) \right)^2$ and $\psi_2^n = \frac{1}{k_n} \sum_{j=1}^{k_n-1} g^2(\frac{j}{k_n})$ appear in the conditional expectation of $\hat{V}_{i\Delta_n}$. As $\Delta_n \to 0$, $\psi_i^n \to \psi_i$ and the error has order $\psi_i^n - \psi_i = O(\Delta_n^{1/2})$, for $i = 1, 2$. This means we can work with ψ_i in the expressions below, and this substitution has no impact on the asymptotic analysis. In contrast, ψ_i^n can differ materially from ψ_i if k_n is small. As a practical recommendation, it is therefore better to work with ψ_i^n.

[4] The corresponding analysis of rates of convergence with the non-truncated spot variance estimator and continuous X appeared in an earlier working paper version of this article (available at request).

3 Asymptotic properties of the REDF

In this section, we cover the asymptotic theory of the REDF. In Section 3.1, we deal first with consistency for the EDF in the infill setting, where we assume that the noisy log-price is recorded over ever shorter intervals but the total time elapsed is constant, as formalized by $\Delta_n \to 0$ with T fixed. In the following Section 3.2, we then further assume high-frequency data are collected on an expanding time window by also letting $T \to \infty$ and deduce convergence toward the marginal distribution function. We exploit these theoretical insights to develop some new goodness-of-fit tests for stochastic volatility models in Section 3.3.

3.1 Infill setting

We start this section with showing that $F_{n,T}$ is a consistent estimator of the EDF.

Theorem 3.1 *Suppose that Assumptions 2 – 3 hold and $r \in [0,2]$. Then, for each $T > 0$, as $\Delta_n \to 0$:*

$$\sup_{x \in \mathbb{R}_+} \left| F_{n,T}(x) - F_T(x) \right| \xrightarrow{\mathbb{P}} 0. \tag{3.1}$$

The consistency of $F_{n,T}$ is derived directly from that of the spot volatility estimator and the continuity of the EDF. It thus inherits the jump robustness of the estimator in (2.13), regardless of the activity level of jumps in X.

We define $Q_{n,T}(\alpha) = \inf \left\{ x : F_{n,T}(x) \geq \alpha \right\}$ and $Q_T(\alpha) = \inf \left\{ x : F_T(x) \geq \alpha \right\}$ as the uniquely determined α-quantile of $F_{n,T}$ and F_T, for $\alpha \in (0,1)$. Then, using the uniform convergence in probability of $F_{n,T}$, we deduce $Q_{n,T}(\alpha)$ is a consistent estimator of $Q_T(\alpha)$. As it is an implication of Lemma 21.2 in van der Vaart (1998), we omit a proof.

Corollary 3.2 *Suppose that Assumptions 2 – 3 hold and $r \in [0,2]$. Then, for each $T > 0$, if $Q_T(\alpha)$ is continuous at α and as $\Delta_n \to 0$:*

$$Q_{n,T}(\alpha) \xrightarrow{\mathbb{P}} Q_T(\alpha). \tag{3.2}$$

3.2 Joint infill and long-span setting

The goal here is to extend the above analysis (with T fixed) to estimation of the stationary distribution function, F. To do so, we study an asymptotic framework, where the time span $T \to \infty$ jointly with $\Delta_n \to 0$. The key result is a feasible CLT for the REDF, which enables us to compute the accuracy with which the marginal distribution can be recovered from the data. To facilitate the derivation of this theory we require some additional assumptions on the volatility process.

Assumption 5 $(V_t)_{t\geq 0}$ *is stationary and strongly mixing with mixing coefficient function* $\alpha(t) = O(t^{-\gamma})$, *as* $t \to \infty$ *and for some* $\gamma > 1$.

The stationarity condition in Assumption 5 is required to make the target of inference well-defined. We then restrict the memory of the volatility, so that $(V_t)_{t\geq 0}$ is not "too" strongly dependent. This implies that F_T (as an estimator of the marginal distribution, F) is consistent, as $T \to \infty$. We exploit this to deduce that $F_{n,T}$ also converges in probability to F, once we show that the discretization error embedded in the recovery of the volatility path is asymptotically negligible (as $\Delta_n \to 0$ fast enough, because the errors accumulate with T). Moreover, it is an essential part in showing the weak convergence of the empirical processes that we construct in the derivation of our goodness-of-fit test statistics in Section 3.3.

The mixing condition is slightly weaker than the comparable assumptions made in related work with joint infill and long-span asymptotics (see, e.g., Todorov and Tauchen, 2012; Andersen, Thyrsgaard, and Todorov, 2019). It encompasses several stochastic volatility models, for instance a large class of processes driven by Brownian motion (e.g., Heston, 1993) or the Lévy-driven Ornstein-Uhlenbeck model of Barndorff-Nielsen and Shephard (2001), where volatility is governed by general (positive) processes.[5,6] As a result, the setup is not restrictive in practice as it can capture a wide variety of marginal distributions.

Next, we replace the pathwise smoothness imposed on the EDF of volatility from Assumption 2 with the following condition.

Assumption 6 F *is differentiable with bounded derivative* f.

Theorem 3.3 *Suppose that Assumptions 4 – 6 hold true and* $r \in [0, 2)$. *If either of the following conditions is fulfilled:*

(i) *Assumption 1(i) with* $H \in (0, 1)$, $h_n \asymp \Delta_n^{-\frac{4H+1}{4H+2}}$ *and* $T^{1/2+\iota}\left(\Delta_n^{\frac{H}{4H+2}-\iota} \vee \Delta_n^{(2-r)\bar{\omega}-\iota}\right) \to 0$ *as* $\Delta_n \to 0$ *and* $T \to \infty$ *for some* $\iota > 0$, $\bar{\omega} \in \left(\frac{H+1}{8H+4}, \frac{1}{4}\right)$, *or*

---

[5]Assume that V is a solution of the stochastic differential equation: $dV_t = \tilde{b}(t, V_t)dt + \tilde{\sigma}(t, V_t)dW_t$, where $\tilde{\sigma}(t, V_t) \leq K\left(1 + |V_t|^{1/2}\right)$, for some $K > 0$. If there exists an S such that for all $|s| \geq S$ and $t \geq 0$, $b(t, s) \leq -\gamma$, for some $\gamma > 0$, Theorem 2 in Veretennikov (1988) implies that $\alpha(t) = O(t^{-\gamma})$.

[6]If V is of the form $dV_t = -\kappa V_t dt + dZ_{\kappa t}$, where $\kappa > 0$ and Z is a positive Lévy process (e.g., a subordinator) with Lévy measure ν, then it follows from Jongbloed, van der Meulen, and van der Vaart (2005) that V is stationary if $\int_2^\infty \ln(x)\nu(dx) < \infty$. If further $\mathbb{E}\left[|V_1|^p\right] < \infty$, for some $p > 0$, there exists $a > 0$ such that $\alpha(t) = O(e^{-at})$. The decay of the mixing coefficient can even be deduced if the driving process in volatility is a fractional Brownian motion (see, e.g., Magdziarz and Weron, 2011).

(ii) *Assumption 1(ii), $h_n \asymp \Delta_n^{-3/4}$ and $T^{1/2+\iota}\left(\Delta_n^{1/8-\iota} \vee \Delta_n^{(2-r)\bar{\omega}-\iota}\right) \to 0$ as $\Delta_n \to 0$ and $T \to \infty$ for some $\iota > 0$, $\bar{\omega} \in \left(\frac{3}{16}, \frac{1}{4}\right)$.*

Then, for fixed $x \in \mathbb{R}_+$, it holds that

$$\sqrt{T}\big(F_{n,T}(x) - F(x)\big) \xrightarrow{d} N(0, \Sigma(x)), \tag{3.3}$$

where $\Sigma(x) = 2\int_0^\infty \big(F_t(x,x) - F(x)^2\big)dt$ with $F_t(x,y) = \mathbb{P}\left(V_0 \le x, V_t \le y\right)$.

In Theorem 3.3, the block length h_n is selected so that the discrepancy between the empirical process associated with the REDF and EDF vanishes at the fastest rate. To control this error, we need to restrict the relative speed at which $T \to \infty$ and $\Delta_n \to 0$.

The first restriction is due to the untangling of the spot variance and Brownian component. The least binding is to take h_n as indicated. Note that h_n is an increasing function of H. In particular, if σ is a continuous semimartingale (i.e., $H = 1/2$), the contribution of this error to $\sqrt{T}\big(F_{n,T}(x) - F_T(x)\big)$ has order $T^{1/2+\iota}\Delta_n^{1/8-\iota}$. With a smoother volatility ($H \to 1$), the rate of convergence improves to $T^{1/2+\iota}\Delta_n^{1/6-\iota}$, whereas it gets arbitrarily slow as $H \to 0$. The intuition is that it is virtually impossible to retrieve spot variance from a "rough" path. On the other hand, if $H = 1/2$ the optimal choice of h_n and the associated speed condition is unchanged, irrespective of whether volatility jumps or not.

The second is due to the presence of jumps in X. It gets more restrictive, the higher r is. Indeed, if $r > \frac{2}{H+1}$ this is the leading term, but for that to happen, the price jumps must not only be of infinite activity, but also infinite variation (i.e., $r > 1$).

The noise-robustness of $F_{n,T}$ plays a critical role in determining the above condition. To compare, if there was no error in our measurement of X, a standard non-noise-robust realized measure suffices for estimation of spot volatility (see, e.g., Jacod and Protter, 2012). Here, the rate condition improves to $T^{1/2+\iota}\Delta_n^{1/4-\iota} \to 0$ with $H = 1/2$, as is consistent with the fixed T asymptotic theory in Li, Todorov, and Tauchen (2013). The deterioration of the speed condition is due to the added complexity from retrieving volatility robustly, which cuts the rate of convergence in half. This problem is well-known in the literature on estimation of integrated variance (see, e.g., Zhang, Mykland, and Aït-Sahalia, 2005; Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008; Jacod, Li, Mykland, Podolskij, and Vetter, 2009). Moreover, while a convergence rate of (almost) $\Delta_n^{1/8}$ may appear slow, it is optimal for extraction of spot variance in noisy diffusion models (e.g., Aït-Sahalia and Jacod, 2014, p. 296).

In addition, note that the error of the spot noise variance estimator in (2.16) is always negligible. For example, in Theorem 3.3(i) $|\hat{\omega}_t^2 - \omega_t^2| = O_p\big(h_n^{-1/2} \vee h_n\Delta_n\big)$, where $h_n \asymp \Delta_n^{-\frac{4H+1}{4H+2}}$, which is smaller than the spot variance estimation error $|\hat{V}_t - V_t|$ of at least order $\Delta_n^{\frac{H}{4H+2}}$ (the latter is bigger than $\Delta_n^{1/6}$, as seen by letting $H \to 1$).

Notice that the asymptotic variance in Theorem 3.3, $\Sigma(x)$, is non-stochastic, so that the usual machinery of stable convergence from high-frequency analysis is not required to conclude that $\sqrt{T}\big(F_{n,T}(x) - F(x)\big)/\sqrt{\Sigma(x)} \xrightarrow{d} N(0,1)$.

$\Sigma(x)$ depends crucially on the memory of the volatility process. If a series is highly dependent, which is true empirically for volatility, it becomes harder to recover its marginal distribution, which results in larger values of $\Sigma(x)$ and wider confidence intervals for $F(x)$. The only way to compensate for this effect is to collect a larger sample.

$\Sigma(x)$ can be computed either analytically or, for example, by numerical integration. As shown by Dehay (2005), however, for finite T:

$$\mathrm{var}\left(\sqrt{T}\big(F_T(x) - F(x)\big)\right) = 2\int_0^T \left(1 - \frac{t}{T}\right)\big(F_t(x,x) - F(x)^2\big)\,dt. \tag{3.4}$$

The reduction by $(1 - t/T)$ in the true variance of F_T is an edge effect. We compute (3.4) in the simulation section, as it provides an infeasible assessment for the accuracy of the approximation in (3.3) for finite T. On the other hand, if Δ_n is not small enough to ignore the pathwise discretization error, there is an extra source of randomness in $F_{n,T}$ induced by the estimation of spot volatility. As such, (3.4) may actually understate the variation of $F_{n,T}$.

To estimate $\Sigma(x)$, we fix $\xi \in (0, 1/3)$ and set:

$$\Sigma_{n,T}(x) = 2\int_0^{T^\xi} \left(1 - \frac{t}{T^\xi}\right)\big(F_{t,n,T}(x) - F_{n,T}(x)^2\big)\,dt, \tag{3.5}$$

where

$$F_{t,n,T}(x) = \frac{1}{T - T^\xi}\int_0^{T-T^\xi} 1_{\{\hat{V}_{s+t}\leq x, \hat{V}_s\leq x\}}\,ds. \tag{3.6}$$

Then, as shown in the proof of Theorem 3.3, $\Sigma_{n,T}(x) \xrightarrow{\mathbb{P}} \Sigma(x)$.

The estimation of $\Sigma(x)$ is related to the problem of estimating the long-run variance in a stationary time series. In that context, the lag length has to grow slower than $\sqrt{T}$ to ensure consistency, and the optimal bandwidth is usually $O(T^{1/3})$, see, e.g., Newey and West (1994). The condition imposed on ξ is identical to that.

In practice, we may also be interested in determining how accurate the associated quantile(s) of F are estimated (e.g., for risk management reporting). These questions are, of course, intertwined.

Indeed, from the assumed existence of the density f and Corollary 21.5 in van der Vaart (1998), we deduce the following result.

Corollary 3.4 *Suppose that the conditions of Theorem 3.3 are fulfilled. Then, for any $\alpha \in (0, 1)$ with $f(Q(\alpha)) > 0$, it holds that*

$$\sqrt{T}(Q_{n,T}(\alpha) - Q(\alpha)) \xrightarrow{d} N\left(0, \frac{\Sigma(Q(\alpha))}{f(Q(\alpha))^2}\right). \tag{3.7}$$

Thus, inference about any given quantile is feasible, so long as an estimator of the density function of the marginal distribution evaluated at that quantile is available.

3.3 Goodness-of-fit of the volatility distribution

The preceding theory shows that $F_{n,T}$ converges to the stationary distribution function of volatility, as $\Delta_n \to 0$ and $T \to \infty$. In this section, we build upon this analysis to construct realized extensions of some classical goodness-of-fit tests for the volatility. To this end, we first derive a functional central limit theorem for the process $G_{n,T} = \left\{\sqrt{T}(F_{n,T}(x) - F(x)), x \in \mathbb{R}_+\right\}$.

Theorem 3.5 *Suppose that the conditions of Theorem 3.3 are fulfilled. Then, $G_{n,T}$ converges weakly in the space $\mathbb{D}(\mathbb{R}_+)$ of càdlàg functions equipped with the uniform topology to a Gaussian process, G_F, with mean zero and covariance function $\Sigma(x, y) = \text{cov}(G_F(x), G_F(y))$.*

Note that Theorem 3.5 can also be extended to the empirical process associated with the quantile estimator: $G_{n,T}^Q = \left\{\sqrt{T}(Q_{n,T}(\alpha) - Q(\alpha)), \alpha \in (0, 1)\right\}$.

As shown in Lemma A.3 in Appendix A, $\Sigma(x, y)$ can be consistently estimated by

$$\Sigma_{n,T}(x, y) = \int_0^{T^\xi} \left(F_{t,n,T}(x, y) + F_{t,n,T}(y, x) - 2F_{n,T}(x)F_{n,T}(y)\right)dt, \tag{3.8}$$

for $\xi \in (0, 1/3)$, where

$$F_{t,n,T}(x, y) = \frac{1}{T - T^\xi} \int_0^{T - T^\xi} 1_{\left\{\hat{V}_{s+t} \leq x, \hat{V}_s \leq y\right\}} ds. \tag{3.9}$$

Appealing to Theorem 3.5, we can now construct the test statistics for our goodness-of-fit tests, which evaluate the fit of an assumed stochastic volatility model:

$$rKS(F) = \sup_{x \in \mathbb{R}_+} |G_{n,T}(x)| \qquad \text{and} \qquad rL^2(F) = \int_{\mathbb{R}_+} G_{n,T}(x)^2 w(x)dx, \tag{3.10}$$

where $w : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous weight function.

rKS resembles a classical Kolmogorov-Smirnov test for an observed process, whereas rL^2 is a realized version of the goodness-of-fit test based on a weighted L^2 norm. The latter nests several existing tests, such as the Cramer-von-Mises (i.e., $w(x) = f(x)$) or Anderson-Darling (i.e., $w(x) = f(x)/[F(x)(1 - F(x))]$). rKS has the advantage that it is rather trivial to implement, once the REDF is constructed. On the other hand, the literature suggests that an L^2 statistic, at least in the i.i.d. setting, may be more powerful in finite samples, but it can also be tedious to compute, if a closed-form expression for w is unavailable. This is, for example, the case for a Cramer-von-Mises test if the marginal density is unknown or hard to evaluate.

Corollary 3.6 *Suppose that the conditions of Theorem 3.3 are fulfilled. Then, it holds that*

$$rKS(F) \overset{d}{\to} \sup_{x\in\mathbb{R}_+} |G_F(x)| \qquad \text{and} \qquad rL^2(F) \overset{d}{\to} \int_{\mathbb{R}_+} |G_F(x)|^2 w(x) dx. \tag{3.11}$$

Now, suppose we are interested in testing the fit of a stochastic volatility model, indexed by a known parameter vector υ. Later in this section, we return to the question of how parameter estimation affects the proposed tests. We let $F_\upsilon(x) = \mathbb{P}(V_t \leq x; \upsilon)$ describe the marginal distribution of the model.[7] The null—and associated alternative—hypothesis is then written as:

$$\mathcal{H}_0 : F = F_\upsilon \qquad \text{and} \qquad \mathcal{H}_a : F \neq F_\upsilon. \tag{3.12}$$

It follows that—under $\mathcal{H}_0$—$rKS(F_\upsilon) \overset{d}{\to} \sup_{x\in\mathbb{R}_+} |G_{F_\upsilon}(x)|$, whereas $rKS(F_\upsilon) \overset{\mathbb{P}}{\to} \infty$ under $\mathcal{H}_a$, because the alternative means there is at least one x, such that $F_{n,T}(x) \overset{\mathbb{P}}{\not\to} F_\upsilon(x)$. An equivalent result holds for $rL^2(F)$. As large values of the t-statistics discredit that $F = F_\upsilon$, the tests are therefore one-sided.

A drawback of this approach is that the limiting distribution of $rKS(F)$ and $rL^2(F)$ depends on the model under $\mathcal{H}_0$. This differs from the classical theory, for example the Kolmogorov-Smirnov statistic in the i.i.d. case with F continuous and known (see, e.g., van der Vaart, 1998). Even there, however, the asymptotic distribution is not readily available if the parameters describing F_υ are estimated, as we do below, although in that setting it often suffices to prepare a single table of family-wise critical values in advance—e.g., as in Lilliefors (1967, 1969)—if F_υ belongs to a location-scale family. This is due to the fact that the limiting distribution, in that case, does not depend on nuisance parameters of F, only the functional form of the probability distribution (e.g., David and Johnson, 1948). This is not true here, and we are therefore forced to retrieve critical values

---

[7]As $\upsilon \mapsto F_\upsilon$ is not one-to-one, in general, the model itself is not fully identified by F_υ, unless further constraints are imposed on the drift and volatility (e.g., Aït-Sahalia, 1996; Bibby, Skovgaard, and Sørensen, 2005).

through simulation on a case-by-case basis. In Appendix B, we offer a detailed recipe of how this procedure works both for Corollary 3.6 and 3.8.

In principle, we can construct a t-statistic that can be evaluated independently of F. To explain how, define $||f||_w = \sqrt{\int_{\mathbb{R}_+} w(x)f(x)^2 dx}$, such that $\mathbb{E}\big[||G_F||_w^2\big] < \infty$.

The other ingredient is a normalizing constant:

$$\tilde{\Sigma} = \int_{\mathbb{R}_+} \Sigma(x)w(x)dx, \tag{3.13}$$

with $\tilde{\Sigma} < \infty$. Then, we set:

$$T(F) = \frac{||G_{n,T}||_w^2}{\tilde{\Sigma}}. \tag{3.14}$$

We can then define a test statistic that rejects $\mathcal{H}_0$, if we observe that $T(F) > \chi_{1-\alpha}^2$, where $\chi_{1-\alpha}^2$ is the $(1-\alpha)$-quantile of a chi-squared distribution with one degree of freedom. Let $\alpha_{T(F)}$ be the significance level achieved by such a test. The following proposition summarizes the details.

Proposition 3.7 *Suppose that the conditions of Theorem 3.3 are fulfilled. Then, for any $\alpha \in (0, 0.215)$, it holds that under $\mathcal{H}_0$:*

$$\lim_{\Delta_n \to 0, T \to \infty} \alpha_{T(F)} \le \alpha, \tag{3.15}$$

while under $\mathcal{H}_a$:

$$\mathbb{P}\big(T(F) > c\big) \to 1, \tag{3.16}$$

for any $c > 0$.

We can thus construct a goodness-of-fit test for which critical values are readily available, however it achieves a significance level of at most α and may therefore be overly conservative for some F. Moreover, while the constant $\tilde{\Sigma}$ can be calculated under $\mathcal{H}_0$, so that no estimation is required, it is complicated to derive for many common stochastic volatility models. We therefore do not explore this test further.

In practice, the parameter vector v is typically unknown and not fixed a priori. It therefore has to be estimated, which invalidates the previous analysis (assuming F_v known). The intuition is that the estimation moves the model-implied marginal distribution closer to the EDF, making inference based on the preceding theory too conservative, as is well-known from the traditional literature on goodness-of-fit testing.

The complication brought about by parameter estimation can be viewed as a different testing problem that applies if we are interested in testing an entire class of stochastic volatility processes, as opposed to a particular member of that class.

The null is thus a composite hypothesis:

$$\mathcal{H}_0 : \exists \upsilon \in \Upsilon : F = F_\upsilon \qquad \text{and} \qquad \mathcal{H}_a : \forall \upsilon \in \Upsilon : F \neq F_\upsilon, \tag{3.17}$$

where Υ is the admissible parameter space (i.e., $\Upsilon = \{(\kappa, v_0, \xi) \in \mathbb{R}_+^3 : 2\kappa v_0 \geq \xi^2\}$ for the Heston (1993) model due to the Feller condition).

We assume a consistent estimator of υ, say $\hat{\upsilon}$, has been found and introduce a pseudo empirical process $\tilde{G}_{n,T} = \left\{ \sqrt{T}\big(F_{n,T}(x) - F_{\hat{\upsilon}}(x)\big), x \in \mathbb{R}_+ \right\}$. At a theoretical level, the problem then arises because the parameters of the underlying distribution typically cannot be recovered at a convergence rate faster than $T^{-1/2}$, i.e $\hat{\upsilon} - \upsilon = O_p(T^{-1/2})$. This cancels with the $\sqrt{T}$ scaling in the construction of $\tilde{G}_{n,T}$, so that the estimation error inevitably affects the asymptotic distribution.[8]

Corollary 3.8 *Suppose that the conditions of Theorem 3.3 are fulfilled. In addition, we assume $\hat{\upsilon}$ is asymptotically linear with influence function ψ and that $\dfrac{\partial F_\upsilon}{\partial \upsilon}(x)$ is bounded and continuous in both υ and x over the set $\mathbb{R}_+ \times \bar{\Upsilon}$ for a neighbourhood $\bar{\Upsilon}$ around υ. Then, $\tilde{G}_{n,T}$ converges weakly in $\mathbb{D}(\mathbb{R}_+)$ equipped with the uniform topology to a Gaussian process of the form $G_{F_\upsilon} - \dfrac{\partial F_\upsilon}{\partial \upsilon}^\top \int \psi dG_{F_\upsilon}$.*

The requirement that $\hat{\upsilon}$ should be asymptotically linear is fulfilled by most $\sqrt{T}$-consistent estimators proposed in the literature (e.g., Bickel, Klaassen, Ritov, and Wellner, 1998).

4 Simulation study

We now appraise, via Monte Carlo simulation, our estimator of the EDF of spot variance introduced in Section 2. We also evaluate the size and power of the goodness-of-fit tests for the marginal distribution of volatility proposed in Section 3.3.

Starting at $X_0 = 0$, and recalling that $\sigma_t = \sqrt{V_t}$, the efficient log-price is a convolution of a Heston (1993)-type stochastic volatility model and a pure-jump process:

$$
\begin{aligned}
dX_t &= \sqrt{V_t}dW_t + dJ_t, \\
dV_t &= \kappa\big(v_0 - V_t\big)dt + \xi\sqrt{V_t}dB_t,
\end{aligned}
\tag{4.1}
$$

---

[8]A standard rate of convergence is required in the following. Although we are not aware about the presence of so-called "super" consistent estimators in the stochastic volatility literature—at least not in the stationary setting—we point out that if $\hat{\upsilon} - \upsilon = o_p(T^{-1/2})$ the sampling error is asymptotically negligible and the problem reverts back to Corollary 3.6. On the other hand, a slower rate of convergence implies the error is blown up by the $\sqrt{T}$ scaling and then our approach just goes out the window.

where W and B are Brownian motions with $\mathrm{corr}[dW_t dB_t] = \rho = -\sqrt{0.5}$. The other parameters are $\kappa = 0.05$, $v_0 = 1$, and $\xi = 0.2$, which is calibrated to our empirical data and broadly aligns with previous studies (e.g, Aït-Sahalia and Kimmel, 2007).

In this framework, the law of V_t (conditional on V_s, for $s < t$) is—up to a scale factor—non-central chi-square with $d = 4v_0\kappa\xi^{-2}$ degrees of freedom and non-centrality parameter $\lambda = \dfrac{4\kappa e^{-\kappa(t-s)}}{\xi^2(1 - e^{-\kappa(t-s)})}V_s$ (e.g., Cox, Ingersoll, and Ross, 1985). As $t \to \infty$, the conditional distribution converges to a Gamma$(2\kappa v_0\xi^{-2}, 2\kappa\xi^{-2})$, which is also the unconditional law of V_t for any finite t, if V_0 is drawn from this distribution. We do that here. The variance of $F_T(x)$ in (3.4) is then computed from these expressions by numerical integration of the joint density function.

J_t is a symmetric tempered stable process with Lévy measure (e.g., Rosiński, 2007):

$$\nu(dx) = c\frac{e^{-\lambda x}}{x^{1+r}}dx, \quad \text{for } x > 0, \tag{4.2}$$

where $c > 0$, $\lambda > 0$ and $r \in [0, 2)$. The degree of price jump activity is measured by r. We choose $r = 0.5$, which produces an infinite-activity, finite-variation process with many small jumps and a few large ones. We assume $\lambda = 3$ and calibrate c such that J_t accounts for 20% of the quadratic variation. This complies with previous work in the field (e.g., Aït-Sahalia, Jacod, and Li, 2012; Aït-Sahalia and Xiu, 2015).

We generate 10,000 "continuous-time" paths with $23,400$ updates per unit of time T. As described in Todorov, Tauchen, and Grynkiv (2014), J_t is simulated as the difference between two spectrally positive tempered stable processes using the acceptance-rejection algorithm of Baeumer and Meerschaert (2010). The rest of the system is discretized by an Euler approach. To gauge how close the REDF is to the population counterpart, we extract a coarser grid of $n = 2,340T$ log-price increments. This emulates a financial market, where new transactions arrive regularly every tenth second over a 6.5 hour trading day, which is aligned to the sample sizes in our empirical work.

We add heteroscedastic noise $Z_{i\Delta_n} = X_{i\Delta_n} + U_{i\Delta_n}$ with $U_{i\Delta_n} \sim N(0, \omega_{i\Delta_n}^2)$, where $\omega_{i\Delta_n}^2$ is set by fixing the noise-to-signal ratio (e.g., Oomen, 2006): $\gamma = \sqrt{\dfrac{1}{\Delta_n}\dfrac{\omega_{i\Delta_n}^2}{V_{i\Delta_n}}}$. As consistent with Christensen, Oomen, and Podolskij (2014), we take $\gamma = 0.5$.[9] Thus, $(Z_{i\Delta_n})_{i=0}^n$ comprises our discretely observed sample used to estimate the spot variances, which we recover with $\tilde{V}_{i\Delta_n}$ from (2.13) combined with (2.15). We pre-average with $g(x) = \min(x, 1 - x)$, $k_n = [\theta/\sqrt{\Delta_n}]$ and $\theta = 1/3$. We also truncate using a standard approach. In particular, a pre-averaged increment is reset to zero if

---

[9]We made a robustness check with $\gamma = 2.0$ as in Aït-Sahalia, Jacod, and Li (2012). The pre-averaging estimator was hardly affected by this change, while the sparse estimator was substantially more biased, due to the amplification of the noise.

Figure 1: Estimation of the EDF.

Panel A: Sample path of V_t.

Panel B: EDF and REDF.
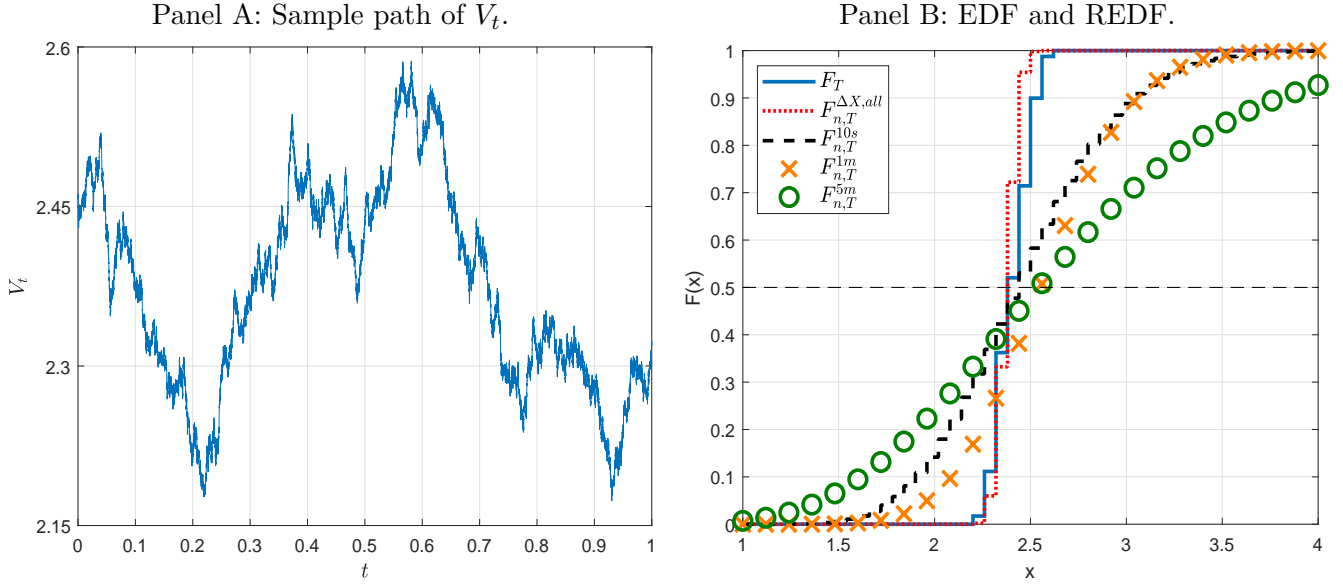
Note. This figure plots a simulated path of V_t in Panel A. The associated EDF of V_t, F_T, appears in Panel B (solid line). We estimate F_T with $F_{n,T}$ over 10,000 simulations and plot the average (dashed line). The non-noise-robust version proposed in Li, Todorov, and Tauchen (2013, 2016) based on 1- and 5-minute sparse sampling is shown as a comparison ('×' and '○'). The dotted line is an infeasible estimate that has access to the whole record of X, $(X_{i\Delta_n})_{i=0}^n$.

$|\bar{Z}_i| \geq q_{1-\alpha}\sqrt{\widehat{IV}}\Delta_n^{\bar{\omega}}$, where $q_{1-\alpha}$ is the $1 - \alpha$ quantile of the standard normal distribution, $\widehat{IV}$ is the pre-averaged bipower variation, and $\bar{\omega} \in (0, 0.25)$ controls how fast the cutoff goes to zero with Δ_n (note $\bar{Z}_i = O_p(\Delta_n^{1/4})$ if X is continuous). Here, $\alpha = 0.001$ and $\bar{\omega} = 0.20$. The debiasing in $\hat{V}_{i\Delta_n}$ is done with $\hat{\omega}_{i\Delta_n}^2$ from (2.16), which we measure locally on a block that covers all the noisy high-frequency data that were included in the calculation of the pre-averaged increments that are part of the computation of $\tilde{V}_{i\Delta_n}$.

In Figure 1, we examine how $F_{n,T}$ recovers F_T. We set $T = 1$ and freeze volatility across replica to keep the estimation target fixed. The sample path of V_t is shown in Panel A, while the associated EDF appears in Panel B. We report our estimator and compare it to Li, Todorov, and Tauchen (2013). We construct the latter statistic from 1- and 5-minute low-frequency returns. As recommended in their work and to facilitate comparison, h_n is set throughout so that it spans about 1.5 hours worth of data.

At the 1-minute resolution, the Li, Todorov, and Tauchen (2013) estimator is severely distorted by market frictions, which leads to a systematic upward bias. This was to be expected, as it is not resistant to noise. On the other hand, while the 5-minute statistic does not accumulate any discernable bias, it delivers imprecise estimates, resulting in a highly overdispersed REDF. In

Figure 2: Mean relative bias and absolute error.

Panel A: Relative bias. Panel B: Absolute error.
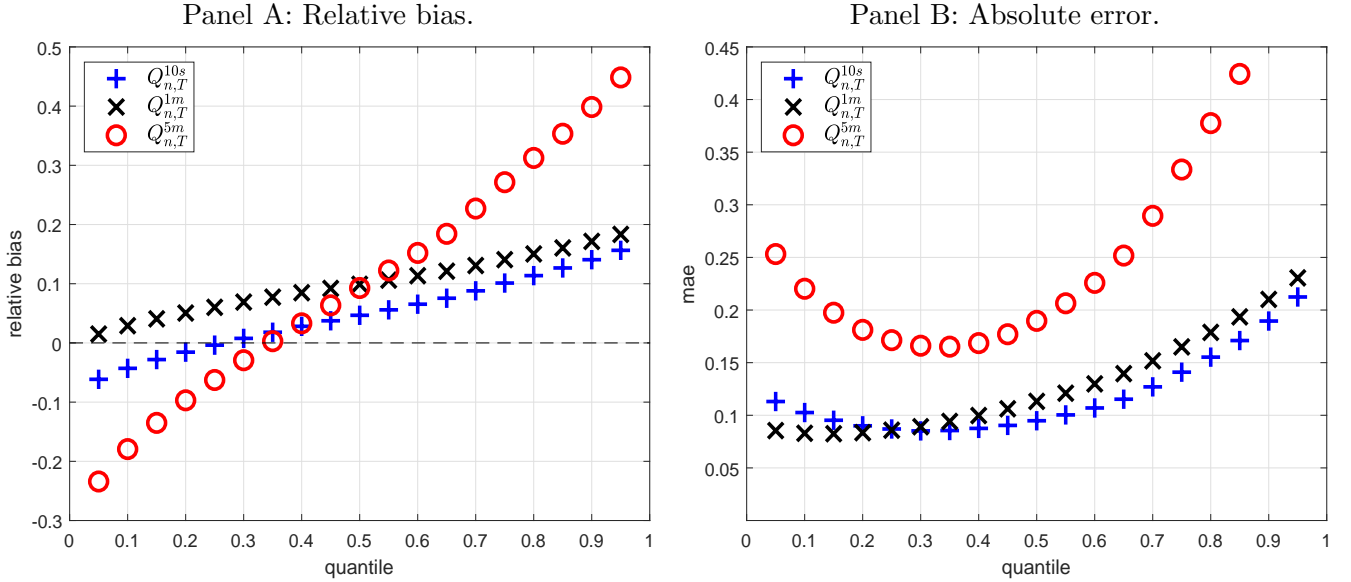
Note. This figure plots the average relative bias in Panel A and absolute error in Panel B of our realized estimator of the EDF of spot variance ('+'). As a comparison, we also plot the Li, Todorov, and Tauchen (2013) no-noise estimator based on 1- and 5-minute sparse sampling ('×' and '○').

contrast, as our noise-robust estimator is able to capitalize on tick-by-tick information, it is more accurate. This suggests the inferior rate of convergence embedded in pre-averaging is more than offset by being able to employ all the data.

Next, we vary V_t randomly in each iteration and estimate F_T as above. In Figure 2, we plot the relative error $Q_{n,T}(\alpha)/Q_T(\alpha) - 1$ in Panel A and absolute error $\left|Q_{n,T}(\alpha) - Q_T(\alpha)\right|$ in Panel B, which are averaged across Monte Carlo trials, for the vector of quantiles $\alpha = (0.05, 0.10, \ldots, 0.95)'$. As evident—and consistent with Figure 1—our estimator is less biased, especially in the tails. The shape of the upward sloping bias, common among estimators, is explained by the fact that the sampling distribution of $\hat{V}_t$, say, is spread out (not necessarily centered) around V_t. Hence, when volatility is low, $\hat{V}_t$ sometimes underestimates V_t, which translates into a downward bias in the low quantiles of the REDF vis-á-vis the EDF, and vice versa. The bias is more pronounced when V_t is high, because it is harder to estimate spot variance in a volatile market. Overall, the REDF proposed in this paper is nevertheless roughly unbiased over a large spectrum of the distribution with a relative error within a few percent. This is further corroborated by Panel B. In general, while all estimators do a relatively poorer job at fitting the tails of the distribution of spot variance, there are again notable gains by applying our estimator.

We now turn to estimation of the marginal distribution of volatility based on Theorem 3.3.

Figure 3: Inference about marginal distribution, $T = 250$.

Panel A: Point estimate.

Panel B: Kernel density.
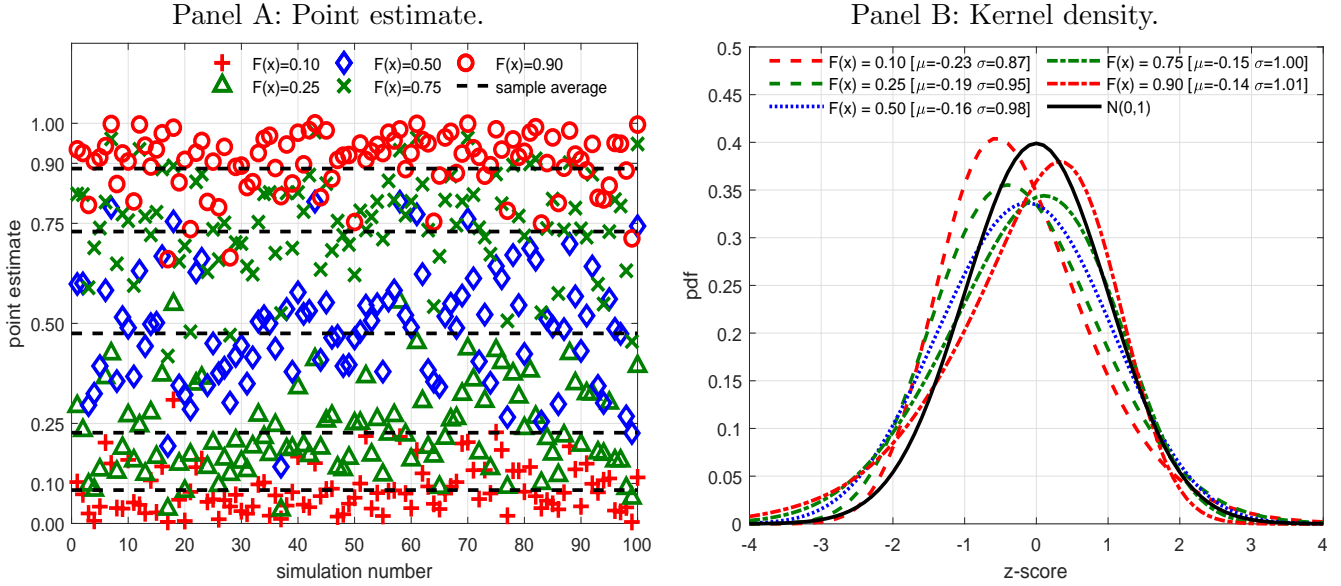
Note. In Panel A, we report point estimates of $F_{n,T}(x)$ for the first one hundred Monte Carlo replica and sample averages across all simulations, for $F(x) = 0.10, 0.25, 0.50, 0.75$ and 0.90. In Panel B, we plot kernel density estimates (with bandwidth selection via Silverman's rule of thumb) of the standardized statistic $\sqrt{T}\big(F_{n,T}(x) - F(x)\big)/\sqrt{\Sigma(x)}$. The Gaussian curve is superimposed as a visual reference point. Throughout, $\Delta_n = 1/2,340$ and $T = 250$.

We work with $T = 250$, corresponding to the number of trading days in about a year. This relatively low value is meant to be conservative and illustrate the potential of our estimator. As a comparison, $T = 2,517$ in our empirical work in Section 5. Note that n is kept fixed relative to T, so that Δ_n is constant. In Figure 3, we report point estimates—for the first 100 realizations—and sample averages across all 10,000 simulations of $F_{n,T}(x)$ and kernel densities of the asymptotic pivot $\sqrt{T}\big(F_{n,T}(x) - F(x)\big)/\sqrt{\Sigma(x)} \xrightarrow{d} N(0,1)$ evaluated at five distinct points, which cover a variety of low-to-high volatility states, i.e. $F(x) = 0.10, 0.25, 0.50, 0.75$ and 0.90. We see that in general the REDF is unbiased and correctly scaled. If we look far in the tails, in particular the right-hand one, the density estimate is slightly skewed towards the center, which is intuitive due to the natural bounds on $F_{n,T}$. Still, the limiting distribution provides a good description of the finite sample variation of $F_{n,T}$.

Last but not least, we explore the properties of the realized Kolmogorov-Smirnov and Cramer-von-Mises (i.e., a weighted L^2 norm with $w(x) = f(x)$) goodness-of-fit test. This piece of analysis is based on a shorter run of 1,000 Monte Carlo replica due to the increased computational cost of the resampling scheme. We draw 500 bootstrap samples in each simulation trial and inspect both the setting where v is known and estimated. We follow Corradi and Distaso (2006) and recover

20

the parameters via $\hat{v}$ using GMM to match the sample mean, variance, plus the first and second autocovariance of the bias-corrected pre-averaged bipower variation of Podolskij and Vetter (2009a) to the corresponding model-based moments of integrated variance (the former is consistent for the latter).

To be able to define a measure of the testing power we simulate from another model, i.e. the log-normal—or exponential Ornstein-Uhlenbeck—process, which is the solution of the stochastic differential equation:

$$d \ln V_t = \kappa(v_0 - \ln V_t)dt + \xi dB_t, \tag{4.3}$$

where $v = (\kappa, v_0, \xi) = (0.08, -0.3, 0.45)$ is comparable to what we observe in the real data. We leave the leverage correlation unchanged at $\rho = -\sqrt{0.5}$.

In Table 1, we report the size and power of the test for a nominal significance level of $\alpha = 5\%$ and $T = 250, 500, 1,000,$ and $2,000$. As seen, the test is roughly unbiased and has a rejection rate close to α. We next comment on the power, as reported in the second half of Table 1. It is calculated as the rejection rate achieved by simulating log-normal volatility and testing the square-root process. $\mathcal{H}_0$ without estimation is here based on the configuration of the Heston (1993) model with the true parameters used to simulate volatility for the size analysis. As shown, the test has moderate power if the sample size is small, but the rejection rate increases steadily as T grows larger. At $T = 2,000$, the power is in the 60–90% range. The largest choice of T corresponds to about eight years worth of high-frequency data, which is widely available for many assets in practice. The impact of parameter estimation is to "flatten" the power curve adding an upward bias for small T and dampening it for large T. We conjecture this effect is caused by the inherent sampling error in the pre-averaged bipower variation, which leads to systematic upward biases in the estimate of both the mean reversion and volatility-of-volatility coefficient κ and ξ.[10] The misspecification of the dynamic properties of the system invariably yields simulated critical values that are slightly off target and this effect appears more pronounced under the alternative. Finally, we notice the realized Kolmogorov-Smirnov test has slightly higher power than our weighted L^2 statistic, which is a somewhat surprising finding compared to the classical results for i.i.d. data.

---

[10]The Heston (1993) model has an ARMA(1,1) representation for the integrated variance. We could exploit this structure by casting the whole system into state-space representation and then use a Kalman filter to extract a filtered (or smoothed) volatility series before estimating the model, as discussed in Barndorff-Nielsen and Shephard (2002) (see also, e.g., Todorov, Tauchen, and Grynkiv, 2011, for an alternative approach). This tends to deliver less biased parameter estimates. However, we do not pursue this idea here.

Table 1: Properties of the realized goodness-of-fit tests.

| | | Size | | | | Power | | | |
| | | $rKS(F)$ | | $rL^2(F)$ | | $rKS(F)$ | | $rL^2(F)$ | |
		v	$\hat{v}$	v	$\hat{v}$	v	$\hat{v}$	v	$\hat{v}$
$T =$	250	0.042	0.086	0.048	0.086	0.091	0.422	0.074	0.314
	500	0.044	0.060	0.047	0.082	0.192	0.502	0.115	0.389
	1000	0.043	0.034	0.043	0.054	0.471	0.596	0.266	0.451
	2000	0.035	0.026	0.049	0.022	0.951	0.793	0.893	0.637

Note. The table reports rejection rates of the goodness-of-fit tests for stochastic volatility models proposed in Section 3.3. $rKS(F)$ ($rL^2(F)$) is the realized Kolmogorov-Smirnov (Cramer-von-Mises) test. The number of Monte Carlo simulations is 1,000 with 500 bootstrap samples used to compute critical values for the t-statistic in each trial. v ($\hat{v}$) is based on the true (estimated) parameter vector. The power is calculated as the rejection rate achieved by simulating log-normal volatility and testing whether the Heston (1993) process is the true model. Further details are available in the main text.

5 Empirical application

We here illustrate the application of our new nonparametric estimator of the EDF of spot variance and the associated goodness-of-fit tests to real high-frequency data. We construct the $F_{n,T}$ measure from tick data based on selected exchange-traded funds (ETFs) that cover different sectors of the U.S. stock market. In addition to a proxy for the market index (ticker symbol SPY), we add the nine industry portfolios of the S&P 500, yielding a total of ten securities.[11] High-frequency data was acquired from the TAQ database and comprise a complete series of trades and quotes for each stock for the sample period July 2008 – June 2018. The analysis here is based on transaction data, which were filtered for outliers as in Christensen, Oomen, and Podolskij (2014) and subsequently pre-ticked to an equidistant 10-second grid from 9:30am to 4:00pm EST, resulting in $T = 2,517$ days of $n = 2,340$ intraday returns.[12] A list of ticker symbols and descriptive statistics of the sample are presented in Table 2.

The calculation of $\hat{V}_t$ proceeds as in Section 2 – 4 in terms of tuning parameters. We further

---

[11]There is also a 10th sector ETF (XLRE), which is a diversified portfolio of companies exposed to real estate. However, as the inception date of this fund is rather recent, we exclude it from consideration due to the limited availability of high-frequency data. Further information about the ETFs is available at http://www.sectorspdr.com/.

[12]Trading at NYSE terminates early at 1:00pm on a few regularly scheduled business days each year in observance of upcoming holidays. We do not collect data after the exchange has officially closed, as there is typically very little liquidity, and therefore use a shorter sample of $n = 1,260$ log-price increments on these days.

Table 2: Descriptive statistics of high-frequency data.

code	effective sample	σ	γ	goodness-of-fit test (p-value)					
				gamma		log-normal		inverse gaussian	
				$rKS(F)$	$rL^2(F)$	$rKS(F)$	$rL^2(F)$	$rKS(F)$	$rL^2(F)$
XLB	0.756	0.204	0.253	0.001	0.000	0.000	0.000	0.080	0.039
XLE	0.938	0.234	0.136	0.000	0.000	0.007	0.005	0.219	0.161
XLF	0.930	0.246	0.895	0.000	0.000	0.000	0.000	0.009	0.004
XLI	0.809	0.176	0.334	0.000	0.000	0.006	0.004	0.130	0.109
XLK	0.761	0.165	0.506	0.000	0.000	0.052	0.025	0.035	0.045
XLP	0.723	0.124	0.558	0.003	0.001	0.021	0.009	0.015	0.022
XLU	0.778	0.167	0.381	0.000	0.000	0.083	0.060	0.041	0.071
XLV	0.768	0.144	0.342	0.000	0.000	0.000	0.000	0.046	0.035
XLY	0.768	0.174	0.245	0.003	0.000	0.000	0.000	0.054	0.021
SPY	0.999	0.153	0.145	0.003	0.000	0.029	0.007	0.300	0.360

Note. We report descriptive statistics of the equity high-frequency data. "code" is the ticker symbol, "effective sample" (a measure of liquidity) is the fraction of previous-tick interpolated 10-second prices originating from a new transaction rather than a repetition, σ is annualized volatility based on a pre-averaged bipower variation of Podolskij and Vetter (2009a), while γ is the noise-to-signal ratio defined in the main text. The numbers are computed daily and averaged over the sample. The p-value is for the realized Kolmogorov-Smirnov and Cramer-von-Mises goodness-of-fit test with marginal distribution under $\mathcal{H}_0$ indicated by the designated column label.

account for the diurnal pattern in intraday spot volatility. We follow the standard approach in the literature, which is to compute the average value of $\hat{V}_t$ at each time point of the day over the sample. The diurnal factor is then estimated by normalizing the sum of these averages to one. We correct $\hat{V}_t$ by this quantity and use the adjusted time series to construct $F_{n,T}$. In Panel A of Figure 4, we show a kernel-based estimate of the probability density function implied by the time series of variance estimates of SPY, as representative of our data.

We compute $rKS(F)$ and $rL^2(F)$ with three choices of F under $\mathcal{H}_0$. The first are the gamma and log-normal distribution inherent in the square-root and log-normal diffusion models that were presented in the simulation section. Furthermore, we inspect the goodness-of-fit of a non-Gaussian Ornstein-Uhlenbeck process (e.g., Barndorff-Nielsen and Shephard, 2001). The latter is a pure-jump specification with SDE:

$$dV_t = -\kappa V_t dt + dL_t, \tag{5.1}$$

where L_t is a subordinator with Lévy measure $\nu_L(dx)$.

Figure 4: Properties of stochastic variance in SPY.

Panel A: PDF of spot variance.
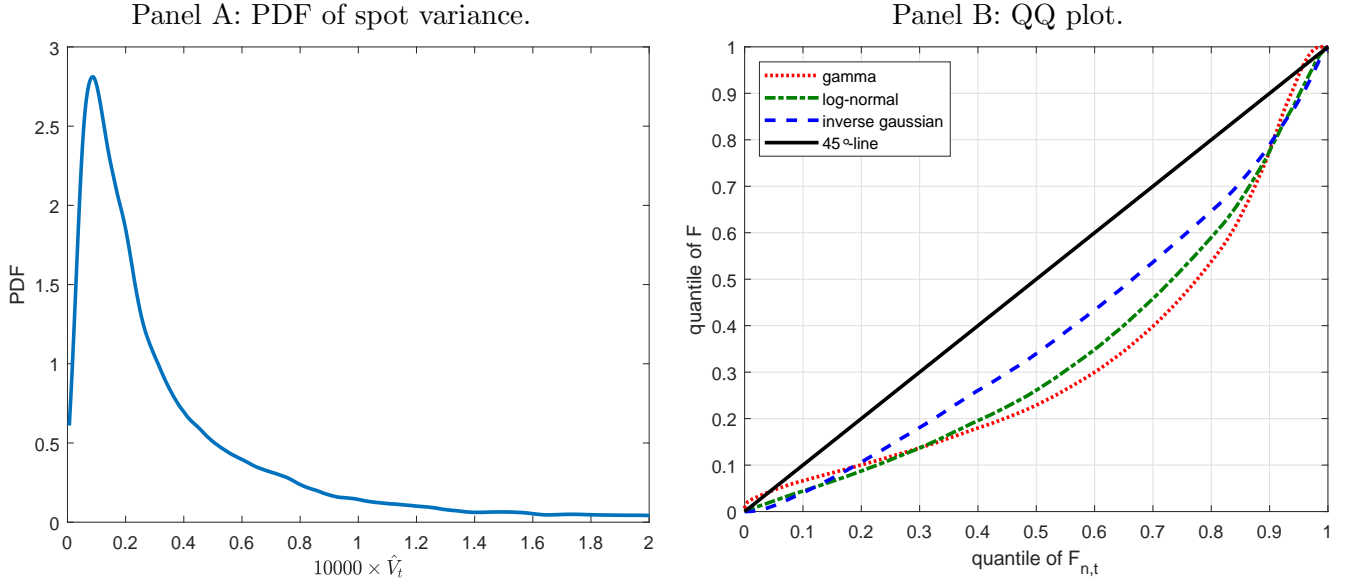
Panel B: QQ plot.

Note. In Panel A, we plot a kernel-based estimate of the probability density function of spot variance for SPY. In Panel B, we create a QQ plot, where the quantiles of $F_{n,T}$ (on the x-axis) are mapped against the quantiles of a fitted distribution function F (on the y-axis). The gamma, log-normal and inverse Gaussian distribution are included for the latter comparison.

To maintain a tractable setting, we proceed as in Todorov, Tauchen, and Grynkiv (2011) by assuming that the distribution of the increments of the volatility process corresponds to that of a tempered stable process. This means it has a Lévy density:

$$\nu_V(dx) = c\frac{e^{-\lambda x}}{x^{1+\beta}}dx, \quad \text{for } x > 0, \tag{5.2}$$

where $c > 0$, $\lambda > 0$ and $\beta < 1$. The Lévy measure of the background driving process $\nu_L(dx)$ can then be backed out from $\nu_V(dx)$, see e.g., Barndorff-Nielsen and Shephard (2001). We fix the activity level of the volatility jumps at $\beta = 0.5$, so that the resulting marginal distribution of volatility is inverse Gaussian, i.e. $V_t \sim IG(\mu, \nu)$ with:

$$\mu = c\sqrt{\frac{\pi}{\lambda}}, \qquad \nu = 2\pi c^2, \tag{5.3}$$

and

$$f(x; \mu, \nu) = \sqrt{\frac{\nu}{2\pi x^3}} \exp\left\{ -\frac{\nu}{2\mu^2 x}(x - \mu)^2 \right\}, \quad \text{for } x > 0, \tag{5.4}$$

is the density function of V_t.

The procedure in Section 4 is applied to report a p-value for the t-statistic. The exception is that we boost the number of boostrap samples to 1,000 for higher precision. The non-Gaussian

Ornstein-Uhlenbeck process is discretized and simulated as explained in the appendix of Todorov, Tauchen, and Grynkiv (2011).[13]

In Table 2, we present the outcome of these efforts, while Panel B of Figure 4 shows a QQ-style plot, where the quantiles of $F_{n,T}$ (on the x-axis) are compared to the quantiles of the best fitting distribution function F in each class (on the y-axis). The graph should align with the 45°-line if the model is correct. As readily seen, the gamma distribution does not capture stochastic equity variance at all. The main problem with the square-root process is that it cannot generate a sufficient level of volatility-of-volatility (via ξ) without violating the Feller restriction, which is highly binding in practice. The p-values of the log-normal model are higher on average, but it also does not deliver an acceptable fit of the data across sectors. Although there are a few close calls, for example XLK, XLU and SPY, the model is generally rejected at standard levels of significance. In contrast, the p-values of the inverse Gaussian are typically much higher. This is in line with Todorov and Tauchen (2012), and it appears relatively consistent across the ETF space. Nevertheless, the latter also struggles (depending a bit on which t-statistic we ask), for instance it does not capture return variation exhibited by the highly volatile financial sector ETF (XLF), but it is generally the best fitting model of the ones estimated in this paper.

6 Conclusion

We construct a nonparametric jump- and noise-robust realized measure of the EDF of the latent volatility of a general Itô semimartingale sampled at high-frequency on a fixed time interval. We extend previous work of Li, Todorov, and Tauchen (2013, 2016) in the noise-free setting and prove that in the presence of microstructure noise the pre-averaged version of their estimator is consistent in the infill asymptotic limit. The resistance to market frictions enables our statistic to fully exploit information about volatility available in tick-by-tick high-frequency data, which improves its accuracy in a simulation study.

In a subsequent analysis, we let the time span tend to infinity. We show that our estimator then converges to the marginal distribution function of volatility. This result is subject to some

---

[13]As in Meddahi (2003), we exploit the ARMA(1,1) structure for the integrated variance of this model to compute the relevant moments: $\mathbb{E}[IV_t] = c\sqrt{\dfrac{\pi}{\lambda}}$, $\mathrm{var}(IV_t) = \dfrac{c\sqrt{\pi}}{\lambda^{3/2}\kappa^2}(e^{-\kappa} + \kappa - 1)$, $\mathrm{cov}(IV_t, IV_{t-1}) = \sqrt{\dfrac{\pi c^2}{4\lambda^3}}\dfrac{(1 - e^{-\kappa})^2}{\kappa^2}$, and $\mathrm{cov}(IV_t, IV_{t+j}) = e^{-\kappa}\mathrm{cov}(IV_t, IV_{t+j-1})$ for $j \geq 2$. We also calculated these expressions based on the ARMA coefficients reported in the appendix of Todorov, Tauchen, and Grynkiv (2011), but the results differ and we suspect there are some typos in that paper.

rate conditions, which are more restrictive in the presence of jumps or roughness in volatility. We establish a functional CLT, from which we prove the validity of a family of goodness-of-fit tests for stochastic volatility models. The limiting distribution of our t-statistic depends on the true model under the null hypothesis, but critical values can be found via simulation. We show how the procedure can be adapted to the case with parameter estimation.

We apply the theory to high-frequency data from several ETFs that track different sectors of the U.S. stock market. The marginal distribution embedded in several classes of popular stochastic volatility models is tested against our nonparametric measure. There is strong evidence against the gamma or log-normal distribution, while the inverse Gaussian—implied by the class of non-Gaussian Ornstein-Uhlenbeck processes of Barndorff-Nielsen and Shephard (2001) with tempered stable increments—has more support. These findings are consistent with Todorov, Tauchen, and Grynkiv (2011), but in disagreement with Corradi and Distaso (2006).

Our paper can serve as a starting point for building more refined models of stochastic volatility. It appears natural, for example, to inspect the properties of the generalized inverse Gaussian (GIG) as the marginal distribution of volatility. It has the PDF:

$$f(x; a, b, p) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})} x^{(p-1)} e^{-(ax+b/x)/2}, \quad \text{for } x > 0, \tag{6.1}$$

where K_p is the modified Bessel function of the second kind, while $a > 0, b > 0$ and p a real number are the parameters. The GIG is self-decomposable (necessary and sufficient for stationary solutions of (5.1) to exist) and nests the gamma distribution as $b \to 0$ (with $a > 0$) and the inverse Gaussian if $p = -1/2$. The extra degree of freedom may just be enough for a complete description of equity variance. We leave this assessment for future research.

A   Appendix

We here prove the theoretical results presented in the main text. As usual, the localization procedure in Jacod and Protter (2012, Section 4.4.1) implies we can assume that b_t, σ_t, $\tilde{b}_t$, $\tilde{\sigma}_t$, $\tilde{\sigma}_t'$, ω_t and the jump component of X are bounded for the infill setting with T fixed, i.e. Lemma 2.1, Theorem 3.1 and Corollary 3.2. Without loss of generality, the proofs are based on $r = 2$. Also, to preserve notation C denotes a generic constant that changes value from one line to the next.

Proof of Lemma 2.1

Fix a $t > 0$. If $t \in [i\Delta_n, (i+1)\Delta_n)$, we set $t_n = i\Delta_n$ to highlight the dependence on n and note that $V_{t_n} \to V_t$ and $\omega_{t_n}^2 \to \omega_t^2$ due to their càdlàg properties. Thus, our job is to establish:

$$\tilde{V}_{i\Delta_n} - \theta\psi_2 V_{t_n} - \frac{1}{\theta}\psi_1\omega_{t_n}^2 \xrightarrow{\mathbb{P}} 0.$$

We start by introducing an approximation of the pre-averaged return series $\bar{Z}_i$, in which the stochastic volatility and noise processes are fixed locally in each computation:

$$\bar{Y}_{i,i+m} = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)\left(\sigma_{i\Delta_n}\Delta_{i+m+j}^n W + \omega_{i\Delta_n}\Delta_{i+m+j}^n\varepsilon\right),$$

and then we deduce the claim based on $\bar{Y}_{i,i+m}$, i.e.

$$\check{V}_{i\Delta_n} \equiv \frac{1}{h_n\sqrt{\Delta_n}}\sum_{m=0}^{h_n-1}\bar{Y}_{i,i+m}^2 - \theta\psi_2 V_{t_n} - \frac{1}{\theta}\psi_1\omega_{t_n}^2 \xrightarrow{\mathbb{P}} 0. \tag{A.1}$$

Now, it is readily seen that for each $m \geq 0$, $\Delta_n^{-1/4}\bar{Y}_{i,i+m}$ has (conditionally on $\mathcal{F}_{(i+m)\Delta_n}$) mean zero and variance equal to $k_n\sqrt{\Delta_n}\psi_2^n V_{t_n} + \frac{1}{k_n\sqrt{\Delta_n}}\psi_1^n\omega_{t_n}^2$. Furthermore,

$$k_n\sqrt{\Delta_n}\psi_2^n V_{t_n} + \frac{1}{k_n\sqrt{\Delta_n}}\psi_1^n\omega_{t_n}^2 - \theta\psi_2 V_{t_n} - \frac{1}{\theta}\psi_1\omega_{t_n}^2 \xrightarrow{\mathbb{P}} 0,$$

as $n \to \infty$. It follows that

$$\frac{1}{h_n\sqrt{\Delta_n}}\sum_{m=0}^{h_n-1}\mathbb{E}\left[\bar{Y}_{i,i+m}^2 \mid \mathcal{F}_{(i+m)\Delta_n}\right] - \theta\psi_2 V_{t_n} - \frac{1}{\theta}\psi_1\omega_{t_n}^2 \xrightarrow{\mathbb{P}} 0,$$

and therefore it suffices to prove that

$$\frac{1}{h_n\sqrt{\Delta_n}}\sum_{m=0}^{h_n-1}\left(\bar{Y}_{i,i+m}^2 - \mathbb{E}\left[\bar{Y}_{i,i+m}^2 \mid \mathcal{F}_{(i+m)\Delta_n}\right]\right) \xrightarrow{\mathbb{P}} 0.$$

To this end, we note that

$$\Delta_n^{-1}\mathbb{E}\left[\left(\bar{Y}_{i,i+m}^2 - \mathbb{E}\left[\bar{Y}_{i,i+m}^2 \mid \mathcal{F}_{(i+m)\Delta_n}\right]\right)\left(\bar{Y}_{i,i+l}^2 - \mathbb{E}\left[\bar{Y}_{i,i+l}^2 \mid \mathcal{F}_{(i+l)\Delta_n}\right]\right)\right] \leq C, \qquad \text{for } |m-l| \leq k_n,$$

$$\Delta_n^{-1}\mathbb{E}\left[\left(\bar{Y}_{i,i+m}^2 - \mathbb{E}\left[\bar{Y}_{i,i+m}^2 \mid \mathcal{F}_{(i+m)\Delta_n}\right]\right)\left(\bar{Y}_{i,i+l}^2 - \mathbb{E}\left[\bar{Y}_{i,i+l}^2 \mid \mathcal{F}_{(i+l)\Delta_n}\right]\right)\right] = 0, \qquad \text{for } |m-l| > k_n.$$

Thus,

$$\mathbb{E}\left[\left(\frac{1}{h_n\sqrt{\Delta_n}}\sum_{m=0}^{h_n-1}\left(\bar{Y}_{i,i+m}^2 - \mathbb{E}\left[\bar{Y}_{i,i+m}^2 \mid \mathcal{F}_{(i+m)\Delta_n}\right]\right)\right)^2\right] \leq C\frac{k_n}{h_n} \to 0,$$

and (A.1) follows.

We are thus left with the assertion:

$$S_n = \tilde{V}_{i\Delta_n} - \check{V}_{i\Delta_n} \xrightarrow{\mathbb{P}} 0. \tag{A.2}$$

To prove (A.2), we define

$$X_t'' = \int_0^t \int_{\mathbb{R}} \delta(s,z)1_{\{|\delta(s,z)|\leq 1\}}(\mu-\nu)(d(s,z)), \quad B_t'' = \int_0^t b_s ds + \int_0^t \int_{\mathbb{R}} \delta(s,z)1_{\{|\delta(s,z)|>1\}}\mu(d(s,z)),$$

$$Y_t' = \int_{i\Delta_n}^t (\sigma_s - \sigma_{i\Delta_n})dW_s + (\omega_t - \omega_{i\Delta_n})\varepsilon_t.$$

Note that $\Delta_j^n Z = \sigma_{i\Delta_n}\Delta_j^n W + \omega_{i\Delta_n}\Delta_j^n \varepsilon + \Delta_j^n B'' + \Delta_j^n Y' + \Delta_j^n X''$, for $j = i+1,\ldots,i+h_n+k_n$.
Now, from the inequality (Jacod and Protter, 2012, p. 258):

$$\left|(x+y+z+w)^2 1_{\{|x+y+z+w|\leq v\}} - x^2\right| \leq C\frac{x^4}{v^2} + \epsilon x^2 + \frac{C}{\epsilon}\left((y^2 \wedge v^2) + z^2 + w^2\right)$$

with $v = v_n\Delta_n^{-1/4} = \alpha\Delta_n^{\bar{\omega}-1/4}$, $\epsilon \in (0,1]$, and

$$x = \frac{\bar{Y}_{i,i+m}}{\Delta_n^{1/4}}, \quad y = \frac{\bar{X}_{i+m}''}{\Delta_n^{1/4}}, \quad z = \frac{\bar{Y}_{i+m}'}{\Delta_n^{1/4}}, \quad w = \frac{\bar{B}_{i+m}''}{\Delta_n^{1/4}},$$

we find that

$$|S_n| \leq \frac{1}{h_n}\sum_{m=0}^{h_n-1}\left[C\Delta_n^{1/2-2\bar{\omega}}\left|\frac{\bar{Y}_{i,i+m}}{\Delta_n^{1/4}}\right|^4 + \epsilon\left|\frac{\bar{Y}_{i,i+m}}{\Delta_n^{1/4}}\right|^2 + \frac{C}{\epsilon}\left(\Delta_n^{2\bar{\omega}-1/2}\left|\frac{\bar{X}_{i+m}''}{\Delta_n^{\bar{\omega}}} \wedge 1\right|^2 + \left|\frac{\bar{Y}_{i+m}'}{\Delta_n^{1/4}}\right|^2 + \left|\frac{\bar{B}_{i+m}''}{\Delta_n^{1/4}}\right|^2\right)\right].$$

As $|\Delta_i^n B''| \leq C\Delta_n$, uniformly in i, we see that

$$\left|\frac{\bar{B}''}{\Delta_n^{1/4}}\right| \leq C\Delta_n^{1/4}.$$

Then, by means of the Burkholder-Davis-Gundy inequality:

$$\mathbb{E}\left[|\bar{Y}_{i+m}'|^2 \mid \mathcal{F}_{(i+m)\Delta_n}\right] \leq Ck_n\Delta_n\mathbb{E}\left[\sup_{s\in[i\Delta_n,(i+h_n+k_n)\Delta_n]}\left(|\sigma_s-\sigma_{i\Delta_n}|^2 + |\omega_s-\omega_{i\Delta_n}|^2\right) \mid \mathcal{F}_{(i+m)\Delta_n}\right]. \tag{A.3}$$

As σ and ω are càdlàg and can be assumed to be bounded, the last term converges to zero. Next, for the components involving $\bar{Y}_{i,i+m}$, it holds that (Podolskij and Vetter, 2009a, Lemma 1):

$$\mathbb{E}\left[|\bar{Y}_{i,i+m}|^4 \mid \mathcal{F}_{(i+m)\Delta_n}\right] \leq C\Delta_n. \tag{A.4}$$

In addition, Corollary 2.1.9 in Jacod and Protter (2012) applied to the pre-averaging case yields:

$$\mathbb{E}\left[\left|\frac{\bar{X}''_{i+m}}{\Delta_n^{\bar{\omega}}} \wedge 1\right|^2 \mid \mathcal{F}_{(i+m)\Delta_n}\right] \leq C\Delta_n^{1/2-2\bar{\omega}}\phi_n,$$

where $(\phi_n)_{n\in\mathbb{N}}$ is a sequence of real numbers such that $\phi_n \to 0$ as $n \to \infty$. This, along with (A.3) − (A.4), implies that:

$$\mathbb{E}\left[|S_n|\right] \leq C\left(\Delta_n^{1/2-2\bar{\omega}} + \epsilon + \frac{1}{\epsilon}\left(\phi_n + \mathbb{E}\left[\sup_{s\in[i\Delta_n,(i+h_n+k_n)\Delta_n]}\left(|\sigma_s - \sigma_{i\Delta_n}|^2 + |\omega_s - \omega_{i\Delta_n}|^2\right)\right] + \Delta_n^{1/2}\right)\right).$$

By letting first n to infinity, and then sending ϵ to zero, we complete the proof of Lemma 2.1. ∎

Proof of Theorem 3.1

First, we prove that $\hat{\omega}_t^2 \xrightarrow{\mathbb{P}} \omega_t^2$ for each $t \in [0, T]$. Due to the definition of $\hat{\omega}_{i\Delta_n}^2$ and the càdlàg property of ω_t, it suffices to show that $\hat{\omega}_{i\Delta_n}^2 \xrightarrow{\mathbb{P}} \omega_{i\Delta_n}^2$ for each i, where

$$\hat{\omega}_{i\Delta_n}^2 = -\frac{1}{h_n-1}\sum_{m=1}^{h_n-1}\Delta_{i+m}^n Z\Delta_{i+m+1}^n Z.$$

Note that $\hat{\omega}_{i\Delta_n}^2 = A_n^{(1)} + A_n^{(2)}$, where

$$A_n^{(1)} = -\frac{1}{h_n-1}\sum_{m=1}^{h_n-1}\Delta_{i+m}^n U\Delta_{i+m+1}^n U,$$

$$A_n^{(2)} = -\frac{1}{h_n-1}\sum_{m=1}^{h_n-1}\Delta_{i+m}^n U\Delta_{i+m+1}^n X + \Delta_{i+m}^n X\Delta_{i+m+1}^n U + \Delta_{i+m}^n X\Delta_{i+m+1}^n X.$$

In view of (2.1.44) in Jacod and Protter (2012), we obtain that $\sup_{1\leq m\leq h_n}\mathbb{E}\left[(\Delta_{i+m}^n X)^2\right] \leq C\Delta_n$. This, together with $\sup_{1\leq m\leq h_n}\mathbb{E}\left[(\Delta_{i+m}^n U)^2\right] \leq C$, yields that $\mathbb{E}\left[|A_n^{(2)}|\right] \leq C\Delta_n^{1/2}$. For the main term $A_n^{(1)}$, we use a further decomposition:

$$A_n^{(1)} - \omega_{i\Delta_n}^2 = -\frac{1}{h_n-1}\sum_{m=1}^{h_n-1}\left(\Delta_{i+m}^n U\Delta_{i+m+1}^n U - \omega_{i\Delta_n}^2\Delta_{i+m}^n\varepsilon\Delta_{i+m+1}^n\varepsilon\right)$$

$$-\omega_{i\Delta_n}^2\frac{1}{h_n-1}\sum_{m=1}^{h_n-1}\left(\Delta_{i+m}^n\varepsilon\Delta_{i+m+1}^n\varepsilon + 1\right) \equiv B_n^{(1)} + B_n^{(2)}.$$

29

We immediately notice that $B_n^{(2)} \xrightarrow{\mathbb{P}} 0$. Moreover,

$$\mathbb{E}\left[|B_n^{(1)}|\right] \leq C\left(\mathbb{E}\left[\sup_{s \in [i\Delta_n, (i+h_n)\Delta_n]} |\omega_s - \omega_{i\Delta_n}|^2\right]\right)^{1/2} \to 0,$$

as ω is càdlàg. This verifies that $\hat{\omega}_{i\Delta_n}^2 \xrightarrow{\mathbb{P}} \omega_{i\Delta_n}^2$.

The rest of the proof follows the structure of Lemma 1 and Theorem 1 in Li, Todorov, and Tauchen (2013). We start by showing that $\mathbb{P}(V_t = x) = 0$. By assumption, F_T is continuous almost surely, so letting $F_T(x-) = \lim_{y \uparrow x} F_T(y)$, it holds that $\mathbb{P}(F_T(x) \neq F_T(x-)) = 0$. Hence,

$$\int_0^T \mathbb{P}(V_t = x)dt = \mathbb{E}\left[\int_0^T (1_{\{V_t \leq x\}} - 1_{\{V_t < x\}})dt\right] = T\mathbb{E}(F_T(x) - F_T(x-)) = 0.$$

Thus, $\mathbb{P}(V_t = x) = 0$ for a.e. $t \in [0, T]$. Next, we show the pointwise consistency of $F_{n,T}(x)$. We therefore fix $x \in \mathbb{R}_+$ and notice that

$$\mathbb{E}\left[|F_{n,T}(x) - F_T(x)|\right] \leq \frac{1}{T}\int_0^T \mathbb{E}\left[|1_{\{\hat{V}_t \leq x\}} - 1_{\{V_t \leq x\}}|\right]dt$$

$$= \frac{1}{T}\int_0^T \mathbb{E}\left[|(1_{\{\hat{V}_t \leq x\}} - 1_{\{V_t \leq x\}})1_{\{V_t \neq x\}}|\right]dt.$$

Applying Lemma 2.1 and Lebesgue's dominated convergence theorem (twice), we deduce that $\mathbb{E}\left[|F_{n,T}(x) - F_T(x)|\right] \to 0$, which implies pointwise consistency, $F_{n,T}(x) \xrightarrow{\mathbb{P}} F_T(x)$. As $F_{n,T}$ and F_T are non-decreasing functions and F_T is continuous a.s., this convergence is locally uniform (see, e.g., Resnick, 1998). Furthermore, as the paths of σ_t are càdlàg, for any $\eta > 0$ there exists a constant M such that $\mathbb{P}(\sup_{t \in [0,T]} V_t > M) < \eta$, which means that $\mathbb{P}(F_T(M) \neq 1) < \eta$. In turn,

$$\mathbb{P}\left(\sup_{x \in \mathbb{R}_+} |F_{n,T}(x) - F_T(x)| > \epsilon\right) \leq \mathbb{P}\left(\sup_{x \in [0,M]} |F_{n,T}(x) - F_T(x)| > \epsilon\right) + \mathbb{P}\left(\sup_{x \geq M} |F_{n,T}(x) - F_T(x)| > \epsilon\right)$$

$$\leq \mathbb{P}\left(\sup_{x \in [0,M]} |F_{n,T}(x) - F_T(x)| > \epsilon\right) + \mathbb{P}\left(\sup_{x \geq M} |F_{n,T}(x) - F_{n,T}(M)| > \frac{\epsilon}{3}\right)$$

$$+ \mathbb{P}\left(\sup_{x \geq M} |F_{n,T}(M) - F_T(M)| > \frac{\epsilon}{3}\right) + \mathbb{P}\left(\sup_{x \geq M} |F_T(M) - F_T(x)| > \frac{\epsilon}{3}\right)$$

$$\leq \mathbb{P}\left(\sup_{x \in [0,M]} |F_{n,T}(x) - F_T(x)| > \epsilon\right) + \mathbb{P}\left(1 - F_{n,T}(M) > \frac{\epsilon}{3}\right)$$

$$+ \mathbb{P}\left(|F_{n,T}(M) - F_T(M)| > \frac{\epsilon}{3}\right) + \mathbb{P}\left(1 - F_T(M) > \frac{\epsilon}{3}\right).$$

Taking $\limsup$ and using the pointwise and locally uniform convergence:

$$\limsup_{n \to \infty} \mathbb{P}\left(\sup_{x \in \mathbb{R}_+} |F_{n,T}(x) - F_T(x)| > \epsilon\right) \leq 2\eta,$$

30

and since η is arbitrary, this shows that

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}_+}\left|F_{n,T}(x)-F_T(x)\right|>\epsilon\right)\to 0,$$

as $\Delta_n\to 0$. ■

Three auxiliary results

In this section, we prove three auxiliary lemmas that are used throughout the subsequent proofs. Before stating the first, we need some additional notation.

Let X' denote the continuous part of X. Next, define $X''=X-X'$ and $Z'_{i\Delta_n}=X'_{i\Delta_n}+U_{i\Delta_n}$, and let $\check{V}_{i\Delta_n}$ be a version of $\hat{V}_{i\Delta_n}$ that uses Z' instead of Z, i.e.

$$\check{V}_{i,n}=\frac{1}{\theta\psi_2}\frac{1}{h_n\sqrt{\Delta_n}}\sum_{m=0}^{h_n-1}\left(\bar{Z}'_{i+m}\right)^2 1_{\{|\bar{Z}'_{i+m}|\leq v_n\}}+\frac{\psi_1}{\theta^2\psi_2}\frac{1}{h_n-1}\sum_{m=1}^{h_n-1}\Delta^n_{i+m}Z'\Delta^n_{i+m+1}Z'.$$

Lemma A.1 *Suppose that Assumption 1 holds, and let $h_n/k_n\to\infty$ as $\Delta_n\to 0$. Then, for any $r\in(0,2)$ and some $\iota>0$,*

$$\mathbb{E}\left[|\hat{V}_{i\Delta_n}-\check{V}_{i\Delta_n}|^p\right]\leq C\left((h_n/k_n)^{1-p}\Delta_n^{1/2-r\bar{\omega}-2p(1/4-\bar{\omega})-\iota}+\Delta_n^{((2-r)\bar{\omega}-\iota)p}+h_n^{-p/2}\Delta_n\right),$$

uniformly in i for $p\geq 2$.

Proof. First, we observe that:

$$\hat{V}_{i\Delta_n}-\check{V}_{i\Delta_n}=S^{(1)}_{i,n}+S^{(2)}_{i,n},$$

where

$$S^{(1)}_{i,n}=\frac{1}{\theta\psi_2}\frac{1}{h_n\sqrt{\Delta_n}}\sum_{m=0}^{h_n-1}\left(\left(\bar{Z}_{i+m}\right)^2 1_{\{|\bar{Z}_{i+m}|\leq v_n\}}-\left(\bar{Z}'_{i+m}\right)^2 1_{\{|\bar{Z}'_{i+m}|\leq v_n\}}\right),$$

$$S^{(2)}_{i,n}=\frac{\psi_1}{\theta^2\psi_2}\frac{1}{h_n-1}\sum_{m=1}^{h_n-1}\left(\Delta^n_{i+m}Z\Delta^n_{i+m+1}Z-\Delta^n_{i+m}Z'\Delta^n_{i+m+1}Z'\right).$$

Next, for any $v\geq 1$ and $a\geq 0$, we note that:

$$\left|\,|x+y|^2 1_{\{|x+y|\leq v\}}-|x|^2 1_{\{|x|\leq v\}}\right|\leq C\left(\frac{|x|^{2+a}}{v^a}+(1+|x|)(|y|\wedge v)^2\right),$$

which can be verified by looking at the cases:

$(1)\,|x+y|\leq v,|x|\leq v\;\;(2)\,|x+y|\leq v,|x|>v,\;\;(3)\,|x+y|>v,v/2<|x|\leq v,\;\;(4)\,|x+y|>v,|x|\leq v/2.$

31

If we apply this inequality with $x = \Delta_n^{-1/4} \bar{Z}'_{i+m}$, $y = \Delta_n^{-1/4} \bar{X}''_{i+m}$, and $v = \Delta_n^{-1/4} v_n = \alpha \Delta_n^{\bar{\omega}-1/4}$ (and n large enough such that $v \geq 1$), it leads to:

$$\left|S_{i,n}^{(1)}\right| \leq \frac{C}{h_n} \sum_{m=0}^{h_n-1} \left(\Delta_n^{a(1/4-\bar{\omega})} \left|\Delta_n^{-1/4} \bar{Z}'_{i+m}\right|^{2+a} + \Delta_n^{2(\bar{\omega}-1/4)} \left(1 + \left|\Delta_n^{-1/4} \bar{Z}'_{i+m}\right|\right) \left(\frac{\bar{X}''_{i+m}}{\Delta_n^{\bar{\omega}}} \wedge 1\right)^2 \right).$$

Another application of Corollary 2.1.9 in Jacod and Protter (2012) to the pre-averaging framework implies that, for any $q \geq 2$,

$$\mathbb{E}\left[\left|\frac{\bar{X}''_{i+m}}{\Delta_n^{\bar{\omega}}} \wedge 1\right|^q \mid \mathcal{F}_{(i+m)\Delta_n}\right] \leq \Delta_n^{1/2-r\bar{\omega}} \phi_n,$$

where $\phi_n \to 0$. Let

$$\chi_i^n = \Delta_n^{a(1/4-\bar{\omega})} \left|\Delta_n^{-1/4} \bar{Z}'_i\right|^{2+a} + \Delta_n^{2(\bar{\omega}-1/4)} \left(1 + \left|\Delta_n^{-1/4} \bar{Z}'_i\right|\right) \left(\frac{\bar{X}''_i}{\Delta_n^{\bar{\omega}}} \wedge 1\right)^2.$$

Then, for any $p \geq 1$ and some arbitrarily small $\iota > 0$, Hölder's inequality leads to:

$$\mathbb{E}\left[|\chi_i^n|^p \mid \mathcal{F}_{i\Delta_n}\right] \leq C\left(\Delta_n^{pa(1/4-\bar{\omega})} + \Delta_n^{(1/2-r\bar{\omega})b-p(1/2-2\bar{\omega})}\right) \leq C\Delta_n^{(1/2-r\bar{\omega})-p(1/2-2\bar{\omega})-\iota},$$

where the last part follows by taking large enough a (recall $1/4 - \bar{\omega} > 0$). A successive conditioning argument combined with Burkholder-Gundis-Davis' inequality results in:

$$\mathbb{E}\left[\left|S_{i,n}^{(1)}\right|^p\right] \leq C\left((h_n/k_n)^{1-p} \Delta_n^{1/2-r\bar{\omega}-p(1/2-2\bar{\omega})} + \Delta_n^{((2-r)\bar{\omega}-\iota)p}\right).$$

We proceed with the term $S_{i,n}^{(2)}$. Notice that:

$$S_{i,n}^{(2)} = \frac{\psi_1}{\theta^2 \psi_2} \frac{1}{h_n - 1} \sum_{m=1}^{h_n-1} \zeta_{i+m}^n,$$

where $\zeta_i^n = \Delta_i^n Z' \Delta_{i+1}^n X'' + \Delta_i^n X'' \Delta_{i+1}^n Z' + \Delta_i^n X'' \Delta_{i+1}^n X''$. Section 2.1 in Jacod and Protter (2012) then implies that, for any $p \geq 2$, $\mathbb{E}\left[|\Delta_i^n Z'|^p \mid \mathcal{F}_{(i-1)\Delta_n}\right] \leq C$ and $\mathbb{E}\left[|\Delta_i^n X''|^p \mid \mathcal{F}_{(i-1)\Delta_n}\right] \leq C\Delta_n^2$. This means that $\mathbb{E}\left[|\zeta_i^n|^p \mid \mathcal{F}_{(i-1)\Delta_n}\right] \leq C\Delta_n$, for any $p \geq 1$. After subtracting the conditional means, an application of the discrete Burkholder inequality then delivers that:

$$\mathbb{E}\left[\left|S_n^{(2)}\right|^p\right] \leq C\left(h_n^{-p/2} \Delta_n + \Delta_n^p\right) \leq C h_n^{-p/2} \Delta_n,$$

and this concludes the proof. ∎

Lemma A.2 *Suppose that Assumption 1(i) with $H \in (0,1)$ and 4 – 6 hold true, as does the rate condition $T^{1/2+\iota}\left(\Delta_n^{\frac{H}{4H+2}-\iota} \vee \Delta_n^{\frac{H}{2H+1}-(1/2-2\bar{\omega})-\iota} \vee \Delta_n^{(2-r)\bar{\omega}-\iota}\right) \to 0$, as $\Delta_n \to 0$ and $T \to \infty$, for some $\iota > 0$ and $h_n \asymp \Delta_n^{-\frac{4H+1}{4H+2}}$. Finally, let $r \in [0,2)$ and $\bar{\omega} \in \left(\frac{1}{8H+4}, \frac{1}{4}\right)$. Then,*

$$\sqrt{T} \sup_{x \in \mathbb{R}_+} \left|F_{n,T}(x) - F_T(x)\right| \xrightarrow{\mathbb{P}} 0. \tag{A.5}$$

Proof. The proof consists of three parts. First, we show that for each i and $p \geq 2$:

$$\mathbb{E}\left[|\check{V}_{i\Delta_n} - V_{i\Delta_n}|^p\right] \leq C\left[(k_n/h_n)^{p/2} + (h_n\Delta_n)^{pH} + \Delta_n^{p/4}\right]. \tag{A.6}$$

Note that

$$\hat{V}_{i\Delta_n} - V_{i\Delta_n} = R_n^{(0)} + R_n^{(1)} + R_n^{(2)} + R_n^{(3)} + R_n^{(4)} + R_n^{(5)}, \tag{A.7}$$

where

$$R_n^{(0)} = -\frac{1}{h_n\sqrt{\Delta_n}} \sum_{m=0}^{h_n-1} \left(\bar{Z}_{i+m}\right)^2 1_{\{|\bar{Z}_{i+m}|>v_n\}},$$

$$R_n^{(1)} = \frac{1}{\theta\psi_2} \frac{1}{h_n\sqrt{\Delta_n}} \sum_{m=0}^{h_n-1} \bar{X}_{i+m}^2 - \sigma_{i\Delta_n}^2 \bar{W}_{i+m}^2,$$

$$R_n^{(2)} = \frac{1}{\theta\psi_2} \frac{2}{h_n} \sum_{m=0}^{h_n-1} \Delta_n^{-1/2} \bar{X}_{i+m} \bar{U}_{i+m},$$

$$R_n^{(3)} = \frac{1}{\theta\psi_2} \frac{1}{h_n} \sum_{m=0}^{h_n-1} \left(\Delta_n^{-1/2} \bar{U}_{i+m}^2 - \frac{\psi_1}{\theta}\omega_{i\Delta_n}^2\right),$$

$$R_n^{(4)} = \frac{\psi_1}{\theta^2\psi_2}(\omega_{i\Delta_n}^2 - \hat{\omega}_{i\Delta_n}^2),$$

$$R_n^{(5)} = \sigma_{i\Delta_n}^2 \left(\frac{k_n\Delta_n^{1/2}\psi_2^n}{\theta\psi_2} - 1\right) + \frac{1}{\theta\psi_2} \frac{1}{h_n} \sum_{m=0}^{h_n-1} \sigma_{i\Delta_n}^2 \left(\Delta_n^{-1/2}\bar{W}_{i+m}^2 - k_n\Delta_n^{1/2}\psi_2^n\right).$$

To deal with the term $R_n^{(0)}$, fix $p \geq 2$. We choose large q satisfying $q(1/4 - \bar{\omega}) \geq p/4$. Then:

$$\mathbb{E}\left[|\left(\bar{Z}_{i+m}\right)^2 1_{\{|\bar{Z}_{i+m}|>v_n\}}|^p\right] \leq v_n^{-q}\mathbb{E}\left[|\left(\bar{Z}_{i+m}\right)^{2p+q}|\right] \leq C\Delta_n^{p/4+p/2}.$$

Invoking the discrete Hölder inequality yields $\mathbb{E}\left[|R_n^{(0)}|^p\right] \leq C\Delta_n^{p/4}$.

To deal with the remaining terms, consider the random variable:

$$S_n = \sum_{m=0}^{l_n k_n - 1} Y_m = \sum_{i=1}^{k_n} S_n^{(i)},$$

where $(Y_m)_{m\geq 1}$ are mean zero and $(k_n - 1)$-dependent, such that for each i $S_n^{(i)} = \sum_{m=0}^{l_n} Y_{i-1+mk_n}$ is a sum of martingale differences and $\sup_{m\in\mathbb{N}} \mathbb{E}\left[|Y_m|^p\right] \leq C$. Applying the discrete Hölder and Burkholder inequality, we obtain for each $p \geq 2$:

$$\mathbb{E}\left[|S_n|^p\right] \leq k_n^{p-1} \sum_{i=1}^{k_n} \mathbb{E}\left[|S_n^{(i)}|^p\right] \leq Ck_n^p l_n^{p/2}.$$

This observation leads to

$$\mathbb{E}\left[|R_n^{(2)}|^p\right] + \mathbb{E}\left[|R_n^{(3)}|^p\right] + \mathbb{E}\left[|R_n^{(5)}|^p\right] \leq C\left((k_n/h_n)^{p/2} + \Delta_n^{p/4}\right).$$

33

In addition, using the notation and results derived in the proof of Theorem 3.1, the discrete Hölder and Burkholder inequalities imply that:

$$\mathbb{E}\Big[\big|R_n^{(4)}\big|^p\Big] \le C\bigg(\mathbb{E}\Big[\big|A_n^{(2)}\big|^p\Big] + \mathbb{E}\Big[\big|B_n^{(1)}\big|^p\Big] + \mathbb{E}\Big[\big|B_n^{(2)}\big|^p\Big]\bigg) \le C\big(\Delta_n^{p/2} + (h_n\Delta_n)^p + h_n^{-p/2}\big).$$

Turning next to $R_n^{(1)}$, we have for each i and m that

$$\bar{X}_{i+m} - \sigma_{i\Delta_n}\bar{W}_{i+m} = \sum_{j=1}^{k_n-1} g(j/k_n)A_{i+m+j},$$

where

$$A_{i+m+j} = \int_{(i+m+j-1)\Delta_n}^{(i+m+j)\Delta_n} b_s ds + \int_{(i+m+j-1)\Delta_n}^{(i+m+j)\Delta_n} (\sigma_s - \sigma_{i\Delta_n})dW_s.$$

The Burkholder inequality then implies that, for each $p \ge 2$,

$$\mathbb{E}\Big[\big|\bar{X}_{i+m} - \sigma_{i\Delta_n}\bar{W}_{i+m}\big|^p\Big] \le C\Big(\big(k_n\Delta_n\big)^p + \big(k_n\Delta_n\big)^{p/2}\big(h_n\Delta_n\big)^{pH}\Big) \le C\Big(\Delta_n^{p/2} + \Delta_n^{p/4}\big(h_n\Delta_n\big)^{pH}\Big),$$

uniformly in i and m. Using the above together with the identity $|a^2 - b^2|^p = |a - b|^p|a + b|^p$, plus the discrete Hölder and Cauchy-Schwarz inequality, we deduce that:

$$\mathbb{E}\Big[\big|R_n^{(1)}\big|^p\Big] \le \frac{C}{h_n} \sum_{m=0}^{h_n-1} \mathbb{E}\Big[\big|\Delta_n^{-1/4}\big(\bar{X}_{i+m} - \sigma_{i\Delta_n}\bar{W}_{i+m}\big)\big|^p\big|\Delta_n^{-1/4}\big(\bar{X}_{i+m} + \sigma_{i\Delta_n}\bar{W}_{i+m}\big)\big|^p\Big]$$
$$\le C\Big(\Delta_n^{p/4} + \big(h_n\Delta_n\big)^{pH}\Big),$$

which finishes the proof of (A.6).

We continue with the second part. We denote $\eta_{n,T} = \sup_{t\in[0,T]}\big|\hat{V}_t - V_t\big|$ and notice that

$$\eta_{n,T} \le \sup_{i\in A_{n,T}} |\hat{V}_{i\Delta_n} - V_{i\Delta_n}| + \sup_{i\in A_{n,T}} \sup_{t\in(i\Delta_n,(i+1)\Delta_n]} |V_t - V_{i\Delta_n}|.$$

where $A_{n,T} = \{i : 0 \le i \le [T/\Delta_n] - h_n - k_n\}$. Now, Assumption 1(i) implies:

$$\mathbb{E}\Big[\big|\sup_{t\in(i\Delta_n,(i+1)\Delta_n]} |V_t - V_{i\Delta_n}|\big|^p\Big] \le C\Delta_n^{pH}.$$

This result, combined with Lemma A.1, inequality in means, the maximal inequality and (A.6), implies that for some $\iota > 0$,

$$\mathbb{E}\Big[\big|\eta_{n,T}\big|^p\Big] \le C\frac{T}{\Delta_n}\Big[\big(k_n/h_n\big)^{p/2} + \big(h_n\Delta_n\big)^{pH} + \big(h_n/k_n\big)^{1-p}\Delta_n^{1/2-r\bar\omega-2p(1/4-\bar\omega)-\iota} + \Delta_n^{((2-r)\bar\omega-\iota)p}\Big],$$

which immediately delivers the bound:

$$\mathbb{E}\Big[\big|\eta_{n,T}\big|^p\Big]^{1/p} \le CT^{1/p}\Delta_n^{-1/p}\Big[\big(k_n/h_n\big)^{1/2}+\big(h_n\Delta_n\big)^{H}+(h_n/k_n)^{(1-p)/p}\Delta_n^{(1/2-r\bar\omega-2p(1/4-\bar\omega)-\iota)/p}+\Delta_n^{(2-r)\bar\omega-\iota}\Big].$$

As p can be arbitrarily large and optimizing $h_n \asymp \Delta_n^{-\frac{4H+1}{4H+2}}$, we deduce that

$$\eta_{n,T} = O_p\left(T^\iota\left(\Delta_n^{\frac{H}{4H+2}-\iota} \vee \Delta_n^{\frac{H}{2H+1}-(1/2-2\bar\omega)-\iota} \vee \Delta_n^{(2-r)\bar\omega-\iota}\right)\right).$$

As for the final step in the proof, recall that by definition of $\eta_{n,T}$:

$$F_T(x - \eta_{n,T}) \le F_{n,T}(x) \le F_T(x + \eta_{n,T}).$$

Then,

$$\sqrt{T}\big|F_{n,T}(x) - F_T(x)\big| \le \sqrt{T}\big|F_T(x + \eta_{n,T}) - F_T(x - \eta_{n,T})\big|$$

$$\le \big|G_T(x + \eta_{n,T}) - G_T(x - \eta_{n,T})\big| + \sqrt{T}\big|F(x + \eta_{n,T}) - F(x - \eta_{n,T})\big|, \tag{A.8}$$

where $G_T(x) = \sqrt{T}(F_T(x) - F(x))$.

As the density of F exists and is bounded:

$$\sqrt{T}\sup_{x\in\mathbb{R}_+}|F(x + \eta_{n,T}) - F(x - \eta_{n,T})| = \sqrt{T}\sup_{x\in\mathbb{R}_+}\int_{x-\eta_{n,T}}^{x+\eta_{n,T}}f(z)dz \le 2K\sqrt{T}\eta_{n,T},$$

which—by the rate condition on $\eta_{n,T}$—implies that $\sqrt{T}\sup_{x\in\mathbb{R}_+}|F(x + \eta_{n,T}) - F(x - \eta_{n,T})| \xrightarrow{\mathbb{P}} 0$. And since $\eta_{n,T} \xrightarrow{\mathbb{P}} 0$, it follows that for any $\epsilon, \delta > 0$ there exists (n', T') such that $\mathbb{P}(\eta_{n,T} > \delta) \le \epsilon/2$ for all pairs (n, T) with $n \ge n'$ and $T \ge T'$, which adheres to the rate condition. Let $\gamma > 0$ be given. Then,

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}_+}\big|G_T(x + \eta_{n,T}) - G_T(x - \eta_{n,T})\big| > \gamma\right) \le \mathbb{P}\left(\{\sup_{x\in\mathbb{R}_+}\big|G_T(x + \eta_{n,T}) - G_T(x - \eta_{n,T})\big| > \gamma\} \cap \{\eta_{n,T} < \delta\}\right)$$

$$+ \mathbb{P}\left(\eta_{n,T} \ge \delta\right)$$

$$\le \mathbb{P}\left(\{\sup_{|x-y|\le 2\delta}\big|G_T(x) - G_T(y)\big| > \gamma\}\right) + \mathbb{P}\left(\eta_{n,T} \ge \delta\right). \tag{A.9}$$

It follows from Theorem 5.2 in Dehay (2005) that for any $\epsilon, \gamma > 0$ there exists $\delta > 0$ such that the first term in (A.9) is smaller than $\epsilon/2$. Hence,

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}_+}\big|G_T(x + \eta_{n,T}) - G_T(x - \eta_{n,T})\big| > \gamma\right) \le \epsilon.$$

As ϵ, γ are arbitrary, this verifies that $\sup_{x\in\mathbb{R}_+}|G_T(x + \eta_{n,T}) - G_T(x - \eta_{n,T})| \xrightarrow{\mathbb{P}} 0$, concluding the proof of Lemma A.2. $\blacksquare$

Lemma A.3 *Suppose that Assumption 1*(ii) *and 4 – 6 hold true, as does the rate condition*
$T^{1/2+\iota} \left(\Delta_n^{1/8-\iota} \vee \Delta_n^{2\bar{\omega}-\frac{1}{4}} \vee \Delta_n^{(2-r)\bar{\omega}-\iota} \right) \to 0$, *as* $\Delta_n \to 0$ *and* $T \to \infty$, *for some* $\iota > 0$ *and* $h_n \asymp \Delta_n^{-3/4}$.
Finally, let $r \in [0, 2)$ *and* $\bar{\omega} \in \left(\frac{1}{8}, \frac{1}{4} \right)$. *Then,*

$$\sqrt{T} \sup_{x \in \mathbb{R}_+} \left| F_{n,T}(x) - F_T(x) \right| \xrightarrow{\mathbb{P}} 0. \tag{A.10}$$

Proof. Recall that if σ_t satisfies Assumption 1(ii), the form of $V_t = \sigma_t^2$ is representable via (the general version of) Itô's Lemma:

$$V_t = V_0 + \int_0^t \bar{b}_s ds + \int_0^t \bar{\sigma}_s dW_s + \int_0^t \bar{\sigma}'_s dB_s + \int_0^t \int_{\mathbb{R}} \bar{\delta}(s, z)(\mu - \nu)(d(s, z)).$$

We split the variance in two: $V_t = V'_t(\epsilon) + V''_t(\epsilon)$, for each $\epsilon > 0$, where

$$V_t^c = V_0 + \int_0^t \bar{b}_s ds + \int_0^t \bar{\sigma}_s dW_s + \int_0^t \bar{\sigma}'_s dB_s,$$

$$V'_t(\epsilon) = V_t^c + \int_0^t \int_{\{z : \tilde{\Gamma}(z) \leq \epsilon\}} \bar{\delta}(s, z)(\mu - \nu)(d(s, z)) - \int_0^t \int_{\{z : \tilde{\Gamma}(z) > \epsilon\}} \bar{\delta}(s, z)\nu(d(s, z)),$$

$$V''_t(\epsilon) = \int_0^t \int_{\{z : \tilde{\Gamma}(z) > \epsilon\}} \bar{\delta}(s, z)\mu(d(s, z)).$$

Moreover, we denote

$$J_n(\epsilon) = \{ i \in \mathbb{Z} : 0 \leq i \leq [T/\Delta_n] - h_n - k_n, \mu([i\Delta_n, (i+1)\Delta_n) \times \{z : \tilde{\Gamma}(z) > \epsilon\}) = 0 \},$$

$$S_n(\epsilon) = \cup_{i \in J_n(\epsilon)} [i\Delta_n, (i+1)\Delta_n),$$

$$A_{n,T} = \{ i : 0 \leq i \leq [T/\Delta_n] - h_n - k_n \}.$$

In words, $J_n(\epsilon)$ is a set of indices that excludes the big jumps and $S_n(\epsilon)$ is the union of these. It leads to a decomposition $F_{n,T}(x) = F_{n,T}(x, \epsilon) + R_{n,T}(x, \epsilon)$ and $F_T(x) = F_T(x, \epsilon) + R_T(x, \epsilon)$ with

$$F_{n,T}(x, \epsilon) = \frac{1}{T} \int_{S_n(\epsilon)} 1_{\{\hat{V}_t \leq x\}} dt, \quad R_{n,T}(x, \epsilon) = \frac{1}{T} \int_{[0,T] \setminus S_n(\epsilon)} 1_{\{\hat{V}_t \leq x\}} dt,$$

$$F_T(x, \epsilon) = \frac{1}{T} \int_{S_n(\epsilon)} 1_{\{V_t \leq x\}} dt, \quad R_T(x, \epsilon) = \frac{1}{T} \int_{[0,T] \setminus S_n(\epsilon)} 1_{\{V_t \leq x\}} dt.$$

Here, F is the main term, while R is a remainder. A final piece of notation:

$$\eta_{n,T}(\epsilon) = \sup_{i \in J_n(\epsilon)} \left| \hat{V}_{i\Delta_n} - V_{i\Delta_n} \right| + \sup_{i \in A_{n,T}} \sup_{t \in (i\Delta_n, (i+1)\Delta_n]} \left| V'_t(\epsilon) - V'_{i\Delta_n}(\epsilon) \right|.$$

Let $\epsilon_n \asymp \Delta_n^{1/8}$. The aim is to prove that

$$\eta_{n,T}(\epsilon_n) = O_p(T^\iota \Delta_n^{1/8-\iota}). \tag{A.11}$$

We deal first with the last term of $\eta_{n,T}(\epsilon_n)$, which we call

$$\tilde{\eta}_{n,T}(\epsilon_n) = \sup_{i \in A_{n,T}} \sup_{t \in (i\Delta_n, (i+1)\Delta_n]} |V_t'(\epsilon_n) - V_{i\Delta_n}'(\epsilon_n)|.$$

Observe that

$$\sup_{t \in (i\Delta_n, (i+1)\Delta_n]} |V_t'(\epsilon_n) - V_{i\Delta_n}'(\epsilon_n)| \leq \gamma_{i,n}^{(1)} + \gamma_{i,n}^{(2)} + \gamma_{i,n}^{(3)},$$

where

$$\gamma_{i,n}^{(1)} = \sup_{t \in (i\Delta_n, (i+1)\Delta_n]} |V_t^c - V_{i\Delta_n}^c|,$$

$$\gamma_{i,n}^{(2)} = \sup_{t \in (i\Delta_n, (i+1)\Delta_n]} \Big| \int_{i\Delta_n}^t \int_{\{z:\tilde{\Gamma}(z) \leq \epsilon_n\}} \bar{\delta}(s,z)(\mu - \nu)(ds, dz) \Big|,$$

$$\gamma_{i,n}^{(3)} = \sup_{t \in (i\Delta_n, (i+1)\Delta_n]} \Big| \int_{i\Delta_n}^t \int_{\{z:\tilde{\Gamma}(z) > \epsilon_n\}} \bar{\delta}(s,z)\nu(ds, dz) \Big|.$$

Application of the Burkholder and maximal inequality yield that, for any $p \geq 2$,

$$\mathbb{E}\Big[|\sup_{i \in A_{n,T}} \gamma_{i,n}^{(1)}|^p \Big] \leq CT\Delta_n^{p/2-1}. \tag{A.12}$$

Moreover, for each $p \geq 2$ and i, Lemma 2.1.5 in Jacod and Protter (2012) implies that

$$\mathbb{E}\Big[|\gamma_{i,n}^{(2)}|^p \Big] \leq C\Delta_n \int_{\{z:\tilde{\Gamma}(z) \leq \epsilon_n\}} \tilde{\Gamma}(z)^p \lambda(dz) + C\Delta_n^{p/2} \Big(\int_{\{z:\tilde{\Gamma}(z) \leq \epsilon_n\}} \tilde{\Gamma}(z)^2 \lambda(dz) \Big)^{p/2}$$

$$\leq C\big(\Delta_n \epsilon_n^{p-2} + \Delta_n^{p/2} \big).$$

In view of the maximal inequality:

$$\mathbb{E}\Big[|\sup_{i \in A_{n,T}} \gamma_{i,n}^{(2)}|^p \Big] \leq CT\big(\epsilon_n^{p-2} + \Delta_n^{p/2-1} \big). \tag{A.13}$$

Finally,

$$|\sup_{i \in A_{n,T}} \gamma_{i,n}^{(3)}| \leq \sup_{i \in A_{n,T}} \int_{i\Delta_n}^{(i+1)\Delta_n} \int_{\{z:\tilde{\Gamma}(z) > \epsilon_n\}} \tilde{\Gamma}(z)\nu(ds, dz) \leq C\Delta_n \epsilon_n^{-1}. \tag{A.14}$$

As a result, for any $p \geq 2$, a combination of (A.12) - (A.14) shows that

$$\mathbb{E}\Big[|\tilde{\eta}_{n,T}(\epsilon_n)|^p \Big]^{1/p} \leq C\big(T^{1/p}\Delta_n^{1/2-1/p} + T^{1/p}\epsilon_n^{1-2/p} + \Delta_n \epsilon_n^{-1} \big). \tag{A.15}$$

Now, turn to the first term of $\eta_{n,T}(\epsilon_n)$, which is $\hat{\eta}_{n,T}(\epsilon_n) \equiv \sup_{i \in J_n(\epsilon_n)} |\hat{V}_{i\Delta_n} - V_{i\Delta_n}|$. As in (A.7) in the proof of Lemma A.2, we disentangle, for each $i \in J_n(\epsilon_n)$,

$$\check{V}_{i\Delta_n} - V_{i\Delta_n} = R_n^{(1)} + R_n^{(2)} + R_n^{(3)} + R_n^{(4)} + R_n^{(5)}.$$

It suffices to look at $R_n^{(1)}$, as the rates of the other four terms are equal. As above, utilizing Burkholder's inequality, we see that for each $p \geq 2$, $i \in J_n(\epsilon_n)$ and $0 \leq m \leq h_n - 1$:

$$\mathbb{E}\Big[|\bar{X}'_{i+m} - \sigma_{i\Delta_n}\bar{W}_{i+m}|^p\Big] \leq C\Big((k_n\Delta_n)^p + (k_n\Delta_n)^{p/2}\mathbb{E}\Big[\sup_{t\in[i\Delta_n,(i+h_n)\Delta_n]}|\sigma_t - \sigma_{i\Delta_n}|^p\Big]\Big)$$

$$\leq C\Big((k_n\Delta_n)^p + (k_n\Delta_n)^{p/2}\big[(h_n\Delta_n)^{p/2} + h_n\Delta_n\epsilon_n^{p-2} + (h_n\Delta_n)^p\epsilon_n^{-p}\big]\Big)$$

$$\leq C\Delta_n^{p/4}\Big[(h_n\Delta_n)^{p/2} + h_n\Delta_n\epsilon_n^{p-2} + (h_n\Delta_n)^p\epsilon_n^{-p}\Big],$$

uniformly in i and m. Proceeding as in the proof of Lemma A.2, it follows that

$$\mathbb{E}\Big[|R_n^{(1)}|^p\Big] \leq \frac{C}{h_n}\sum_{m=0}^{h_n-1}\mathbb{E}\Big[|\Delta_n^{-1/4}(\bar{X}'_{i+m} - \sigma_{i\Delta_n}\bar{W}_{i+m})|^p|\Delta_n^{-1/4}(\bar{X}'_{i+m} + \sigma_{i\Delta_n}\bar{W}_{i+m})|^p\Big]$$

$$\leq C\Big[(h_n\Delta_n)^{p/2} + h_n\Delta_n\epsilon_n^{p-2} + (h_n\Delta_n)^p\epsilon_n^{-p}\Big].$$

In conclusion, for each $p \geq 2$ and $i \in J_n(\epsilon_n)$, we get

$$\mathbb{E}\Big[|\check{V}_{i\Delta_n} - V_{i\Delta_n}|^p\Big] \leq C\Big[(k_n/h_n)^{p/2} + (h_n\Delta_n)^{p/2} + h_n\Delta_n\epsilon_n^{p-2} + (h_n\Delta_n)^p\epsilon_n^{-p}\Big].$$

Hence, for any $p \geq 2$, the maximal inequality, Lemma A.1, along with $k_n \asymp \Delta_n^{-1/2}$ and $h_n \asymp \Delta_n^{-3/4}$ yield that

$$\mathbb{E}\Big[|\sup_{i\in J_n(\epsilon)}|\hat{V}_{i\Delta_n} - V_{i\Delta_n}||^p\Big]^{1/p} \leq CT^{1/p}\Delta_n^{-1/p}\Big[\Delta_n^{1/8} + \epsilon_n^{1-2/p} + \Delta_n^{1/4}\epsilon_n^{-1}$$

$$+ \Delta_n^{((p-1)/4+1/2-r\bar{\omega}-p(1/2-2\bar{\omega}))/p} + \Delta_n^{(2-r)\bar{\omega}-\iota}\Big].$$

Merging this with (A.15):

$$\mathbb{E}\Big[|\eta_{n,T}(\epsilon_n)|^p\Big]^{1/p} \leq CT^{1/p}\Delta_n^{-1/p}\Big[\Delta_n^{1/8} + \epsilon_n^{1-2/p} + \Delta_n^{1/4}\epsilon_n^{-1} + \Delta_n^{((p-1)/4+1/2-r\bar{\omega}-p(1/2-2\bar{\omega}))/p} + \Delta_n^{(2-r)\bar{\omega}-\iota}\Big]$$

$$\leq CT^{1/p}\Delta_n^{-1/p}\Big[\Delta_n^{1/8} + \Delta_n^{1/8-1/4p} + \Delta_n^{((p-1)/4+1/2-r\bar{\omega}-p(1/2-2\bar{\omega}))/p} + \Delta_n^{(2-r)\bar{\omega}-\iota}\Big],$$

where $\epsilon_n \asymp \Delta_n^{1/8}$ was used in the last line. Now, choosing large enough p implies (A.11).

Passing back to the main proof, it holds that

$$\mathbb{E}\Big[T^{-1}\int_{[0,T]\backslash S_n(\epsilon_n)} dt\Big] \leq C\Delta_n\epsilon_n^{-2}.$$

This delivers a bound for the remainder term(s):

$$R_{n,T}(x,\epsilon_n) + R_T(x,\epsilon_n) = O_p(\Delta_n\epsilon_n^{-2}). \tag{A.16}$$

As $V_t''(\epsilon_n)$ vanishes on $S_n(\epsilon_n)$, $\sup_{t \in S_n(\epsilon_n)} |\hat{V}_t - V_t| \leq \eta_{n,T}(\epsilon_n)$. Then, proceeding as in (A.8) we deduce that

$$
\begin{aligned}
\sqrt{T}|F_{n,T}(x, \epsilon_n) - F_T(x, \epsilon_n)| &\leq \sqrt{T}\big|F_T(x + \eta_{n,T}(\epsilon_n), \epsilon_n) - F_T(x - \eta_{n,T}(\epsilon_n), \epsilon_n)\big| \\
&\leq \sqrt{T}\big|F_T(x + \eta_{n,T}(\epsilon_n)) - F_T(x - \eta_{n,T}(\epsilon_n))\big| \\
&\leq \big|G_T(x + \eta_{n,T}(\epsilon_n)) - G_T(x - \eta_{n,T}(\epsilon_n))\big| + \sqrt{T}\big|F(x + \eta_{n,T}(\epsilon_n)) - F(x - \eta_{n,T}(\epsilon_n))
\end{aligned}
$$

where $G_T(x) = \sqrt{T}(F_T(x) - F(x))$. Here, the second inequality is due to integrating over a larger set. Now, arguing as in the end of the proof of Lemma A.2 and using (A.16), we conclude that

$$
\sqrt{T} \sup_{x \in \mathbb{R}_+} |F_{n,T}(x) - F_T(x)| = o_p(1) + O_p(T^{1/2-\iota}\Delta_n^{1/8-\iota}) + O_p(T^{1/2}\Delta_n\epsilon_n^{-2}).
$$

As $\epsilon_n \asymp \Delta_n^{1/8}$, our work is done. $\blacksquare$

Proof of Theorem 3.3

Notice that

$$
\sqrt{T}\big(F_{n,T}(x) - F(x)\big) = \sqrt{T}\big(F_{n,T}(x) - F_T(x)\big) + \sqrt{T}\big(F_T(x) - F(x)\big). \tag{A.17}
$$

Appealing to Lemma A.2 – A.3, it holds that $\sup_{x \in \mathbb{R}_+} \sqrt{T}\big(F_{n,T}(x) - F_T(x)\big) \xrightarrow{\mathbb{P}} 0$. Furthermore, by Theorem 4.1 in Dehay (2005) it follows that $\sqrt{T}\big(F_T(x) - F(x)\big) \xrightarrow{d} N(0, \Sigma(x))$, where

$$
\Sigma(x) = 2 \int_0^\infty \big(F_t(x, x) - F(x)^2\big) dt
$$

and $F_t(x, y) = \mathbb{P}\big(\sigma_0^2 \leq x, V_t \leq y\big)$. We next prove the proposed estimator of the asymptotic variance $\Sigma_{n,T}(x) \xrightarrow{\mathbb{P}} \Sigma(x)$, for all $x \in \mathbb{R}_+$. Thus, let $\xi \in (0, 1/3)$ be fixed and observe that:

$$
\begin{aligned}
\Sigma_{n,T}(x) &= 2 \int_0^{T^\xi} \big(F_{t,n,T}(x) - F_{n,T}(x)^2\big) dt \\
&= 2 \int_0^{T^\xi} \big(F_{t,n,T}(x) - F_{t,T}(x)\big) dt + 2 \int_0^{T^\xi} \big(F_{t,T}(x) - F_T(x)^2\big) dt + 2T^\xi\big(F_T(x)^2 - F_{n,T}(x)^2\big),
\end{aligned}
\tag{A.18}
$$

where $F_{t,T}(x) = \frac{1}{T} \int_0^{T-T^\xi} 1_{\{V_{s+t} \leq x, V_s \leq x\}} ds$. The last term in the above display is asymptotically negligible by Lemma A.2 – A.3:

$$
T^\xi\big|F_T(x)^2 - F_{n,T}(x)^2\big| = T^\xi\big|F_T(x) - F_{n,T}(x)\big|\big|F_T(x) + F_{n,T}(x)\big|
$$

$$\leq 2T^\xi \big| F_T(x) - F_{n,T}(x) \big| \xrightarrow{\mathbb{P}} 0.$$

Next,

$$F_{t,T}(x - \eta_{n,T}) \leq F_{t,n,T}(x) \leq F_{t,T}(x + \eta_{n,T}),$$

hence $\big| F_{t,n,T}(x) - F_{t,T}(x) \big| \leq F_{t,T}(x + \eta_{n,T}) - F_{t,T}(x - \eta_{n,T})$. By definition of $F_{t,T}$, it follows that for any $z, y \in \mathbb{R}_+$ with $z > y$:

$$
\begin{aligned}
F_{t,T}(z) - F_{t,T}(y) &= \frac{1}{T} \int_0^{T - T^\xi} 1_{\{y < \sigma_{s+t}^2 \leq z, \, y < \sigma_s^2 \leq z\}} ds \\
&\leq \frac{1}{2T} \int_0^{T - T^\xi} 1_{\{y < \sigma_{s+t}^2 \leq z\}} + \frac{1}{2T} \int_0^{T - T^\xi} 1_{\{y < \sigma_s^2 \leq z\}} ds \\
&= F_T(z) - F_T(y) + O\big(T^{\xi - 1}\big).
\end{aligned}
$$

Taking $z = x + \eta_{n,T}$ and $y = x - \eta_{n,T}$:

$$\int_0^{T^\xi} \big| F_{t,n,T}(x) - F_{t,T}(x) \big| dt = T^\xi \big(F_T(x + \eta_{n,T}) - F_T(x - \eta_{n,T}) \big) + O\big(T^{2\xi - 1}\big).$$

As in the proof of Lemma A.2 – A.3, we can show that

$$\int_0^{T^\xi} \big| F_{t,n,T}(x) - F_{t,T}(x) \big| dt \xrightarrow{\mathbb{P}} 0.$$

Finally, for the remaining term in (A.18), we first note that stationarity of $(V_t)_{t \geq 0}$ implies that $\mathbb{E}\big[F_{t,T}(x)\big] = \mathbb{E}\big[F_t(x, x)\big] + O\big(T^{\xi - 1}\big)$ and $\mathbb{E}\big[F_T(x)\big] = F(x)$. Hence,

$$\mathbb{E}\left[2 \int_0^{T^\xi} \big(F_{t,T}(x) - F_T(x)^2 \big) dt\right] = 2 \int_0^{T^\xi} \big(F_t(x, x) - F(x)^2 \big) dt - 2T^\xi \mathrm{var}\big(F_T(x)\big) + O\big(T^{2\xi - 1}\big).$$

Since $\alpha(t) = O(t^{-\gamma})$ and $\xi \in (0, 1/3)$, Theorem 3.3 in Dehay (2005) implies that $T^\xi \mathrm{var}\big(F_T(x)\big) \to 0$ as $T \to \infty$. Therefore,

$$\mathbb{E}\left[2 \int_0^{T^\xi} \big(F_{t,T}(x) - F_T(x)^2 \big) dt\right] \to \Sigma(x),$$

as $T \to \infty$. It thus suffices to show that $2 \int_0^{T^\xi} \big(F_{t,T}(x) - F_T(x)^2 \big) dt - \mathbb{E}\left[2 \int_0^{T^\xi} \big(F_{t,T}(x) - F_T(x)^2 \big) dt\right] \to 0$ in L^2, as $\Delta_n \to 0$ and $T \to \infty$, in order to conclude that $\Sigma_{n,T}(x) \xrightarrow{\mathbb{P}} \Sigma(x)$. To this end, we decompose the difference as follows

$$
\begin{aligned}
2 \int_0^{T^\xi} \big(F_{t,T}(x) - F_T(x)^2 \big) dt - \mathbb{E}\left[2 \int_0^{T^\xi} \big(F_{t,T}(x) - F_T(x)^2 \big) dt\right] &= 2 \int_0^{T^\xi} \big(F_{t,T}(x) - F_t(x, x) \big) dt \\
&\quad + T^\xi \big(F_T(x)^2 - F(x)^2 \big)
\end{aligned}
$$

$$+ 2T^\xi \mathrm{var}\big(F_T(x)\big) + O\big(T^{2\xi - 1}\big).$$

By Minkowski's inequality, it is enough to look at each term on the right-hand side separately. We define $\tilde{F}_{t,T}(x) = \frac{1}{T}\int_0^T 1_{\{V_{s+t} \le x, V_s \le x\}} ds$ and note that $\tilde{F}_{t,T}(x) = F_{t,T}(x) + O\big(T^{\xi-1}\big)$, $\mathbb{E}\big[\tilde{F}_{t,T}(x)\big] = F_t(x,x)$ and

$$2\int_0^{T^\xi} \big(F_{t,T}(x) - F_t(x,x)\big)dt = 2\int_0^{T^\xi} \big(\tilde{F}_{t,T}(x) - F_t(x,x)\big)dt + O\big(T^{2\xi-1}\big).$$

The Cauchy-Schwarz inequality yields that

$$\mathbb{E}\left[\left(\int_0^{T^\xi} \big(\tilde{F}_{t,T}(x) - F_t(x,x)\big)dt\right)^2\right] \le T^\xi \int_0^{T^\xi} \mathrm{var}\big(\tilde{F}_{t,T}(x)\big)dt$$

$$= O\big(T^{2\xi-1}\big),$$

where the last inequality follows from the calculations on p. 941 in Dehay (2005). Therefore,

$$2\int_0^{T^\xi} \big(F_{t,T}(x) - F_t(x,x)\big)dt \to 0$$

in L^2. As both F_T and F are bounded by 1, the bounded convergence theorem and Theorem 3.3 in Dehay (2005) implies that

$$T^\xi \big(F_T(x)^2 - F(x)^2\big) \to 0 \tag{A.19}$$

in L^2, as $T \to \infty$. Thus,

$$2\int_0^{T^\xi} \big(F_{t,T}(x) - F_T(x)^2\big)dt - \mathbb{E}\left[2\int_0^{T^\xi} \big(F_{t,T}(x) - F_T(x)^2\big)dt\right] \to 0$$

in L^2, given the rate condition from Theorem 3.3. This implies $\Sigma_{n,T}(x) \xrightarrow{\mathbb{P}} \Sigma(x)$. ∎

We now state a multivariate version of Theorem 3.3.

Lemma A.4 *Suppose that the conditions of Theorem 3.3 are fulfilled. Let $p \in \mathbb{N}$ be given. Then, for any $x_1, \ldots, x_p \in \mathbb{R}_+$,*

$$\big(G_{n,T}(x_1), \ldots, G_{n,T}(x_p)\big) \xrightarrow{d} N\big(0, \Sigma(x_1, \ldots, x_p)\big),$$

where $\Sigma(x_1, \ldots, x_p)$ is the $p \times p$ covariance matrix with elements of the form

$$\Sigma(x_1, \ldots, x_p)_{ij} = \int_0^\infty \big(F_t(x_i, x_j) + F_t(x_i, x_j) - 2F(x_i)F(x_j)\big)dt,$$

for $i, j = 1, \ldots, p$. Furthermore, let $\xi \in (0, 1/3)$ be given. Then,

$$\Sigma_{n,T}(x_i, x_j) = \int_0^{T^\xi} \big(F_{t,n,T}(x_i, x_j) + F_{t,n,T}(x_j, x_i) - 2F_{n,T}(x_i)F_{n,T}(x_j)\big)dt \xrightarrow{\mathbb{P}} \Sigma(x_1, \ldots, x_p)_{ij}.$$

Proof. The first part follows by a Crámer-Rao device. Also, proving consistency of the covariance estimator goes along the lines of the univariate case. We therefore skip it. ∎

Proof of Theorem 3.5

In view of Lemma A.2-A.3, we obtain that $\sup_{x \in \mathbb{R}_+} \sqrt{T}\big(F_{n,T}(x) - F_T(x)\big) \overset{\mathbb{P}}{\to} 0$. On the other hand, Theorem 5.2 in Dehay (2005) implies a functional central limit theorem for the process $G_T = \big\{\sqrt{T}\big(F_T(x) - F(x)\big), x \in \mathbb{R}_+\big\}$. ■

Proof of Corollary 3.6

This follows from Theorem 3.5 by noting that the sup and weighted L^2 norm are both continuous mappings from $\mathbb{D}(\mathbb{R}_+) \to \mathbb{R}$. ■

Proof of Proposition 3.7

From Corollary 3.6 the limiting distribution of rL^2 under $\mathcal{H}_0$ is $||G_F||_w^2$, where G_F is a Gaussian process with covariance operator Σ. The distribution can therefore be written as $||G_F||_w^2 = \sum_{i=1}^{\infty} \lambda_i \Phi_i^2$, where Φ_i are independent standard normal random variables, and λ_i are the eigenvalues of Σ (see, e.g., Kuo, 1975). Moreover, from Bosq (2000, chapter 1) we deduce that $\tilde{\Sigma} = \sum_{i=1}^{\infty} \lambda_i < \infty$, so that $\mathbb{E}\big[||G_F||_w^2\big] = \tilde{\Sigma}$. Combining this result with the fact that $T(F)$ is a quadratic form of mean zero Gaussian random variables, it follows that for all $\alpha \in (0, 0.215)$:

$$\mathbb{P}\big(T(F) > \alpha\big) = 1 - \mathbb{P}\big(T(F) \le \alpha\big) \le 1 - \mathbb{P}\big(\chi^2 \le \alpha\big) = \mathbb{P}\big(\chi^2 > \alpha\big),$$

where the inequality is due to Székely and Bakirov (2003). This concludes the proof of the first statement, while the second part follows upon observing that $||G_F||_w^2 \overset{\mathbb{P}}{\to} \infty$ under $\mathcal{H}_a$. ■

Proof of Corollary 3.8

The proof follows from Theorem 19.23 in van der Vaart (1998). ■

B   Simulation of critical values for goodness-of-fit testing

To describe how critical values for our goodness-of-fit tests are generated, we again fix the sampling frequency at Δ_n and the time horizon at $[0, T]$.

The stochastic volatility model is assumed to be indexed by a parameter vector υ. We discriminate between the case where υ is known (I), and the more realistic setup where υ is unknown, but a preliminary estimate $\hat{\upsilon}$ is available along with an estimate of the noise variance $\hat{\omega}$ (II). As a

biproduct, we emphasize that the latter approach also accounts for the estimation error induced by the infill step, which is negligible in the limit but can be sizable in small samples.

We proceed as follows:

I. Without parameter estimation

1. Simulate a volatility path of length T based on υ.

2. Sample the path at frequency Δ_n.

3. Construct the EDF, F_T.

4. Calculate the t-statistic with F based on υ.

5. Repeat step $1 - 4$ B times.

II. With parameter estimation

1. Simulate a volatility path of length T based on $\hat{\upsilon}$. Call it $(\sigma_t)_{t \in [0,T]}$.

2. Construct noisy log-returns at sampling frequency Δ_n according to: $\Delta_i^n Z = \sigma_{(i-1)\Delta_n} \sqrt{\Delta_n} \Phi_1 + \hat{\omega} \Phi_2$, where Φ_1 and Φ_2 are independent standard normal random variables.

3. Construct the REDF, $F_{n,T}$.

4. Retrieve a new estimate of the parameter vector, denoted by $\tilde{\upsilon}$, from the artificial data.

5. Calculate the t-statistic with F based on $\tilde{\upsilon}$.

6. Repeat step $1 - 5$ B times.

In both I. and II., the critical values are subsequently found as the α-quantile of the distribution of t-statistics based on the above B repetitions of the simulation experiment.

References

Aït-Sahalia, Y., 1996, "Testing continuous-time models of the spot interest rate," *Review of Financial Studies*, 9(2), 385–426.

Aït-Sahalia, Y., and J. Jacod, 2014, *High-frequency Financial Econometrics*. Princeton University Press, 1st edn.

Aït-Sahalia, Y., J. Jacod, and J. Li, 2012, "Testing for jumps in noisy high frequency data," *Journal of Econometrics*, 168(2), 207–222.

Aït-Sahalia, Y., and R. Kimmel, 2007, "Maximum likelihood estimation of stochastic volatility models," *Journal of Financial Economics*, 83(2), 413–452.

Aït-Sahalia, Y., and D. Xiu, 2015, "Increased correlation among asset classes: Are volatility or jumps to blame, or both?," *Journal of Econometrics*, 194(2), 205–219.

Andersen, T. G., L. Benzoni, and J. Lund, 2002, "An empirical investigation of continuous-time equity return models," *Journal of Finance*, 57(4), 1239–1284.

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.

Andersen, T. G., M. Thyrsgaard, and V. Todorov, 2019, "Time-varying periodicity in intraday volatility," *Journal of the American Statistical Association*, (Forthcoming).

Baeumer, B., and M. M. Meerschaert, 2010, "Tempered stable Lévy motion and transient super-diffusion," *Journal of Computational and Applied Mathematics*, 233(10), 2438–2448.

Bandi, F. M., and J. R. Russell, 2006, "Separating microstructure noise from volatility," *Journal of Financial Economics*, 79(3), 655–692.

Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, M. Podolskij, and N. Shephard, 2006, "A central limit theorem for realized power and bipower variations of continuous semimartingales," in *From Stochastic Calculus to Mathematical Finance: The Shiryaev Festschrift*, ed. by Y. Kabanov, R. Lipster, and J. Stoyanov. Springer-Verlag, Heidelberg, pp. 33–68.

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.

Bibby, B. M., I. M. Skovgaard, and M. Sørensen, 2005, "Diffusion-type models with given marginal distribution and autocorrelation function," *Bernoulli*, 11(2), 191–220.

Bibinger, M., and L. Winkelmann, 2018, "Common price and volatility jumps in noisy high-frequency data," *Electronic Journal of Statistics*, 12(1), 2018–2073.

Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner, 1998, *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag, 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.

Bosq, D., 2000, *Linear Processes in Function Spaces: Theory and Applications*. Springer-Verlag, 1st edn.

Bull, A. D., 2017, "Semimartingale detection and goodness-of-fit tests," *Annals of Statistics*, 45(3), 1254–1283.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen, 2003, "Alternative models for stock price dynamics," *Journal of Econometrics*, 116(1-2), 225–257.

Christensen, K., R. C. A. Oomen, and M. Podolskij, 2014, "Fact or friction: Jumps at ultra high frequency," *Journal of Financial Economics*, 114(3), 576–599.

Christoffersen, P. F., K. Jacobs, and K. Mimouni, 2010, "Volatility dynamics for the S&P500: Evidence from realized volatility, daily returns, and option prices," *Review of Financial Studies*, 23(8), 3141–3189.

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.

Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, "A theory of the term structure of interest rates," *Econometrica*, 53(2), 385–407.

David, F. N., and N. L. Johnson, 1948, "The probability integral transformation when parameters are estimated from the sample," *Biometrika*, 35(1-2), 182–190.

Dehay, D., 2005, "On invariant distribution function estimation for continuous-time stationary processes," *Bernoulli*, 11(5), 933–948.

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

Dette, H., and M. Podolskij, 2008, "Testing the parametric form of the volatility in continuous time diffusion models–A stochastic process approach," *Journal of Econometrics*, 143(1), 56–73.

Dette, H., M. Podolskij, and M. Vetter, 2006, "Estimation of integrated volatility in continuous-time financial models with applications to goodness-of-fit testing," *Scandinavian Journal of Statistics*, 33(2), 259–278.

Diebold, F. X., and G. H. Strasser, 2013, "On the correlation structure of microstructure noise: A financial economic approach," *Review of Economic Studies*, 80(4), 1304–1337.

Foster, D. P., and D. B. Nelson, 1996, "Continuous record asymptotics for rolling sample variance estimators," *Econometrica*, 64(1), 139–174.

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

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

Gatheral, J., and R. C. A. Oomen, 2010, "Zero-intelligence realized variance estimation," *Finance and Stochastics*, 14(2), 249–283.

Geman, D., and J. Horowitz, 1980, "Occupation densities," *Annals of Probability*, 8(1), 1–67.

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

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.

Hull, J., and A. White, 1987, "The pricing of options on assets with stochastic volatilities," *Journal of Finance*, 42(2), 281–300.

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.

Jacod, J., and P. E. Protter, 2012, *Discretization of Processes.* Springer-Verlag, Berlin, 2nd edn.

Jacod, J., and M. Rosenbaum, 2013, "Quarticity and other functionals of volatility: Efficient estimation," *Annals of Statistics*, 41(3), 1462–1484.

Jongbloed, G., F. H. van der Meulen, and A. W. van der Vaart, 2005, "Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes," *Bernoulli*, 11(5), 759–791.

Kuo, H.-H., 1975, *Gaussian Measures in Banach Spaces.* Springer-Verlag, 1st edn.

Li, J., V. Todorov, and G. Tauchen, 2013, "Volatility occupation times," *Annals of Statistics*, 41(4), 1865–1891.

——— , 2016, "Estimating the volatility occupation time via regularized Laplace inversion," *Econometric Theory*, 32(5), 1253–1288.

——— , 2017, "Robust jump regressions," *Journal of the American Statistical Association*, 112(517), 332–341.

Lilliefors, H. W., 1967, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American Statistical Association*, 62(318), 399–402.

——— , 1969, "On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown," *Journal of the American Statistical Association*, 64(325), 387–389.

Lin, L.-C., S. Lee, and M. Guo, 2013, "Goodness-of-fit test for stochastic volatility models," *Journal of Multivariate Analysis*, 116(1), 473–498.

——— , 2016, "Goodness-of-fit test for the SVM based on noisy observations," *Statistica Sinica*, 26(3), 1305–1329.

Magdziarz, M., and A. Weron, 2011, "Ergodic properties of anomalous diffusion processes," *Annals of Physics*, 326(9), 2431–2443.

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

Merton, R. C., 1980, "On estimating the expected return on the market: An exploratory investigation," *Journal of Financial Economics*, 8(4), 323–361.

Newey, W. K., and K. D. West, 1994, "Automatic lag selection in covariance matrix estimation," *Review of Economic Studies*, 61(4), 631–653.

Oomen, R. C. A., 2006, "Comment on 2005 JBES invited address "Realized variance and market microstructure noise" by Peter R. Hansen and Asger Lunde," *Journal of Business and Economic Statistics*, 24(2), 195–202.

Podolskij, M., and M. Vetter, 2009a, "Bipower-type estimation in a noisy diffusion setting," *Stochastic Processes and their Applications*, 119(9), 2803–2831.

——— , 2009b, "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli*, 15(3), 634–658.

Resnick, S. I., 1998, *A probability path*. Birkhäuser, 1 edn.

Rosiński, J., 2007, "Tempering stable processes," *Stochastic Processes and their Applications*, 117(6), 677–707.

Székely, G. J., and N. K. Bakirov, 2003, "Extremal probabilities for Gaussian quadratic forms," *Probability Theory and Related Fields*, 126(2), 184–202.

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, 2012, "The realized Laplace transform of volatility," *Econometrica*, 80(3), 1105–1127.

Todorov, V., G. Tauchen, and I. Grynkiv, 2011, "Realized Laplace transforms for estimation of jump diffusive volatility models," *Journal of Econometrics*, 164(2), 367–381.

——— , 2014, "Volatility activity: Specification and estimation," *Journal of Econometrics*, 178(1), 180–193.

van der Vaart, A. W., 1998, *Asymptotic Statistics*. Cambridge University Press, 1 edn.

van der Vaart, A. W., and J. A. Wellner, 2007, "Empirical processes indexed by estimated functions," *Lecture Notes-Monograph Series*, 55, 234–252.

Veretennikov, Y. A., 1988, "Bounds for the mixing rate in the theory of stochastic equations," *Theory of Probability and Its Applications*, 32(2), 273–281.

Vetter, M., and H. Dette, 2012, "Model checks for the volatility under microstructure noise," *Bernoulli*, 18(4), 1421–1447.

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.

Zu, Y., 2015, "Nonparametric specification tests for stochastic volatility models based on volatility density," *Journal of Econometrics*, 1(323-344), 187.

Zu, Y., and H. P. Boswijk, 2014, "Estimating spot volatility with high-frequency financial data," *Journal of Econometrics*, 181(2), 117–135.