

# NONPARAMETRIC STOCHASTIC VOLATILITY\*

Federico M. Bandi<sup>†</sup>

Roberto Reno<sup>‡</sup>

*First draft: October 2007*

*This draft: June 2009*

## 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), the Far Eastern and South Asian Meetings of the Econometric Society (Singapore, July 16-18, 2008), the Hitotsubashi University's International Conference on "High Frequency Data Analysis in Financial Markets" (Tokyo, October 25-26), the LSE Conference on "Recent Advances in High-Frequency Financial Econometrics" (London, November 15, 2008), and the North American Winter Meetings of the Econometric Society (San Francisco, January 3-5, 2009) for discussions. We are especially grateful to Valentina Corradi, Fulvio Corsi, Cecilia Mancini, and Eric Renault for useful comments and suggestions. The usual disclaimers apply.

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

<sup>‡</sup>*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:

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

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

where  $\{dW_t^r, dW_t^\sigma\} = \{\rho(\sigma_t^2)dW_t^1 + \sqrt{1 - \rho^2(\sigma_t^2)}dW_t^2, dW_t^1\}$  with  $-1 \leq \rho(\cdot) \leq 1$ ,  $\{W_t^1, W_t^2\}$  are standard Brownian motions,  $\{J_t^r, J_t^\sigma\}$  are Poisson jump processes independent of each other and independent of  $\{W_t^1, W_t^2\}$  with intensities  $\lambda^r(\cdot)$  and  $\lambda_{f(\cdot)}^\sigma(\cdot)$ , and  $\mu(\cdot)$ ,  $m_{f(\cdot)}(\cdot)$ ,  $\Lambda_{f(\cdot)}(\cdot)$ , and  $\rho(\cdot)$  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.,  $\lambda_{f(\cdot)}^\sigma(\cdot)$ ,  $m_{f(\cdot)}(\cdot)$ ,  $\Lambda_{f(\cdot)}(\cdot)$  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 et 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.<sup>1</sup> Second, differently from much existing work on stochastic volatility modelling, we avoid imposing tight (possibly affine) parametric structures on  $\lambda_{f(\cdot)}^\sigma(\cdot)$ ,  $m_{f(\cdot)}(\cdot)$ , and  $\Lambda_{f(\cdot)}(\cdot)$ . 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

---

<sup>1</sup>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).

estimates. However, as we illustrate below, extensions to alternative functional estimation methods are rather 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 (2007). 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. By simulation only, Reno' (2006) evaluates the performance of nonparametric drift and diffusion estimates for stochastic volatility models without jumps applied to high-frequency variance measures in the absence of noise. This simulation paper can be viewed as providing empirical support for the theory proposed in the present work, at least for a sub-class of our broader family of models and high-frequency spot variance estimates. Finally, Reno' (2008) provides a nonparametric theory of dif-

fusion function estimation for stochastic volatility models without jumps and noise using, contrary to the present work, a single frequency to estimate spot variance by virtue of the Fourier estimator of Malliavin and Mancino (2002), which is suitable for daily data.

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.,

$$\theta^j(x) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \mathbf{E} \left[ (f(\sigma_{t+\Delta}^2) - f(\sigma_t^2))^j \mid \sigma_t^2 = x \right] \quad j = 1, \dots, \quad (3)$$

by virtue of

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

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

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

**Assumption 1.**  $\mathbf{K}(\cdot)$  is a bounded, continuously-differentiable, symmetric, and nonnegative function whose derivative  $\mathbf{K}'(\cdot)$  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

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

and

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

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  $\eta$  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  $T \rightarrow \infty$  (long span) and  $n \rightarrow \infty$  with  $\Delta_{n,T} \rightarrow 0$  (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.,  $k \rightarrow \infty$  with  $\phi_{n,T} \rightarrow 0$ ). 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  $\tilde{\sigma}_{iT/n}^2$  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  $\tau$  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., 2008) with a kernel function  $g(\cdot)$  satisfying  $g'(0) = 0$  and  $g'(1) = 0$ , for instance. We find that, if the number of autocovariances  $q$  is such that  $q = \tau \left(\frac{k}{\phi_{n,T}}\right)^{1/2}$  and  $\phi_{n,T} = k^\theta$  with  $-1 < \theta < 0$ , then  $\alpha = \frac{1}{4}$ ,  $\beta = \frac{1}{4}$ ,  $a \neq 0$ ,  $b = 0$ , and  $\eta = \frac{3}{4}$ . 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  $\alpha$ ,  $\beta$ ,  $a$ ,  $b$ , and  $\eta$ , for each estimator.

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

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

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

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

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

$$k^{\frac{1}{2}} \left( \frac{\widehat{V}_{iT/n}}{\phi_{n,T}} - \sigma_{iT/n}^2 \right) \xrightarrow[k \rightarrow \infty, \phi_{n,T} \rightarrow 0]{} MN \left( 0, 2\sigma_{iT/n}^4 \right), \quad (6)$$

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,  $\widetilde{\sigma}_{iT/n}^2 = \frac{\widehat{V}_{iT/n}}{\phi_{n,T}}$  converges in probability to  $\sigma_{iT/n}^2$  (at speed  $\phi_{n,T}^\beta k^\alpha$  if  $\phi_{n,T}^\beta k^\alpha \rightarrow \infty$ ) provided  $k^\alpha \phi_{n,T}^\beta \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} \rightarrow 0$ . 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

$$\phi_{n,T}^\beta k^\alpha \left( \frac{\widehat{V}_{iT/n}}{\phi_{n,T}} - \sigma_{iT/n}^2 \right) \xrightarrow[k \rightarrow \infty, \phi_{n,T} \rightarrow 0]{} MN \left( 0, a \left( \sigma_{iT/n}^4 \right)^\eta + b \right). \quad (7)$$

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 may be  $k^{1/10}$ . Provided the kernel function  $g(\cdot)$  satisfies  $g'(0) = 0$  and  $g'(1) = 0$ , the spot variance estimates constructed using realized kernels may converge at rate  $k^{1/6}$ .

**Remark 4. (More on spot volatility estimation.)** As emphasized above, the quantity  $\widetilde{\sigma}_{iT/n}^2 = \frac{\widehat{V}_{iT/n}}{\phi_{n,T}}$  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  $m(\cdot)$  and  $\Lambda(\cdot)$  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.,

$$\widehat{\sigma}_{i/n}^2 = \frac{1}{h} \sum_{j=1}^{nk} \mathbf{K} \left( \frac{j - i/n}{h} \right) r_j^2 \quad i = 1, \dots, n,$$

where  $\mathbf{K}(\cdot)$  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  $nhk \rightarrow \infty$ ,  $\widehat{\sigma}_{i/n}^2$  converges to the spot variance at  $i/n$  with a standard nonparametric speed  $\sqrt{nhk}$ . Specifically, the weak convergence result

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

holds if, in addition,  $nk h^{1+2\gamma} \rightarrow 0$ , where  $0 < \gamma \leq 1$  is the order of smoothness of  $\sigma_t^2$  (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 as in Kristensen (2007). Write

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

where  $\mathbf{1}_{\{\cdot\}}$  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  $\widehat{\sigma}_{i/n}^2$ . This said, our derived asymptotic distribution in Eq. (6) and the asymptotic distribution in Eq. (8) have to be consistent with each other when  $\mathbf{K}(\cdot) = \mathbf{1}_{\{\cdot\}}$ . In other words, if  $nk\phi_{n,1} \rightarrow \infty$  and  $nk\phi_{n,1}^{1+2\gamma} \rightarrow 0$ ,

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

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  $2\sigma_{i/n}^4$ , 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.

Importantly, the properties of (either discontinuous or, more efficient, smooth) 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. Our treatment also allows us to draw from the recent literature on integrated variance estimation in terms of finite sample adjustments required for more accurate empirical implementations. The latter have been advocated by Bandi and Russell (2008, 2009) and are employed in Section 8 below.

**Intuition.** We now turn to the logic behind our estimation procedure. Given Remark 2 and 3, the rate of convergence of  $\widehat{\sigma}_{iT/n}^2$  to  $\sigma_{iT/n}^2$  is  $k^\alpha \phi_{n,T}^\beta$  (if  $k^\alpha \phi_{n,T}^\beta \rightarrow \infty$ , of course, and  $k^\alpha \phi_{n,T}^\beta \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} \rightarrow 0$ ), where  $\alpha \in (0, \frac{1}{2}]$  and  $\beta = [0, 1]$ . Loosely speaking, if  $k \rightarrow \infty$  at a fast enough pace as  $h_{n,T} \rightarrow 0$ , 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  $j = 1, \dots$ , 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 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

$$\tilde{\theta}_u^j(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \tilde{\mathbf{K}}_i(x, h_{n,T}) \left[ f(\tilde{\sigma}_{(i+1)T/n}^2) - f(\tilde{\sigma}_{iT/n}^2) \right]^j}{\sum_{i=1}^n \tilde{\mathbf{K}}_i(x, h_{n,T})} \quad j = 1, \dots, \quad (9)$$

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

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

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

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

where  $\tilde{\theta}_u^j(x) = \hat{\alpha}_0(x)$  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  $(\Omega, \mathfrak{F}, P, \{\mathfrak{F}_t\}_{t \geq 0})$  and the compensated  $N$ -dimensional jump-diffusion process  $X_t$  defined as

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

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

$$\begin{aligned} \bar{\nu}(dt, dy) &= N(dt, dy) - \mathbf{E}(N(dt, dy)) \\ &= N(dt, dy) - \tilde{\Pi}(dy) dt \end{aligned}$$

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

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

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

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

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

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

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

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

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  $x \in R^N$ ,

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

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  $\boldsymbol{\mu}(\cdot) = \{\mu(\cdot), m_{f(\cdot)}(\cdot)\}^\top - \{\lambda^\top(\cdot), \lambda_{f(\cdot)}^\sigma(\cdot)\}^\top \circ \{\mathbf{E}[c^\top(\cdot, y^r)], \mathbf{E}[c^\sigma(\cdot, y^\sigma)]\}^\top$ , 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: $dJ_t^r = 0$ and $dJ_t^\sigma = 0$

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  $\hat{\theta}^1(x) \xrightarrow{p} m(x)$  and  $\hat{\theta}^2(x) \xrightarrow{p} \Lambda^2(x)$  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 ( $T = \bar{T}$ ). If  $k, n \rightarrow \infty$  and  $h_{n,\bar{T}}, \phi_{n,\bar{T}} \rightarrow 0$  so that*

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

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

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

where  $\bar{L}_{\sigma^2}(\bar{T}, x)$  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  $\bar{L}_{\sigma^2}(T, x)/T \xrightarrow{p} p(x)$ , 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  $\bar{L}_{\sigma^2}(T, x) \propto v(T)$ , 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  $k, n, T \rightarrow \infty$  and  $h_{n,T}, \phi_{n,T} \rightarrow 0$  so that*

$$\lim_{n, T \rightarrow \infty} h_{n,T} v(T) = \infty, \quad (11)$$

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

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

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

$$\begin{aligned} \lim_{n, T \rightarrow \infty} h_{n, T} v(T) &= \infty, \\ \lim_{n, T \rightarrow \infty} h_{n, T}^5 v(T) &= C_1, \end{aligned} \quad (14)$$

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

then,

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

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  $k, n, T \rightarrow \infty$  and  $h_{n, T}, \phi_{n, T} \rightarrow 0$  so that

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

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

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

then,

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

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  $nh \rightarrow \infty$ , 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  $\Delta_{n,T} \rightarrow 0$ , 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: $dJ_t^r = 0$ and $dJ_t^\sigma \neq 0$

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).<sup>2</sup> 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  $\sigma_t^2$ ),  $\Lambda_{\sigma^2}(\sigma_t^2)$  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  $N_t^\sigma$ ) is constant and, hence, independent of the state (see, also, Andersen et al. (2002) for an affine stochastic volatility model with  $\lambda_{\sigma^2}(\sigma_t^2) = 0^3$ ). Provided the variance drift, diffusion, and intensity satisfy the conditions laid out in Section 3, we leave their functional forms unspecified. Now, notice that

$$\theta^1(x) = m_{\sigma^2}(x) + \mu_\xi \lambda_{\sigma^2}(x) \quad (16)$$

$$\theta^2(x) = \Lambda_{\sigma^2}^2(x) + 2\mu_\xi^2 \lambda_{\sigma^2}(x) \quad (17)$$

$$\theta^3(x) = 6\mu_\xi^3 \lambda_{\sigma^2}(x) \quad (18)$$

$$\theta^4(x) = 24\mu_\xi^4 \lambda_{\sigma^2}(x) \quad (19)$$

...

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

$$\hat{\mu}_\xi = \frac{1}{n} \sum_{i=1}^n \frac{\hat{\theta}^4(\hat{\sigma}_{iT/n}^2)}{4\hat{\theta}^3(\hat{\sigma}_{iT/n}^2)}, \quad (20)$$

$$\hat{\lambda}_{\sigma^2}(x) = \frac{\hat{\theta}^4(x)}{24\hat{\mu}_\xi^4}, \quad (21)$$

$$\hat{\Lambda}_{\sigma^2}^2(x) = \hat{\theta}^2(x) - 2\hat{\mu}_\xi^2 \hat{\lambda}_{\sigma^2}(x), \quad (22)$$

$$\hat{m}_{\sigma^2}(x) = \hat{\theta}^1(x) - \hat{\mu}_\xi \hat{\lambda}_{\sigma^2}(x). \quad (23)$$

<sup>2</sup>A different methodology based on pre-filtering the data with a threshold function is explored in Mancini and Renò (2009).

<sup>3</sup>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.<sup>4</sup> 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

$$\theta^1(x) = m_{\log \sigma^2}(x), \quad (24)$$

$$\theta^2(x) = \Lambda_{\log \sigma^2}^2(x) + \sigma_\xi^2 \lambda_{\log \sigma^2}(x), \quad (25)$$

$$\theta^4(x) = 3\sigma_\xi^4 \lambda_{\log \sigma^2}(x), \quad (26)$$

$$\theta^6(x) = 15\sigma_\xi^6 \lambda_{\log \sigma^2}(x), \quad (27)$$

...

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

$$\widehat{\sigma}_\xi^2 = \frac{1}{n} \sum_{i=1}^n \frac{\widehat{\theta}^6(\widehat{\sigma}_{iT/n}^2)}{5\widehat{\theta}^4(\widehat{\sigma}_{iT/n}^2)}, \quad (28)$$

$$\widehat{\lambda}_{\log \sigma^2}(x) = \frac{\widehat{\theta}^4(x)}{3\widehat{\sigma}_\xi^4}, \quad (29)$$

$$\widehat{\Lambda}_{\log \sigma^2}^2(x) = \widehat{\theta}^2(x) - \widehat{\sigma}_\xi^2 \widehat{\lambda}_{\log \sigma^2}(x), \quad (30)$$

$$\widehat{m}_{\log \sigma^2}(x) = \widehat{\theta}^1(x). \quad (31)$$

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  $k, n, T \rightarrow \infty$  and  $h_{n,T}, \phi_{n,T} \rightarrow 0$  so that*

---

<sup>4</sup>We do so when evaluating the return dynamics in order to achieve more accurate finite sample performance (see Section 10).

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

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

$$\tilde{\theta}^j(x) \xrightarrow{p} \theta^j(x) \quad j \geq 1,$$

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

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

then,

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

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

**Theorem 5. (Variance moments: Weak convergence.)**

*Assume*

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

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

**Generalized Duffie et al.'s model**

*Expected jump size:*

$$(\Upsilon(T))^{-1/2} \bar{T} \{ \hat{\mu}_\xi - \mu_\xi \} \Rightarrow \mathbf{N}(0, 1),$$

where

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

*Jump intensity:*

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

*Diffusive function:*

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

<sup>5</sup>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:

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

**Log-variance model**

Jump standard deviation:

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

where

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

Jump intensity:

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

Diffusive function:

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

Drift function:

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

**Proof.** See Appendix B.

**Remark 12 (The ergodic case)** In the positive recurrent and strictly stationary case,  $v(T) = T$  and  $\frac{\widehat{\bar{L}}_{\sigma^2}(T,x)}{T} \xrightarrow{p} p(x)$ , 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  $\sqrt{h_{n,T}\bar{T}}$ . 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  $\bar{n} = n \rightarrow \infty$  and  $\bar{T} = T \rightarrow \infty$  with  $\Delta_{n,T} = \Delta_{\bar{n},\bar{T}} \rightarrow 0$ . In this case, in effect,

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

and

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

**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:

$$\begin{aligned} & \frac{h_{n,T}}{h_{\bar{n},\bar{T}}} \frac{(\Delta_{\bar{n},\bar{T}})^2}{\Delta_{n,T}} \sum_{i=1}^{\bar{n}} \left( \frac{\hat{\theta}^8}{16(\hat{\theta}^3)^2} - \frac{2\hat{\theta}^4\hat{\theta}^7}{16(\hat{\theta}^3)^3} + \frac{(\hat{\theta}^4)^2\hat{\theta}^6}{16(\hat{\theta}^3)^4} \right) \frac{\sum_{j=1}^{\bar{n}} \mathbf{K} \left( \frac{\hat{\sigma}_{jT/n}^2 - \hat{\sigma}_{i\bar{T}/\bar{n}}^2}{h_{n,T}} \right)}{\sum_{j=1}^n \mathbf{K} \left( \frac{\hat{\sigma}_{jT/n}^2 - \hat{\sigma}_{i\bar{T}/\bar{n}}^2}{h_{n,T}} \right)} \\ & \xrightarrow{p} \int_{-\infty}^{\infty} \lambda_{\sigma^2}(x) \left( \frac{\bar{L}_{\sigma^2}(\bar{T}, x)}{\bar{L}_{\sigma^2}(T, x)} \right) \mathbf{E} \left( \left( \frac{1}{4\theta^3(x)} (\xi^\sigma)^4 - \frac{\theta^4(x)}{4(\theta^3(x))^2} (\xi^\sigma)^3 \right)^2 \right) dx, \\ & \frac{\hat{\theta}^8(x)}{(24)^2 \hat{\mu}_\xi^8} \xrightarrow{p} \frac{\lambda_{\sigma^2}(x) \mathbf{E} \left( (\xi^\sigma)^8 \right)}{(24)^2 \mu_\xi^8}, \\ & \hat{\theta}^4(x) - \frac{2\hat{\theta}^6(x)}{12\hat{\mu}_\xi^2} + \frac{\hat{\theta}^8(x)}{(12)^2 \hat{\mu}_\xi^4} \xrightarrow{p} \lambda_{\sigma^2}(x) \mathbf{E} \left( \left( (\xi^\sigma)^2 - \frac{1}{12\mu_\xi^2} (\xi^\sigma)^4 \right)^2 \right), \end{aligned}$$

and

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

Furthermore,

$$\begin{aligned} & \frac{h_{n,T}}{h_{\bar{n},\bar{T}}} \frac{(\Delta_{\bar{n},\bar{T}})^2}{\Delta_{n,T}} \sum_{i=1}^{\bar{n}} \left( \frac{\hat{\theta}^{12}}{25(\hat{\theta}^4)^2} - \frac{2\hat{\theta}^6\hat{\theta}^{10}}{25(\hat{\theta}^4)^3} + \frac{(\hat{\theta}^6)^2\hat{\theta}^8}{25(\hat{\theta}^4)^4} \right) \frac{\sum_{j=1}^{\bar{n}} \mathbf{K} \left( \frac{\hat{\sigma}_{jT/n}^2 - \hat{\sigma}_{i\bar{T}/\bar{n}}^2}{h_{n,T}} \right)}{\sum_{j=1}^n \mathbf{K} \left( \frac{\hat{\sigma}_{jT/n}^2 - \hat{\sigma}_{i\bar{T}/\bar{n}}^2}{h_{n,T}} \right)} \\ & \xrightarrow{p} \int_{-\infty}^{\infty} \lambda_{\log \sigma^2}(x) \left( \frac{\bar{L}_{\log \sigma^2}(\bar{T}, x)}{\bar{L}_{\log \sigma^2}(T, x)} \right) \mathbf{E} \left( \left( \frac{1}{5\theta^4(x)} (\xi^\sigma)^6 - \frac{\theta^6(x)}{5(\theta^4(x))^2} (\xi^\sigma)^4 \right)^2 \right) dx, \\ & \frac{\hat{\theta}^8(x)}{9\hat{\sigma}_\xi^8} \xrightarrow{p} \frac{\lambda_{\log \sigma^2}(x) \mathbf{E} \left( (\xi^\sigma)^8 \right)}{9\sigma_\xi^8}, \\ & \hat{\theta}^4(x) - \frac{2\hat{\theta}^6(x)}{3\hat{\sigma}_\xi^2} + \frac{\hat{\theta}^8(x)}{9\hat{\sigma}_\xi^4} \xrightarrow{p} \lambda_{\log \sigma^2}(x) \mathbf{E} \left( \left( (\xi^\sigma)^2 - \frac{1}{3\sigma_\xi^2} (\xi^\sigma)^4 \right)^2 \right), \end{aligned}$$

and, of course,

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

in the log-variance case.<sup>6</sup>

<sup>6</sup>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: $dJ_t^r \neq 0$ and $dJ_t^\sigma \neq 0$

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  $\frac{1}{\phi_{n,T}^{1/2}} \left( \frac{1}{k} \log \left( \frac{k}{\phi_{n,T}} \right) \right)^{1/2} \rightarrow 0$  with  $\phi_{n,T} \rightarrow 0$  and  $k \rightarrow \infty$ , the quantity  $\tilde{\sigma}_{iT/n}^2 = \frac{\hat{V}_{iT/n}}{\phi_{n,T}}$ , where  $\hat{V}_{iT/n}$  is bipower variation, is consistent for  $\sigma_{iT/n}^2$  in the presence of jumps in returns and

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

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),<sup>7</sup> 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 $dJ_t^r \neq 0$ , $dJ_t^\sigma \neq 0$ , 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  $\tilde{\sigma}_{iT/n}^2$  (obtained by using bipower variation or alternative identification methods robust to jumps in returns) as well as infinitesimal moment estimates for the return process  $(\hat{\theta}_r^j(\sigma^2))$  with  $j = 1, 2, \dots$ , 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  $\psi \sim \mathbf{N}(0, \sigma_\psi^2)$ ,<sup>8</sup> and employ

$$\hat{\sigma}_\psi^2 = \frac{1}{\bar{n}} \sum_{i=1}^{\bar{n}} \frac{\hat{\theta}_r^6(\tilde{\sigma}_{iT/\bar{n}}^2)}{5\hat{\theta}_r^4(\tilde{\sigma}_{iT/\bar{n}}^2)}, \quad (33)$$

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

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

<sup>7</sup>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.

<sup>8</sup>See, e.g., Eraker et al. (2003).

where,

$$\hat{\theta}_r^j(\sigma^2) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K} \left( \frac{\tilde{\sigma}_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) (\log p_{(i+1)T/n} - \log p_{iT/n})^j}{\sum_{i=1}^n \mathbf{K} \left( \frac{\tilde{\sigma}_{iT/n}^2 - \sigma^2}{h_{n,T}} \right)} \quad j = 1, \dots$$

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

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

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

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

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

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

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

then,

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

with

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

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<sup>9</sup> can be identified as follows:

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

where  $\widehat{\Lambda}_{f(\cdot)}^2(\sigma^2)$  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  $k, n, T \rightarrow \infty$  and  $h_{n,T}, \phi_{n,T} \rightarrow 0$  so that*

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

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

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

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

$$\begin{aligned}\lim_{n,T \rightarrow \infty} \frac{h_{n,T}^5 v(T)}{\Delta_{n,T}} &= C_5, \\ \lim_{n,T \rightarrow \infty} \frac{v(T)}{h_{n,T}} \left( \Delta_{n,T} \log \frac{1}{\Delta_{n,T}} \right)^{1/2} &= 0, \\ \lim_{k,n,T \rightarrow \infty} \frac{Tv(T)^{-1/2} \log(n)}{\Delta_{n,T}^{3/2} h_{n,T}^{1/2} k^\alpha \phi_{n,T}^\beta} + \frac{Tv(T)^{-1/2}}{\Delta_{n,T}^{3/2} h_{n,T}^{1/2}} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} &= 0,\end{aligned}$$

then,

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

<sup>9</sup>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

$$\theta^C(\sigma^2) = \Lambda_{f(\cdot)}^2(\sigma^2)\sigma^2(1 + \rho^2(\sigma^2)) + \sigma^2\lambda_{f(\cdot)}^\sigma(\sigma^2)\mathbf{E}[\xi^2] + \Lambda_{f(\cdot)}^2(\sigma^2)\lambda^r(\sigma^2)\mathbf{E}[\psi^2] + \lambda_{f(\cdot)}^\sigma(\sigma^2)\mathbf{E}[\xi^2]\lambda^r(\sigma^2)\mathbf{E}[\psi^2],$$

and

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

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

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

then

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

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

$$\text{Asyvar} \left( \widehat{\Lambda}_{f(\cdot)}^2(\sigma^2) \right) = \lambda_{\sigma^2}^\sigma(x) \mathbf{E} \left( \left( (\xi^\sigma)^2 - \frac{1}{12\mu_\xi^2} (\xi^\sigma)^4 \right)^2 \right)$$

if  $f(\cdot) = \sigma^2$  or

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

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

**Proof.** See Appendix B.

## 8 High-frequency 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 functional estimates of the infinitesimal moments and parameters driving return and variance evolution in the context of the generalized Duffie et al's model discussed in Section 5.

## 8.1 The variance estimates

As mentioned, we consider three high-frequency variance estimates with various degrees of robustness to market microstructure noise and different economic justifications depending on the assumed return dynamics, i.e., the two-scale estimator of Zhang et al. (2005), the kernel estimator of Barndorff-Nielsen et al. (2006), and the bipower variation estimator of Barndorff-Nielsen and Shephard (2004). After a discussion of the intra-daily price formation process justifying these alternative approaches, we turn to their description. We only focus on the features that are relevant for our analysis and refer the reader to the original articles, as well as to the review papers by Bandi and Russell (2007), Barndorff-Nielsen and Shephard (2007), and McAleer and Medeiros (2008), for further comments.

**Intra-daily price formation.** As earlier, we assume availability of  $k$  price observations in each interval  $[i\Delta_{n,T}, i\Delta_{n,T} + \phi_{n,T}]$  with  $i = 1, \dots, n$ . The intra-daily price formation mechanism is defined as:

$$\log(p_j)^* = \log(p_j) + \eta_j \quad j = 1, \dots, k$$

or, in terms of continuously-compounded returns,

$$\underbrace{\log(p_j)^* - \log(p_{(j-1)})^*}_{r_{j\delta}^*} = \underbrace{\log(p_j) - \log(p_{(j-1)})}_{r_{j\delta}} + \underbrace{\eta_j - \eta_{(j-1)}}_{\varepsilon_{j\delta}},$$

where  $\log(p)$  denotes the *unobservable* true price and  $\eta$  denotes *unobservable* market microstructure noise. We assume that the noise is independent of the true price process and IID mean zero with second moment  $\mathbf{E}(\eta^2) = \sigma_\eta^2$  (or, equivalently,  $\mathbf{E}(\varepsilon^2) = 2\sigma_\eta^2$ ).<sup>10</sup> Importantly, the second moment of the noise is allowed to be time-varying across periods. The  $k$  intra-period price observations are employed to evaluate variance over each sub-interval of size  $\phi_{n,T}$ .

**The two-scale estimator.** Define  $q$  non-overlapping sub-grids  $\Psi^{(i)}$  of the original grid of  $k$  arrival times with  $i = 1, \dots, q$ . The first sub-grid starts at  $t_0$  and takes every  $q$ -th arrival time, i.e.,  $\Psi^{(1)} = (t_0, t_{0+q}, t_{0+2q}, \dots)$ , the second sub-grid starts at  $t_1$  and also takes every  $q$ -th arrival time, i.e.,  $\Psi^{(2)} = (t_1, t_{1+q}, t_{1+2q}, \dots)$ , and so on. Given the generic  $i$ -th sub-grid of arrival times, define

$$\widehat{V}^{(i)} = \sum_{t_j, t_{j+} \in \Psi^{(i)}} (\log(p_{t_{j+}})^* - \log(p_{t_j})^*)^2,$$

where  $t_j$  and  $t_{j+}$  denote adjacent elements in  $\Psi^{(i)}$ . The two-scale estimator is constructed as

$$\widehat{V}^{ZMA} = \frac{\sum_{i=1}^q \widehat{V}^{(i)}}{q} - \bar{k} \widehat{\mathbf{E}}(\varepsilon^2),$$

where  $\bar{k} = \frac{k-q+1}{q}$  and  $\widehat{\mathbf{E}}(\varepsilon^2) = \frac{\sum_{j=1}^k (\log(p_{t_{j+}})^* - \log(p_{t_j})^*)^2}{k}$ .

Under the previous price formation mechanism, *in the absence of jumps in the return process*, the estimator is consistent for  $\int \sigma_s^2 ds$  at speed  $k^{1/6}$  provided  $q, k \rightarrow \infty$  with  $\frac{q}{k} \rightarrow 0$  and  $\frac{q^2}{k} \rightarrow \infty$  (Zhang et al., 2005). The estimator is biased in finite samples. We optimize its finite sample properties by (1) bias-correcting it and (2)

<sup>10</sup>We adopt the classical set of assumptions under which asymptotic results (Zhang et al., 2005, and Barndorff-Nielsen et al., 2006) and finite sample results (Bandi and Russell, 2009) have been derived. For a discussion about the empirical validity of these assumptions, we refer the reader to the review paper by Bandi and Russell (2007) and the references therein.

optimizing its finite sample variance as suggested by Bandi and Russell (2009). Define  $\widehat{V}^{ZMAadj} = c(q, k)\widehat{V}^{ZMA}$ , where

$$c(q, k) = \left( \frac{qk - 1 + 2q - q^2 - k}{qk} \right)^{-1}$$

is a (bias-)correction term. Now write,

$$var\left(\widehat{V}^{ZMAadj}\right) = (c(q, k))^2 var\left(\widehat{V}^{ZMA}\right),$$

where, if  $\frac{q}{k} \leq 1/2$ ,

$$\begin{aligned} var\left(\widehat{V}^{ZMA}\right) &= \Theta^{ZMA} - \frac{1}{3}(Q + V^2) \left(\frac{q}{k}\right)^2 + \left(-\frac{1}{3}V^2\frac{1}{k} - 4V^2\frac{1}{k^2} + \frac{4}{3}Q\right) \frac{q}{k} \\ &+ \left[-\frac{4}{k^4}(Q + V^2) + \left(\frac{8\sigma_\eta^4 + 16\sigma_\eta^2V - 8Q - \frac{56}{3}V^2}{k^3}\right) + \left(\frac{24\sigma_\eta^2V - \frac{10}{3}Q + 8\sigma_\eta^4}{k^2}\right) + \left(\frac{-8\sigma_\eta^4 + 8\sigma_\eta^2V}{k}\right)\right] \frac{k}{q} \\ &+ \left[\frac{2}{k^5}Q + \left(\frac{-4\sigma_\eta^4 - 8\sigma_\eta^2V + 4Q - 8V^2}{k^4}\right) + \left(\frac{-4\sigma_\eta^4 - 16\sigma_\eta^2V + 2Q}{k^3}\right) + \left(\frac{8\sigma_\eta^4 - 8\sigma_\eta^2V}{k^2}\right) + \frac{8}{k}\sigma_\eta^4\right] \frac{k^2}{q^2}, \end{aligned}$$

and

$$\Theta^{ZMA} = (-4\sigma_\eta^4 - 8V\sigma_\eta^2)\frac{1}{k} + \left(-4\sigma_\eta^4 - 8\sigma_\eta^2V + \frac{13}{3}Q + \frac{79}{3}V^2\right) \frac{1}{k^2} + \frac{1}{k^3}(2Q + 8V^2)$$

with  $\sigma_\eta^2 = \mathbf{E}(\eta^2)$ ,  $V = \int \sigma_s^2 ds$  and  $Q = \int \sigma_s^4 ds$  (Bandi and Russell, 2009). Given preliminary estimates of  $V$ , the quarticity term  $Q$ , and  $\mathbf{E}(\eta^2)$ ,<sup>11</sup> the optimal  $q$  is defined as the minimizer of  $var\left(\widehat{V}^{ZMAadj}\right)$ . Thus,

$$\widehat{\sigma}_{iT/n}^2 = \frac{\widehat{V}_{iT/n}^{ZMAadj}(q^*)}{\phi_{n,T}} = \frac{c(q^*, k)\widehat{V}_{iT/n}^{ZMA}(q^*)}{\phi_{n,T}},$$

where  $q^* = \arg \min \left( (c(q, k))^2 var\left(\widehat{V}_{iT/n}^{ZMA}\right) \right)$ .

**Kernel estimators.** 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  $s = -q, \dots, q$ ,  $w_s = g\left(\frac{s-1}{q}\right)$ , and  $g(\cdot)$  is a kernel function on  $[0, 1]$  satisfying  $g(0) = 1$  and  $g(1) = 0$ . As in Barndorff-Nielsen et al. (2008), the number of jittered observations at the end points is equal to  $m$  (set equal to 1 when deriving the estimators' finite sample variance).

Under the previous price formation mechanism, *in the absence of jumps in the return process*, this family of estimators is consistent at rate  $k^{1/6}$  if  $q = \tau k^{2/3}$ . Provided  $g(\cdot)$  is also chosen in such a way as to guarantee that  $g'(0) = 0$  and  $g'(1) = 0$ , then the number of autocovariances can be selected as  $q = \tau k^{1/2}$  and the corresponding estimator is consistent at rate  $k^{1/4}$  (Barndorff-Nielsen et al., 2008). One such function is the modified Tukey-Hanning kernel  $g(x) = (1 - \cos \pi(1 - x)^2)/2$  which we use in what follows. Given our assumptions, the kernel estimators are unbiased for  $\int \sigma_s^2 ds$  in finite samples. Their finite sample variance, however, can (and should) be optimized (Bandi and Russell, 2009). Write  $\varsigma = \frac{q}{k}$ . Hence,

<sup>11</sup>In Subsection 8.2, we discuss a simple method of inference for these moments.

$$\text{var}_\zeta \left( \widehat{V}^{BNHLS} \right) = \frac{Q}{k} w^\top \Omega_1 w + 4 \left( \mathbf{E}(\eta^2) \right)^2 k (w^\top \Omega_2 w) + 4 \left( \mathbf{E}(\eta^2) \right)^2 (w^\top \Omega_3 w) + (2\mathbf{E}(\eta^2)V)4(w^\top \Omega_4 w),$$

with

$$w = \left( 1, 1, g \left( \frac{1}{\zeta k} \right), \dots, g \left( \frac{\zeta k - 1}{\zeta k} \right) \right)^\top,$$

and  $\Omega_a$   $a = 1, \dots, 4$  are  $(\zeta k + 1, \zeta k + 1)$  square matrices. For  $j \leq \zeta k$ , the matrices  $\Omega_1$  and  $\Omega_4$  are defined as follows:

$$\begin{aligned} \Omega_1[1, 1] &= 2, \quad \Omega_1[1 + j, 1 + j] = 4, \\ \Omega_4[1, 1] &= 1, \quad \Omega_4[2, 1] = -1, \quad \Omega_4[1, 2] = -1, \quad \Omega_4[2, 2] = 2, \quad \Omega_4[1 + j, 1 + j] = 2, \\ \Omega_4[1 + j, j] &= -1, \quad \Omega_4[j, j + 1] = -1, \end{aligned}$$

and 0 everywhere else. For  $j \leq \zeta M - 1$ , the matrices  $\Omega_2$  and  $\Omega_3$  are defined as follows:

$$\begin{aligned} \Omega_2[1, 1] &= 3, \quad \Omega_2[1, 2] = -4, \quad \Omega_2[2, 1] = -4, \quad \Omega_2[2, 2] = 7, \\ \Omega_2[2 + j, 2 + j] &= 6, \quad \Omega_2[2 + j, 1 + j] = -4, \quad \Omega_2[1 + j, 2 + j] = -4, \\ \Omega_2[2 + j, j] &= 1, \quad \Omega_2[j, 2 + j] = 1, \\ \Omega_3[1, 1] &= -1, \quad \Omega_3[1, 2] = 2, \quad \Omega_3[2, 1] = 2, \quad \Omega_3[2, 2] = -4.5, \quad \Omega_3[j + 2, j + 2] = -3(j + 1) - 1, \\ \Omega_3[2 + j, 1 + j] &= 2(j + 1), \quad \Omega_3[1 + j, 2 + j] = 2(j + 1), \\ \Omega_3[2 + j, j] &= -(j + 1)/2, \quad \Omega_3[j, 2 + j] = -(j + 1)/2, \end{aligned}$$

and 0 everywhere else. Thus, using preliminary estimates of  $\sigma_\eta^2$ ,  $V$ , and  $Q$ ,

$$\tilde{\sigma}_{iT/n}^2 = \frac{\widehat{V}_{iT/n}^{BNHLS}(\zeta^*)}{\phi_{n,T}},$$

where  $\zeta^* = \left( \frac{q}{k} \right)^* = \arg \min \left( \text{var}_\zeta \left( \widehat{V}^{BNHLS} \right) \right)$  and  $g(x) = (1 - \cos \pi(1 - x)^2) / 2$ .

**Bipower variation.** Write

$$\widehat{V}^{BNS} = \mu^{-2} \sum_{j=2}^k |r_j| |r_{j-1}|$$

with  $\mu = E(|Z|)$ , 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*, Barndorff-Nielsen and Shephard (2004, 2006) show that  $\widehat{V}^{BNS}$  is consistent for  $\int \sigma_s^2 ds$  as  $k$  increases over the period.

To break the first-order dependence in the observed returns induced by noise, Andersen et al. (2007) and Huang and Tauchen (2005) consider a staggered version of the same estimator defined as

$$\widehat{V}^{BNS(stag.)} = \mu^{-2} \left( \frac{k-2}{k} \right)^{-1} \sum_{j=3}^k |r_j| |r_{j-2}|.$$

This correction does not yield consistency in the presence of noise but has been shown to perform very satisfactorily in practise.<sup>12</sup> When estimating models with discontinuities in the return process, we employ

$$\tilde{\sigma}_{iT/n}^2 = \frac{\widehat{V}_{iT/n}^{BNS(stag.)}}{\phi_{n,T}}.$$

## 8.2 The 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 volatility estimates. 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. 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%. Table 1 contains descriptive statistics about the Spiders data. 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).

The noise-return second moment  $\mathbf{E}(\varepsilon^2)$ ,  $V$ , and  $Q$  are necessary inputs in the optimization procedures laid out in the previous subsection. Since all variance estimates are optimized for each day in our sample, these quantities are also computed daily. We estimate  $\mathbf{E}(\varepsilon^2)$  using sample second moments of quote-to-quote continuously-compounded returns. The variance and quarticity estimates are obtained by using realized variance and realized quarticity with fixed, 15-minute, calendar-time intervals (Bandi and Russell, 2008, for discussions). The "prevailing quote" method is used in the absence of a quote.

We express the variance estimates in daily terms ( $\phi_{n,T} = \frac{9}{2053}$ ). This choice is standard in the existing work on stochastic volatility modelling and, importantly, has the added benefit in our context of yielding robustness to intra-daily effects (as discussed, e.g., in Andersen and Bollerslev, 1998, and the references therein). Since the original estimates are for an intra-daily 6-hour period, we multiply them by a constant factor  $\delta$  defined 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  $\delta$ .<sup>13</sup>

Table 2 provides descriptive statistics about  $r_{iT/n}^{S\&P500}$  and  $\tilde{\sigma}_{iT/n}^2$  expressed in daily terms. Daily returns are further expressed in percentage terms ( $\times 100$ ). Consistently with this scaling, the daily variances are multiplied by 10,000. As always, market returns display little autocorrelation, little skewness, and excess kurtosis. The variance estimates are strongly right-skewed and persistent. For a graphical representation, see Fig. 1. Importantly for our purposes, the two-scale estimator and the modified Tukey-Hanning estimator have a higher mean and standard deviation than the staggered bipower variation estimator. Because in the presence of jumps in the return process

<sup>12</sup>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).

<sup>13</sup>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.

the former two estimators identify the sum of the squared jumps in addition to integrated variance, the difference between their sample means and the mean of bipower variation might suggest the presence of jumps in the return dynamics. In what follows we report a formal test of discontinuous return dynamics relying on the continuous-time properties of the full return/variance system.

## 9 Stochastic volatility dynamics

We estimate the generalized Duffie et al.’s jump-diffusion model presented in Section 5. We choose a straightforward 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.<sup>14</sup> Only for the time being, we assume absence of jumps in the return process. Fig. 2 and Fig. 3 present the results pertaining to spot variance estimated using the two-scale estimator and the modified Tukey-Hanning kernel estimator. Drift function, diffusion function, and intensity of the jumps are reported in annual terms.

As expected, the results associated with these alternative spot variance estimators are rather similar. The estimated drift denotes mildly nonlinear mean-reversion. The diffusive function is nonlinear. Below, we show that it conforms more naturally with a nonlinear constant-elasticity-of-variance (CEV) specification than with a square-root process, as introduced by Heston (1993) and adopted by several others. In the relevant variance range, the intensity point estimates suggest between 1 and 5 volatility jumps per year. The estimated expected size of the jumps is about 5 (with larger asymptotic bands for the case of kernel spot volatilities as shown in Fig. 3).

We now turn to the empirically more compelling case of jumps in the return process and use staggered bipower variation to identify integrated variance. The corresponding results are in Fig. 4. As expected in light of the temporal dynamics of the relevant series (in Fig. 1) and the economic interpretation of bipower variation in the presence of price jumps, the estimated mean size of the variance jumps is now smaller.

We compare our findings to the parametric estimates (converted to annual figures) in Table III, Column 5, in Eraker et al. (2003). 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 diffusion function is better represented by a more flexible CEV specification (i.e.,  $\Lambda^2(x) \propto x^{3/2}$ ) than by a square-root model for stochastic volatility. 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.

To further illustrate the empirical validity of our more flexible variance of variance, Fig. 5(a) presents the ratio  $(\Delta(\sigma_t^2) - m(\sigma_t^2)\Delta t) / (\Lambda(\sigma_t^2)\sqrt{\Delta t})$  with a functional specification for  $m(\sigma_t^2)$  and the linear parametric specification for  $\Lambda(\sigma_t^2)$  reported in Eraker et al. (2003). Fig. 5(b) displays the same ratio with a functional drift and a nonlinear CEV diffusion consistent with our data ( $\Lambda^2(x) = 0.1x^{3/2}$ ).<sup>15</sup> The ratio’s dynamics appear to be better

<sup>14</sup> The infinitesimal moments’ bandwidths are set equal to  $c_j * stdc(\tilde{\sigma}^2) * n^{-1/5}$ , 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).

<sup>15</sup> The parameters are estimated by GMM on the infinitesimal first, second, third, and fourth moment. We assume a linear mean-reverting drift, a constant jump intensity, and exponential jumps. The corresponding t-statistics are equal to about 3 and 4.

behaved in the latter case with occasional (positive) jumps as implied, for example, by our assumed exponential jump distribution. Importantly, descriptive statistic from the data nicely conform with descriptive statistics from simulated observations<sup>16</sup> from a nonlinear CEV stochastic variance model as described in Section 10 below. It is now of interest to assess whether the reported differences in the variance-of-variance's continuous component 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 Eraker 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. 6). The identification scheme is therefore consistent with Eq. (28) and Eq. (31) applied to returns rather than to log-variances, see Eq. (33) through Eq. (35) in Section 7.<sup>17</sup> 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. Write

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

and

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

Hence,

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

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

To define  $\widehat{\sigma}_\psi^2$  we further weigh the quantities in the sum by the corresponding local density. These additional terms are asymptotically negligible (thereby not affecting our limiting results), but play an important role in finite samples particularly when estimating return dynamics, as we show below by simulation.

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

<sup>16</sup>Not reported, but available upon request.

<sup>17</sup>The bandwidths are set as described in Footnote 14 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.

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. 6(a)). We find mildly hump-shaped (in the 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).<sup>18</sup>

Before turning to the simulations, we report a nonparametric test for jumps in returns and variances.

## 9.2 Testing nonparametrically for jumps in returns and volatility

The test compares the conditional kurtosis from data to the conditional kurtosis implied by alternative data generating processes for returns and variances. The intuition is straightforward: if discontinuities play a role in practise, the empirical conditional kurtosis should be larger than that implied by diffusive models for returns and variances. This same logic has been applied by Johannes (2004) to test for jumps in the bond market. Consider the bi-variate system:

$$r_{t,t+\Delta_t} = b\Delta_t + \sqrt{\sigma_t^2 \Delta_t} \varepsilon_t^r + \psi_t J_t^r, \quad (38)$$

$$\sigma_{t+\Delta_t}^2 - \sigma_t^2 = \kappa(\theta - \sigma_t^2)\Delta_t + \sigma_v \sqrt{\sigma_t^2 \Delta_t} \varepsilon_t^\sigma + \xi_t^\sigma J_t^\sigma, \quad (39)$$

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  $\rho$ ,  $\psi_t$  is a mean zero Gaussian random variable with standard deviation  $\sigma_\psi$ ,  $\xi_t^\sigma$  is an exponential random variable with mean  $\mu_\sigma$ , and  $\Delta_t$  is a time-discretization (one day). The parameters are those in Table III, Column 5, of Eraker et al. (2003). We start off by simulating the model under the null of no jumps, i.e.,  $J_t^r := 0$  and  $J_t^\sigma := 0$ . Consistent with data, we generate 2,053 daily observations for each simulated path of the system and replicate the simulations 1,000 times. Fig. 7 reports the mean and 95% coverage bands of the estimated conditional kurtosis  $\frac{\hat{\theta}^4(\sigma^2)}{(\hat{\theta}^2(\sigma^2))^2}$  from simulations. When compared to the estimated kurtosis from data, it is apparent that the null of no jumps is rejected overwhelmingly. This is particularly true for the variance process. By simulating the model under the alternative (with jumps in returns and variance, as implied by the more general specification in Eq. (38) and Eq. (39)), it is immediate to notice that the alternative is considerably more consistent with our empirical statistics (Fig. 8). Importantly, this coherence increases when replacing  $\sigma_v \sqrt{\sigma^2}$  with a nonparametrically-estimated, nonlinear variance of variance (Fig. 9). This result is, of course, expected in light of our previous findings in favor of a richer specification for the variance process and further supports this specification.

<sup>18</sup>The negative value of the estimated jump intensity at high variance levels should not be surprising. It is a by-product of the scarcity of observations in this range combined with bias-corrected estimates (as in the previous footnote) 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.

## 10 Simulations

The simulations rely on the bi-variate system in Eq. (38) and Eq. (39). We again generate 2,053 observations for every sample path and 1,000 paths. However, in agreement with data, the diffusive volatility of variance is specified as being equal to  $\sigma_v \sqrt{(\sigma_t^2)^{1.5} \Delta_t}$ . As earlier, the parameters are those in Table III, Column 5, of Eraker et al. (2003) with the exception of  $\sigma_v$  (set equal to 0.31). The specification  $\Lambda^2(\sigma_t^2) = 0.1(\sigma_t^2)^{1.5}$  provides superior fit for our data, as illustrated previously. Consistent with the empirical work, we use cross-validated bandwidths.

Figs. 10 and 11 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 estimates tend to be slightly downward biased, thereby reinforcing our previous results about the need for a higher variance of variance. The jump features are accurately estimated. The expected jump size is only slightly downward biased (with the true value being well within the 95% band), while the intensity of the jumps is 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 information 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 successful, 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  $\tilde{\sigma}_{iT/n}^2 = \frac{\hat{V}_{iT/n}}{\phi_{n,T}}$ , where  $\hat{V}_{iT/n}$  is a consistent *integrated* (over  $\phi_{n,T} \rightarrow 0$ ) variance estimator constructed using  $k$  intra-daily observations (with  $k \rightarrow \infty$ ). In this Appendix we show that Assumption 2 and the stronger weak convergence result

$$\phi_{n,T}^\beta k^\alpha \{ \tilde{\sigma}_{i/n}^2 - \sigma_{iT/n}^2 \} \xrightarrow[k \rightarrow \infty, \phi_{n,T} \rightarrow 0]{} MN \left( 0, a (\sigma_{iT/n}^4)^\eta + b \right)$$

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

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 specifying the values of  $\alpha$ ,  $\beta$ ,  $a$ ,  $b$ , and  $\eta$  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. $dJ_t^i = 0$ , 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. $dJ_t^i \neq 0$ , without microstructure noise.

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

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

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  $\frac{1}{\phi_{n,T}} \left( \frac{\phi_{n,T}}{k} \log \left( \frac{k}{\phi_{n,T}} \right) \right)^{(z-1)/z} \rightarrow 0$ , then  $\tilde{\sigma}_{i/n}^2 = \sigma_{iT/n}^2 + o_p(1)$  and*

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

If  $k^{\frac{1}{2}} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} \rightarrow 0$  and  $\frac{k^{\frac{1}{2}}}{\phi_{n,T}} \left( \frac{\phi_{n,T}}{k} \log \left( \frac{k}{\phi_{n,T}} \right) \right)^{(z-1)/z} \rightarrow 0$ , 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.**  $dJ_t^r = 0$ , with microstructure noise.

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

**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 (2009), Theorem 2, when  $\phi_{n,T} \rightarrow 0$  and  $k \rightarrow \infty$ , 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

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

provided  $v = \frac{q}{k} \rightarrow 0$ . 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  $q = \tau k^{2/3}$  (i.e., the optimal order in the *integrated* variance case), then  $v = \frac{q}{k} = \tau \frac{1}{k^{1/3}}$  and

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

or

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

Notice, also, that

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

■

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

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

and

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

Importantly, for  $\phi_{n,T}^{1/3} k^{1/6} \rightarrow \infty$  and  $\phi_{n,T} \rightarrow 0$ , 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  $\frac{q}{k} = \frac{k^{2/3}}{\phi_{n,T}^{2/3} k} = \frac{1}{\phi_{n,T}^{2/3} k^{1/3}} \rightarrow 0$  is also satisfied. Now write

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

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

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

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

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

and

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

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

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

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

be re-defined as  $q = \tau \phi_{n,T} \left( \frac{k}{\phi_{n,T}} \right)^{2/3} (\phi_{n,T})^{2/3} \sim q^\circ$ . ■

**Remark. (The rate of convergence)** The condition  $\phi_{n,T} = k^\theta$  with  $-\frac{1}{2} < \theta < 0$  (which implies  $\phi_{n,T}^{1/3} k^{1/6} \rightarrow \infty$ ) combined with  $\phi_{n,T}^{1/3} k^{1/6} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} \rightarrow 0$  (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  $\theta = -\frac{1}{5} - \varepsilon$ , and the implied rate of convergence of the two-scale *spot* variance estimator would be  $k^{1/10-\delta}$ , which is slower than the rate of convergence of the corresponding *integrated* variance estimator ( $k^{1/6}$ ).

**Proof of 11.** Using Barndorff-Nielsen et al. (2008), Eq. (12), and Bandi and Russell (2009), Theorem 3, write

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

where the kernel-related  $g$  terms are defined in Barndorff-Nielsen et al. (2008). In particular, if  $q = \tau k^{2/3}$  (i.e., the optimal order in the *integrated* variance case), then

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

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  $q^o = \tau \left( \frac{k}{\phi_{n,T}} \right)^{2/3}$ , which is the optimal order for this problem. Hence,

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

and

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

The optimal  $\tau (= \tau^o)$  is now

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

This choice implies

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

■

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

$$\begin{aligned} \mathbf{V}_{\sigma^2} \left\{ \left( \widehat{V}^{BNHLS} - \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds \right) \right\} &\sim 4g_{\bullet}^{0,0} \left( \phi_{n,T} \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^4 ds \right) v \\ &+ 4 \frac{1}{vk} \{g_{\bullet}^{1,1}\} \{2\mathbf{E}(\varepsilon^2) \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds\} \\ &+ 4\mathbf{E}(\varepsilon^2)^2 \frac{1}{v^3 k^2} \{g_{\bullet}^{2,2}\} + 4\mathbf{E}(\varepsilon^2)^2 \frac{1}{vkm} \{g_{\bullet}^{1,1}\}, \end{aligned}$$

If  $q = \tau k^{1/2}$  (i.e., the optimal order in the *integrated* variance case) and  $m \rightarrow \infty$ , then

$$\begin{aligned} \mathbf{V}_{\sigma^2} \left\{ \left( \frac{\widehat{V}^{BNHLS}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \\ \sim 4\tau g_{\bullet}^{0,0} \left( \frac{\phi_{n,T} \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^4 ds}{(\phi_{n,T})^2} \right) \frac{1}{k^{1/2}} + 4 \frac{\mathbf{E}(\varepsilon^2)^2}{(\phi_{n,T})^2} \left[ \frac{1}{\tau^3 k^{1/2}} \{g_{\bullet}^{2,2}\} \right] \end{aligned}$$

and

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

■

**Proof of 14.** If  $q = \tau \left( \frac{k}{\phi_{n,T}} \right)^{1/2}$ , which is the optimal order for this problem, and  $m \rightarrow \infty$ , then

$$\begin{aligned} \mathbf{V}_{\sigma^2} \left\{ \left( \frac{\widehat{V}^{BNHLS}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \\ \sim 4g_{\bullet}^{0,0} \left( \frac{\phi_{n,T} \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^4 ds}{\phi_{n,T}^2} \right) \tau \frac{1}{k^{1/2} \phi_{n,T}^{1/2}} \\ + 4 \frac{1}{\tau k^{1/2} \phi_{n,T}^{1/2}} \{g_{\bullet}^{1,1}\} \left\{ \frac{2\mathbf{E}(\varepsilon^2) \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right\} \\ + 4\mathbf{E}(\varepsilon^2)^2 \left[ \frac{1}{\tau^3 k^{1/2} \phi_{n,T}^{1/2}} \{g_{\bullet}^{2,2}\} \right] \end{aligned}$$

and

$$\begin{aligned} \mathbf{V}_{\sigma^2} \left\{ \phi_{n,T}^{1/4} k^{1/4} \left( \frac{\widehat{V}^{BNHLS}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \\ \sim 4\tau g_{\bullet}^{0,0} (\sigma_{iT/n}^4) + 4 \frac{1}{\tau} \{g_{\bullet}^{1,1}\} \{2\mathbf{E}(\varepsilon^2) \sigma_{iT/n}^2\} + 4 \frac{1}{\tau^3} \{g_{\bullet}^{2,2}\} (\mathbf{E}(\varepsilon^2))^2, \end{aligned}$$

where  $\tau$  can be optimized in the usual fashion. The optimal  $\tau (= \tau^o)$  is now

$$\tau^o = \left\{ \frac{g_{\bullet}^{1,1} + \sqrt{(g_{\bullet}^{1,1})^2 + 3g_{\bullet}^{0,0} g_{\bullet}^{2,2}}}{g_{\bullet}^{0,0}} \right\}^{1/2} \left( \frac{\mathbf{E}(\varepsilon^2)}{\sigma_{iT/n}^2} \right)^{1/2}.$$

Thus,

$$\begin{aligned} & \mathbf{V}_{\sigma^2} \left\{ \phi_{n,T}^{1/4} k^{1/4} \left( \frac{\widehat{V}^{BNHLS}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \\ & \sim \left\{ 4\tau^* g_{\bullet}^{0,0} + \frac{4}{\tau^*} g_{\bullet}^{1,1} (2\mathbf{E}(\varepsilon^2)) + \frac{4}{(\tau^*)^3} \{g_{\bullet}^{2,2}\} (\mathbf{E}(\varepsilon^2))^2 \right\} (\sigma_{iT/n}^4)^{3/4} \end{aligned}$$

with  $\tau^* = \tau^\circ \sigma_{iT/n}$ .

■

**Remark. (The rate of convergence)** The condition  $\phi_{n,T} = k^\theta$  with  $-1 < \theta < 0$  (which implies  $\phi_{n,T}^{1/4} k^{1/4} \rightarrow \infty$ ) combined with  $\phi_{n,T}^{1/4} k^{1/4} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} \rightarrow 0$  (which is necessary to eliminate the asymptotic bias term) leads to (roughly)  $\phi_{n,T} = k^\theta$  with  $-1 < \theta < -\frac{1}{3}$ . Hence, one could choose  $\theta = -\frac{1}{3} - \varepsilon$ , and the implied rate of convergence of the *spot* variance estimator would now be  $k^{1/6-\delta}$ , which is slower than the rate of convergence of the corresponding *integrated* variance estimator ( $k^{1/4}$ ).

**Remark. (The limiting variances)** Naturally, the look of the limiting variances depends crucially on the choice of  $q$ . When choosing the order of  $q$  as in the *integrated* variance case ( $q \propto k^{2/3}$  or  $q \propto k^{1/2}$ ), and thus suboptimally, the asymptotic variances do not depend on the quarticity but do depend on the second noise moment ( $a = 0$  and  $b \neq 0$ ).

When choosing the order of  $q$  optimally for our *spot* variance problem ( $q \propto \left(\frac{k}{\phi_{n,T}}\right)^{2/3}$  or  $q \propto \left(\frac{k}{\phi_{n,T}}\right)^{1/2}$ ), the asymptotic variances depend both on the quarticity and on the second moment of the noise. In this case,  $a \neq 0$  and  $b = 0$ .

**Lemma A.1.** Assume  $\tilde{\sigma}_{iT/n}^2 = \left( \sum_{j=1}^k r_j^2 \right) / \phi_{n,T}$ . Then,

$$\max_{1 \leq i \leq n} |\tilde{\sigma}_{iT/n}^2 - \sigma_{iT/n}^2| = O_p(\phi_{n,T}^{1/2} \log^{1/2}(1/\phi_{n,T})) + O_p\left(\frac{\log(n)}{\sqrt{(kn\phi_{n,T})/T}}\right).$$

**Proof of Lemma A.1.** Write, as in Remark 1,

$$\tilde{\sigma}_{iT/n}^2 = \left( \sum_{j=1}^k r_j^2 \right) / \phi_{n,T} = \frac{1}{\phi_{n,T}} \sum_{j=1}^{nk} \mathbf{1}_{\{0 \leq \frac{j-iT/n}{\phi_{n,T}} \leq 1\}} r_j^2.$$

Ignoring the higher-order drift term, by Doob's decomposition,  $r_j^2 \sim \left( \int_{j-1}^j \sigma_s dW_s \right)^2 = \int_{j-1}^j \left( \int_{j-1}^s \sigma_u dW_u \right) \sigma_s dW_s + \int_{j-1}^j \sigma_s^2 ds$ . Hence,

$$\begin{aligned} \tilde{\sigma}_{iT/n}^2 - \sigma_{iT/n}^2 & \sim \frac{1}{\phi_{n,T}} \sum_{j=1}^{nk} \mathbf{1}_{\{0 \leq \frac{j-iT/n}{\phi_{n,T}} \leq 1\}} \int_{j-1}^j \left( \int_{j-1}^s \sigma_u dW_u \right) \sigma_s dW_s \\ & \quad + \frac{1}{\phi_{n,T}} \sum_{j=1}^{nk} \mathbf{1}_{\{0 \leq \frac{j-iT/n}{\phi_{n,T}} \leq 1\}} \left( \int_{j-1}^j \sigma_s^2 ds \right) - \sigma_{iT/n}^2 \\ & = \Xi_{iT/n}^1 + \Xi_{iT/n}^2. \end{aligned}$$

Using Boole's inequality,

$$\Pr \left[ \max_{1 \leq i \leq n} |\Xi_{iT/n}^1| > \delta \right] \leq \sum_{i=1}^n \Pr [|\Xi_{iT/n}^1| > \delta] = n \Pr [|\Xi_{iT/n}^1| > \delta] \text{ for all } 1 \leq i \leq n.$$

By the modulus of continuity of Brownian motion, the local martingale  $\int_{j-1}^t \left( \int_{j-1}^s \sigma_u dW_u \right) \sigma_s dW_s$  is so that

$$\Pr \left\{ \left| \frac{1}{\phi_{n,T}} \mathbf{1}_{\{0 \leq \frac{j-iT/n}{\phi_{n,T}} \leq 1\}} \int_{j-1}^j \left( \int_{j-1}^s \sigma_u dW_u \right) \sigma_s dW_s \right| \leq \frac{TM}{kn\phi_{n,T}} \mathbf{1}_{\{0 \leq \frac{j-iT/n}{\phi_{n,T}} \leq 1\}} \log \left( \frac{kn}{T} \right) = c \right\} = 1, \quad (41)$$

for some  $M > 0$ . Also,

$$V_{\Xi_{iT/n}^1} \sim \frac{T}{\phi_{n,T} nk} 2 \sum_{j=1}^{nk} \mathbf{1}_{\{0 \leq \frac{j-iT/n}{\phi_{n,T}} \leq 1\}} \int_{j-1}^j \sigma_s^4 ds = \frac{T}{\phi_{n,T} nk} V$$

However, Eq. (41) implies that, by De La Peña's exponential inequality for sums of almost surely bounded martingale difference sequences (De La Peña, 1999, Theorem 1.2A. and Theorem 1.2B) and, again, by Boole's inequality:

$$\begin{aligned} \Pr [|\Xi_{iT/n}^1| > \delta] &\leq 2 \exp \left( \frac{-\delta^2}{V_{\Xi_{iT/n}^1} \left(1 + \sqrt{1 + 2c\delta/V_{\Xi_{iT/n}^1}}\right) + c\delta} \right) \\ &\sim 2 \exp \left( \frac{-\delta^2 (\phi_{n,T}nk/T)}{V \left(1 + \sqrt{1 + 2\frac{c\delta}{V}(\phi_{n,T}nk/T)}\right) + c\delta(\phi_{n,T}nk/T)} \right). \end{aligned}$$

Now set  $\delta_{k,n,T} = \frac{\sqrt{\log n^{(1+\varepsilon)}}}{\sqrt{(kn\phi_{n,T})/T}}$ , where  $\varepsilon > 0$  is an arbitrarily small number. Then, since

$$c\delta(\phi_{n,T}nk/T) \sim \frac{T}{kn\phi_{n,T}} \log(kn/T) \frac{\sqrt{\log n^{(1+\varepsilon)}}}{\sqrt{(kn\phi_{n,T})/T}} \left(\frac{\phi_{n,T}nk}{T}\right) \rightarrow 0,$$

we have

$$\begin{aligned} \Pr \left[ \max_{1 \leq i \leq n} |\Xi_{iT/n}^1| > \delta_{k,n,T} \right] &\leq n \Pr [|\Xi_{iT/n}^1| > \delta_{k,n,T}] \\ &\sim n \exp \left( -\log n^{(1+\varepsilon)} \right) \\ &= n \frac{1}{n^{1+\varepsilon}} \rightarrow 0. \end{aligned}$$

Thus,  $\max_{1 \leq i \leq n} |\Xi_{iT/n}^1| = O_p \left( \frac{\sqrt{\log n^{1+\varepsilon}}}{\sqrt{(kn\phi_{n,T})/T}} \right) = O_p \left( \frac{\sqrt{(1+\varepsilon) \log n}}{\sqrt{(kn\phi_{n,T})/T}} \right)$ . Finally,

$$\begin{aligned} \max_{1 \leq i \leq n} |\Xi_{iT/n}^2| &= \max_{1 \leq i \leq n} \left| \frac{1}{\phi_{n,T}} \int_{iT/n}^{iT/n + \phi_{n,T}} \sigma_s^2 ds - \sigma_{iT/n}^2 \right| \\ &= \max_{1 \leq i \leq n} \sup_{iT/n \leq s \leq iT/n + \phi_{n,T}} |\sigma_s^2 - \sigma_{iT/n}^2| \\ &= o_{a.s.}(\phi_{n,T}^{1/2} \log^{1/2}(1/\phi_{n,T})). \end{aligned}$$

■

**Remark.** The result is intuitive. The terms  $\Xi_{iT/n}^1$  can be embedded into asymptotically independent (across  $i \leq n$ ) Gaussian random variables. However, consider  $n$  independent normal random variables  $\Psi_i$  with variance  $E(\Psi_i^2)$ . It is known that, almost surely,

$$\max_{1 \leq i \leq n} \Psi_i \leq \sqrt{\log(n) \times \max_{1 \leq i \leq n} E(\Psi_i^2)}.$$

Similarly, the denominator in  $O_p \left( \frac{\sqrt{\log n}}{\sqrt{(kn\phi_{n,T})/T}} \right)$  is the square root of the asymptotic order of the quadratic variation of  $\Xi_{iT/n}^1$ .

## B Appendix B

We begin with useful preliminary lemmas.

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

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

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

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

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.

**Proof.** Straightforward using Bandi and Phillips (2003).

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

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

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

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

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.

**Proof.** Straightforward using Bandi and Phillips (2003).

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

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

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

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

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  $\forall j \geq 1$ .

**Proof.** Straightforward using Bandi and Nguyen (2003).

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

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

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

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

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

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

by the Brownian modulus of continuity. Also, given Lemma A.1.,

$$\tilde{\sigma}_{i\Delta_{n,T}}^2 - \frac{\int_{iT/n}^{iT/n + \phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} = O_p \left( \frac{\log(n)}{k^\alpha \phi_{n,T}^\beta} \right)$$

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

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

This implies that

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

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

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

by the occupation time formula for semimartingales (see, e.g., Protter, 1995, Corollary 1, p.168) and the integrability of  $\mathbf{K}'(\cdot)$  from Assumption 1. If  $g(T, n, k, \phi_{n,T}) \rightarrow 0$ , then  $\tilde{L}_n(x) \xrightarrow{p} 0$ , 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

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

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

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

and

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

Now, write

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

and, since

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

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

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

Since  $\hat{m}(x) = m(x) + o_p(1)$  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

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

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

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

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

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

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

and

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

since

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

As for weak convergence, note that

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

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

$$\left\{ \widehat{\lambda}_{\sigma^2}(x) - \lambda_{\sigma^2}(x) \right\} = \frac{\widehat{\theta}^4(x)}{24\widehat{\mu}_\xi^4} - \frac{\theta^4(x)}{24\mu_\xi^4} = \frac{\widehat{\theta}^4(x) - \theta^4(x)}{24\mu_\xi^4} - \frac{4\theta^4(x)(\widehat{\mu}_\xi - \mu_\xi)}{24\mu_\xi^5}. \quad (42)$$

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

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

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

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

$$\begin{aligned}
\mathbf{U}_{n,T}^2 &= \sum_{j=1}^{n-1} w^2(\sigma_{jT/n}^2) \mathbf{E} \left\{ \left( \int_{jT/n}^{(j+1)T/n} \int (\xi^\sigma)^4 \bar{v}_\sigma(ds, d\xi^\sigma) \right)^2 \middle| \mathfrak{S}_{j,n} \right\} \\
&= \frac{1}{h_{n,T}} \sum_{j=1}^{n-1} \mathbf{K}^2 \left( \frac{\sigma_{jT/n}^2 - x}{h_{n,T}} \right) \mathbf{E} \left\{ \left( \int_{jT/n}^{(j+1)T/n} \lambda(\sigma_{s-}^2) \mathbf{E}((\xi^\sigma)^8) ds \right) \middle| \mathfrak{S}_{j,n} \right\} \\
&= \frac{\Delta_{n,T}}{h_{n,T}} \sum_{j=1}^{n-1} \mathbf{K}^2 \left( \frac{\sigma_{jT/n}^2 - x}{h_{n,T}} \right) \lambda(\sigma_{jT/n}^2) \mathbf{E}((\xi^\sigma)^8) + o_p(1) \\
&\xrightarrow[\Delta_{n,T} \rightarrow 0]{p} \frac{1}{h_{n,T}} \int_0^T \mathbf{K}^2 \left( \frac{\sigma_{s-}^2 - x}{h_{n,T}} \right) \lambda(\sigma_{s-}^2) \mathbf{E}((\xi^\sigma)^8) ds = \tilde{\mathbf{U}}.
\end{aligned}$$

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

$$\forall \epsilon > 0, \sum_{j=1}^{n-1} \mathbf{E} \left\{ u_{jT/n, (j+1)T/n}^2 \mathbf{1}_{(|u_{jT/n, (j+1)T/n}| > \epsilon)} \middle| \mathfrak{S}_{j,n} \right\} \xrightarrow{p} 0, \quad (43)$$

then

$$\mathbf{U}_{n,T} \Rightarrow Z, \text{ with characteristic function } \mathbf{E}(\exp(-\frac{1}{2} \tilde{\mathbf{U}} t^2)), \quad (44)$$

which is a generalized martingale central limit theorem (Hall and Heyde, 1980, Theorem 3.2, Corollary 3.1). We now show that Eq. (43) holds. Write

$$\begin{aligned}
&\sum_{j=1}^{n-1} \mathbf{E} \left\{ u_{jT/n, (j+1)T/n}^2 \mathbf{1}_{(|u_{jT/n, (j+1)T/n}| > \epsilon)} \middle| \mathfrak{S}_{jT/n} \right\} \\
&= \sum_{j=1}^{n-1} \mathbf{E} \left\{ u_{jT/n, (j+1)T/n}^2 \middle| \mathfrak{S}_{jT/n} \right\} - \sum_{j=1}^{n-1} \mathbf{E} \left\{ u_{jT/n, (j+1)T/n}^2 \mathbf{1}_{(|u_{jT/n, (j+1)T/n}| \leq \epsilon)} \middle| \mathfrak{S}_{jT/n} \right\} \\
&= \tilde{\mathbf{U}}_{n,T}^2 - \sum_{j=1}^{n-1} \mathbf{E} \left\{ u_{jT/n, (j+1)T/n}^2 \mathbf{1}_{\left( \left| \mathbf{K} \left( \frac{\sigma_{jT/n}^2 - x}{h_{n,T}} \right) \int_{jT/n}^{(j+1)T/n} \int (\xi^\sigma)^4 \bar{v}_\sigma(ds, d\xi^\sigma) \right| \leq \epsilon \sqrt{h_{n,T}} \right)} \middle| \mathfrak{S}_{jT/n} \right\}, \quad (45)
\end{aligned}$$

but  $\mathbf{K} \left( \frac{\sigma_{jT/n}^2 - x}{h_{n,T}} \right) \int_{jT/n}^{(j+1)T/n} \int (\xi^\sigma)^4 \bar{v}_\sigma(ds, d\xi^\sigma) = O_p(\sqrt{\Delta_{n,T}})$ . Hence, the indicator converges in probability to 1 and, given boundedness of  $\tilde{\mathbf{U}}$ , Eq. (45) converges in probability to 0 (as  $\Delta_{n,T} \rightarrow 0$ ). Thus, by Eq. (44),  $\mathbf{U}_{n,T} \xrightarrow[\Delta_{n,T} \rightarrow 0]{p} W_{\tilde{\mathbf{U}}}$ , where  $W$  denotes Brownian motion. In addition,

$$\frac{\mathbf{U}_{n,T}}{\sqrt{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{j=1}^n \mathbf{K} \left( \frac{\sigma_{jT/n}^2 - x}{h_{n,T}} \right)}} \xrightarrow[\Delta_{n,T} \rightarrow 0]{p} W \frac{\tilde{\mathbf{U}}}{\frac{1}{h_{n,T}} \int_0^T \mathbf{K} \left( \frac{\sigma_{s-}^2 - x}{h_{n,T}} \right) ds}$$

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

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

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

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

It then follows that

$$\left( \sqrt{h_{n,T} \widehat{L}_{\sigma^2}(T, x)} \right) \left\{ \widehat{\lambda}_{\sigma^2}(x) - \lambda_{\sigma^2}(x) \right\} \Delta_{n,T \rightarrow 0, T \rightarrow \infty, h_{n,T} \rightarrow 0} \left( \frac{\sqrt{\mathbf{K}_2 \lambda(x) \mathbf{E}((\xi^\sigma)^8)}}{24\mu_\xi^4} \right) Z, \quad (46)$$

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

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

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

Write

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

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

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

Thus,

$$\left( \tilde{\mathbf{A}} \right)^{-1/2} (\mathbf{A}_{n,\bar{n},T}) \xrightarrow[\Delta_{n,T} \rightarrow 0, T \rightarrow \infty, h_{n,T} \rightarrow 0]{\Rightarrow} Z, \tag{49}$$

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

In this case, write

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

where  $w(\sigma_{jT/n}^2) = \frac{\bar{T}}{\bar{n}} \sum_{i=1}^{\bar{n}} \frac{\theta^4(\sigma_{iT/\bar{n}}^2)}{4(\theta^3(\sigma_{iT/\bar{n}}^2))^2} \left( \frac{\frac{1}{h_{n,T}} \mathbf{K} \left( \frac{\sigma_{jT/n}^2 - \sigma_{iT/\bar{n}}^2}{h_{n,T}} \right)}{\left( \frac{\Delta_{n,T}}{h_{n,T}} \sum_{j=1}^n \mathbf{K} \left( \frac{\sigma_{jT/n}^2 - \sigma_{iT/\bar{n}}^2}{h_{n,T}} \right) \right)} \right)$ . Following the same steps leading to Eq. (49) we

obtain:

$$\left( \tilde{\mathbf{B}} \right)^{-1/2} (\mathbf{B}_{n,\bar{n},T}) \xrightarrow[\Delta_{n,T} \rightarrow 0, T \rightarrow \infty, h_{n,T} \rightarrow 0]{\Rightarrow} Z,$$

where

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

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

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

In sum,

$$\left(\tilde{\mathbf{A}} + 2\tilde{\mathbf{C}} + \tilde{\mathbf{B}}\right)^{-1/2} \bar{T} \{\hat{\mu}_\xi - \mu_\xi\} \xrightarrow{\Delta_{n,T} \rightarrow 0, T \rightarrow \infty, h_{n,T} \rightarrow 0} Z, \quad (50)$$

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

$$\left(\hat{\mathbf{D}}\right)^{-1/2} \bar{T} \{\hat{\mu}_\xi - \hat{\mu}_\xi\} \xrightarrow{\Delta_{n,T} \rightarrow 0, T \rightarrow \infty, h_{n,T} \rightarrow 0} Z,$$

where

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

(as in Remark 13) since  $(\tilde{\mathbf{A}} + 2\tilde{\mathbf{C}} + \tilde{\mathbf{B}}) / \hat{\mathbf{D}} \xrightarrow{P} 1$ . Also, from Eq. (50), the rate of convergence of  $\hat{\mu}_\xi$  to  $\mu_\xi$  is effectively  $O_p((\bar{L}_{\sigma^2}(T, x))^{1/2})$ . This (semiparametric) rate is of course faster than the (nonparametric) rate of convergence of  $\hat{\theta}^j(x)$  to  $\theta^j(x) \forall j$ , namely  $O_p((h_{n,T} \bar{L}_{\sigma^2}(T, x))^{1/2})$ , thereby justifying the asymptotic negligibility of the second term in Eq. (42).  $\blacksquare$   
**Proof of Theorem 6.** The result follows from using the method of proof of Theorem 2 combined with Lemma B.3.  $\blacksquare$   
**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

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

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

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

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

$$\begin{aligned}
\mathbf{R}_1 &= \frac{\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K} \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} (\sigma_{s-}^2 - \sigma_{iT/n}^2) \sigma_{s-} dW_s^r \right]}{\underbrace{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^n \mathbf{K} \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right)}_{\mathbf{R}_{1,a}}} \\
&+ \frac{\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K} \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} (\log(p_{s-}) - \log(p_{iT/n})) \Lambda(\sigma_{s-}^2) dW_s^\sigma \right]}{\underbrace{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^n \mathbf{K} \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right)}_{\mathbf{R}_{1,b}}} = O_p \left( \sqrt{\frac{\Delta_{n,T}}{h_{n,T} \bar{L}_{\sigma^2}(T, \sigma^2)}} \right) \quad (51)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{R}_2 &= \frac{\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K} \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} \int (\sigma_{s-}^2 - \sigma_{iT/n}^2) \psi \bar{v}_r(ds, d\psi) \right]}{\underbrace{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^n \mathbf{K} \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right)}_{\mathbf{R}_{2,a}}} \\
&+ \frac{\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K} \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} \int (\log(p_{s-}) - \log(p_{iT/n})) \xi^\sigma \bar{v}_\sigma(ds, d\xi^\sigma) \right]}{\underbrace{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^n \mathbf{K} \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right)}_{\mathbf{R}_{2,b}}} = O_p \left( \sqrt{\frac{\Delta_{n,T}}{h_{n,T} \bar{L}_{\sigma^2}(T, \sigma^2)}} \right) \quad (52)
\end{aligned}$$

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

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

Notice also that,

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

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

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

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. (51) and Eq. (52). Start with  $\mathbf{R}_{1,b}$ . Express the conditional variance of the standardized numerator  $\widetilde{\mathbf{R}}_{1,b}^{num} = \left( \sqrt{\frac{h_{n,T}}{\Delta_{n,T}}} \right) \mathbf{R}_{1,b}^{num}$  as

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

Noting that

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

write

$$\begin{aligned} & \left( \widetilde{\mathbf{R}}_{1,b}^{num} \right)^2 \\ = & \frac{1}{h_{n,T} \Delta_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) \mathbf{E} \left\{ \int_{iT/n}^{(i+1)T/n} (\log p_{s-} - \log p_{iT/n})^2 \Lambda^2(\sigma_{s-}^2) ds | \mathfrak{S}_{i,n} \right\} \\ = & \frac{1}{h_{n,T} \Delta_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) \mathbf{E} \left\{ \int_{iT/n}^{(i+1)T/n} \frac{\Lambda^2(\sigma_{s-}^2)}{(\sigma_{s-}^2 + \lambda_r(\sigma_{s-}^2) \mathbf{E}(\psi^2))} \{ \log p_{s-} - \log p_{iT/n} \} d \{ \log p - \log p_{iT/n} \}_s | \mathfrak{S}_{i,n} \right\} \\ & + o_p(1) \\ = & \frac{1}{h_{n,T} \Delta_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) \frac{\Lambda^2(\sigma_{iT/n}^2 + o_{a.s.}(1))}{(\sigma_{iT/n}^2 + \lambda_r(\sigma_{iT/n}^2) \mathbf{E}(\psi^2) + o_{a.s.}(1))} \mathbf{E} \left\{ \frac{\{ \log p_{s-} - \log p_{iT/n} \}^2}{2} \Big|_{iT/n}^{(i+1)T/n} | \mathfrak{S}_{i,n} \right\} \\ & + o_p(1) \\ = & \frac{1}{h_{n,T} \Delta_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) \frac{\Lambda^2(\sigma_{iT/n}^2 + o_{a.s.}(1)) \mathbf{E} \left( \left( \int_{iT/n}^{(i+1)T/n} \sigma_{s-}^2 + \lambda_r(\sigma_{s-}^2) \mathbf{E}(\psi^2) \right)^2 | \mathfrak{S}_{i,n} \right)}{2(\sigma_{iT/n}^2 + \lambda_r(\sigma_{iT/n}^2) \mathbf{E}(\psi^2) + o_{a.s.}(1))} \\ & + o_p(1) \\ & \xrightarrow{p} \frac{1}{\Delta_{n,T} \rightarrow 0} \frac{1}{2} \frac{1}{h_{n,T}} \int_0^T \mathbf{K}^2 \left( \frac{\sigma_{s-}^2 - \sigma^2}{h_{n,T}} \right) \Lambda^2(\sigma_{s-}^2) [\sigma_{s-}^2 + \lambda_r(\sigma_{s-}^2) \mathbf{E}(\psi^2)] ds, \end{aligned} \tag{53}$$

where the term  $\{ \log p_{s-} - \log p_{iT/n} \}$  defines the measure  $\int_{iT/n}^s (\sigma_{u-}^2 + \lambda_r(\sigma_{u-}^2) \mathbf{E}(\psi^2)) du$ . We now justify the  $o_p(1)$  term above by showing that the neglected terms (in  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ ) 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

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

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

$$\begin{aligned}
& \frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) \mathbf{E} \left( \left( \frac{\int_{iT/n}^{(i+1)T/n} \Lambda^2(\sigma_{s-}^2) ds}{\Delta_{n,T}} \right) \left( \int_{iT/n}^{(i+1)T/n} \int \psi^2 \bar{v}(ds, d\psi) \right) | \mathfrak{S}_{i,n} \right) \\
&= \frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) \mathbf{E} \left( \left( \Lambda^2(\sigma_{iT/n}^2) + o_{a.s.} \left( \Delta_{n,T}^{1/2} \log \left( \frac{1}{\Delta_{n,T}} \right) \right) \right) \left( \sum_{iT/n \leq s \leq (i+1)T/n} (\Delta \log p_s)^2 \right) | \mathfrak{S}_{i,n} \right) \\
&\quad - \frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left( \frac{\sigma_{iT/n}^2 - \sigma^2}{h_{n,T}} \right) \mathbf{E} \left( \left( \Lambda^2(\sigma_{iT/n}^2) + o_{a.s.} \left( \Delta_{n,T}^{1/2} \log \left( \frac{1}{\Delta_{n,T}} \right) \right) \right) \left( \int_{iT/n}^{(i+1)T/n} \mathbf{E}(\psi^2) \lambda_r(\sigma_{s-}^2) ds \right) | \mathfrak{S}_{i,n} \right) \\
&= o_{a.s.} \left( \Delta_{n,T}^{1/2} \log \left( \frac{1}{\Delta_{n,T}} \right) \right) O_p \left( \frac{1}{h_{n,T}} \int_0^T \mathbf{K}^2 \left( \frac{\sigma_{s-}^2 - \sigma^2}{h_{n,T}} \right) \mathbf{E}(\psi^2) \lambda_r(\sigma_{s-}^2) ds \right) \xrightarrow{\Delta_{n,T} \rightarrow 0} 0,
\end{aligned}$$

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

Finally, note that  $\left\{ \int_{iT/n}^{(i+1)T/n} (\log p_{s-} - \log p_{iT/n}) \Lambda(\sigma_{s-}^2) dW_s^\sigma, \mathfrak{S}_{i,n} \right\}$  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  $\tilde{\mathbf{R}}_{1,b}^{num} = \sqrt{\frac{h_{n,T}}{\Delta_{n,T}}} \mathbf{R}_{1,b}^{num}$ . Hence, using a generalized martingale convergence theorem and Skorohod embedding as in the proof of Theorem 5:

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

Similarly,

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

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

$$\begin{aligned}
& asycov \left( \sqrt{\frac{h_{n,T} \widehat{L}_{\sigma^2}(T, \sigma^2)}{\Delta_{n,T}}} \frac{\mathbf{R}_{1,b}^{num}}{\widehat{L}_{\sigma^2}(T, \sigma^2)}, \sqrt{\frac{h_{n,T} \widehat{L}_{\sigma^2}(T, \sigma^2)}{\Delta_{n,T}}} \frac{\mathbf{R}_{1,a}^{num}}{\widehat{L}_{\sigma^2}(T, \sigma^2)} \right) \\
& \xrightarrow{\Delta_{n,T} \rightarrow 0, T \rightarrow \infty, h_{n,T} \rightarrow 0} \frac{1}{2} \mathbf{K}_2 \rho^2(\sigma^2) \sigma^2 \Lambda^2(\sigma^2).
\end{aligned}$$

This result yields:

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

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

$$\begin{aligned}
& \sqrt{\frac{h_{n,T} \widehat{L}_{\sigma^2}(T, \sigma^2)}{\Delta_{n,T}}} \mathbf{R}_2 \\
& \xrightarrow{\Delta_{n,T} \rightarrow 0, T \rightarrow \infty, h_{n,T} \rightarrow 0} \sqrt{\mathbf{K}_2 \frac{1}{2} (\Lambda^2(\sigma^2) + \lambda_\sigma(\sigma^2) \mathbf{E}[\xi^2]) \lambda_r(\sigma^2) \mathbf{E}[\psi^2] + (\sigma^2 + \lambda_r(\sigma^2) \mathbf{E}[\psi^2]) \lambda_\sigma(\sigma^2) \mathbf{E}[\xi^2]} Z.
\end{aligned}$$

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

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

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

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

Thus,

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

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,

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

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,  $\lim_{n,T \rightarrow \infty} h_{n,T}^5 v(T) = o(1)$ ),  $\widehat{\rho}(\sigma^2) - \rho(\sigma^2)$  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 (2007). Volatility. In *Handbook of Financial Engineering* (V. Linetski and J. Birge, Eds.) Elsevier Science, New York.
- [15] Bandi, F.M., and J.R. Russell (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies* 75, 339-369.
- [16] Bandi, F.M., and J.R. Russell (2009). Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *Journal of Econometrics*, forthcoming.
- [17] Barndorff-Nielsen, O.E., P. Hansen, A. Lunde and N. Shephard (2008). Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481-1536.
- [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 (2002). Fourier series methods for measurement of multivariate volatilities. *Finance and Stochastics* 6, 49-61.
- [51] Malliavin, P. and M.E. Mancino (2008). A Fourier transform method for nonparametric estimation of volatility. *Annals of Statistics*, forthcoming.
- [52] Mancini, C. (2007). Nonparametric threshold estimation for models with stochastic diffusion coefficients and jumps. *Scandinavian Journal of Statistics*, forthcoming.
- [53] Mancini, C., and R. Renò (2009). Threshold estimation of Markov models with jumps and interest rate modeling. *Journal of Econometrics*, forthcoming.
- [54] McAleer, M., and M. Medeiros (2008). Realized volatility: A review. *Econometric Reviews*, forthcoming.
- [55] Moloche, G. (2002). Local nonparametric estimation of scalar diffusions. *Working paper*.
- [56] Pan, J. (2002). The jump-risk premia implicit in options. Evidence from an integrated time-series study. *Journal of Financial Economics* 63, 3-50.
- [57] Protter, P. (1995). *Stochastic integration and differential equations*. Springer, New York.

- [58] Renò, R. (2006). Nonparametric estimation of stochastic volatility models. *Economics Letters* 90, 390-395.
- [59] Renò, R. (2008). Nonparametric estimation of the diffusion coefficient of stochastic volatility models. *Econometric Theory* 24, 1174-1206.
- [60] Revuz, D. and M. Yor (1998). *Continuous Martingales and Brownian Motion*. Springer-Verlag.
- [61] Shephard, N. (2005). *Stochastic volatility: Selected Readings*. Edited volume. Oxford University Press.
- [62] Shephard, N. (2006). Stochastic volatility. In *New Palgrave Dictionary of Economics* (S. Durlauf and L. Blume, Eds.) 2nd Edition.
- [63] Todorov, V. (2007). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *Working paper*.
- [64] Van Zanten, J.H. (2000). A multivariate central limit theorem for continuous local martingales. *Statistics and Probability Letters* 50, 229-235.
- [65] Wee, I-S. (2000). Recurrence and transience for jump-diffusion models. *Stochastic Analysis and Applications* 18, 1055-1064.
- [66] Woerner, J. (2006). Power and multipower variation: inference for high frequency data. In A. N. Shiryaev, M. do Rosario Grossinho, P. Oliveira, and M. Esquivel (Eds.). *Proceedings of the International Conference on Stochastic Finance 2004*, 343-364. Berlin: Springer-Verlag.
- [67] Xu, K.-L. (2006). Empirical likelihood based inference for nonlinear diffusions. *Working paper*.
- [68] Xu, K.-L. (2007). Re-weighted functional estimation of diffusion models. *Working paper*.
- [69] Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics* 127, 165-178.
- [70] 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)  
S&P500 returns

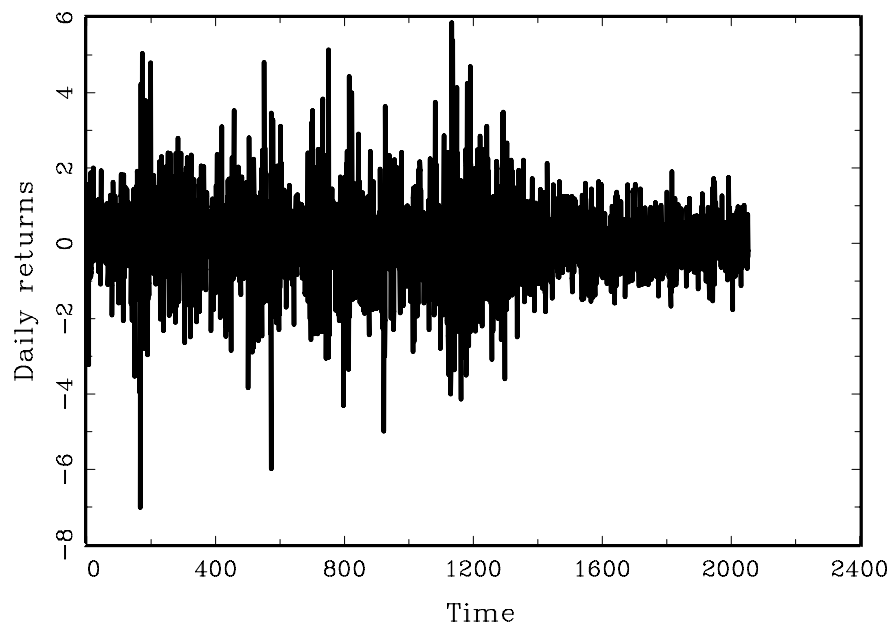


Figure 1(b)  
ZMAadj variance estimates

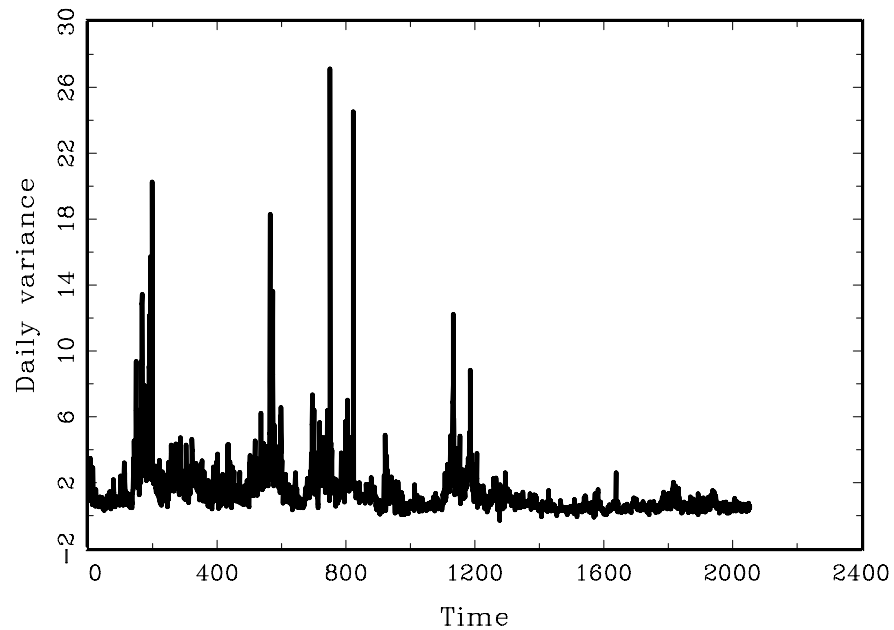


Figure 1(c)  
BNHLS variance estimates

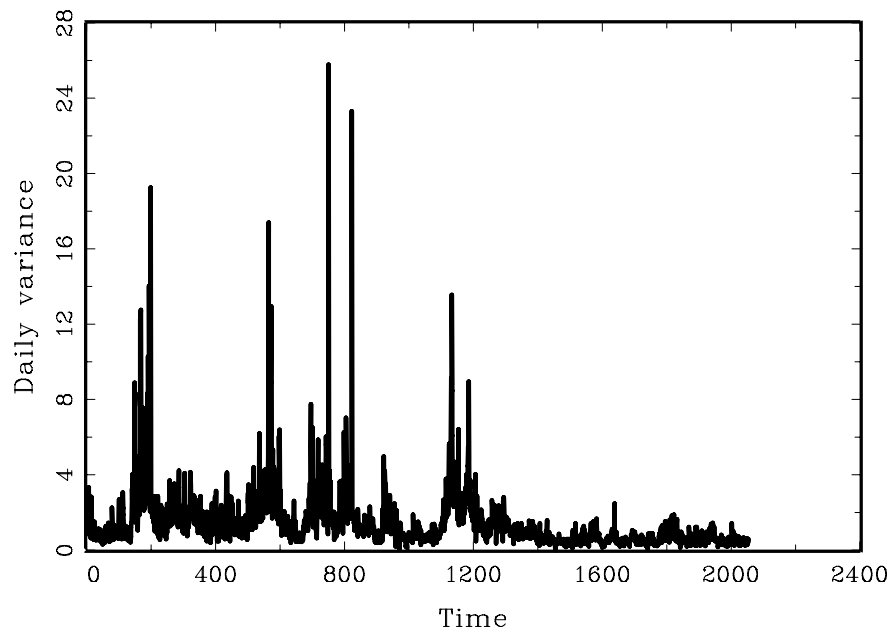


Figure 1(d)  
BNS(stag.) variance estimates

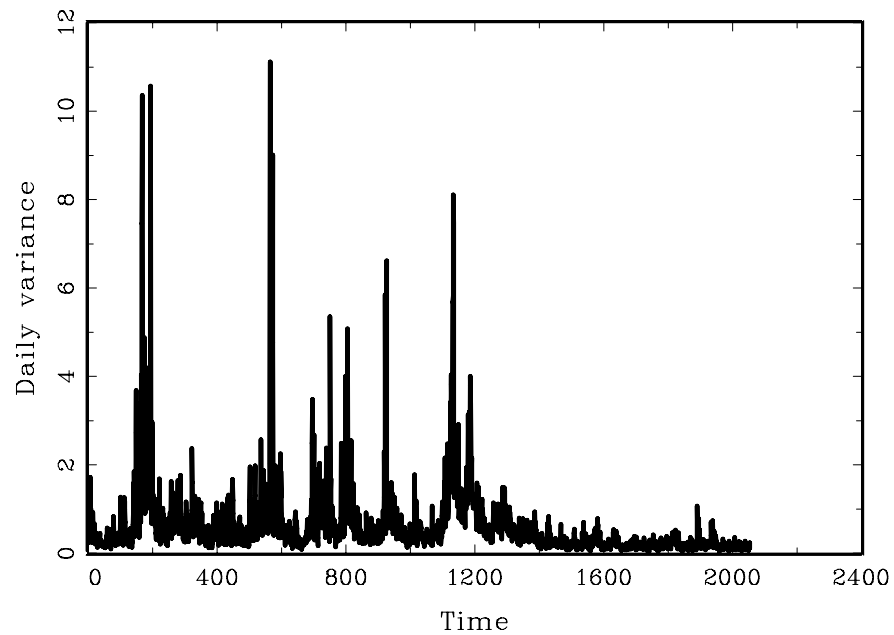


Figure 2(a)  
Variance drift estimates

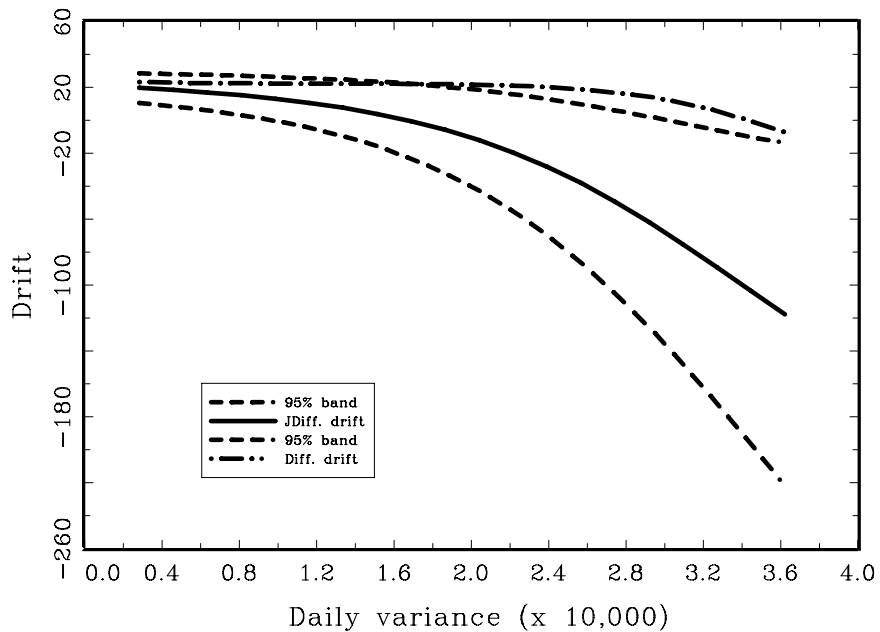


Figure 2(b)  
Variance diffusion estimates

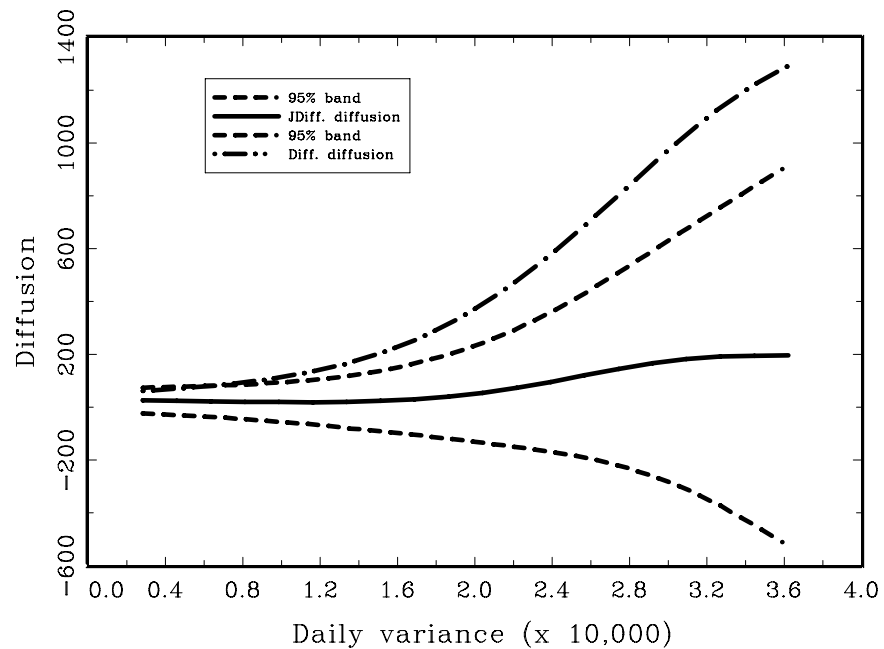


Figure 2(c)  
Variance jump intensity

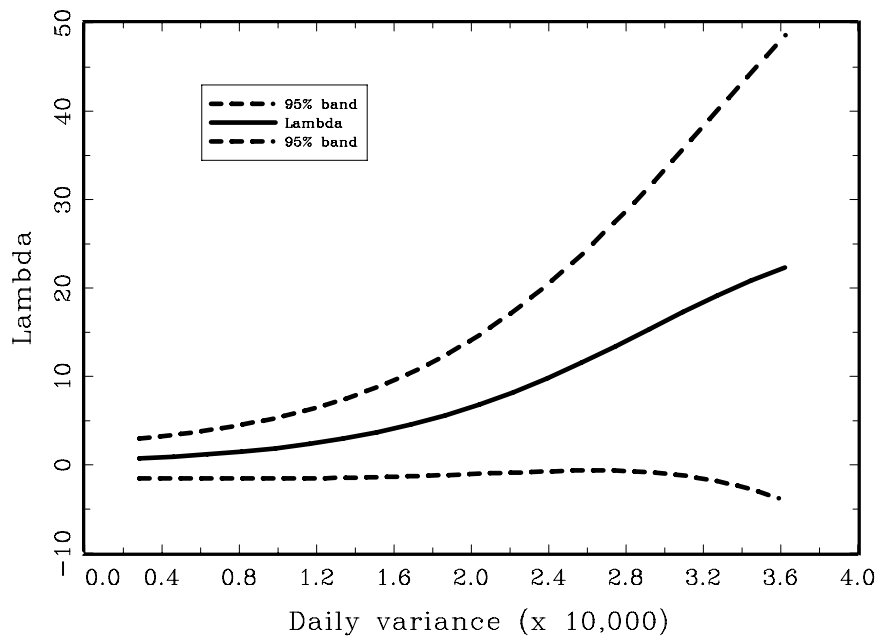


Figure 2(d)  
Variance expected jump size

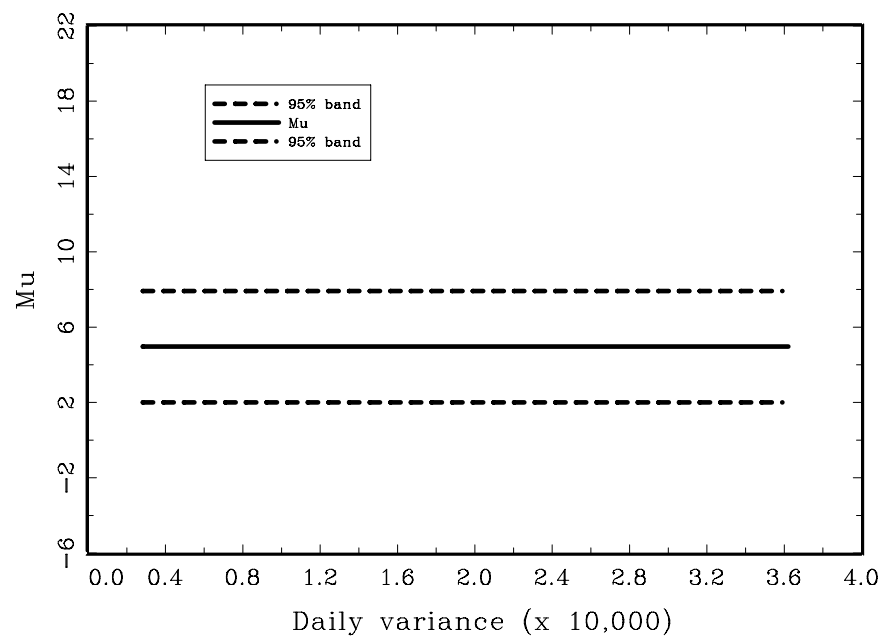


Figure 3(a)  
Variance drift estimates

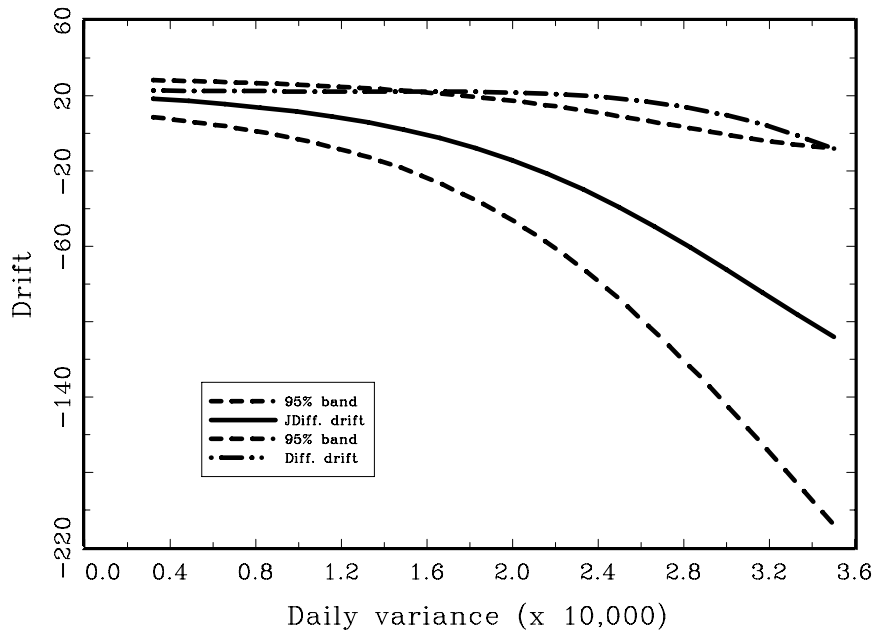


Figure 3(b)  
Variance diffusion estimates

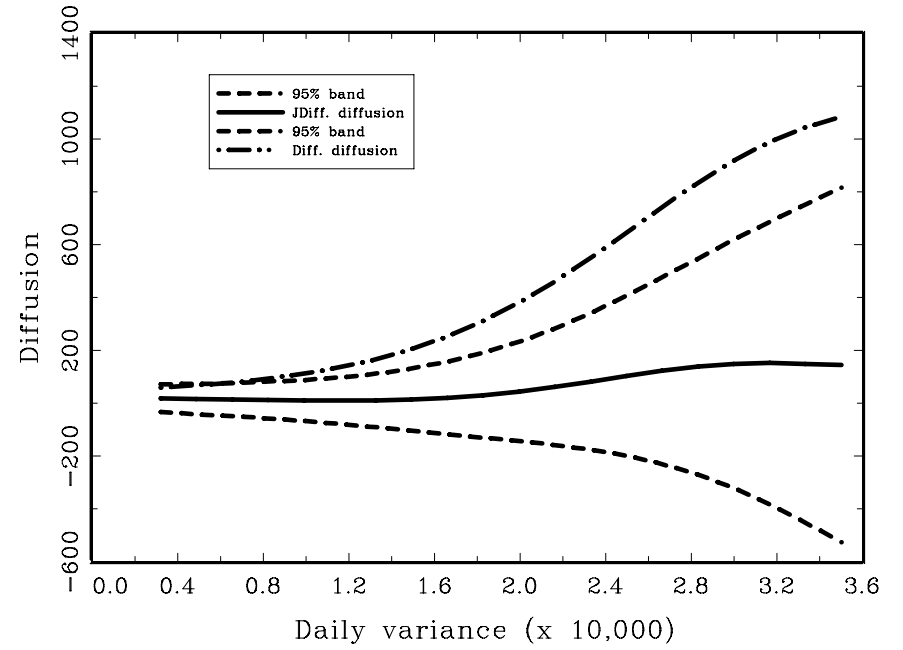


Figure 3(c)  
Variance jump intensity

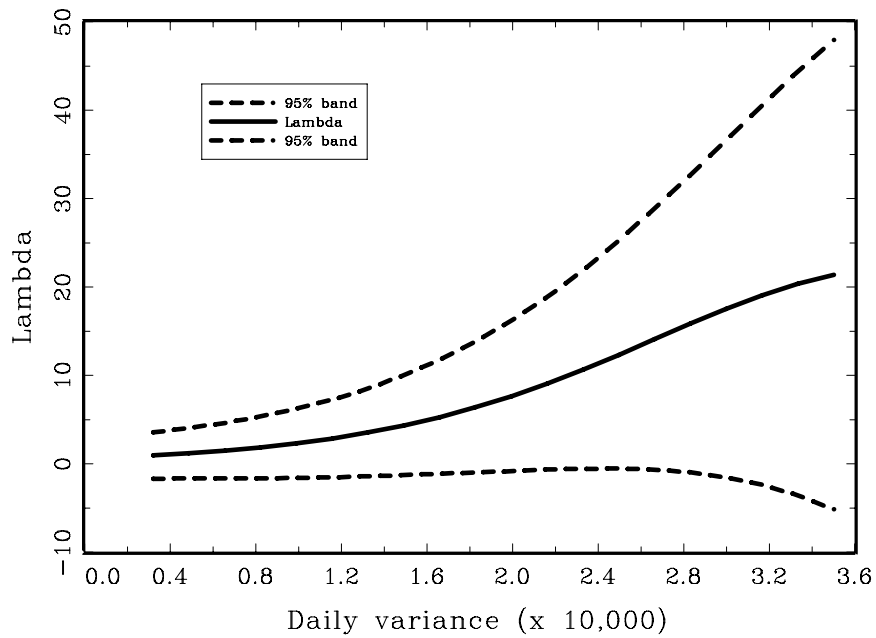


Figure 3(d)  
Variance expected jump size

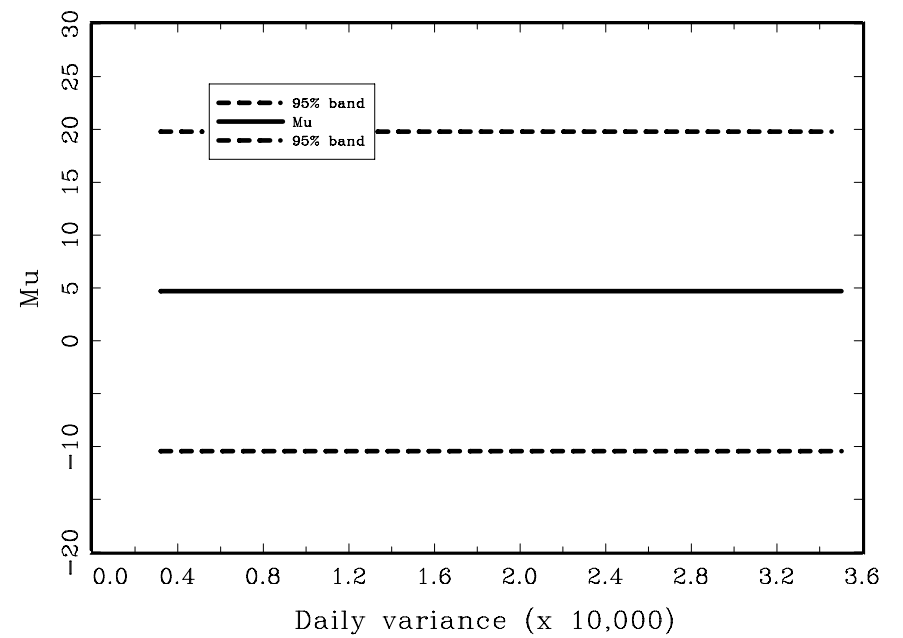


Figure 4(a)  
Variance drift estimates

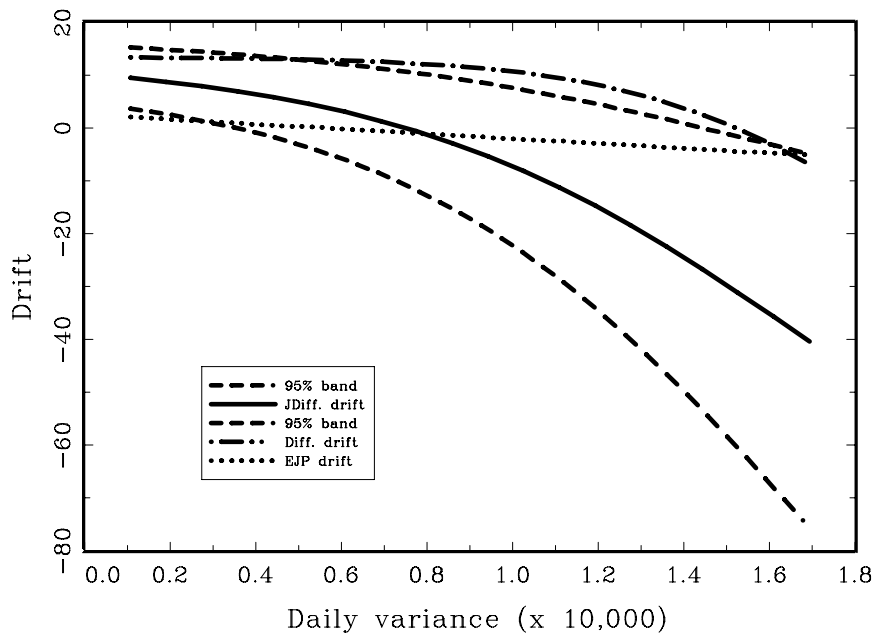


Figure 4(b)  
Variance diffusion estimates

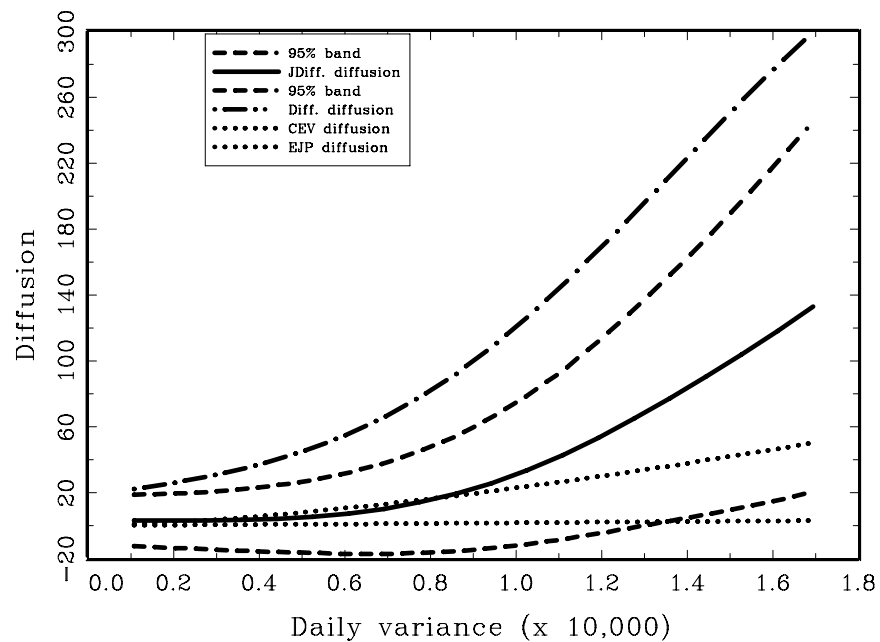


Figure 4(c)  
Variance jump intensity

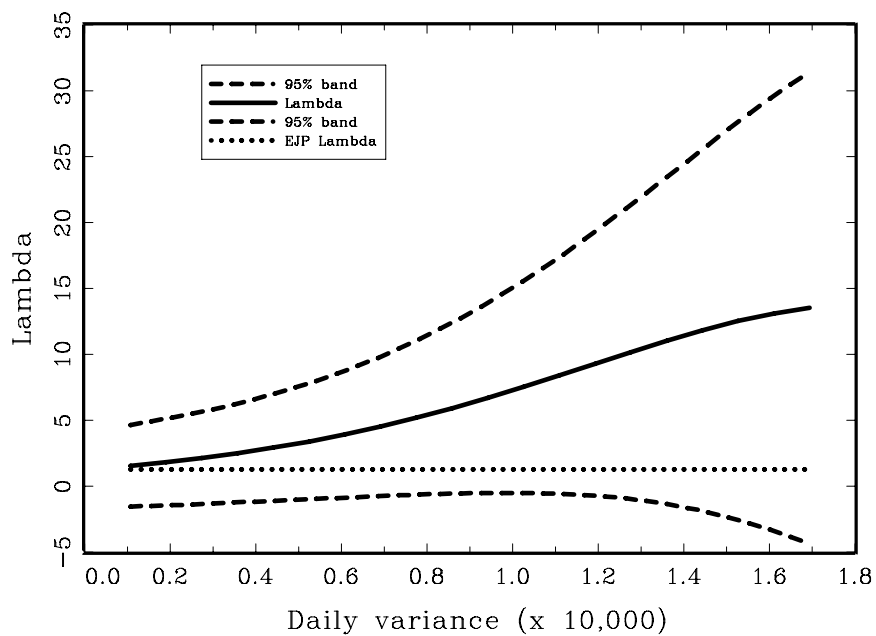


Figure 4(d)  
Variance expected jump size

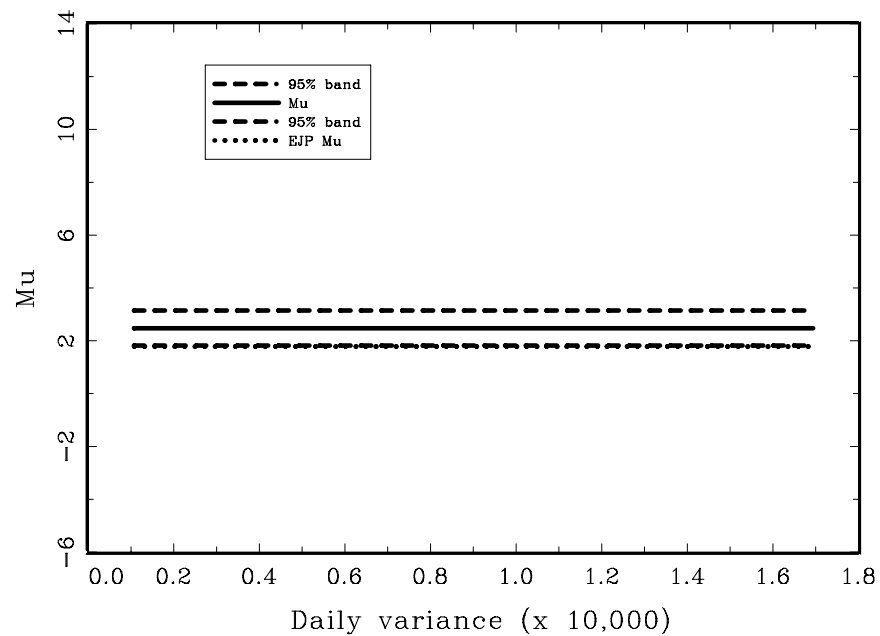


Figure 5(a)  
Standardized variance estimates  
using estimated drift and EJP diffusion

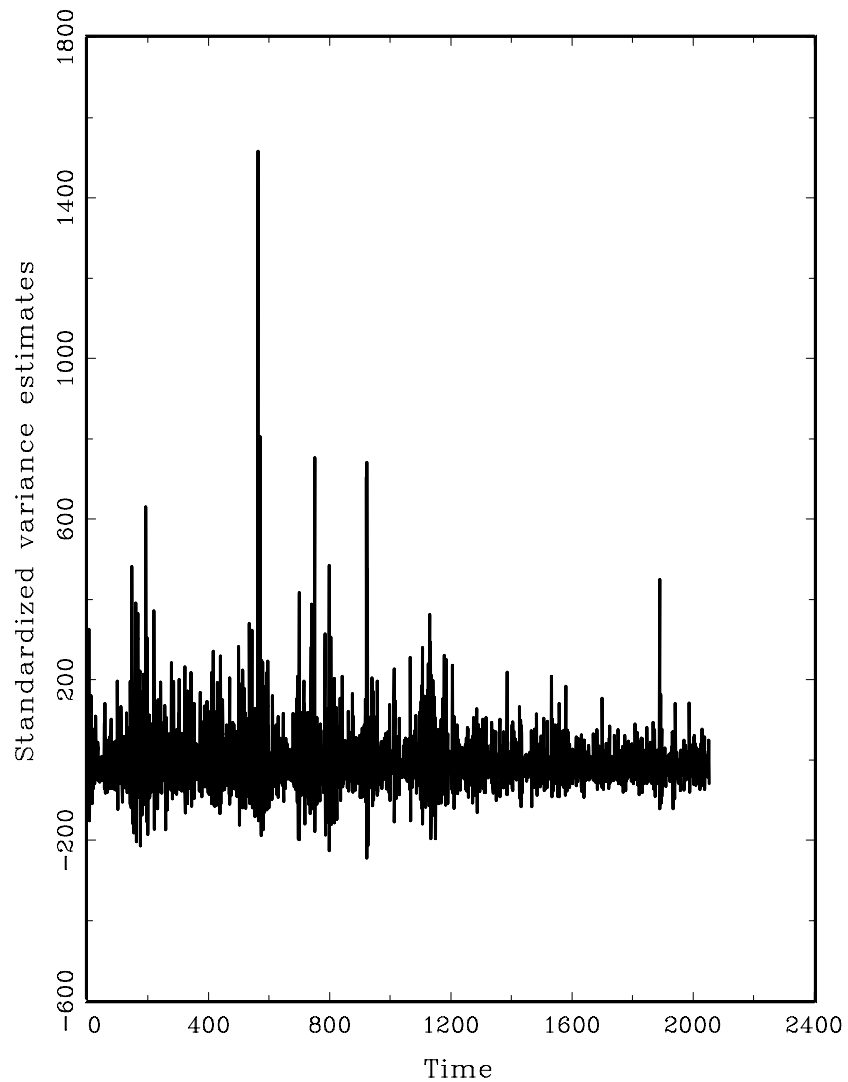


Figure 5(b)  
Standardized variance estimates  
using estimated drift and CEV diffusion

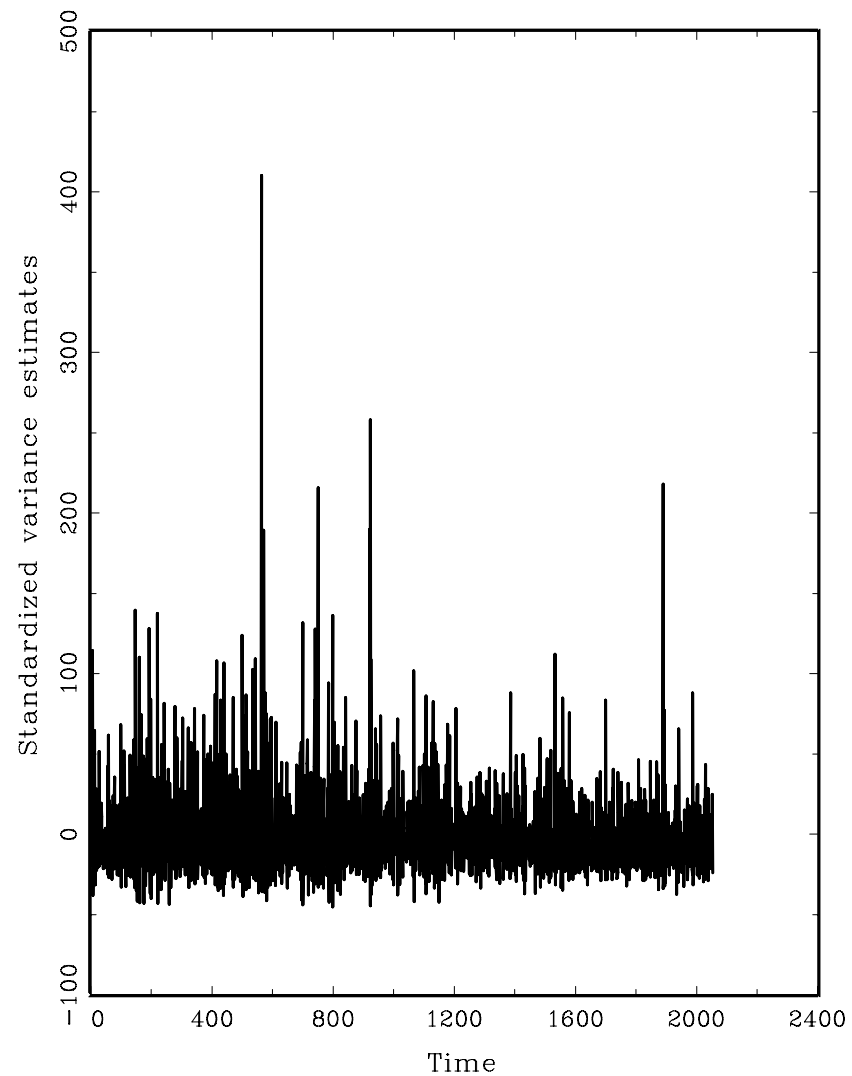


Figure 6(a)  
Price drift estimates

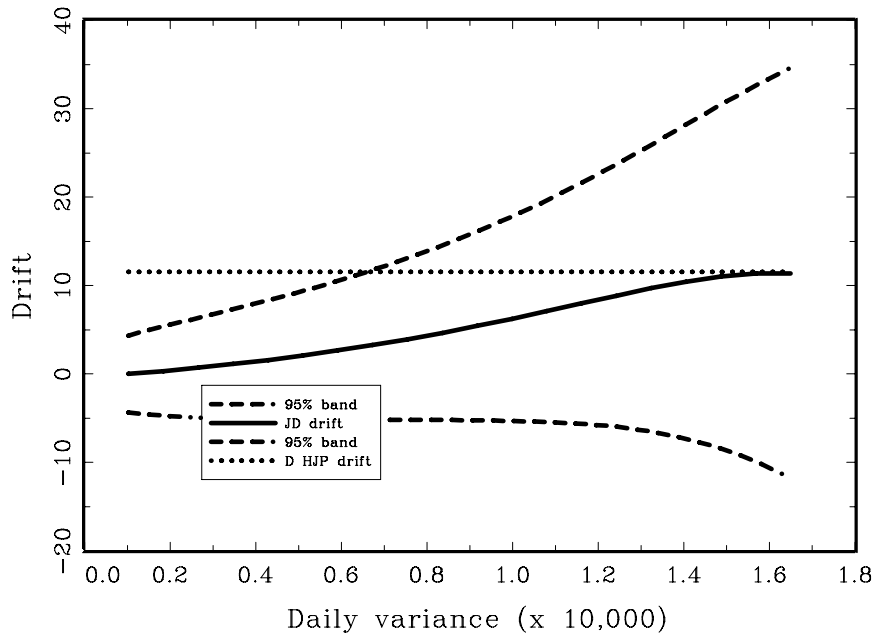


Figure 6(b)  
Leverage estimates

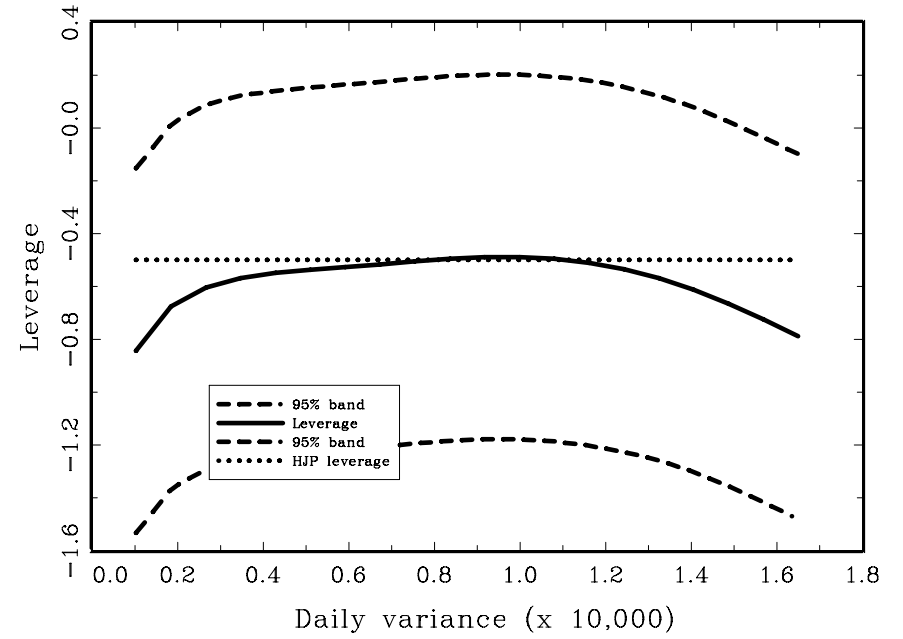


Figure 6(c)  
Price jump intensity

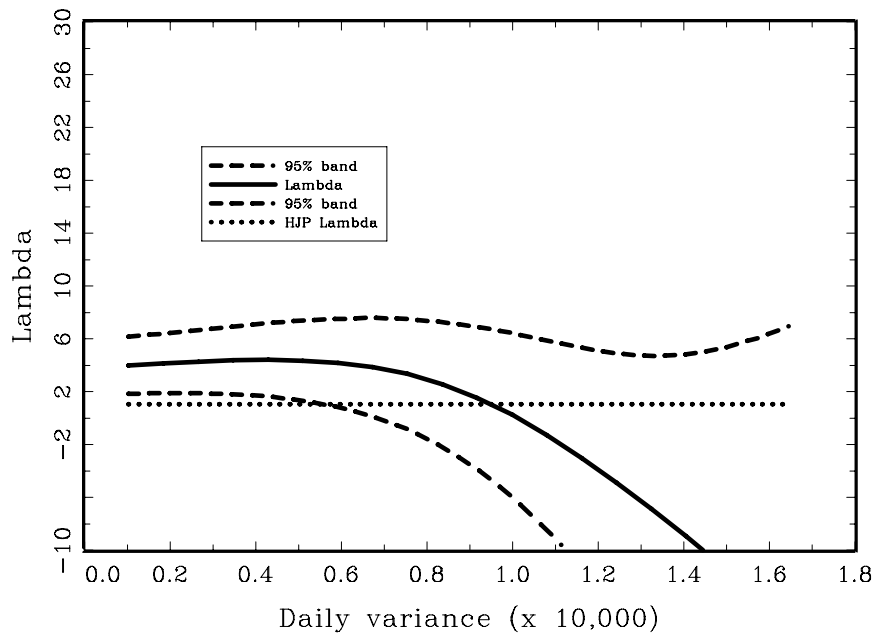


Figure 6(d)  
Price jump standard deviation

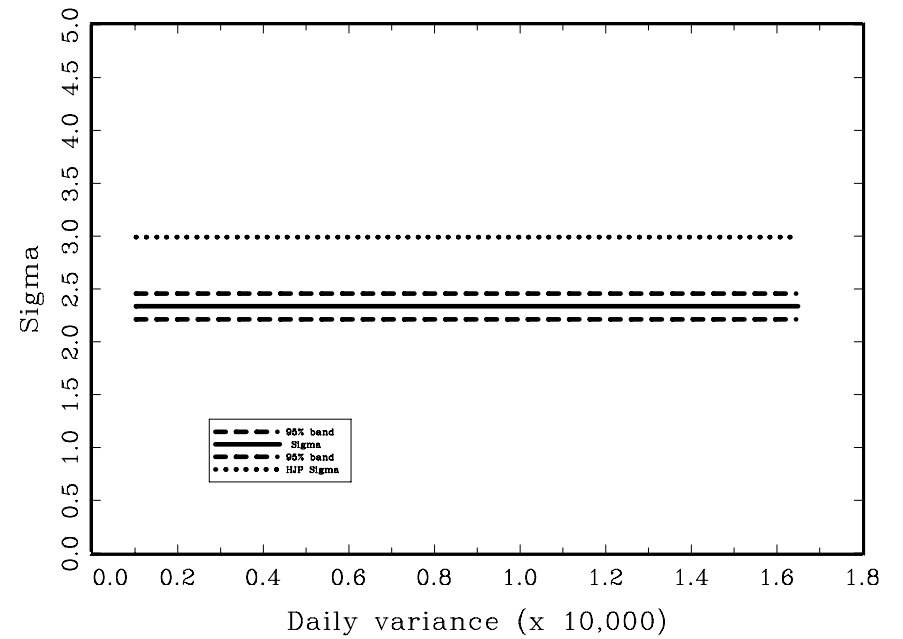


Figure 7(a)  
Price conditional kurtosis

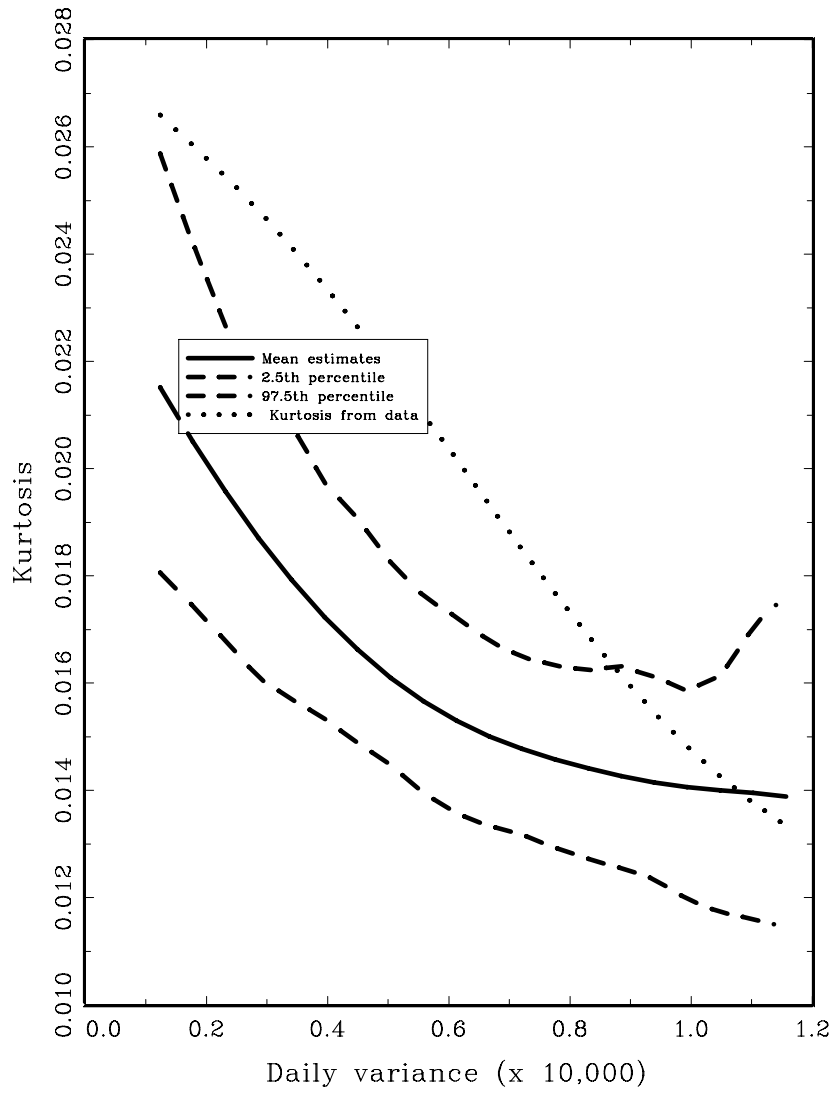


Figure 7(b)  
Variance conditional kurtosis

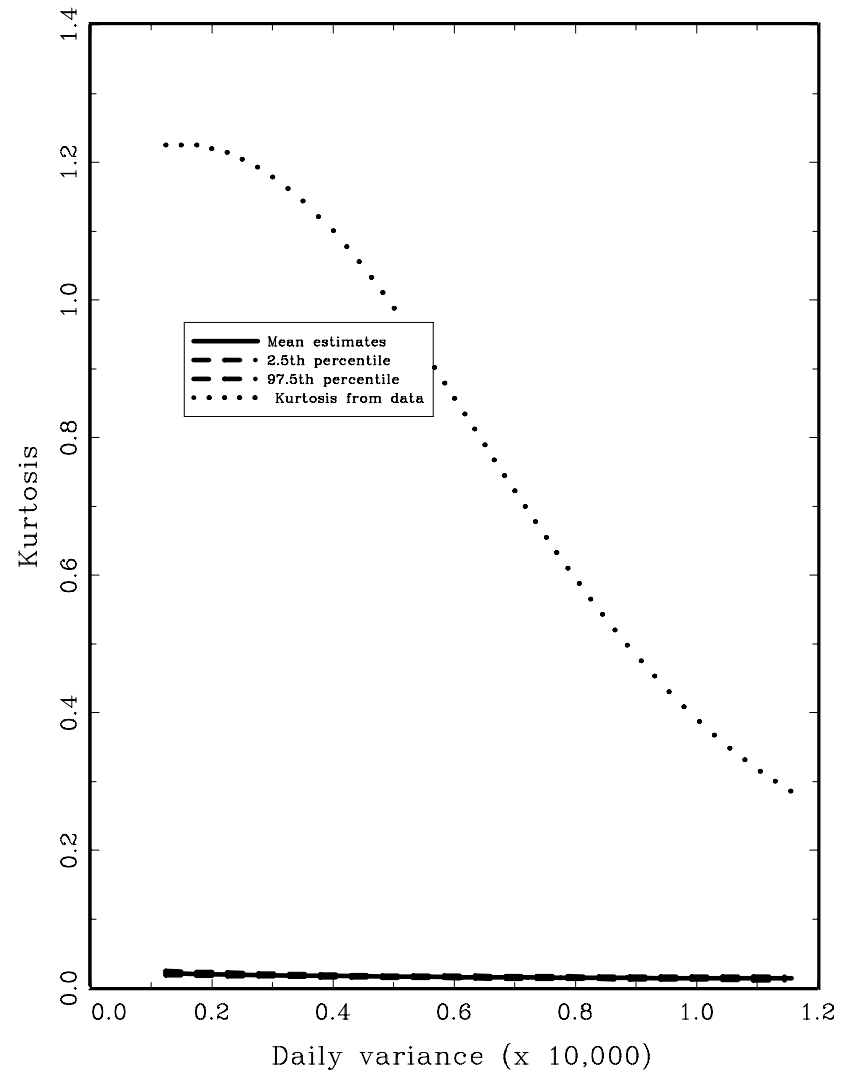


Figure 8(a)  
Price conditional kurtosis

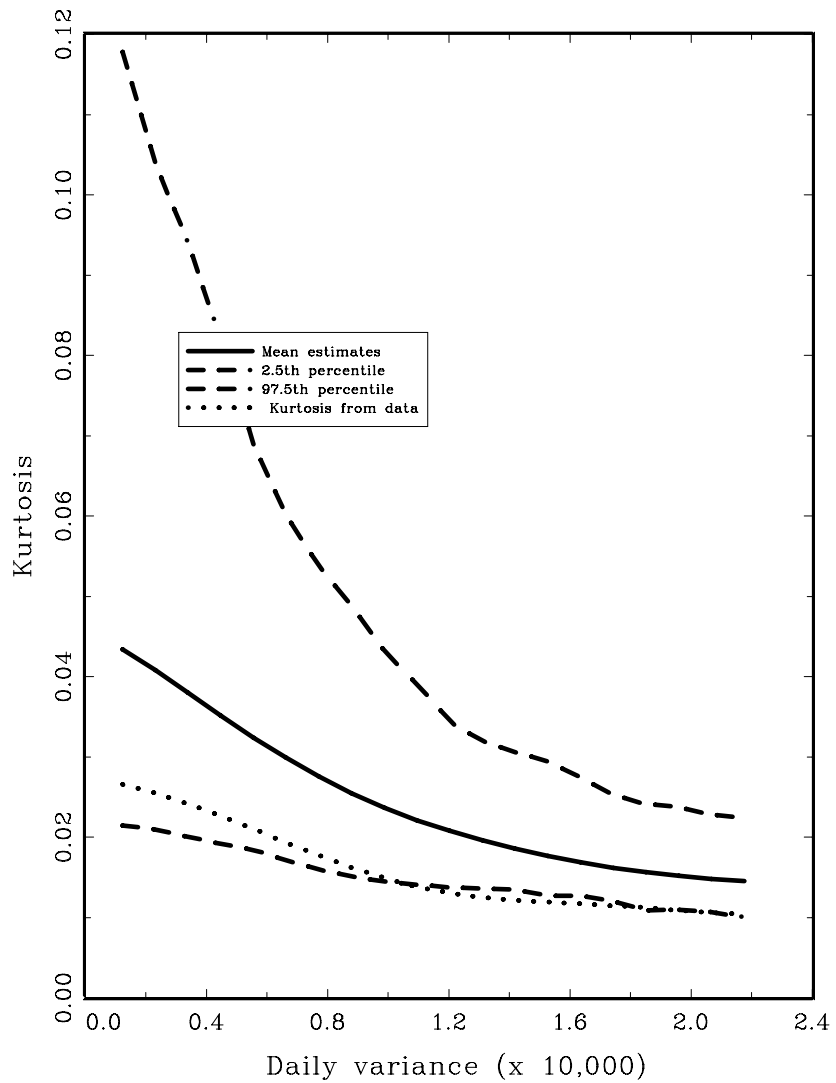


Figure 8(b)  
Variance conditional kurtosis

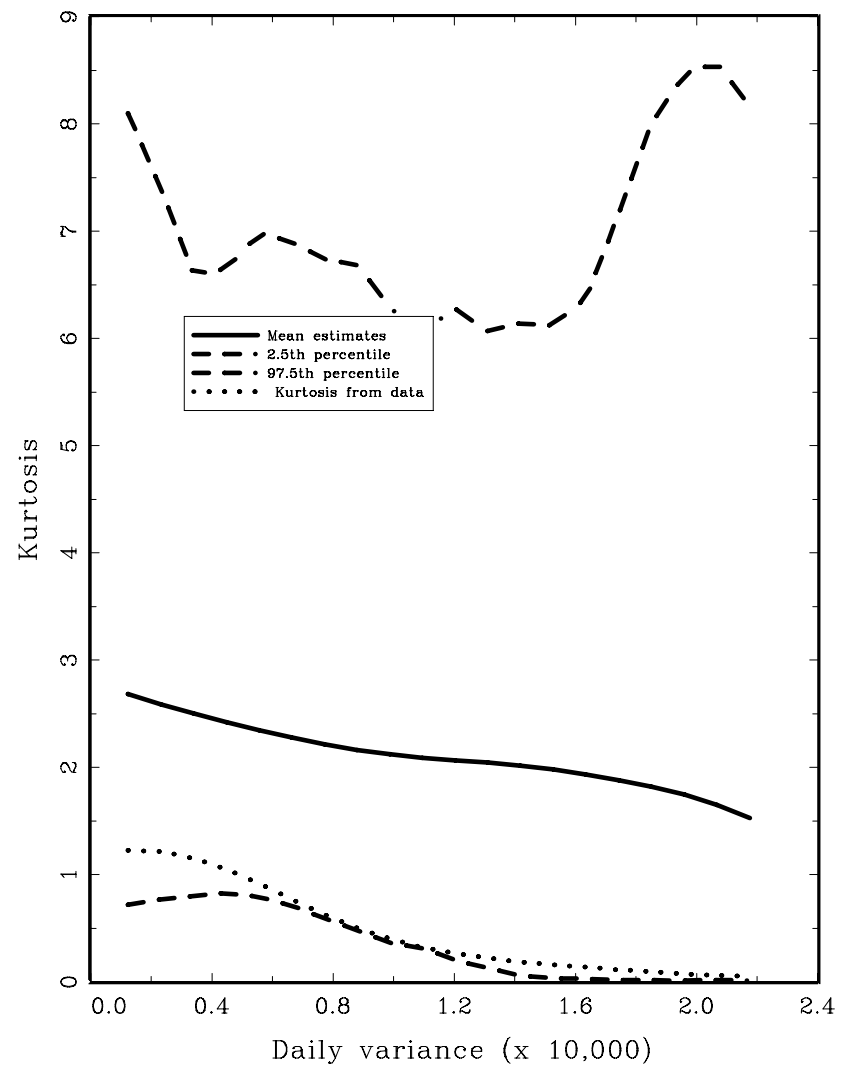


Figure 9(a)  
Price conditional kurtosis

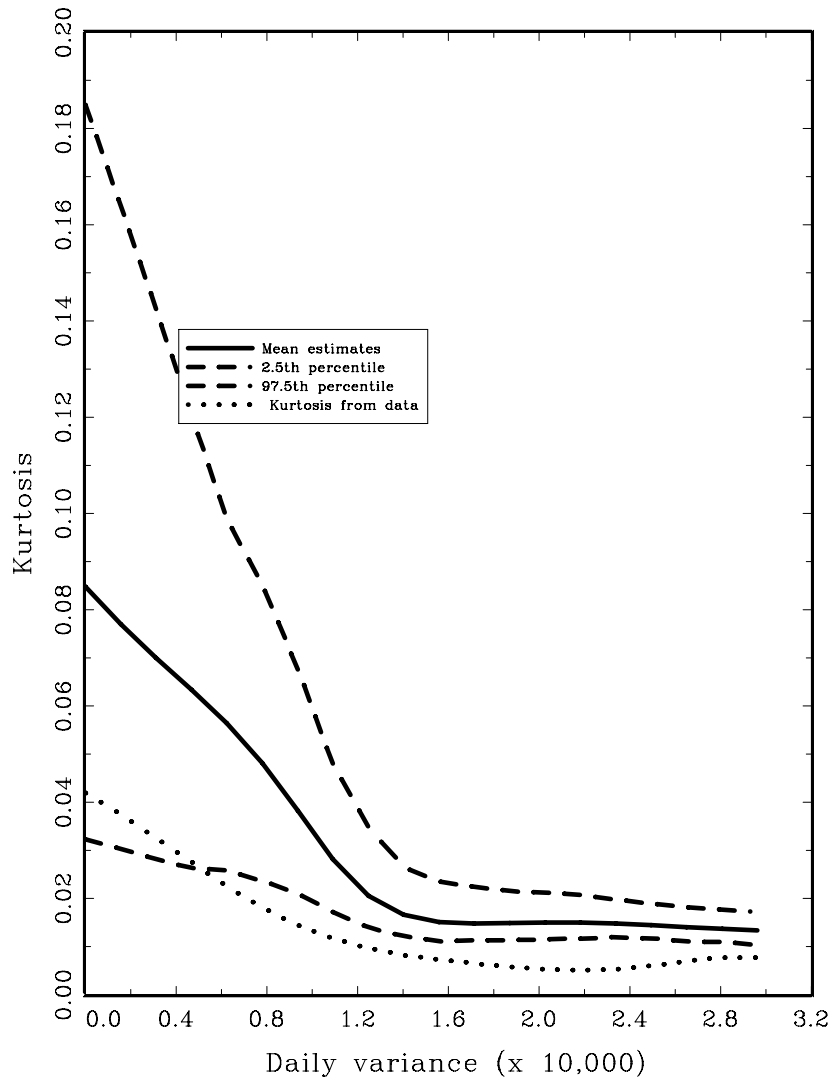


Figure 9(b)  
Variance conditional kurtosis

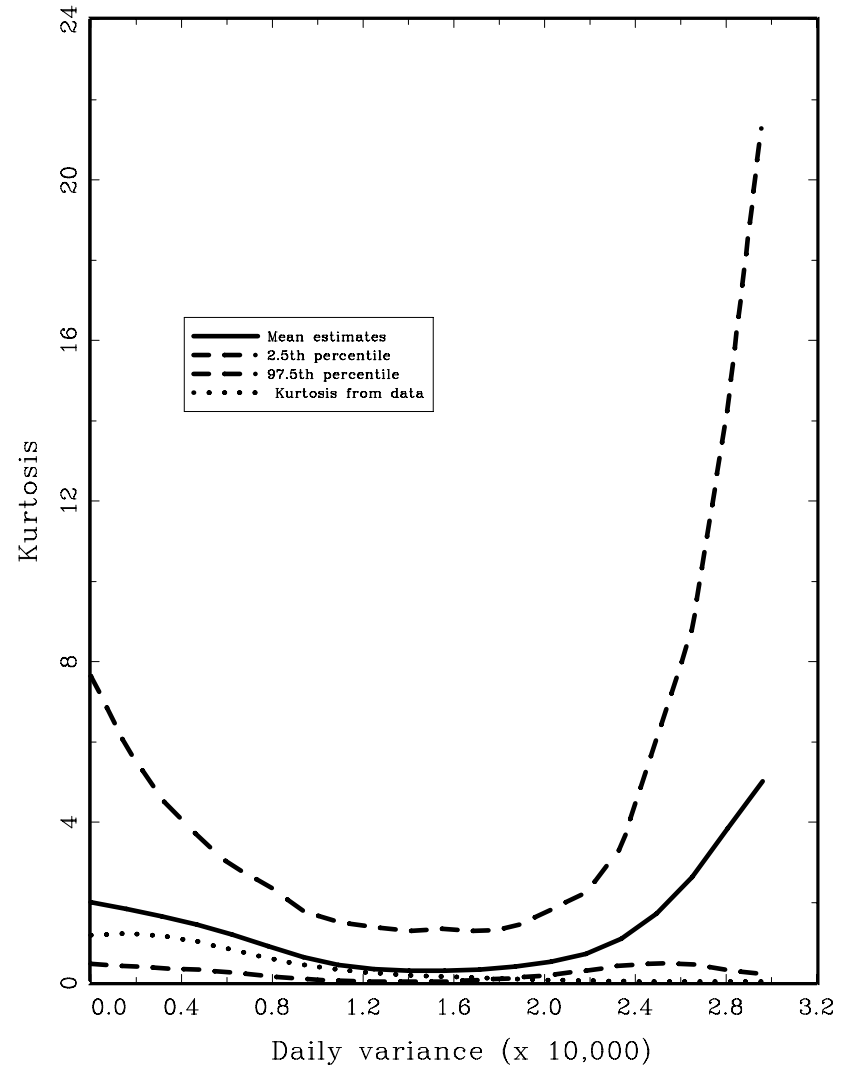


Figure 10(a)  
Drift

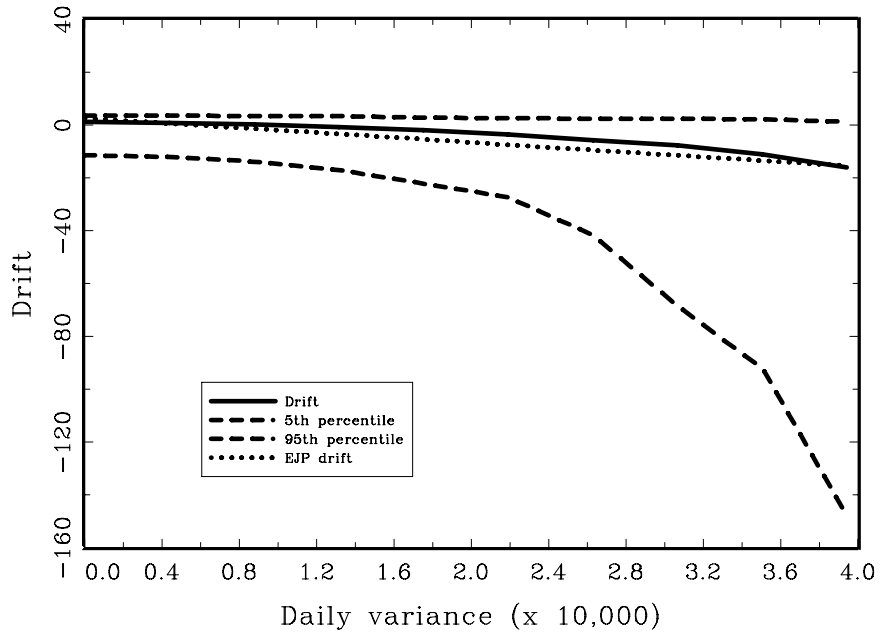


Figure 10(b)  
Diffusion

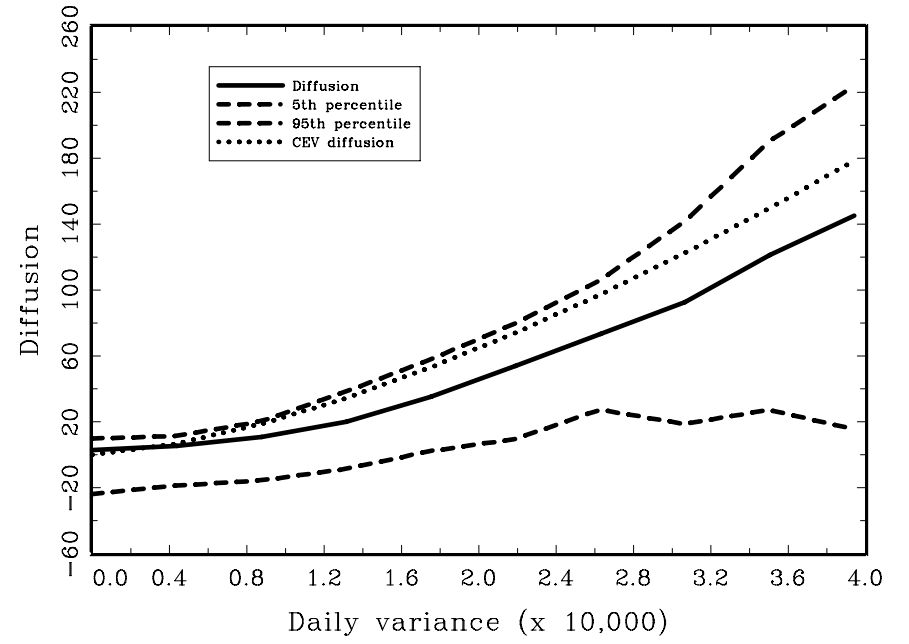


Figure 10(c)  
Lambda

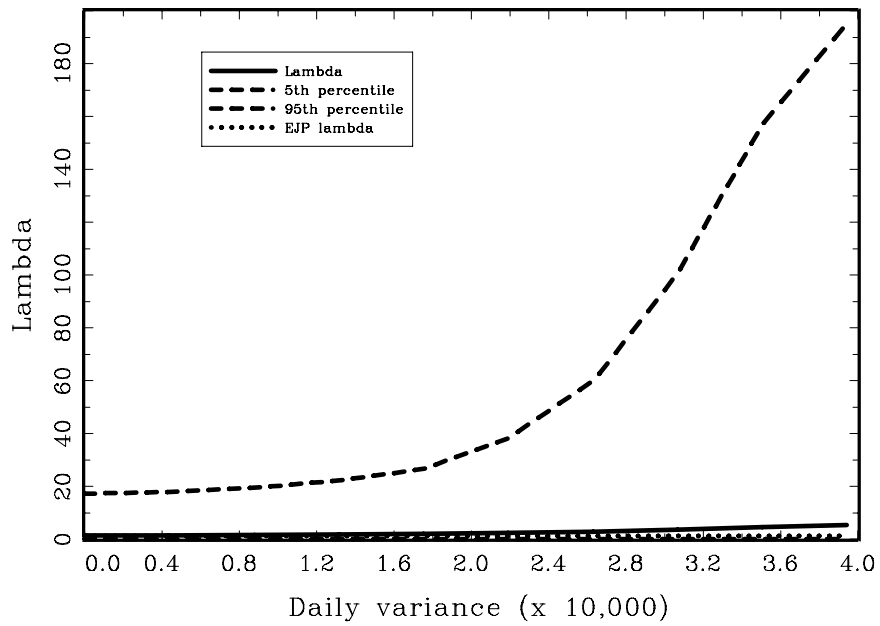


Figure 10(d)  
Variance expected jump size

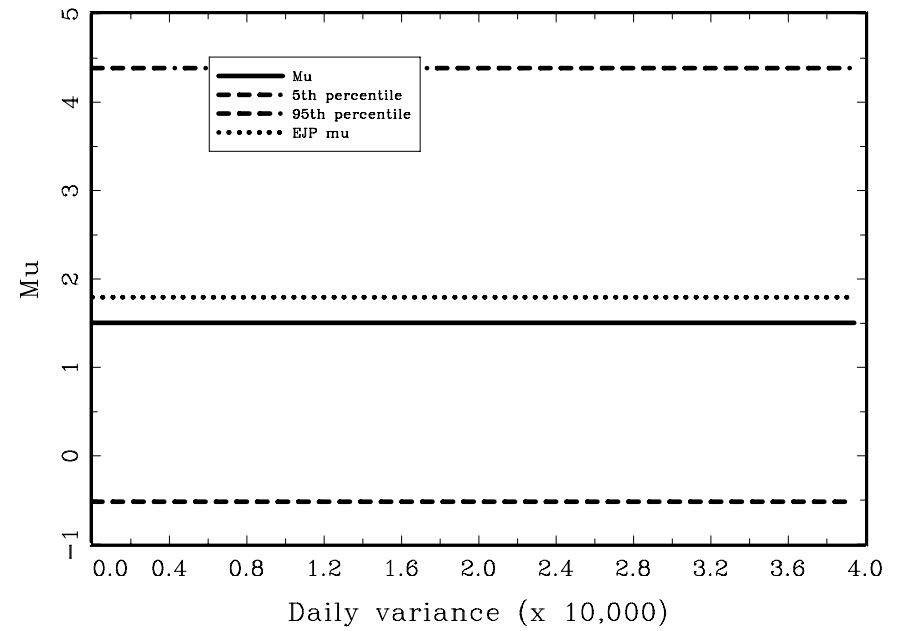


Figure 11(a)  
Return drift

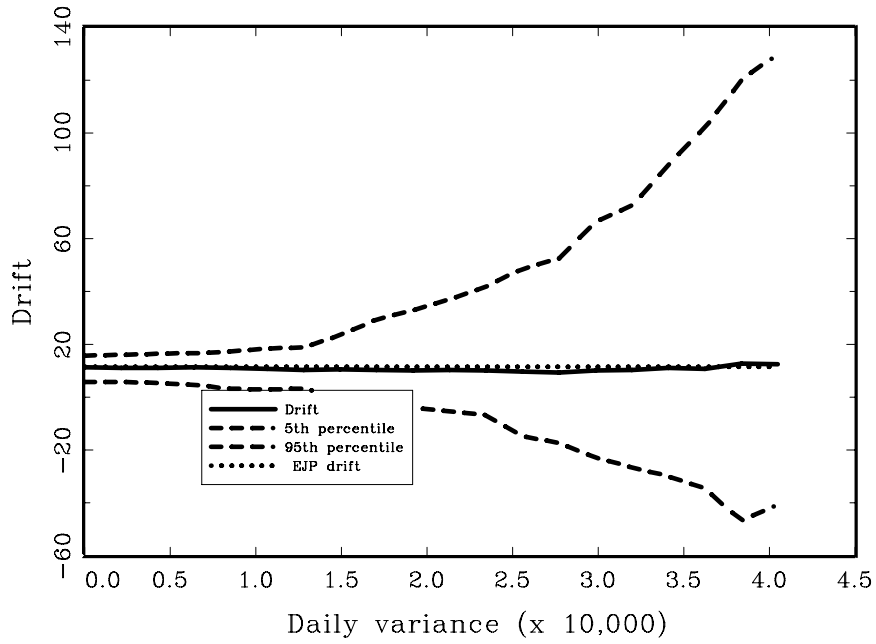


Figure 11(b)  
Leverage

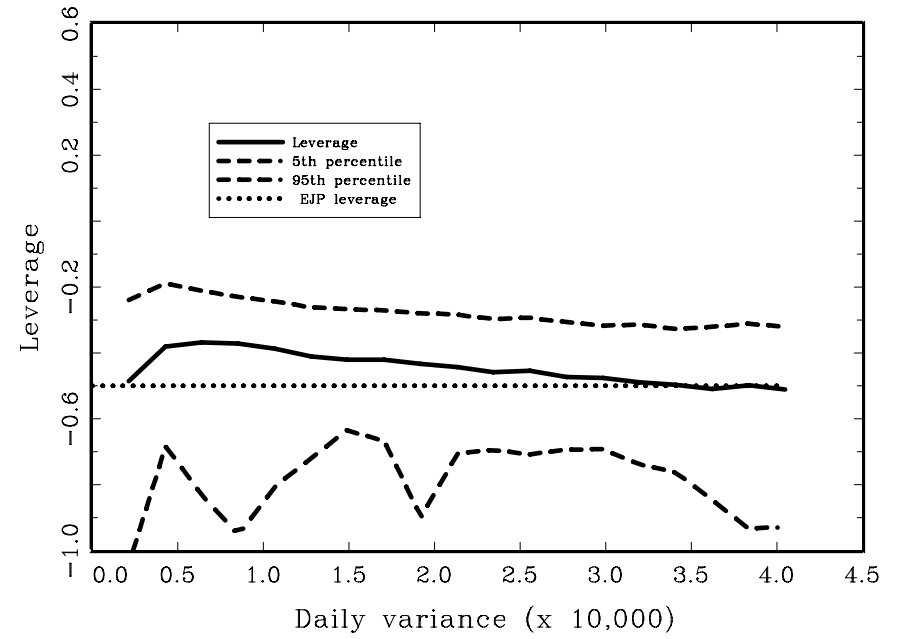


Figure 11(c)  
Price jump intensity

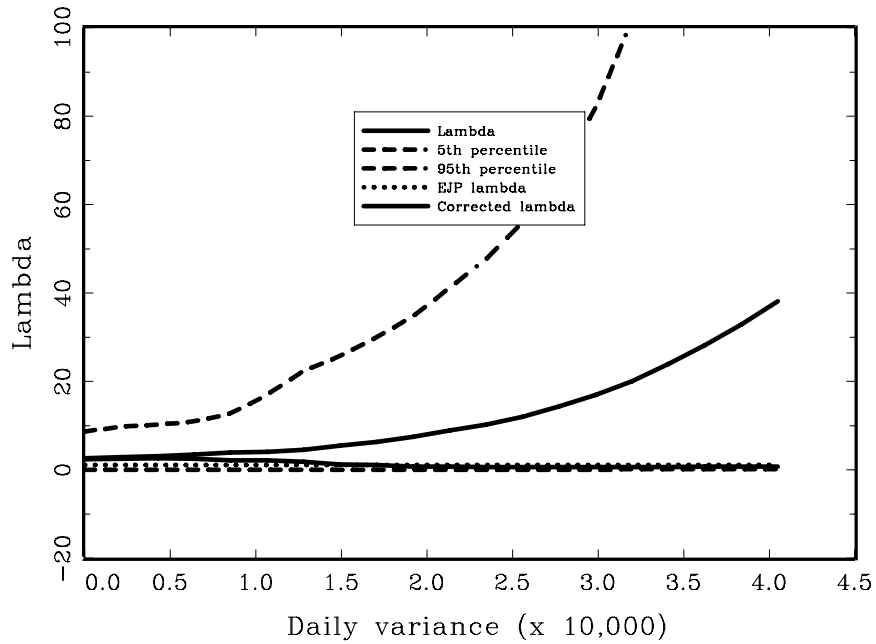


Figure 11(d)  
Price jump standard deviation

