NONPARAMETRIC STOCHASTIC VOLATILITY

Author(s): Federico M. Bandi and Roberto Renò
Source: *Econometric Theory*, December 2018, Vol. 34, No. 6 (December 2018), pp. 1207-1255
Published by: Cambridge University Press
Stable URL: https://www.jstor.org/stable/10.2307/26613606

Econometric Theory, **34**, 2018, 1207–1255.
doi:10.1017/S0266466617000457

NONPARAMETRIC STOCHASTIC VOLATILITY

FEDERICO M. BANDI
Johns Hopkins University and Edhec-Risk Institute

ROBERTO RENÒ
Università di Verona

We provide nonparametric methods for stochastic volatility modeling. Our methods allow for the joint evaluation of return and volatility dynamics with nonlinear drift and diffusion functions, nonlinear leverage effects, and jumps in returns and volatility with possibly state-dependent jump intensities, among other features. In the first stage, we identify spot volatility by virtue of jump-robust nonparametric estimates. Using observed prices and estimated spot volatilities, the second stage extracts the functions and parameters driving price and volatility dynamics from nonparametric estimates of the bivariate process' *infinitesimal moments*. For these infinitesimal moment estimates, we report an asymptotic theory relying on *joint* in-fill and long-span arguments which yields consistency and weak convergence under mild assumptions.

1. INTRODUCTION

Understanding volatility dynamics is important for effective portfolio choice, derivative pricing, and risk management, among other issues. A successful strand of the literature on volatility estimation has focused on stochastic volatility modelling either in continuous time or in discrete time (see, e.g., Bauwens, Hafner, and Laurent, 2012). This literature provides alternative methods to filter volatility—an inherently unobservable state variable—by using return data sampled at relatively low, generally daily, frequencies. An equally successful, but alternative, recent strand of the literature on volatility estimation has recognized the identification potential of return data sampled at intra-daily frequencies to treat daily volatility (estimated by "aggregating" squared intra-daily returns in various

We thank the participants at the SoFiE Inaugural conference (New York, June 4–6, 2008), the Festschrift in honor of Peter C.B. Phillips (Singapore, July 14–15, 2008), the Far Eastern and South Asian Meetings of the Econometric Society (Singapore, July 16–18, 2008), the Hitotsubashi University's International Conference on High Frequency Data Analysis in Financial Markets (Tokyo, October 25–26, 2008), the LSE Conference on Recent Advances in High-Frequency Financial Econometrics (London, November 15, 2008), and the North American Winter Meetings of the Econometric Society (San Francisco, January 3–5, 2009) for discussions. We are grateful to three anonymous referees, the Editors Oliver Linton and Peter C.B. Phillips, Valentina Corradi, Fulvio Corsi, Cecilia Mancini, and Eric Renault for useful comments and suggestions. The usual disclaimers apply. Address correspondence to Roberto Renò, Università di Verona, Via Cantarane 24, 37129 Verona, Italy; e-mail: roberto.reno@univr.it.

ways) as an "observable" (up to measurement error) quantity, without the need for filtering on the basis of low-frequency return data (Aït-Sahalia and Jacod, 2014, provide a thorough coverage). This second body of work has seldom investigated the implications of high-frequency volatility estimation for stochastic volatility modelling. The *parametric* approaches of, e.g., Bollerslev and Zhou (2002), Corradi and Distaso (2006), and Todorov (2009) are exceptions and promising contributions in this area.

We further bridge the gap between arguably the two main strands of the current volatility literature by providing functional inferential methods. Specifically, we study *nonparametric* stochastic volatility modeling in continuous time using high-frequency asset price data. High-frequency price data is employed for the purpose of spot volatility estimation prior to nonparametric modeling of the dynamics.

We consider the price/variance system:

$$\mathrm{d}\log p_t = \mu(\sigma_t^2)\mathrm{d}t + \sigma_t \mathrm{d}W_t^r + \mathrm{d}J_t^r, \tag{1}$$

$$\mathrm{d}f(\sigma_t^2) = m_{f(\sigma^2)}(\sigma_t^2)\mathrm{d}t + \Lambda_{f(\sigma^2)}(\sigma_t^2)\mathrm{d}W_t^\sigma + \mathrm{d}J_t^\sigma, \tag{2}$$

where $\{\mathrm{d}W_t^r, \mathrm{d}W_t^\sigma\} = \{\rho(\sigma_t^2)dW_t^1 + \sqrt{1-\rho^2(\sigma_t^2)}dW_t^2, dW_t^1\}$ with $-1 \le \rho(.) \le 1$, $\{W_t^1, W_t^2\}$ are independent, standard Brownian motions, $\{J_t^r, J_t^\sigma\}$ are Poisson jump processes independent of each other and independent of $\{W_t^1, W_t^2\}$ with intensities $\lambda^r(.)$ and $\lambda^{f(\sigma^2)}(.)$ respectively, $f(.)$ is a monotonically non-decreasing transformation of variance, and $\mu(.)$, $m_{f(\sigma^2)}(.)$, $\Lambda_{f(\sigma^2)}(.)$, and $\rho(.)$ are generic functions satisfying suitable smoothness conditions described in what follows. Equations (1)–(2) nest and generalize, as we discuss below, a large number of popular stochastic volatility models in the extant literature. For instance, given assumptions on the distributions of the jump sizes and suitable choices of the function $f(.)$, the price/variance system can be viewed as a nonlinear generalization of the stochastic volatility model with exponential jumps in variance in, e.g., Duffie, Pan, and Singleton (2000) and Eraker, Johannes, and Polson (2003), as well as of the logarithmic volatility model with normal jumps in variance in, e.g., Jacquier, Polson, and Rossi (1994) and Andersen, Benzoni, and Lund (2002).

A similar system, but without price/variance jumps, has been considered by Renò (2006, 2008) and in recent, concurrent, work by Kanaya and Kristensen (2016), Comte, Genon-Catalot, and Rozenholc (2010) and, in discrete time, by Comte, Lacour, and Rozenholc (2010). The presence of price/variance discontinuities in this paper results in a different approach to spot volatility filtering and nonparametric identification of the dynamics.

The unobservable nature of variance leads to the two substantive aspects of the procedure proposed in this work. In the first stage, we filter *spot variance* by localizing (in time) an appropriately-chosen high-frequency estimate of *integrated variance* (i.e., $\int_t^{t+\phi} \sigma_s^2 \mathrm{d}s$, with $\phi > 0$) robust to the presence of price jumps. In the second stage, we use the resulting spot variance estimates to identify the parameters and functions driving return and variance dynamics (for instance, focusing solely on the variance equation, $m_{f(\sigma^2)}(.)$, $\Lambda_{f(\sigma^2)}(.)$, $\lambda^{f(\sigma^2)}(.)$ and, given assumptions on the jump size distribution, the moments of the volatility jumps). The second stage requires controlling the estimation error introduced by the first-step spot variance estimates. We provide conditions to do so asymptotically.

Unlike existing work on parametric stochastic volatility modeling, we avoid imposing (possibly affine) parametric structures. Specifically, after filtering spot variance, we identify the relevant functions/parameters, through estimates of the system's *infinitesimal moments*, using nonparametric kernel methods for jump-diffusion processes as originally studied by Bandi and Nguyen (2003) and Johannes (2004) in a more traditional framework, that of scalar specifications with observables.

We extend the observable, scalar approach in Bandi and Nguyen (2003) and Johannes (2004) along two main dimensions: we allow for a bivariate system with one unobservable state variable (i.e., variance) and employ a method of proof which relaxes the bandwidth conditions in Bandi and Nguyen (2003). Differently from Bandi and Nguyen (2003), we do not rely on the notion of local time to derive limiting results but use, instead, occupation densities (Bandi and Moloche, 2017). The latter are defined more generally. Contrary to local time, they can be defined for systems with multiple state variables. In this sense, they are suitable for multi-state extensions of the (single state) theory presented in this work.

Identification is (infinitesimal) moment-based. In order to present ideas in the context of a classical estimation problem, we use Nadaraya-Watson kernel estimates. However, as we illustrate below, extensions to alternative functional estimation methods may be readily derived. To this end, we discuss the local linear and the local polynomial case explicitly. Because of the persistent behavior of daily variance, methods which do not rely on the information contained in a potentially inaccurately estimated (in finite samples) stationary density, and hinge on local moments instead, are particularly suitable for our problem.

We first discuss the properties of the infinitesimal moment estimates under stationarity (Section 2 through 4). In the Appendix, we turn to the general recurrent case. Recurrence is known to be a milder assumption than stationarity and mixing (Meyn and Tweedie, 1993). We refer the reader to Bandi and Phillips (2010) for a review of identification methods for recurrent continuous-time processes.

In what follows, the symbols $\Rightarrow$, $\xrightarrow{a.s.}$, $\xrightarrow{p}$ and $\overset{\mathrm{d}}{=}$ stand for weak convergence, convergence with probability one, convergence in probability, and distributional equivalence. The notation $A \sim B$ and $A \overset{p}{\sim} B$ signifies that A is $O(B)$ and A is $O_p(B)$, respectively.

Supplementary material to this article is provided in "Online Supplement to Nonparametric Stochastic Volatility," which is available at Cambridge Journals Online (journals.cambridge.org/ect).

2. INFINITESIMAL MOMENT ESTIMATORS: ASSUMPTIONS AND ASYMPTOTICS

We begin with the assumptions on the bivariate process.

Assumption 1. (a) Define the system in equations (1)–(2) on a filtered probability space $(\Omega, (\mathcal{F}_t)_{0\leq t<\infty}, \mathcal{P})$ satisfying the usual conditions. Let $\mu(.)$, $m_{f(\sigma^2)}(.)$, $\Lambda_{f(\sigma^2)}(.)$, $\rho(.)$, $\lambda^r(.)$ and $\lambda^{f(\sigma^2)}(.)$ be uniformly bounded and, at least, twice continuously-differentiable with bounded derivatives of any order. The moments of the jump sizes are allowed to depend on the variance state. In this case, they are also assumed to be uniformly bounded and, at least, twice continuously-differentiable with bounded derivatives of any order. The variance transformation $f(.)$ is monotonically non-decreasing and continuously-differentiable with a bounded first derivative.

(b) We assume that the spot variance process is stationary and with a bounded state space. Let $\lim_{T\to\infty}\int_0^T \left(1-\frac{u}{T}\right)||g_u(a,b)||_r du < \infty$ for some $r \in [1,\infty]$ and $\forall a,b$, with $g_u(a,b) = g_{|t-s|}(a,b) = p_{|t-s|}(a,b) - p(a)p(b)$, where $p_{|t-s|}(.,.)$ is the joint unconditional density of σ_t^2 and σ_s^2 and $p(.)$ is the time-invariant marginal density of σ_t^2.[1] Let $Th_{n,T}^{2/r} \to \infty$, where $h_{n,T}$ is the smoothing parameter used in kernel estimation in this paper and T is the sampling horizon.

The requirement in Assumption 1(a) of uniformly bounded driving functions is a classical, often implicitly-made, requirement in the continuous-time nonparametric literature when letting the sampling horizon T grow without bound along with a vanishing distance between discretely-sampled observations. The requirement in Assumption 1(b) of volatility recurrence in a bounded set is less traditional. The latter implies stationarity of the variance process. In addition, because variance is the system's only state variable, it implies boundedness of all of the system's functions, if the functions are uniformly continuous. We remark that both conditions are not *empirically* restrictive since the data, and its transformations, are clearly bounded. In what follows, we show that these assumptions are not *theoretically* restrictive either. Both assumptions are relaxed in the paper's proofs which are, in the interest of theoretical generality, written for a recurrent (rather than stationary) variance process and unbounded driving functions.

Assumption 1(b) also controls dependence in the variance sample path. It corresponds to assumption $A'(r)$ in Bosq (1998, page 105). If $r = \infty$, the condition on the bandwidth (i.e., $Th_{n,T}^{2/r} \to \infty$) is always satisfied since $T \to \infty$. This is the case if, for example, Castellana and Leadbetter's condition (Castellana and Leadbetter, 1986) is met: $|g_u(a,b)| \leq \Psi(u) \in L_1((0,\infty))$ for all a and b. If $r = \infty$, in fact,

$$\lim_{T\to\infty}\int_0^T \left(1-\frac{u}{T}\right)||g_u(a,b)||_\infty \mathrm{d}u = \int_0^\infty ||g_u(a,b)||_\infty \mathrm{d}u < \infty.$$

The assumption relates a bandwidth requirement (i.e., $Th_{n,T}^{2/r} \to \infty$) to the "regularity" of the process sample path (which is captured by the unknown parameter r) and is known to be valid for a large class of ergodic processes (Leblanc, 1995; Kutoyants, 1997; Veretennikov, 1999, among others). The more locally informative the path ($r \to \infty$), the less stringent the assumption (since, when $r \to \infty$, $Th_{n,T}^{2/r} \to T$). By capturing an array of local behaviors of the variance process, Assumption 1(b) will prove helpful for the estimation of the variance occupation density.

We now turn to the sampling scheme. The graph in Figure 1 provides a visual representation.

Assumption 2 (The sampling scheme)**.** We assume availability of $(k+1)n$ discretely-sampled price observations over the time span $[0, T]$. In particular, we observe prices in the form of n buckets of $k+1$ high-frequency equispaced observations in each interval $[i\Delta_{n,T} - \phi, i\Delta_{n,T}]$ with $\phi < \Delta_{n,T}$, $\Delta_{n,T} = T/n$ and $i = 1, \dots, n$. In the limit, $T \to \infty$, $n \to \infty$, $k \to \infty$, $\Delta_{n,T} \to 0$ and $\phi \to 0$, *jointly*.

The identification strategy works as follows. For any sub-interval $[i\Delta_{n,T} - \phi, i\Delta_{n,T}]$ of length ϕ, the $k+1$ intra-period price observations are employed to estimate spot variance ($\widehat{\sigma}^2_{i\Delta_{n,T}}$ with $i = 1, \dots, n$). The n resulting spot variance estimates $\left\{\widehat{\sigma}^2_{\Delta_{n,T}}, \widehat{\sigma}^2_{2\Delta_{n,T}}, \dots, \widehat{\sigma}^2_{n\Delta_{n,T}}\right\}$ along with the observable logarithmic price observations $\left\{\log p_{\Delta_{n,T}}, \log p_{2\Delta_{n,T}}, \dots, \log p_{n\Delta_{n,T}}\right\}$ are then used to identify the dynamics of the bi-variate system in equations (1)–(2). One may, for instance, set $\Delta_{n,T} = 1$ day and $\phi = 1$ hour. Asymptotic conditions on $T, n, k, \Delta_{n,T}$ and ϕ for consistency and asymptotic normality of the relevant estimates will be listed below.

The functions driving the dynamics of the system have infinitesimal conditional moment representations which can be exploited for the purpose of nonparametric identification. We first define the infinitesimal moment estimates of the system, for which we discuss an asymptotic theory in this section. We then identify the relevant functions/parameters from the estimated infinitesimal moments (in Section 3). In what follows, x is used to denote a generic interior point of the domain of σ_t^2.

We begin with the variance process in equation (2). We will turn to the price process and its connection with the variance process in Section 4.2 below.

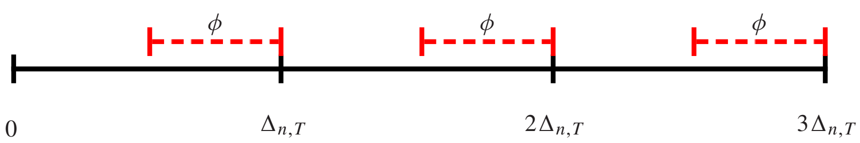

FIGURE 1. (Color online) We assume $k+1$ intra-period price observations (on dashed lines of total length $\phi < \Delta_{n,T}$) for each price observation sampled at $i\Delta_{n,T}$ with $i = 1, \dots, n$. In the graph, we consider 3 periods ($n = 3$). Spot variance is estimated for every time $i\Delta_{n,T}$ using $k+1$ intra-period observations over ϕ.

We identify the R-th infinitesimal moment of the variance process, i.e.,

$$\theta_{f(\sigma^2),R}(x) = \lim_{\Delta\to 0} \frac{1}{\Delta} \mathrm{E}\left[\left(f(\sigma^2_{t+\Delta}) - f(\sigma^2_t)\right)^R |\sigma^2_t = x\right] \qquad R = 1,\ldots, \tag{3}$$

by virtue of

$$\widehat{\theta}_{f(\sigma^2),R}(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)\left(f(\widehat{\sigma}^2_{(i+1)\Delta_{n,T}}) - f(\widehat{\sigma}^2_{i\Delta_{n,T}})\right)^R}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)} \qquad R = 1,\ldots,$$

where $\mathbf{K}(\cdot)$ is a kernel function, whose properties are listed below, and the unobservable spot variances $\sigma^2_{i\Delta_{n,T}}$ are replaced by suitable estimates $\widehat{\sigma}^2_{i\Delta_{n,T}}$ for all $i = 1,\ldots,n$.

In order to estimate $\sigma^2_{i\Delta_{n,T}}$, we use a localized (in time) threshold realized variance estimator (Mancini, 2009), namely

$$\widehat{\sigma}^2_{i\Delta_{n,T}} = \frac{1}{\phi}\sum_{j=1}^{k} r^2_{i,j}\, \mathbf{1}_{\{r^2_{i,j}\le\vartheta\}}, \tag{4}$$

where $r_{i,j} = \log p_{i\Delta_{n,T}-\phi+j\phi/k} - \log p_{i\Delta_{n,T}-\phi+(j-1)\phi/k}$ is the return between two consecutive observations in the interval ϕ associated with the generic time $i\Delta_{n,T}$ and ϑ is a vanishing threshold. By eliminating, under conditions, discontinuities larger than the threshold ϑ, this estimator is robust to jumps in returns and, therefore, identifies spot variance (as $\phi \to 0$ with $k \to \infty$) in the presence of price jumps. Adapting results in Mancini (2009), which are for *integrated* variance, to the *spot* variance case, it can be shown that

$$k^{1/2}\left\{\widehat{\sigma}^2_t - \sigma^2_t\right\} \underset{k\to\infty,\phi\to 0,\vartheta\to 0}{\Rightarrow} \mathrm{MN}\left(0, 2\sigma^4_t\right) \qquad \forall t > 0, \tag{5}$$

where MN stands for mixed normal, provided $\frac{1}{\vartheta}\left(\frac{\phi}{k}\log\left(\frac{k}{\phi}\right)\right) \to 0$ and $k^{1/2}\phi^{1/2} \to 0$, as $\phi, \vartheta \to 0$ with $k \to \infty$. The former condition guarantees that the threshold goes to zero more slowly than the squared modulus of continuity of Brownian motion, thereby only annihilating the genuine jumps, while the latter condition eliminates the estimator's asymptotic bias. A similar approach to spot variance estimation, but in the absence of jumps, has been followed by Fan and Wang (2008), Mykland and Zhang (2008), and Kristensen (2010). Bandi and Renò (2012) discuss an analogous result for a spot variance estimator based on threshold bipower variation (Corsi et al., 2010).

Importantly, equation (5) leads to a probability bound $O_p\left(\frac{1}{k^{1/2}}\right)$ on the estimation error $\left\{\widehat{\sigma}^2_t - \sigma^2_t\right\}$ which is pointwise and, therefore, insufficient for nonparametric inference on the infinitesimal moments of the system. Lemma A.17 in the Appendix provides a uniform result which is, also, of separate interest.

The kernel function $\mathbf{K}(.)$ satisfies the following property:

Assumption 3. The function $\mathbf{K}(.)$ is a bounded, continuous, and symmetric density defined on a compact set S. The kernel's derivatives are bounded.[2] In what follows, $\mathbf{K}_1 = \int_{\mathrm{S}} s^2 \mathbf{K}(s)\mathrm{d}s$ and $\mathbf{K}_2 = \int_{\mathrm{S}} \mathbf{K}^2(s)\mathrm{d}s$.

Finally, we turn to conditions on $T, n, k, \Delta_{n,T}$, the length ϕ, the bandwidth $h_{n,T}$ and the threshold ϑ for consistency and weak convergence of the infinitesimal moment estimates when spot variance is estimated. In the statement of each theorem, we will specify which of the following conditions is requested.

Assumption 4. Define $\widehat{L}_{\sigma^2}(T,x) = \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)$, an infeasible nonparametric estimator of the spot variance's occupation density.

Assume k, n, $T \to \infty$, $\Delta_{n,T} \to 0$, $\phi \to 0$, $h_{n,T} \to 0$, and $\vartheta \to 0$, *jointly*. The symbol C defines a suitable constant. Let

$$\frac{1}{\vartheta}\left(\frac{\phi}{k}\log\left(\frac{k}{\phi}\right)\right) \to 0, \tag{4.0}$$

$$h_{n,T}\widehat{L}_{\sigma^2}(T,x) \xrightarrow{p} \infty, \tag{4.1}$$

$$\frac{\Delta_{n,T}}{h^2_{n,T}} \to 0, \tag{4.2}$$

$$h^5_{n,T}\widehat{L}_{\sigma^2}(T,x) \xrightarrow{p} C, \tag{4.3}$$

$$\frac{\Delta_{n,T}\sqrt{\widehat{L}_{\sigma^2}(T,x)}}{h^{3/2}_{n,T}} \xrightarrow{p} 0, \tag{4.4}$$

$$g(n,T,k,\phi) = \frac{(\log n \log nk)^{1/2}}{h_{n,T}k^{1/2}} + \frac{\phi^{1/2}(\log n)^{1/2}}{h_{n,T}} + \frac{n\phi}{h_{n,T}} \to 0, \tag{4.5}$$

$$g(n,T,k,\phi)\frac{h_{n,T}}{\Delta_{n,T}} \to 0, \tag{4.6}$$

$$g(n,T,k,\phi)\frac{1}{\Delta^{1/2}_{n,T}} \to 0, \tag{4.7}$$

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\, g(n,T,k,\phi)\frac{h_{n,T}}{\Delta_{n,T}} \xrightarrow{p} 0, \tag{4.8}$$

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\, g(n,T,k,\phi)\frac{1}{\Delta^{1/2}_{n,T}} \xrightarrow{p} 0. \tag{4.9}$$

Assumption 4.0 is needed to guarantee that the threshold spot variance estimates in equation (4) eliminate discontinuities whose order is larger than that of the modulus of continuity of Brownian motion. Assumptions 4.1 and 4.2 are necessary for consistency (in probability) of the kernel estimates of the infinitesimal variance moments under observability of the spot variance process. Assumptions 4.6 and 4.7, along with Assumptions 4.1 and 4.2, yield consistency of the infinitesimal variance moments when spot variance is not observable but is, instead, estimated (Assumption 4.6 applies to the case $R = 1$, Assumption 4.7 applies to the case $R > 1$). Assumptions 4.1, 4.3 and 4.4, guarantee weak convergence—at the optimal MSE rate $\widehat{L}_{\sigma^2}(T, x)^{-4/5}$—of the kernel estimates, again under observability of the spot variance process. Assumptions 4.8 ($R = 1$) and 4.9 ($R > 1$), along with Assumptions 4.1, 4.3, and 4.4, yield weak convergence when spot variance is estimated. It will be shown that Assumption 4.9 yields weak convergence of all infinitesimal *price* moment estimates ($R \geq 1$) in equation (32) below. It is weaker than Assumption 4.8 (due to Assumption 4.2) since prices are observable and estimated spot variance enters the infinitesimal price moment estimates only through the kernel.

Remark 1. Invoking Lemma A.5 in the Appendix, call $\upsilon(T)$ the probability divergence rate of the spot variance occupation density ($\upsilon(T) = T$ in the stationary case discussed in the main text). Consider the case $R > 1$ (i.e., estimation of all infinitesimal moments excluding the first). To make the relevant orders more explicit, neglecting for convenience the logarithmic terms, it is worth rewriting Assumptions 4.1, 4.2, 4.3, and 4.7 as follows: $h_{n,T} \sim \upsilon(T)^{-\beta}$ with $\frac{1}{5} \leq \beta < 1$, $n \sim T^{1+2\beta+\epsilon}$ with $\epsilon > 0$, $\phi = o(\upsilon(T)^{-2\beta}\Delta_{n,T} + \upsilon(T)^{-\beta}\Delta_{n,T}^{1/2}/n)$ and $o(k) = \Delta_{n,T}^{-1}\upsilon(T)^{2\beta}$. For $R > 1$, the derived orders yield consistency of the infinitesimal moment estimates, for a reasonable (possibly MSE-optimal) bandwidth choice, with discretization error (due to sampling) and measurement error (due to spot variance estimation). They are compatible with (n) low-frequency (daily) data for the purpose of estimating the dynamics and (k) high-frequency (intra-daily) data on a liquid stock for the purpose of spot variance estimation in the first stage. Since the more stringent Assumption 4.6 (given Assumption 4.2 or $h_{n,T}/\Delta_{n,T}^{1/2} \to \infty$) replaces Assumption 4.7, the case $R = 1$ (first moment estimation) may require lower (than daily) frequencies. Focusing now on weak convergence, the conditions on the measurement error in spot variance estimation (Assumption 4.8 for $R = 1$ and Assumption 4.9 for $R > 1$) necessarily imply tighter restrictions on ϕ and k. The condition on vanishing discretization for weak convergence (Assumption 4.4) is, in general, easily satisfied with daily or, even, with monthly sampling. We refer the reader to the Appendix for further details.

We now turn to the asymptotics. Below, Assumptions 1, 2, and 3 are assumed to hold. We will be specific about which conditions in Assumption 4 are required for each subsequent result to hold.

THEOREM 1 (The feasible estimator of the variance occupation density). *Define* $\widetilde{L}_{\sigma^2}(T,x) = \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)$, *a feasible version of* $\widehat{L}_{\sigma^2}(T,x)$. *Given Assumption 4.0, we have*

$$\frac{\widetilde{L}_{\sigma^2}(T,x)}{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{\sigma_s^2 - x}{h_{n,T}}\right) \mathrm{d}s} = 1 + O_p\left(\frac{\Delta_{n,T}}{h^2_{n,T}}\right) + O_p(g(n,T,k,\phi)).$$

Thus, if Assumptions 4.0, 4.2, and 4.5 hold, then

$$\frac{\widetilde{L}_{\sigma^2}(T,x)}{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{\sigma_s^2 - x}{h_{n,T}}\right) \mathrm{d}s} \xrightarrow{p} 1.$$

Proof. See Appendix. ■

Remark 2. Under the listed conditions, the discretized feasible estimator $\widetilde{L}_{\sigma^2}(T,x)$ is asymptotically equivalent to the variance continuous occupation density, even when spot variance is estimated.

Remark 3. The conditions in Assumption 4 are written with respect to the infeasible estimator $\widehat{L}_{\sigma^2}(T,x)$. This is not problematic. In light of Theorem 1 and its method of proof, we can write

$$\frac{\widetilde{L}_{\sigma^2}(T,x)}{\widehat{L}_{\sigma^2}(T,x)} = \frac{\frac{\widetilde{L}_{\sigma^2}(T,x)}{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{\sigma_s^2 - x}{h_{n,T}}\right)\mathrm{d}s}}{\frac{\widehat{L}_{\sigma^2}(T,x)}{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{\sigma_s^2 - x}{h_{n,T}}\right)\mathrm{d}s}} = \frac{1 + O_p\left(\frac{\Delta_{n,T}}{h^2_{n,T}}\right) + O_p(g(n,T,k,\phi))}{1 + O_p\left(\frac{\Delta_{n,T}}{h^2_{n,T}}\right)} \xrightarrow{p} 1,$$

which yields asymptotic equivalence between $\widetilde{L}_{\sigma^2}(T,x)$ and $\widehat{L}_{\sigma^2}(T,x)$ if $\frac{\Delta_{n,T}}{h^2_{n,T}} \to 0$ and $g(n,T,k,\phi) \to 0$.

Next, we present limiting results for the infinitesimal moment estimates.

THEOREM 2 (The feasible infinitesimal moment estimates). *If Assumptions 4.0, 4.1, 4.2, and 4.6 (for $R = 1$) or 4.7 (for $R > 1$) hold, then*

$$\widehat{\theta}_{f(\sigma^2),R}(x) \xrightarrow{p} \theta_{f(\sigma^2),R}(x) \quad \forall R \geq 1.$$

If Assumptions 4.0, 4.1, 4.3, 4.4, and 4.8 (for $R = 1$) or 4.9 (for $R > 1$) hold, then

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\left\{\widehat{\theta}_{f(\sigma^2),R}(x) - \theta_{f(\sigma^2),R}(x) - \Gamma_{\theta_{f(\sigma^2),R}}(x)\right\} \Rightarrow \mathrm{N}\left(0, \mathbf{K}_2\theta_{f(\sigma^2),2R}(x)\right) \quad \forall R \geq 1,$$

where, given the invariant density $s(\mathrm{d}x)$ of the variance process,

$$\Gamma_{\theta_{f(\sigma^2),R}}(x) = h^2_{n,T}\mathbf{K}_1\left[\frac{\partial\theta_{f(\sigma^2),R}(x)}{\partial x}\frac{\partial s(x)/\partial x}{s(x)} + \frac{1}{2}\frac{\partial^2\theta_{f(\sigma^2),R}(x)}{\partial x \partial x}\right].$$

Proof. See Appendix. ■

Remark 4. The result has a similar look to classical findings in nonparametric kernel estimation. Under stationarity of the variance process, which we assume in the main text, we would also have $\frac{1}{T}\widehat{L}_{\sigma^2}(T,x) \overset{p}{\sim} p(x)$, where $p(.)$ denotes the time-invariant density of the variance process, and $\frac{\partial s(x)/\partial x}{s(x)} = \frac{\partial p(x)/\partial x}{p(x)}$. Because the result, in its general form above, will also apply to the recurrent case discussed in the Appendix, we leave its expression unrestricted in the statement of the theorem. Similarly, we express all results below in their more general form. Importantly, the theorem provides an intuitive representation of statistical uncertainty by inversely relating the width of the asymptotic confidence bands to the number of visits to the interior level x at which pointwise estimation is performed.

Remark 5 (Local linear and local polynomial estimates)**.** As indicated in the Introduction, while we present the results for traditional Nadaraya-Watson kernel estimates, extensions to kernel estimators with superior asymptotic mean-squared error properties can be conducted similarly. Among other methods, $\widehat{\theta}_{f(\sigma^2),R}(x)$ could be a local linear estimator of the form

$$\widehat{\theta}_{R,ll}(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \widetilde{\mathbf{K}}_{i\Delta_{n,T}}\left(x, h_{n,T}\right)\left(f(\widehat{\sigma}^2_{(i+1)\Delta_{n,T}}) - f(\widehat{\sigma}^2_{i\Delta_{n,T}})\right)^R}{\sum_{i=1}^{n} \widetilde{\mathbf{K}}_{i\Delta_{n,T}}\left(x, h_{n,T}\right)} \qquad R = 1, \ldots,$$

where

$$\widetilde{\mathbf{K}}_{i\Delta_{n,T}}\left(x, h_{n,T}\right) = \frac{1}{h_{n,T}}\mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)\Xi_{n,T,2} - (\widehat{\sigma}^2_{i\Delta_{n,T}} - x)\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)\Xi_{n,T,1}$$

with

$$\Xi_{n,T,s} = \sum_{i=1}^{n}(\widehat{\sigma}^2_{i\Delta_{n,T}} - x)^s \frac{1}{h_{n,T}}\mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)$$

for $s = 1, 2$. More generally, it could be a local polynomial estimator defined by minimizing over $\left\{\chi_0, \chi_1, \ldots, \chi_q\right\}$ the criterion

$$\sum_{i=1}^{n-1}\left(\frac{\left(f(\widehat{\sigma}^2_{(i+1)\Delta_{n,T}}) - f(\widehat{\sigma}^2_{i\Delta_{n,T}})\right)^R}{\Delta_{n,T}} - \sum_{u=0}^{q}\chi_u(\widehat{\sigma}^2_{i\Delta_{n,T}} - x)^u\right)^2 \mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right),$$

with $\widehat{\chi}_0 = \widehat{\theta}_{R,ll}(x)$ for $q = 1$ (see, e.g., Fan and Gijbels, 1996). Local polynomial methods for diffusions without jumps are studied in, e.g., Fan and Zhang (2003), Moloche (2004) and Aït-Sahalia and Park (2016).

Consider the local linear estimates. The statement in Theorem 2 would remain unchanged with the exception of an intuitive, given existing work in discrete time, modification to the asymptotic bias: $\Gamma_{\theta_{f(\sigma^2),R}}(x)$ would become $h_{n,T}^2 \mathbf{K}_1 \left[\frac{1}{2} \frac{\partial^2 \theta_{f(\sigma^2),R}(x)}{\partial x \partial x} \right]$.

3. ON THE GENERALITY OF BOUNDED PROCESSES

Two models are observationally-equivalent if they generate the same likelihood of observing the data. We now show that, given an unbounded stochastic variance process, one may specify a bounded process, in terms of its domain, which is arbitrarily close to the original unbounded process and, hence, is approximately observationally-equivalent to it. In light of this observation, Assumption 1(a) may be viewed as being without loss of generality and may handle virtually all stochastic volatility models studied in the literature. In this sense, the treatment in this section is of general interest.

Consider a model in which the variance process follows

$$\mathrm{d}\sigma_t^2 = m_{\sigma^2}(\sigma_t^2)\mathrm{d}t + \Lambda_{\sigma^2}(\sigma_t^2)\mathrm{d}W_t^{\sigma} + \mathrm{d}J_t^{\sigma}, \tag{6}$$

where $m_{\sigma^2}(.)$ and $\Lambda_{\sigma^2}(.)$ are *bounded* functions satisfying Assumption 1(a). Let $dJ_t^{\sigma} = \xi^{\sigma} dN_t^{\sigma}$ with $\xi^{\sigma} \overset{\mathrm{d}}{=} \exp\{\mu_{\xi}\}$.

Importantly, the domain of the variance process σ_t^2 might be unbounded. However, if we consider the *modified* process

$$\widetilde{\sigma}_t^2 = \overline{\sigma}^2 \left(1 - \exp\{-\sigma_t^2/\overline{\sigma}^2\}\right), \tag{7}$$

where $\overline{\sigma}^2$ is a positive constant, the resulting process is so that $0 \leq \widetilde{\sigma}_t^2 < \overline{\sigma}^2$. Also, for any *fixed* σ_t^2, we have that $\widetilde{\sigma}_t^2$ is arbitrarily close to σ_t^2 for a sufficiently large $\overline{\sigma}^2$:[3]

$$\widetilde{\sigma}_t^2 = \overline{\sigma}^2 \left(1 - \left(1 - \sigma_t^2/\overline{\sigma}^2 + \sigma_t^4 O\left(\frac{1}{\overline{\sigma}^4}\right)\right)\right) = \sigma_t^2 + O\left(\frac{1}{\overline{\sigma}^2}\right).$$

Using Itô's lemma for discontinuous processes (see, e.g., Protter, 2004, Theorem 31, page 78), we can characterize the dynamics of the modified process, since

$$\begin{aligned} \mathrm{d}\widetilde{\sigma}_t^2 = & \underbrace{\left(1 - \frac{\widetilde{\sigma}_t^2}{\overline{\sigma}^2}\right)\left[m_{\sigma^2}\left(-\overline{\sigma}^2 \log\left(1 - \frac{\widetilde{\sigma}_t^2}{\overline{\sigma}^2}\right)\right) - \frac{1}{2\overline{\sigma}^2}\Lambda_{\sigma^2}^2\left(-\overline{\sigma}^2 \log\left(1 - \frac{\widetilde{\sigma}_t^2}{\overline{\sigma}^2}\right)\right)\right]}_{m'_{\sigma^2}(\widetilde{\sigma}_t^2)} \mathrm{d}t \\ & + \underbrace{\left(1 - \frac{\widetilde{\sigma}_t^2}{\overline{\sigma}^2}\right)\Lambda_{\sigma^2}\left(-\overline{\sigma}^2 \log\left(1 - \frac{\widetilde{\sigma}_t^2}{\overline{\sigma}^2}\right)\right)}_{\Lambda'_{\sigma^2}(\widetilde{\sigma}_t^2)} \mathrm{d}W_t^{\sigma} + \xi'^{,\sigma}\mathrm{d}N_t^{\sigma}. \end{aligned} \tag{8}$$

where, now, the jump sizes are given by

$$\xi'^{,\sigma} = \overline{\sigma}^2 \left[1 - \left(1 - \frac{\widetilde{\sigma}_t^2}{\overline{\sigma}^2} \right) \exp\{-\xi^{\sigma} / \overline{\sigma}^2\} \right] - \widetilde{\sigma}_t^2.$$

The modified process has a bounded domain and can be made arbitrarily close to the original one in equation (6) by letting $\overline{\sigma}^2$ be sufficiently large. Asking for bounded driving functions to begin with, which was done earlier, is clearly not even needed: a bounded domain would imply bounded functions, if the functions are uniformly continuous.

Figure 2 illustrates these ideas using the popular Heston model, for which $\Lambda_{\sigma^2}(x) = \eta\sqrt{x}$. We plot $\Lambda_{\sigma^2}(x) = \eta\sqrt{x}$ along with the diffusion coefficient $\Lambda'_{\sigma^2}(x)$ for two modified specifications: one with a low $\overline{\sigma}^2$ value, showing that the function $\Lambda'_{\sigma^2}(x)$ of the bounded process $\widetilde{\sigma}_t^2$ is also bounded, and one with a large $\overline{\sigma}^2$ value, showing that $\Lambda'_{\sigma^2}(x)$ and $\Lambda_{\sigma^2}(x)$ can be made virtually indistinguishable.

More explicitly, Taylor expansions (for a *fixed* x value) yield

$$m'_{\sigma^2}(x) = m_{\sigma^2}(x) + O\left(\frac{1}{\overline{\sigma}^2}\right),$$

$$\Lambda'_{\sigma^2}(x) = \Lambda_{\sigma^2}(x) + O\left(\frac{1}{\overline{\sigma}^2}\right),$$

and

$$\xi'^{,\sigma} = \xi^{\sigma} + O_p\left(\frac{1}{\overline{\sigma}^2}\right).$$

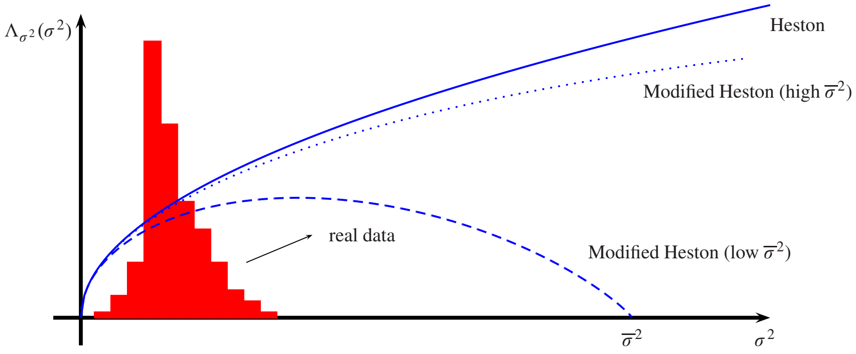

FIGURE 2. (Color online) Solid line: Original Heston model (unbounded diffusion coefficient, unbounded domain for σ_t^2). Dashed line: Modified Heston model with low $\overline{\sigma}^2$. Dotted line: Modified Heston model with high $\overline{\sigma}^2$. The modified versions of the Heston model have bounded coefficients and a bounded domain for σ_t^2.

Again, given boundedness of $\widetilde{\sigma}_t^2$, all of the system's driving functions, including those of the price equation, are automatically bounded since $\widetilde{\sigma}_t^2$ is the sole state variable.

In sum, within the bounded domain of the data (i.e., for a *fixed* σ_t^2 value), observations on σ_t^2 and $\widetilde{\sigma}_t^2$ can be made arbitrarily close by choosing $\overline{\sigma}^2$ suitably large. Thus, the *bounded* jump-diffusion process $\widetilde{\sigma}_t^2$ with large enough $\overline{\sigma}^2$ can not be identified as being different from the original *unbounded* process σ_t^2. Hence, assuming a bounded process for spot variance, as well as bounded functions for the full system as in Assumption 1(a), is without loss of generality. Naturally, the transformation we use in equation (7) is only a possible one delivering boundedness.

4. PRICE/VARIANCE DYNAMICS: ASYMPTOTICS

Recent empirical work has emphasized the importance of models allowing for rapid increases in the conditional volatility of stock returns (see, e.g., Bates, 2000; Duffie et al., 2000; Pan, 2002; Eraker et al., 2003). Such increases cannot be yielded by the small Gaussian changes implied by classical diffusive stochastic volatility models. Jumps in volatility provide an important channel through which sudden volatility jumps translate, due to persistence in the volatility dynamics, into lasting, higher volatility levels (see Eraker et al., 2003, for discussions). In the presence of jumps in volatility, the high-order infinitesimal moments of the volatility process can be employed to learn about the intensity of the jumps and the moments of the jump size distribution as suggested, in more conventional scalar contexts with observables, by Johannes (2004) and studied formally by Bandi and Nguyen (2003).

To this end, we now turn to the identification of nonlinear generalizations of two stochastic volatility models which have drawn attention in recent years, namely the square-root stochastic volatility model with exponential jumps of Duffie et al. (2000) and a logarithmic variance model with Gaussian jumps in the spirit of Jacquier et al. (1994) and Jacquier, Polson, and Rossi (2004).

The original parametric models do not satisfy Assumption 1(a) (for instance, the drift and the diffusion function are, as is often the case, not bounded). Their functional counterparts considered here may also give rise to an unbounded variance process, in general. For the purpose of stating asymptotic results with implementable bandwidth conditions, we therefore work with specifications satisfying Assumption 1(a). In light of the observations in the previous section, this is without loss of generality from a theoretical and an empirical standpoint. To this extent, assume the bounded process $\widetilde{\sigma}_t^2$ (defined in equation (7)) replaces the process σ_t^2 in equations (1)–(2):

$$\mathrm{d}\log p_t = \mu(\widetilde{\sigma}_t^2)\mathrm{d}t + \widetilde{\sigma}_t \mathrm{d}W_t^r + \mathrm{d}J_t^r, \tag{9}$$

$$\mathrm{d}f(\widetilde{\sigma}_t^2) = m'_{f(\widetilde{\sigma}^2)}(\widetilde{\sigma}_t^2)\mathrm{d}t + \Lambda'_{f(\widetilde{\sigma}^2)}(\widetilde{\sigma}_t^2)\mathrm{d}W_t^\sigma + \mathrm{d}J_t^{\prime,\sigma}, \tag{10}$$

where $\mathrm{d}J_t^{\prime,\sigma} = \xi^{\prime,\sigma}\mathrm{d}N_t^\sigma$.

The Generalized Duffie et al. (2000) Model. Write equation (2) with $f(\sigma_t^2) = \sigma_t^2$ and $dJ_t^\sigma = \xi^\sigma dN_t^\sigma$, where $\xi^\sigma \overset{d}{=} \exp(\mu_\xi)$. In Duffie et al. (2000) and Eraker et al. (2003), $m_{\sigma^2}(\sigma_t^2)$ is affine, $\Lambda_{\sigma^2}(\sigma_t^2)$ is a square-root process ($\Lambda_{\sigma^2}^2(\sigma_t^2)$ is also affine) as in Heston (1993) and $\lambda^{\sigma^2}(\sigma_t^2)$ (i.e., the intensity of the Poisson jump N_t^σ) is constant and, hence, independent of the state (see, also, Andersen et al., 2002 for an affine stochastic volatility model with $\lambda^{\sigma^2}(\sigma_t^2) = 0$). Other papers allowing for jumps in returns and stochastic volatility, but without jumps in volatility are, for example, Bakshi, Cao, and Chen (1997), Bates (2000), and Pan (2002). These papers find evidence of misspecifications in the volatility dynamics pointing to the likely presence of discontinuities in the volatility sample path. We leave the functional form of the variance drift, diffusion, and jump intensity unspecified.

Consider the moments $\theta_{\widetilde{\sigma}^2,R}(x)$ of the bounded process $\widetilde{\sigma}_t^2$. We have

$$\theta_{\widetilde{\sigma}^2,1}(x) \asymp m_{\sigma^2}(x) + \mu_\xi(x)\lambda^{\sigma^2}(x), \tag{11}$$

$$\theta_{\widetilde{\sigma}^2,2}(x) \asymp \Lambda_{\sigma^2}^2(x) + 2\mu_\xi^2(x)\lambda^{\sigma^2}(x), \tag{12}$$

$$\theta_{\widetilde{\sigma}^2,3}(x) \asymp 6\mu_\xi^3(x)\lambda^{\sigma^2}(x), \tag{13}$$

$$\theta_{\widetilde{\sigma}^2,4}(x) \asymp 24\mu_\xi^4(x)\lambda^{\sigma^2}(x), \tag{14}$$

where, here and below, the symbol $\asymp$ signifies "arbitrarily close to" for a constant $\overline{\sigma}^2 > 0$ large enough.[4] Identification of the relevant functions may be conducted by computing

$$\widehat{\mu}_\xi(x) = \frac{\widehat{\theta}_{\sigma^2,4}(x)}{4\widehat{\theta}_{\sigma^2,3}(x)}, \tag{15}$$

$$\widehat{\lambda}^{\sigma^2}(x) = \frac{\widehat{\theta}_{\sigma^2,4}(x)}{24\widehat{\mu}_\xi^4(x)}, \tag{16}$$

$$\widehat{\Lambda}_{\sigma^2}^2(x) = \widehat{\theta}_{\sigma^2,2}(x) - 2\widehat{\mu}_\xi^2(x)\widehat{\lambda}^{\sigma^2}(x), \tag{17}$$

$$\widehat{m}_{\sigma^2}(x) = \widehat{\theta}_{\sigma^2,1}(x) - \widehat{\mu}_\xi(x)\widehat{\lambda}^{\sigma^2}(x). \tag{18}$$

Alternative identification schemes using different infinitesimal moments may, of course, be employed. Here, we chose to present the main ideas by considering the most intuitive scheme.

The Generalized Logarithmic Variance Model. Write equation (2) with $f(\sigma_t^2) = \log\sigma_t^2$ and $dJ_t^\sigma = \xi^\sigma dN_t^\sigma$, where $\xi^\sigma \overset{d}{=} \mathrm{N}(0,\sigma_\xi^2)$. We extend the traditional log-linear model by allowing for a nonlinear drift, diffusion, and intensity of the jumps.[5]

Selected moments $\theta_{\log\widetilde{\sigma}^2,R}(x)$ of the bounded process can now be expressed as

$$\theta_{\log\widetilde{\sigma}^2,1}(x) \asymp m_{\log\sigma^2}(x), \tag{19}$$

$$\theta_{\log\widetilde{\sigma}^2,2}(x) \asymp \Lambda^2_{\log\sigma^2}(x) + \sigma^2_{\xi}(x)\lambda^{\log\sigma^2}(x), \tag{20}$$

$$\theta_{\log\widetilde{\sigma}^2,4}(x) \asymp 3\sigma^4_{\xi}(x)\lambda^{\log\sigma^2}(x), \tag{21}$$

$$\theta_{\log\widetilde{\sigma}^2,6}(x) \asymp 15\sigma^6_{\xi}(x)\lambda^{\log\sigma^2}(x). \tag{22}$$

A potential identification method, along the lines of Johannes (2004) and Bandi and Nguyen (2003) in other contexts, is:

$$\widehat{\sigma}^2_{\xi}(x) = \frac{\widehat{\theta}_{\log\sigma^2,6}(x)}{5\widehat{\theta}_{\log\sigma^2,4}(x)}, \tag{23}$$

$$\widehat{\lambda}^{\log\sigma^2}(x) = \frac{\widehat{\theta}_{\log\sigma^2,4}(x)}{3\widehat{\sigma}^4_{\xi}(x)}, \tag{24}$$

$$\widehat{\Lambda}^2_{\log\sigma^2}(x) = \widehat{\theta}_{\log\sigma^2,2}(x) - \widehat{\sigma}^2_{\xi}(x)\widehat{\lambda}^{\log\sigma^2}(x), \tag{25}$$

$$\widehat{m}_{\log\sigma^2}(x) = \widehat{\theta}_{\log\sigma^2,1}(x). \tag{26}$$

4.1. Variance Process: Drift, Diffusion, Intensity of the Jumps and Jump Size Moments

For both models, we now discuss asymptotic inference on the functions of interest. We note that, for the sake of generality, we let the moments of the jump sizes be functions of the latent variance process (c.f., Assumption 1(a)) and estimate them point-wise. If we were to assume that these moments are parameters, we could identify them by, for instance, averaging higher-order infinitesimal moments over any fixed time period $\overline{T}$: $\widehat{\mu}_{\xi} = \frac{1}{\overline{n}}\sum_{i=1}^{\overline{n}} \frac{\widehat{\theta}_{\sigma^2,4}\left(\widehat{\sigma}^2_{i\Delta_{\overline{n},\overline{T}}}\right)}{4\widehat{\theta}_{\sigma^2,3}\left(\widehat{\sigma}^2_{i\Delta_{\overline{n},\overline{T}}}\right)}$ and $\widehat{\sigma}^2_{\xi} = \frac{1}{\overline{n}}\sum_{i=1}^{\overline{n}} \frac{\widehat{\theta}_{\log\sigma^2,6}\left(\widehat{\sigma}^2_{i\Delta_{\overline{n},\overline{T}}}\right)}{5\widehat{\theta}_{\log\sigma^2,4}\left(\widehat{\sigma}^2_{i\Delta_{\overline{n},\overline{T}}}\right)}$, with $\Delta_{\overline{n},\overline{T}} = \overline{T}/\overline{n} \to 0$ as $\overline{n} \to \infty$ over $\overline{T} < T$.

THEOREM 3 (Variance moments: weak convergence in the generalized Duffie et al.'s model). *Let Assumptions 4.0, 4.1, 4.3* (*with* $C = 0$), *4.4, and 4.9 hold.*

1. *Expected jump size*:

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\left\{\widehat{\mu}_{\xi}(x) - \mu^{\diamond}_{\xi}(x)\right\} \Rightarrow \mathrm{N}\left(0, \mathrm{V}(\widehat{\mu}_{\xi}(x))\right),$$

where $\mu^{\diamond}_{\xi}(x) \asymp \mu_{\xi}(x)$ *and* $\mathrm{V}(\widehat{\mu}_{\xi}(x)) \asymp 20\mathbf{K}_2\frac{\mu^2_{\xi}(x)}{\lambda^{\sigma^2}(x)}$.

2. *Jump intensity*:

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\left\{\widehat{\lambda}^{\sigma^2}(x)-\lambda^{\diamond,\sigma^2}(x)\right\}\Rightarrow \mathrm{N}\left(0,\mathrm{V}\left(\widehat{\lambda}^{\sigma^2}(x)\right)\right),$$

where $\lambda^{\diamond,\sigma^2}(x)\asymp\lambda^{\sigma^2}(x)$ *and* $\mathrm{V}\left(\widehat{\lambda}^{\sigma^2}(x)\right)\asymp 110\mathbf{K}_2\lambda^{\sigma^2}(x)$.

3. *Diffusive function*:

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\left\{\widehat{\Lambda}^2_{\sigma^2}(x)-\left(\Lambda^{\diamond}_{\sigma^2}(x)\right)^2\right\}\Rightarrow \mathrm{N}\left(0,\mathrm{V}\left(\widehat{\Lambda}^2_{\sigma^2}(x)\right)\right),$$

where $\left(\Lambda^{\diamond}_{\sigma^2}(x)\right)^2\asymp\Lambda^2_{\sigma^2}(x)$ *and* $\mathrm{V}\left(\widehat{\Lambda}^2_{\sigma^2}(x)\right)\asymp 24\mathbf{K}_2\lambda^{\sigma^2}(x)\mu^4_{\xi}(x)$.

4. *Drift function*:

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\left\{\widehat{m}_{\sigma^2}(x)-m^{\diamond}_{\sigma^2}(x)\right\}\Rightarrow \mathrm{N}\left(0,\mathrm{V}\left(\widehat{m}_{\sigma^2}(x)\right)\right),$$

where $m^{\diamond}_{\sigma^2}(x)\asymp m_{\sigma^2}(x)$ *and* $\mathrm{V}\left(\widehat{m}_{\sigma^2}(x)\right)\asymp\mathbf{K}_2\left(\Lambda^2_{\sigma^2}(x)+38\lambda^{\sigma^2}(x)\mu^2_{\xi}(x)\right)$.

Proof. See Appendix. ■

THEOREM 4 (Variance moments: weak convergence in a generalized logarithmic variance model). *Let Assumptions 4.0, 4.1, 4.3* (*with* $C=0$)*, 4.4, and 4.9 hold.*

1. *Jump standard deviation*:

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\left\{\widehat{\sigma}^2_{\xi}(x)-\left(\sigma^{\diamond}_{\xi}(x)\right)^2\right\}\Rightarrow \mathrm{N}\left(0,\mathrm{V}(\widehat{\sigma}^2_{\xi}(x))\right),$$

where $\left(\sigma^{\diamond}_{\xi}(x)\right)^2\asymp\sigma^2_{\xi}(x)$ *and* $\mathrm{V}(\widehat{\sigma}^2_{\xi}(x))\asymp\frac{238}{15}\mathbf{K}_2\frac{\sigma^4_{\xi}(x)}{\lambda^{\log\sigma^2}(x)}$.

2. *Jump intensity*:

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\left\{\widehat{\lambda}^{\log\sigma^2}(x)-\lambda^{\diamond,\log\sigma^2}(x)\right\}\Rightarrow \mathrm{N}\left(0,\mathrm{V}\left(\widehat{\lambda}^{\log\sigma^2}(x)\right)\right),$$

where $\lambda^{\diamond,\log\sigma^2}(x)\asymp\lambda^{\log\sigma^2}(x)$ *and* $\mathrm{V}\left(\widehat{\lambda}^{\log\sigma^2}(x)\right)\asymp\frac{189}{5}\mathbf{K}_2\lambda^{\log\sigma^2}(x)$.

3. *Diffusive function*:

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\left\{\widehat{\Lambda}^2_{\log\sigma^2}(x)-\left(\Lambda^{\diamond}_{\log\sigma^2}(x)\right)^2\right\}\Rightarrow \mathrm{N}\left(0,\mathrm{V}\left(\widehat{\Lambda}^2_{\log\sigma^2}(x)\right)\right),$$

where $\left(\Lambda^{\diamond}_{\log\sigma^2}(x)\right)^2\asymp\Lambda^2_{\log\sigma^2}(x)$ *and* $\mathrm{V}\left(\widehat{\Lambda}^2_{\log\sigma^2}(x)\right)\asymp\frac{88}{15}\mathbf{K}_2\lambda^{\log\sigma^2}(x)\sigma^4_{\xi}(x)$.

4. *Drift function (with (4.8) replacing (4.9))*:

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\left\{\widehat{m}_{\log\sigma^2}(x)-m^{\diamond}_{\log\sigma^2}(x)\right\}\Rightarrow \mathrm{N}\left(0,\mathrm{V}\left(\widehat{m}_{\log\sigma^2}(x)\right)\right),$$

where $m^{\diamond}_{\log\sigma^2}(x) \asymp m_{\log\sigma^2}(x)$ *and* $\mathrm{V}\left(\widehat{m}_{\log\sigma^2}(x)\right) \asymp \mathbf{K}_2\left(\Lambda^2_{\log\sigma^2}(x)+\lambda^{\log\sigma^2}(x)\sigma^2_{\xi}(x)\right)$.

Proof. See Appendix.

Remark 6. We are assuming that unbounded variance processes with exponential or Gaussian jumps are approximations to bounded (with large $\overline{\sigma}^2$) data generating processes for variance, as in equation (10). The advantage of working with a bounded process for variance is that the procedure results in informative bandwidth conditions (c.f., Assumption 4 and Remark 1). Dealing directly with unbounded processes (a case discussed in the Appendix) would instead yield bandwidth conditions which cannot, in general, be characterized in closed form (due to their dependence on the unknown rates of divergence of certain maximal processes), thereby rendering empirical implementations infeasible (see, e.g., Remark A.4). In addition, writing, as we do, the bounded process as a transformation of an unbounded process for which the infinitesimal moment conditions can be easily expressed (c.f., equations (11)–(14) and equations (19)–(22)) lends itself to simple identification schemes. These schemes do not immediately deliver consistency for the true functions of the unbounded and bounded processes. They, however, deliver consistency for functions which are arbitrarily close (for a large $\overline{\sigma}^2$) to the true functions. Consider, for example, equations (15) and (16). We have:

$$\widehat{\mu}_{\xi}(x)=\frac{\widehat{\theta}_{\sigma^2,4}(x)}{4\widehat{\theta}_{\sigma^2,3}(x)}\xrightarrow{p}\frac{\mathrm{E}((\xi^{\prime,\sigma})^4)\lambda^{\sigma^2}(x)}{4\mathrm{E}((\xi^{\prime,\sigma})^3)\lambda^{\sigma^2}(x)}:=\mu^{\diamond}_{\xi}(x)\asymp\frac{24\mu^4_{\xi}(x)\lambda^{\sigma^2}(x)}{4\times 6\mu^3_{\xi}(x)\lambda^{\sigma^2}(x)}=\mu_{\xi}(x), \quad \mathbf{(27)}$$

$$\widehat{\lambda}^{\sigma^2}(x)=\frac{\widehat{\theta}_{\sigma^2,4}(x)}{24\widehat{\mu}^4_{\xi}(x)}\xrightarrow{p}\frac{\mathrm{E}((\xi^{\prime,\sigma})^4)\lambda^{\sigma^2}(x)}{24\left(\frac{\mathrm{E}((\xi^{\prime,\sigma})^4)\lambda^{\sigma^2}(x)}{4\mathrm{E}((\xi^{\prime,\sigma})^3)\lambda^{\sigma^2}(x)}\right)^4}:=\widehat{\lambda}^{\diamond,\sigma^2}(x)\asymp\frac{24\mu^4_{\xi}(x)\lambda^{\sigma^2}(x)}{24\mu^4_{\xi}(x)}=\lambda^{\sigma^2}(x). \quad \mathbf{(28)}$$

Remark 7. In light of Theorems 3 and 4, asymptotic statistical inference on all functions and parameters of interest is now straightforward given consistent estimates of the asymptotic variances. These estimates are easy to obtain given the closed-form expressions of the asymptotic variances in terms of estimable functions.

4.2. Risk-Return Trade-Off, Price Jumps and Leverage Effects

We now focus on the full system inclusive of the return dynamics. Given spot variance estimates and infinitesimal moment estimates for the return process (which we denote by $\widehat{\theta}_{r,R}(\sigma^2)$ with $R \geq 1$), the relevant functions, inclusive of the features of the return jump distribution, can be estimated by using an identification scheme in the spirit of those reported above.

Let $dJ_t^r = \psi^r dN_t^r$, where $\psi^r \stackrel{d}{=} N(0, \sigma_\psi^2)$. We may employ

$$\widehat{\mu}(x) = \widehat{\theta}_{r,1}(x), \tag{29}$$

$$\widehat{\sigma}_\psi^2(x) = \frac{\widehat{\theta}_{r,6}(x)}{5\widehat{\theta}_{r,4}(x)}, \tag{30}$$

$$\widehat{\lambda}^r(x) = \frac{\widehat{\theta}_{r,4}(x)}{3\widehat{\sigma}_\psi^4(x)}, \tag{31}$$

with

$$\widehat{\theta}_{r,R}(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)\left(\log p_{(i+1)\Delta_{n,T}} - \log p_{i\Delta_{n,T}}\right)^R}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)} \qquad R = 1, \ldots. \tag{32}$$

Should $\widehat{\mu}(x)$ be found to be a statistically increasing function of x, there would be evidence of compensation for variance risk in expected returns. Theorem 5 discusses the weak convergence of $\widehat{\mu}(x)$.

THEOREM 5 (Risk-return trade-off: weak convergence). *Let Assumptions 4.0, 4.1, 4.3, 4.4, and 4.9 hold. Then,*

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\left\{\widehat{\mu}(x) - \mu(x) - \Gamma_\mu(x)\right\} \Rightarrow N(0, V(\widehat{\mu}(x))),$$

with

$$\Gamma_\mu(x) = h_{n,T}^2 \mathbf{K}_1 \left[\frac{\partial \mu(x)}{\partial x}\frac{\partial s(x)/\partial x}{s(x)} + \frac{1}{2}\frac{\partial^2 \mu(x)}{\partial x \partial x}\right],$$

where $V(\widehat{\mu}(x)) = \mathbf{K}_2\left(x + \lambda^r(x)\sigma_\psi^2(x)\right)$.

Proof. See Appendix. ■

In order to complete the system, we turn to the identification of the leverage function. Recent work on leverage estimation in parametric stochastic volatility models includes Harvey and Shephard (1996), Jacquier et al. (2004), and Yu (2005), among others. Yu (2005) provides a thorough discussion of influential parametric approaches and their connections. Write

$$\widehat{\rho}(x) = \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)\left(\log p_{(i+1)\Delta_{n,T}} - \log p_{i\Delta_{n,T}}\right)\left(f(\widehat{\sigma}^2_{(i+1)\Delta_{n,T}}) - f(\widehat{\sigma}^2_{i\Delta_{n,T}})\right)}{\Delta_{n,T}\sum_{i=1}^{n} \mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)\sqrt{x\widehat{\Lambda}^2_{f(\sigma^2)}(x)}},$$

where $\widehat{\Lambda}^2_{f(\sigma^2)}(x)$ may be estimated by virtue of equation (17) or equation (25) depending on the assumed variance model.

Theorem 6 shows that $\widehat{\rho}(x)$ identifies $\rho(x)$ consistently. The theorem assumes that the same bandwidth is used to estimate the numerator and the denominator of $\widehat{\rho}(x)$ and explicitly distinguishes between the linear variance case and the logarithmic variance case.

THEOREM 6 (Leverage function: weak convergence). *If Assumption 4.0, 4.1, 4.3 (with $C = 0$), 4.4, and 4.9 hold, then*

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}\left\{\widehat{\rho}(x) - \rho^{\diamond}(x)\right\} \Rightarrow \mathrm{N}(0, \mathrm{V}(\widehat{\rho}(x)))$$

with $\rho^{\diamond}(x) \asymp \rho(x)$ and

$$\mathrm{V}(\widehat{\rho}(x)) \asymp \begin{cases} \frac{\rho^2(x)}{4\Lambda^4_{\sigma^2}(x)}\left(40\mathbf{K}_2\lambda^{\sigma^2}(x)\mu^2_{\xi}(x)\right) & \textit{with } f(x) = x, \\ \frac{\rho^2(x)}{4\Lambda^4_{\log\sigma^2}(x)}\left(\frac{88}{15}\mathbf{K}_2\lambda^{\log\sigma^2}(x)\sigma^4_{\xi}(x)\right) & \textit{with } f(x) = \log x. \end{cases}$$

Proof. See Appendix. ■

5. FINITE SAMPLE ADJUSTMENTS

Identification of the quantities which characterize the high order moments, namely the jump intensity and the moments of the distribution of the jump sizes, may be improved upon by allowing for higher order terms in the infinitesimal (conditional) moment representations. We consider only the price process, for conciseness, but similar expansions may be derived for the variance moments. Write

$$\theta_{r,4}(x) = 3\sigma^4_{\psi}(x)\lambda^r(x) + \underbrace{3\left[\theta_{r,2}(x)\right]^2\Delta_{n,T}}_{correction} + O(\Delta^2_{n,T})$$

and

$$\begin{aligned}\theta_{r,6}(x) &= 15\sigma^6_{\psi}(x)\lambda^r(x) + \underbrace{15\theta_{r,2}(x)(3\sigma^4_{\psi}(x)\lambda^r(x))\Delta_{n,T}}_{correction} + O(\Delta^2_{n,T}) \\ &= 15\sigma^6_{\psi}(x)\lambda^r(x) + \underbrace{15\theta_{r,2}(x)(\theta_{r,4}(x) - 3\left[\theta_{r,2}(x)\right]^2\Delta_{n,T})\Delta_{n,T}}_{correction} + O(\Delta^2_{n,T}).\end{aligned}$$

Hence, if we assume that σ^2_{ψ} is a parameter, rather than a function, as in the simulations in Section 6 below, we have

$$\widehat{\sigma}_{\psi}^{2*} = \frac{1}{5n} \sum_{i=1}^{n} \frac{\widehat{\theta}_{r,6}\left(\widehat{\sigma}^2_{i\Delta_{n,T}}\right) - 15\widehat{\theta}_{r,2}\left(\widehat{\sigma}^2_{i\Delta_{n,T}}\right)\left(\widehat{\theta}_{r,4}\left(\widehat{\sigma}^2_{i\Delta_{n,T}}\right) - 3\left[\widehat{\theta}_{r,2}\left(\widehat{\sigma}^2_{i\Delta_{n,T}}\right)\right]^2 \Delta_{n,T}\right)\Delta_{n,T}}{\widehat{\theta}_{r,4}\left(\widehat{\sigma}^2_{i\Delta_{n,T}}\right) - 3\left[\widehat{\theta}_{r,2}\left(\widehat{\sigma}^2_{i\Delta_{n,T}}\right)\right]^2 \Delta_{n,T}},$$

$$\widehat{\lambda}^{r*}(x) = \frac{\widehat{\theta}_{r,4}(x) - 3\left[\widehat{\theta}_{r,2}(x)\right]^2 \Delta_{n,T}}{3\widehat{\sigma}_{\psi}^{4*}}.$$

The corrections are asymptotically negligible, but may play an important role in a finite sample. Because of the noisy nature of returns, they do so particularly when estimating return moments, as we show next.

6. SIMULATIONS

We simulate S&P 500 returns from the following bi-variate system:

$$\log p_{t+\Delta_t} - \log p_t = b\Delta_t + \sqrt{\sigma_t^2 \Delta_t}\varepsilon_t^r + \psi^r J_t^r, \tag{33}$$

$$\sigma^2_{t+\Delta_t} - \sigma_t^2 = \kappa(\theta - \sigma_t^2)\Delta_t + \sigma_v\sqrt{\sigma_t^2 \Delta_t}\varepsilon_t^\sigma + \xi^\sigma J_t^\sigma, \tag{34}$$

where $\{J_t^r, J_t^\sigma\}$ are Bernoulli random variables with constant intensities $\lambda^r \Delta_t$ and $\lambda^{\sigma^2}\Delta_t$, $\{\varepsilon_t^r, \varepsilon_t^\sigma\}$ are standard Gaussian random variables with correlation ρ, ψ^r is a zero mean Gaussian random variable with standard deviation σ_ψ, ξ^σ is an exponential random variable with mean μ_ξ, and Δ_t is a time discretization. All random shocks are independent, unless otherwise specified. The parameters are those in Table III, Column 5, of Eraker et al. (2003). We generate $1,000$ replications of a data set with $5,000$ days. The spot variance estimates in equation (4) are obtained with $k = 360$ and $\phi = 1/7$, mimicking 10-second observations sampled over a one-hour time interval. This choice of k and ϕ is consistent with the typical number of intra-daily observations available for liquid financial assets. The threshold ϑ is chosen with the data-driven method suggested by Corsi et al. (2010). We use an Epanechnikov kernel for all infinitesimal moments. The bandwidth $h_{n,T}$ is selected, once more for all infinitesimal moments, using the rule of thumb $\widehat{h}_{n,T} = 1.06\text{std}\left(\widehat{\sigma}^2\right)n^{-1/5}$.

Figure 3 reports the means and the empirical 95% confidence bands of the estimates along with the true functions. The variance dynamics are accurately estimated. The most significant deviations from the true functions occur in correspondence with the intensity of the price jumps and the standard deviation of the price jump size. Panel (a), which is in logarithmic scale to display with clarity the true price intensity and its unadjusted estimate on the same plot, illustrates this result. The use of a first-order correction (as described in Section 5) restores approximate unbiasedness of the price intensity estimates. Corresponding findings (with and without a first-order correction) for the standard deviation of the price jump size are reported in the legend. The corrections are implemented solely for the intensity of the price jumps and for the standard deviation of the price jump size.

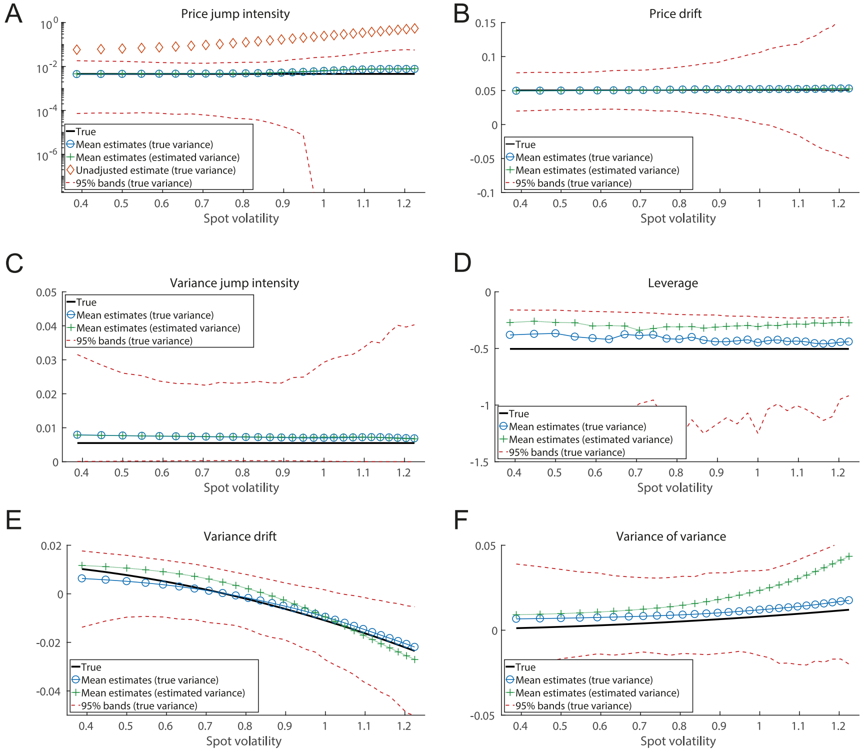

FIGURE 3. We display the true functions of the system in equations (33) and (34) along with mean estimates across 1,000 replications. On the x axis we report the square root of the spot variance in daily units and in percentage terms. The true value of σ_ψ^2 is 8.94. When using the true spot variance, the mean of the unadjusted estimator $\widehat{\sigma}_\psi^2$ is 3.6, while the mean of the adjusted estimator $\widehat{\sigma}_\psi^{2*}$ is 16.21 (with a 95% confidence interval given by $[4.57, 42.04]$). With an estimated spot variance, i.e., with measurement error, these numbers are virtually unchanged. The true value of μ_ξ is 1.798. When using the true spot variance, the mean estimate of μ_ξ is 1.60 (with a 95% confidence interval given by $[0.770, 3.15]$). With an estimated spot variance, the mean estimate of μ_ξ is 1.62 with an almost identical confidence interval.

The impact on the functional estimates of the introduction of measurement (i.e., estimation) error in spot variance is small, the most relevant effect being on the leverage estimator (reported in Panel (d)). This is a small-sample issue which is rather well-known for alternative approaches to nonparametric leverage estimation with high-frequency data, see, e.g., Mykland and Zhang (2009). Bandi and Renò (2012) further discuss this issue.

Finally, even though our bandwidth choice is sensible, the use of smoothing sequences tailored to the divergence properties of the variance's occupation density may deliver superior performance. The design of optimal data-driven bandwidth selection procedures for infinitesimal moment estimation (with and without jumps) is an important open problem which is beyond the scope of the present paper and is, therefore, better left for future work.

7. CONCLUSIONS

A very successful literature has focused on the efficient use of intra-period price observations for the purpose of estimating variance over the period (Aït-Sahalia and Jacod, 2014). This literature aims at being as much as possible "model-free." In the same model-free spirit, we view this paper as an initial effort to render this literature's contributions operative in the context of continuous-time finance modeling under weak (largely nonparametric) assumptions in terms of model specification and under mild (recurrence) conditions needed for identification.

The paper's approach may be employed to focus on specific features of the price/variance dynamics, our work on time-varying leverage estimation being solely an example (Bandi and Renò, 2012). It may also be put to work to identify richer systems in which, e.g., the price/variance discontinuities are correlated (Bandi and Renò, 2016) and the variance process displays multiple components with different levels of persistence.

Finally, in spite of the bivariate nature of the process studied in this paper, consistent with traditional stochastic volatility modeling, the conditioning variable used in this study (variance) is univariate. The methods are, however, naturally suited for extensions to multivariate discontinuous systems with observable or, as in the present paper, unobservable states.

NOTES

1. Given the measure η, we recall that $||g||_r = \left(\int |f|^r d\eta\right)^{1/r}$ and $||g||_\infty = \inf\{x : \eta\{g > x\} = 0\}$.
2. The polynomial family $\mathbf{K}(s) = \frac{(2q+1)!!}{2^{q+1}q!}(1-s^2)^q \mathbf{1}_{(|s|\leq 1)}$, for instance, satisfies these properties. The family includes, e.g., the uniform kernel ($q = 0$), the Epanechnikov kernel ($q = 1$), the biweight kernel ($q = 2$) and the triweight kernel ($q = 3$). The Gaussian kernel is the standardized limit of $\mathbf{K}(s)$ for $q \to \infty$.
3. The statement should, of course, not be viewed as being "uniform" or "path-wise." In particular, nothing is said about the goodness of the approximation for large values of σ_t^2 close to $\overline{\sigma}^2$. The data, however, lays in a bounded domain by definition. All we are saying to justify nonparametric identification using the modified process is that, for any *fixed* value of σ_t^2 *within* the bounded domain of the data, $\widetilde{\sigma}_t^2$ is arbitrarily close to σ_t^2 for a large enough $\overline{\sigma}^2$ (c.f., Figure 2)
4. We note that equations (11) and (12), for instance, can be written as follows:

$$\theta_{\widetilde{\sigma}^2,1}(x) = m'_{\widetilde{\sigma}^2}(x) + \mu_\xi(x)\lambda^{\sigma^2}(x)\frac{\overline{\sigma}^2 - x}{\overline{\sigma}^2 + \mu_\xi(x)},$$

$$\theta_{\widetilde{\sigma}^2,2}(x) = (\Lambda'_{\widetilde{\sigma}^2}(x))^2 + 2\mu_\xi^2(x)\lambda^{\sigma^2}(x)\frac{(\overline{\sigma}^2 - x)^2}{(\overline{\sigma}^2 + 2\mu_\xi(x))(\overline{\sigma}^2 + \mu_\xi(x))}.$$

5. The logarithmic transformation does not have a bounded first derivative for $\widetilde{\sigma}_t^2$ approaching zero (c.f., Assumption 1(a)). This is easily corrected by writing, e.g., $\log(\widetilde{\sigma}_t^2) \vee \log(\epsilon)$, for an ϵ arbitrarily small.

REFERENCES

Aït-Sahalia, Y. & J. Jacod (2014) *High-Frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y. & J.Y. Park (2016) Bandwidth selection and asymptotic properties of local nonparametric estimators in possibly nonstationary continuous-time models. *Journal of Econometrics* 192, 119–138.

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

Bakshi, G., C. Cao, & Z. Chen (1997) Empirical performance of alternative option pricing models. *Journal of Finance* 52, 2003–2049.

Bandi, F.M. & G. Moloche (2017) On the functional estimation of multivariate diffusion processes. *Econometric Theory*, forthcoming.

Bandi, F.M. & T. Nguyen (2003) On the functional estimation of jump-diffusion models. *Journal of Econometrics* 116, 293–328.

Bandi, F.M. & P.C.B. Phillips (2010) Nonstationary continuous-time processes. In *Handbook of Financial Econometrics*, pp. 139–202. North-Holland.

Bandi, F.M. & R. Renò (2012) Time-varying leverage effects. *Journal of Econometrics* 169, 94–113.

Bandi, F.M. & R. Renò (2016) Price and volatility co-jumps. *Journal of Financial Economics* 119, 107–146.

Bates, D. (2000) Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics* 94, 181–238.

Bauwens, L., C.M. Hafner, & S. Laurent (2012) *Handbook of Volatility Models and Their Applications*, vol. 3. John Wiley & Sons.

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

Borkovec, M. & C. Klüppelberg (1998) Extremal behaviour of diffusion models in finance. *Extremes* 1, 47–80.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer-Verlag.

Castellana, J. & M. Leadbetter (1986) On smoothed probability density estimation for stationary processes. *Stochastic Processes and Their Applications* 21, 179–193.

Comte, F., V. Genon-Catalot, & Y. Rozenholc (2010) Nonparametric estimation for a stochastic volatility model. *Finance and Stochastics* 14, 49–80.

Comte, F., C. Lacour, & Y. Rozenholc (2010) Adaptive estimation of the dynamics of a discrete time stochastic volatility model. *Journal of Econometrics* 154, 49–73.

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

Corsi, F., D. Pirino, & R. Renò (2010) Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* 159, 276–288.

Dubins, L.E., L.A. Shepp, & A.N. Shiryaev (1993) Optimal stopping rules and maximal inequalities for bessel processes. *Theory of Probability & its Applications* 38, 226–261.

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

Dzhaparidze, K. & J. Van Zanten (2001) On bernstein-type inequalities for martingales. *Stochastic Processes and Their Applications* 93, 109–117.

Eraker, B., M. Johannes, & N. Polson (2003) The impact of jumps in volatility and returns. *Journal of Finance* 58, 1269–1300.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and its Applications*. Chapman & Hall/CRC.

Fan, J. & Y. Wang (2008) Spot volatility estimation for high-frequency data. *Statistics and its Interface* 1, 279–288.

Fan, J. & C. Zhang (2003) A re-examination of Stanton's diffusion estimations with applications to financial model validation. *Journal of the American Statistical Association* 98, 118–134.

Fasen, V. & C. Klüppelberg (2007) Extremes of supOU processes. In F.E. Benth, G. Di Nunno, T. Lindstrom, B. Oksendal, & T. Zhang (eds.), *Stochastic Analysis and Applications: The Abel Symposium 2005*, pp. 340–349. Springer.

Fasen, V., C. Klüppelberg, & A. Lindner (2006) Extremal behavior of stochastic volatility models. In A.N. Shiryaev, M.d.R. Grossinho, P. Oliveira, & M. Esquivel (eds.), *Stochastic Finance*, pp. 107–155. Springer.

Gapeev, P.V. (2006) On maximal inequalities for some jump processes. Working paper, Humboldt University.

Harvey, A. & N. Shephard (1996) Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics* 14, 429–434.

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

Höpfner, R. & E. Löcherbach (2003) *Limit Theorems for Null Recurrent Markov Processes*, vol. 768. American Mathematical Society.

Jacquier, E., N. Polson, & P. Rossi (1994) Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12, 371–389.

Jacquier, E., N. Polson, & P. Rossi (2004) Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122, 185–212.

Jeong, M. & J. Park (2010) Asymptotic theory of maximum likelihood estimator for diffusion model. Working paper, Indiana University.

Johannes, M. (2004) The statistical and economic role of jumps in continuous-time interest rate models. *Journal of Finance* 59, 227–260.

Kanaya, S. & D. Kristensen (2016) Estimation of stochastic volatility models by nonparametric filtering. *Econometric Theory* 34, 861–916.

Karatzas, I. & E. Shreve (1991) *Brownian Motion and Stochastic Calculus*. Springer-Verlag.

Kristensen, D. (2010) Nonparametric filtering of the realised spot volatility: A kernel-based approach. *Econometric Theory* 26, 60–93.

Kutoyants, Y.A. (1997) On density estimation by the observations of ergodic diffusion processes. In Y.A. Kabanov, B.L. Rozovskii, & A.N. Shiryayev (eds.), *Statistic and Control of Stochastic Processes*, pp. 253–274. World Scientific.

Leblanc, F. (1995) *Estimation par ondelettes de la densité marginale d'un processus stochastique: temps discret, temps continu et discrétisation*. Ph.D. thesis, Univ. Paris VI.

Löcherbach, E. & D. Loukianova (2008) On Nummelin splitting for continuous time Harris recurrent Markov processes and application to kernel estimation for multi-dimensional diffusions. *Stochastic Processes and their Applications* 118, 1301–1321.

Mancini, C. (2009) Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* 36, 270–296.

Meyn, S.P. & R.L. Tweedie (1993) Stability of markovian processes II: Continuous-time processes and sampled chains. *Advances in Applied Probability* 25, 487–517.

Moloche, G. (2004) Local nonparametric estimation of scalar diffusions. Working paper, University of Chicago.

Mykland, P. & L. Zhang (2008) Inference for volatility-type objects and implications for hedging. *Statistics and its Interface* 1, 255–278.

Mykland, P. & L. Zhang (2009) The econometrics of high-frequency data. In M. Kesssler, A. Lindner, & M. Sorensen (eds.), *Statistical Methods for Stochastic Differential Equations*. Chapman & Hall/CRC Press.

Nummelin, E. (1984) *Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press.

Pan, J. (2002) The jump-risk premia implicit in options: Evidence from an integrated time series study. *Journal of Financial Economics* 63, 3–50.

Protter, P. (2004) *Stochastic Integration and Differential Equations*. Springer.

Renò, R. (2006) Nonparametric estimation of stochastic volatility models. *Economics Letters* 90, 390–395.

Renò, R. (2008) Nonparametric estimation of the diffusion coefficient of stochastic volatility models. *Econometric Theory* 24, 1174–1206.

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

Veretennikov, A.Y. (1999) On Castellana–Leadbetter's condition for diffusion density estimation. *Statistical Inference for Stochastic Processes* 2, 1–9.

Yan, L. & Y. Li (2004) Maximal inequalities for CIR processes. *Letters in Mathematical Physics* 67, 111–124.

Yan, L. & B. Zhu (2005) l^p-estimates on diffusion processes. *Journal of Mathematical Analysis and Applications* 303, 418–435.

Yu, J. (2005) On leverage in a stochastic volatility model. *Journal of Econometrics* 127, 165–178.

APPENDIX A: Mathematical Appendix

The proofs are for the general case of driving functions which are, possibly, not uniformly bounded and a volatility process which is recurrent in an unbounded set. This is a case of theoretical interest when letting T grow without bound. The bandwidth conditions in the statements of the theorems in this Appendix will be easily recast in the more familiar uniformly bounded case discussed in the main text. Specifically, the conditions in Assumption 4 will be obtained as sub-cases of the general unbounded case discussed here.

Hereafter, we assume that Assumptions 2 and 3 hold. Assumption 5 replaces Assumption 1 and relaxes it by permitting unboundedness of the system's driving functions, with the exception of the intensities of the jumps $\lambda^r(.)$ and $\lambda^{f(.)}(.)$. The jump random measures of the price and variance process are denoted by $\upsilon_r(\mathrm{d}s, \mathrm{d}\psi^r)$ and $\upsilon_\sigma(\mathrm{d}s, \mathrm{d}\xi^\sigma)$, respectively. Their compensated counterparts are $\overline{\upsilon}_r(\mathrm{d}s, \mathrm{d}\psi^r)$ and $\overline{\upsilon}_\sigma(\mathrm{d}s, \mathrm{d}\xi^\sigma)$. In what follows, the symbol $\Delta^*_{n,T}$ stands for $\mathrm{E}\left(\sup_{s\leq T}\left|F(\sigma_s^2)\right|\right)\Delta_{n,T}$, where $F(\cdot)$ is a function defined next.

Assumption A.1. (a.0) Define the system in equations (1)–(2) on a filtered probability space $(\Omega, (\mathcal{F}_t)_{0\leq t<\infty}, \mathcal{P})$ satisfying the usual conditions. Let $\mu(.)$, $m_{f(\sigma^2)}(.)$, $\Lambda_{f(\sigma^2)}(.)$, $\rho(.)$, $\lambda^r(.)$ and $\lambda^{f(\sigma^2)}(.)$ be, at least, twice continuously-differentiable and Lipschitz. (a.1) There exists a function $F(\cdot)$ satisfying the property $\frac{\mathrm{E}[\sup_{s\leq T}|f^*(\sigma_s^2)|]}{\mathrm{E}[\sup_{s\leq T}|F(\sigma_s^2)|]} = \frac{\mathrm{E}[\mathcal{M}(f^*(\sigma_s^2))]}{\mathrm{E}[\mathcal{M}(F(\sigma_s^2))]} \leq 1$, where $\mathcal{M}(g(\sigma_s^2))$ is the maximal process of the generic variance transformation $g(\sigma_s^2)$ and $f^*(\sigma^2)$ is a function of the variance state σ^2 like σ^2 itself, $\mu(\sigma^2)$, $\rho(\sigma^2)$, $m_{f(\sigma^2)}(\sigma^2)$, $\Lambda^2_{f(\sigma^2)}(\sigma^2)$, σ^4, $\Lambda^4_{f(\sigma^2)}(\sigma^2)$ and so on. (a.2) We further assume that $(\mathrm{E}[\mathcal{M}(F)])^2\,\Delta_{n,T} \to 0$. (a.3) The intensities of the jumps $\lambda^r(.)$ and $\lambda^{f(\sigma^2)}(.)$ and the moments of the jump sizes (which are allowed to be twice continuously-differentiable functions of the variance state) are uniformly bounded. (The domains of the jump distributions can, however, be unbounded.) The intensity of the variance jumps is integrable with respect to the invariant density of the variance process. (a.4) The variance transformation $f(.)$ is monotonically non-decreasing and continuously-differentiable with a bounded first derivative. (b) The spot variance process is Harris recurrent strong Markov and such that Nummelin splitting can be applied (see, e.g., Höpfner and Löcherbach, 2003). The rate of divergence of additive functionals of the process is $\upsilon(T) = U(T)T^\alpha$, with $U(T)$ denoting a slowly-varying function at infinity and the parameter α capturing the recurrence properties of the variance process. Also, let $\lim_{T\to\infty}\frac{1}{\upsilon(T)}\int_0^T\int_0^T ||g_{s,t}(a,b)||_r \mathrm{d}s\mathrm{d}t<\infty$, with $g_{s,t}(a,b)=p_{s,t}(a,b)-p_s(a)p_t(b)$, where $p_{s,t}(.,.)$ is the joint density of the variance process at times s and t and $p_s(.)$ is the marginal density of the variance process at time s, for some $r\in[1,\infty]$ and $\forall a,b$. We, finally, assume $\upsilon(T)h_{n,T}^{2/r}\to\infty$.

Remark A.2. If the functions $\mu(.), m_{f(\sigma^2)}(.),\ldots$ and the process σ_t^2 are not uniformly bounded, then the generic function $f^*(\sigma_t^2)$ will, of course, not be uniformly bounded. Thus, the condition

$$\max_{1\le i\le n}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left|f^*(\sigma_s^2)\right|\mathrm{d}s\overset{p}{\sim}\Delta_{n,T}(\to 0),\tag{A.1}$$

which is a routine approximation in this literature, may not be valid as $T\to\infty$, in general. However, $\max\limits_{1\le i\le n}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left|f^*(\sigma^2)\right|\mathrm{d}s\overset{p}{\sim}\mathrm{E}\left(\mathcal{M}(f^*(\sigma_s^2))\right)\Delta_{n,T}$ will be valid with $\mathcal{M}(f^*(\sigma_s^2))=\sup\limits_{s\le T}\left|f^*(\sigma_s^2)\right|$ defining the maximal process of $f^*(\sigma_s^2)$. Here, we simply assume the existence of a "*maximal*" maximal process $\mathcal{M}(F(\sigma_s^2))$ and write, by virtue of Assumption A.1(a.1), $\max_{1\le i\le n}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left|f^*(\sigma_s^2)\right|\mathrm{d}s\overset{p}{\sim}\mathrm{E}\left(\mathcal{M}(f^*(\sigma_s^2))\right)\Delta_{n,T}\le\mathrm{E}\left(\mathcal{M}(F(\sigma_s^2))\right)\Delta_{n,T}:=\Delta_{n,T}^*$. By Assumption A.1(a.2), therefore, the condition $\Delta_{n,T}^*\to 0$ is invoked to replace the traditional condition $\Delta_{n,T}\to 0$ in equation (A.1). In fact, the assumption implies $\Delta_{n,T}^*=o\left(\frac{1}{\mathrm{E}\left(\mathcal{M}(F(\sigma_s^2))\right)}\right)$, a slightly stronger requirement used in some derivations. Assumption A.1(b) is a generalization to the possibly nonstationary (but recurrent) case of the condition on the "regularity" of the variance sample path in Assumption 1(b).

Remark A.3. The class of continuous-time processes for which a maximal process, i.e., $\sup\limits_{s\le T}|X_s|$, is well-defined in terms of its probability divergence rate is rather well established. In the case of diffusion processes, interesting discussions are provided by, e.g., Borkovec and Klüppelberg (1998) and Jeong and Park (2010). Jeong and Park (2010) employ the notion of maximal process in the context of parametric estimation methods. For some examples of maximal processes in the case of jump processes, we refer the reader to Fasen, Klüppelberg, and Lindner (2006), and Fasen and Klüppelberg (2007). (We thank Claudia Klüpperberg and Vicky Fasen for useful communications on this issue.) Here we focus on moments and appeal to maximal inequalities providing explicit bounds on $\mathrm{E}\left(\sup\limits_{s\le T}|X_s|\right)$ for a specific process X_t.

Next, we report some examples. It is known that, for a standard Brownian motion W_t, the Doob-type inequality

$$\mathrm{E}\left(\sup_{s\le\tau}|W_s|\right)\le\sqrt{2}\sqrt{\mathrm{E}(\tau)},$$

holds for all stopping times τ, see, e.g., Dubins, Shepp, and Shiryaev (1993). A Bessel process of dimension δ starting at zero is such that $\mathrm{E}^{\frac{1}{p}}\left(\left(\sup\limits_{s\le\tau}X_s\right)^p\right)\le$

$\sqrt{\delta}\left(\frac{4-p}{2-p}\right)^{1/p}\mathrm{E}\left(\tau^{p/2}\right)$ for all $0 < p < 2$ (see, e.g., Yan and Zhu, 2005). The bound can be viewed as a refinement of the classical result $\mathrm{E}\left(\sup_{s\leq\tau} X_s\right) \leq \gamma(\delta)\sqrt{\mathrm{E}(\tau)}$ with $\frac{\gamma(\delta)}{\sqrt{\delta}} \to 1$ as $\delta \to \infty$ in, e.g., Dubins et al. (1993). We note that if $\delta = 2$, the process is null recurrent. It is transient if $\delta > 2$ (see the discussion in Bandi and Phillips, 2010). Consider, now, a geometric Brownian motion with percentage drift μ and percentage volatility σ. We have

$$\mathrm{E}\left(\sup_{s\leq\tau} X_s\right) \leq 1 - \frac{\sigma^2}{2\mu} + \frac{\sigma^2}{2\mu}\exp^{\left(-\frac{(\sigma^2-2\mu)^2}{2\sigma^2}\mathrm{E}(\tau)-1\right)}.$$

It is known that the process is not recurrent if $\frac{\sigma^2}{2\mu} \neq 1$. For the square-root process $\mathrm{d}X_t = (a + bX_t)\mathrm{d}t + c\sqrt{X_t}\mathrm{d}W_t$ with $a, c > 0$ and $b < 0$ (and, by a straightforward extension, for Heston's variance model in Section 4 above) we can write

$$\mathrm{E}^{\frac{1}{p}}\left(\left(\sup_{s\leq\tau} X_s\right)^p\right) \leq \gamma_p \frac{c^2 2^{\frac{2a}{c^2}}}{|b|}\mathrm{E}\left(\log^p\left(1 + \frac{a|b|}{c^2}\tau\right)\right),$$

see Yan and Li (2004). Much less is known for processes with jumps. For $X_t = J_t - t/\lambda^2$, where J_t is a compound Poisson with intensity λ and exponential sizes with parameter 1, we have

$$\mathrm{E}\left(\sup_{s\leq\tau} X_s\right) \leq \frac{\sqrt{1 + 2\mathrm{E}[\tau]/\lambda} - 1}{\lambda},$$

see, e.g., Gapeev (2006).

Importantly, we do not need to characterize the rate of divergence of the expected maximal processes $\mathrm{E}\left(\mathcal{M}(f^*(\sigma_s^2))\right)$, something which would in general be *infeasible* given the nonparametric nature of this study. We are simply requesting the expected maximal process to be well-defined for a sensible variance process σ_t^2 and suitable transformations $f^*(\sigma_t^2)$ defined in Assumption A.1(a.1), which is rather innocuous. In order to uniformize the resulting bandwidth conditions, and therefore only for convenience, we are also requesting existence of a "*maximal*" maximal function $F(\sigma_t^2)$ so that $\frac{\mathrm{E}[\sup_{s\leq T}|f^*(\sigma_s^2)|]}{\mathrm{E}[\sup_{s\leq T}|F(\sigma_s^2)|]} \leq 1$, for all $f^*(\sigma_t^2)$. This is, again, an innocuous requirement since $F(\sigma_t^2)$ can be chosen as being equal to the function $f^*(\sigma_t^2)$ in Assumption A.1(a.1) for which the divergence order of $\mathrm{E}\left(\mathcal{M}(f^*(\sigma_s^2))\right)$ is, indeed, "*maximal.*"

Remark A.4. In the discretization conditions, when needed, we will use $\Delta^*_{n,T}$ along with $\Delta_{n,T}$. The symbol $\Delta^*_{n,T}$ should always be interpreted as having the same order as $\mathrm{E}\left(\mathcal{M}(F(\sigma_s^2))\right)\Delta_{n,T}$. Hence, in general, $\frac{\Delta_{n,T}}{\Delta^*_{n,T}} = o(1)$ (or $O(1)$, as in the uniformly bounded case in the main text). In practice, since $\Delta^*_{n,T}$ is process-specific and cannot be known, one always has to empirically treat $\Delta^*_{n,T}$ as being $\Delta_{n,T}$, i.e., the uniformly bounded case. Treating $\Delta^*_{n,T}$ as $\Delta_{n,T}$, or assuming uniform boundedness of the relevant functions is—as we have shown in the main text—not restrictive in practice. Since asymptotics are approximations of data behavior, and the data is bounded no matter how large the time span, using $\Delta^*_{n,T}$ instead of $\Delta_{n,T}$, when needed, is simply an issue of theoretical interest, one which we tackle explicitly in this Appendix.

LEMMA A.5. *Assume $\mathcal{S}(.)$ is a non-negative, continuous and bounded function defined on a compact set* S. *Write $M_{h_{n,T},T}(x) = \frac{1}{h_{n,T}} \int_0^T \mathcal{S}\left(\frac{\sigma_s^2 - x}{h_{n,T}}\right) \mathrm{d}s$. Let Assumption A.1(b) be satisfied. Then, as $T \to \infty$ and $h_{n,T} \to 0$ jointly, we have*

$$\frac{M_{h_{n,T},T}(x)}{\upsilon(T)} \Rightarrow C_{\sigma^2} s(x) \left(\int_{\mathrm{S}} \mathcal{S}(u) \mathrm{d}u \right) \mathrm{g}_\alpha,$$

where C_{σ^2} is a process-specific constant, g_α is the Mittag-Leffler density with parameter α and $s(\mathrm{d}x) = s(x)\mathrm{d}x$ is the invariant measure of the variance process.

Proof of Lemma A.5. The process σ_t^2 is a Harris recurrent strong Markov process with Poisson jumps and, therefore, is càdlàg. If the process has a recurrent atom, Nummelin splitting (Nummelin, 1984) would imply that the increments $\zeta_m := \frac{1}{h_{n,T}} \int_{R_{m-1}}^{R_m} \mathcal{S}\left(\frac{\sigma_s^2 - x}{h_{n,T}}\right) \mathrm{d}s$, where $\{R_m : m \geq 1\}$ are the variance process' regeneration times, are independent. If the process does not have a recurrent atom, e.g., in the case of general multivariate processes, then the construction in Löcherbach and Loukianova (2008) may be used to define regeneration-like times so that the increments constitute a stationary and strong mixing sequence for which $\sigma\{\zeta_m, m \leq g\}$ is independent of $\sigma\left\{\zeta_{g+l}, l \geq 2\right\}$. In this case, the increments would not be independent but ζ_m would be independent of ζ_{m+2}. Irrespective of the splitting, the same method of proof as in Lemma 1 of Bandi and Moloche (2017) leads to the result. ■

Remark A.6. Write $L_{h_{n,T},T}(x) = \frac{1}{h_{n,T}} \int_0^T \mathbf{K}\left(\frac{\sigma_s^2 - x}{h_{n,T}}\right) \mathrm{d}s$ and assume Assumption A.1(b) is satisfied. Then, as $T \to \infty$ and $h_{n,T} \to 0$ *jointly*, we have

$$\frac{L_{h_{n,T},T}(x)}{\upsilon(T)} \Rightarrow C_{\sigma^2} s(x) \mathrm{g}_\alpha,$$

as a straightforward consequence of Lemma A.5. One simply has to recognize that, because $\mathbf{K}(.)$ is a density defined on the compact set S, $\int_{\mathrm{S}} \mathbf{K}(u)\mathrm{d}u = 1$ (c.f., Assumption 3).

Remark A.7. The previous remark provides a weak convergence result for the empirical occupation density (in its continuous version) of a càdlàg spot variance process. The remark makes explicit that, under Assumption A.1(d), $L_{h_{n,T},T}(x) = O_p(\upsilon(T))$ and, therefore, provides a probability rate for occupation densities which depends on its *degree of recurrence* through the parameter α. The link between the empirical occupation density in its continuous version and the empirical occupation density in its discrete version is given by Lemma A.9. Lemma A.5, in its general form with a function $\mathcal{S}(.)$, will be employed in what follows.

LEMMA A.8. *Given Assumption A.1,*

$$\frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right) m_{f(\sigma^2)}\left(\sigma^2_{i\Delta_{n,T}}\right)}{\frac{1}{h_{n,T}} \int_0^T \mathbf{K}\left(\frac{\sigma_s^2 - x}{h_{n,T}}\right) \mathrm{d}s} - \frac{\frac{1}{h_{n,T}} \int_0^T \mathbf{K}\left(\frac{\sigma_s^2 - x}{h_{n,T}}\right) m_{f(\sigma^2)}(\sigma_s^2) \mathrm{d}s}{\frac{1}{h_{n,T}} \int_0^T \mathbf{K}\left(\frac{\sigma_s^2 - x}{h_{n,T}}\right) \mathrm{d}s} = O_p\left(\frac{\Delta^*_{n,T}}{h^2_{n,T}}\right).$$

Proof of Lemma A.8. See the Online Supplement. ■

LEMMA A.9. *Given Assumption A.1,*

$$\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{\sigma^2_s-x}{h_{n,T}}\right)\mathrm{d}s} = 1 + O_p\left(\frac{\Delta_{n,T}}{h^2_{n,T}}\right).$$

Proof of Lemma A.9. Straightforward from Lemma A.8 by setting $m_{f(\sigma^2)}(.) = 1$. It is also easy to show that the discretization's order here is $O_p\left(\frac{\Delta_{n,T}}{h^2_{n,T}}\right)$ rather than $O_p\left(\frac{\Delta^*_{n,T}}{h^2_{n,T}}\right)$. The difference in the rates is induced by the absence of the potentially unbounded function $m_{f(\sigma^2)}(.)$. Finally, the result applies immediately to the ratio $\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{\sigma^2_s-x}{h_{n,T}}\right)\mathrm{d}s}$ inclusive of the n-th term $\frac{\frac{\Delta_{n,T}}{h_{n,T}}\mathbf{K}\left(\frac{\sigma^2_{n\Delta_{n,T}}-x}{h_{n,T}}\right)}{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{\sigma^2_s-x}{h_{n,T}}\right)}$ since this term is of order $O_p\left(\frac{\Delta_{n,T}}{\overline{v}(T)h_{n,T}}\right)$ by the boundedness of the kernel function. Hence, it is of higher order than $O_p\left(\frac{\Delta_{n,T}}{h^2_{n,T}}\right)$. ■

LEMMA A.10. *Given Assumption A.1, if* $\frac{\Delta^*_{n,T}}{h^2_{n,T}} \to 0$ *jointly with* $T, n \to \infty$ *and* $\Delta_{n,T}, h_{n,T} \to 0$*, then*

1. $$\frac{\frac{1}{\sqrt{h_{n,T}}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\int_{\xi}\xi^{\sigma}\overline{v}_{\sigma}(\mathrm{d}s,\mathrm{d}\xi^{\sigma})}{\sqrt{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}} \Rightarrow \mathrm{N}\left(0, \mathbf{K}_2\mathrm{E}_{\xi}((\xi^{\sigma})^2)\lambda^{f(\sigma^2)}(x)\right),$$
2. $$\frac{\frac{1}{\sqrt{h_{n,T}}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\Lambda_{f(\sigma^2)}(\sigma^2_s)\mathrm{d}W^{\sigma}_s}{\sqrt{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}} \Rightarrow \mathrm{N}\left(0, \mathbf{K}_2\Lambda^2_{f(\sigma^2)}(x)\right).$$

Proof of Lemma A.10. See the Online Supplement. ■

LEMMA A.11. *Given Assumption A.1, if* $\frac{\Delta^*_{n,T}}{h^2_{n,T}} \to 0$ *jointly with* $T, n \to \infty$ *and* $\Delta_{n,T}, h_{n,T} \to 0$*, then*

$$\frac{\frac{1}{\sqrt{h_{n,T}}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\left[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\int_{\xi}\xi^{\sigma}\overline{v}_{\sigma}(\mathrm{d}s,\mathrm{d}\xi^{\sigma}) + \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\Lambda_{f(\sigma^2)}(\sigma^2_s)\mathrm{d}W^{\sigma}_s\right]}{\sqrt{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}}$$

$$\Rightarrow \mathrm{N}\left(0, \mathbf{K}_2\left[\mathrm{E}_\xi((\xi^\sigma)^2)\lambda^{f(\sigma^2)}(x) + \Lambda^2_{f(\sigma^2)}(x)\right]\right).$$

Proof of Lemma A.11. See the Online Supplement. ■

LEMMA A.12 (Consistency of the infeasible estimator of the first-order moment). *Given Assumption A.1,*

$$\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\left(f\left(\sigma^2_{(i+1)\Delta_{n,T}}\right)-f\left(\sigma^2_{i\Delta_{n,T}}\right)\right)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} - \theta_{f(\sigma^2),1}(x)$$

$$= O_p(h^2_{n,T}) + O_p\left(\frac{1}{\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}}\right) + O_p\left(\frac{\Delta^*_{n,T}}{h^2_{n,T}}\right).$$

Thus, if $\frac{\Delta^*_{n,T}}{h^2_{n,T}} \to 0$ *and* $h_{n,T}\widehat{L}_{\sigma^2}(T,x) \xrightarrow{p} \infty$ *jointly with* $T, n \to \infty$ *and* $\Delta_{n,T}, h_{n,T} \to 0$*, then*

$$\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\left(f\left(\sigma^2_{(i+1)\Delta_{n,T}}\right)-f\left(\sigma^2_{i\Delta_{n,T}}\right)\right)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \xrightarrow{p} \theta_{f(\sigma^2),1}(x).$$

Proof of Lemma A.12. See the Online Supplement. ■

LEMMA A.13 (Asymptotic normality of the infeasible estimator of the first-order moment). *Given Assumption A.1, if* $h_{n,T}\widehat{L}_{\sigma^2}(T,x) \xrightarrow{p} \infty$, $h^5_{n,T}\widehat{L}_{\sigma^2}(T,x) \xrightarrow{p} C$, *and* $\frac{\Delta^*_{n,T}\sqrt{\widehat{L}_{\sigma^2}(T,x)}}{h^{3/2}_{n,T}} \xrightarrow{p} 0$ *jointly with* $T, n \to \infty$ *and* $\Delta_{n,T}, h_{n,T} \to 0$*, we have*

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}$$
$$\times\left\{\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\left(f\left(\sigma^2_{(i+1)\Delta_{n,T}}\right)-f\left(\sigma^2_{i\Delta_{n,T}}\right)\right)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} - \theta_{f(\sigma^2),1}(x) - \Gamma_{\theta_{f(\sigma^2),1}}(x)\right\}$$
$$\Rightarrow \mathrm{N}(0, \mathbf{K}_2\theta_{f(\sigma^2),2}(x)),$$

where $\Gamma_{\theta_{f(\sigma^2),1}}(x) = h^2_{n,T}\mathbf{K}_1\left(\frac{\partial\theta_{f(\sigma^2),1}(x)}{\partial x}\frac{\partial s(x)/\partial x}{s(x)} + \frac{1}{2}\frac{\partial^2\theta_{f(\sigma^2),1}(x)}{\partial x\partial x}\right)$.

Proof of Lemma A.13. See the Online Supplement. ■

LEMMA A.14 (Consistency of the infeasible estimator of the higher-order moments). *Let $R \geq 2$. Given Assumption A.1,*

$$\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\left(f\left(\sigma^2_{(i+1)\Delta_{n,T}}\right)-f\left(\sigma^2_{i\Delta_{n,T}}\right)\right)^R}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}-\theta_{f(\sigma^2),R}(x)$$

$$=O_p(h^2_{n,T})+O_p\left(\frac{1}{\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}}\right)+O_p\left(\frac{\Delta^*_{n,T}}{h^2_{n,T}}\right).$$

Thus, if $\frac{\Delta^*_{n,T}}{h^2_{n,T}} \to 0$ and $h_{n,T}\widehat{L}_{\sigma^2}(T,x)\xrightarrow{p}\infty$ jointly with $T,n\to\infty$ and $\Delta_{n,T},h_{n,T}\to 0$, then

$$\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\left(f\left(\sigma^2_{(i+1)\Delta_{n,T}}\right)-f\left(\sigma^2_{i\Delta_{n,T}}\right)\right)^R}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}\xrightarrow{p}\theta_{f(\sigma^2),R}(x).$$

Proof of Lemma A.14. See the Online Supplement. ■

LEMMA A.15 (Asymptotic normality of the infeasible estimator of the higher-order moments). *Let $R \geq 2$. Given Assumption A.1, if $h_{n,T}\widehat{L}_{\sigma^2}(T,x)\xrightarrow{p}\infty$, $h^5_{n,T}\widehat{L}_{\sigma^2}(T,x)\xrightarrow{p}C$ and $\frac{\Delta^*_{n,T}\sqrt{\widehat{L}_{\sigma^2}(T,x)}}{h^{3/2}_{n,T}}\xrightarrow{p}0$ jointly with $T,n\to\infty$ and $\Delta_{n,T},h_{n,T}\to 0$, we have*

$$\sqrt{h_{n,T}\widehat{L}_{\sigma^2}(T,x)}$$
$$\times\left\{\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\left(f\left(\sigma^2_{(i+1)\Delta_{n,T}}\right)-f\left(\sigma^2_{i\Delta_{n,T}}\right)\right)^R}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}-\theta_{f(\sigma^2),R}(x)-\Gamma_{\theta_{f(\sigma^2),R}}(x)\right\}$$
$$\Rightarrow \mathrm{N}(0,\mathbf{K}_2\theta_{f(\sigma^2),2R}(x)),$$

where $\Gamma_{\theta_{f(\sigma^2),R}}(x)=h^2_{n,T}\mathbf{K}_1\left(\frac{\partial\theta_{f(\sigma^2),R}(x)}{\partial x}\frac{\partial s(x)/\partial x}{s(x)}+\frac{1}{2}\frac{\partial^2\theta_{f(\sigma^2),R}(x)}{\partial x\partial x}\right).$

Proof of Lemma A.15. See the Online Supplement. ■

Remark A.16. For the classical uniformly bounded case with observable states ($\Delta^*_{n,T}=\Delta_{n,T}$), Lemmas A.13 and A.15 relax the bandwidth conditions provided in Bandi and Nguyen (2003).

LEMMA A.17 (Uniform consistency of the jump-robust spot variance estimator). *Assume* $\widehat{\sigma}^2_{i\Delta_{n,T}}$ *is given by equation (4). If, as* $\vartheta \to 0$, *we have* $\frac{1}{\sqrt{\vartheta}}\left(\sqrt{\frac{\phi}{k}\log\left(\frac{k}{\phi}\right)\mathcal{M}(\sigma^2)+\frac{\phi}{k}\mathcal{M}(|\mu|)}\right)\stackrel{a.s.}{\to} 0$, *then*

$$
\begin{aligned}
&\max_{1\le i\le n}\left|\widehat{\sigma}^2_{i\Delta_{n,T}}-\sigma^2_{i\Delta_{n,T}}\right| \\
&= O_p\left(n\left(\frac{\phi}{k}\right)^2\mathcal{M}(\mu^2)\right)+O_p\left(n\left(\frac{\phi}{k}\right)\mathcal{M}(\sigma^2)\right)+O_p\left(n\left(\frac{\phi}{k}\right)^{3/2}\mathcal{M}^{*1/2}(\sigma^4,\mu^4)\sqrt{\log nk}\right)\\
&\quad+O_p\left(n\frac{\phi}{k}\mathcal{M}^*(\sigma^2)\log nk\right)+O_p\left(n\left(\frac{\phi}{k}\right)^{3/2}\mathcal{M}^*(|\mu|)\mathcal{M}^{*1/2}(\sigma^2)\sqrt{\log nk}\right)\\
&\quad+O_p\left(\frac{\phi}{k}\mathcal{M}(\mu^2)\right)+O_p\left(\mathcal{M}^*(\sigma^2)\sqrt{\frac{\log n\log nk}{k}}\right)+O_p\left(\left(\frac{\phi}{k}\right)^{1/2}\frac{\mathcal{M}^{*1/2}(\sigma^4,\mu^4)\sqrt{\log n}}{\sqrt{k}}\right)\\
&\quad+O_p\left(\mathcal{M}^*(|\mu|)\mathcal{M}^{*1/2}(\sigma^2)\sqrt{\frac{\log nk}{k}}\phi^{1/2}\right)+O_p\left(\mathcal{M}^{*1/2}(\Lambda^2)\phi^{1/2}\sqrt{\log n}+n\phi+\mathcal{M}(m)\phi\right)\\
&= O_p\left(\Theta(n,T,k,\phi)\right),
\end{aligned}
$$

where $\mathcal{M}^{*k}(f)=(\mathrm{E}[\mathcal{M}(f)])^k$ *and* $\mathcal{M}^{*k}(f,g)=\max(\mathrm{E}[\mathcal{M}(f)],\mathrm{E}[\mathcal{M}(g)])^k$.

Proof of Lemma A.17. Write the price process as $\log p=\log\widetilde{p}+J$, where $\widetilde{p}$ is the continuous component and J is the jump component. Recall that $\phi<\Delta_{n,T}$. Also, write $\widetilde{p}_j=\widetilde{p}_{t_j}$, where $t_j=\left(\lfloor\frac{j-1}{k+1}\rfloor+1\right)\Delta_{n,T}-\phi+\left(\frac{j-1}{k}-\lfloor\frac{j-1}{k+1}\rfloor\left(1+\frac{1}{k}\right)\right)\phi$, with $j=1,\dots,n(k+1)$, where $\lfloor . \rfloor$ is the "floor" function. The symbol $\mathbf{1}_{i,j}$ signifies $\mathbf{1}_{\left\{i-1<\frac{j}{k+1}\le i\right\}}$. Here, we are explicit about the case $f(\sigma^2)=\sigma^2$ and, to simplify notation, drop the subscripts from $m_{\sigma^2}(.)$ and $\Lambda_{\sigma^2}(.)$. The case of a general function $f(.)$ satisfying the properties of Assumption A.1(a.4) can be treated similarly.

Let, for $i=1,\dots,n$,

$$
\widetilde{\sigma}^2_{i\Delta_{n,T}}:=\frac{1}{\phi}\sum_{j=1}^{n(k+1)-1}\mathbf{1}_{i,j}\left(\log\widetilde{p}_{j+1}-\log\widetilde{p}_j\right)^2,
$$

which is a spot variance estimator applied to the continuous component of the price process (c.f., equation (4) without threshold). Using Itô's Lemma,

$$
\begin{aligned}
&\max_{1\le i\le n}\left|\widehat{\sigma}^2_{i\Delta_{n,T}}-\sigma^2_{i\Delta_{n,T}}\right|\le\max_{1\le i\le n}\left|\widetilde{\sigma}^2_{i\Delta_{n,T}}-\sigma^2_{i\Delta_{n,T}}\right|+\max_{1\le i\le n}\left|\widehat{\sigma}^2_{i\Delta_{n,T}}-\widetilde{\sigma}^2_{i\Delta_{n,T}}\right|\\
&\le\underbrace{\max_{1\le i\le n}\left|\frac{2}{\phi}\sum_{j=1}^{n(k+1)-1}\mathbf{1}_{i,j}\int_{t_j}^{t_{j+1}}\left(\int_{t_j}^{s}\mu(\sigma^2_v)\mathrm{d}v\right)\mu(\sigma^2_s)\mathrm{d}s\right|}_{V_{1,T,k,\phi}}\\
&\quad+\underbrace{\max_{1\le i\le n}\left|\frac{2}{\phi}\sum_{j=1}^{n(k+1)-1}\mathbf{1}_{i,j}\int_{t_j}^{t_{j+1}}\left(\int_{t_j}^{s}\sigma_v\mathrm{d}W^r_v\right)\mu(\sigma^2_s)\mathrm{d}s\right|}_{V_{2,n,T,k,\phi}}
\end{aligned}
$$

$$
\begin{aligned}
&+\underbrace{\max_{1\le i\le n}\left|\frac{2}{\phi}\sum_{j=1}^{n(k+1)-1}\mathbf{1}_{i,j}\int_{t_j}^{t_{j+1}}\left(\int_{t_j}^{s}\sigma_v\mathrm{d}W_v^r\right)\sigma_s\mathrm{d}W_s^r\right|}_{V_{3,n,T,k}}\\
&+\underbrace{\max_{1\le i\le n}\left|\frac{2}{\phi}\sum_{j=1}^{n(k+1)-1}\mathbf{1}_{i,j}\int_{t_j}^{t_{j+1}}\left(\int_{t_j}^{s}\mu(\sigma_v^2)\mathrm{d}v\right)\sigma_s\mathrm{d}W_s^r\right|}_{V_{4,n,T,k,\phi}}\\
&+\underbrace{\max_{1\le i\le n}\left|\frac{1}{\phi}\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}}\sigma_s^2\mathrm{d}s-\sigma_{i\Delta_{n,T}}^2\right|}_{B_{n,T,\phi}}+\underbrace{\max_{1\le i\le n}\left|\widehat{\sigma}_{i\Delta_{n,T}}^2-\widetilde{\sigma}_{i\Delta_{n,T}}^2\right|}_{J_{n,T,k,\phi}}.
\end{aligned}
$$

We start with the bias term $B_{n,T,\phi}$. Write

$$
\begin{aligned}
&\max_{1\le i\le n}\left|\frac{1}{\phi}\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}}\sigma_s^2\mathrm{d}s-\sigma_{i\Delta_{n,T}}^2\right|=\max_{1\le i\le n}\left|\frac{1}{\phi}\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}}\left(\sigma_s^2-\sigma_{i\Delta_{n,T}}^2\right)\mathrm{d}s\right|\\
&\le\max_{1\le i\le n}\sup_{i\Delta_{n,T}-\phi\le s\le i\Delta_{n,T}}\left|\sigma_s^2-\sigma_{i\Delta_{n,T}}^2\right|\\
&\le\max_{1\le i\le n}\Bigg(\sup_{i\Delta_{n,T}-\phi\le s\le i\Delta_{n,T}}\left|\int_s^{i\Delta_{n,T}}m(\sigma_u^2)\mathrm{d}u\right|+\sup_{i\Delta_{n,T}-\phi\le s\le i\Delta_{n,T}}\left|\int_s^{i\Delta_{n,T}}\Lambda(\sigma_u^2)\mathrm{d}W_u^\sigma\right|\\
&\qquad+\sup_{i\Delta_{n,T}-\phi\le s\le i\Delta_{n,T}}\left|\int_s^{i\Delta_{n,T}}\int_\xi\xi^\sigma v_\sigma(\mathrm{d}u,\mathrm{d}\xi^\sigma)\right|\Bigg)\\
&\le\max_{1\le i\le n}\left(\left|\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}}m(\sigma_u^2)\mathrm{d}u\right|+\left|\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}}\Lambda(\sigma_u^2)\mathrm{d}W_u^\sigma\right|+\left|\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}}\int_\xi\xi^\sigma v_\sigma(\mathrm{d}u,\mathrm{d}\xi^\sigma)\right|\right)\\
&\quad+\max_{1\le i\le n}\Bigg(\sup_{i\Delta_{n,T}-\phi\le s\le i\Delta_{n,T}}\left|\int_{i\Delta_{n,T}-\phi}^{s}m(\sigma_u^2)\mathrm{d}u\right|+\sup_{i\Delta_{n,T}-\phi\le s\le i\Delta_{n,T}}\left|\int_{i\Delta_{n,T}-\phi}^{s}\Lambda(\sigma_u^2)\mathrm{d}W_u^\sigma\right|\\
&\qquad+\sup_{i\Delta_{n,T}-\phi\le s\le i\Delta_{n,T}}\left|\int_{i\Delta_{n,T}-\phi}^{s}\int_\xi\xi^\sigma v_\sigma(\mathrm{d}u,\mathrm{d}\xi^\sigma)\right|\Bigg).
\end{aligned}
$$

We only consider the last three terms since the first three are dominated asymptotically. For the first one, we have

$$
\max_{1\le i\le n}\sup_{i\Delta_{n,T}-\phi\le s\le i\Delta_{n,T}}\left|\int_{i\Delta_{n,T}-\phi}^{s}m(\sigma_u^2)\mathrm{d}u\right|\le\sup_{s\le T}\left|m(\sigma_s^2)\right|\phi.
$$

For the second term, we use the fact that, given any two events A and B, the following holds:

$$
\Pr(A)=\Pr(A\cap B)+\Pr(A\cap\overline{B})\le\Pr(A\cap B)+P(\overline{B}),
$$

so that

$$
\Pr\left(\max_{1\le i\le n}\sup_{i\Delta_{n,T}-\phi\le s\le i\Delta_{n,T}}\left|\int_{i\Delta_{n,T}-\phi}^{s}\Lambda(\sigma_u^2)\mathrm{d}W_u^\sigma\right|\ge C_{n,T,\phi}\right)
$$

$$\leq \Pr\left(\max_{1\leq i\leq n} \sup_{i\Delta_{n,T}-\phi\leq s\leq i\Delta_{n,T}} \left|\int_{i\Delta_{n,T}-\phi}^{s} \Lambda(\sigma_u^2)\mathrm{d}W_u^\sigma\right| \geq C_{n,T,\phi},\ \max_{1\leq i\leq n}\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \Lambda^2(\sigma_u^2)\mathrm{d}u \leq \beta_{T,\phi}\right)$$
$$+\Pr\left(\max_{1\leq i\leq n}\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \Lambda^2(\sigma_u^2)\mathrm{d}u > \beta_{T,\phi}\right).$$

Using Boole's inequality, and the exponential inequality in Corollary 3.4, page 115, of Dzhaparidze and Van Zanten (2001), for $C_{n,T,\phi} > 0$ and $\beta_{T,\phi} > 0$, we obtain

$$\Pr\left(\max_{1\leq i\leq n} \sup_{i\Delta_{n,T}-\phi\leq s\leq i\Delta_{n,T}} \left|\int_{i\Delta_{n,T}-\phi}^{s} \Lambda(\sigma_u^2)\mathrm{d}W_u^\sigma\right| \geq C_{n,T,\phi},\ \max_{1\leq i\leq n}\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \Lambda^2(\sigma_u^2)\mathrm{d}u \leq \beta_{T,\phi}\right)$$
$$\leq \sum_{i=1}^{n}\Pr\left(\sup_{i\Delta_{n,T}-\phi\leq s\leq i\Delta_{n,T}} \left|\int_{i\Delta_{n,T}-\phi}^{s} \Lambda(\sigma_u^2)\mathrm{d}W_u^\sigma\right| \geq C_{n,T,\phi},\ \max_{1\leq i'\leq n}\int_{i'\Delta_{n,T}-\phi}^{i'\Delta_{n,T}} \Lambda^2(\sigma_u^2)\mathrm{d}u \leq \beta_{T,\phi}\right)$$
$$\leq \sum_{i=1}^{n}\Pr\left(\sup_{i\Delta_{n,T}-\phi\leq s\leq i\Delta_{n,T}} \left|\int_{i\Delta_{n,T}-\phi}^{s} \Lambda(\sigma_u^2)\mathrm{d}W_u^\sigma\right| \geq C_{n,T,\phi},\ \int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \Lambda^2(\sigma_u^2)\mathrm{d}u \leq \beta_{T,\phi}\right)$$
$$\leq 2n\exp\left\{-\frac{C_{n,T,\phi}^2}{2\beta_{T,\phi}}\right\}.$$

By Markov's inequality, we also have

$$\Pr\left(\max_{1\leq i\leq n}\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \Lambda^2(\sigma_u^2)\mathrm{d}u > \beta_{T,\phi}\right) \leq \frac{\mathrm{E}\left[\max_{1\leq i\leq n}\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \Lambda^2(\sigma_u^2)\mathrm{d}u\right]}{\beta_{T,\phi}} \leq \mathrm{E}\left[\mathcal{M}(\Lambda^2)\right]\frac{\phi}{\beta_{T,\phi}}.$$

Now, setting $\beta_{T,\phi} = c\mathrm{E}\left[\mathcal{M}(\Lambda^2)\right]\phi$ and $C_{n,T,\phi} = c\left(\mathrm{E}\left[\mathcal{M}(\Lambda^2)\right]\right)^{1/2}\sqrt{\phi\log n} = cC^*_{n,T,\phi}$, we obtain

$$\Pr\left(\frac{1}{C^*_{n,T,\phi}}\max_{1\leq i\leq n} \sup_{i\Delta_{n,T}-\phi\leq s\leq i\Delta_{n,T}} \left|\int_{i\Delta_{n,T}-\phi}^{s} \Lambda(\sigma_u^2)\mathrm{d}W_u^\sigma\right| \geq c\right) \leq 2n^{1-c/2}+\frac{1}{c},$$

which gives $\max_{1\leq i\leq n} \sup_{i\Delta_{n,T}-\phi\leq s\leq i\Delta_{n,T}} \left|\int_{i\Delta_{n,T}-\phi}^{s} \Lambda(\sigma_u^2)\mathrm{d}W_u^\sigma\right| = O_p\left(C^*_{n,T,\phi}\right) = O_p\left(\mathcal{M}^{*1/2}(\Lambda^2)\sqrt{\phi\log n}\right)$.

Regarding the third term, using again Boole's inequality, followed by Markov's inequality, we have

$$\Pr\left(\frac{\max_{1\leq i\leq n}\sup_{i\Delta_{n,T}-\phi\leq s\leq i\Delta_{n,T}}\left|\int_{i\Delta_{n,T}-\phi}^{s}\int_\xi \xi^\sigma\,\upsilon_\sigma(\mathrm{d}u,\mathrm{d}\xi^\sigma)\right|}{n\phi} \geq c\right)$$
$$\leq \sum_{i=1}^{n}\Pr\left(\frac{\sup_{i\Delta_{n,T}-\phi\leq s\leq i\Delta_{n,T}}\left|\int_{i\Delta_{n,T}-\phi}^{s}\int_\xi \xi^\sigma\,\upsilon_\sigma(\mathrm{d}u,\mathrm{d}\xi^\sigma)\right|}{n\phi} \geq c\right)$$

$$\leq \sum_{i=1}^{n} \Pr\left(\frac{\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \int_{\xi} |\xi^{\sigma}| \upsilon_{\sigma}(\mathrm{d}u, \mathrm{d}\xi^{\sigma})}{n\phi} \geq c \right)$$
$$\leq \sum_{i=1}^{n} \frac{\mathrm{E}\left[\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \int_{\xi} |\xi^{\sigma}| \upsilon_{\sigma}(\mathrm{d}u, \mathrm{d}\xi^{\sigma})\right]}{cn\phi} \leq \frac{\mathrm{E}\left[|\xi^{\sigma}|\right]\overline{\lambda}^{\sigma^2}}{c},$$

where $\overline{\lambda}^{\sigma^2}$ is the upper bound on the intensity of the variance jumps, i.e., $\lambda^{\sigma^2}(.)$ (c.f., Assumption A.1(a.3)). This proves that

$$\max_{1\leq i\leq n} \sup_{i\Delta_{n,T}-\phi\leq s\leq i\Delta_{n,T}} \left| \int_{i\Delta_{n,T}-\phi}^{s} \int_{\xi} \xi^{\sigma} \upsilon_{\sigma}(\mathrm{d}u, \mathrm{d}\xi^{\sigma}) \right| = O_p(n\phi).$$

Thus,

$$B_{n,T,\phi} = O_p\left(\mathcal{M}^{*1/2}(\Lambda^2)\phi^{1/2}\sqrt{\log n} + n\phi + \mathcal{M}(m)\phi \right).$$

Next, consider $V_{1,T,k,\phi}$. Write

$$\begin{aligned}
\frac{2}{\phi}\sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j} \int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \mu(\sigma_{\upsilon}^2)\mathrm{d}\upsilon \right) \mu(\sigma_s^2)\mathrm{d}s &\leq \frac{2}{\phi}\sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j} \int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} |\mu(\sigma_{\upsilon}^2)|\mathrm{d}\upsilon \right) |\mu(\sigma_s^2)|\mathrm{d}s \\
&\leq \frac{2}{\phi}\sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j} \int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{t_{j+1}} |\mu(\sigma_{\upsilon}^2)|\mathrm{d}\upsilon \right) |\mu(\sigma_s^2)|\mathrm{d}s \\
&\leq \frac{2}{\phi}\sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j} \left(\int_{t_j}^{t_{j+1}} |\mu(\sigma_s^2)|\mathrm{d}s \right)^2 \\
\text{(by Jensen's inequality)} &\leq \frac{2}{k}\sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j} \left(\int_{t_j}^{t_{j+1}} \mu^2(\sigma_s^2)\mathrm{d}s \right) \\
&= \frac{2}{k}\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \mu^2(\sigma_s^2)\mathrm{d}s \\
&\leq \frac{2\phi\mathcal{M}(\mu^2)}{k},
\end{aligned}$$

since $\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \mu^2(\sigma_s^2)\mathrm{d}s \leq \sup_{s\leq T} \mu^2(\sigma_s^2)\phi = \mathcal{M}(\mu^2)\phi$.

We now turn to $V_{3,n,T,k}$. We have

$$V_{3,n,T,k} \leq \max_{1\leq i\leq n} \sup_{i\Delta_{n,T}-\phi\leq\tau\leq i\Delta_{n,T}} \left| \underbrace{\frac{2}{\phi}\sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j}\mathbf{1}_{\{t_j<\tau\}} \int_{t_j}^{t_{j+1}\wedge\tau} \left(\int_{t_j}^{s} \sigma_{\upsilon}\mathrm{d}W_{\upsilon}^r \right) \sigma_s \mathrm{d}W_s^r}_{V_{3,n,T,k}^i(\tau)} \right|.$$

As before, write

$$
\begin{aligned}
&\Pr\left(V_{3,n,T,k} \geq C_{n,T,k}\right) \\
&\leq \Pr\left(\max_{1\leq i\leq n} \sup_{i\Delta_{n,T}-\phi\leq\tau\leq i\Delta_{n,T}} \left|V^i_{3,n,T,k}(\tau)\right| \geq C_{n,T,k}\right) \\
&\leq \Pr\left(\max_{1\leq i\leq n} \sup_{i\Delta_{n,T}-\phi\leq\tau\leq i\Delta_{n,T}} \left|V^i_{3,n,T,k}(\tau)\right| \geq C_{n,T,k},\ \max_{1\leq i\leq n}[V^i_{3,n,T,k}(i\Delta_{n,T})] \leq \beta_{n,T,k,\phi}\right) \\
&\quad + \Pr\left(\max_{1\leq i\leq n}[V^i_{3,n,T,k}(i\Delta_{n,T})] > \beta_{n,T,k,\phi}\right) \\
&\leq \sum_{i=1}^{n}\Pr\left(\sup_{i\Delta_{n,T}-\phi\leq\tau\leq i\Delta_{n,T}} \left|V^i_{3,n,T,k}(\tau)\right| \geq C_{n,T,k},\ \max_{1\leq i'\leq n}[V^{i'}_{3,n,T,k}(i'\Delta_{n,T})] \leq \beta_{n,T,k,\phi}\right) \\
&\quad + \Pr\left(\max_{1\leq i\leq n}[V^i_{3,n,T,k}(i\Delta_{n,T})] > \beta_{n,T,k,\phi}\right) \\
&\leq \sum_{i=1}^{n}\Pr\left(\sup_{i\Delta_{n,T}-\phi\leq\tau\leq i\Delta_{n,T}} \left|V^i_{3,n,T,k}(\tau)\right| \geq C_{n,T,k}, [V^i_{3,n,T,k}(i\Delta_{n,T})] \leq \beta_{n,T,k,\phi}\right) \\
&\quad + \Pr\left(\max_{1\leq i\leq n}[V^i_{3,n,T,k}(i\Delta_{n,T})] > \beta_{n,T,k,\phi}\right),
\end{aligned}
$$

for $C_{n,T,k} > 0$ and $\beta_{n,T,k,\phi} > 0$, where again, in the third inequality, we employ Boole's inequality to handle the maximum term.

To the first term, we apply, once more, the exponential inequality in Dzhaparidze and Van Zanten (2001) obtaining

$$
\begin{aligned}
&\Pr\left(\sup_{i\Delta_{n,T}-\phi\leq\tau\leq i\Delta_{n,T}} \left|V^i_{3,n,T,k}(\tau)\right| \geq C_{n,T,k}, [V^i_{3,n,T,k}(i\Delta_{n,T})] \leq \beta_{n,T,k,\phi}\right) \\
&\leq 2\exp\left\{-\frac{C^2_{n,T,k}}{2\beta_{n,T,k,\phi}}\right\}.
\end{aligned}
$$

For the second term, we note that each quantity $V^i_{3,n,T,k}(i\Delta_{n,T})$ is a martingale whose quadratic variation satisfies

$$
\begin{aligned}
&\max_{1\leq i\leq n}[V^i_{3,n,T,k}(i\Delta_{n,T})] \\
&= \frac{4}{\phi^2}\max_{1\leq i\leq n}\sum_{j=1}^{n(k+1)-1}\mathbf{1}_{i,j}\left\{\int_{t_j}^{t_{j+1}}\left(\int_{t_j}^{s}\sigma_\upsilon \mathrm{d}W^r_\upsilon\right)^2 \sigma^2_s \mathrm{d}s\right\} \\
&\leq \frac{4}{\phi^2}\mathcal{M}(\sigma^2)\max_{1\leq i\leq n}\sum_{j=1}^{n(k+1)-1}\mathbf{1}_{i,j}\left\{\int_{t_j}^{t_{j+1}}\left(\sup_{t_j\leq s\leq t_j+\phi/k}\left|\int_{t_j}^{s}\sigma_\upsilon \mathrm{d}W^r_\upsilon\right|\right)^2 \mathrm{d}s\right\} \\
&\leq \frac{4}{\phi^2}\mathcal{M}(\sigma^2)\frac{\phi}{k}\max_{1\leq i\leq n}\sum_{j=1}^{n(k+1)-1}\mathbf{1}_{i,j}\left(\sup_{t_j\leq s\leq t_j+\phi/k}\left|\int_{t_j}^{s}\sigma_\upsilon \mathrm{d}W^r_\upsilon\right|\right)^2 \\
&\leq \frac{4}{\phi^2}\mathcal{M}(\sigma^2)\frac{\phi}{k}k\max_{1\leq i\leq n}\left(\max_{i-1\leq j/(k+1)\leq i}\sup_{t_j\leq s\leq t_j+\phi/k}\left|\int_{t_j}^{s}\sigma_\upsilon \mathrm{d}W^r_\upsilon\right|\right)^2
\end{aligned}
$$

$$= \frac{4}{\phi}\mathcal{M}(\sigma^2)\left(\max_{1\le j\le(k+1)n}\sup_{t_j\le s\le t_j+\phi/k}\left|\int_{t_j}^{s}\sigma_v \,\mathrm{d}W_v^r\right|\right)^2,$$

so that

$$\begin{aligned}
&\Pr\left(\max_{1\le i\le n}[V_{3,n,T,k}^i(i\Delta_{n,T})] > \beta_{n,T,k,\phi}\right)\\
&\le \Pr\left(\frac{4}{\phi}\mathcal{M}(\sigma^2)\left(\max_{1\le j\le(k+1)n}\sup_{t_j\le s\le t_j+\phi/k}\left|\int_{t_j}^{s}\sigma_v \,\mathrm{d}W_v^r\right|\right)^2 > \beta_{n,T,k,\phi}\right)\\
&= \Pr\left(\max_{1\le j\le(k+1)n}\sup_{t_j\le s\le t_j+\phi/k}\left|\int_{t_j}^{s}\sigma_v \,\mathrm{d}W_v^r\right| > \beta_{n,T,k,\phi}^{1/2}\left(\frac{4}{\phi}\mathcal{M}(\sigma^2)\right)^{-1/2}\right)\\
&\le \Pr\left(\max_{1\le j\le(k+1)n}\sup_{t_j\le s\le t_j+\phi/k}\left|\int_{t_j}^{s}\sigma_v \,\mathrm{d}W_v^r\right| > \beta_{n,T,k,\phi}^{1/2}\left(\frac{4}{\phi}\mathcal{M}(\sigma^2)\right)^{-1/2},\ \mathcal{M}(\sigma^2)\le M_T\right)\\
&\quad + \Pr\left(\mathcal{M}(\sigma^2) > M_T\right)\\
&\le \Pr\left(\max_{1\le j\le(k+1)n}\sup_{t_j\le s\le t_j+\phi/k}\left|\int_{t_j}^{s}\sigma_v \,\mathrm{d}W_v^r\right| > \beta_{n,T,k,\phi}^{1/2}\left(\frac{4}{\phi}M_T\right)^{-1/2}\right) + \frac{\mathrm{E}\left[\mathcal{M}(\sigma^2)\right]}{M_T},
\end{aligned}$$

for $M_T > 0$. The first term in the last inequality can be bounded using, again, the exponential inequality in Dzhaparidze and Van Zanten (2001), thereby obtaining

$$\begin{aligned}
&\Pr\left(\max_{1\le i\le n}[V_{3,n,T,k}^i(i\Delta_{n,T})] > \beta_{n,T,k,\phi}\right)\\
&\le 2n(k+1)\exp\left\{-\frac{\beta_{n,T,k,\phi}\phi}{4M_T\gamma_{T,k,\phi}}\right\} + \Pr\left(\max_{1\le j\le(k+1)n}\int_{t_j}^{t_{j+1}}\sigma_s^2 ds > \gamma_{T,k,\phi}\right) + \frac{\mathrm{E}\left[\mathcal{M}(\sigma^2)\right]}{M_T}\\
&\le 2n(k+1)\exp\left\{-\frac{\beta_{n,T,k,\phi}\phi}{4M_T\gamma_{T,k,\phi}}\right\} + \frac{\mathrm{E}[\mathcal{M}(\sigma^2)]\frac{\phi}{k}}{\gamma_{T,k,\phi}} + \frac{\mathrm{E}\left[\mathcal{M}(\sigma^2)\right]}{M_T},
\end{aligned}$$

for $\gamma_{T,k,\phi} > 0$. Now, set $M_T = \sqrt{c}\mathcal{M}^*(\sigma^2)$, $\gamma_{T,k,\phi} = \sqrt{c}\mathcal{M}^*(\sigma^2)\frac{\phi}{k}$, $\beta_{n,T,k,\phi} = \gamma_{T,k,\phi}M_T\frac{\log n(k+1)}{\phi}$ and $C_{n,T,k} = c\mathcal{M}^*(\sigma^2)\sqrt{\frac{\log n \log n(k+1)}{k}}$ to achieve

$$\Pr\left(V_{3,n,T,k} > C_{n,T,k}\right) \le 2n^{1-c/2} + 2(n(k+1))^{1-c/4} + \frac{2}{\sqrt{c}},$$

which implies $V_{3,n,T,k} = O_p\left(\mathcal{M}^*(\sigma^2)\sqrt{\frac{\log n \log n(k+1)}{k}}\right) = O_p\left(\mathcal{M}^*(\sigma^2)\sqrt{\frac{\log n \log nk}{k}}\right)$.

As for the term $V_{4,n,T,k,\phi}$, we similarly have

$$V_{4,n,T,k,\phi} \le \max_{1\le i\le n}\sup_{i\Delta_{n,T}-\phi\le\tau\le i\Delta_{n,T}}\left|\underbrace{\frac{2}{\phi}\sum_{j=1}^{n(k+1)-1}\mathbf{1}_{i,j}\mathbf{1}_{\{t_j<\tau\}}\int_{t_j}^{t_{j+1}\wedge\tau}\left(\int_{t_j}^{s}\mu(\sigma_v^2)\mathrm{d}v\right)\sigma_s \,\mathrm{d}W_s^r}_{V_{4,n,T,k,\phi}^i(\tau)}\right|,$$

where

$$\begin{aligned}
\frac{\phi^2}{4}[V^i_{4,n,T,k,\phi}(\tau)] &= \sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j}\mathbf{1}_{\{t_j<\tau\}} \left\{ \int_{t_j}^{t_{j+1}\wedge\tau} \left(\int_{t_j}^{s} \mu(\sigma_v^2)\mathrm{d}v \right)^2 \sigma_s^2 \mathrm{d}s \right\} \\
&\leq \frac{\phi}{k} \sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j}\mathbf{1}_{\{t_j<\tau\}} \left(\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \mu^2(\sigma_v^2)\mathrm{d}v \right) \sigma_s^2 \mathrm{d}s \right) \\
&\leq \frac{1}{2}\frac{\phi}{k} \sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j}\mathbf{1}_{\{t_j<\tau\}} \left\{ \left(\int_{t_j}^{t_{j+1}} \mu^2(\sigma_s^2)\mathrm{d}s \right)^2 + \left(\int_{t_j}^{t_{j+1}} \sigma_s^2 \mathrm{d}s \right)^2 \right\} \\
&\leq \frac{1}{2}\left(\frac{\phi}{k}\right)^2 \sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j}\mathbf{1}_{\{t_j<\tau\}} \left\{ \left(\int_{t_j}^{t_{j+1}} \mu^4(\sigma_s^2)\mathrm{d}s \right) + \left(\int_{t_j}^{t_{j+1}} \sigma_s^4 \mathrm{d}s \right) \right\} \\
&\leq \frac{1}{2}\left(\frac{\phi}{k}\right)^2 \left(\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \mu^4(\sigma_s^2)\mathrm{d}s + \int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \sigma_s^4 \mathrm{d}s \right).
\end{aligned}$$

Reasoning as for terms $B_{n,T,\phi}$ and $V_{3,n,T,k}$, this result implies that $V_{4,n,T,k,\phi} = O_p\left(\left(\frac{\phi^{1/2}}{k^{1/2}}\right) \frac{\mathcal{M}^{*1/2}(\sigma^4,\mu^4)\sqrt{\log n}}{\sqrt{k}} \right)$.

Next, we turn to $V_{2,n,T,k,\phi}$. Write

$$\begin{aligned}
&\max_{1\leq i\leq n} \left| \frac{2}{\phi} \sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j} \int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \sigma_v \mathrm{d}W_v^r \right) \mu(\sigma_s^2)\mathrm{d}s \right| \\
&\leq \max_{1\leq i\leq n} \frac{2}{\phi} \sum_{j=1}^{n(k+1)-1} \mathbf{1}_{i,j} \int_{t_j}^{t_{j+1}} \left(\sup_{t_j\leq s\leq t_j+\phi/k} \left| \int_{t_j}^{s} \sigma_v \mathrm{d}W_v^r \right| \right) |\mu(\sigma_s^2)|\mathrm{d}s \\
&\leq 2\mathcal{M}(|\mu|) \max_{1\leq j\leq (k+1)n} \sup_{t_j\leq s\leq t_j+\phi/k} \left| \int_{t_j}^{s} \sigma_v \mathrm{d}W_v^r \right| = O_p\left(\mathcal{M}^*(|\mu|)\mathcal{M}^{*1/2}(\sigma^2) \sqrt{\frac{\log(nk)}{k}} \phi^{1/2} \right),
\end{aligned}$$

where, to obtain the order term, we apply the same logic used for term $V_{3,n,T,k}$.

Finally, we analyze $J_{n,T,k,\phi}$. Provided $\frac{1}{\sqrt{\vartheta}}\left(\sqrt{\frac{\phi}{k}\log\left(\frac{k}{\phi}\right)\mathcal{M}(\sigma^2)} + \frac{\phi}{k}\mathcal{M}(|\mu|) \right) \overset{a.s.}{\to} 0$, by an application of Theorem 1, point 3, in Mancini (2009), we have that the maximum difference between $\widehat{\sigma}^2_{i\Delta_{n,T}}$ and $\widetilde{\sigma}^2_{i\Delta_{n,T}}$ satisfies, almost surely,

$$\begin{aligned}
&\max_{1\leq i\leq n} \left| \widehat{\sigma}^2_{i\Delta_{n,T}} - \widetilde{\sigma}^2_{i\Delta_{n,T}} \right| = \max_{1\leq i\leq n} \left| \frac{1}{\phi} \sum_{\left\{ \left(\log \frac{p_{j+1}}{p_j} \right)^2 > \vartheta \right\}} \mathbf{1}_{i,j} \left(\log \widetilde{p}_{j+1} - \log \widetilde{p}_j \right)^2 \right| \\
&\leq \frac{1}{\phi} \max_{1\leq i\leq n} N_i^r \max_{1\leq j\leq (k+1)n-1} \left(\log \widetilde{p}_{j+1} - \log \widetilde{p}_j \right)^2,
\end{aligned}$$

with N_i^r defining the number of price Poisson jumps in the interval $\left[i\Delta_{n,T}-\phi, i\Delta_{n,T}\right]$. Denote, now, by $\widetilde{N}_i^r$ a Poisson process with intensity $\overline{\lambda}^r$, i.e., the upper bound on $\lambda^r(.)$. We have, using Markov inequality,

$$\Pr\left(\max_{1\le i\le n} N_i^r \ge c\right) \le \Pr\left(\max_{1\le i\le n} \widetilde{N}_i^r \ge c\right) \le \frac{\mathrm{E}\left[\max_{1\le i\le n} \widetilde{N}_i^r\right]}{c} \le \sum_{i=1}^{n} \frac{\mathrm{E}\left[\widetilde{N}_i^r\right]}{c} \le n\frac{\overline{\lambda}^r \phi}{c}$$

for $0 < c < \infty$. Hence, $\max_{1\le i\le n} \left|N_i^r\right| = O_p(n\phi)$. Next, notice that, by Itô's lemma,

$$\begin{aligned}
&\left(\log \widetilde{p}_{j+1} - \log \widetilde{p}_j\right)^2 \\
&= 2\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \mu(\sigma_v^2)\mathrm{d}v\right)\mu(\sigma_s^2)\mathrm{d}s + 2\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \sigma_v \mathrm{d}W_v^r\right)\mu(\sigma_s^2)\mathrm{d}s \\
&\quad + 2\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \sigma_v \mathrm{d}W_v^r\right)\sigma_s \mathrm{d}W_s^r + 2\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \mu(\sigma_v^2)\mathrm{d}v\right)\sigma_s \mathrm{d}W_s^r + \int_{t_j}^{t_{j+1}} \sigma_s^2 \mathrm{d}s.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\max_{1\le j\le (k+1)n-1} 2\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \mu(\sigma_v^2)\mathrm{d}v\right)\mu(\sigma_s^2)\mathrm{d}s \\
&\le \max_{1\le j\le (k+1)n-1} 2\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{t_{j+1}} \left|\mu(\sigma_v^2)\right| \mathrm{d}v\right)\left|\mu(\sigma_s^2)\right| \mathrm{d}s \\
&\le \max_{1\le j\le (k+1)n-1} \frac{\phi}{k} \int_{t_j}^{t_{j+1}} \mu^2(\sigma_s^2)\mathrm{d}s = O_p\left(\left(\frac{\phi}{k}\right)^2 \mathcal{M}(\mu^2)\right),
\end{aligned}$$

and

$$\max_{1\le j\le (k+1)n-1} \int_{t_j}^{t_{j+1}} \sigma_s^2 \mathrm{d}s = O_p\left(\left(\frac{\phi}{k}\right)\mathcal{M}(\sigma^2)\right).$$

The term $2\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \mu(\sigma_v^2)\mathrm{d}v\right)\sigma_s \mathrm{d}W_s^r$ can be treated like $V_{4,n,T,k,\phi}$, thus

$$\max_{1\le j\le (k+1)n-1} \left|\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \mu(\sigma_v^2)\mathrm{d}v\right)\sigma_s \mathrm{d}W_s^r\right| = O_p\left(\left(\frac{\phi}{k}\right)^{3/2} \mathcal{M}^{*1/2}(\sigma^4, \mu^4)\sqrt{\log nk}\right).$$

The term $2\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \sigma_v \mathrm{d}W_v^r\right)\sigma_s \mathrm{d}W_s^r$ can be treated like $V_{3,n,T,k}$, obtaining

$$\max_{1\le j\le (k+1)n-1} \left|\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \sigma_v \mathrm{d}W_v^r\right)\sigma_s \mathrm{d}W_s^r\right| = O_p\left(\frac{\phi}{k}\mathcal{M}^*(\sigma^2)\log nk\right).$$

The term $2\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \sigma_v \mathrm{d}W_v^r\right)\mu(\sigma_s^2)\mathrm{d}s$ can be treated like $V_{2,n,T,k,\phi}$, giving

$$\max_{1\le j\le (k+1)n-1} 2\int_{t_j}^{t_{j+1}} \left(\int_{t_j}^{s} \sigma_v \mathrm{d}W_v^r\right)\mu(\sigma_s^2)\mathrm{d}s = O_p\left(\left(\frac{\phi}{k}\right)^{3/2} \mathcal{M}^*(|\mu|)\mathcal{M}^{*1/2}(\sigma^2)\sqrt{\log nk}\right).$$

Summarizing,

$$J_{n,T,k,\phi} = O_p\left(n\left(\frac{\phi}{k}\right)^2 \mathcal{M}(\mu^2)\right) + O_p\left(n\left(\frac{\phi}{k}\right)\mathcal{M}(\sigma^2)\right)$$
$$+ O_p\left(n\left(\frac{\phi}{k}\right)^{3/2} \mathcal{M}^{*1/2}(\sigma^4,\mu^4)\sqrt{\log nk}\right)$$
$$+ O_p\left(n\frac{\phi}{k}\mathcal{M}^*(\sigma^2)\log nk\right) + O_p\left(n\left(\frac{\phi}{k}\right)^{3/2} \mathcal{M}^*(|\mu|)\mathcal{M}^{*1/2}(\sigma^2)\sqrt{\log nk}\right),$$

which completes the proof. ■

Remark A.18. When the driving functions and σ_t^2 are uniformly bounded, the statement of Lemma A.17 simplifies to the following. If $\frac{1}{\vartheta}\frac{\phi}{k}\log\left(\frac{k}{\phi}\right) \to 0$, then

$$\max_{1\le i\le n}\left|\widehat{\sigma}^2_{i\Delta_{n,T}} - \sigma^2_{i\Delta_{n,T}}\right|$$
$$= O_p\left(n\left(\frac{\phi}{k}\right)^2\right) + O_p\left(n\left(\frac{\phi}{k}\right)\right) + O_p\left(n\left(\frac{\phi}{k}\right)^{3/2}\sqrt{\log nk}\right) + O_p\left(n\frac{\phi}{k}\log nk\right)$$
$$+ O_p\left(n\left(\frac{\phi}{k}\right)^{3/2}\sqrt{\log nk}\right) + O_p\left(\frac{\phi}{k}\right) + O_p\left(\sqrt{\frac{\log n \log nk}{k}}\right) + O_p\left(\left(\frac{\phi}{k}\right)^{1/2}\frac{\sqrt{\log n}}{\sqrt{k}}\right)$$
$$+ O_p\left(\sqrt{\frac{\log nk}{k}}\phi^{1/2}\right) + O_p\left(\phi^{1/2}\sqrt{\log n} + n\phi + \phi\right) = O_p\left(\sqrt{\frac{\log n \log nk}{k}} + \phi^{1/2}\sqrt{\log n} + n\phi\right).$$

The last expression justifies Assumption 4.5 in the main text.

Remark A.19. For a continuously-differentiable transformation $f(.)$ of variance with a bounded first derivative, as implied by Assumption A.1(a.4), by the mean-value theorem, we have

$$\max_{1\le i\le n}\left|f(\widehat{\sigma}^2_{i\Delta_{n,T}}) - f(\sigma^2_{i\Delta_{n,T}})\right| = \max_{1\le i\le n}\left|f'(s^2_{i\Delta_{n,T}})(\widehat{\sigma}^2_{i\Delta_{n,T}} - \sigma^2_{i\Delta_{n,T}})\right|$$
$$\le \max_{1\le i\le n}\left|f'\left(s^2_{i\Delta_{n,T}}\right)\right| \max_{1\le i\le n}\left|(\widehat{\sigma}^2_{i\Delta_{n,T}} - \sigma^2_{i\Delta_{n,T}})\right| = O_p\left(\Theta(n,T,k,\phi)\right),$$

where $s^2_{i\Delta_{n,T}}$ is a value on the line segment connecting $\widehat{\sigma}^2_{i\Delta_{n,T}}$ and $\sigma^2_{i\Delta_{n,T}}$. The probability order is a consequence of Lemma A.17.

Proof of Theorem 1. Using the mean-value theorem and letting, once more, $s^2_{i\Delta_{n,T}}$ be a value on the line segment connecting $\widehat{\sigma}^2_{i\Delta_{n,T}}$ and $\sigma^2_{i\Delta_{n,T}}$, write

$$\widetilde{L}_{\sigma^2}(T,x) = \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right)$$
$$= \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right) + \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - \sigma^2_{i\Delta_{n,T}}}{h_{n,T}}\right)\mathbf{K}'\left(\frac{s^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right).$$

Thus, we have

$$\left|\frac{\widetilde{L}_{\sigma^2}(T,x)}{\upsilon(T)} - \frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\right|$$
$$\leq \max_{1\leq i\leq n}\left|\frac{\widehat{\sigma}^2_{i\Delta_{n,T}}-\sigma^2_{i\Delta_{n,T}}}{h_{n,T}}\right| \frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n}\left|\mathbf{K}'\left(\frac{\sigma^2_{i\Delta_{n,T}}+s^2_{i\Delta_{n,T}}-\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\right|.$$

Recalling that $s^2_{i\Delta_{n,T}}-\sigma^2_{i\Delta_{n,T}} = O_p(\Theta(n,T,k,\phi))$ uniformly over $i=1,\ldots,n$ by Lemma A.17, using the same method of proof as for Lemma A.9, we have

$$\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n}\left|\mathbf{K}'\left(\frac{\sigma^2_{i\Delta_{n,T}}+O_p(\Theta(n,T,k,\phi))-x}{h_{n,T}}\right)\right| = O_p(1)+O_p\left(\frac{\Delta_{n,T}}{h^2_{n,T}}\right).$$

Also, by Lemma A.17 again,

$$\max_{1\leq i\leq n}\left|\frac{\widehat{\sigma}^2_{i\Delta_{n,T}}-\sigma^2_{i\Delta_{n,T}}}{h_{n,T}}\right| = O_p\left(g(n,T,k,\phi)\right),$$

where $g(n,T,k,\phi)=\frac{1}{h_{n,T}}\Theta(n,T,k,\phi)$. Finally,

$$\left|\frac{\widetilde{L}_{\sigma^2}(T,x)}{\upsilon(T)} - \frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\right|$$
$$=O_p(1)O_p\left(g(n,T,k,\phi)\right)+O_p\left(\frac{\Delta_{n,T}}{h^2_{n,T}}\right)O_p\left(g(n,T,k,\phi)\right). \qquad \blacksquare$$

Proof of Theorem 2. We deal explicitly with the measurement error induced by the estimation of spot variance (as in Lemma A.17). The statement of the theorem then follows from an application of Lemmas A.12, A.13, A.14 and A.15.

For notational brevity, we write $\widehat{\Delta f}_{i\Delta_{n,T}} = f(\widehat{\sigma}^2_{(i+1)\Delta_{n,T}})-f(\widehat{\sigma}^2_{i\Delta_{n,T}})$, $\Delta f_{i\Delta_{n,T}} = f(\sigma^2_{(i+1)\Delta_{n,T}})-f(\sigma^2_{i\Delta_{n,T}})$, $\widehat{\mathbf{K}}_{i\Delta_{n,T}} = \mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)$, $\mathbf{K}_{i\Delta_{n,T}} = \mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-x}{h_{n,T}}\right)$ and $\mathbf{K}_s=\mathbf{K}\left(\frac{\sigma^2_s-x}{h_{n,T}}\right)$. Also, $\mathbf{M}_{n,k,T,\phi}=\max_{1\leq i\leq n}\left|\widehat{\sigma}^2_{i\Delta_{n,T}}-\sigma^2_{i\Delta_{n,T}}\right| = O_p(g(n,T,k,\phi)h_{n,T})$.

We have

$$\widehat{\theta}_{f(\sigma^2),R}(x) - \frac{\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\left(\Delta f_{i\Delta_{n,T}}\right)^R}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i\Delta_{n,T}}} = \underbrace{\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\left(\Delta f_{i\Delta_{n,T}}\right)^R}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}} - \frac{\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\left(\Delta f_{i\Delta_{n,T}}\right)^R}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i\Delta_{n,T}}}}_{\Pi_{1,n,T,k,\phi}}$$

$$+\underbrace{\frac{\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\left(\Delta f_{i\Delta_{n,T}}\right)^R}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}-\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\left(\Delta f_{i\Delta_{n,T}}\right)^R}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}}_{\Pi_{2,n,T,k,\phi}}$$

$$+\underbrace{\frac{\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\left(\widehat{\Delta f}_{i\Delta_{n,T}}\right)^R}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}-\frac{\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\left(\Delta f_{i\Delta_{n,T}}\right)^R}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}}_{\Pi_{3,n,T,k,\phi}}.$$

First, we recall that, by Remark A.19,

$$\max_{1\le i\le n}\left|\widehat{\Delta f}_{i\Delta_{n,T}}-\Delta f_{i\Delta_{n,T}}\right|=O_p\left(\mathbf{M}_{n,k,T,\phi}\right)=O_p(g(n,T,k,\phi)h_{n,T}).$$

We begin with $\Pi_{1,n,T,k,\phi}$. Using the mean-value theorem as in the proof of Theorem 1,

$$\Pi_{1,n,T,k,\phi}=-\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\left(\Delta f_{i\Delta_{n,T}}\right)^R}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i\Delta_{n,T}}}\times\frac{\Delta_{n,T}\left(\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}-\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\right)}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$

$$\le\left|\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\left(\Delta f_{i\Delta_{n,T}}\right)^R}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i\Delta_{n,T}}}\right|\frac{\max_{1\le i\le n}\left|\frac{\widehat{\sigma}^2_{i\Delta_{n,T}}-\sigma^2_{i\Delta_{n,T}}}{h_{n,T}}\right|\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\left|\mathbf{K}'\left(\frac{\sigma^2_{i\Delta_{n,T}}+O_p(\mathbf{M}_{n,k,T,\phi})-x}{h_{n,T}}\right)\right|}{\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$

$$=O_p(1)O_p\left(\frac{\mathbf{M}_{n,k,T,\phi}}{h_{n,T}}\right).$$

We now turn to $\Pi_{3,n,T,k,\phi}$. In the case $R=1$, we immediately have

$$\Pi_{3,n,T,k,\phi}=\frac{\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\left(\widehat{\Delta f}_{i\Delta_{n,T}}-\Delta f_{i\Delta_{n,T}}\right)}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}\le\max_{1\le i\le n-1}\left|\widehat{\Delta f}_{i\Delta_{n,T}}-\Delta f_{i\Delta_{n,T}}\right|\frac{1}{\Delta_{n,T}}$$

$$=O_p(\mathbf{M}_{n,k,T,\phi}/\Delta_{n,T}).$$

For comparison with other order terms, it will however be convenient to write $\Pi_{3,n,T,k,\phi}=O_p(\Delta^*_{n,T}\mathbf{M}_{n,k,T,\phi}/\Delta^2_{n,T})$ when $R=1$, which is more conservative (since $\frac{\Delta_{n,T}}{\Delta^*_{n,T}}=o(1)$, see Remark A.4). If $R=2$, however,

$$\Pi_{3,n,T,k,\phi}=\frac{\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\left(\left(\widehat{\Delta f}_{i\Delta_{n,T}}\right)^2-\left(\Delta f_{i\Delta_{n,T}}\right)^2\right)}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$

$$= \frac{\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \left(\widehat{\Delta f}_{i\Delta_{n,T}} - \Delta f_{i\Delta_{n,T}}\right)\left(\widehat{\Delta f}_{i\Delta_{n,T}} + \Delta f_{i\Delta_{n,T}}\right)}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$

$$= \frac{\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \left(\widehat{\Delta f}_{i\Delta_{n,T}} - \Delta f_{i\Delta_{n,T}}\right)\left(\widehat{\Delta f}_{i\Delta_{n,T}} - \Delta f_{i\Delta_{n,T}} + 2\Delta f_{i\Delta_{n,T}}\right)}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$

$$\leq \max_{1\leq i\leq n-1} \frac{\left|\widehat{\Delta f}_{i\Delta_{n,T}} - \Delta f_{i\Delta_{n,T}}\right|^2}{\Delta_{n,T}} + \frac{2\max_{1\leq i\leq n-1}\left|\widehat{\Delta f}_{i\Delta_{n,T}} - \Delta f_{i\Delta_{n,T}}\right| \sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \left(\left|\Delta f_{i\Delta_{n,T}}\right|\right)}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$

$$= O_p\left(\mathbf{M}_{n,k,T,\phi}^2/\Delta_{n,T} + \mathbf{M}_{n,k,T,\phi}\Delta_{n,T}^*/\Delta_{n,T}^{3/2}\right) = O_p\left(\mathbf{M}_{n,k,T,\phi}\Delta_{n,T}^*/\Delta_{n,T}^{3/2}\right).$$

The driving order term (i.e., $O_p\left(\mathbf{M}_{n,k,T,\phi}\Delta_{n,T}^*/\Delta_{n,T}^{3/2}\right)$) is obtained by writing

$$\frac{\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \left|\Delta f_{i\Delta_{n,T}}\right|}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}} \leq \frac{\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \left|\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \overline{m}_{f(\sigma^2)}(\sigma_s^2)\mathrm{d}s\right|}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$
$$+ \frac{\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \left|\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \Lambda_{f(\sigma^2)}(\sigma_s^2)\mathrm{d}W_s^\sigma\right|}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$
$$+ \frac{\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \left|\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \int_\xi \xi^\sigma \overline{v}_\sigma(\mathrm{d}s, \mathrm{d}\xi^\sigma)\right|}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}},$$

where $\overline{m}_{f(\sigma^2)}(.) = m_{f(\sigma^2)}(.) + \mathrm{E}_\xi(\xi^\sigma)\lambda^\sigma(.)$. The first quantity is such that

$$\frac{\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \left|\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \overline{m}_{f(\sigma^2)}(\sigma_s^2)\mathrm{d}s\right|}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}} \leq \frac{\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left|\overline{m}_{f(\sigma^2)}(\sigma_s^2)\right|\mathrm{d}s}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}} = O_p\left(\frac{\Delta_{n,T}^*}{\Delta_{n,T}}\right).$$

As for the second quantity, the denominator is $O_p(1)$ when standardized by $h_{n,T}$ and $\upsilon(T)$. Using the law of iterated expectations and Burkholder-Davis-Gundy (BDG) inequality (see, e.g., Karatzas and Shreve, 1991, Theorem 3.28, page 166), the order of the numerator (also standardized by $h_{n,T}$ and $\upsilon(T)$) is given by the following:

$$\frac{1}{\upsilon(T)h_{n,T}}\mathrm{E}\left(\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \mathrm{E}_{i\Delta_{n,T}}\left(\left|\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \Lambda_{f(\sigma^2)}(\sigma_s^2)\mathrm{d}W_s^\sigma\right|\right)\right)$$

$$\leq \frac{1}{\upsilon(T)h_{n,T}}\mathrm{E}\left(\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\mathrm{E}_{i\Delta_{n,T}}\left(\left|\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)\mathrm{d}s\right|^{1/2}\right)\right)$$

$$\leq \frac{1}{\upsilon(T)h_{n,T}}\mathrm{E}\left(\sqrt{\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)}\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\left(\left|\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathrm{d}s\right|^{1/2}\right)\right)$$

$$= \frac{1}{\Delta_{n,T}^{1/2}}\mathrm{E}\left(\sqrt{\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)}\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\right),$$

where $\mathrm{E}_{i\Delta_{n,T}}[.] = \mathrm{E}[.|\mathcal{F}_{i\Delta_{n,T}}]$. Now, write

$$\frac{1}{\Delta_{n,T}^{1/2}}\mathrm{E}\left(\sqrt{\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)}\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\right)$$

$$= \frac{1}{\Delta_{n,T}^{1/2}}\mathrm{E}\left(\sqrt{\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)}\left(\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}} - \frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\right)\right)$$

$$+ \frac{1}{\Delta_{n,T}^{1/2}}\mathrm{E}\left(\sqrt{\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)}\left(\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\right)\right).$$

The asymptotic order is driven by the second term. We have

$$\frac{1}{\Delta_{n,T}^{1/2}}\mathrm{E}\left(\sqrt{\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)}\left(\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\right)\right)$$

$$= \frac{1}{\Delta_{n,T}^{1/2}}\mathrm{E}\left(\sqrt{\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)}\left(\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}} - \mathrm{E}\left(\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\right)\right)\right)$$

$$+ \frac{1}{\Delta_{n,T}^{1/2}}\mathrm{E}\left(\sqrt{\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)}\right)\mathrm{E}\left(\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\right)$$

$$\leq \frac{1}{\Delta_{n,T}^{1/2}}\left(\mathrm{E}\left(\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)\right)\right)^{1/2}$$

$$\times\left(\mathrm{E}\left(\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}} - \mathrm{E}\left(\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\right)\right)^2\right)^{1/2} \tag{A.2}$$

$$+ \frac{1}{\Delta_{n,T}^{1/2}}\sqrt{\mathrm{E}\left(\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)\right)}\mathrm{E}\left(\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}_{i\Delta_{n,T}}\right) \tag{A.3}$$

$$\leq \frac{1}{\Delta_{n,T}^{1/2}}\left(\mathrm{E}\left(\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma_s^2)\right)\right)^{1/2} O\left(\frac{1}{\upsilon^{1/2}(T)h_{n,T}^{1/r}}\right) \tag{A.4}$$

$$+\frac{\Delta^*_{n,T}}{\Delta^{3/2}_{n,T}}\mathrm{E}\left(\frac{1}{\upsilon(T)h_{n,T}}\int_0^T \mathbf{K}_s \mathrm{d}s\right)+O\left(\frac{\Delta^*_{n,T}}{\Delta^{3/2}_{n,T}}\frac{\Delta_{n,T}}{h^2_{n,T}}\right) \tag{A.5}$$

$$\sim\frac{1}{\Delta^{1/2}_{n,T}}\left(\mathrm{E}\left(\sup_{s\leq T}\Lambda^2_{f(\sigma^2)}(\sigma^2_s)\right)\right)^{1/2} O\left(\frac{1}{\upsilon^{1/2}(T)h^{1/r}_{n,T}}\right)$$

$$+\frac{\Delta^*_{n,T}}{\Delta^{3/2}_{n,T}}\frac{\mathrm{E}[N_T]}{\upsilon(T)}\mathrm{E}\left(\frac{1}{h_{n,T}}\int_{R_m}^{R_{m+1}} \mathbf{K}_s \mathrm{d}s\right)+O\left(\frac{\Delta^*_{n,T}}{\Delta^{3/2}_{n,T}}\frac{\Delta_{n,T}}{h^2_{n,T}}\right). \tag{A.6}$$

The inequality in equation (A.2) is Cauchy-Schwartz. Equation (A.3) follows from Jensen's inequality. The order term in equation (A.4) (i.e., $O\left(\frac{1}{\upsilon^{1/2}(T)h^{1/r}_{n,T}}\right)=o(1)$, since $\upsilon^{1/2}(T)h^{1/r}_{n,T}\to\infty$ by Assumption A.1(b)) derives from the same method of proof as for Lemma 1 of Bandi and Moloche (2017). The order term in equation (A.5) derives from Lemma A.8 in this paper. Recalling (from Lemma A.5 in this paper) that the notation $\{R_m : m\geq 1\}$ denotes the regeneration times of the variance process and using the symbol N_T to represent the random number of regenerations up to time T, the expression in equation (A.6) derives from Lemma 1 in Bandi and Moloche (2017). Since, $\mathrm{E}[N_T]\sim\upsilon(T)$, we readily have $\frac{\mathrm{E}[N_T]}{\upsilon(T)}=O(1)$. Thus,

$$\frac{\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\left|\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\Lambda_{f(\sigma^2)}(\sigma^2_s)\mathrm{d}W^\sigma_s\right|}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}=O_p\left(\frac{\Delta^*_{n,T}}{\Delta^{3/2}_{n,T}}\right).$$

Using now Meyer's version of BDG inequality, the third quantity can be treated in the same way. Because, however, the jumps have a bounded intensity by Assumption A.1(a.3), the order is $O_p\left(\frac{1}{\Delta^{1/2}_{n,T}}\right)$.

When $R>2$, write

$$\Pi_{3,n,T,k,\phi}=\frac{\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\left(\left(\widehat{\Delta f}_{i\Delta_{n,T}}\right)^R-\left(\Delta f_{i\Delta_{n,T}}\right)^R\right)}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$

$$=\frac{\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\left(\left(\Delta f_{i\Delta_{n,T}}+\widehat{\Delta f}_{i\Delta_{n,T}}-\Delta f_{i\Delta_{n,T}}\right)^R-\left(\Delta f_{i\Delta_{n,T}}\right)^R\right)}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$

$$=\sum_{j=1}^{R}\binom{R}{j}\frac{\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_{i\Delta_{n,T}}\left(\widehat{\Delta f}_{i\Delta_{n,T}}-\Delta f_{i\Delta_{n,T}}\right)^j\left(\Delta f_{i\Delta_{n,T}}\right)^{R-j}}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}.$$

For each term in the sum above, we obtain

$$\frac{\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \left(\widehat{\Delta f}_{i\Delta_{n,T}} - \Delta f_{i\Delta_{n,T}}\right)^j \left(\Delta f_{i\Delta_{n,T}}\right)^{R-j}}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$

$$\leq \max_{1\leq i\leq n-1} \left|\widehat{\Delta f}_{i\Delta_{n,T}} - \Delta f_{i\Delta_{n,T}}\right|^j \frac{\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_{i\Delta_{n,T}} \left|\left(\Delta f_{i\Delta_{n,T}}\right)^{R-j}\right|}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}}.$$

It is immediate to see that, if $j = R$, the order of the latter term is $O_p\left(\mathbf{M}_{n,k,T,\phi}^R / \Delta_{n,T}\right)$. If $j < R$, it is instead $O_p\left(\mathbf{M}_{n,k,T,\phi}^j \Delta_{n,T}^* / \Delta_{n,T}^{3/2}\right)$. Intuitively, this is because, using Itô's lemma, for all $k \geq 1$, it is the case that

$$\begin{aligned}
&\left|\left(\Delta f_{i\Delta_{n,T}}\right)^k\right| \\
&= \left| k \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(f(\sigma_s^2) - f(\sigma_{i\Delta_{n,T}}^2)\right)^{k-1} m_{f(\sigma^2)}(\sigma_s^2)\mathrm{d}s \right. \\
&\quad + k \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(f(\sigma_s^2) - f(\sigma_{i\Delta_{n,T}}^2)\right)^{k-1} \Lambda_{f(\sigma^2)}(\sigma_s^2)\mathrm{d}W_s^\sigma \\
&\quad + \frac{1}{2}k(k-1) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(f(\sigma_s^2) - f(\sigma_{i\Delta_{n,T}}^2)\right)^{k-2} \Lambda_{f(\sigma^2)}^2(\sigma_s^2)\mathrm{d}s \\
&\quad \left. + \sum_{\Delta\sigma_s^2\neq 0, i\Delta_{n,T}\leq s\leq (i+1)\Delta_{n,T}} \left[\left(f(\sigma_{s-}^2 + \Delta\sigma_s^2) - f(\sigma_{i\Delta_{n,T}}^2)\right)^k - \left(f(\sigma_{s-}^2) - f(\sigma_{i\Delta_{n,T}}^2)\right)^k\right] \right| \\
&= O_p\left(\frac{\Delta_{n,T}^*}{\Delta_{n,T}^{1/2}}\right).
\end{aligned}$$

The final order follows from the fact that, as in the proof of Lemma A.14, the dominating term in the expansion of the k-th power of the absolute value of $\Delta f_{i\Delta_{n,T}}$ is the compensated jump component. This component can be handled as in the case $R = 2$ and is, when weighted by the kernel, of order $O_p\left(\frac{\Delta_{n,T}^*}{\Delta_{n,T}^{3/2}}\right)$. Thus, when $R \geq 2$, $\Pi_{3,n,T,k,\phi} = O_p\left(\mathbf{M}_{n,k,T,\phi} \Delta_{n,T}^* / \Delta_{n,T}^{3/2}\right)$.

Finally, by the mean-value theorem and the same reasoning as above,

$$\Pi_{2,n,T,k,\phi} = \frac{\sum_{i=1}^{n-1} \left(\widehat{\mathbf{K}}_{i\Delta_{n,T}} - \mathbf{K}_{i\Delta_{n,T}}\right) \left(\Delta f_{i\Delta_{n,T}}\right)^R}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$

$$\leq \max_{1\leq i\leq n}\left|\frac{\widehat{\sigma}^2_{i\Delta_{n,T}}-\sigma^2_{i\Delta_{n,T}}}{h_{n,T}}\right| \frac{\frac{1}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n-1}\left|\mathbf{K}'\left(\frac{\sigma^2_{i\Delta_{n,T}}+O_p(\mathbf{M}_{n,k,T,\phi})-x}{h_{n,T}}\right)\right|\left|\left(\Delta f_{i\Delta_{n,T}}\right)^R\right|}{\frac{\Delta_{n,T}}{\upsilon(T)h_{n,T}}\sum_{i=1}^{n}\widehat{\mathbf{K}}_{i\Delta_{n,T}}}$$

$$= O_p\left(\frac{\mathbf{M}_{n,k,T,\phi}\,\Delta^*_{n,T}}{h_{n,T}\,\Delta^{3/2}_{n,T}}\right).$$

This is the dominating order when $R > 1$ whereas, for the case $R = 1$, the dominating order is $O_p\left(\frac{\mathbf{M}_{n,k,T,\phi}\,\Delta^*_{n,T}}{\Delta^2_{n,T}}\right)$, due to Assumption 4.2. The latter derives from $\Pi_{3,n,T,k,\phi}$. ■

Proof of Theorems 3 and 4. See the Online Supplement. ■

Proof of Theorem 5. The same method of proof as for Theorem 2 leads to the result. The conditions on the vanishing discretization error for consistency and weak converge depend on the fact that prices, contrary to spot variance, are observed. Hence, only terms analogous to $\Pi_{1,n,T,k,\phi}$ and $\Pi_{2,n,T,k,\phi}$ in the proof of Theorem 2 would appear. Thus, the order of the estimation error is, in this case, $O_p\left(\frac{\mathbf{M}_{n,k,T,\phi}\,\Delta^*_{n,T}}{h_{n,T}\,\Delta^{3/2}_{n,T}}\right)$. ■

Proof of Theorem 6. See the Online Supplement. ■

Proof and Further Discussion of Remark 1. First, we focus on conditions which are needed for weak convergence of the moments in the presence of discretely-sampled observations but in the absence of measurement error due to spot variance estimation: 1) $h_{n,T}\widehat{L}_{\sigma^2}(T,x) \xrightarrow{p} \infty$ (consistency), 2) $h^5_{n,T}\widehat{L}_{\sigma^2}(T,x) \xrightarrow{p} C$ (centering of the limiting distribution), and 3) $\frac{\Delta_{n,T}\sqrt{\widehat{L}_{\sigma^2}(T,x)}}{h^{3/2}_{n,T}} \xrightarrow{p} 0$ (vanishing discretization error). Given Lemma A.9, if $\frac{\Delta_{n,T}}{h^2_{n,T}} \to 0$, then $\widehat{L}_{\sigma^2}(T,x) = O_p(L_{h_{n,T},T}(x))$, where $L_{h_{n,T},T}(x)$ is defined in Remark A.6. This implies that 1') $h_{n,T}L_{h_{n,T},T}(x) \xrightarrow{p} \infty$, 2') $h^5_{n,T}L_{h_{n,T},T}(x) \xrightarrow{p} C$, and 3') $\frac{\Delta_{n,T}\sqrt{L_{h_{n,T},T}(x)}}{h^{3/2}_{n,T}} \xrightarrow{p} 0$ yield 1), 2), and 3) since 1') and 3') guarantee $\frac{\Delta_{n,T}}{h^2_{n,T}} \to 0$. Furthermore, in light of Remark A.6, $L_{h_{n,T},T}(x) = O_p(\upsilon(T))$ if $\upsilon(T)h^{2/r}_{n,T} \to \infty$. Now, if $\upsilon(T)h^{2/r}_{n,T} \to \infty$, the above conditions become 1") $h_{n,T}\upsilon(T) \to \infty$, 2") $h^5_{n,T}\upsilon(T) \to C$, and 3") $\frac{\Delta_{n,T}\sqrt{\upsilon(T)}}{h^{3/2}_{n,T}} \to 0$.

Notice that, for 1") and 2") to be satisfied (with C possibly equal to zero), $h_{n,T} \sim \upsilon(T)^{-\beta}$ with $\frac{1}{5} \leq \beta < 1$. However, 3") also has to be satisfied. Write

$$\frac{\Delta_{n,T}\sqrt{\upsilon(T)}}{\upsilon(T)^{-\frac{3}{2}\beta}} = \frac{T}{n}\upsilon(T)^{\frac{1}{2}+\frac{3}{2}\beta} = o(1). \tag{A.7}$$

Since $T \succ \upsilon(T)$, where $\succ$ signifies larger in order, a sufficient condition for equation (A.7) to be met is

$$n \succ T^{\frac{3}{2}+\frac{3}{2}\beta}. \tag{A.8}$$

Next, we add measurement error in the spot variance estimates. The first part of Assumption 4.8 (for $R = 1$) reads $\frac{h_{n,T}^{1/2}\upsilon^{1/2}(T)(\log n \log nk)^{1/2}}{\Delta_{n,T}k^{1/2}} \to 0$, which implies

$$\frac{\upsilon^{(1-\beta)}(T)\,(\log n \log nk)}{\Delta_{n,T}^2} \prec k. \tag{A.9}$$

The second part is $\frac{h_{n,T}^{1/2}\upsilon^{1/2}(T)}{\Delta_{n,T}}\phi^{1/2}(\log n)^{1/2} \to 0$, which requires

$$\phi \prec \frac{\Delta_{n,T}^2}{\log n}\upsilon^{\beta-1}(T). \tag{A.10}$$

The third part (due to jumps) reads $\frac{h_{n,T}^{1/2}\upsilon^{1/2}(T)}{\Delta_{n,T}}n\phi \to 0$, which requires

$$\phi \prec \frac{\Delta_{n,T}}{n}\upsilon^{\frac{\beta-1}{2}}(T).$$

Similarly, the first part of Assumption 4.9 (for $R > 1$) reads $\frac{\upsilon^{1/2}(T)(\log n \log nk)^{1/2}}{h_{n,T}^{1/2}\Delta_{n,T}^{1/2}k^{1/2}} \to 0$, which implies

$$\frac{\upsilon^{(1+\beta)}(T)\,(\log n \log nk)}{\Delta_{n,T}} \prec k. \tag{A.11}$$

The second part is $\frac{\upsilon^{1/2}(T)}{h_{n,T}^{1/2}\Delta_{n,T}^{1/2}}\phi^{1/2}(\log n)^{1/2} \to 0$, which requires

$$\phi \prec \frac{\Delta_{n,T}}{\log n}\upsilon^{-\beta-1}(T). \tag{A.12}$$

The third part (due, again, to jumps) reads $\frac{\upsilon^{1/2}(T)}{h_{n,T}^{1/2}\Delta_{n,T}^{1/2}}n\phi \to 0$, which requires

$$\phi \prec \frac{\Delta_{n,T}^{1/2}}{n}\upsilon^{-\frac{\beta+1}{2}}(T).$$

Given Assumption 4.6 the corresponding conditions for consistency in the case $R = 1$ are:

$$k \succ \frac{\log n \log nk}{\Delta_{n,T}^2}, \quad \phi \prec \frac{\Delta_{n,T}^2}{\log n}, \quad \phi \prec \frac{\Delta_{n,T}}{n}. \tag{A.13}$$

In light of Assumption 4.7, if $R > 1$, they are:

$$k \succ \frac{\log n \log nk}{\Delta_{n,T}}\upsilon^{2\beta}(T), \quad \phi \prec \frac{\Delta_{n,T}\upsilon^{-2\beta}(T)}{\log n}, \quad \phi \prec \frac{\Delta_{n,T}^{1/2}\upsilon^{-\beta}(T)}{n}. \tag{A.14}$$

To understand the different orders, let us now set $T = 20$ (years) and $n = 250 \times T$ (daily observations). In all computations, we will assume absence of jump contributions from Assumption 4.6, 4.7, 4.8, and 4.9, something which is reasonable should the jumps be infrequent. Based on the discussion below, adding the corresponding conditions induced by jumps is straightforward and simply leads to more stringent requirements on ϕ and k. For conciseness, we will also neglect the logarithmic terms.

Assume that $v(T) = T$ (the ergodic or strictly stationary case) and consider the orders for consistency first (in addition to Assumption 4.3). For $R > 1$, setting $\beta = \frac{1}{5}$ (the MSE optimal value) and recalling equation (A.14), we obtain $\phi \prec 1.2 \times 10^{-3}$ and $k \succ 829$. Since, for a liquid stock, roughly 23,000 transactions over a 6.5 hour trading day are not uncommon, the orders may be compatible with the use of an interval ϕ equal to a fraction of a day ($\Delta_{n,T} = 1/250 = 0.004$, so that the calculated ϕ corresponds to roughly one third of a day) and a very typical number of high-frequency observations for many stocks.

For $R = 1$, which is needed for drift estimation only, setting again $\beta = \frac{1}{5}$ and using equation (A.13), we obtain $\phi \prec 1.6 \times 10^{-5}$ and $k \succ 62,500$. This pair of conditions is not empirically feasible since over a very small fraction of a day one cannot hope to make use of so many observations. Imposing, however, $n = 26 \times 20$ (amounting to bi-weekly observations over 20 years of data), one obtains $\phi \prec 1.4 \times 10^{-3}$ and $k \succ 676$, thereby implying that high-frequency observations over a fraction of a day can again be used for spot variance estimation at lower (than daily) frequencies.

Next, we turn to the orders for weak convergence, which are necessarily more stringent. Again, we begin with $T = 20$ (years) and $n = 250 \times T$ (daily observations). For $R > 1$, using equations (A.11) and (A.12), we have $\phi \prec 1.1 \times 10^{-4}$ (roughly 1/40 of a day) and $k \succ 9,102$. These conditions are hard to meet in typical situations, but one must remember that they provide orders, rather than actual numbers. In the case of bi-weekly observations ($n = 26 \times T$), we have $\phi \prec 1.06 \times 10^{-3}$ and $k \succ 946$, which is again compatible with high-frequency observations over a fraction of a day for the computation of spot variance. For $R = 1$, using equations (A.9) and (A.10), similar computations show that the orders are consistent with the use of monthly frequencies to estimate the dynamics, and a fraction of a day of high-frequency data, as earlier, for the evaluation of the spot variance estimates. We also notice that the use of daily, bi-weekly or monthly data to estimate the dynamics is compatible with the discretization condition in equation (A.8).

These are back-of-the-envelope calculations based on asymptotic orders and typical data sets. While they should be taken with caution, they are suggestive of the identification potential of a combination of high- and low-frequency data to estimate spot variance in the first stage and return/variance dynamics in the second stage. As stressed above, the inclusion of jumps would naturally lead to more stringent conditions. ■