NONPARAMETRIC STOCHASTIC VOLATILITY*

Federico M. Bandi† Roberto Renò‡

First draft: October 2007
This draft: June 2008

Abstract

Using recent advances in the nonparametric estimation of continuous-time processes under mild statistical assumptions as well as recent developments on nonparametric volatility estimation by virtue of market microstructure noise-contaminated high-frequency asset price data, we provide (i) a theory of spot variance estimation and (ii) functional methods for stochastic volatility modelling. Our methods allow for the joint evaluation of return and volatility dynamics with nonlinear drift and diffusion functions, nonlinear leverage effects, jumps in returns and volatility with possibly state-dependent jump intensities, as well as nonlinear risk-return trade-offs. Our identification approach and asymptotic results apply under weak *recurrence* assumptions and, hence, accommodate the persistence properties of variance in finite samples. Functional estimation of a generalized (i.e., nonlinear) version of the square-root stochastic variance model with jumps in both volatility and returns for the S&P500 index suggests the need for richer variance dynamics than in existing work. We find a linear specification for the variance's diffusive variance to be misspecified (and inferior to a more flexible CEV specification) even when allowing for jumps in the variance dynamics.

Keywords: Spot variance, stochastic volatility, jumps in returns, jumps in volatility, leverage effects, risk-return trade-offs, kernel methods, recurrence, market microstructure noise.

*We thank the participants at the SoFiE Inaugural conference (New York, June 4-6, 2008), the Festschrift in honor of Peter Phillips (Singapore, July 14-15, 2008), and the Far Eastern and South Asian Meetings of the Econometric Society (Singapore, July 16-18, 2008) for discussions. We are especially grateful to Valentina Corradi, Fulvio Corsi, and Cecilia Mancini for useful comments and suggestions. The usual disclaimers apply.

†*Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA. E-mail: federico.bandi@chicagogsb.edu*

‡*Dipartimento di Economia Politica, Università di Siena, Piazza San Francesco 7, 53100, Siena, Italy. E-mail: reno@unisi.it.*

1 Introduction

Understanding volatility is of fundamental importance 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 (for a review, Shephard, 2005, 2006). 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 effectively treat daily volatility (estimated by aggregating squared intra-daily returns) as an "observable" quantity, without need for filtering on the basis of low-frequency return data (for a review, Andersen et al., 2004). This second body of work has seldom investigated the implications of high-frequency variance estimation for stochastic volatility modelling. The *parametric* approaches of Barndorff-Nielsen and Shephard (2002), Bollerslev and Zhao (2002), Corradi and Distaso (2006), and Todorov (2007), however, are important exceptions and very promising contributions in this area.

We further bridge the gap between arguably the two main strands of the current literature on financial markets volatility by providing functional inferential methods. Specifically, we study *nonparametric* stochastic volatility modelling in continuous time using high-frequency asset price data for the purpose of spot volatility estimation.

Write continuously-compounded returns as $r_{t,t+1} = \log(p_{t+1}) - \log(p_t)$ and consider the system:

$$r_{t,t+dt} = d\log(p_t) = \mu(\sigma_t^2)dt + \sigma_t dW_t^r + dJ_t^r, \tag{1}$$

$$df(\sigma_t^2) = m_{f(.)}(\sigma_t^2)dt + \Lambda_{f(.)}(\sigma_t^2)dW_t^\sigma + dJ_t^\sigma, \tag{2}$$

where $\{W_t^r, W_t^\sigma\}$ are possibly correlated Brownian motions, $\{J_t^r, J_t^\sigma\}$ are Poisson jump processes independent of each other and independent of $\{W_t^r, W_t^\sigma\}$ with intensities $\lambda^r(.)$ and $\lambda^\sigma_{f(.)}(.)$, and $\mu(.)$, $m_{f(.)}(.)$, and $\Lambda_{f(.)}(.)$ are generic functions satisfying smoothness conditions laid out in the following sections.

Our procedures have three main features. First, we filter *spot variance* by localizing (in time) high-frequency estimates of integrated variance $\int \sigma_s^2 ds$. We then use spot variance to identify the parameters and functions driving variance dynamics (i.e., $\lambda^\sigma_{f(.)}(.)$, $m_{f(.)}(.)$, $\Lambda_{f(.)}(.)$ and, given parametric assumptions on the jump size distribution, the moments of the volatility jumps). Since the classical realized variance estimator (i.e., the sum of squared intra-daily returns over the day) may contain substantial contaminations due to market microstructure noise (as emphasized by Bandi and Russell, 2008, and Zhang at al., 2005, in recent work), we employ robust (to noise) integrated variance estimates. In other words, when possible, we allow for market microstructure noise and control for it.[1] Second, differently from much existing work on stochastic volatility modelling, we avoid imposing tight (possibly affine) parametric structures on $\lambda^\sigma_{f(.)}(.)$, $m_{f(.)}(.)$, and $\Lambda_{f(.)}(.)$. Specifically, we identify the relevant functions (through estimates of the system's infinitesimal moments) using nonparametric kernel methods for diffusion and jump-diffusion processes as proposed by Bandi and Nguyen (2003), Bandi and Phillips (2003), and Johannes (2004) in simpler frameworks, namely in the context of scalar models with observables. In order to lay out the main ideas in the context of a well-understood estimation problem, we use classical Nadaraya-Watson kernel estimates. However, as we illustrate below, extensions to alternative functional estimation methods are rather

[1] For recent surveys of nonparametric methods for integrated variance estimation using market microstructure noise-contaminated high-frequency asset price data, we refer the reader to the review papers by Bandi and Russell (2007), Barndorff-Nielsen and Shephard (2007), and McAleer and Medeiros (2008).

straightforward given our procedures. Third, identification does not require stationarity. Rather, it relies on recurrence, which is known to be a milder assumption than stationarity and mixing (see Bandi and Phillips, 2004, for a review of identification methods for recurrent continuous-time processes). In light of the persistent behavior of daily volatility series, methods which only hinge on recurrence and do not rely on the information contained in a potentially inaccurately estimated (in finite samples) stationary density may, arguably, be particularly suitable for our problem.

We present preliminary ideas in the no jump case ($dJ_t^r = 0$, $dJ_t^\sigma = 0$) - Section 4. We then consider the empirically-important case of jumps in volatility ($dJ_t^r = 0$, $dJ_t^\sigma \neq 0$) - Section 5. For clarity, two alternative models (and corresponding identification methods) are presented. We first discuss a nonlinear version of the square-root specification with exponential jump sizes of Duffie et al. (2000). Having received important empirical validation in recent studies (see, e.g., Eraker et al., 2003), this is the specification we analyze in our empirical work. We then discuss a nonlinear log-volatility model ($f(\sigma^2) = \log(\sigma^2)$) with Gaussian jump sizes in the spirit of Jacquier et al. (2001). Finally, we consider the case of jumps in both the return and the volatility process ($dJ_t^r \neq 0$, $dJ_t^\sigma \neq 0$) - Section 6. When focusing on the full system (in Section 7) we study nonparametric identification of risk-return trade-offs ($\mu(\sigma_t^2)$) and (possibly nonlinear) leverage effects.

Our empirical work evaluates the S&P500 joint return/variance dynamics. Using intra-daily Spiders data sampled between the beginning of January 1998 and the end of March 2006, we provide further evidence for the need of jumps in both returns and variance. Estimation of a generalized (i.e., nonlinear) version of the square-root stochastic variance model with exponential jumps in variance and Gaussian jumps in returns suggests the need for richer (diffusive) variance dynamics than in existing parametric work. We show that a linear specification for the variance's diffusive variance is likely misspecified (and inferior to a CEV specification) even when allowing for discontinuities in the variance dynamics.

We conclude this Introduction by pointing out that, in independent and concurrent work, Kanaya and Kristensen (2008) have also tackled estimation of stochastic volatility models in the presence of spot volatility filtered nonparametrically by virtue of the functional estimator proposed by Kristensen (2006). The focus of their stimulating work is however somewhat different from ours. They study the impact of the measurement error induced by high-frequency kernel estimates of spot volatility on both nonparametric and parametric estimates of stochastic volatility models. We concentrate on the nonparametric case but allow for market microstructure noise (when handling the first-stage spot volatility estimates) as well as for discontinuities in the volatility and return dynamics. As we discuss below, these differences naturally result in different approaches to spot volatility estimation and nonparametric modelling. In particular, while not being the substantive core of our analysis but only an input for later developments, we view our theory of microstructure noise-robust and jump-robust spot variance estimation (in Appendix A) to be a promising contribution of our approach. An interesting, recent paper related to ours is also that of Comte et al. (2007) who, in the presence of a continuous stochastic volatility local martingale price process, study least-squares functional techniques to identify the drift and diffusion function of their assumed diffusive volatility while providing bounds for the estimators' risk.

We begin with a description of the infinitesimal moment estimators and their logic.

2 The infinitesimal moment estimators

We assume availability of n equi-spaced price observations in the time interval $[0, T]$ with $\Delta_{n,T} = \frac{T}{n}$. We also assume availability of k (not necessarily equi-spaced) price observations in each interval $[i\Delta_{n,T}, i\Delta_{n,T} + \phi_{n,T}]$. The k intra-period observations are employed to evaluate integrated variance ($\widehat{V}_{iT/n}$) over each sub-interval of size $\phi_{n,T}$.

The functions driving the dynamics of diffusion and jump-diffusion models are known to have infinitesimal conditional moment representations which can be exploited for the purpose of nonparametric identification (Bandi and Phillips, 2004, for discussions). We identify the j-th infinitesimal moment of the variance process, i.e.,

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

by virtue of

$$\widehat{\theta}^j(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{iT/n} - x}{h_{n,T}}\right)\left[f(\widetilde{\sigma}^2_{(i+1)T/n}) - f(\widetilde{\sigma}^2_{iT/n})\right]^j}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{iT/n} - x}{h_{n,T}}\right)} \qquad j = 1, ...,$$

where $\widetilde{\sigma}^2_{iT/n} = \frac{\widehat{V}_{iT/n}}{\phi_{n,T}}$ and $\widehat{V}_{iT/n}$ is a consistent estimate of $\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma^2_s ds$ for fixed n, T, and $\phi_{n,T}$.

Renò (2006) provides simulation evidence for the performance of $\widehat{\theta}^j(x)$ with $j = 1, 2$ (i.e., the drift and the diffusion case) while dealing with stochastic volatility models without discontinuities. This paper develops the necessary theory for specifications with and without jumps in the presence of preliminary high-frequency spot variance estimates $\widetilde{\sigma}^2_{iT/n}$.

The kernel function $\mathbf{K}(.)$ and the integrated variance estimates $\widehat{V}_{iT/n}$ satisfy the following properties:

Assumption 1. $\mathbf{K}(.)$ *is a bounded, continuously-differentiable, symmetric, and nonnegative function whose derivative* $\mathbf{K}'(.)$ *is absolutely integrable and bounded, and for which* $\int \mathbf{K}(s)ds = 1$, $\mathbf{K}_1 = \int s^2\mathbf{K}(s)ds < \infty$, *and* $\mathbf{K}_2 = \int \mathbf{K}^2(s)ds < \infty$.

Assumption 2. $\widehat{V}_{iT/n}$ *is such that*

$$\mathbf{E}_{\sigma^2}\left(\phi^{\beta}_{n,T} k^{\alpha}\left(\frac{\widehat{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma^2_s ds}{\phi_{n,T}}\right)\right) \overset{a}{\approx} 0 \tag{4}$$

and

$$\mathbf{V}_{\sigma^2}\left(\phi^{\beta}_{n,T} k^{\alpha}\left(\frac{\widehat{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma^2_s ds}{\phi_{n,T}}\right)\right) \overset{a}{\approx} \left(a\left(\sigma^4_{iT/n}\right)^{\eta} + b\right) \tag{5}$$

with $\alpha \in (0, \frac{1}{2}]$ and $\beta \in [0, 1]$ *given* T *and* n. $\mathbf{E}_{\sigma^2}$ *and* $\mathbf{V}_{\sigma^2}$ *denote expectation and variance conditional on the spot volatility path.* a, b, *and* η *are numbers. The symbol* $\overset{a}{\approx}$ *denotes asymptotic equivalence for a large* k *and a small* $\phi_{n,T}$.

Coherently with Bandi and Nguyen (2003) and Bandi and Phillips (2003), the asymptotics are derived under $T \to \infty$ (*long span*) and $n \to \infty$ with $\Delta_{n,T} \to 0$ (*infill*). We also assume asymptotic increases in the number of observations for every time span of size $\phi_{n,T}$ with $\phi_{n,T}$ vanishing to zero (i.e., $k \to \infty$ with $\phi_{n,T} \to 0$). The relation between T, n, k, and $\phi_{n,T}$ is made precise in the theorems. Assumption 2 deserves some attention. Its meaning is spelled out in Remarks 1 and 2.

Remark 1. In Appendix A we show that the spot variance estimates $\widetilde{\sigma}^2_{iT/n}$ constructed by virtue of (almost) all recently-proposed integrated variance estimators $\widehat{V}_{iT/n}$ have asymptotic variances and biases which may be represented as in Eq. (5) and Eq. (4) (sometimes for a specific - large - number of subsamples or autocovariances). Consider, for instance, the classical realized variance estimator (Andersen et al., 2003, and Barndorff-Nielsen and Shephard, 2002) in the absence of market microstructure noise. In this case, $\alpha = \frac{1}{2}$, $\beta = 0$, $a = 2$, $b = 0$, and $\eta = 1$. When allowing for noise, in the case of the two-scale estimator (Zhang et al., 2005) with a number of subsamples q equal to $\tau k^{2/3}$ (with τ fixed), we show that $\alpha = \frac{1}{6}$, $\beta = 1$, $a = 0$, and $b \neq 0$. If $q = \tau\left(\frac{k}{\phi_{n,T}}\right)^{2/3}$ and $\phi_{n,T} = k^{\theta}$ with $-\frac{1}{2} < \theta < 0$, then $\alpha = \frac{1}{6}$, $\beta = \frac{1}{3}$, $a \neq 0$, $b = 0$, and $\eta = \frac{2}{3}$. Consider now the class of flat-top realized kernels (Barndorff-Nielsen et al., 2006) with a kernel function $g(.)$ satisfying $g'(0) = 0$ and $g'(1) = 0$, for instance. We find that, if the number of autocovariances q is such that $q = \tau\frac{k^{1/2}}{\phi_{n,T}}$ and $\phi_{n,T} = k^{\theta}$ with $-\frac{1}{2} < \theta < 0$, then $\alpha = \frac{1}{4}$, $\beta = \frac{1}{2}$, $a \neq 0$, $b = 0$, and $\eta = \frac{1}{2}$. Explicit expressions for these estimators, as well as derivations, are provided in Appendix A. In particular, the Appendix relates Assumption 2 to a broader class of integrated variance estimators $\widehat{V}_{iT/n}$ recently proposed in the literature while offering details on the form of the relevant parameters α, β, a, b, and η, for each estimator.

Remark 2. (Spot volatility estimation using realized variance.) In the absence of market microstructure noise, for realized variance we have:

$$\mathbf{V}_{\sigma^2}\left(k^{\frac{1}{2}}\left(\frac{\widehat{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right)\right) \overset{a}{\approx} \left(2\left(\frac{\phi_{n,T}\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^4 ds}{\phi_{n,T}^2}\right)\right) \overset{a}{\approx} 2\sigma^4_{iT/n}$$

since $\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^4 ds}{\phi_{n,T}} \overset{a.s.}{\to} \sigma^4_{iT/n}$ as $\phi_{n,T} \to 0$. By the modulus of continuity of Brownian motion, notice that

$$\begin{aligned}
& k^{\frac{1}{2}}\left(\frac{\widehat{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}} + \frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}} - \sigma^2_{iT/n}\right) \\
= \; & k^{\frac{1}{2}}\left(\frac{\widehat{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right) + k^{\frac{1}{2}} o_{a.s.}\left(\sup_{i\frac{T}{n}\le s\le i\frac{T}{n}+\phi_{n,T}}\left|\sigma_s^2 - \sigma^2_{i\Delta_{n,T}}\right|\right) \\
= \; & k^{\frac{1}{2}}\left(\frac{\widehat{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right) + o_{a.s.}\left(k^{\frac{1}{2}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\right) \\
= \; & O_p(1)
\end{aligned}$$

if $k^{\frac{1}{2}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$. Thus, if $k^{\frac{1}{2}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$ with $k \to \infty$ and $\phi_{n,T} \to 0$, then $\frac{\widehat{V}_{iT/n}}{\phi_{n,T}}$ converges in probability to $\sigma^2_{iT/n}$ (at speed $k^{\frac{1}{2}}$). In addition, using classical weak convergence results (see, e.g., Jacod, 1994, and Jacod and Protter, 1998):

$$k^{\frac{1}{2}}\left(\frac{\widehat{V}_{iT/n}}{\phi_{n,T}} - \sigma^2_{iT/n}\right) \underset{k\to\infty,\phi_{n,T}\to 0}{\Rightarrow} MN\left(0, 2\sigma^4_{iT/n}\right), \tag{6}$$

where MN denotes a mixed Gaussian distribution.

Remark 3. (Spot volatility estimation using more general estimators.) Using Remark 1, by the same argument as in Remark 2 above, $\widetilde{\sigma}^2_{iT/n} = \frac{\widehat{V}_{iT/n}}{\phi_{n,T}}$ converges in probability to $\sigma^2_{iT/n}$ (at speed $\phi^{\beta}_{n,T}k^{\alpha}$ if $\phi^{\beta}_{n,T}k^{\alpha} \to \infty$) provided $k^{\alpha}\phi^{\beta}_{n,T}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$. Furthermore, if the distribution of $\widehat{V}_{iT/n}$ is mixed normal (as is the case for virtually all integrated variance estimators studied in the literature thus far), then

$$\phi^{\beta}_{n,T}k^{\alpha}\left(\frac{\widehat{V}_{iT/n}}{\phi_{n,T}} - \sigma^2_{iT/n}\right) \underset{k\to\infty,\phi_{n,T}\to 0}{\Rightarrow} MN\left(0, a\left(\sigma^4_{iT/n}\right)^{\eta} + b\right). \tag{7}$$

In the case of (kernel-based) estimators that are robust to market microstructure noise (such as the two-scale estimator and the class of flat-top realized kernel estimators), the result also requires appropriate (limiting) relations between the number of subsamples/autocovariances, the number of intra-period observations k, and the length $\phi_{n,T}$. Again, Appendix A provides details while specializing the weak convergence result in Eq. (7) to a variety of estimators recently proposed to evaluate integrated variance in the absence as well as in the presence of noise. The Appendix, for instance, shows that, in the presence of noise, the rate of convergence of the spot variance estimates constructed using the two-scale estimator is $k^{1/10}$. Provided the kernel function $g(.)$ satisfies $g'(0) = 0$ and $g'(1) = 0$, the spot variance estimates constructed using realized kernels may converge at rate $k^{1/8}$.

Remark 4. (More on spot volatility estimation.) As emphasized above, the quantity $\widetilde{\sigma}^2_{iT/n} = \frac{\widehat{V}_{iT/n}}{\phi_{n,T}}$ is a spot variance estimator constructed using an integrated variance measure. Alternative spot volatility estimates have been recently proposed, for instance, by Malliavin and Mancino (2008) and Kristensen (2007). Renò (2008) uses the former to identify the functions $m(.)$ and $\Lambda(.)$ in Eq. (2) for the case without jumps in either volatility or returns. When just aggregating squared continuously-compounded returns (i.e., the realized variance case) in the absence of market microstructure noise, there is an important connection between the interesting approach advocated by Kristensen (2007) and the one adopted here for the purpose of evaluating the full return/variance system. Kristensen's estimator uses all of the observations in the sample and smoothes squared continuously-compounded returns locally, i.e.,

$$\widehat{\sigma}^2_{i/n} = \frac{1}{h}\sum_{j=1}^{nk}\mathbf{K}\left(\frac{j-i/n}{h}\right)r^2_j \qquad i = 1, ..., n,$$

where $\mathbf{K}(.)$ is a kernel function (largely) satisfying Assumption 1 and nk is the total number of observations in $[0, T]$ with $T = 1$, for simplicity. If $nkh \to \infty$, $\widehat{\sigma}^2_{i/n}$ converges to the spot variance at i/n with a standard nonparametric speed $\sqrt{nkh}$. Specifically, the weak convergence result

$$\sqrt{nkh}(\widehat{\sigma}^2_{i/n} - \sigma^2_{i/n}) \Rightarrow MN(0, 2\mathbf{K}_2\sigma^4_{i/n}) \tag{8}$$

holds if, in addition, $nkh^{1+2\gamma} \to 0$, where $0 < \gamma \leq 1$ is the order of smoothness of σ^2_t (see Kristensen, 2007, Theorem 2). The latter condition guarantees disappearance of the asymptotic bias term. We now turn to our approach when $\widehat{V}_{i/n}$ is realized variance and noise is absent. Write

$$\frac{\widehat{V}_{i/n}}{\phi_{n,1}} = \frac{\sum_{j=1}^{k} r^2_j}{\phi_{n,1}} = \frac{1}{\phi_{n,1}}\sum_{j=1}^{nk}\mathbf{1}_{\left\{0\leq\frac{j-i/n}{\phi_{n,1}}\leq 1\right\}}r^2_j,$$

where $\mathbf{1}_{\{.\}}$ is the indicator kernel. Hence, $\frac{\widehat{V}_{i/n}}{\phi_{n,1}}$ has an interpretation in terms of kernel smoother. We are simply aggregating (using equal weights) observations in a local neighborhood of i/n, i.e., $[i/n, i/n + \phi_{n,1}]$. Thus, $\phi_{n,1}$ is effectively a bandwidth playing the same role as h in the case of $\widehat{\sigma}^2_{i/n}$. This said, our derived asymptotic distribution in Eq. (6) and the asymptotic distribution in Eq. (8) have to be consistent with each other when $\mathbf{K}(.) = \mathbf{1}_{\{.\}}$. In other words, if $nk\phi_{n,1} \to \infty$ and $nk\phi_{n,1}^{1+2\gamma} \to 0$,

$$\sqrt{nk\phi_{n,1}}\left(\frac{\widehat{V}_{i/n}}{\phi_{n,1}} - \sigma^2_{i/n}\right) \Rightarrow MN\left(0, 2\left(\int \mathbf{1}^2_{\{0\leq s\leq 1\}} ds\right)\sigma^4_{i/n}\right).$$

Now, notice that $\phi_{n,T}$ is defined as an interval containing k observations, i.e., $n\phi_{n,1} = 1$. Thus, the (effective) rate becomes $\sqrt{k}$, which is coherent with Eq. (6). Similarly, the asymptotic variance becomes $2\sigma^4_{i/n}$, which is also consistent with Eq. (6). Finally, the condition for a vanishing asymptotic bias term $nk\phi_{n,1}^{1+2\gamma} = k\phi_{n,1}^{2\gamma} = k^{1/2}\phi_{n,1}^{\gamma} = o(1)$ is equivalent to $k^{\frac{1}{2}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} = o(1)$ for our assumed degree of smoothness of the spot volatility process.

From a theoretical standpoint, the use of smooth kernels, as in Kristensen (2007), yields efficiency gains over the equal weighting implicitly delivered by our methods. In fact, the term $\mathbf{K}_2$ is generally smaller than 1 (it is, for example, equal to $\frac{1}{2\sqrt{\pi}}$ for a second-order Gaussian kernel). From an empirical standpoint, the presence of intraday seasonalities (see, e.g., Andersen and Bollerslev, 1998, and the references therein) might affect estimates based on smooth kernels in ways that are difficult to predict. Diurnal effects appear more likely to average out when using equal weighting over a trading day as implied by integrated variance-type measures, such as realized variance.

Importantly, the properties of (either smooth or discountinuous) kernel estimates of spot variance have not been previously studied for the cases with jumps and market microstructure noise, which are relevant for our purposes. We do so in Appendix A. Using equal weighting allows us to draw from the recent literature on integrated variance estimation both in terms of asymptotic results and in terms of finite sample adjustments required for more accurate empirical implementations. The latter have been advocated by Bandi and Russell (2006, 2008).[2] We leave the study of smooth kernels for spot variance estimation in the presence of noise and return jumps for future work.

Intuition. We now turn to the logic behind our estimation procedure. Given Remark 2 and 3, the rate of convergence of $\widetilde{\sigma}^2_{iT/n}$ to $\sigma^2_{iT/n}$ is $k^{\alpha}\phi^{\beta}_{n,T}$ (if $k^{\alpha}\phi^{\beta}_{n,T} \to \infty$, of course, and $k^{\alpha}\phi^{\beta}_{n,T}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$), where $\alpha \in (0, \frac{1}{2}]$ and $\beta = [0, 1]$. Loosely speaking, if $k \to \infty$ at a fast enough pace as $h_{n,T} \to 0$, then one may hope to control (asymptotically) the estimation error induced by the preliminary spot variance estimates so that $\widehat{\theta}^j(x)$ identifies $\theta^j(x)$ consistently (in probability). A set of conditions that are sufficient for this to happen (and for the moments to have well-defined limiting distributions) is listed in the theorems below. Identification of all the functions (and parameters of interest) of the model in Eq. (1) and Eq. (2) will rely on consistent estimation of $\widehat{\theta}^j(x)$, for $j = 1, ...,$ as we discuss in Section 5 below.

This paper presents the main ideas in the context of classical Nadaraya-Watson kernel estimates. Extensions to functional estimates with improved asymptotic and finite sample properties are rather immediate given our

[2] For the case with no jumps in returns, the working paper version of this paper (Bandi and Reno', 2008) estimates the S&P 500 index's spot variance using the two-scale estimator and appropriate realized kernels. In both cases, the finite-sample properties of the estimators are optimized by minimizing the estimators' MSEs (under microstructure noise) as suggested by Bandi and Russell (2008). We refer the interested reader to that version of the paper for details.

procedures and are, in some cases, discussed below (see, e.g., Remark 10). Among other methods, $\widehat{\theta}^j(x)$ could be a local linear estimator of the form

$$\widehat{\theta}^j_{ll}(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \widetilde{\mathbf{K}}_i(x, h_{n,T}) \left[f(\widetilde{\sigma}^2_{(i+1)T/n}) - f(\widetilde{\sigma}^2_{iT/n}) \right]^j}{\sum_{i=1}^{n} \widetilde{\mathbf{K}}_i(x, h_{n,T})} \qquad j = 1, ..., \tag{9}$$

where $\widetilde{\mathbf{K}}_i(x, h_{n,T}) = \frac{1}{h_{n,T}} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{iT/n} - x}{h_{n,T}} \right) \Gamma_{n,2} - (\widetilde{\sigma}^2_{iT/n} - x) \frac{1}{h_{n,T}} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{iT/n} - x}{h_{n,T}} \right) \Gamma_{n,1}$ with

$$\Gamma_{n,s} = \sum_{i=1}^{n} (\widetilde{\sigma}^2_{iT/n} - x)^s \frac{1}{h_{n,T}} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{iT/n} - x}{h_{n,T}} \right)$$

for $s = 1, 2$. More generally, it could be a local polynomial estimator defined as the solution $\{\alpha_0, \alpha_1, ..., \alpha_p\}$ to

$$\sum_{i=1}^{n} \left(\frac{1}{\Delta_{n,T}} (f(\widetilde{\sigma}^2_{(i+1)T/n}) - f(\widetilde{\sigma}^2_{iT/n}))^j - \sum_{u=0}^{p} \alpha_u (\widetilde{\sigma}^2_{iT/n} - x)^u \right) \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{iT/n} - x}{h_{n,T}} \right) \qquad j = 1, ...,$$

where $\widehat{\theta}^j_{ll}(x) = \widehat{\alpha}_0(x)$ for $p = 1$. Local polynomial methods for diffusions are studied by Fan and Zhang (2003) and, under recurrence, by Moloche (2002). Alternative, interesting approaches for diffusion estimation under a variety of assumptions, including weak recurrence assumptions, have been recently proposed, *inter alia*, by Kristensen (2008) and Xu (2006, 2007).

3 Recurrence

Consider a complete probability space $(\Omega, \Im, P, \{\Im_t\}_{t \geq 0})$ and the compensated N-dimensional jump-diffusion process X_t defined as

$$X_t = X_0 + \int_0^t \boldsymbol{\mu}(X_{s-}) ds + \int_0^t \boldsymbol{\sigma}(X_{s-}) dW_s + \int_0^t \int \mathbf{c}(X_{s-}, y) \overline{v}(ds, dy),$$

where $\{W_t, \Im_t\}$ is a standard m-dimensional Brownian motion and

$$\begin{aligned} \overline{v}(dt, dy) &= N(dt, dy) - \mathbf{E}(N(dt, dy)) \\ &= N(dt, dy) - \widetilde{\Pi}(dy) dt \end{aligned}$$

is a compensated Poisson random measure on $[0, \infty) \times R^N$ independent of W_t.

Assumption 3. *The terms $\boldsymbol{\mu}(.)$, $\boldsymbol{\sigma}(.)$, and $\mathbf{c}(., y)$ are at least twice continuously-differentiable vector functions of the Markov state. $\boldsymbol{\mu}(.) = \{\mu_i(.)\}_{1 \leq i \leq N}$, and $\mathbf{c}(., y) = \{c_i(., y)\}_{1 \leq i \leq N}$, are $N \times 1$ Borel measurable vectors, and $\boldsymbol{\sigma}(.) = \{\sigma_{ij}(.)\}_{\substack{1 \leq i \leq N \\ 1 \leq j \leq m}}$ is a $N \times m$ Borel measurable matrix. There exists a constant C such that, for any $x, z \in R^N$,*

$$\begin{aligned} |\boldsymbol{\mu}(x) - \boldsymbol{\mu}(z)|^2 + ||\boldsymbol{\sigma}(x) - \boldsymbol{\sigma}(z)||^2 + \int |\mathbf{c}(x, y) - \mathbf{c}(z, y)|^2 \widetilde{\Pi}(dy) &\leq C|x - z|^2, \\ |\boldsymbol{\mu}(x)|^2 + ||\boldsymbol{\sigma}(x)||^2 + \int |\mathbf{c}(x, y)|^2 \widetilde{\Pi}(dy) &\leq C(x - |z|)^2. \end{aligned}$$

Write $\mathbf{a}(x) = \boldsymbol{\sigma}(x)\boldsymbol{\sigma}(x)^\intercal$. *In addition, there exists a number* $\alpha > 0$ *so that*

$$z^\intercal \mathbf{a}(x) z \geq \alpha |z|^2 \text{ for all } x \text{ and } z.$$

Assumption 3 guarantees the existence of a nondegenerate strong solution X_t.

Assumption 4. *For each* $x \in R^N$,

$$\sup_{z \neq 0} \int \left(\ln \frac{|z + c(z+x, y)|}{|y|} \right)^2 \widetilde{\Pi}(dy) = C_x.$$

If Assumption 3 and 4 are satisfied and, for $r > 0$, there exists $\varepsilon_1 = \varepsilon_1(r) > 0$ and $\eta_1 = \eta_1(r) > 0$ so that, for any $|z| \geq r$ and $x \in R^N$,

$$\frac{\sum z_i \mu_i(z+x)}{|z|^2} - \frac{\sum a_{ij}(z+x) z_i z_j}{2|z|^2} + \frac{\sum a_{ii}(z+x)}{2|z|^2} +$$
$$+ \int \left(\ln \frac{|z + \mathbf{c}(z+x, y)|}{|y|} - \frac{z^\intercal \mathbf{c}(z+x, y)}{|y|^2} \right) \widetilde{\Pi}(dy) < (1 - \varepsilon_1) \frac{\sum a_{ij}(z+x) z_i z_j}{2|z|^4} - \eta_1,$$

then the process X_t is *recurrent* (In-Suk Wee, 2000).

Remark 5. The model in Eq. (1) and Eq. (2) is not compensated. This is of course not problematic since we could compensate it and redefine the drift vector as being equal to $\boldsymbol{\mu}(.) = \{\mu(.), m_{f(.)}(.)\}^\intercal - \{\lambda^r(.), \lambda^\sigma_{f(.)}(.)\}^\intercal \circ \{\mathbf{E}[(c^r(., y^r)], \mathbf{E}[(c^\sigma(., y^\sigma)]\}^\intercal$, where $\circ$ denotes element-by-element multiplication. The conditions in Assumption 4 would therefore have to apply to the system with a re-defined drift term.

Under recurrence, for any $x \in R^N$ and $r > 0$,

$$P_x(|X_t - x| < r \text{ for a sequence of times increasing to } \infty) = 1.$$

In other words, the process returns to open sets in its range an infinite number of times over time, thereby making consistent point-wise kernel estimation possible even in the absence of a time-invariant stationary density. Recurrent processes for which a stationary density exists converge to it and are called *positive recurrent* (or *ergodic*). They are called *strictly stationary* when started at the stationary density. Recurrent processes which are not endowed with a stationary density are called *null recurrent*. See, e.g., Bandi and Phillips (2004) for discussions.

Importantly for our purposes, while it is of course hard to argue against the stationarity properties of return and variance series, the persistence features of variance should be a concern when identification is conducted by heavily relying on the informational content of the variance process' stationary density. The use of identification methods which do not hinge on stationarity is expected to lead to less distorted estimates in regions where the variance's stationary density cannot be estimated reliably and, as we will show below, a somewhat more objective representation of statistical uncertainty. In effect, the size of the point-wise (asymptotic) confidence bands of each infinitesimal moment estimate will be shown to be an inverse function of the number of visits made by the spot variance process in the local neighborhood of each spatial point, i.e., the local time of the process.

Here we provide conditions for recurrence only in the case of our most general system with jumps. When specializing to individual equations (either variance or returns) and/or when considering the benchmark framework without jumps, we refer the reader to the conditions for multivariate diffusion processes in Hasminskii (1960) and Bhattacharya (1978).

4 A preliminary case: $\mathbf{dJ}_t^r = 0$ and $\mathbf{dJ}_t^\sigma = 0$

In the absence of jumps, the estimated infinitesimal moments are known to directly identify the functions of interest since $\theta^1(x) = m(x)$, $\theta^2(x) = \Lambda^2(x)$, and $\theta^j(x) = 0 \; \forall j \geq 3$. Theorems 2 and 3 below present conditions on $T, n, k, \phi_{n,T}$, and the bandwidth $h_{n,T}$, which guarantee $\widehat{\theta}^1(x) \xrightarrow{p} m(x)$ and $\widehat{\theta}^2(x) \xrightarrow{p} \Lambda^2(x)$ while yielding asymptotic Gaussian distributions. We begin with the limiting properties of the averaged kernel function.

Theorem 1 (Convergence to the chronological local time.) *Assume T is fixed ($T = \overline{T}$). If $k, n \to \infty$ and $h_{n,\overline{T}}, \phi_{n,\overline{T}} \to 0$ so that*

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{h_{n,\overline{T}}} \left(\Delta_{n,\overline{T}} \log \frac{1}{\Delta_{n,\overline{T}}} \right)^{1/2} &= 0, \\ \lim_{k,n\to\infty} \frac{1}{h_{n,\overline{T}} k^\alpha \phi_{n,\overline{T}}^\beta} + \frac{1}{h_{n,\overline{T}}} \left(\phi_{n,\overline{T}} \log \left(\frac{1}{\phi_{n,\overline{T}}} \right) \right)^{1/2} &= 0, \end{aligned} \tag{10}$$

with $\alpha \in (0, \frac{1}{2}]$ and $\beta = [0, 1]$, then,

$$\widehat{\overline{L}}_{\sigma^2}(\overline{T}, x) = \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=1}^n \mathbf{K} \left(\frac{\widetilde{\sigma}^2_{i\overline{T}/n} - x}{h_{n,\overline{T}}} \right) \xrightarrow{p} \overline{L}_{\sigma^2}(\overline{T}, x),$$

where $\overline{L}_{\sigma^2}(\overline{T}, x)$ is the chronological local time of the spot variance process.

Proof. *See Appendix B.*

Remark 6. In functional estimation methods for recurrent continuous-time semimartingales, chronological local time (i.e., the time spent by the process in the vicinity of a point) drives the rate of convergence of the functional estimates (see, e.g., Theorem 3 and 4, below). Since recurrent processes visit each open neighborhood of a point infinitely often over time (Section 3), then local time diverges with T. The divergence rate is linear (in T) for positive recurrent (ergodic) or stationary processes (since $\overline{L}_{\sigma^2}(T, x)/T \xrightarrow{p} p(x)$, where $p(x)$ is the time-invariant stationary density at x) but is lower for null recurrent processes and, importantly, unknown in general. One important exception in the null recurrent class is Brownian motion for which $v(T) = T^{1/2}$. In what follows, we write $\overline{L}_{\sigma^2}(T, x) \propto v(T)$, where $v(T)$ is a regularly-varying function at infinity (see, e.g., Bandi and Moloche, 2004, for discussions).

Remark 7. In practise, the nature of the divergence properties of local time is immaterial for our purposes. All we will need, in order to express the (effective) rate of convergence of our functional estimates (and, of course, their limiting variance), is an in-sample characterization of the local time factor. As Theorem 1 implies, one can do so by using kernel methods similar to those employed for estimating classical stationary densities.

Theorem 2 (The volatility drift.) *If $k, n, T \to \infty$ and $h_{n,T}, \phi_{n,T} \to 0$ so that*

$$\lim_{n,T\to\infty} h_{n,T} v(T) = \infty, \tag{11}$$

$$\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}} \left(\Delta_{n,T} \log \frac{1}{\Delta_{n,T}} \right)^{1/2} = 0, \tag{12}$$

$$\lim_{k,n,T\to\infty} \frac{T v(T)^{-1}}{\Delta_{n,T} h_{n,T} k^\alpha \phi_{n,T}^\beta} + \frac{T v(T)^{-1}}{\Delta_{n,T} h_{n,T}} \left(\phi_{n,T} \log \left(\frac{1}{\phi_{n,T}} \right) \right)^{1/2} = 0, \tag{13}$$

with $\alpha \in (0, \frac{1}{2}]$ *and* $\beta = [0,1]$ *, then,*

$$\widehat{\theta}^1(x) \xrightarrow{p} m(x),$$

where $\overline{L}_{\sigma^2}(T,x) \propto v(T)$*. If*

$$\begin{aligned}
\lim_{n,T\to\infty} h_{n,T} v(T) &= \infty, \\
\lim_{n,T\to\infty} h_{n,T}^5 v(T) &= C_1, \qquad (14)\\
\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}} \left(\Delta_{n,T} \log \frac{1}{\Delta_{n,T}}\right)^{1/2} &= 0, \\
\lim_{k,n,T\to\infty} \frac{T v(T)^{-1/2}}{\Delta_{n,T} h_{n,T}^{1/2} k^{\alpha} \phi_{n,T}^{\beta}} + \frac{T v(T)^{-1/2}}{\Delta_{n,T} h_{n,T}^{1/2}} \left(\phi_{n,T} \log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} &= 0, \qquad (15)
\end{aligned}$$

then,

$$\sqrt{h_{n,T} \widehat{\overline{L}}_{\sigma^2}(T,x)} \left\{\widehat{\theta}^1(x) - m(x) - \Gamma_m(x)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 \Lambda^2(x)\right),$$

with

$$\Gamma_m(x) = h_{n,T}^2 \mathbf{K}_1 \left[m^{'}(x) \frac{s^{'}(x)}{s(x)} + \frac{1}{2} m^{''}(x)\right],$$

where $s(dx)$ *is the diffusion's speed measure and* C_1 *is a constant.*

Proof. *See Appendix B.*

Theorem 3 (The volatility diffusion.) *If* $k, n, T \to \infty$ *and* $h_{n,T}, \phi_{n,T} \to 0$ *so that*

$$\begin{aligned}
\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}} \left(\Delta_{n,T} \log \frac{1}{\Delta_{n,T}}\right)^{1/2} &= 0, \\
\lim_{k,n,T\to\infty} \frac{T v(T)^{-1}}{\Delta_{n,T} h_{n,T} k^{\alpha} \phi_{n,T}^{\beta}} + \frac{T v(T)^{-1}}{\Delta_{n,T} h_{n,T}} \left(\phi_{n,T} \log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} &= 0,
\end{aligned}$$

with $\alpha \in (0, \frac{1}{2}]$ *and* $\beta = [0,1]$ *, then,*

$$\widehat{\theta}^2(x) \xrightarrow{p} \Lambda^2(x),$$

where $\overline{L}_{\sigma^2}(T,x) \propto v(T)$*. If*

$$\begin{aligned}
\lim_{n,T\to\infty} \frac{h_{n,T}^5 v(T)}{\Delta_{n,T}} &= C_2, \\
\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}} \left(\Delta_{n,T} \log \frac{1}{\Delta_{n,T}}\right)^{1/2} &= 0, \\
\lim_{k,n,T\to\infty} \frac{T v(T)^{-1/2}}{\Delta_{n,T}^{3/2} h_{n,T}^{1/2} k^{\alpha} \phi_{n,T}^{\beta}} + \frac{T v(T)^{-1/2}}{\Delta_{n,T}^{3/2} h_{n,T}^{1/2}} \left(\phi_{n,T} \log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} &= 0,
\end{aligned}$$

then,

$$\sqrt{\frac{h_{n,T}\widehat{\widehat{L}}_{\sigma^2}(T,x)}{\Delta_{n,T}}}\left\{\widehat{\theta}^2(x)-\Lambda^2(x)-\Gamma_\Lambda(x)\right\}\Rightarrow \mathbf{N}\left(0,2\mathbf{K}_2\Lambda^4(x)\right),$$

with

$$\Gamma_\Lambda(x)=h_{n,T}^2\mathbf{K}_1\left[\Lambda^{2'}(x)\frac{s^{'}(x)}{s(x)}+\frac{1}{2}\Lambda^{2''}(x)\right],$$

where $s(dx)$ is the diffusion's speed measure and C_2 is a constant.

Proof. *See Appendix B.*

Remark 8 (The bandwidth conditions) In order to discuss the meaning of the bandwidth conditions, we simply focus on the drift estimator. Similar observations can be made in the diffusion case and will apply to the more general infinitesimal moment estimators in the next section. Eq. (11) is analogous to the classical condition $nh\to\infty$, which is necessary for consistency of the Nadaraya-Watson kernel estimator in discrete-time. Here the local time's divergence rate ($v(T)$) replaces the divergence rate of the number of observations (n). Eq. (12) hinges on the Brownian modulus of continuity and guarantees that the continuous sample path of the process can be approximated (in the limit) by virtue of discretely-sampled data. This condition is important to replicate the "infinitesimal" features of the moment(s), as implied by Eq. (3). Eq. (12) is the condition which allows us to eliminate (asymptotically) the measurement error which is, necessarily, induced by the preliminary spot variance estimates. This condition ought to be slightly strengthened (in Eq. (15)) for weak convergence to hold. When deriving weak convergence, the additional condition in Eq. (14) guarantees optimality of $h_{n,T}$ by appropriately balancing the estimator's asymptotic bias and variance. It is analogous to $nh^5=O(1)$ in the case of classical Nadaraya-Watson kernel estimators in discrete-time.

Remark 9 Since $\Delta_{n,T}\to 0$, the diffusion estimator has a faster rate of convergence than the drift estimator. In both cases, optimal rate selection for the smoothing parameter $h_{n,T}$ yields an asymptotic bias term which has a familiar form (from more conventional kernel estimation in discrete time) but, in light of the mildness of our assumptions, depends on the process' invariant (speed) measure rather on the process' time-invariant stationary density, which is not assumed to exist. The drift's optimal bandwidth rate is $\left(\frac{1}{v(T)}\right)^{1/5}$. The corresponding diffusion's value is $\left(\frac{\Delta_{n,T}}{v(T)}\right)^{1/5}$.

Remark 10 (Local linear estimates) As pointed out in the Introduction and in Section 2, while we illustrate the main issues by virtue of traditional Nadaraya-Watson estimates, extensions to kernel estimators with superior asymptotic mean-squared error properties can be conducted similarly. Consider, for example, local linear estimates of the drift and diffusion as in Eq. (9). All bandwidth conditions would be preserved. In fact, the statements of Theorems 2 and 3 would remain unchanged with the exception of intuitive (given existing work in discrete time) modifications of the asymptotic biases: $\Gamma_m(x)$ would become $h_{n,T}^2\mathbf{K}_1\left(\frac{1}{2}m^{''}(x)\right)$ and $\Gamma_\Lambda(x)$ would become $h_{n,T}^2\mathbf{K}_1\left(\frac{1}{2}\Lambda^{2''}(x)\right)$.

5 Jumps in volatility: $\mathbf{dJ}_t^r = 0$ and $\mathbf{dJ}_t^\sigma \neq 0$

Recent empirical work has emphasized the importance of models allowing for rapid increases in stock returns' conditional volatility (see, e.g., Bates, 2000, Duffie et al., 2000, Pan, 2002, and 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 means by 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 other contexts, by Johannes (2004) and studied formally by Bandi and Nguyen (2003).[3] To clarify ideas, we consider nonlinear versions of two stochastic volatility models which have drawn particular attention in recent years, namely the square-root stochastic volatility model with exponential jumps of Duffie et al. (2000) and a log-volatility model with Gaussian jumps in the spirit of Jacquier et al. (2002). Alternative specifications may of course be easily adopted provided the identification scheme is modified accordingly.

Generalized Duffie, Pan, and Singleton (2000) model. Write Eq. (2) with $f(\sigma_t^2) = \sigma_t^2$ and $dJ_t^\sigma = \xi^\sigma dN_t^\sigma$, where $\xi^\sigma \sim \exp(\mu_\xi)$. In Duffie et al. (2000) and Eraker et al. (2003), $m_{\sigma^2}(\sigma_t^2)$ is affine (i.e., linear in σ_t^2), $\Lambda_{\sigma^2}(\sigma_t^2)$ 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 $\lambda_{\sigma^2}(\sigma_t^2) = 0$[4]). Provided the variance drift, diffusion, and intensity satisfy the conditions laid out in Section 3, we leave their functional forms unspecified. Now, notice that

$$\begin{aligned}
\theta^1(x) &= m_{\sigma^2}(x) + \mu_\xi \lambda_{\sigma^2}(x) && (16)\\
\theta^2(x) &= \Lambda_{\sigma^2}^2(x) + 2\mu_\xi^2 \lambda_{\sigma^2}(x) && (17)\\
\theta^3(x) &= 6\mu_\xi^3 \lambda_{\sigma^2}(x) && (18)\\
\theta^4(x) &= 24\mu_\xi^4 \lambda_{\sigma^2}(x) && (19)\\
& \ldots
\end{aligned}$$

Hence, consistent (in probability) identification of the relevant functions may be conducted by computing:

$$\begin{aligned}
\widehat{\mu}_\xi &= \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{\theta}^4(\widetilde{\sigma}_{iT/n}^2)}{4\widehat{\theta}^3(\widetilde{\sigma}_{iT/n}^2)}, && (20)\\
\widehat{\lambda}_{\sigma^2}(x) &= \frac{\widehat{\theta}^4(x)}{24\widehat{\mu}_\xi^4}, && (21)\\
\widehat{\Lambda}_{\sigma^2}^2(x) &= \widehat{\theta}^2(x) - 2\widehat{\mu}_\xi^2 \widehat{\lambda}_{\sigma^2}(x), && (22)\\
\widehat{m}_{\sigma^2}(x) &= \widehat{\theta}^1(x) - \widehat{\mu}_\xi \widehat{\lambda}_{\sigma^2}(x). && (23)
\end{aligned}$$

[3] A different methodology based on pre-filtering the data with a threshold function is explored in Mancini and Renò (2006).

[4] Other papers allowing for jumps in returns, stochastic volatility, but no jumps in volatility are, for example, Bakshi et al. (1997), Bates (2000), and Pan (2002). These papers find evidence for misspecification in the volatility dynamics pointing to the likely presence of discontinuities in the volatility sample path.

Alternative (possibly superior) identification methods can of course be employed. One could, for instance, consider higher-order approximations.[5] Here we lay out the main ideas by considering the most intuitive identification scheme. Below, we show empirically, and by simulation, that this approach may perform very satisfactorily in practise.

Log-variance. Write Eq. (2) with $f(\sigma_t^2) = \log(\sigma_t^2)$ and $dJ_t^\sigma = \xi^\sigma dN_t^\sigma$, where $\xi^\sigma \sim \mathbf{N}(0, \sigma_\xi^2)$. This model is in the spirit of Jacquier et al. (2002), among others. As earlier, we generalize it by allowing for a nonlinear drift, diffusion, and intensity of the jumps. Write

$$\begin{aligned}
\theta^1(x) &= m_{\log\sigma^2}(x), && (24)\\
\theta^2(x) &= \Lambda^2_{\log\sigma^2}(x) + \sigma_\xi^2 \lambda_{\log\sigma^2}(x), && (25)\\
\theta^4(x) &= 3\sigma_\xi^4 \lambda_{\log\sigma^2}(x), && (26)\\
\theta^6(x) &= 15\sigma_\xi^6 \lambda_{\log\sigma^2}(x), && (27)\\
& \ldots
\end{aligned}$$

A potential identification method (Bandi and Nguyen, 2003, and Johannes, 2004) is now:

$$\begin{aligned}
\widehat{\sigma}_\xi^2 &= \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{\theta}^6(\widetilde{\sigma}^2_{iT/n})}{5\widehat{\theta}^4(\widetilde{\sigma}^2_{iT/n})}, && (28)\\
\widehat{\lambda}_{\log\sigma^2}(x) &= \frac{\widehat{\theta}^4(x)}{3\widehat{\sigma}_\xi^4}, && (29)\\
\widehat{\Lambda}^2_{\log\sigma^2}(x) &= \widehat{\theta}^2(x) - \widehat{\sigma}_\xi^2 \widehat{\lambda}_{\log\sigma^2}(x), && (30)\\
\widehat{m}_{\log\sigma^2}(x) &= \widehat{\theta}^1(x). && (31)
\end{aligned}$$

This identification procedure has proved successful in the analysis of the temporal dynamics of spot interest rate series in continuous time (Johannes, 2004). As discussed earlier, it can be generalized when needed (see, e.g., Section 9, Eq. (36) and Eq. (37)).

Using linear specifications with no jumps in either returns or variance, Andersen et al. (2002) and Chernov et al. (2002) find that the log-volatility and the square-root model provide very similar fit to the data. In light of the recent empirical validation provided by Eraker et al. (2003) to the affine square-root model with jumps in both volatility and returns, a nonlinear version of this model will be the subject of our empirical work.

Theorem 4 presents conditions on $T, n, k, \phi_{n,T}$, and the bandwidth $h_{n,T}$, guaranteeing $\widehat{\theta}^j(x) \xrightarrow{p} \theta^j(x)$ for all j, and by an application of Slutsky's theorem, consistency (in probability) of the relevant functions and jump size moments. Our discussion in the previous section provides intuition for several aspects of the Theorem. Below, we focus on what is specific to the case with discontinuous jumps.

Theorem 4. (The infinitesimal moments.) *If $k, n, T \to \infty$ and $h_{n,T}, \phi_{n,T} \to 0$ so that*

[5] We do so when evaluating the return dynamics in order to achieve more accurate finite sample performance (see Section 10).

$$\begin{aligned}
\lim_{n,T\to\infty} h_{n,T}v(T) &= \infty, \\
\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}}\left(\Delta_{n,T}\log\frac{1}{\Delta_{n,T}}\right)^{1/2} &= 0, \\
\lim_{k,n,T\to\infty} \frac{Tv(T)^{-1}}{\Delta_{n,T}h_{n,T}k^{\alpha}\phi_{n,T}^{\beta}} + \frac{Tv(T)^{-1}}{\Delta_{n,T}h_{n,T}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} &= 0,
\end{aligned}$$

with $\alpha \in (0, \frac{1}{2}]$ *and* $\beta = [0,1]$, *then,*

$$\widehat{\theta}^{j}(x) \xrightarrow{p} \theta^{j}(x) \qquad j \geq 1,$$

where $\overline{L}_{\sigma^2}(T,x) \propto v(T)$. *If*

$$\begin{aligned}
\lim_{n,T\to\infty} h_{n,T}v(T) &= \infty, \\
\lim_{n,T\to\infty} h_{n,T}^{5}v(T) &= C_3, \\
\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}}\left(\Delta_{n,T}\log\frac{1}{\Delta_{n,T}}\right)^{1/2} &= 0, \\
\lim_{k,n,T\to\infty} \frac{Tv(T)^{-1/2}}{\Delta_{n,T}h_{n,T}^{1/2}k^{\alpha}\phi_{n,T}^{\beta}} + \frac{Tv(T)^{-1/2}}{\Delta_{n,T}h_{n,T}^{1/2}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} &= 0,
\end{aligned}$$

then,

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

with

$$\Gamma_{\theta^{j}}(x) = h_{n,T}^{2}\mathbf{K}_1\left[\theta^{j'}(x)\frac{s^{'}(x)}{s(x)} + \frac{1}{2}\theta^{j''}(x)\right],$$

where $s(dx)$ *is the process' invariant measure and* C_3 *is a constant.*
Proof. *See Appendix B.*

Remark 11. Contrary to the no jump case, all infinitesimal moments converge at the same rate. In particular, an enlarging span of data ($T \to \infty$) is necessary to guarantee $h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x) \overset{a.s.}{\to} \infty$ and, hence, consistency of all moments. As earlier, selection of the optimal bandwidth rate $\left(\left(\frac{1}{v(T)}\right)^{1/5}\right)$ yields an asymptotic bias term which depends on the process' invariant measure and may be eliminated by slight undersmoothing.

5.1 The implied drift, diffusion, intensity of the jumps, and jump size: asymptotic properties

For both models presented earlier, we now discuss asymptotic inference on the functions and parameters of interest. In all cases, the bandwidth $h_{n,T}$ is set so as to avoid the presence of an asymptotic bias term (as implied by the condition in Eq. (32)).

We initially assume that the moments of the jump sizes are estimated by averaging higher-order infinitesimal moments over a fixed time period $\overline{T}$. In other words, $\widehat{\mu}_\xi = \frac{1}{\overline{n}} \sum_{i=1}^{\overline{n}} \frac{\widehat{\theta}^4(\widetilde{\sigma}^2_{i\overline{T}/\overline{n}})}{4\widehat{\theta}^3(\widetilde{\sigma}^2_{i\overline{T}/\overline{n}})}$ and $\widehat{\sigma}^2_\xi = \frac{1}{\overline{n}} \sum_{i=1}^{\overline{n}} \frac{\widehat{\theta}^6(\widetilde{\sigma}^2_{i\overline{T}/\overline{n}})}{5\widehat{\theta}^4(\widetilde{\sigma}^2_{i\overline{T}/\overline{n}})}$ with $\overline{T}/\overline{n} \to 0$ as $\overline{n} \to \infty$ over a fixed $\overline{T} < T$. Importantly, for consistency (see Remark 11), the higher-order moments $\widehat{\theta}^j$ used to compute $\widehat{\mu}_\xi$ and $\widehat{\sigma}^2_\xi$ continue to be estimated (before the averaging) over an asymptotically expanding T. As we will show, the fixed $\overline{T}$ case is theoretically interesting when dealing with generic (stationary and nonstationary) recurrent processes. Specifically, it guarantees that the averages $\widehat{\mu}_\xi$ and $\widehat{\sigma}^2_\xi$ are asymptotically well-behaved in the null recurrent (nonstationary) case.[6] We will relax the fixed $\overline{T}$ condition (and let $\overline{T}$ diverge with T) when focusing on ergodic (or strictly stationary) systems (Remark 12).

Theorem 5. (Variance moments: Weak convergence.)

Assume

$$\begin{aligned}
\lim_{n,T\to\infty} h_{n,T} v(T) &= \infty, \\
\lim_{n,T\to\infty} h^5_{n,T} v(T) &= 0, \\
\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}} \left(\Delta_{n,T} \log \frac{1}{\Delta_{n,T}} \right)^{1/2} &= 0, \\
\lim_{k,n,T\to\infty} \frac{T v(T)^{-1/2}}{\Delta_{n,T} h^{1/2}_{n,T} k^\alpha \phi^\beta_{n,T}} + \frac{T v(T)^{-1/2}}{\Delta_{n,T} h^{1/2}_{n,T}} \left(\phi_{n,T} \log \left(\frac{1}{\phi_{n,T}} \right) \right)^{1/2} &= 0,
\end{aligned} \tag{32}$$

with $\alpha \in (0, \frac{1}{2}]$ *and* $\beta = [0,1]$, *where* $\overline{L}_{\sigma^2}(T,x) \propto v(T)$.

Generalized Duffie et al.'s model

Expected jump size:

$$(\Upsilon(T))^{-1/2}\, \overline{T} \left\{ \widehat{\mu}_\xi - \mu_\xi \right\} \Rightarrow \mathbf{N}(0,1),$$

where

$$\Upsilon(T) = \int_{-\infty}^{\infty} \lambda_{\sigma^2}(x) \left(\frac{\overline{L}^2_{\sigma^2}(\overline{T},x)}{\overline{L}_{\sigma^2}(T,x)} \right) \mathbf{E} \left(\left(\frac{1}{4\theta^3(x)} (\xi^\sigma)^4 - \frac{\theta^4(x)}{4\left(\theta^3(x)\right)^2} (\xi^\sigma)^3 \right)^2 \right) dx.$$

Jump intensity:

$$\sqrt{h_{n,T} \widehat{\overline{L}}_{\sigma^2}(T,x)} \left\{ \widehat{\lambda}_{\sigma^2}(x) - \lambda_{\sigma^2}(x) \right\} \Rightarrow \mathbf{N} \left(0, \mathbf{K}_2 \frac{\lambda_{\sigma^2}(x) \mathbf{E}\left((\xi^\sigma)^8 \right)}{(24)^2 \mu^8_\xi} \right).$$

Diffusive function:

$$\sqrt{h_{n,T} \widehat{\overline{L}}_{\sigma^2}(T,x)} \left\{ \widehat{\Lambda}^2_{\sigma^2}(x) - \Lambda^2_{\sigma^2}(x) \right\} \Rightarrow \mathbf{N} \left(0, \mathbf{K}_2 \lambda_{\sigma^2}(x) \mathbf{E} \left(\left((\xi^\sigma)^2 - \frac{1}{12\mu^2_\xi} (\xi^\sigma)^4 \right)^2 \right) \right).$$

[6] A similar asymptotic design is adopted in Bandi and Phillips (2007) in a different context. We refer the interested reader to that paper for details. In particular, Bandi and Phillips (2007) emphasize that the assumption is made only for theoretical convenience but is empirically immaterial.

Drift function:

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\left\{\widehat{m}_{\sigma^2}(x)-m_{\sigma^2}(x)\right\}\Rightarrow\mathbf{N}\left(0,\mathbf{K}_2\left(\Lambda^2_{\sigma^2}(x)+\lambda_{\sigma^2}(x)\mathbf{E}\left(\left(\xi^\sigma-\frac{1}{24\mu_\xi^3}\left(\xi^\sigma\right)^4\right)^2\right)\right)\right).$$

Log-variance model

Jump standard deviation:

$$\left(\Upsilon(T)\right)^{-1/2}2\overline{T}\sigma_\xi\left\{\widehat{\sigma}_\xi-\sigma_\xi\right\}\Rightarrow\mathbf{N}(0,1),$$

where

$$\Upsilon(T)=\int_{-\infty}^{\infty}\lambda_{\log\sigma^2}(x)\left(\frac{\overline{L}^2_{\sigma^2}(\overline{T},x)}{\overline{L}_{\sigma^2}(T,x)}\right)\mathbf{E}\left(\left(\frac{1}{5\theta^4(x)}\left(\xi^\sigma\right)^6-\frac{\theta^6(x)}{5\left(\theta^4(x)\right)^2}\left(\xi^\sigma\right)^4\right)^2\right)dx.$$

Jump intensity:

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\left\{\widehat{\lambda}_{\log\sigma^2}(x)-\lambda_{\log\sigma^2}(x)\right\}\Rightarrow\mathbf{N}\left(0,\mathbf{K}_2\frac{\lambda_{\log\sigma^2}(x)\mathbf{E}\left(\left(\xi^\sigma\right)^8\right)}{9\sigma_\xi^8}\right).$$

Diffusive function:

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\left\{\widehat{\Lambda}^2_{\log\sigma^2}(x)-\Lambda^2_{\log\sigma^2}(x)\right\}\Rightarrow\mathbf{N}\left(0,\mathbf{K}_2\lambda_{\log\sigma^2}(x)\mathbf{E}\left(\left(\left(\xi^\sigma\right)^2-\frac{1}{3\sigma_\xi^2}\left(\xi^\sigma\right)^4\right)^2\right)\right).$$

Drift function:

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\left\{\widehat{m}_{\log\sigma^2}(x)-m_{\log\sigma^2}(x)\right\}\Rightarrow\mathbf{N}\left(0,\mathbf{K}_2\left(\Lambda^2_{\log\sigma^2}(x)+\lambda_{\log\sigma^2}(x)\mathbf{E}\left(\left(\xi^\sigma\right)^2\right)\right)\right).$$

Proof. *See Appendix B.*

Remark 12 (The ergodic case) In the positive recurrent and strictly stationary case, $v(T)=T$ and $\frac{\widehat{\overline{L}}_{\sigma^2}(T,x)}{T}\xrightarrow{p}p(x)$, where $p(x)$ is the stationary density of the spot variance process. Hence, the rate of convergence of the point-wise estimates and the denominator of their asymptotic variances have a familiar look. The former is $\sqrt{h_{n,T}T}$. The later depends on the volatility process' time-invariant probability distribution, $p(x)$.

As expected, due to the averaging, the moments of the jump components converge at a faster (parametric) rate than that of the remaining functions. Even in this case, the look of their asymptotic distributions is more recognizable when setting $\overline{n}=n\to\infty$ and $\overline{T}=T\to\infty$ with $\Delta_{n,T}=\Delta_{\overline{n},\overline{T}}\to 0$. In this case, in effect,

$$\sqrt{T}\left\{\widehat{\mu}_\xi-\mu_\xi\right\}\Rightarrow\mathbf{N}\left(0,\int_{-\infty}^{\infty}\lambda_{\sigma^2}(x)\mathbf{E}\left(\left(\frac{1}{4\theta^3(x)}\left(\xi^\sigma\right)^4-\frac{\theta^4(x)}{4\left(\theta^3(x)\right)^2}\left(\xi^\sigma\right)^3\right)^2\right)p(x)dx\right),$$

and

$$\sqrt{T}\left\{\widehat{\sigma}_\xi-\sigma_\xi\right\}\Rightarrow\mathbf{N}\left(0,\frac{1}{4\sigma_\xi^2}\int_{-\infty}^{\infty}\lambda_{\log\sigma^2}(x)\mathbf{E}\left(\left(\frac{1}{5\theta^4(x)}\left(\xi^\sigma\right)^6-\frac{\theta^6(x)}{5\left(\theta^4(x)\right)^2}\left(\xi^\sigma\right)^4\right)^2\right)p(x)dx\right).$$

Remark 13 (Asymptotic covariance estimation) In light of Theorem 5, statistical inference on all functions and parameters of interest is now straightforward given estimates of the relevant asymptotic variances. To this extent, assume that the bandwidth conditions yielding consistency of the infinitesimal moments (as expressed in Theorem 4) are satisfied. Also, assume the usual asymptotic scheme. In the case of the generalized Duffie et al.'s model, we notice that:

$$\frac{h_{n,T}}{h_{\overline{n},\overline{T}}}\frac{\left(\Delta_{\overline{n},\overline{T}}\right)^2}{\Delta_{n,T}}\sum_{i=1}^{\overline{n}}\left(\frac{\widehat{\theta}^8}{16\left(\widehat{\theta}^3\right)^2}-\frac{2\widehat{\theta}^4\widehat{\theta}^7}{16\left(\widehat{\theta}^3\right)^3}+\frac{\left(\widehat{\theta}^4\right)^2\widehat{\theta}^6}{16\left(\widehat{\theta}^3\right)^4}\right)\frac{\sum_{j=1}^{\overline{n}}\mathbf{K}\left(\frac{\widetilde{\sigma}^2_{jT/n}-\widetilde{\sigma}^2_{i\overline{T}/\overline{n}}}{h_{n,\overline{T}}}\right)}{\sum_{j=1}^{n}\mathbf{K}\left(\frac{\widetilde{\sigma}^2_{jT/n}-\widetilde{\sigma}^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)}$$

$$\xrightarrow{p}\int_{-\infty}^{\infty}\lambda_{\sigma^2}(x)\left(\frac{\overline{L}^2_{\sigma^2}(\overline{T},x)}{\overline{L}_{\sigma^2}(T,x)}\right)\mathbf{E}\left(\left(\frac{1}{4\theta^3(x)}\left(\xi^\sigma\right)^4-\frac{\theta^4(x)}{4\left(\theta^3(x)\right)^2}\left(\xi^\sigma\right)^3\right)^2\right)dx,$$

$$\frac{\widehat{\theta}^8(x)}{(24)^2\widehat{\mu}_\xi^8}\xrightarrow{p}\frac{\lambda_{\sigma^2}(x)\mathbf{E}\left(\left(\xi^\sigma\right)^8\right)}{(24)^2\mu_\xi^8},$$

$$\widehat{\theta}^4(x)-\frac{2\widehat{\theta}^6(x)}{12\widehat{\mu}_\xi^2}+\frac{\widehat{\theta}^8(x)}{(12)^2\widehat{\mu}_\xi^4}\xrightarrow{p}\lambda_{\sigma^2}(x)\mathbf{E}\left(\left(\left(\xi^\sigma\right)^2-\frac{1}{12\mu_\xi^2}\left(\xi^\sigma\right)^4\right)^2\right),$$

and

$$\widehat{\theta}^2(x)-\frac{2\widehat{\theta}^5(x)}{24\widehat{\mu}_\xi^3}+\frac{\widehat{\theta}^8(x)}{(24)^2\widehat{\mu}_\xi^6}\xrightarrow{p}\Lambda^2_{\sigma^2}(x)+\lambda_{\sigma^2}(x)\mathbf{E}\left(\left(\xi^\sigma-\frac{1}{24\mu_\xi^3}\left(\xi^\sigma\right)^4\right)^2\right).$$

Furthermore,

$$\frac{h_{n,T}}{h_{\overline{n},\overline{T}}}\frac{\left(\Delta_{\overline{n},\overline{T}}\right)^2}{\Delta_{n,T}}\sum_{i=1}^{\overline{n}}\left(\frac{\widehat{\theta}^{12}}{25\left(\widehat{\theta}^4\right)^2}-\frac{2\widehat{\theta}^6\widehat{\theta}^{10}}{25\left(\widehat{\theta}^4\right)^3}+\frac{\left(\widehat{\theta}^6\right)^2\widehat{\theta}^8}{25\left(\widehat{\theta}^4\right)^4}\right)\frac{\sum_{j=1}^{\overline{n}}\mathbf{K}\left(\frac{\widetilde{\sigma}^2_{jT/n}-\widetilde{\sigma}^2_{i\overline{T}/\overline{n}}}{h_{n,\overline{T}}}\right)}{\sum_{j=1}^{n}\mathbf{K}\left(\frac{\widetilde{\sigma}^2_{jT/n}-\widetilde{\sigma}^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)}$$

$$\xrightarrow{p}\int_{-\infty}^{\infty}\lambda_{\log\sigma^2}(x)\left(\frac{\overline{L}^2_{\sigma^2}(\overline{T},x)}{\overline{L}_{\sigma^2}(T,x)}\right)\mathbf{E}\left(\left(\frac{1}{5\theta^4(x)}\left(\xi^\sigma\right)^6-\frac{\theta^6(x)}{5\left(\theta^4(x)\right)^2}\left(\xi^\sigma\right)^4\right)^2\right)dx,$$

$$\frac{\widehat{\theta}^8(x)}{9\widehat{\sigma}_\xi^8}\xrightarrow{p}\frac{\lambda_{\log\sigma^2}(x)\mathbf{E}\left(\left(\xi^\sigma\right)^8\right)}{9\sigma_\xi^8},$$

$$\widehat{\theta}^4(x)-\frac{2\widehat{\theta}^6(x)}{3\widehat{\sigma}_\xi^2}+\frac{\widehat{\theta}^8(x)}{9\widehat{\sigma}_\xi^4}\xrightarrow{p}\lambda_{\log\sigma^2}(x)\mathbf{E}\left(\left(\left(\xi^\sigma\right)^2-\frac{1}{3\sigma_\xi^2}\left(\xi^\sigma\right)^4\right)^2\right),$$

and, of course,

$$\widehat{\theta}^2(x)\xrightarrow{p}\Lambda^2_{\log\sigma^2}(x)+\lambda_{\log\sigma^2}(x)\mathbf{E}\left(\left(\xi^\sigma\right)^2\right)$$

in the log-variance case.[7]

[7] The proofs of these results follow from the methods laid out in Appendix B. For brevity, we do not report them here. However, they may be provided by the authors upon request.

6 Jumps in returns and volatility: $\mathbf{dJ}_t^r \neq 0$ and $\mathbf{dJ}_t^\sigma \neq 0$

When allowing for the empirically-important case of discontinuities in the price process, realized variance, realized kernels, and the two-scale estimator, *inter alia*, identify the continuous quadratic variation component of the price process $\int \sigma_s^2 ds$, as earlier, in addition to the sum of the squared jumps. We therefore need to consider estimators which solely identify integrated variance. The realized bipower variation measure of Barndorff-Nielsen and Shephard (2004, 2005), which we formally define in Section 8 below, achieves, among other procedures, this goal.

Remark 14 (Bipower variation in the no noise case.) Provided $\frac{1}{\phi_{n,T}^{1/2}}\left(\frac{1}{k}\log\left(\frac{k}{\phi_{n,T}}\right)\right)^{1/2} \to 0$ with $\phi_{n,T} \to 0$ and $k \to \infty$, the quantity $\widetilde{\sigma}_{iT/n}^2 = \frac{\widehat{V}_{iT/n}}{\phi_{n,T}}$, where $\widehat{V}_{iT/n}$ is bipower variation, is consistent for $\sigma_{iT/n}^2$ in the presence of jumps in returns and

$$\widetilde{\sigma}_{iT/n}^2 - \sigma_{iT/n}^2 = O_p\left(\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\right) + O_p\left(\frac{1}{\phi_{n,T}}\left(\frac{\phi_{n,T}}{k}\log\left(\frac{k}{\phi_{n,T}}\right)\right)^{1/2}\right) + O_p\left(\frac{1}{\sqrt{k}}\right).$$

In other words, while Assumption 2 is technically not satisfied (when jumps play a role) in the bipower variation case due to the presence of an asymptotic (jump-induced) bias-component (whose order is given by the second term in the above expression),[8] our theory continues to be valid with $\alpha = \frac{1}{2}$ and $\beta = \frac{1}{2}$. As done earlier when using spot variance estimates for which Assumption 2 (and a classical mixed normal theory of inference) hold, these choices, in fact, are sufficient to eliminate the measurement error induced by the estimated spot variances. Appendix A provides more details and discusses the more general case of multipower variation-based spot variance estimation.

7 $\mathbf{dJ}_t^r \neq 0$, $\mathbf{dJ}_t^\sigma \neq 0$, risk-return trade-offs, and leverage effects

We now turn to the full system for our more general case with both jumps in returns and in volatility. Given spot variance estimates $\widetilde{\sigma}_{iT/n}^2$ (obtained by using bipower variation or alternative identification methods robust to jumps in returns) as well as infinitesimal moment estimates for the return process ($\widehat{\theta}_r^j(\sigma^2)$ with $j = 1, 2, ...$), the relevant functions and the features of the return jump distribution can be identified by using a scheme similar to those in Section 5. One could assume, for instance, Gaussian mean-zero jumps, i.e., $dJ_t^r = \psi dN_t^r$ with $\psi \sim \mathbf{N}(0, \sigma_\psi^2)$,[9] and employ

$$\widehat{\sigma}_\psi^2 = \frac{1}{\overline{n}}\sum_{i=1}^{\overline{n}} \frac{\widehat{\theta}_r^6(\widetilde{\sigma}_{iT/\overline{n}}^2)}{5\widehat{\theta}_r^4(\widetilde{\sigma}_{iT/\overline{n}}^2)}, \tag{33}$$

$$\widehat{\lambda}^r(\sigma^2) = \frac{\widehat{\theta}_r^4(\sigma^2)}{3\widehat{\sigma}_\psi^4}, \tag{34}$$

$$\widehat{\mu}(\sigma^2) = \widehat{\theta}_r^1(\sigma^2). \tag{35}$$

[8]The presence of jump-induced limiting biases in bipower (and multipower) estimates of integrated variance as been discussed by Barndorff-Nielsen et al. (2006) and Woerner (2006). Here, of course, we focus on the spot variance case.

[9]See, e.g., Eraker et al. (2003).

where,

$$\widehat{\theta}_r^j(\sigma^2) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{iT/n} - \sigma^2}{h_{n,T}}\right) \left(\log p_{(i+1)T/n} - \log p_{iT/n}\right)^j}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{iT/n} - \sigma^2}{h_{n,T}}\right)} \qquad j = 1, \ldots.$$

Should $\widehat{\mu}(\sigma^2)$ be a statistically increasing function of σ^2, then a risk-return trade-off would exist. Theorem 6 discusses consistency and weak convergence of $\widehat{\mu}(\sigma^2)$.

Theorem 6. (Risk-return trade-offs: consistency and weak convergence.) *If $k, n, T \to \infty$ and $h_{n,T}, \phi_{n,T} \to 0$ so that*

$$\begin{aligned}
\lim_{n,T\to\infty} h_{n,T} v(T) &= \infty, \\
\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}} \left(\Delta_{n,T} \log \frac{1}{\Delta_{n,T}}\right)^{1/2} &= 0, \\
\lim_{k,n,T\to\infty} \frac{T v(T)^{-1}}{\Delta_{n,T} h_{n,T} k^{\alpha} \phi^{\beta}_{n,T}} + \frac{T v(T)^{-1}}{\Delta_{n,T} h_{n,T}} \left(\phi_{n,T} \log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} &= 0,
\end{aligned}$$

with $\alpha \in (0, \frac{1}{2}]$ and $\beta = [0, 1]$, then,

$$\widehat{\mu}(\sigma^2) \xrightarrow{p} \mu(\sigma^2),$$

where $\overline{L}_{\sigma^2}(T, \sigma^2) \propto v(T)$. If

$$\begin{aligned}
\lim_{n,T\to\infty} h_{n,T} v(T) &= \infty, \\
\lim_{n,T\to\infty} h^5_{n,T} v(T) &= C_4, \\
\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}} \left(\Delta_{n,T} \log \frac{1}{\Delta_{n,T}}\right)^{1/2} &= 0, \\
\lim_{k,n,T\to\infty} \frac{T v(T)^{-1/2}}{\Delta_{n,T} h^{1/2}_{n,T} k^{\alpha} \phi^{\beta}_{n,T}} + \frac{T v(T)^{-1/2}}{\Delta_{n,T} h^{1/2}_{n,T}} \left(\phi_{n,T} \log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} &= 0,
\end{aligned}$$

then,

$$\sqrt{h_{n,T} \widehat{\overline{L}}_{\sigma^2}(T, \sigma^2)} \left\{\widehat{\mu}(\sigma^2) - \mu(\sigma^2) - \Gamma_{\mu}(\sigma^2)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 \theta^2(\sigma^2)\right),$$

with

$$\Gamma_{\mu}(\sigma^2) = h^2_{n,T} \mathbf{K}_1 \left[\mu^{'}(\sigma^2) \frac{s^{'}(\sigma^2)}{s(\sigma^2)} + \frac{1}{2} \mu^{''}(\sigma^2)\right],$$

where

$$\theta^2(\sigma^2) = \sigma^2 + \lambda^r(\sigma^2) \mathbf{E}(\psi^2),$$

$s(d\sigma^2)$ *is the variance process' invariant measure, and* C_4 *is a constant.*

Proof. *See Appendix B.*

The leverage function[10] can be identified as follows:

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

where $\widehat{\Lambda}^2_{f(.)}(\sigma^2)$ may be estimated by virtue of Eq. (22) or Eq. (30) (depending, of course, on the assumed variance model). Our empirical work will use Eq. (22). In light of the independence of the jumps in returns and volatility and the independence between jumps and Brownian shocks, $\widehat{\rho}(\sigma^2)$ identifies $\rho(\sigma^2)$ consistently, as we show in Theorem 7. The theorem assumes that the same bandwidth is used to estimate numerator and denominator of $\widehat{\rho}(\sigma^2)$.

Theorem 7. (Leverage: consistency and weak convergence.) *If* $k, n, T \to \infty$ *and* $h_{n,T}, \phi_{n,T} \to 0$ *so that*

$$\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}} \left(\Delta_{n,T} \log \frac{1}{\Delta_{n,T}}\right)^{1/2} = 0,$$
$$\lim_{k,n,T\to\infty} \frac{Tv(T)^{-1}}{\Delta_{n,T} h_{n,T} k^{\alpha} \phi^{\beta}_{n,T}} + \frac{Tv(T)^{-1}}{\Delta_{n,T} h_{n,T}} \left(\phi_{n,T} \log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} = 0,$$

with $\alpha \in (0, \frac{1}{2}]$ *and* $\beta = [0, 1]$, *then*

$$\widehat{C}(\sigma^2) \xrightarrow{p} C(\sigma^2),$$

where $\overline{L}_{\sigma^2}(T, \sigma^2) \propto v(T)$. *If*

$$\lim_{n,T\to\infty} \frac{h^5_{n,T} v(T)}{\Delta_{n,T}} = C_5,$$
$$\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}} \left(\Delta_{n,T} \log \frac{1}{\Delta_{n,T}}\right)^{1/2} = 0,$$
$$\lim_{k,n,T\to\infty} \frac{Tv(T)^{-1/2}}{\Delta^{3/2}_{n,T} h^{1/2}_{n,T} k^{\alpha} \phi^{\beta}_{n,T}} + \frac{Tv(T)^{-1/2}}{\Delta^{3/2}_{n,T} h^{1/2}_{n,T}} \left(\phi_{n,T} \log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} = 0,$$

then,

$$\sqrt{\frac{h_{n,T} \widehat{\overline{L}}_{\sigma^2}(T, \sigma^2)}{\Delta_{n,T}}} \left\{\widehat{C}(\sigma^2) - C(\sigma^2) - \Gamma_C(\sigma^2)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 \theta^C(\sigma^2)\right),$$

[10] Influential recent work on leverage estimation in stochastic volatility models includes Harvey and Shephard (1996), Jacquier et al. (2004), and Yu (2005), among others. Yu (2005) also provides a thorough discussion of the existing literature.

with

$$\theta^C(\sigma^2) = \Lambda^2_{f(.)}(\sigma^2)\sigma^2(1+\rho^2(\sigma^2)) + \sigma^2\lambda^\sigma_{f(.)}(\sigma^2)\mathbf{E}[\xi^2] + \Lambda^2_{f(.)}(\sigma^2)\lambda^r(\sigma^2)\mathbf{E}[\psi^2] + \lambda^\sigma_{f(.)}(\sigma^2)\mathbf{E}[\xi^2]\lambda^r(\sigma^2)\mathbf{E}[\psi^2],$$

and

$$\Gamma_{\widehat{C}}(\sigma^2) = h^2_{n,T}\mathbf{K}_1\left[z^{'}(\sigma^2)\frac{s^{'}(\sigma^2)}{s(\sigma^2)} + \frac{1}{2}z^{''}(\sigma^2)\right],$$

where $z(x) = \sqrt{x}\Lambda_{f(.)}(x)\rho(x)$, $s(dx)$ *is the variance process' invariant measure, and* C_5 *is a constant. Finally, if*

$$\begin{aligned}
\lim_{n,T\to\infty} h_{n,T}v(T) &= \infty, \\
\lim_{n,T\to\infty} h^5_{n,T}v(T) &= 0, \\
\lim_{n,T\to\infty} \frac{v(T)}{h_{n,T}}\left(\Delta_{n,T}\log\frac{1}{\Delta_{n,T}}\right)^{1/2} &= 0, \\
\lim_{k,n,T\to\infty} \frac{Tv(T)^{-1/2}}{\Delta_{n,T}h^{1/2}_{n,T}k^\alpha\phi^\beta_{n,T}} + \frac{Tv(T)^{-1/2}}{\Delta_{n,T}h^{1/2}_{n,T}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} &= 0,
\end{aligned}$$

then

$$\sqrt{h_{n,T}\widehat{\widehat{L}}_{\sigma^2}(T,\sigma^2)}\left\{\widehat{\rho}(\sigma^2) - \rho(\sigma^2)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\theta^\rho(\sigma^2)\right)$$

with $\theta^\rho(\sigma^2) = \frac{\rho^2(\sigma^2)}{4\Lambda^4_{f(.)}(\sigma^2)} Asyvar\left(\widehat{\Lambda}^2_{f(.)}(\sigma^2)\right)$ *and*

$$Asyvar\left(\widehat{\Lambda}^2_{f(.)}(\sigma^2)\right) = \lambda^\sigma_{\sigma^2}(x)\mathbf{E}\left(\left((\xi^\sigma)^2 - \frac{1}{12\mu^2_\xi}(\xi^\sigma)^4\right)^2\right)$$

if $f(.) = \sigma^2$ *or*

$$Asyvar\left(\widehat{\Lambda}^2_{f(.)}(\sigma^2)\right) = \lambda^\sigma_{\log\sigma^2}(x)\mathbf{E}\left(\left((\xi^\sigma)^2 - \frac{1}{3\sigma^2_\xi}(\xi^\sigma)^4\right)^2\right)$$

if $f(.) = \log\sigma^2$.

Proof. *See Appendix B.*

8 High-frequency spot variance estimates and data

We are interested in the joint S&P500 return/variance dynamics. We start with a description of the high-frequency variance estimates used to identify spot variance. We then present the data. The next section reports estimates of the functions and parameters driving the S&P500 return/variance evolution in the context of the generalized Duffie et al's model discussed in Section 5.

8.1 Bipower-based spot variance

As earlier, we assume availability of k price observations $\log(p_j)$ in each interval $[i\Delta_{n,T}, i\Delta_{n,T} + \phi_{n,T}]$ with $i = 1, ..., n$. Write $r_j = \log(p_j) - \log(p_{j-1})$ with $j = 2, ..., k$ and

$$\widetilde{\sigma}^2_{iT/n} = \frac{\widehat{V}_{iT/n}}{\phi_{n,T}} = \frac{\mu^{-2}\sum_{j=2}^{k}|r_j||r_{j-1}|}{\phi_{n,T}} \qquad i = 1, ..., n$$

with $\mu = \mathbf{E}(|Z|)$, where Z denotes the standard normal random variable. In the absence of market microstructure noise but, importantly, regardless of the presence of jumps in the return process, Appendix A shows that $\widetilde{\sigma}^2_{iT/n}$ is consistent (when $\phi_{n,T} \to 0$ and $k \to \infty$ at appropriate rates) for $\widetilde{\sigma}^2_{iT/n}$ for all i (c.f., Section 6).

Importantly, alternative spot variance estimates (such as those based on the two-scale estimator and the family of realized kernels) would give us robustness to market microstructure noise but would be affected by price discontinuities (when focusing on the variation of the continuous price component). Here, in agreement with our more general specification in Eqs. (1) and (2), as well as much recent evidence in empirical asset pricing, we opt for being general in terms of return dynamics and allow for discontinuities in the return sample path.[11]

However, in order to break the first-order dependence in the observed high-frequency returns induced by microstructure noise contaminations in the k price observations, we employ a "staggered version" of bipower variation and write

$$\widetilde{\sigma}^2_{iT/n} = \frac{\widehat{V}^{Stag.}_{iT/n}}{\phi_{n,T}} = \frac{\mu^{-2}\left(\frac{k-2}{k}\right)^{-1}\sum_{j=3}^{k}|r_j||r_{j-2}|}{\phi_{n,T}} \qquad i = 1, ..., n.$$

This correction does not yield theoretical consistency of the estimator in the presence of noise but preserves consistency in the no noise case and, importantly for our empirical purposes in this section, has been shown to perform well when applied to noise-contaminated price observations, as is the case in practice (see, e.g., Andersen et al., 2007, and Huang and Tauchen, 2005).[12]

8.2 Data

Our sample period is January 2, 1998 to March 31, 2006. We employ daily returns on the S&P500 index and high-frequency price data on the Standard and Poor's depository receipts (Spiders) to construct the index's (daily) variance estimates.[13] Specifically, we use Spiders mid-quotes on the NYSE sampled between 10am and 4pm. We delete quotes whose associated price changes and/or spreads are larger than 10%. In our sample, the average duration between quote updates is 11.53 seconds. The average spread and the average price level are 0.0015 and 117.27, respectively. Hence, in terms of our previous notation, $T = 9$, $n = 2,053$, and the average k is about $1,873$ (the number of seconds in a 6-hour period divided by 11.53).

We follow common practise in the literature and convert the integrated variance estimates $\widehat{V}$ into daily measures. Since the original estimates are for an intra-daily 6-hour period, we multiply them by a constant factor δ defined

[11] As is well-known, the study of integrated (and spot, in our case) variance estimates which are robust to *both* return jumps and market microstructure noise is an open area of research.

[12] For an interesting, alternative approach to integrated variance estimation using bipower variation we refer the reader to Corsi et al., 2008. Their approach employs the threshold methods proposed, in other contexts, by Mancini (2007).

[13] Spiders are shares in a trust which owns stocks in the same proportion as that found in the S&P500 index. Spiders trade like a stock (with the ticker symbol SPY) at approximately one-tenth of the level of the S&P500 index. They are widely used by institutions and traders as bets on the overall direction of the market or as a means of passive management.

as $\delta = \sum_{i=1}^{n} \left(r_{iT/n}^{S\&P500} \right)^2 / \sum_{i=1}^{n} \widehat{V}_{iT/n}$, where $r_{iT/n}^{S\&P500}$ is the return on the S&P500 index over day i. This procedure ensures that the average of the transformed variances, i.e., $\delta\widehat{V}$, is equal to the average of the squared daily returns. Alternatively, one could add the squared overnight returns to the original estimates. Qualitatively, we find similar results when using the latter procedure and only report results relying on the adjustment δ.[14]

In what follows, the S&P500 returns and the spot variance estimates are expressed in daily terms ($\phi_{n,T} = \frac{9}{2053}$). The returns are further expressed as percentages ($\times 100$). In agreement with this scaling, the daily spot variances are multiplied by $10,000$. As always, market returns display little autocorrelation (-0.026), little skewness (0.069), and excess kurtosis (5.57). The bipower-based spot variance estimates are strongly right-skewed (5.943) and persistent (0.761).

9 Stochastic volatility dynamics

We estimate the generalized Duffie et al.'s jump-diffusion model presented in Section 5. We choose a simple identification scheme, as laid out in Eq. (20) through Eq. (23), but, contrary to existing parametric work, allow for a nonlinear drift, diffusion, and jump intensity.[15] The validity of this scheme is verified below by simulation. Drift function, diffusion function, and intensity of the jumps are reported in annual terms. Figure 1 contains the functional estimates (along with the corresponding drift and diffusion function for the case with no variance jumps). The (asymptotic) confidence bands are obtained by using the limiting results in Theorem 5 along with asymptotic covariance estimates evaluated as discussed in Remark 13. We note that the (daily) spot variance process makes most of its visits at levels between about 0.1 and 0.8, i.e., for a volatility of annual S&P500 returns between about 5% and 15%, as implied by the variance's estimated local time. Not surprisingly, the point-wise confidence bands are relatively tighter in this range (c.f. Figure 1).

The estimated drift denotes mildly nonlinear mean-reversion. The diffusive function conforms more naturally with a nonlinear constant-elasticity-of-variance (CEV) specification than with a linear structure (i.e., a square-root specification for diffusive volatility), as introduced by Heston (1993) and adopted by several others. In the relevant variance range, the intensity estimates suggest between 0 and 6 volatility jumps per year (with point estimates around 2 annual jumps). The estimated expected size of the jumps is about 2.5.

We compare our findings to the parametric estimates (converted to annual figures) of Eraker et al. (2003, Table III, Column 5). In Eraker et al. (2003) the drift is linear, the diffusive volatility is square-root, and the intensity of the jumps is constant. Our nonparametric (nonlinear) drift implies more mean-reversion. Despite differences in the point estimates, their jump intensity and average jump size are *statistically* supported by our data. So is their variance's diffusion function. This said, our diffusive variance's point estimates differ from those in Eraker et al. (2003) in important ways. We find more volatility associated with the process' continuous component. As indicated above, we also find that the variance's diffusive function is more accurately represented by a flexible CEV specification (i.e., $\Lambda^2(x) \propto x^{3/2}$)[16] than by a square-root model for diffusive volatility (or a linear model for

---

[14] Hansen and Lunde (2005) provide a theoretical justification for this traditional adjustment while studying the optimal combination of overnight squared returns and intra-daily realized variance for the purpose of daily integrated variance estimation.

[15] The infinitesimal moments' bandwidths are set equal to $c_j * stdc\left(\widetilde{\sigma}^2\right) * n^{-1/5}$, where c_j is chosen by cross-validation. In general, $c_1 > c_2$ and $c_j > c_2$ for $j > 2$ (the first and higher moment's bandwidths are larger than the second moment bandwidth). We use a second-order Gaussian kernel for all moments.

[16] We assume a linear mean-reverting drift, a constant jump intensity, and exponential jumps for spot variance. Applying GMM to the infinitesimal first, second, third, and fourth moments, we find $\widehat{\Lambda}^2(x) = 0.1x^{3/2}$. The t-statistics associated with these estimates are equal to about 3 and 4.

diffusive variance). Using specifications without jumps in variance, Chacko and Viceira (2001) and Jones (2002) also emphasize the need for nonlinear structures in the variance of variance. In Chacko and Viceira (2001) such a need diminishes with the addition of jumps in returns. For brevity, given the largely methodological nature of this paper, we do not dwell on this important empirical point here. However, we refer the reader to the broader empirical treatment in Bandi and Renò (2008)[17] for residual-based procedures and nonparametric tests based on high-order moments which confirm (i) the need for richer variance-of-variance structures and (ii) the importance of variance jumps.

It is now of interest to assess whether the reported differences between our estimates and those obtained in Eraker et al. (2003) are simply due to the use of different sample periods (Eraker et al., 2003, employ S&P500 return data sampled between January 2, 1980, and December 31, 1999) or whether they are a genuine by-product of alternative variance filtering methods (in Eraket et al., 2003, daily variance is filtered from daily returns using MCMC methods).

9.1 The joint volatility/return dynamics

We estimate a (possibly) nonlinear model for returns with Gaussian jumps (Fig. 2). The identification scheme is therefore consistent with Eq. (28)-Eq. (31) applied to returns rather than to log-variances, see Eq. (33) through Eq. (35) in Section 7.[18] However, identification of the quantities which heavily hinge on high-order moments, namely the standard deviation of the jump size and the price jump intensity, is conducted by also allowing for higher (first) order terms in the relevant conditional moment representations. Specifically, write

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

and

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

Hence,

$$\widehat{\sigma}_\psi^2 = \frac{1}{5\overline{n}}\sum_{i=1}^{\overline{n}} \frac{\widehat{\theta}_r^6(\widetilde{\sigma}_{i\overline{T}/\overline{n}}^2) - 15\widehat{\theta}_r^2(\widetilde{\sigma}_{i\overline{T}/\overline{n}}^2)(\widehat{\theta}_r^4(\widetilde{\sigma}_{i\overline{T}/\overline{n}}^2) - 3\left[\widehat{\theta}_r^2(\widetilde{\sigma}_{i\overline{T}/\overline{n}}^2)\right]^2\Delta_{n,T})\Delta_{n,T}}{\widehat{\theta}_r^4(\widetilde{\sigma}_{i\overline{T}/\overline{n}}^2) - 3\left[\widehat{\theta}_r^2(\widetilde{\sigma}_{i\overline{T}/\overline{n}}^2)\right]^2\Delta_{n,T}}, \tag{36}$$

$$\widehat{\lambda}^r(\sigma^2) = \frac{\widehat{\theta}_r^4(\widetilde{\sigma}_{iT/n}^2) - 3\left[\widehat{\theta}_r^2(\widetilde{\sigma}_{iT/n}^2)\right]^2\Delta_{n,T}}{3\widehat{\sigma}_\psi^4}. \tag{37}$$

To define $\widehat{\sigma}_\psi^2$ we further weigh the quantities in the sum by the corresponding local density. The additional terms (in $\widehat{\sigma}_\psi^2$ and $\widehat{\lambda}^r(\sigma^2)$) are asymptotically negligible (thereby not affecting our limiting results), but play an important

[17] This is the unpublished working paper version of the current paper.

[18] The bandwidths are set using cross-validation as indicated in Footnote 15 above. In particular, the constants c_1 and c_j, with $c_j > 2$, are equal to about 3, whereas c_2 is equal to about 2. The corresponding bandwidths of the variance process are generally slightly larger. A second-order Gaussian kernel is again used to define all infinitesimal moment estimators.

role in finite samples (when estimating return dynamics) as we show below by simulation. As earlier, we employ the results in Theorem 1 (and Remark 13) to characterize the sampling error by virtue of asymptotic confidence bands.

For clarity, we again compare our estimates to the parametric estimates of the affine model with Gaussian jumps in returns of Eraker et al. (2003). In Eraker et al. (2003) the return drift and the intensity of the Gaussian jumps are constant. The return dynamics suggest the presence of a statistically-insignificant nonlinear risk-return trade-off. The insignificance of the trade-off is of course not surprising and fully consistent with much empirical work on the evaluation of the relation between conditional mean returns and conditional variance at low (daily, here) frequencies (see, e.g., Bandi and Perron, 2008, for references). Importantly, in our case the use of high-frequency data does not yield a stronger dependence between conditional mean and conditional variance as suggested in some recent work (see, e.g., Bali and Peng, 2006). Similarly, Eraker et al. (2003) stress that experimentation with a linear risk-return model did not deliver significant estimates using their filtering methods and, therefore, resorted to a specification with a constant mean (whose numerical value is reported in Fig. 2(a)). We find mildly hump-shaped (in the spot variance level) leverage effects around -0.5. We also find smaller jump sizes (implying about 95% jumps between 4.5% and -4.5%) and a slightly higher number of jumps (between 2 and 6 in the relevant variance range) than in Eraker et al. (2003).[19] We again refer the interested reader to Bandi and Renò (2008) for a broader set of empirical results providing additional evidence in favor of discontinuities in the return (and variance) process. We now turn to simulations.

10 Simulations

Consider the bi-variate system:

$$
\begin{aligned}
r_{t,t+\Delta_t} &= b\Delta_t + \sqrt{\sigma_t^2\Delta_t}\varepsilon_t^r + \psi_t J_t^r, && (38)\\
\sigma_{t+\Delta_t}^2 - \sigma_t^2 &= \kappa(\theta - \sigma_t^2)\Delta_t + \sigma_\upsilon\sqrt{\sigma_t^2\Delta_t}\varepsilon_t^\sigma + \xi_t^\sigma J_t^\sigma, && (39)
\end{aligned}
$$

where $\{J_t^r, J_t^\sigma\}$ are Bernoulli random variables with constant intensities $\lambda^r\Delta_t$ and $\lambda^\sigma\Delta_t$, $\{\varepsilon_t^r, \varepsilon_t^\sigma\}$ are standard Gaussian random variables with correlation ρ, ψ_t is a mean zero Gaussian random variable with standard deviation σ_ψ, ξ_t^σ is an exponential random variable with mean μ_σ, and Δ_t is a time-discretization (one day). We generate 2,053 observations (as in our sample) for every sample path and 1,000 paths. The parameters are those in Table III, Column 5, of Eraker et al. (2003) with the exception of σ_υ (set equal to 0.31). The specification $\Lambda^2(\sigma_t^2) = 0.1(\sigma_t^2)^{3/2}$ provides superior fit for our data, as indicated previously (Bandi and Renò, 2008, for further discussions). Consistent with the empirical work, we use cross-validated bandwidths.

Figs. 3 and 4 report the 10th, 50th, and 90th percentile of the distribution of the estimates. We start with the variance dynamics. Drift and diffusion function are estimated fairly accurately. In light of our empirical results regarding the shape of the variance's diffusive variance, this is an important finding. If anything, the diffusion

[19] The negative value of the estimated jump intensity at high spot variance levels (i.e., at values that are hardly visited in-sample) should not be surprising. It is a by-product of the scarcity of observations in this range combined with bias-corrected estimates (as discussed in the main text) which are not guaranteed to remain positive in finite samples. This is a general issue in the nonparametric literature sometimes caused by the identification scheme, sometimes even caused by the (implied) kernel function used (as in the case of local polynomial estimates, for example, see Xu, 2007). While this paper lays out an identification methodology which is shown (through simulations and applied work) to be very informative even in a relatively simple form, future research should focus on identification schemes which explicitly address the non-negativity issue in finite samples.

estimates tend to be slightly downward biased, thereby possibly reinforcing our previous results about the need for a higher variance of variance. The jump features are accurately estimated. The expected jump size is very slightly downward biased (with the true value being well within the 95% band), while the intensity of the jumps is only mildly upward biased and tends to increase when moving away from the bulk of the data.

We now turn to the return and joint dynamics. The return drift and the leverage parameter are fairly accurately estimated. However, interestingly, if one were to use a straightforward identification scheme as in Section 7, the standard deviation of the Gaussian jumps would be excessively downward biased whereas the return jump intensities would be biased upward with, again, an increasing nonlinear trend when moving to values away from the center of the simulated data. Importantly, similar patterns would be observed in the data had this simple identification scheme been employed. The use of first-order adjustments (as done with data previously) improves finite sample performance drastically. The "corrected" sigma estimates are now much closer to the true value resulting, in conjunction with a bias-corrected fourth moment, in substantially more accurate lambda estimates.

We conclude with three observations. First, we find that the most important departure from affine stochastic volatility models with Gaussian jumps in returns and exponential jumps in variance is the nonlinear shape of the variance's diffusion function. Simulations show that this function is estimated accurately. Second, the features of the jumps (their probability and jump distribution) appear to be more easily identifiable for variance than for the (noisier) return process, thereby requiring finite sample corrections in the later case. We provide these corrections, when needed, and emphasize their potential importance for applied purposes, in general. Third, we stress that, while very informative, natural identification schemes and straightforward bandwidth choices have been used throughout. More efficient schemes potentially making use of the informational content of alternative infinitesimal moments could have been employed. Different bandwidth choices capable of adapting to the sparsity of the data (as implied by our asymptotic results) may also have been used. These issues are better left for future work.

11 Conclusions

We study stochastic volatility modelling in continuous time by employing functional estimation procedures. Preliminary nonparametric estimates of spot variance (for which we provide a theory of inference) are employed, in conjunction with functional estimates of the model's conditional moments, to learn about the functions and parameters driving the joint return/variance dynamics. The model and identification methods allow for nonlinearities in the drift and diffusion functions as well as in the intensities of the return/variance jumps. Nonlinear leverage effects are also permitted. Our limiting results control for the measurement error induced by the preliminary spot variance estimates and are derived under the weaker (than stationarity) assumption of recurrence. We show consistency and weak convergence of all the relevant functions and jump parameters of generalized (i.e., nonlinear) versions of two widely-employed stochastic volatility models.

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

A Appendix A: A theory of spot variance estimation

Write $\widetilde{\sigma}^2_{iT/n} = \frac{\widehat{V}_{iT/n}}{\phi_{n,T}}$, where $\widehat{V}_{iT/n}$ is an integrated (over $\phi_{n,T} \to 0$) variance estimator constructed using k intra-daily observations (with $k \to \infty$). In this Appendix we show that Assumption 2 and the stronger weak convergence result

$$\phi^{\beta}_{n,T} k^{\alpha} \left\{ \widetilde{\sigma}^2_{i/n} - \sigma^2_{iT/n} \right\} \underset{k\to\infty,\phi_{n,T}\to 0}{\Rightarrow} MN\left(0, a\left(\sigma^4_{iT/n}\right)^{\eta} + b\right)$$

are satisfied, under assumptions, by a large class of integrated variance estimators $\widehat{V}_{iT/n}$ recently proposed in the literature if $\phi^{\beta}_{n,T} k^{\alpha} \to \infty$ and $k^{\alpha}\phi^{\beta}_{n,T}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$.

As we clarify below, depending on the estimator, the result hinges on the absence of jumps in the return process, on the absence of market microstructure noise, or both. For some (kernel-based) estimators, the result also requires appropriate conditions on the number of autocovariances. We provide these conditions for the two-scale estimator and for the family of flat-top kernel estimators. More generally, by specifing the values of α, β, a, b, and η for various estimators, we provide an inferential theory for spot variance estimation under alternative estimators and different data generating processes allowing (or not) for discontinuities in the price process and market microstructure noise. Extensions to additional estimators can be conducted along similar lines.

Case 1. $dJ^r_t = 0$**, without microstructure noise.**

1. **Realized variance** (Andersen et al., 2003, and Barndorff-Nielsen and Shephard, 2002): $\alpha = \frac{1}{2}$, $\beta = 0$, $a = 2$, $b = 0$, *and* $\eta = 1$.
2. **Bipower variation** (Barndorff-Nielsen and Shephard, 2004, 2005): $\alpha = \frac{1}{2}$, $\beta = 0$, $a \approx 2.6$, $b = 0$, *and* $\eta = 1$.
3. **Realized range** (Christensen and Podolskij, 2007): $\alpha = \frac{1}{2}$, $\beta = 0$, $a \approx 0.4$, $b = 0$, *and* $\eta = 1$.
4. **Fourier estimator** (Malliavin and Mancino, 2008): *same as realized variance.*

Case 2. $dJ^r_t \neq 0$**, without microstructure noise.**

5. **Bipower variation** (Barndorff-Nielsen and Shephard, 2004, 2005): *The above weak convergence result does not hold. However, if* $\frac{1}{\phi^{1/2}_{n,T}}\left(\frac{1}{k}\log\left(\frac{k}{\phi_{n,T}}\right)\right)^{1/2} \to 0$, *then* $\widetilde{\sigma}^2_{i/n} = \sigma^2_{iT/n} + o_p(1)$ *and*

$$\widetilde{\sigma}^2_{iT/n} - \sigma^2_{iT/n} = O_p\left(\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\right) + O_p\left(\frac{1}{\phi_{n,T}}\left(\frac{\phi_{n,T}}{k}\log\left(\frac{k}{\phi_{n,T}}\right)\right)^{1/2}\right) + O_p\left(\frac{1}{\sqrt{k}}\right).$$

6. **Multipower variation**: *Assume multipower variation is computed by summing up z adjacent absolute values of equilibrium returns* $|r_{j\delta}|^{2/z}$ *with* $z \geq 2$. *If* $\frac{1}{\phi_{n,T}}\left(\frac{\phi_{n,T}}{k}\log\left(\frac{k}{\phi_{n,T}}\right)\right)^{(z-1)/z} \to 0$, *then* $\widetilde{\sigma}^2_{i/n} = \sigma^2_{iT/n} + o_p(1)$ *and*

$$\widetilde{\sigma}^2_{iT/n} - \sigma^2_{iT/n} = O_p\left(\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\right) + O_p\left(\frac{1}{\phi_{n,T}}\left(\frac{\phi_{n,T}}{k}\log\left(\frac{k}{\phi_{n,T}}\right)\right)^{(z-1)/z}\right) + O_p\left(\frac{1}{\sqrt{k}}\right).$$

   If $k^{\frac{1}{2}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$ *and* $\frac{k^{\frac{1}{2}}}{\phi_{n,T}}\left(\frac{\phi_{n,T}}{k}\log\left(\frac{k}{\phi_{n,T}}\right)\right)^{(z-1)/z} \to 0$, *then the above weak convergence result holds with* $\alpha = \frac{1}{2}$, $\beta = 0$, $a \neq 0$, $\eta = 1$, *and* $b = 0$.
7. **Threshold realized variance** (Mancini, 2007): $\alpha = \frac{1}{2}$, $\beta = 0$, $a = 2$, $b = 0$, *and* $\eta = 1$.
8. **Threshold bipower variation** (Corsi et al., 2008): $\alpha = \frac{1}{2}$, $\beta = 0$, $a \approx 2.6$, $b = 0$, *and* $\eta = 1$.

Case 3. $dJ^r_t = 0$**, with microstructure noise.**

The market microstructure noise is assumed to satisfy the assumptions that are common to Zhang et al. (2005) and Barndorff-Nielsen et al. (2006). Write $\log(p_j)^* = \log(p_j) + \eta_j$ (for $j = 1, ..., k$ over $\phi_{n,T}$), where the shocks η_j are iid in discrete time with mean zero and variance $\mathbf{E}(\eta^2)$. Below, the symbol $\mathbf{E}(\varepsilon^2)$ denotes the variance of the contaminations in the return process (i.e., $\mathbf{E}(\varepsilon^2) = 2\mathbf{E}(\eta^2)$).

The two-scale estimator (Zhang et al., 2005): Define q non-overlapping sub-grids $\Psi^{(i)}$ of the original grid of k arrival times with $i = 1, ..., q$. The first sub-grid starts at t_0 and takes every q-th arrival time, i.e., $\Psi^{(1)} = (t_0, t_{0+q}, t_{0+2q}, ...,)$, the second sub-grid starts at t_1 and also takes every q-th arrival time, i.e., $\Psi^{(2)} = (t_1, t_{1+q}, t_{1+2q}, ...,)$, and so on. Given the generic i-th sub-grid of arrival times, define $\widehat{V}^{(i)} = \sum_{t_j, t_{j+} \in \Psi^{(i)}} \left(\log(p_{t_{j+}})^* - \log(p_{t_j})^*\right)^2$, where t_j and t_{j+} denote adjacent elements in $\Psi^{(i)}$. The two-scale estimator is constructed as $\widehat{V}^{ZMA} = \frac{\sum_{i=1}^{q}\widehat{V}^{(i)}}{q} - \overline{k}\widehat{\mathbf{E}}(\varepsilon^2)$, where $\overline{k} = \frac{k-q+1}{q}$ and $\widehat{\mathbf{E}}(\varepsilon^2) = \frac{\sum_{j=1}^{k}\left(\log(p_{t_{j+}})^* - \log(p_{t_j})^*\right)^2}{k}$.

Realized kernels (Barndorff-Nielsen et al., 2006): Write $\widehat{V}^{BNHLS} = \widehat{\gamma}_0 + \sum_{s=1}^{q} w_s(\widehat{\gamma}_s + \widehat{\gamma}_{-s})$, where $\widehat{\gamma}_s = \sum_{j=1}^{k} r_j^* r_{j-s}^*$ with $s = -q, ..., q$, $r_j^* = \log(p_j)^* - \log(p_{j-1})^*$, $w_s = g\left(\frac{s-1}{q}\right)$, and $g(.)$ is a kernel function on $[0,1]$ satisfying $g(0) = 1$ and $g(1) = 0$.

9. **Two-scale estimator**: *If* $q = \tau k^{2/3}$, *then* $\beta = 1$, $\alpha = \frac{1}{6}$, $a = 0$, *and* $b = \left(\frac{8}{\tau^2}\right)\left(\mathbf{E}(\varepsilon^2)\right)^2$.
10. **Two-scale estimator**: *If* $q^o = \tau\left(\frac{k}{\phi_{n,T}}\right)^{2/3}$ *with* $\tau = \left(\frac{12\left(\mathbf{E}(\varepsilon^2)\right)^2}{\sigma^4_{iT/n}}\right)^{1/3}$, *then* $\beta = \frac{1}{3}$, $\alpha = \frac{1}{6}$, $a = 2\left(12\left(\mathbf{E}(\varepsilon^2)\right)^2\right)^{1/3}$, $\eta = \frac{2}{3}$, *and* $b = 0$. *The optimal rate is* $k^{1/10}$.
11. **Realized kernels**: *If* $q = \tau k^{2/3}$, *then* $\beta = 1$, $\alpha = \frac{1}{6}$, $a = 0$, *and* $b = 4\left(\mathbf{E}(\varepsilon^2)\right)^2 \frac{1}{\tau^2}\{g'(0)^2 + g'(1)^2\}$.
12. **Realized kernels**: *If* $q^o = \tau\left(\frac{k}{\phi_{n,T}}\right)^{2/3}$ *with* $\tau = \left(\frac{2\{g'(0)^2+g'(1)^2\}}{g_\bullet^{0,0}}\right)^{1/3}\left(\frac{\left(\mathbf{E}(\varepsilon^2)\right)^2}{\sigma^4_{iT/n}}\right)^{1/3}$, *then* $\beta = \frac{1}{3}$, $\alpha = \frac{1}{6}$, $a = 6\left(\frac{2\{g'(0)^2+g'(1)^2\}}{g_\bullet^{0,0}}\right)^{1/3} g_\bullet^{0,0}\left(\mathbf{E}(\varepsilon^2)^2\right)^{1/3}$, $\eta = \frac{2}{3}$, *and* $b = 0$.
13. **Realized kernels**: *Assume* $g'(0) = 0$ *and* $g'(1) = 0$. *If* $q = \tau k^{1/2}$, *then* $\beta = 1$, $\alpha = \frac{1}{4}$, $a = 0$, *and* $b = 4\left[\frac{1}{\tau^3}\{g'''(0)^2 + g_\bullet^{0,4}\} - \frac{1}{\tau}\left\{g_\bullet^{0,2}\right\}\right]\left(\mathbf{E}(\varepsilon^2)\right)^2$.
14. **Realized kernels**: *Assume* $g'(0) = 0$ *and* $g'(1) = 0$. *If* $q^o = \tau\frac{k^{1/2}}{\phi_{n,T}}$ *with* $\tau = \left(\frac{-g_\bullet^{0,2}}{g_\bullet^{0,0}}\right)^{1/2}\left(\frac{(\mathbf{E}(\varepsilon^2))^2}{\sigma^4_{iT/n}}\right)^{1/2}$, *then* $\beta = \frac{1}{2}$, $\alpha = \frac{1}{4}$, $a = 8\left\{-g_\bullet^{0,2}\right\}^{1/2}\left\{g_\bullet^{0,0}\right\}^{1/2}\left(\mathbf{E}(\varepsilon^2)^2\right)^{1/2}$, $\eta = \frac{1}{2}$, *and* $b = 0$. *The optimal rate is* $k^{1/8}$.

Proof of 5. Write

$$\frac{\widehat{V}^{BNS}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}} = \frac{\widehat{V}_c^{BNS}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}} + O_p\left(\frac{1}{\phi_{n,T}}\left(\frac{\phi_{n,T}}{k}\log\left(\frac{k}{\phi_{n,T}}\right)\right)^{1/2}\right),$$

where $\widehat{V}_c^{BNS}$ is the bipower variation estimator without jump components and the last term is the order of the jump term. Clearly, as in 2. above,

$$\frac{\widehat{V}_c^{BNS}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}} = O_p\left(\frac{1}{\sqrt{k}}\right).$$

Finally, using the Brownian modulus of continuity, we obtain the result in the statement. ■

Proof of 6. Obvious using the same expansion as in the proof of 5. ■

Proof of 9. From Bandi and Russell (2005), Theorem 2, when $\phi_{n,T} \to 0$ and $k \to \infty$, the dominating terms of the estimator's variance decomposition, i.e., $\mathbf{V}_{\sigma^2}\left(\widehat{V}^{ZMA} - \int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds\right)$, are

$$\frac{4}{3}\left(\phi_{n,T}\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^4 ds\right)v + 8\frac{\left(\mathbf{E}(\varepsilon^2)\right)^2}{v^2 k}$$

provided $v = \frac{q}{k} \to 0$. As for the bias term, i.e., $\mathbf{E}_{\sigma^2}\left(\widehat{V}^{ZMA} - \int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds\right)$, the dominating terms are

$$-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{vk} - v\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds.$$

If $q = \tau k^{2/3}$, then $v = \frac{q}{k} = \tau\frac{1}{k^{1/3}}$ and

$$\mathbf{V}_{\sigma^2}\left\{k^{1/6}\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right)\right\} \sim \frac{4}{3}\tau\phi_{n,T}\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^4 ds}{\phi_{n,T}^2} + \left(\frac{8}{\tau^2}\right)\frac{\left(\mathbf{E}(\varepsilon^2)\right)^2}{\phi_{n,T}^2},$$

or

$$\mathbf{V}_{\sigma^2}\left\{\phi_{n,T}k^{1/6}\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right)\right\}\sim\left(\frac{8}{\tau^2}\right)\left(\mathbf{E}(\varepsilon^2)\right)^2.$$

Notice, also, that

$$\mathbf{E}_{\sigma^2}\left\{\phi_{n,T}k^{1/6}\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right)\right\}\sim -k^{1/6}\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\tau k^{2/3}}-\tau\frac{k^{1/6}}{k^{1/3}}\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds\to 0.$$

■

Proof of 10. If $q=\tau_{\phi_{n,T}}k^{2/3}$ with $\tau_{\phi_{n,T}}=\left(\frac{12\left(\mathbf{E}(\varepsilon^2)\right)^2}{\phi_{n,T}\left(\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^4 ds\right)}\right)^{1/3}$ and $\phi_{n,T}$ is so that $\frac{q}{k}\to 0$, then

$$\begin{aligned}\mathbf{V}_{\sigma^2}\left\{k^{1/6}\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right)\right\} &\sim \frac{4}{3}\tau_{\phi_{n,T}}\phi_{n,T}\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^4 ds}{\phi_{n,T}^2}+\left(\frac{8}{\tau_{\phi_{n,T}}^2}\right)\frac{\left(\mathbf{E}(\varepsilon^2)\right)^2}{\phi_{n,T}^2}\\ &\sim \frac{2\left(12\left(\mathbf{E}(\varepsilon^2)\right)^2\right)^{1/3}\left(\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^4 ds\right)^{2/3}}{\left(\phi_{n,T}\right)^{4/3}}\end{aligned}$$

and

$$\mathbf{V}_{\sigma^2}\left\{\phi_{n,T}^{1/3}k^{1/6}\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right)\right\}\sim 2\left(12\left(\mathbf{E}(\varepsilon^2)\right)^2\right)^{1/3}\left(\sigma_{iT/n}^4\right)^{2/3}. \quad (40)$$

Importantly, for $\phi_{n,T}^{1/3}k^{1/6}\to\infty$ and $\phi_{n,T}\to 0$, it has to be the case that $\phi_{n,T}=k^{\theta}$ with $-\frac{1}{2}<\theta<0$. If this condition is satisfied, then the condition $\frac{q}{k}=\frac{k^{2/3}}{\phi_{n,T}^{2/3}k}=\frac{1}{\phi_{n,T}^{2/3}k^{1/3}}\to 0$ is also satisfied. Now write

$$\begin{aligned}&\mathbf{E}_{\sigma^2}\left\{\left(\phi_{n,T}\right)^{1/3}k^{1/6}\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right)\right\}\sim -k^{1/6}\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\left(\frac{k}{\phi_{n,T}}\right)^{2/3}\phi_{n,T}^{2/3}}-\left(\frac{k}{\phi_{n,T}}\right)^{2/3}\frac{k^{1/6}}{k}\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}^{2/3}}\\ \sim\ & -\frac{k^{1/6}}{k^{2/3}}\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds-\frac{1}{k_{n,T}^{1/6}\phi^{1/3}}\sigma_{iT/n}^2\to 0,\end{aligned}$$

since $\phi_{n,T}^{2/3}k^{1/3}\to 0$. The optimal rate can be derived more explicitly. Since

$$\widehat{V}^{ZMA}-\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds=O_p\left(\frac{q}{k}\left(\phi_{n,T}\right)^2\right)^{1/2}+O_p\left(\frac{k}{q^2}\right)^{1/2}$$

the optimal number of subsamples q^o is such that $q^o=\tau\left(\frac{k}{\phi_{n,T}}\right)^{2/3}$. Hence, provided $\frac{q^o}{k}\to 0$,

$$\mathbf{V}_{\sigma^2}\left\{\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right)\right\}\sim\frac{4}{3}\tau\frac{\left(\frac{k}{\phi_{n,T}}\right)^{2/3}}{k}\phi_{n,T}\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^4 ds}{\phi_{n,T}^2}+8\frac{\left(\mathbf{E}(\varepsilon^2)\right)^2}{\tau^2\phi_{n,T}^2}k\left(\frac{\phi_{n,T}}{k}\right)^{4/3}$$

and

$$\mathbf{V}_{\sigma^2}\left\{\phi_{n,T}^{1/3}k^{1/6}\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right)\right\}\sim\frac{4}{3}\tau\sigma_{iT/n}^4+8\frac{\left(\mathbf{E}(\varepsilon^2)\right)^2}{\tau^2}.$$

The optimal τ is therefore equal to $\left(\frac{12\left(\mathbf{E}(\varepsilon^2)\right)^2}{\sigma_{iT/n}^4}\right)^{1/3}$ which leads to

$$\mathbf{V}_{\sigma^2}\left\{\phi_{n,T}^{1/3}k^{1/6}\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right)\right\}\sim\frac{4}{3}\tau\sigma_{iT/n}^4+8\frac{\left(\mathbf{E}(\varepsilon^2)\right)^2}{\tau^2}\sim 2\left(12\left(\mathbf{E}(\varepsilon^2)\right)^2\right)^{1/3}\left(\sigma_{iT/n}^4\right)^{2/3}.$$

The final expression is the same as that in Eq. (40) above since $q = \tau_{\phi_{n,T}} k^{2/3}$ with $\tau_{\phi_{n,T}} = \left(\frac{12\left(\mathbf{E}(\varepsilon^2)\right)^2}{\phi_{n,T}\left(\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^4 ds\right)} \right)^{1/3}$ can be re-defined as $q = \tau_{\phi_{n,T}} \left(\frac{k}{\phi_{n,T}}\right)^{2/3} \left(\phi_{n,T}\right)^{2/3} \sim q^o$. ■

Remark (i) (The rate of convergence.) The condition $\phi_{n,T} = k^{\theta}$ with $-\frac{1}{2} < \theta < 0$ (which implies $\phi_{n,T}^{1/3} k^{1/6} \to \infty$) combined with $\phi_{n,T}^{1/3} k^{1/6} \left(\phi_{n,T} \log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$ (which is necessary to eliminate the asymptotic bias term), leads to (roughly) $\phi_{n,T} = k^{\theta}$ with $-\frac{1}{2} < \theta < -\frac{1}{5}$. Hence, one could choose $\theta = -\frac{1}{5}$, and the implied rate of convergence of the spot variance estimator would be $k^{1/10}$, which is slower than the rate of convergence of integrated variance ($k^{1/6}$).

Proof of 11. Using Barndorff-Nielsen et al. (2006), Eq. (15), and Bandi and Russell (2005), Theorem 3, write

$$
\begin{aligned}
&\mathbf{V}_{\sigma^2}\left\{\left(\widehat{V}^{BNHLS} - \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds\right)\right\} \sim 4 g_{\bullet}^{0,0} \left(\phi_{n,T} \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^4 ds\right) v \\
&-4\frac{1}{vk}\{g^{'}(0) + g_{\bullet}^{0,2}\}\{2\mathbf{E}(\varepsilon^2) \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds + \mathbf{E}(\varepsilon^2)^2\} \\
&+4\mathbf{E}(\varepsilon^2)^2 \left[\frac{1}{v^2 k}\{g^{'}(0)^2 + g^{'}(1)^2\} + \frac{1}{v^3 k^2}\{g^{'''}(0)^2 + g_{\bullet}^{0,4}\}\right] - 4\mathbf{E}(\varepsilon^2)^2 \frac{1}{vk}\frac{1}{2} g^{'}(0)^2,
\end{aligned}
$$

where the kernel-related g terms are defined in Barndorff-Nielsen et al. (2006). Hence, as earlier, in the general case there are two dominating terms (provided $v = \frac{q}{k} \to 0$) and

$$
\mathbf{V}_{\sigma^2}\left\{\left(\widehat{V}^{BNHLS} - \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds\right)\right\} \sim 4 g_{\bullet}^{0,0} \left(\phi_{n,T} \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^4 ds\right) v + 4\mathbf{E}(\varepsilon^2)^2 \frac{1}{v^2 k}\{g^{'}(0)^2 + g^{'}(1)^2\}.
$$

Similar expressions as in the two-scale case arise. In particular, if $q = \tau k^{2/3}$, then

$$
\begin{aligned}
\mathbf{V}_{\sigma^2}\left\{k^{1/6}\left(\frac{\widehat{V}^{BNHLS}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}}\right)\right\} \quad &\sim \quad 4\tau g_{\bullet}^{0,0} \left(\frac{\phi_{n,T} \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^4 ds}{\left(\phi_{n,T}\right)^2}\right) \\
&\quad + \frac{1}{\left(\phi_{n,T}\right)^2} \mathbf{E}(\varepsilon^2)^2 \frac{1}{\tau^2} 4\{g^{'}(0)^2 + g^{'}(1)^2\}
\end{aligned}
$$

and

$$
\mathbf{V}_{\sigma^2}\left\{\phi_{n,T} k^{1/6}\left(\frac{\widehat{V}^{BNHLS}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}}\right)\right\} \sim 4\mathbf{E}(\varepsilon^2)^2 \frac{1}{\tau^2}\{g^{'}(0)^2 + g^{'}(1)^2\}.
$$

■

Proof of 12. Assume now $q^o = \tau\left(\frac{k}{\phi_{n,T}}\right)^{2/3}$. Hence,

$$
\begin{aligned}
\mathbf{V}_{\sigma^2}\left\{\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}}\right)\right\} \quad &\sim \quad 4\tau g_{\bullet}^{0,0} \frac{\left(\frac{k}{\phi_{n,T}}\right)^{2/3}}{k} \phi_{n,T} \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^4 ds}{\phi_{n,T}^2} \\
&\quad + \frac{\left(\mathbf{E}(\varepsilon^2)\right)^2}{\tau^2 \phi_{n,T}^2} k \left(\frac{\phi_{n,T}}{k}\right)^{4/3} 4\{g^{'}(0)^2 + g^{'}(1)^2\}
\end{aligned}
$$

and

$$
\mathbf{V}_{\sigma^2}\left\{\phi_{n,T}^{1/3} k^{1/6}\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}}\right)\right\} \sim 4\tau g_{\bullet}^{0,0} \sigma_{iT/n}^4 + \frac{\left(\mathbf{E}(\varepsilon^2)\right)^2}{\tau^2} 4\{g^{'}(0)^2 + g^{'}(1)^2\}.
$$

The optimal τ is now

$$\left(\frac{2\{g^{'}(0)^2+g'(1)^2\}}{g_{\bullet}^{0,0}}\right)^{1/3}\left(\frac{\left(\mathbf{E}(\varepsilon^2)\right)^2}{\sigma_{iT/n}^4}\right)^{1/3}.$$

This choice implies

$$\mathbf{V}_{\sigma^2}\left\{\phi_{n,T}^{1/3}k^{1/6}\left(\frac{\widehat{V}^{ZMA}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2ds}{\phi_{n,T}}\right)\right\}\sim 6\left(\frac{2\{g^{'}(0)^2+g'(1)^2\}}{g_{\bullet}^{0,0}}\right)^{1/3}g_{\bullet}^{0,0}\left(\mathbf{E}(\varepsilon^2)^2\right)^{1/3}\left(\sigma_{iT/n}^4\right)^{2/3}.$$

■

Proof of 13. Now notice that if $g^{'}(0)=0$ and $g'(1)=0$, then

$$\begin{aligned}\mathbf{V}_{\sigma^2}\left\{\left(\widehat{V}^{BNHLS}-\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2ds\right)\right\} \quad\sim\quad & 4g_{\bullet}^{0,0}\left(\phi_{n,T}\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^4ds\right)v \\ & -4\frac{1}{vk}\{g_{\bullet}^{0,2}\}\{\mathbf{E}(\varepsilon^2)^2+2\mathbf{E}(\varepsilon^2)\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2ds\} \\ & +4\mathbf{E}(\varepsilon^2)^2\left[\frac{1}{v^3k^2}\{g^{'''}(0)^2+g_{\bullet}^{0,4}\}\right].\end{aligned}$$

If $q=\tau k^{1/2}$, then

$$\begin{aligned}& \mathbf{V}_{\sigma^2}\left\{\left(\frac{\widehat{V}^{BNHLS}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2ds}{\phi_{n,T}}\right)\right\} \\ \sim\quad & 4\tau g_{\bullet}^{0,0}\left(\frac{\phi_{n,T}\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^4ds}{\left(\phi_{n,T}\right)^2}\right)\frac{1}{k^{1/2}}-4\frac{1}{\tau k^{1/2}\left(\phi_{n,T}\right)^2}\{g_{\bullet}^{0,2}\}\{\mathbf{E}(\varepsilon^2)^2\} \\ +\quad & 4\frac{\mathbf{E}(\varepsilon^2)^2}{\left(\phi_{n,T}\right)^2}\left[\frac{1}{\tau^3k^{1/2}}\{g^{'''}(0)^2+g_{\bullet}^{0,4}\}\right]\end{aligned}$$

and

$$\mathbf{V}_{\sigma^2}\left\{\phi_{n,T}k^{1/4}\left(\frac{\widehat{V}^{BNHLS}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2ds}{\phi_{n,T}}\right)\right\}\sim 4\left[\frac{1}{\tau^3}\{g^{'''}(0)^2+g_{\bullet}^{0,4}\}-\frac{1}{\tau}\{g_{\bullet}^{0,2}\}\right]\left(\mathbf{E}(\varepsilon^2)\right)^2.$$

■

Proof of 14. If $q=\tau\frac{k^{1/2}}{\phi_{n,T}}$, then

$$\begin{aligned}\mathbf{V}_{\sigma^2}\left\{\left(\frac{\widehat{V}^{BNHLS}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2ds}{\phi_{n,T}}\right)\right\} \quad\sim\quad & 4g_{\bullet}^{0,0}\left(\frac{\phi_{n,T}\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^4ds}{\phi_{n,T}^2}\right)\tau\frac{1}{k^{1/2}\phi_{n,T}} \\ & -4\frac{1}{\tau k^{1/2}\phi_{n,T}}\{g_{\bullet}^{0,2}\}\{\left(\mathbf{E}(\varepsilon^2)\right)^2+2\mathbf{E}(\varepsilon^2)\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2ds\} \\ & +4\left(\mathbf{E}(\varepsilon^2)\right)^2\left[\frac{\phi_{n,T}}{\tau k^{1/2}}\{g^{'''}(0)^2+g_{\bullet}^{0,4}\}\right].\end{aligned}$$

and

$$\mathbf{V}_{\sigma^2}\left\{\phi_{n,T}^{1/2}k^{1/4}\left(\frac{\widehat{V}^{BNHLS}}{\phi_{n,T}}-\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2ds}{\phi_{n,T}}\right)\right\}\sim 4\tau g_{\bullet}^{0,0}\left(\sigma_{iT/n}^4\right)-4\frac{1}{\tau}\{g_{\bullet}^{0,2}\}\left(\mathbf{E}(\varepsilon^2)\right)^2.$$

Now select $\tau = \left(\frac{-g_{\bullet}^{0,2}}{g_{\bullet}^{0,0}}\right)^{1/2} \left(\frac{(\mathbf{E}(\varepsilon^2))^2}{\sigma^4_{iT/n}}\right)^{1/2}$. This choice yields,

$$\mathbf{V}_{\sigma^2}\left\{\phi_{n,T}^{1/2}k^{1/4}\left(\frac{\widehat{V}^{BNHLS}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma_s^2 ds}{\phi_{n,T}}\right)\right\} \sim 8\left\{-g_{\bullet}^{0,2}\right\}^{1/2}\left\{g_{\bullet}^{0,0}\right\}^{1/2}\left(\sigma^4_{iT/n}\right)^{1/2}\left(\mathbf{E}(\varepsilon^2)^2\right)^{1/2}.$$

■

Remark (ii) (The rate of convergence.) The condition $\phi_{n,T} = k^{\theta}$ with $-\frac{1}{2} < \theta < 0$ (which implies $\phi_{n,T}^{1/2}k^{1/4} \to \infty$) combined with $\phi_{n,T}^{1/2}k^{1/4}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$ (which is necessary to eliminate the asymptotic bias term), leads to (roughly) $\phi_{n,T} = k^{\theta}$ with $-\frac{1}{2} < \theta < -\frac{1}{4}$. Hence, one could choose $\theta = -\frac{1}{4}$, and the implied rate of convergence of the spot variance estimator would now be $k^{1/8}$, which appears to be optimal for this problem.

B Appendix B

We begin with useful preliminary lemmas.

Lemma B.1. (Bandi and Phillips, 2003) *Assume $dJ_t^{\sigma} = 0$ $\forall t$. Also, assume $\Delta_{n,T} = \frac{T}{n} \to 0$ and $h_{n,T} \to 0$ (as $n, T \to \infty$) in such a way as to guarantee that $\frac{\overline{L}_{\sigma^2}(T,x)}{h_{n,T}}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} \stackrel{a.s.}{\to} 0$ and $h_{n,T}\overline{L}_{\sigma^2}(T,x) \stackrel{a.s.}{\to} \infty$, where $\overline{L}_{\sigma^2}(T,x)$ is the chronological local time of σ^2. Then,*

$$\widehat{m}(x) = \frac{1}{\Delta_{n,T}}\frac{\sum_{i=0}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-x}{h_{n,T}}\right)(f(\sigma^2_{(i+1)T/n}) - f(\sigma^2_{iT/n}))}{\sum_{i=0}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-x}{h_{n,T}}\right)} \stackrel{a.s.}{\to} m(x).$$

If, in addition, $h_{n,T}^5\overline{L}_{\sigma^2}(T,x) = O_{a.s.}(1)$, then

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\left\{\widehat{m}(x) - m(x) - \Gamma_m(x)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\Lambda^2(x)\right),$$

with

$$\Gamma_m(x) = h_{n,T}^2\mathbf{K}_1\left[m^{'}(x)\frac{s^{'}(x)}{s(x)} + \frac{1}{2}m^{''}(x)\right],$$

where $s(x)$ is the diffusion's speed measure.

Lemma B.2. (Bandi and Phillips, 2003) *Assume $dJ_t^{\sigma} = 0$ $\forall t$. Also, assume $\Delta_{n,T} = \frac{T}{n} \to 0$ and $h_{n,T} \to 0$ (as $n, T \to \infty$) in such a way as to guarantee that $\frac{\overline{L}_{\sigma^2}(T,x)}{h_{n,T}}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} \stackrel{a.s.}{\to} 0$, where $\overline{L}_{\sigma^2}(T,x)$ is the chronological local time of σ^2. Then,*

$$\widehat{\Lambda}^2(x) = \frac{1}{\Delta_{n,T}}\frac{\sum_{i=0}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-x}{h_{n,T}}\right)(f(\sigma^2_{(i+1)T/n}) - f(\sigma^2_{iT/n}))^2}{\sum_{i=0}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-x}{h_{n,T}}\right)} \stackrel{a.s.}{\to} \Lambda^2(x).$$

If, in addition, $\frac{h_{n,T}^5\overline{L}_{\sigma^2}(T,x)}{\Delta_{n,T}} = O_{a.s.}(1)$, then

$$\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}{\Delta_{n,T}}}\left\{\widehat{\Lambda}^2(x) - \Lambda^2(x) - \Gamma_{\Lambda}(x)\right\} \Rightarrow \mathbf{N}\left(0, 2\mathbf{K}_2\Lambda^4(x)\right),$$

with

$$\Gamma_{\Lambda}(x) = h_{n,T}^2\mathbf{K}_1\left[\Lambda^{2'}(x)\frac{s^{'}(x)}{s(x)} + \frac{1}{2}\Lambda^{2''}(x)\right],$$

where $s(x)$ is the diffusion's speed measure.

Lemma B.3. (Bandi and Nguyen, 2003) *Assume $dJ_t^\sigma \neq 0$. Also, assume $\Delta_{n,T} = \frac{T}{n} \to 0$ and $h_{n,T} \to 0$ (as $n, T \to \infty$) in such a way as to guarantee that $\frac{\overline{L}_{\sigma^2}(T,x)}{h_{n,T}}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} \overset{a.s.}{\to} 0$ and $h_{n,T}\overline{L}_{\sigma^2}(T,x) \overset{a.s.}{\to} \infty$, where $\overline{L}_{\sigma^2}(T,x)$ is the chronological local time of σ^2. Then,*

$$\widehat{\theta}^j(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=0}^{n-1} \mathbf{K}\left(\frac{\sigma^2_{iT/n} - x}{h_{n,T}}\right)(f(\sigma^2_{(i+1)T/n}) - f(\sigma^2_{iT/n}))^j}{\sum_{i=0}^{n} \mathbf{K}\left(\frac{\sigma^2_{iT/n} - x}{h_{n,T}}\right)} \overset{a.s.}{\to} \theta^j(x) \qquad \forall j \geq 1.$$

If, in addition, $h^5_{n,T}\overline{L}_{\sigma^2}(T,x) = O_{a.s.}(1)$, then

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\left\{\widehat{\theta}^j(x) - \theta^j(x) - \Gamma_{\theta^j}(x)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\theta^{2j}(x)\right),$$

with

$$\Gamma_{\theta^j}(x) = h^2_{n,T}\mathbf{K}_1\left[\theta^{j'}(x)\frac{s'(x)}{s(x)} + \frac{1}{2}\theta^{j''}(x)\right],$$

where $s(dx)$ is the process' invariant measure $\forall j \geq 1$.

Proof of Theorem 1. Fix T. We wish to show that

$$\widetilde{L}_n(x) = \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right) - \int_0^T \frac{1}{h_{n,T}}\mathbf{K}\left(\frac{\sigma^2_s - x}{h_{n,T}}\right) ds = o_p(1).$$

Since $\mathbf{K}(.)$ is continuously-differentiable and bounded by Assumption 1, then

$$\begin{aligned}
\widetilde{L}_n(x) \quad &\leq \quad \frac{1}{h_{n,T}}\left|\sum_{i=0}^{n-1}\int_{iT/n}^{(i+1)T/n}\left[\mathbf{K}\left(\frac{\widetilde{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right) - \mathbf{K}\left(\frac{\sigma^2_s - x}{h_{n,T}}\right)\right] ds\right| \\
&\quad + \left|\frac{\Delta_{n,T}}{h_{n,T}}\mathbf{K}\left(\frac{\widetilde{\sigma}^2_0 - x}{h_{n,T}}\right)\right| + \left|\frac{\Delta_{n,T}}{h_{n,T}}\mathbf{K}\left(\frac{\widetilde{\sigma}^2_{n\Delta_{n,T}} - x}{h_{n,T}}\right)\right| \\
&\leq \quad \frac{1}{h_{n,T}}\sum_{i=0}^{n-1}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left|\mathbf{K}'\left(\frac{\widetilde{\sigma}^2_{is} - x}{h_{n,T}}\right)\right|\left|\left(\frac{\sigma^2_s - \widetilde{\sigma}^2_{i\Delta_{n,T}}}{h_{n,T}}\right)\right| ds + 2C_1 O_{a.s.}\left(\frac{\Delta_{n,T}}{h_{n,T}}\right),
\end{aligned}$$

where, by the mean-value theorem, σ^2_{is} is a value on the line segment connecting $\sigma^2_{i\Delta_{n,T}}$ with σ^2_s. Now notice that

$$\sup_{i\leq n}\left|\frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma^2_s ds}{\phi_{n,T}} - \sigma^2_{i\Delta_{n,T}}\right| \leq \sup_{i\leq n}\sup_{i\frac{T}{n}\leq s\leq i\frac{T}{n}+\phi_{n,T}}\left|\sigma^2_s - \sigma^2_{i\Delta_{n,T}}\right| = o_{a.s.}\left(\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\right)$$

by the Brownian modulus of continuity. Also, given Assumption 2,

$$\widetilde{\sigma}^2_{i\Delta_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}}\sigma^2_s ds}{\phi_{n,T}} = O_p\left(\frac{1}{k^\alpha\phi^\beta_{n,T}}\right)$$

uniformly over $i = 1, ..., n$. Finally,

$$\sup_{i\frac{T}{n}\leq s\leq i\frac{T}{n}+\Delta_{n,T}}\left|\sigma^2_s - \sigma^2_{i\Delta_{n,T}}\right| = o_{a.s.}\left(\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2}\right).$$

This implies that

$$\begin{aligned}
&\sup_{i\leq n}\left|\frac{\sigma^2_s - \widetilde{\sigma}^2_{i\Delta_{n,T}}}{h_{n,T}}\right| = O_p\left(\frac{1}{h_{n,T}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} + \frac{1}{h_{n,T}}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} + \frac{1}{k^\alpha\phi^\beta_{n,T}h_{n,T}}\right) \\
= \quad & O_p(g(T, n, k, \phi_{n,T}))
\end{aligned}$$

and, ignoring the negligible (under our assumptions on the bandwidth $h_{n,T}$) term $O_{a.s.}\left(\frac{\Delta_{n,T}}{h_{n,T}}\right)$,

$$\begin{aligned}
\widetilde{L}_n(x) &\leq \frac{O_p(g(T,n,k,\phi_{n,T}))}{h_{n,T}} \sum_{i=0}^{n-1} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left| \mathbf{K}^{'}\left(\frac{\sigma_s^2 - x}{h_{n,T}} + o_p(1)\right)\right| ds \\
&= \frac{O_p(g(T,n,k,\phi_{n,T}))}{h_{n,T}} \int_0^T \left| \mathbf{K}^{'}\left(\frac{\sigma_s^2 - x}{h_{n,T}} + o_p(1)\right)\right| ds \\
&= O_p(g(T,n,k,\phi_{n,T})) \int_{-\infty}^{\infty} \left| \mathbf{K}^{'}\left(v + o_p(1)\right)\right| \overline{L}_{\sigma^2}\left(T, vh_{n,T} + x\right) dv \\
&= C_2 O_p(g(T,n,k,\phi_{n,T})) \overline{L}_{\sigma^2}\left(T, x\right),
\end{aligned}$$

by the occupation time formula for semimartingales (see, e.g., Protter, 1995, Corollary 1, p.168) and the integrability of $\mathbf{K}^{'}(.)$ from Assumption 1. If $g(T,n,k,\phi_{n,T}) \to 0$, then $\widetilde{L}_n(x) \xrightarrow{p} 0$, and convergence to the chronological local time (at T and x) of the spot variance process derives from another application of the occupation time formula to $\int_0^T \frac{1}{h_{n,T}} \mathbf{K}\left(\frac{\sigma_s^2 - x}{h_{n,T}}\right) ds$.
■

Proof of Theorem 2. We show how the estimation error induced by the preliminary spot variance estimates is handled asymptotically in the drift case. Similar derivations apply to all infinitesimal moments (in both the diffusion and the jump-diffusion case) and are omitted for brevity. Write

$$\widetilde{\mathbf{K}}_i = \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right), \quad \mathbf{K}_i = \mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right), \quad \mathbf{K}_i^* = \mathbf{K}\left(\frac{\frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} - x}{h_{n,T}}\right)$$

Immediately, given the arguments in Theorem 1 above, $\widetilde{\mathbf{K}}_i - \mathbf{K}_i^* = O_p\left(\frac{1}{h_{n,T} k^{\alpha} \phi^{\beta}_{n,T}}\right)$, $\mathbf{K}_i^* - \mathbf{K}_i = O_p\left(\frac{1}{h_{n,T}}\left(\phi_{n,T} \log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\right)$,

$$\widetilde{\mathbf{K}}_i - \mathbf{K}_i = O_p\left(\frac{1}{h_{n,T} k^{\alpha} \phi^{\beta}_{n,T}} + \frac{1}{h_{n,T}}\left(\phi_{n,T} \log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\right) = O_p(q(T,n,k,\phi_{n,T})),$$

and

$$f(\widetilde{\sigma}^2_{(i+1)\Delta_{n,T}}) - f(\widetilde{\sigma}^2_{i\Delta_{n,T}}) = f(\sigma^2_{(i+1)\Delta_{n,T}}) - f(\sigma^2_{i\Delta_{n,T}}) + O_p\left(h_{n,T} q(T,n,k,\phi_{n,T})\right).$$

Now, write

$$\widehat{\theta}^1(x) = \frac{\sum_{i=1}^n \left[\mathbf{K}_i + O_p\left(q(T,n,k,\phi_{n,T})\right)\right]\left[f(\sigma^2_{(i+1)\Delta_{n,T}}) - f(\sigma^2_{i\Delta_{n,T}}) + O_p\left(h_{n,T} q(T,n,k,\phi_{n,T})\right)\right]}{\Delta_{n,T} \sum_{i=1}^n \left[\mathbf{K}_i + O_p\left(q(T,n,k,\phi_{n,T})\right)\right]}$$

and, since

$$\frac{1}{\Delta_{n,T}\sum_{i=1}^n \left[\mathbf{K}_i + O_p(q(T,n,k,\phi_{n,T}))\right]} = \frac{1}{\Delta_{n,T}\sum_{i=1}^n \mathbf{K}_i} - \frac{T O_p(q(T,n,k,\phi_{n,T}))}{\left(\Delta_{n,T}\sum_{i=1}^n \mathbf{K}_i\right)^2} + o_p(q(T,n,k,\phi_{n,T})),$$

then, neglecting smaller-order terms in $q(T,n,k,\phi_{n,T})$,

$$\begin{aligned}
&\widehat{\theta}^1(x) = \widehat{m}(x) \\
&-\frac{T O_p\left(q(T,n,k,\phi_{n,T})\right)\sum_{i=1}^n \left[\mathbf{K}_i + O_p\left(q(T,n,k,\phi_{n,T})\right)\right]\left[f(\sigma^2_{(i+1)\Delta_{n,T}}) - f(\sigma^2_{i\Delta_{n,T}}) + O_p\left(h_{n,T} q(T,n,k,\phi_{n,T})\right)\right]}{\left(\Delta_{n,T}\sum_{i=1}^n \mathbf{K}_i\right)^2} \\
&+\frac{\sum_{i=1}^n O_p(q(T,n,k,\phi_{n,T}))[f(\sigma^2_{(i+1)\Delta_{n,T}}) - f(\sigma^2_{i\Delta_{n,T}}) + O_p\left(h_{n,T} q(T,n,k,\phi_{n,T})\right)]}{\Delta_{n,T}\sum_{i=1}^n \mathbf{K}_i} \\
&+\frac{\sum_{i=1}^n \mathbf{K}_i O_p(h_{n,T} q(T,n,k,\phi_{n,T}))}{\Delta_{n,T}\sum_{i=1}^n \mathbf{K}_i} \\
= \;& \widehat{m}(x) + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3.
\end{aligned}$$

Since $\widehat{m}(x) = m(x) + o_p(1)$ from Lemma B.1, we only need to show that the remaining terms are so that $\mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3 = o_p(1)$. Write

$$\mathbf{R}_3 = O_p\left(\frac{h_{n,T}q(T,n,k,\phi_{n,T})}{\Delta_{n,T}}\right) = O_p\left(\frac{1}{\Delta_{n,T}k^{\alpha}\phi_{n,T}^{\beta}} + \frac{1}{\Delta_{n,T}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\right) \xrightarrow{p} 0$$

under our assumptions (see Eq. (15)). Now write

$$\begin{aligned}
\mathbf{R}_1 \quad &= \quad \frac{TO_p(q(T,n,k,\phi_{n,T}))\sum_{i=1}^{n}\mathbf{K}_i[f(\sigma^2_{(i+1)\Delta_{n,T}}) - f(\sigma^2_{i\Delta_{n,T}}) + O_p\left(h_{n,T}q(T,n,k,\phi_{n,T})\right)]}{\left(\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_i\right)^2} \\
&\quad + \frac{TO_p(q(T,n,k,\phi_{n,T}))\sum_{i=1}^{n}O_p(q(T,n,k,\phi_{n,T}))[f(\sigma^2_{(i+1)\Delta_{n,T}}) - f(\sigma^2_{i\Delta_{n,T}}) + O_p\left(h_{n,T}q(T,n,k,\phi_{n,T})\right)]}{\left(\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_i\right)^2} \\
&= \quad \mathbf{R}_4 + \mathbf{R}_5.
\end{aligned}$$

Since $\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}_i \xrightarrow{p} \overline{L}_{\sigma^2}(T,x)$ from Theorem 1 and $\overline{L}_{\sigma^2}(T,x) \propto v(T)$, then

$$\mathbf{R}_4 = O_p\left(\frac{Tq(T,n,k,\phi_{n,T})}{v(T)h_{n,T}\Delta_{n,T}}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} + \frac{Tq^2(T,n,k,\phi_{n,T})}{v(T)\Delta_{n,T}}\right) \xrightarrow{p} 0,$$

$$\begin{aligned}
\mathbf{R}_5 \quad &= \quad O_p\left(\frac{Tnq^2(T,n,k,\phi_{n,T})}{v(T)^2h_{n,T}^2}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} + \frac{Tnq^3(T,n,k,\phi_{n,T})h_{n,T}}{v(T)^2h_{n,T}^2}\right) \\
&= \quad O_p\left(\frac{T^2q^2(T,n,k,\phi_{n,T})}{v(T)^2h_{n,T}^2\Delta_{n,T}}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} + \frac{T^2q^3(T,n,k,\phi_{n,T})h_{n,T}}{v(T)^2h_{n,T}^2\Delta_{n,T}}\right) \\
&= \quad O_p\left(\frac{T^2v(T)^{-2}q^2(T,n,k,\phi_{n,T})\Delta_{n,T}}{h_{n,T}^2\Delta_{n,T}^2}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} + \frac{T^2v(T)^{-2}q^3(T,n,k,\phi_{n,T})}{h_{n,T}\Delta_{n,T}}\right) \xrightarrow{p} 0,
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{R}_2 \quad &= \quad O_p\left(\frac{q(T,n,k,\phi_{n,T})n}{h_{n,T}v(T)}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} + \frac{nq^2(T,n,k,\phi_{n,T})h_{n,T}}{h_{n,T}v(T)}\right) \\
&= \quad O_p\left(\frac{Tq(T,n,k,\phi_{n,T})}{h_{n,T}v(T)\Delta_{n,T}}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} + \frac{Tq^2(T,n,k,\phi_{n,T})h_{n,T}}{h_{n,T}v(T)\Delta_{n,T}}\right) \\
&= \quad O_p\left(\frac{Tv(T)^{-1}q(T,n,k,\phi_{n,T})}{h_{n,T}\Delta_{n,T}}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} + \frac{Tv(T)^{-1}q^2(T,n,k,\phi_{n,T})}{\Delta_{n,T}}\right) \xrightarrow{p} 0,
\end{aligned}$$

since

$$\frac{Tv(T)^{-1}q(T,n,k,\phi_{n,T})}{\Delta_{n,T}} = \frac{Tv(T)^{-1}}{\Delta_{n,T}h_{n,T}k^{\alpha}\phi_{n,T}^{\beta}} + \frac{Tv(T)^{-1}}{\Delta_{n,T}h_{n,T}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0.$$

As for weak convergence, note that

$$\begin{aligned}
&\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\left\{\widehat{\theta}^1(x) - m(x)\right\} \\
= \quad &\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\left\{\widehat{m}(x) - m(x)\right\} + \sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\left\{\mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3\right\} \\
= \quad &\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\left\{\widehat{m}(x) - m(x)\right\} + O_p\left(\frac{Tv(T)^{-1/2}}{\Delta_{n,T}h_{n,T}^{1/2}k^{\alpha}\phi_{n,T}^{\beta}} + \frac{Tv(T)^{-1/2}}{\Delta_{n,T}h_{n,T}^{1/2}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\right) \\
= \quad &\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\left\{\widehat{m}(x) - m(x)\right\} + o_p(1).
\end{aligned}$$

Lemma B.1. now gives the limiting distribution. ■

Proof of Theorem 3. The result follows from using the method of proof of Theorem 2 combined with Lemma B.2. ■

Proof of Theorem 4. The result follows from using the method of proof of Theorem 2 combined with Lemma B.3. ■

Proof of Theorem 5. For brevity, we focus on two representative cases in the context of the generalized Duffie et al.'s specification ($f(\sigma^2)=\sigma^2$ and $\xi\sim\exp(\mu_\xi)$). In particular, in order to illustrate the relevant issues, we detail the method of proof for *(1)* a point-wise functional estimator ($\widehat{\lambda}_{\sigma^2}(x)$) and *(2)* a semiparametric estimator ($\widehat{\mu}_\xi$). Similar methods apply to all estimators in Theorem 5. In all cases below, the estimation error induced by the spot variance estimates is handled as in proof of Theorem 2.

We begin with $\widehat{\lambda}_{\sigma^2}(x)$. Using a Taylor expansion, write the dominating terms in the estimation error decomposition as

$$\left\{\widehat{\lambda}_{\sigma^2}(x)-\lambda_{\sigma^2}(x)\right\}=\frac{\widehat{\theta}^4(x)}{24\widehat{\mu}_\xi^4}-\frac{\theta^4(x)}{24\mu_\xi^4}=\frac{\widehat{\theta}^4(x)-\theta^4(x)}{24\mu_\xi^4}-\frac{4\theta^4(x)(\widehat{\mu}_\xi-\mu_\xi)}{24\mu_\xi^5}. \tag{41}$$

By an application of Ito's lemma for Lèvy processes (see, e.g., Cont and Tankov, 2004, Proposition 8.15) to $\left(x-\sigma^2_{jT/n}\right)^4$, the dominating term in the limiting distribution of $\widehat{\theta}^4(x)-\theta^4(x)$ is easily shown to be $\dfrac{\frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{jT/n}-x}{h_{n,T}}\right)\int_{jT/n}^{(j+1)T/n}(\xi^\sigma)^4\overline{v}_\sigma(ds,d\xi^\sigma)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{jT/n}-x}{h_{n,T}}\right)}$.

Write the (standardized by $\sqrt{h_{n,T}}$) numerator of the previous ratio as

$$\mathbf{U}_{n,T}=\sum_{j=1}^{n-1}w(\sigma^2_{jT/n})\int_{jT/n}^{(j+1)T/n}\int(\xi^\sigma)^4\,\overline{v}_\sigma(ds,d\xi^\sigma)=\sum_{j=1}^{n-1}u_{jT/n,(j+1)T/n},$$

where $w(\sigma^2_{jT/n})=\frac{1}{\sqrt{h_{n,T}}}\mathbf{K}\left(\frac{\sigma^2_{jT/n}-x}{h_{n,T}}\right)$. For all T, the objects $\left\{u_{jT/n,(j+1)T/n},\Im_{j,n},1\le j\le n,\ n\ge 1\right\}$ constitute a zero-mean, square-integrable, martingale difference array. The conditional variance of $\mathbf{U}_{n,T}$ is

$$\begin{aligned}
\mathbf{U}^2_{n,T} &= \sum_{j=1}^{n-1}w^2(\sigma^2_{jT/n})\mathbf{E}\left\{\left(\int_{jT/n}^{(j+1)T/n}\int(\xi^\sigma)^4\overline{v}_\sigma(ds,d\xi^\sigma)\right)^2|\Im_{j,n}\right\}\\
&= \frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{jT/n}-x}{h_{n,T}}\right)\mathbf{E}\left\{\left(\int_{jT/n}^{(j+1)T/n}\lambda(\sigma^2_{s-})\mathbf{E}((\xi^\sigma)^8)ds\right)|\Im_{j,n}\right\}\\
&= \frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{jT/n}-x}{h_{n,T}}\right)\lambda(\sigma^2_{jT/n})\mathbf{E}((\xi^\sigma)^8)+o_p(1)\\
&\underset{\Delta_{n,T}\to 0}{\overset{p}{\to}} \frac{1}{h_{n,T}}\int_0^T\mathbf{K}^2\left(\frac{\sigma^2_{s-}-x}{h_{n,T}}\right)\lambda(\sigma^2_{s-})\mathbf{E}((\xi^\sigma)^8)ds=\widetilde{\mathbf{U}}.
\end{aligned}$$

If an appropriate negligibility conditions (i.e., a conditional version of the Lindeberg condition) is satisfied, namely

$$\forall\epsilon>0,\sum_{j=1}^{n-1}\mathbf{E}\left\{u^2_{jT/n,(j+1)T/n}\mathbf{1}_{(|u_{jT/n,(j+1)T/n}|>\epsilon)}|\Im_{j,n}\right\}\overset{p}{\to}0, \tag{42}$$

then

$$\mathbf{U}_{n,T}\Rightarrow Z,\ \text{with characteristic function}\ \mathbf{E}(\exp(-\frac{1}{2}\widetilde{\mathbf{U}}t^2)), \tag{43}$$

which is a generalized martingale central limit theorem (Hall and Heyde, 1980, Theorem 3.2, Corollary 3.1). We now show that Eq. (42) holds. Write

$$\begin{aligned}
&\sum_{j=1}^{n-1}\mathbf{E}\left\{u^2_{jT/n,(i+1)T/n}\mathbf{1}_{(|u_{jT/n,(j+1)T/n}|>\epsilon)}|\Im_{jT/n}\right\}\\
=&\sum_{j=1}^{n-1}\mathbf{E}\left\{u^2_{jT/n,(j+1)T/n}|\Im_{jT/n}\right\}-\sum_{j=1}^{n-1}\mathbf{E}\left\{u^2_{jT/n,(i+1)T/n}\mathbf{1}_{(|u_{jT/n,(j+1)T/n}|\le\epsilon)}|\Im_{iT/n}\right\}\\
=&\widetilde{\mathbf{U}}^2_{n,T}-\sum_{j=1}^{n-1}\mathbf{E}\left\{u^2_{jT/n,(j+1)T/n}\mathbf{1}_{\left(\left|\mathbf{K}\left(\frac{\sigma^2_{jT/n}-x}{h_{n,T}}\right)\int_{jT/n}^{(j+1)T/n}\int(\xi^\sigma)^4\overline{v}_\sigma(ds,d\xi^\sigma)\right|\le\epsilon\sqrt{h_{n,T}}\right)}|\Im_{jT/n}\right\},
\end{aligned} \tag{44}$$

but $\mathbf{K}\left(\frac{\sigma^2_{jT/n}-x}{h_{n,T}}\right)\int_{jT/n}^{(j+1)T/n}\int(\xi^\sigma)^4\,\overline{v}_\sigma(ds,d\xi^\sigma) = O_p(\sqrt{\Delta_{n,T}})$. Hence, the indicator converges in probability to 1 and, given boundedness of $\widetilde{\mathbf{U}}$, Eq. (44) converges in probability to 0 (as $\Delta_{n,T}\to 0$). Thus, by Eq. (43), $\mathbf{U}_{n,T} \underset{\Delta_{n,T}\to 0}{\Rightarrow} W_{\widetilde{\mathbf{U}}}$, where W denotes Brownian motion. In addition,

$$\frac{\mathbf{U}_{n,T}}{\sqrt{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^n \mathbf{K}\left(\frac{\sigma^2_{jT/n}-x}{h_{n,T}}\right)}} \underset{\Delta_{n,T}\to 0}{\Rightarrow} W_{\frac{\widetilde{\mathbf{U}}}{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{\sigma^2_{s-}-x}{h_{n,T}}\right)ds}}$$

Now, by the ratio-limit theorem for Harris recurrent processes (see, e.g., Revuz and Yor, 1994, Theorem 3.12),

$$\begin{aligned}
&\frac{\widetilde{\mathbf{U}}}{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{\sigma^2_{s-}-x}{h_{n,T}}\right)ds} \\
&\underset{T\to\infty,\ h_{n,T}\to 0}{\overset{p}{\to}} \frac{\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}^2\left(\frac{a-x}{h_{n,T}}\right)\lambda(a)\mathbf{E}((\xi^\sigma)^8)s(a)da}{\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{a-x}{h_{n,T}}\right)s(a)da} \\
&= \frac{\int_{-\infty}^{\infty}\mathbf{K}^2(u)\,\lambda(uh_{n,T}+x)\mathbf{E}((\xi^\sigma)^8)s(uh_{n,T}+x)du}{\int_{-\infty}^{\infty}\mathbf{K}(u)\,s(uh_{n,T}+x)du} \\
&= \left(\int_{-\infty}^{\infty}\mathbf{K}^2(u)\,du\right)\lambda(x)\mathbf{E}((\xi^\sigma)^8).
\end{aligned}$$

Finally, $W_{\frac{\widetilde{\mathbf{U}}}{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{\sigma^2_{s-}-x}{h_{n,T}}\right)ds}} \underset{\Delta_{n,T}\to 0,\ T\to\infty,\ h_{n,T}\to 0}{\Rightarrow} \left(\sqrt{\mathbf{K}_2\lambda(x)\mathbf{E}((\xi^\sigma)^8)}\right)Z$, where Z is a standard normal random variable, by Skorohod embedding (see, e.g., Van Zanten, 2000, Theorem 4.1). Hence,

$$\left(\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\right)\left\{\widehat{\theta}^4(x)-\theta^4(x)\right\} \underset{\Delta_{n,T}\to 0,\ T\to\infty,\ h_{n,T}\to 0}{\Rightarrow} \left(\sqrt{\mathbf{K}_2\lambda(x)\mathbf{E}((\xi^\sigma)^8)}\right)Z.$$

It then follows that

$$\left(\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}\right)\left\{\widehat{\lambda}_{\sigma^2}(x)-\lambda_{\sigma^2}(x)\right\} \underset{\Delta_{n,T}\to 0,\ T\to\infty,\ h_{n,T}\to 0}{\Rightarrow} \left(\frac{\sqrt{\mathbf{K}_2\lambda(x)\mathbf{E}((\xi^\sigma)^8)}}{24\mu_\xi^4}\right)Z, \tag{45}$$

since, as we will show below, $\widehat{\mu}_\xi-\mu_\xi$ converges at a faster rate than $\widehat{\theta}^4(x)-\theta^4(x)$ and, therefore, has an estimation error which does not affect the resulting limiting distribution. This proves the stated result for $\widehat{\lambda}_{\sigma^2}(x)$. We now turn to $\widehat{\mu}_\xi$. Write the estimation error as

$$\begin{aligned}
\overline{T}(\widehat{\mu}_\xi-\mu_\xi) &= \frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\left(\frac{\widehat{\theta}^4(\sigma^2_{i\overline{T}/\overline{n}})}{4\widehat{\theta}^3(\sigma^2_{i\overline{T}/\overline{n}})}-\frac{\theta^4(\sigma^2_{i\overline{T}/\overline{n}})}{4\theta^3(\sigma^2_{i\overline{T}/\overline{n}})}\right)+o_p(1) \\
&= \frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\left(\underbrace{\frac{1}{4\theta^3(\sigma^2_{i\overline{T}/\overline{n}})}\left(\widehat{\theta}^4(\sigma^2_{i\overline{T}/\overline{n}})-\theta^4(\sigma^2_{i\overline{T}/\overline{n}})\right)}_{A_{n,\overline{n},T,i}}-\underbrace{\frac{\theta^4(\sigma^2_{i\overline{T}/\overline{n}})}{4\left(\theta^3(\sigma^2_{i\overline{T}/\overline{n}})\right)^2}\left(\widehat{\theta}^3(\sigma^2_{i\overline{T}/\overline{n}})-\theta^3(\sigma^2_{i\overline{T}/\overline{n}})\right)}_{B_{n,\overline{n},T,i}}\right) \qquad (46)\\
&\quad +o_p(1). \qquad (47)
\end{aligned}$$

Again, the dominating term in the limiting distribution of $\widehat{\theta}^4(\sigma^2_{i\overline{T}/\overline{n}})-\theta^4(\sigma^2_{i\overline{T}/\overline{n}})$ is $\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{jT/n}-\sigma^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)\int_{jT/n}^{(j+1)T/n}(\xi^\sigma)^4\overline{v}_\sigma(ds,d\xi^\sigma)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^n\mathbf{K}\left(\frac{\sigma^2_{jT/n}-\sigma^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)}$.

Write

$$\mathbf{A}_{n,\overline{n},T}=\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}A_{n,\overline{n},T,i}=\sum_{j=1}^{n-1}w(\sigma^2_{jT/n})\int_{j\overline{T}/n}^{(j+1)\overline{T}/n}(\xi^\sigma)^4\,\overline{v}_\sigma(ds,d\xi^\sigma),$$

where $w(\sigma^2_{jT/n})=\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\frac{\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{\sigma^2_{jT/n}-\sigma^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)}{4\theta^3(\sigma^2_{i\overline{T}/\overline{n}})\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^n\mathbf{K}\left(\frac{\sigma^2_{jT/n}-\sigma^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)\right)}$. Given previous results, $\mathbf{A}_{n,\overline{n},T}$ is an average of asymptotically normal random variables. Repeated applications of the occupation time formula yield its limiting variance:

$$\begin{aligned}
&\mathbf{A}^2_{n,\overline{n},T}\\
=\;&\Delta_{n,T}\sum_{j=1}^{n-1}w^2(\sigma^2_{jT/n})\lambda(\sigma^2_{jT/n})\mathbf{E}((\xi^\sigma)^8)+o_p(1)\\
=\;&\Delta_{n,T}\sum_{j=1}^{n-1}\left(\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\frac{\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{\sigma^2_{jT/n}-\sigma^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)}{4\theta^3(\sigma^2_{i\overline{T}/\overline{n}})\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^n\mathbf{K}\left(\frac{\sigma^2_{jT/n}-\sigma^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)\right)}\right)^2\lambda(\sigma^2_{jT/n})\mathbf{E}((\xi^\sigma)^8)+o_p(1)\\
=\;&\int_0^T\left(\int_0^{\overline{T}}\frac{\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{\sigma^2_{s-}-\sigma^2_{u-}}{h_{n,T}}\right)}{4\theta^3(\sigma^2_{u-})\left(\frac{1}{h_{n,T}}\int_0^T\mathbf{K}\left(\frac{\sigma^2_{e-}-\sigma^2_{u-}}{h_{n,T}}\right)de\right)}du\right)^2\lambda(\sigma^2_{s-})\mathbf{E}((\xi^\sigma)^8)ds+o_p(1)\\
=\;&\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{b-a}{h_{n,T}}\right)}{\frac{4\theta^3(a)}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{c-a}{h_{n,T}}\right)\overline{L}_{\sigma^2}(T,c)dc}\overline{L}_{\sigma^2}(\overline{T},a)da\right)^2\lambda(b)\mathbf{E}((\xi^\sigma)^8)\overline{L}_{\sigma^2}(T,b)db+o_p(1)\\
=\;&\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{\mathbf{K}(f)}{4\theta^3(b-fh_{n,T})\int_{-\infty}^{\infty}\mathbf{K}(u)\,\overline{L}_{\sigma^2}(T,uh+b-fh_{n,T})du}\overline{L}_{\sigma^2}(\overline{T},b-fh_{n,T})df\right)^2\lambda(b)\mathbf{E}((\xi^\sigma)^8)\overline{L}_{\sigma^2}(T,b)db\\
&+o_p(1)\\
=\;&\int_{-\infty}^{\infty}\frac{1}{16(\theta^3(b))^2}\left(\frac{\left(\overline{L}_{\sigma^2}(\overline{T},b)\right)^2}{\left(\overline{L}_{\sigma^2}(T,b)\right)^2}\right)\lambda(b)\mathbf{E}((\xi^\sigma)^8)\overline{L}_{\sigma^2}(T,b)db+o_p(1)\\
=\;&\widetilde{\mathbf{A}}+o_p(1).
\end{aligned}$$

Thus,

$$\left(\widetilde{\mathbf{A}}\right)^{-1/2}(\mathbf{A}_{n,\overline{n},T})\underset{\Delta_{n,T}\to0,\;T\to\infty,\;h_{n,T}\to0}{\Rightarrow}Z,\tag{48}$$

Similarly, the dominating term in the limiting distribution of $\widehat{\theta}^3(\sigma^2_{i\overline{T}/\overline{n}})-\theta^3(\sigma^2_{i\overline{T}/\overline{n}})$ is $\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{jT/n}-\sigma^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)\int_{j\overline{T}/n}^{(j+1)\overline{T}/n}(\xi^\sigma)^3\overline{v}_\sigma(ds,d\xi^\sigma)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^n\mathbf{K}\left(\frac{\sigma^2_{jT/n}-\sigma^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)}$.

In this case, write

$$\mathbf{B}_{n,\overline{n},T}=\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}B_{n,\overline{n},T,i}=\sum_{j=1}^{n-1}w(\sigma^2_{jT/n})\int_{jT/n}^{(j+1)T/n}(\xi^\sigma)^3\,\overline{v}_\sigma(ds,d\xi^\sigma),$$

where $w(\sigma^2_{jT/n})=\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\frac{\theta^4(\sigma^2_{i\overline{T}/\overline{n}})}{4\left(\theta^3(\sigma^2_{i\overline{T}/\overline{n}})\right)^2}\left(\frac{\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{\sigma^2_{jT/n}-\sigma^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^n\mathbf{K}\left(\frac{\sigma^2_{jT/n}-\sigma^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)\right)}\right)$. Following the same steps leading to Eq. (48) we obtain:

$$\left(\widetilde{\mathbf{B}}\right)^{-1/2}(\mathbf{B}_{n,\overline{n},T})\underset{\Delta_{n,T}\to0,\;T\to\infty,\;h_{n,T}\to0}{\Rightarrow}Z,$$

where

$$\widetilde{\mathbf{B}} = \int_{-\infty}^{\infty} \frac{\left(\theta^4(b)\right)^2}{16\left(\theta^3(b)\right)^4} \left(\frac{\left(\overline{L}_{\sigma^2}(\overline{T},b)\right)^2}{\left(\overline{L}_{\sigma^2}(T,b)\right)^2} \right) \lambda(b)\mathbf{E}((\xi^\sigma)^6)\overline{L}_{\sigma^2}(T,b)db.$$

Finally, the asymptotic covariance between $\mathbf{A}_{n,\overline{n},T}$ and $\mathbf{B}_{n,\overline{n},T}$ can be expressed as

$$\widetilde{\mathbf{C}} = -\int_{-\infty}^{\infty} \frac{\left(\theta^4(b)\right)}{16\left(\theta^3(b)\right)^3} \left(\frac{\left(\overline{L}_{\sigma^2}(\overline{T},b)\right)^2}{\left(\overline{L}_{\sigma^2}(T,b)\right)^2} \right) \lambda(b)\mathbf{E}((\xi^\sigma)^7)\overline{L}_{\sigma^2}(T,b)db.$$

In sum,

$$\left(\widetilde{\mathbf{A}} + 2\widetilde{\mathbf{C}} + \widetilde{\mathbf{B}}\right)^{-1/2} \overline{T}\left\{\widehat{\mu}_\xi - \mu_\xi\right\} \underset{\Delta_{n,T}\to 0,\ T\to\infty,\ h_{n,T}\to 0}{\Rightarrow} Z, \tag{49}$$

which proves the result as stated in the main text. We conclude by noting that

$$\left(\widehat{\mathbf{D}}\right)^{-1/2} \overline{T}\left\{\widehat{\mu}_\xi - \widehat{\mu}_\xi\right\} \underset{\Delta_{n,T}\to 0,\ T\to\infty,\ h_{n,T}\to 0}{\Rightarrow} Z,$$

where

$$\widehat{\mathbf{D}} = \frac{h_{n,T}}{h_{\overline{n},\overline{T}}} \frac{\left(\Delta_{\overline{n},\overline{T}}\right)^2}{\Delta_{n,T}} \sum_{i=1}^{\overline{n}} \left(\frac{\widehat{\theta}^8}{16\left(\widehat{\theta}^3\right)^2} - \frac{2\widehat{\theta}^4\widehat{\theta}^7}{16\left(\widehat{\theta}^3\right)^3} + \frac{\left(\widehat{\theta}^4\right)^2\widehat{\theta}^6}{16\left(\widehat{\theta}^3\right)^4} \right) \frac{\sum_{j=1}^{\overline{n}} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{jT/n} - \widetilde{\sigma}^2_{i\overline{T}/\overline{n}}}{h_{n,\overline{T}}}\right)}{\sum_{j=1}^{n} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{jT/n} - \widetilde{\sigma}^2_{i\overline{T}/\overline{n}}}{h_{n,T}}\right)}$$

(as in Remark 13) since $\left(\widetilde{\mathbf{A}} + 2\widetilde{\mathbf{C}} + \widetilde{\mathbf{B}}\right)/\widehat{\mathbf{D}} \xrightarrow{p} 1$. Also, from Eq. (49), the rate of convergence of $\widehat{\mu}_\xi$ to μ_ξ is effectively $O_p(\left(\overline{L}_{\sigma^2}(T,x)\right)^{1/2})$. This (semiparametric) rate is of course faster than the (nonparametric) rate of convergence of $\widehat{\theta}^j(x)$ to $\theta^j(x)$ $\forall j$, namely $O_p(\left(h_{n,T}\overline{L}_{\sigma^2}(T,x)\right)^{1/2})$, thereby justifying the asymptotic negligibility of the second term in Eq. (41). ■

Proof of Theorem 6. The result follows from using the method of proof of Theorem 2 combined with Lemma B.3. ■

Proof of Theorem 7. For brevity, we only consider the case $f(\sigma^2) = \sigma^2$, $\xi^\sigma \sim \exp(\mu_\xi)$. The case $f(\sigma^2) = \log\sigma^2$, $\xi^\sigma \sim \mathbf{N}(0,\sigma_\xi^2)$ can of course be handled analogously. We start by separating measurement error as in the proof of Theorem 2 and write

$$\begin{aligned}
\widehat{C}(\sigma^2) &= \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{iT/n} - \sigma^2}{h_{n,T}}\right) \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right) \left(\widetilde{\sigma}^2_{(i+1)T/n} - \widetilde{\sigma}^2_{iT/n}\right)}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}\left(\frac{\widetilde{\sigma}^2_{iT/n} - \sigma^2}{h_{n,T}}\right)} \\
&= \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\sigma^2_{iT/n} - \sigma^2}{h_{n,T}}\right) \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right) \left(\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n}\right)}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}\left(\frac{\sigma^2_{iT/n} - \sigma^2}{h_{n,T}}\right)} + \mathbf{R}_0,
\end{aligned}$$

where $\mathbf{R}_0 \to 0$, under our assumed bandwidth conditions. Now notice that Ito's lemma yields

$$\begin{aligned}
&\left(\log\left(p_{(i+1)T/n}\right) - \log\left(p_{iT/n}\right)\right)\left(\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n}\right) = \int_{iT/n}^{(i+1)T/n} \left(\sigma^2_{s^-} - \sigma^2_{iT/n}\right)\mu(\sigma^2_{s^-})ds \\
&+ \int_{iT/n}^{(i+1)T/n} \left(\sigma^2_{s^-} - \sigma^2_{iT/n}\right)\sigma_{s^-}dW^r_s + \int_{iT/n}^{(i+1)T/n} \int \left(\sigma^2_{s^-} - \sigma^2_{iT/n}\right)\psi v_r(ds,d\psi) \\
&+ \int_{iT/n}^{(i+1)T/n} \left(\log(p_{s^-}) - \log\left(p_{iT/n}\right)\right) m(\sigma^2_{s^-})ds + \int_{iT/n}^{(i+1)T/n} \left(\log(p_{s^-}) - \log\left(p_{iT/n}\right)\right) \Lambda(\sigma^2_{s^-})dW^\sigma_s \\
&+ \int_{iT/n}^{(i+1)T/n} \int \left(\log(p_{s^-}) - \log\left(p_{iT/n}\right)\right) \xi^\sigma v_\sigma(ds,d\xi^\sigma) \\
&+ \sum_{i\Delta_{n,T}\le s\le(i+1)\Delta_{n,T}} \left(\Delta\log\left(p_s\right)\Delta\sigma^2_s\right) + \int_{iT/n}^{(i+1)T/n} \rho(\sigma^2_{s^-})\sigma_{s^-}\Lambda(\sigma^2_{s^-})ds,
\end{aligned}$$

where $\Delta\log p_s = \log p_s - \log p_{s^-}$ and $\Delta\sigma^2_s = \sigma^2_s - \sigma^2_{s^-}$. For convenience, in what follows we compensate the random measures $v_r(ds,d\psi)$, $v_\sigma(ds,d\xi^\sigma)$ and, consequently, write $m(\sigma^2_{s^-}) + \lambda_\sigma(\sigma^2_{s^-})\mathbf{E}[\xi^\sigma]$ in place of $m(\sigma^2_{s^-})$ and $\widetilde{\mu}(\sigma^2_{s^-}) = \mu(\sigma^2_{s^-}) +$

$\lambda_r(\sigma^2_{s-})\mathbf{E}[\psi]$ in place of $\mu(\sigma^2_{s-})$. In other words, we add and subtract the conditional first moment of the jump size to render the discontinuous components martingale difference sequences. With this, one can show (see below for an explicit derivation of the rates of convergence) that

$$\mathbf{R}_1 = \underbrace{\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\left[\int_{iT/n}^{(i+1)T/n}\left(\sigma^2_{s-}-\sigma^2_{iT/n}\right)\sigma_{s-}dW^r_s\right]}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)}}_{\mathbf{R}_{1,a}} + \underbrace{\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\left[\int_{iT/n}^{(i+1)T/n}\left(\log(p_{s-})-\log(p_{iT/n})\right)\Lambda(\sigma^2_{s-})dW^\sigma_s\right]}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)}}_{\mathbf{R}_{1,b}} = O_p\left(\sqrt{\frac{\Delta_{n,T}}{h_{n,T}\overline{L}_{\sigma^2}(T,\sigma^2)}}\right) \quad (50)$$

and

$$\mathbf{R}_2 = \underbrace{\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\left[\int_{iT/n}^{(i+1)T/n}\int\left(\sigma^2_{s-}-\sigma^2_{iT/n}\right)\psi\overline{v}_r(ds,d\psi)\right]}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)}}_{\mathbf{R}_{2,a}} + \underbrace{\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\left[\int_{iT/n}^{(i+1)T/n}\int\left(\log(p_{s-})-\log(p_{iT/n})\right)\xi^\sigma\overline{v}_\sigma(ds,d\xi^\sigma)\right]}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)}}_{\mathbf{R}_{2,b}} = O_p\left(\sqrt{\frac{\Delta_{n,T}}{h_{n,T}\overline{L}_{\sigma^2}(T,\sigma^2)}}\right) \quad (51)$$

Thus, $\mathbf{R}_1+\mathbf{R}_2 \stackrel{p}{\rightarrow} 0$, under our conditions on the bandwidth. Using the process' modulus of continuity and the ratio-limit theorem, write

$$\begin{aligned}\mathbf{R}_3 &= \frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\left[\int_{iT/n}^{(i+1)T/n}\left(\sigma^2_{s-}-\sigma^2_{iT/n}\right)\left(\mu(\sigma^2_{s-})+\lambda_r(\sigma^2_{s-})\mathbf{E}[\psi]\right)ds\right]}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)} \\ &\quad + \frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\left[\int_{iT/n}^{(i+1)T/n}\left(\log(p_{s-})-\log(p_{iT/n})\right)\left(m(\sigma^2_{s-})+\lambda_\sigma(\sigma^2_{s-})\mathbf{E}[\xi^\sigma]\right)ds\right]}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)} \\ &\leq O_{a.s.}\left(\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} \stackrel{a.s.}{\rightarrow} 0.\end{aligned}$$

Notice also that,

$$\mathbf{R}_4 = \frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\left[\sum_{iT/n\leq s\leq(i+1)T/n}\Delta\log p_s\Delta\sigma^2_s\right]}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)} = o_{a.s.}(1),$$

since, by independence of the jumps, the probability of common jumps is zero (see, e.g., Cont and Tankov, 2004, Proposition 5.3). Finally,

$$\widehat{C}(\sigma^2) = \mathbf{R}_0 + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3 + \mathbf{R}_4 + \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\sigma^2_{iT/n} - \sigma^2}{h_{n,T}}\right)\left[\int_{iT/n}^{(i+1)T/n} \rho(\sigma^2_{s-})\sigma_{s-}\Lambda(\sigma^2_{s-})ds\right]}{\Delta_{n,T}\sum_{i=1}^{n} \mathbf{K}\left(\frac{\sigma^2_{iT/n} - \sigma^2}{h_{n,T}}\right)} \xrightarrow{p} \rho(\sigma^2)\sqrt{\sigma^2}\Lambda(\sigma^2),$$

by the ratio-limit theorem. This implies, by Slutsky's theorem, that $\widehat{\rho}(\sigma^2) = \frac{\widehat{C}(\sigma^2)}{\sqrt{\sigma^2}\widehat{\Lambda}(\sigma^2)} \xrightarrow{p} \rho(\sigma^2)$, provided $\widehat{\Lambda}(\sigma^2)$ is a consistent estimate of $\Lambda(\sigma^2)$, as implied by Theorem 5. We now turn to weak convergence and clarify the origin of the rates in Eq. (50) and Eq. (51). Start with $\mathbf{R}_{1,b}$. Express the conditional variance of the standardized numerator $\widetilde{\mathbf{R}}^{num}_{1,b} = \left(\sqrt{\frac{h_{n,T}}{\Delta_{n,T}}}\right)\mathbf{R}^{num}_{1,b}$ as

$$\left(\widetilde{\mathbf{R}}^{num}_{1,b}\right)^2 = \frac{1}{h_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n} - \sigma^2}{h_{n,T}}\right)\mathbf{E}\left\{\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-} - \log p_{iT/n}\right)^2\Lambda^2(\sigma^2_{s-})ds|\Im_{i,n}\right\}.$$

Noting that

$$\begin{aligned}
& \left(\log p_{s-} - \log p_{iT/n}\right)^2 \\
= \; & \underbrace{2\int_{iT/n}^{s}\left(\log p_{u-} - \log p_{iT/n}\right)\widetilde{\mu}(\sigma^2_{u-})du + 2\int_{iT/n}^{s}\left(\log p_{u-} - \log p_{iT/n}\right)\sigma_{u-}dW^r_u}_{\mathbf{A}} \\
& \underbrace{+2\int_{iT/n}^{s}\int(\log p_{u-} - \log p_{iT/n})\psi\overline{v}(du,d\psi)}_{\mathbf{C}} + \int_{iT/n}^{s}\sigma^2_{u-}du + \int_{iT/n}^{s}\lambda_r(\sigma^2_{u-})\mathbf{E}(\psi^2)du + \underbrace{\int_{iT/n}^{s}\int\psi^2\overline{v}(du,d\psi)}_{\mathbf{B}},
\end{aligned}$$

write

$$\begin{aligned}
& \left(\widetilde{\mathbf{R}}^{num}_{1,b}\right)^2 \\
= \; & \frac{1}{h_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n} - \sigma^2}{h_{n,T}}\right)\mathbf{E}\left\{\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-} - \log p_{iT/n}\right)^2\Lambda^2(\sigma^2_{s-})ds|\Im_{i,n}\right\} \\
= \; & \frac{1}{h_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n} - \sigma^2}{h_{n,T}}\right)\mathbf{E}\left\{\int_{iT/n}^{(i+1)T/n}\frac{\Lambda^2(\sigma^2_{s-})}{(\sigma^2_{s-} + \lambda_r(\sigma^2_{s-})\mathbf{E}(\psi^2))}\{\log p_{s-} - \log p_{iT/n}\}d\{\log p - \log p_{iT/n}\}_s|\Im_{i,n}\right\} \\
& +o_p(1) \\
= \; & \frac{1}{h_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n} - \sigma^2}{h_{n,T}}\right)\frac{\Lambda^2(\sigma^2_{iT/n} + o_{a.s.}(1))}{(\sigma^2_{iT/n} + \lambda_r(\sigma^2_{iT/n})\mathbf{E}(\psi^2) + o_{a.s.}(1))}\mathbf{E}\left\{\frac{\{\log p_{s-} - \log p_{iT/n}\}^2}{2}|_{iT/n}^{(i+1)T/n}|\Im_{i,n}\right\} \\
& +o_p(1) \\
= \; & \frac{1}{h_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n} - \sigma^2}{h_{n,T}}\right)\frac{\Lambda^2(\sigma^2_{iT/n} + o_{a.s.}(1))\mathbf{E}\left(\left(\int_{iT/n}^{(i+1)T/n}\sigma^2_{s-} + \lambda_r(\sigma^2_{s-})\mathbf{E}(\psi^2)\right)^2|\Im_{i,n}\right)}{2(\sigma^2_{iT/n} + \lambda_r(\sigma^2_{iT/n})\mathbf{E}(\psi^2) + o_{a.s.}(1))} \\
& +o_p(1) \\
& \xrightarrow[\Delta_{n,T}\to 0]{p} \frac{1}{2}\frac{1}{h_{n,T}}\int_0^T\mathbf{K}^2\left(\frac{\sigma^2_{s-} - \sigma^2}{h_{n,T}}\right)\Lambda^2(\sigma^2_{s-})\left[\sigma^2_{s-} + \lambda_r(\sigma^2_{s-})\mathbf{E}(\psi^2)\right]ds,
\end{aligned} \tag{52}$$

where the term $\{\log p_{s-} - \log p_{iT/n}\}$ defines the measure $\int_{iT/n}^{s}\left(\sigma^2_{u-} + \lambda_r(\sigma^2_{u-})\mathbf{E}(\psi^2)\right)du$. We now justify the $o_p(1)$ term above by showing that the neglected terms (in A, B, and C) are negligible. Write $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$. The term $\mathbf{A}_1$ is clearly not problematic being of higher order than $\{\log p_{s-} - \log p_{iT/n}\}$. The terms $\mathbf{C}$ and $\mathbf{A}_2$ are of higher order than $\mathbf{B}$. We therefore focus on $\mathbf{B}$. Using integration by parts (see, e.g., Protter, 2005), we obtain

$$\frac{1}{h_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\mathbf{E}\left(\int_{iT/n}^{(i+1)T/n}\mathbf{B}\Lambda^2(\sigma^2_{s-})ds|\Im_{i,n}\right)$$
$$= \frac{1}{h_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\mathbf{E}\left(\left(\int_{iT/n}^{(i+1)T/n}\Lambda^2(\sigma^2_{s-})ds\right)\left(\int_{iT/n}^{(i+1)T/n}\int\psi^2\overline{v}(ds,d\psi)\right)|\Im_{i,n}\right)$$
$$-\frac{1}{h_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\mathbf{E}\left(\left[\int_{iT/n}^{(i+1)T/n}\left(\int_{iT/n}^{s}\Lambda^2(\sigma^2_{u-})du\right)\int\psi^2\overline{v}(ds,d\psi)\right]|\Im_{i,n}\right)$$
$$= \frac{1}{h_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\mathbf{E}\left(\left(\int_{iT/n}^{(i+1)T/n}\Lambda^2(\sigma^2_{s-})ds\right)\left(\int_{iT/n}^{(i+1)T/n}\int\psi^2\overline{v}(ds,d\psi)\right)|\Im_{i,n}\right).$$

since $\mathbf{E}\left(\left[\int_{iT/n}^{(i+1)T/n}\left(\int_{iT/n}^{s}\Lambda^2(\sigma^2_{u-})du\right)\int\psi^2\overline{v}(ds,d\psi)\right]|\Im_{i,n}\right)=0$. Now write

$$\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\mathbf{E}\left(\left(\frac{\int_{iT/n}^{(i+1)T/n}\Lambda^2(\sigma^2_{s-})ds}{\Delta_{n,T}}\right)\left(\int_{iT/n}^{(i+1)T/n}\int\psi^2\overline{v}(ds,d\psi)\right)|\Im_{i,n}\right)$$
$$= \frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\mathbf{E}\left(\left(\Lambda^2(\sigma^2_{iT/n})+o_{a.s.}\left(\Delta_{n,T}^{1/2}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)\right)\left(\sum_{iT/n\le s\le(i+1)T/n}(\Delta\log p_s)^2\right)|\Im_{i,n}\right)$$
$$-\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\mathbf{E}\left(\left(\Lambda^2(\sigma^2_{iT/n})+o_{a.s.}\left(\Delta_{n,T}^{1/2}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)\right)\left(\int_{iT/n}^{(i+1)T/n}\mathbf{E}(\psi^2)\lambda_r(\sigma^2_{s-})ds\right)|\Im_{i,n}\right)$$
$$= o_p\left(\Delta_{n,T}^{1/2}\log\left(\frac{1}{\Delta_{n,T}}\right)\right)O_p\left(\frac{1}{h_{n,T}}\int_0^T\mathbf{K}^2\left(\frac{\sigma^2_{s-}-\sigma^2}{h_{n,T}}\right)\mathbf{E}(\psi^2)\lambda_r(\sigma^2_{s--})ds\right)\underset{\Delta_{n,T}\to 0}{\to}0,$$

since $\mathbf{E}\left(\sum_{iT/n\le s\le(i+1)T/n}(\Delta\log p_s)^2|\Im_{i,n}\right)-\mathbf{E}\left(\int_{iT/n}^{(i+1)T/n}\mathbf{E}(\psi^2)\lambda_r(\sigma^2_{s-})ds|\Im_{i,n}\right)=0.$

Finally, note that $\left\{\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-}-\log p_{iT/n}\right)\Lambda(\sigma^2_{s-})dW^\sigma_s,\Im_{i,n}\right\}$ is a square-integrable martingale difference array. The same methods as in the proof of Theorem 5 show that a conditional Lindeberg condition is satisfied for $\widetilde{\mathbf{R}}^{num}_{1,b}=\sqrt{\frac{h_{n,T}}{\Delta_{n,T}}}\mathbf{R}^{num}_{1,b}$. Hence, using a generalized martingale convergence theorem and Skorohod embedding as in the proof of Theorem 5:

$$\left(\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\right)\frac{\mathbf{R}^{num}_{1,b}}{\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\underset{\Delta_{n,T}\to 0,\ T\to\infty,\ h_{n,T}\to 0}{\Rightarrow}\left(\sqrt{\mathbf{K}_2\frac{1}{2}\left(\sigma^2+\lambda_r(\sigma^2)\mathbf{E}[\psi^2]\right)\Lambda^2(\sigma^2)}\right)Z.$$

Similarly,

$$\left(\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\right)\frac{\mathbf{R}^{num}_{1,a}}{\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\underset{\Delta_{n,T}\to 0,\ T\to\infty,\ h_{n,T}\to 0}{\Rightarrow}\left(\sqrt{\mathbf{K}_2\frac{1}{2}\left(\lambda_\sigma(\sigma^2)\mathbf{E}[\xi^2]+\Lambda^2(\sigma^2)\right)\sigma^2}\right)Z.$$

In addition, the asymptotic conditional covariance term can be expressed as:

$$asycov\left(\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\frac{\mathbf{R}^{num}_{1,b}}{\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)},\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\frac{\mathbf{R}^{num}_{1,a}}{\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\right)$$
$$\underset{\Delta_{n,T}\to 0,\ T\to\infty,\ h_{n,T}\to 0}{\overset{p}{\to}}\frac{1}{2}\mathbf{K}_2\rho^2(\sigma^2)\sigma^2\Lambda^2(\sigma^2).$$

This result yields:

$$\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\mathbf{R}_1$$
$$\underset{\Delta_{n,T}\to 0,\ T\to\infty,\ h_{n,T}\to 0}{\Rightarrow}\sqrt{\mathbf{K}_2\left(\frac{1}{2}\lambda_\sigma(\sigma^2)\mathbf{E}[\xi^2]\sigma^2+\frac{1}{2}\lambda_r(\sigma^2)\mathbf{E}[\psi^2]\Lambda^2(\sigma^2)+\left(1+\rho^2(\sigma^2)\right)\sigma^2\Lambda^2(\sigma^2)\right)}Z.$$

Using the same reasoning (combined with the independence of the jumps), we obtain:

$$\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\mathbf{R}_2 \underset{\Delta_{n,T}\to 0,\ T\to\infty,\ h_{n,T}\to 0}{\Rightarrow} \sqrt{\mathbf{K}_2\frac{1}{2}\left(\Lambda^2(\sigma^2)+\lambda_\sigma(\sigma^2)\mathbf{E}[\xi^2]\right)\lambda_r(\sigma^2)\mathbf{E}[\psi^2]+\left(\sigma^2+\lambda_r(\sigma^2)\mathbf{E}[\psi^2]\right)\lambda_\sigma(\sigma^2)\mathbf{E}[\xi^2]\big)}Z.$$

As earlier, the bandwidth conditions guarantee that $\mathbf{R}_0 = o_{a.s.}(\mathbf{R}_1+\mathbf{R}_2)$ and, by independence of the jumps, $\mathbf{R}_3 = o_{a.s.}(\mathbf{R}_1+\mathbf{R}_2)$. Finally, consider the bias term

$$\Gamma_C(\sigma^2) = \frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\int_{iT/n}^{(i+1)T/n}(\rho(\sigma^2_{s-})\sigma_{s-}\Lambda(\sigma^2_{s-}))ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)} - \rho(\sigma^2)\sqrt{\sigma^2}\Lambda(\sigma^2).$$

Define $z(.) = \sqrt{.}\rho(.)\Lambda(.)$. An application of the ratio-limit theorem combined with a second-order Taylor expansion leads to

$$\Gamma_C(\sigma^2) = h^2_{n,T}\left(\frac{1}{2}z''(\sigma^2)+z'(\sigma^2)\frac{s'(\sigma^2)}{s(\sigma^2)}\right)+o_{a.s.}(1).$$

Thus,

$$\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\left\{\widehat{C}(\sigma^2)-z(\sigma^2)-\Gamma_C(\sigma^2)\right\} \underset{\Delta_{n,T}\to 0,\ T\to\infty,\ h_{n,T}\to 0}{\Rightarrow} \mathbf{N}\left(0,\mathbf{K}_2\theta^C(\sigma^2)\right),$$

with

$$\theta^C(\sigma^2) = \Lambda^2(\sigma^2)\sigma^2(1+\rho^2(\sigma^2))+\sigma^2\lambda_\sigma(\sigma^2)\mathbf{E}[\xi^2]+\Lambda^2(\sigma^2)\lambda_r(\sigma^2)\mathbf{E}[\psi^2]+\lambda_\sigma(\sigma^2)\mathbf{E}[\xi^2]\lambda_r(\sigma^2)\mathbf{E}[\psi^2].$$

Now, by the delta method,

$$\begin{aligned}
&\left\{\widehat{\rho}(\sigma^2)-\rho(\sigma^2)\right\} \\
=\ & \left\{\frac{\widehat{C}(\sigma^2)}{\sigma\widehat{\Lambda}(\sigma^2)}-\frac{C(\sigma^2)}{\sigma\Lambda(\sigma^2)}\right\} \\
=\ & \frac{1}{\sigma\Lambda(\sigma^2)}\left\{\widehat{C}(\sigma^2)-C(\sigma^2)\right\}-\frac{C(\sigma^2)}{2\sigma\Lambda^3(\sigma^2)}\left\{\widehat{\Lambda}(\sigma^2)-\Lambda(\sigma^2)\right\}+o_p(1) \\
=\ & \frac{1}{\sigma\Lambda(\sigma^2)}\left\{O(h^2_{n,T})+O_p\left(\sqrt{\frac{\Delta_{n,T}}{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}}\right)\right\}-\frac{\rho(\sigma^2)}{2\Lambda^2(\sigma^2)}\left\{O(h^2_{n,T})+O_p\left(\sqrt{\frac{1}{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}}\right)\right\}+o_p(1).
\end{aligned}$$

Hence, the asymptotic distribution of the leverage function is driven by $\widehat{\Lambda}^2(\sigma^2)-\Lambda^2(\sigma^2)$ and has variance

$$Asyvar(\widehat{\rho}(\sigma^2)) = \frac{\rho^2(\sigma^2)}{4\Lambda^4(\sigma)}\left(Asivar\left(\widehat{\Lambda}^2(\sigma^2)\right)\right) = \frac{\rho^2(\sigma^2)}{4\Lambda^4(\sigma)}\left(\frac{\lambda_\sigma(x)\mathbf{E}\left(\left((\xi^\sigma)^2-\frac{1}{12\mu_\xi^2}(\xi^\sigma)^4\right)^2\right)}{h_{n,T}\overline{L}_{\sigma^2}(T,\sigma^2)}\right)+o_p(1).$$

Finally, under one of our assumptions on the bandwidth sequence (namely, $\lim_{n,T\to\infty} h^5_{n,T}v(T) = o(1)$), $\widehat{\rho}(\sigma^2)-\rho(\sigma^2)$ has a vanishing asymptotic bias. ■

References

[1] Andersen, T.G., L. Benzoni, and J. Lund (2002). Towards an empirical foundation for continuous-time equity return models. *Journal of Finance* 57, 1047-1091.

[2] Andersen, T.G. and T. Bollerslev (2008). Deutsche mark-dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance* 53, 219-265.

[3] Andersen, T.G., T. Bollerslev, and F.X. Diebold (2004). Parametric and non-parametric volatility measurements. In *Handbook of Financial Econometrics* (L.P. Hansen and Y. Ait-Sahalia, Eds.) Elsevier Science, New York, forthcoming.

[4] Andersen, T.G., T. Bollerslev, and F.X. Diebold (2007). Roughing it up: Including jump components in the measurement, modelling, and forecasting of return volatility. *Review of Economics and Statistics,* forthcoming.

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

[6] Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *Journal of Finance* 52, 2003-2049.

[7] Bali, T. and L. Peng (2006). Is there a risk-return trade-off? Evidence from high-frequency data. *Journal of Applied Econometrics* 21, 1169-1198.

[8] Bandi, F.M. and G. Moloche (2004). On the functional estimation of multivariate diffusion models. *Working paper.*

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

[10] Bandi, F.M. and B. Perron (2008). Long-run risk-return trade-offs. *Journal of Econometrics*, 143 349-374.

[11] Bandi, F.M. and P.C.B. Phillips (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 241-283.

[12] Bandi, F.M. and P.C.B. Phillips (2004). Nonstationary continuous-time models. In *Handbook of Financial Econometrics* (L.P. Hansen and Y. Ait-Sahalia, Eds.) Elsevier Science, New York, forthcoming.

[13] Bandi, F. M. and R. Reno' (2008). Nonparametric Stochastic Volatility. Available at SSRN: http://ssrn.com/abstract=1158438.

[14] Bandi, F.M., and J.R. Russell (2005). Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *Working paper.*

[15] Bandi, F.M. and J.R. Russell (2007). Volatility. In *Handbook of Financial Engineering* (V. Linetski and J. Birge, Eds.) Elsevier Science, New York.

[16] Bandi, F.M., and J.R. Russell (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies* 75, 339-369.

[17] Barndorff-Nielsen, O.E., P. Hansen, A. Lunde and N. Shephard (2006). Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. *Econometrica*, forthcoming.

[18] Barndorff-Nielsen, O.E., and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*, 253-280.

[19] Barndorff-Nielsen, O.E., and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1-37.

[20] Barndorff-Nielsen, O.E., and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4, 1-30.

[21] Barndorff-Nielsen, O.E., and N. Shephard (2007). Variation, jumps, market frictions and high-frequency data in financial econometrics. In R. Blundell, P. Torsten and W. K. Newey (Eds.) *Advances in Economics and Econometrics. Theory and Applications*. Ninth World Congress. Econometric Society Monographs, Cambridge University Press.

[22] Barndorff-Nielsen, O.E., N. Shephard, and M. Winkel (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic processes and their applications* 116, 796-806.

[23] Bates, D. (2000). Post-'87 Crash fears in S&P500 futures options. *Journal of Econometrics* 94, 181-238.

[24] Bhattacharya, R.N. (1980). Criteria for recurrence and existence of invariant measures for multidimensional diffusions. *Annals of Probability* 6, 541-553.

[25] Bollerslev, T. and H. Zhao (2002). Estimating stochastic volatility diffusions using conditional moments of integrated volatility. *Journal of Econometrics* 109, 33-65.

[26] Chacko, G. and L. Viceira (2001). Spectral GMM estimation of continuous-time processes. *Journal of Econometrics* 116, 259-292.

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

[28] Christensen, K. and M. Podolskij (2007). Asymptotic theory for range-based estimation of integrated volatility of a continuous semimartingale. *Working paper.*

[29] Comte, F., Genon-Catalot, V. and Y. Rozenholc (2007). Nonparametric estimation for a stochastic volatility model. *Working paper.*

[30] Cont, R. and P. Tankov (2004). *Financial Modelling with Jump Processes*. Chapman & Hall, CRC.

[31] Corradi, V. and W. Distaso (2006). Semiparametric comparison of stochastic volatility models using realized measures. *Review of Economic Studies* 73, 635-667.

[32] Corsi, F., D. Pirino, and R. Renò (2008). Volatility forecasting: The jumps do matter. *Working paper.*

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

[34] Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *Journal of Finance* 1269-1300.

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

[36] Hansen, P.R., and A. Lunde (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics* 4, 525-554.

[37] Harvey, A.C. and N. Shephard (1996). The estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* 14, 429-434.

[38] Hasminskii, R.Z. (1960). Ergodic properties of recurrent diffusion processes and stabilization of the Cauchy problem for parabolic equations. *Probability Theory and Applications* 5, 179-196.

[39] 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.

[40] Huang, X. and G. Tauchen (2005). The relative contribution of jumps to total price variation. *Journal of Financial Econometrics* 3, 456-499.

[41] Jacod, J. (1994). Limit of random measures associated with the increments of a Brownian semimartingale. *Working paper.*

[42] Jacod, J. and P. Protter (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability 26,* 267-307.

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

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

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

[46] Jones, C. (2003). The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics* 116, 181-224.

[47] Kanaya, S. and D. Kristensen (2008). Estimation of stochastic volatility models by nonparametric filtering. *Working paper.*

[48] Kristensen, D. (2007). Nonparametric filtering of the realized spot volatility: a kernel-based approach. *Econometric Theory,* forthcoming.

[49] Kristensen, D. (2008). Nonparametric estimation and misspecification testing of diffusion models. *Working paper.*

[50] Malliavin, P. and M.E. Mancino (2008). A Fourier transform method for nonparametric estimation of volatility. *Annals of Statistics,* forthcoming.

[51] Mancini, C. (2007). Nonparametric threshold estimation for models with stochastic diffusion coefficients and jumps. *Scandinavian Journal of Statistics*, forthcoming.

[52] Mancini, C., and R. Renò (2006). Threshold estimation of jump-diffusion models and interest rate modeling. *Working paper.*

[53] McAleer, M., and M. Medeiros (2008). Realized volatility: A review. *Econometric Reviews,* forthcoming.

[54] Moloche, G. (2002). Local nonparametric estimation of scalar diffusions. *Working paper.*

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

[56] Protter, P. (1995). *Stochastic integration and differential equations.* Springer, New York.

[57] Renò, R. (2006). Nonparametric estimation of stochastic volatility models. *Economics Letters* 90, 390-395.

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

[59] Revuz, D. and M. Yor (1998). *Continuous Martingales and Brownian Motion.* Springer-Verlag.

[60] Shephard, N. (2005). *Stochastic volatility: Selected Readings.* Edited volume. Oxford University Press.

[61] Shephard, N. (2006). Stochastic volatility. In *New Pelgrave Dictionary of Economics* (S. Durlauf and L. Blume, Eds.) 2nd Edition.

[62] Todorov, V. (2007). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *Working paper.*

[63] Van Zanten, J.H. (2000). A multivariate central limit theorem for continuous local martingales. *Statistics and Probability Letters* 50, 229-235.

[64] Wee, I-S. (2000). Recurrence and transience for jump-diffusion models. *Stochastic Analysis and Applications* 18, 1055-1064.

[65] Woerner, J. (2006). Power and multipower variation: inference for high frequency data. In A. N. Shiryaev, M. do Rosario Grossinho, P. Oliviera, and M. Esquivel (Eds.). *Proceedings of the International Conference on Stochastic Finance 2004*, 343-364. Berlin: Springer-Verlag.

[66] Xu, K.-L. (2006). Empirical likelihood based inference for nonlinear diffusions. *Working paper.*

[67] Xu, K.-L. (2007). Re-weighted functional estimation of diffusion models. *Working paper.*

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

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

Figure 1(a)
Variance drift estimates

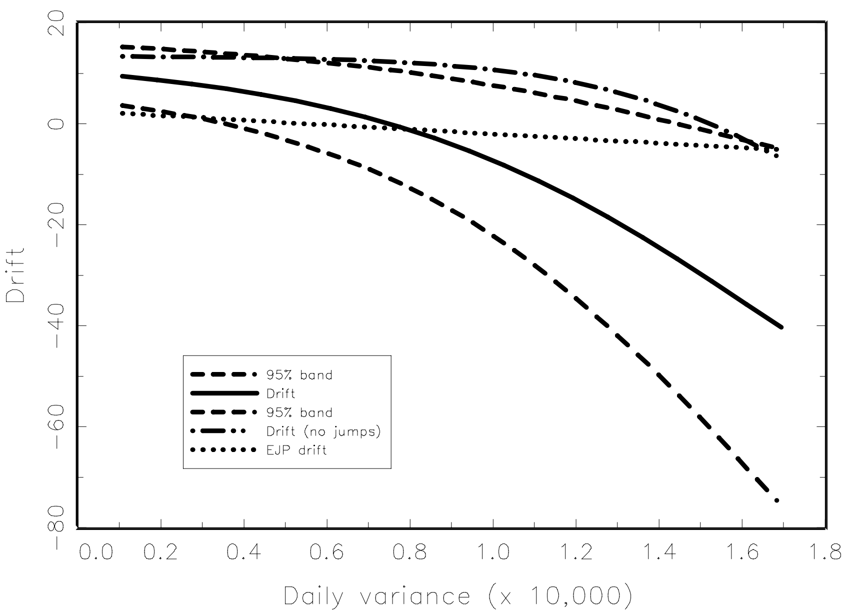

Figure 1(b)
Variance diffusion estimates

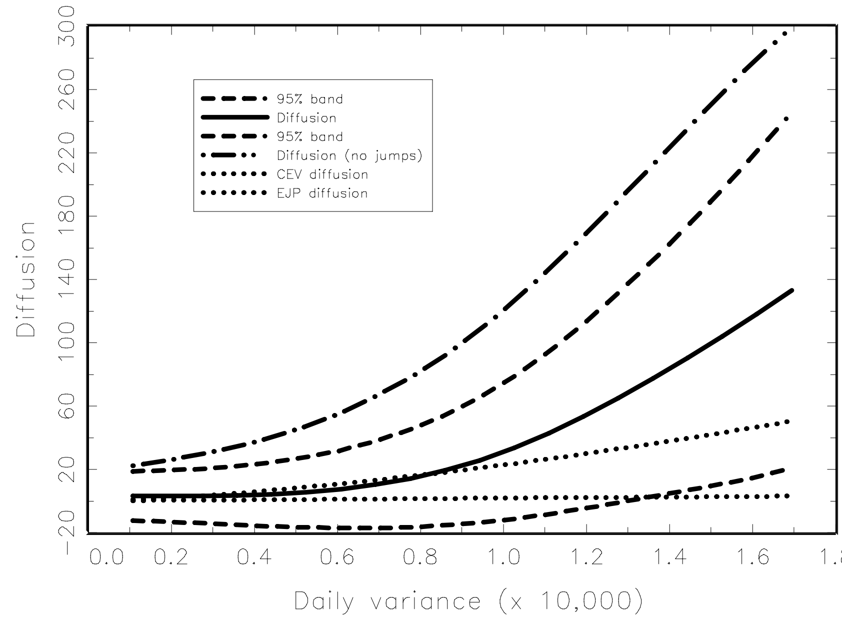

Figure 1(c)
Variance jump intensity

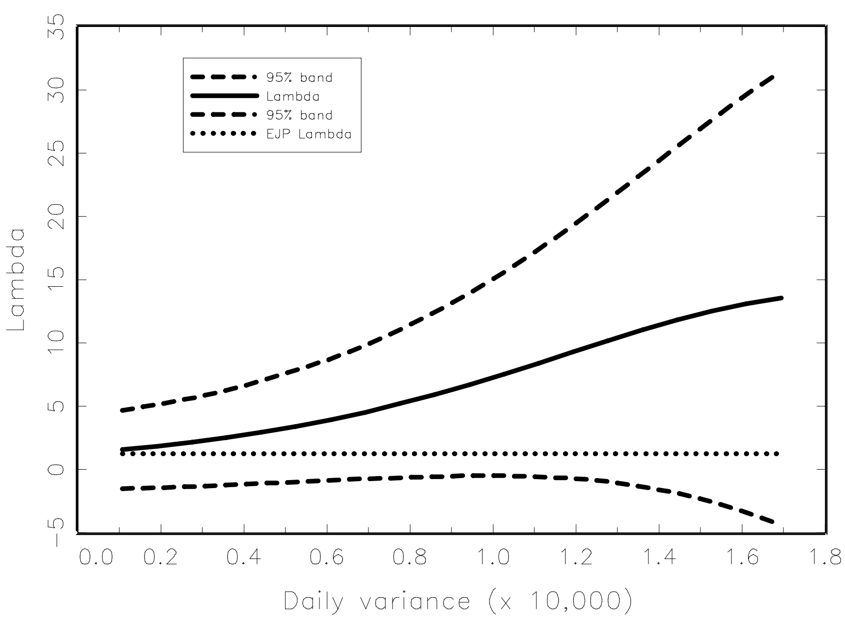

Figure 1(d)
Variance expected jump size

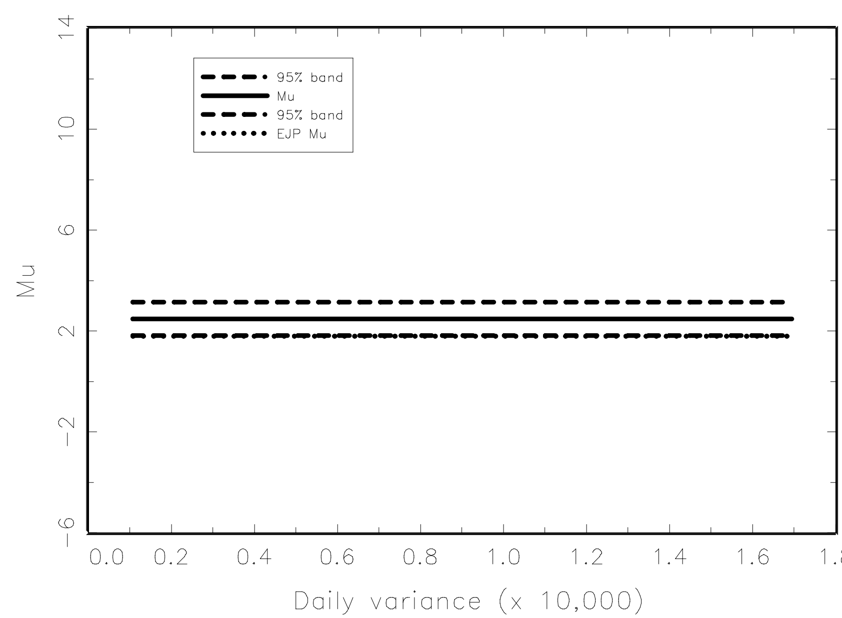

Figure 2(a)
Price drift estimates

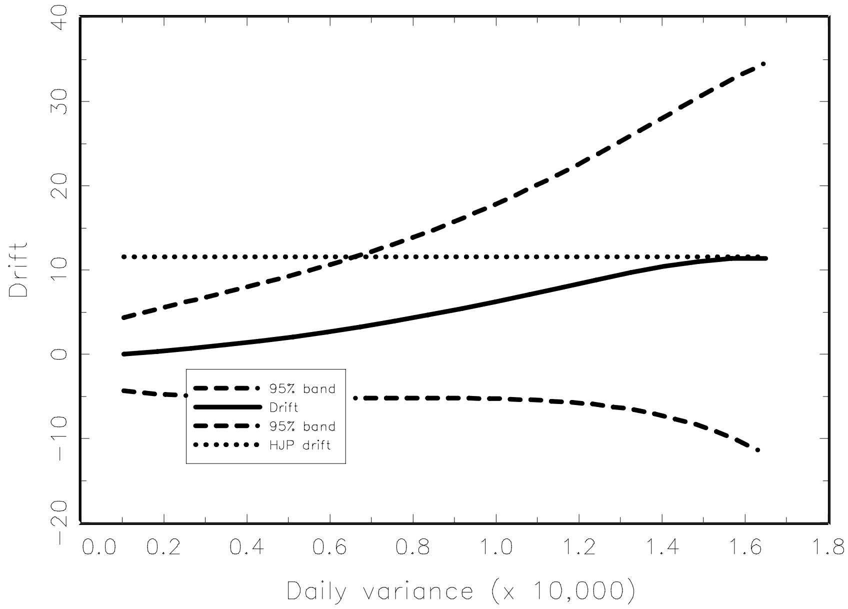

Figure 2(b)
Leverage estimates

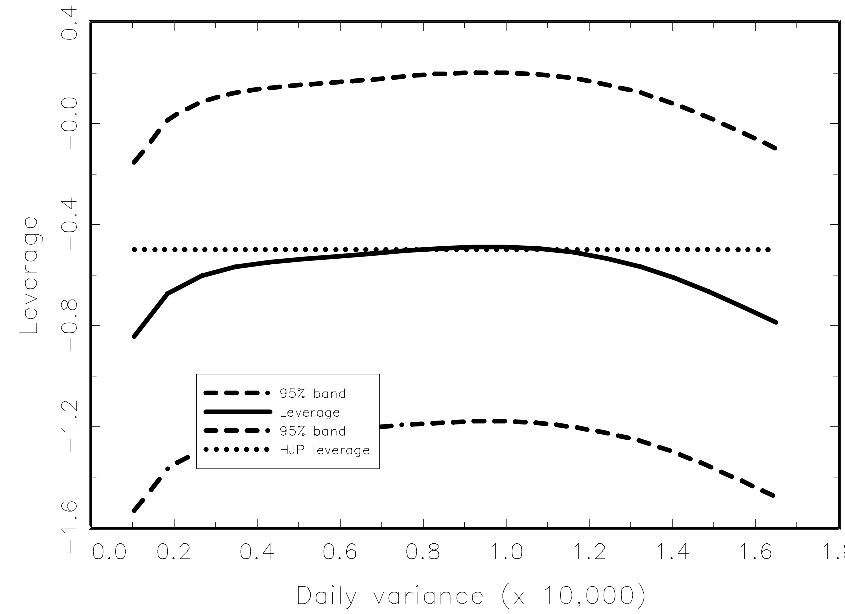

Figure 2(c)
Price jump intensity

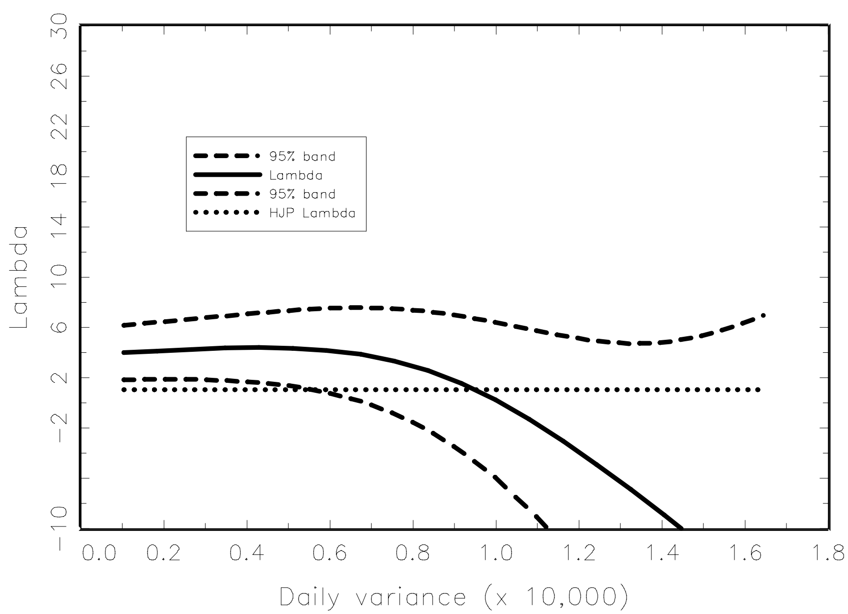

Figure 2(d)
Price jump standard deviation

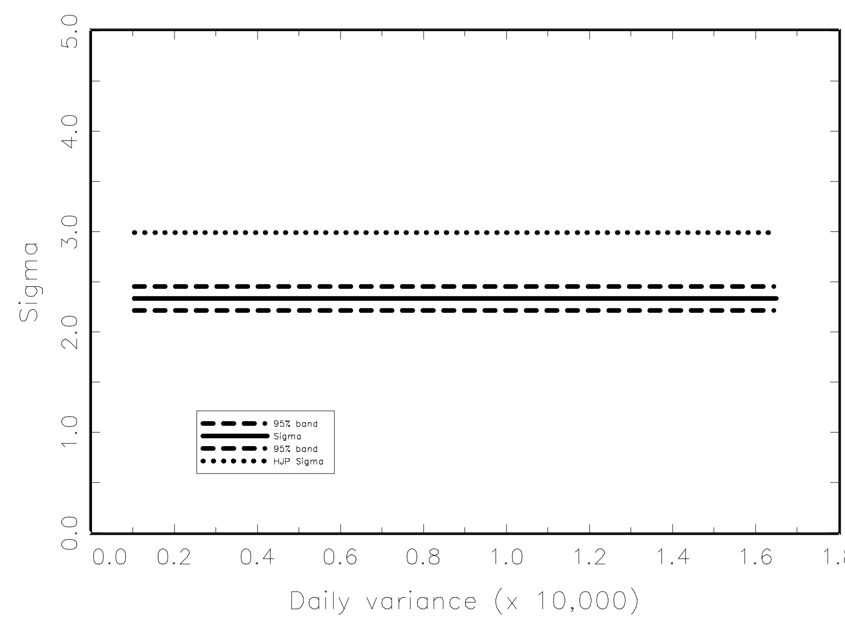

Figure 3(a)
Drift

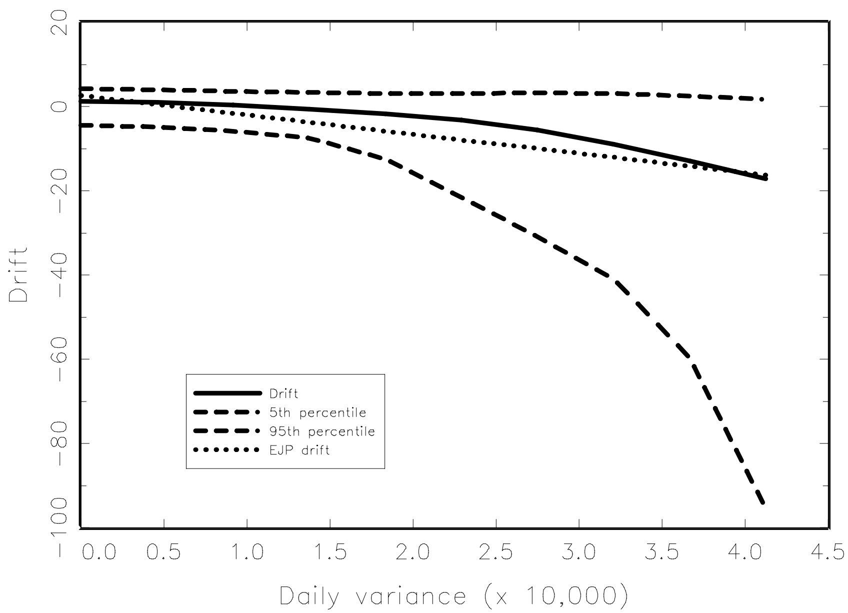

Figure 3(b)
Diffusion

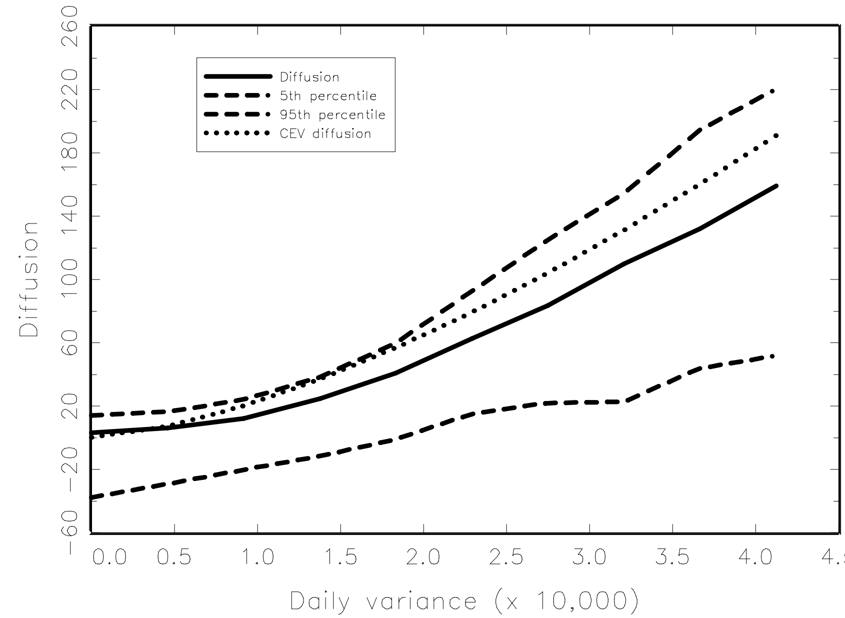

Figure 3(c)
Lambda

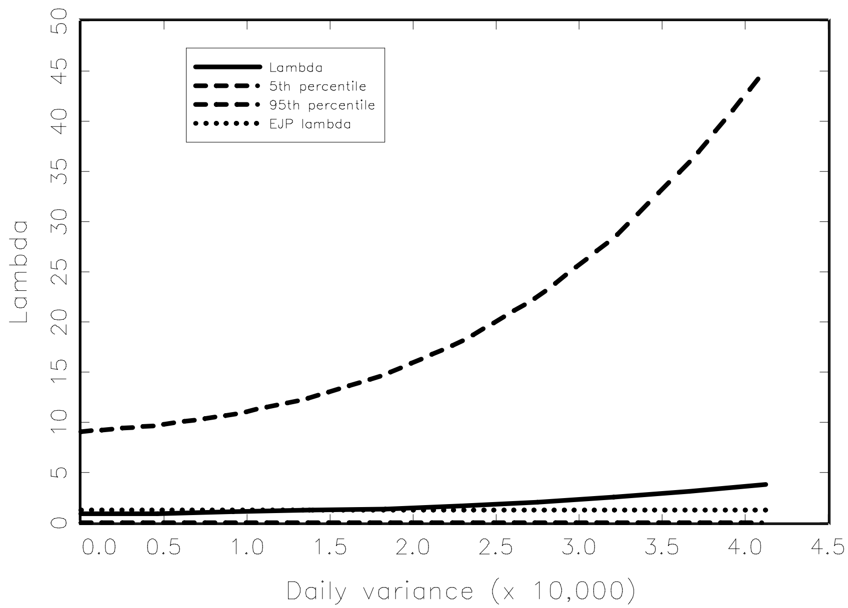

Figure 3(d)
Variance expected jump size

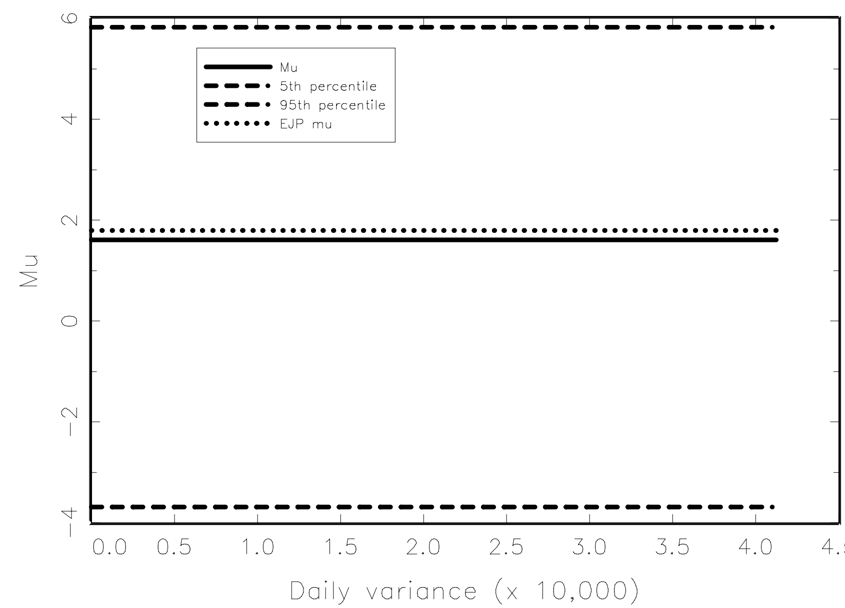

Figure 4(a)
Return drift

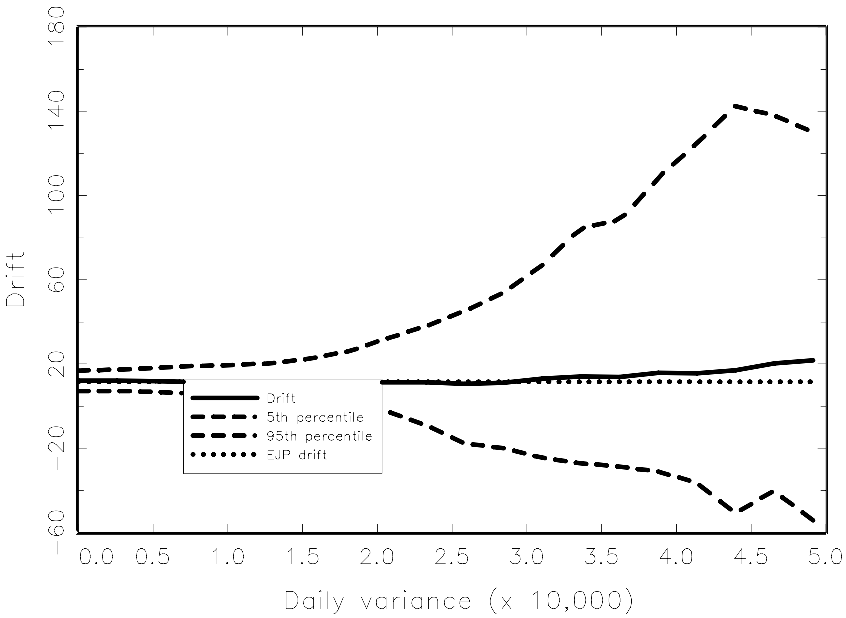

Figure 4(b)
Leverage

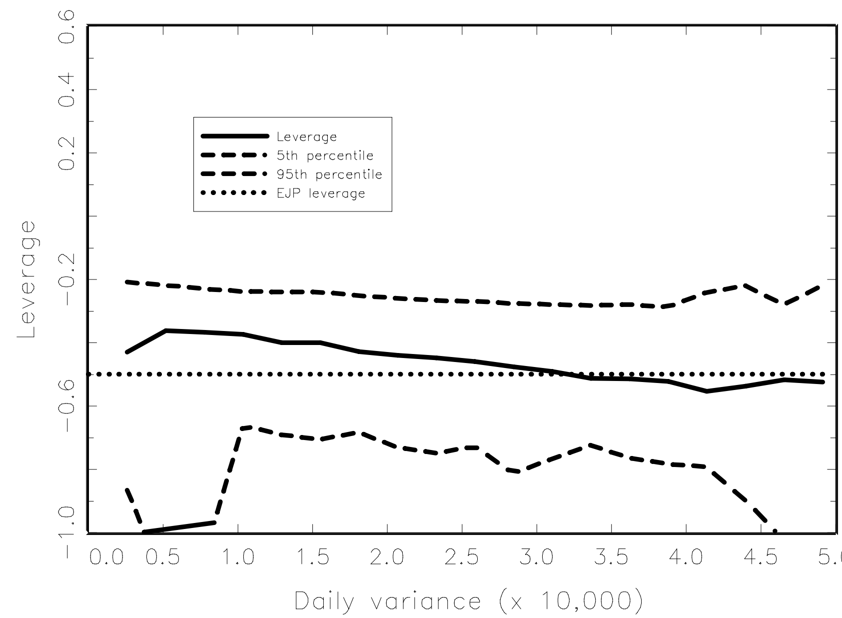

Figure 4(c)
Price jump intensity

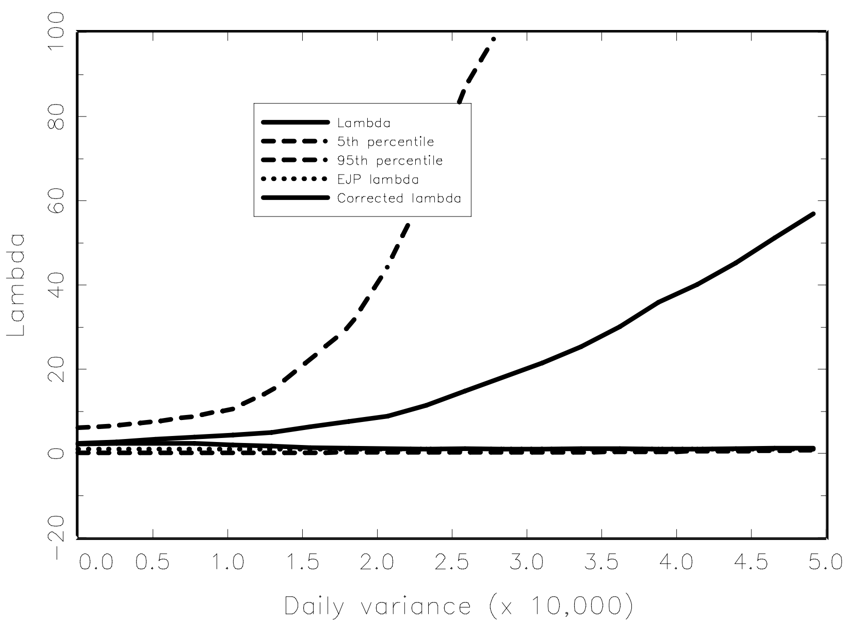

Figure 4(d)
Price jump standard deviation