NONPARAMETRIC STOCHASTIC VOLATILITY*

Federico M. Bandi† Roberto Renò‡

First draft: October 2007
This draft: February 2008

Abstract

Using recent advances in the nonparametric estimation of continuous-time processes under mild statistical assumptions as well as recent developments on nonparametric volatility estimation by virtue of market microstructure noise-contaminated high-frequency asset price data, we provide functional methods for stochastic volatility modelling. Our methods allow for the joint evaluation of return and volatility dynamics with nonlinear drift and diffusion functions, leverage effects, jumps in returns and volatility with possibly state-dependent jump intensities, as well as risk-return trade-offs. Our identification approach and asymptotic results apply under mild 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 CEV specification) even when allowing for jumps in the variance dynamics.

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

---

*Preliminary and incomplete. Comments very welcome.
†Graduate School of Business, University of Chicago.
‡Dipartimento di Economia Politica, Università di Siena, Piazza San Francesco 7, 53100, Siena, Italy.
1 Introduction

Understanding volatility is of fundamental importance for effective portfolio choice, derivative pricing, and risk management, among other issues. A successful, recent 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) are important exceptions and 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:

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

$$df(\sigma_t^2) = m_f(.)dW_t^\sigma + \lambda_f(.)dJ_t^\sigma;$$

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

Our procedures have three main features. First, we use high-frequency estimates of daily integrated variance $\int \sigma_s^2 ds$ to identify the parameters and functions driving volatility dynamics (i.e., $\lambda_f(.)$, $m_f(.)$, $\lambda_f(.)$ and, given parametric assumptions on the jump size distribution, the moments of the volatility jumps). Since the classical realized variance estimator (i.e.,
the sum of squared intra-daily returns over the day) may contain substantial contaminations due to market microstructure noise (as emphasized by Bandi and Russell, 2008, and Zhang at al., 2005, in recent work), we employ robust (to noise) integrated variance estimates. In other words, we allow for market microstructure noise and control for it.\footnote{For recent surveys of nonparametric methods for daily variance estimation using market microstructure noise-contaminated high-frequency asset price data, we refer the reader to the review papers of Bandi and Russell (2007), Barndorff-Nielsen and Shephard (2007), and McAleer and Medeiros (2008).} Second, differently from much existing work on stochastic volatility modelling, we avoid imposing tight (possibly affine) parametric structures on $\lambda_f(.)$, $m_f(.)$, and $\Lambda_f(.)$. Specifically, we identify the relevant functions (through estimates of the system’s infinitesimal moments) using nonparametric kernel methods for diffusion and jump-diffusion processes as proposed by Bandi and Nguyen (2003), Bandi and Phillips (2003), and Johannes (2004) in recent work. In order to lay out the main ideas in the context of a well-understood estimation framework, we use classical Nadaraya-Watson kernel estimates. However, extensions to alternative functional estimation methods are straightforward, as shown in other contexts and as discussed below. 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 are arguably particularly suitable for our problem.

We present preliminary ideas in the no jump case ($dJ^r_t = 0$, $dJ^\sigma_t = 0$) - Section 4. We then consider the empirically-important case of jumps in volatility ($dJ^r_t = 0$, $dJ^\sigma_t \neq 0$) - Section 5. For clarity, two alternative models (and corresponding identification methods) are presented. First, we 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. Second, we 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 ($dJ^r_t \neq 0$, $dJ^\sigma_t \neq 0$) - Section 6. When focusing on the full system (in Section 7) we discuss nonparametric evaluation of nonlinear risk-return trade-offs ($\mu_t(\sigma^2_t)$) and leverage effects.

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

2 The 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 over each sub-interval of size $\phi_{n,T}$.

The functions driving the dynamics of diffusions and jump-diffusions models 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 volatility process, i.e.,

$$\theta^j(x) = \frac{1}{\Delta_{n,T}} \mathbb{E}\left[ (f(\sigma_{i+\Delta}^2) - f(\sigma_i^2))^j \middle| \sigma_i^2 = x \right], \quad j = 1, \ldots,$$

by virtue of

$$\hat{\theta}^j(x) = \frac{1}{\Delta_{n,T}} \sum_{i=1}^{n-1} K\left( \frac{\bar{\sigma}_{iT/n}^2 - x}{\phi_{n,T}} \right) \frac{f(\bar{\sigma}_{i+1T/n}^2) - f(\bar{\sigma}_{iT/n}^2)}{\phi_{n,T}}, \quad j = 1, \ldots,$$

where $\bar{\sigma}_{iT/n}^2 = \frac{\hat{V}_{iT/n}}{\phi_{n,T}}$ and $\hat{V}_{iT/n}$ is a consistent estimate of $\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds$ for fixed $n, T,$ and $\phi_{n,T}$. The kernel function $K(.)$ satisfies the following properties:

**Assumption 1.** $K(.)$ is a bounded, continuously-differentiable, symmetric, and nonnegative function whose derivative $K'(.)$ is absolutely integrable and bounded, and for which $\int K(s)ds = 1$, $K_1 = \int s^2 K(s)ds < \infty$, and $K_2 = \int K^2(s)ds < \infty$.

**Assumption 2.** $\hat{V}_{iT/n}$ is such that
\[ E_{\sigma^2} \left( \phi_{n,T}^\beta \left( \frac{\tilde{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{i/T/n}^{i/T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right)^a \approx 0 \]  

(3)

and

\[ V_{\sigma^2} \left( \phi_{n,T}^\beta \left( \frac{\tilde{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{i/T/n}^{i/T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right)^a \approx \left( a \left( \sigma_{iT/n}^A \right)^\eta + b \right) \]  

(4)

with \( \alpha \in (0, \frac{1}{2}] \) and \( \beta \in [0, 1] \) given \( T \) and \( n \). \( E_{\sigma^2} \) and \( V_{\sigma^2} \) denote expectation and variance conditional on the spot volatility path. \( a, b, \) and \( \eta \) are numbers. The symbol \( \approx \) denotes asymptotic equivalence for a large \( k \) and a small \( \phi_{n,T} \).

Reno’ (2006) provides simulation evidence for the performance of \( \hat{\theta}_j^j(x) \) with \( j = 1, 2 \) (i.e., drift and diffusion) in the case of stochastic volatility models without jumps. Coherently with Bandi and Nguyen (2003) and Bandi and Phillips (2003), the asymptotics are derived here 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 the \( T, n, k, \) 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.** Virtually all recently proposed integrated variance estimators have asymptotic (for a specific - large - number of subsamples/autocovariances - see Section 8) variances and biases which may be represented as in Eq. (4) and Eq. (3). Consider the classical realized variance estimator in the absence of market microstructure noise, for instance. In this case, \( \alpha = \frac{1}{2}, \beta = 0, a = 2, b = 0, \) and \( \eta = 1 \). The two-scale estimator of Zhang et al. (2005) with a number of subsamples \( K \) equal to \( \tau k^{2/3} \) (with \( \tau \) fixed) has \( \alpha = \frac{1}{6}, \beta = 1, a = 0, \) and \( b \neq 0 \). If \( K = \tau \left( \frac{k}{\phi_{n,T}} \right)^{2/3}, \phi_{n,T} \) is such that \( \frac{K}{k} \rightarrow 0, \) and \( \tau \) is chosen optimally, then \( \alpha = \frac{1}{6}, \beta = \frac{1}{3}, a \neq 0, b = 0, \) and \( \eta = \frac{2}{3} \). The realized kernels of Barndorff-Nielsen et al. (2006) with a number of autocovariances equal to \( \tau k^{1/2} \) (with \( \tau \) fixed) and a kernel function \( g(.) \) satisfying \( g'(0) = 0 \) and \( g'(1) = 0 \) have \( \alpha = \frac{1}{4}, \beta = 1, a = 0, \) and \( b \neq 0 \). Explicit expressions for these estimators are provided in Section 8 below. In our empirical work, we optimize their finite sample properties by virtue of the MSE-based methods proposed by Bandi and Russell (2006, 2008). Appendix A relates Assumption 2 to a broader class of estimators (and choices of the number of subsamples/autocovariances) recently proposed in the literature while, importantly, providing details on the relevant parameters \( \alpha, \beta, a, b, \) and \( \eta \).
Remark 2. (Spot volatility estimation using realized variance.) In the absence of market microstructure noise, for realized variance we have:

\[
V_{\sigma^2}\left(k_{1/2}^2 \left( \frac{\hat{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right) \approx \left( 2 \left( \frac{\phi_{n,T} \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}^2} \right) \right) \approx 2\sigma_{iT/n}^4
\]

since \(\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds \overset{a.s.}{\to} \sigma_{iT/n}^4\) as \(\phi_{n,T} \to 0\). By the modulus of continuity of Brownian motion, notice that

\[
k_{1/2}^2 \left( \frac{\hat{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) + \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} - \sigma_{iT/n}^2
\]

\[
= k_{1/2}^2 \left( \frac{\hat{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) + k_{1/2}^2 \text{a.s.} \left( \sup_{\frac{1}{n} \leq i \leq \frac{T}{n} + \phi_{n,T}} \sigma_s^2 - \sigma_{iT/n}^2 \right)
\]

\[
= k_{1/2}^2 \left( \frac{\hat{V}_{iT/n}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) + o_{a.s.} \left( k_{1/2} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} \right)
\]

\[
= O_p(1)
\]

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

\[
k_{1/2} \left( \frac{\hat{V}_{iT/n}}{\phi_{n,T}} - \sigma_{iT/n}^2 \right) \underset{k \to \infty, \phi_{n,T} \to 0}{\Rightarrow} MN \left( 0, 2\sigma_{iT/n}^4 \right) \tag{5}
\]

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

\[
\phi_{n,T}^\alpha \left( \frac{\hat{V}_{iT/n}}{\phi_{n,T}} - \sigma_{iT/n}^2 \right) \underset{k \to \infty, \phi_{n,T} \to 0}{\Rightarrow} MN \left( 0, a \left( \sigma_{iT/n}^4 \right)^{\eta} + b \right) \tag{6}
\]
Appendix A provides further details while specializing this result to a variety of estimators recently proposed to estimate integrated variance in the absence of noise as well as in the presence of noise. In the case of the two-scale estimator of Zhang et al. (2005) and the class of flat-top kernel estimators of Barndorff-Nielsen et al. (2006) the result also requires a vanishing ratio between the number of subsamples/autocovariances and the number of observations which is, of course, standard (see Appendix A).

**Remark 4. (More on spot volatility estimation.)** The quantity \( \tilde{\sigma}_{i/n}^2 = \frac{\hat{v}_{i/T}}{\phi_{n,T}} \) is a spot variance estimator constructed using an integrated measure. Different spot volatility estimators have been recently proposed by Malliavin and Mancino (2002) and Kristensen (2007). Renò (2007) uses the former to estimate the functions \( m(.) \) and \( \Lambda(.) \) in Eq. (2) for the case without jumps in either volatility or returns. When just averaging 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 smooths squared continuously-compounded returns locally, i.e.,

\[
\tilde{\sigma}_{i/n}^2 = \frac{1}{h} \sum_{j=1}^{nk} K\left(\frac{j-i/n}{h}\right) r_j^2 \quad i = 1, \ldots, n,
\]

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

\[
\sqrt{nk h}(\tilde{\sigma}_{i/n}^2 - \sigma_{i/n}^2) \Rightarrow MN(0, 2K2\sigma_i^4)
\]

holds if, in addition, \( nk h^{1+2\gamma} \to 0 \), 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 \( \hat{v}_{i/n} \) is realized variance and noise is absent. Write

\[
\hat{V}_{i/n} = \frac{\sum_{j=1}^{k} r_j^2}{\phi_{n,1}} = \frac{1}{\phi_{n,1}} \sum_{j=1}^{nk} \sum_{\{0 \leq i-j/n \leq 1\}} 1(r_j) r_j^2,
\]
where $1_{\{\cdot\}}$ is the indicator kernel. Hence, $\hat{V}_{i/n}$ has an interpretation in terms of kernel smoother. We are simply averaging (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 $\hat{\sigma}_{i/n}^2$. This said, our derived asymptotic distribution in Eq. (5) ought to be consistent with the asymptotic distribution in Eq. (7) for the case where $K(\cdot) = 1_{\{\cdot\}}$. In other words, if $nk\phi_{n,1} \to \infty$ and $nk\phi_{n,1}^{1+2\gamma} \to 0$,

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

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 consistent with Eq. (5). Similarly, the asymptotic variance becomes $2\sigma_{i/n}^4$, which is also consistent with Eq. (5). Finally, the condition for a vanishing asymptotic bias term $nk\phi_{n,1}^{1+2\gamma} = k\phi_{n,1}^{2\gamma} = k^{1/2}\phi_{n,1} = o(1)$ is equivalent to $k^{2 \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2}} = o(1)$ for our assumed degree of smoothness of the spot volatility process.

From a theoretical standpoint, the use of smooth kernels as in Kristensen (2007) yields efficiency gains over the equal weighting implicitly delivered by our methods. In fact, the term $K_2$ is generally smaller than 1 (it is, for example, equal to $\frac{1}{2\sqrt{\pi}}$ for a second-order Gaussian kernel). From an empirical standpoint, the presence of intraday seasonalities (see, e.g., Andersen and Bollerslev, 1998, and the references therein) might affect estimates based on smooth kernels in ways that are difficult to predict. Diurnal effects appear more likely to average out when using equal weighting over a trading day as implied by integrated variance-type measures, such as realized variance. Importantly, the properties of kernel estimates of spot variance have not been studied for the case with jumps and market microstructure noise, which is relevant for our purposes. Using integrated variance to evaluate spot variance allows us to borrow from the recent literature on integrated variance estimation both in terms of limiting results and in terms of finite sample adjustments required for more accurate empirical implementation (see Appendix A and Section 8, respectively). The latter have been recently advocated by Bandi and Russell (2006, 2008).

**Intuition.** We now turn to the logic behind our estimation procedure. Given Remark 2 and 3, the rate of convergence of $\hat{\sigma}_{i/T,n}^2$ to $\sigma_{i/T,n}^2$ is $k^{\alpha}\phi_{n,T}^\beta$ (if $k^{\alpha}\phi_{n,T} \to \infty$, of course), and $k^{\alpha}\phi_{n,T}^\beta \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} \to 0$, where $\alpha \in (0, \frac{1}{2}]$ and $\beta = [0, 1]$. If $k \to \infty$ at a fast enough
pace as $h_{n,T} \to 0$, then $\hat{j}_j(x)$ estimates $\theta^j(x)$ consistently (in probability). Identification of the functions of interest will rely on consistent estimation of $\hat{j}_j(x)$, for $j = 1, \ldots$, 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 immediate. Among other methods, $\hat{j}_j(x)$ could be a local linear estimate of the form

$$\hat{j}_j(x) = \frac{1}{\Delta n,T} \sum_{i=1}^{n-1} \bar{K}_i(x, h_{n,T}) \left( f(\tilde{\sigma}_{(i+1)T/n}) - f(\tilde{\sigma}_{iT/n}) \right)^j$$

where $\bar{K}_i(x, h_{n,T}) = \frac{1}{h_{n,T}} K \left( \frac{\tilde{\sigma}_{iT/n} - x}{h_{n,T}} \right) \Gamma_{n,2} - \frac{1}{h_{n,T}} K \left( \frac{\tilde{\sigma}_{iT/n} - x}{h_{n,T}} \right) \Gamma_{n,1}$ with $\Gamma_{n,s} = \sum_{i=1}^{n} (\tilde{\sigma}_{iT/n} - x)^s \frac{1}{h_{n,T}} K \left( \frac{\tilde{\sigma}_{iT/n} - x}{h_{n,T}} \right)$ for $s = 1, 2$. More generally, it could be a local polynomial estimator defined as the solution $\{\alpha_0, \alpha_1, \ldots, \alpha_p\}$ to

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

where $\hat{j}_j(x) = \hat{\alpha}_0(x)$ for $p = 1$. Local polynomial methods for diffusions are discussed by Fan and Zhang (2003) and, under recurrence, by Moloche (2002), among others. Alternative, interesting approaches for diffusion estimation under mild recurrence assumptions have been recently proposed, inter alia, by Xu (2006, 2007).

### 3 Recurrence

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

$$X_t = X_0 + \int_0^t \mu(X_s^-)ds + \int_0^t \sigma(X_s^-)dW_s + \int_0^t \int \sigma(X_s, y)\nu(ds, dy),$$

where $\{W_t, \mathcal{F}_t\}$ is a standard $m$ dimensional Brownian motion and
\[ \overline{\mu}(dt, dy) = N(dt, dy) - \mathbb{E}(N(dt, dy)) \]

\[ = N(dt, dy) - \overline{\Pi}(dy) dt \]

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

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

\[
|\mu(x) - \mu(z)|^2 + ||\sigma(x) - \sigma(z)||^2 + \int |c(x, y) - c(z, y)|^2 \overline{\Pi}(dy) \leq C|x - z|^2,
\]

\[
|\mu(x)|^2 + ||\sigma(x)||^2 + \int |c(x, y)|^2 \overline{\Pi}(dy) \leq C(x - |z|)^2.
\]

Write \(a(x) = \sigma(x)\sigma(x)^T\). In addition, there exists a number \(\alpha > 0\) so that

\[
z^T a(x) z \geq \alpha |z|^2 \text{ for all } x \text{ and } z.
\]

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

**Assumption 4.** For each \(x \in \mathbb{R}^N,\)

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

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 \(x \in \mathbb{R}^N,\)

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

then the process \(X_t\) is recurrent (In-Suk Wee, 2000).
Remark 5. The model in Eq. (1) and Eq. (2) is not compensated. This is of course not problematic since we could compensate it and redefine the drift vector as being equal to \( \mu(.) = \{ \mu(.), m_{f(.)}(.) \}^T - \{ X^*(.), \lambda_f^*(.) \}^T \circ \{ E[(c^*(., y^*)], E[(c^*(., y^*))] \}^T \), 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.

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 stationary density. Bandi and Phillips (2004) provide further discussions.

We present 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: \( dJ^r_t = 0 \) and \( dJ^\sigma_t = 0 \)

In the absence of jumps, the estimated infinitesimal moments 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 \( \tilde{\theta}^1(x) \xrightarrow{p} m(x) \) and \( \tilde{\theta}^2(x) \xrightarrow{p} \Lambda^2(x) \) 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 \( k, n \to \infty \) and \( h_{n,T}, \phi_{n,T} \to 0 \) so that

\[
\lim_{n \to \infty} \frac{1}{h_{n,T}} \left( \Delta_{n,T} \log \frac{1}{\Delta_{n,T}} \right)^{1/2} = 0,
\]

\[
\lim_{k,n \to \infty} \frac{1}{h_{n,T}} \frac{k^\alpha \phi_{n,T}^{\beta}}{h_{n,T}} + \frac{1}{h_{n,T}} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} = 0,
\]

with \( \alpha \in (0, \frac{1}{2}] \) and \( \beta = [0, 1] \), then,
\[ \widehat{L}_{a^2}(T, x) = \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K} \left( \frac{\sigma^2_{i,T/n} - x}{h_{n,T}} \right) \xrightarrow{p} L_{a^2}(T, x), \]

where \( L_{a^2}(T, x) \) is the chronological local time of the spot volatility 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 (e.g., Bandi and Nguyen, 2003, and Bandi and Phillips, 2003). 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 stationary processes (since \( L_{a^2}(T, x)/T \xrightarrow{p} p(x) \), where \( p(x) \) is the time-invariant stationary density at \( x \)) but is lower for generic recurrent processes and, importantly, unknown in general. In what follows, we write \( L_{a^2}(T, x) \propto v(T) \), where \( v(T) \) is a regularly-varying function at infinity (see, e.g., Bandi and Moloche, 2004, for discussions). As said, \( v(T) = T \) if the process is strictly stationary or positive recurrent (ergodic). If the process is Brownian motion, then \( v(T) = T^{1/2} \).

**Remark 7.** In practice, the nature of the divergence properties of local time is immaterial for our purposes. All we need, in order to define the 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. We refer the reader to Bandi and Phillips (2004) for further discussions of theory and applications in alternative continuous-time frameworks.

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

\[
\lim_{n,T \to \infty} h_{n,T}v(T) = \infty,
\]

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

\[
\lim_{k,n,T \to \infty} \frac{T \sigma(T)^{-1}}{\Delta_{n,T} h_{n,T} k^n \phi_{n,T}^{\beta}} + \frac{T \sigma(T)^{-1}}{\Delta_{n,T} h_{n,T}} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} = 0,
\]

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

\[ \frac{\widehat{\theta}^1(x)}{p} m(x), \]
where \( \mathcal{L}_{\alpha^2}(T, x) \propto v(T) \). If

\[
\lim_{n,T \to \infty} h_{n,T}v(T) = \infty, \\
\lim_{n,T \to \infty} h_{n,T}^5 v(T) = C_1, \\
\lim_{n,T \to \infty} \frac{v(T)}{h_{n,T}} \left( \Delta_{n,T} \log \frac{1}{\Delta_{n,T}} \right)^{1/2} = 0, \\
\lim_{k,n,T \to \infty} \frac{T^{3/2}v(T)^{-1}}{\Delta_{n,T}h_{n,T}^{1/2}} + \frac{T^{3/2}v(T)^{-1}}{\Delta_{n,T}h_{n,T}^{1/2}} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} = 0,
\]
then,

\[
\sqrt{h_{n,T} \mathcal{L}_{\alpha^2}(T, x)} \left\{ \frac{\theta^1(x)}{m(x)} - m(x) - \Gamma_m(x) \right\} \Rightarrow \mathcal{N} (0, K_2 \Lambda^2(x)),
\]
with

\[
\Gamma_m(x) = h_{n,T}^2 K_1 \left[ m'(x) \frac{s'(x)}{s(x)} + \frac{1}{2} m''(x) \right],
\]
where \( s(dx) \) is the diffusion's speed measure and \( C_1 \) is a constant.

**Proof.** See Appendix B.

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

\[
\lim_{n,T \to \infty} \frac{v(T)}{h_{n,T}} \left( \Delta_{n,T} \log \frac{1}{\Delta_{n,T}} \right)^{1/2} = 0, \\
\lim_{k,n,T \to \infty} \frac{T^{3/2}v(T)^{-1}}{\Delta_{n,T}h_{n,T}^{1/2}} + \frac{T^{3/2}v(T)^{-1}}{\Delta_{n,T}h_{n,T}^{1/2}} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} = 0,
\]
with \( \alpha \in (0, \frac{1}{2}] \) and \( \beta = [0, 1] \), then,

\[
\tilde{\theta}^2(x) \Rightarrow \mathcal{L}_2(x),
\]
where \( \mathcal{L}_{\alpha^2}(T, x) \propto v(T) \). If
\begin{align*}
\lim_{n,T \to \infty} \frac{h_n^5 v(T)}{\Delta_n} &= C_2, \\
\lim_{n,T \to \infty} \frac{v(T)}{h_n^3} \left( \Delta_n \log \frac{1}{\Delta_n} \right)^{1/2} &= 0, \\
\lim_{k,n,T \to \infty} \frac{T^{3/2} v(T)^{-1}}{\Delta_{n,T}^{3/2} h_n^{1/2} \phi_{n,T}^{3/2}} + \frac{T^{3/2} v(T)^{-1}}{\Delta_{n,T}^{3/2} h_n^{1/2} \phi_{n,T}^{3/2}} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} &= 0,
\end{align*}

then,

\[
\sqrt{\frac{h_n T \sigma^2(T,x)}{\Delta_n}} \left\{ \tilde{\sigma}^2(x) - \Lambda^2(x) - \Gamma_A(x) \right\} \Rightarrow \mathcal{N}(0, 2K_2 \Lambda^4(x)),
\]

with

\[
\Gamma_A(x) = h_n^2 K_1 \left[ \Lambda' \left( \frac{s'}{s(x)} s'' + \frac{1}{2} \Lambda'' \right) \right],
\]

where \( s(dx) \) is the diffusion’s speed measure and \( C_2 \) is a constant.

**Proof.** See Appendix B.

**Remark 8.** Since \( \Delta_{n,T} \to 0 \), the diffusion estimator has a faster rate of convergence than the drift estimator. In both cases, optimal rate selection for the smoothing parameter \( h_{n,T} \) yields an asymptotic bias term which has a familiar form (from more conventional kernel estimation in discrete time) but, in light of the mildness of our assumptions, depends on the process’ invariant (speed) measure rather on the process’ time-invariant stationary density. The drift’s optimal bandwidth rate is \( \left( \frac{1}{v(T)} \right)^{1/5} \). The corresponding diffusion’s value is \( \left( \frac{\Delta_{n,T}}{v(T)} \right)^{1/5} \).

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). 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 $\xi^\sigma \sim \exp(\mu)$. 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 $\lambda_{\sigma^2}(\sigma_t^2) = 0^2$). Provided the variance drift, diffusion, and intensity satisfy the conditions laid out in Section 3, we leave their functional forms unspecifed. Notice that

\begin{align}
\hat{\theta}_1(x) &= m_{\sigma^2}(x) + \mu \lambda_{\sigma^2}(x) \\
\hat{\theta}_2(x) &= \Lambda_{\sigma^2}^2(x) + 2\mu^2 \lambda_{\sigma^2}(x) \\
\hat{\theta}_3(x) &= 6\mu^2 \lambda_{\sigma^2}(x) \\
\hat{\theta}_4(x) &= 24\mu^4 \lambda_{\sigma^2}(x) \\
\vdots
\end{align}

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

\begin{align}
\hat{\mu} &= \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\theta}_4(\hat{\sigma}_{iT/n})}{4\hat{\theta}^3(\hat{\sigma}_{iT/n})}, \\
\hat{\lambda}_{\sigma^2}(x) &= \frac{\hat{\theta}_4(x)}{24\hat{\mu}^4}, \\
\hat{\Lambda}_{\sigma^2}^2(x) &= \frac{\hat{\theta}_2(x) - 2\hat{\lambda}_{\sigma^2}(x)\hat{\mu}^2}{\hat{\theta}(x) - \hat{\lambda}_{\sigma^2}(x)\hat{\mu}}, \\
\hat{m}_{\sigma^2}(x) &= \frac{\hat{\theta}(x) - \hat{\lambda}_{\sigma^2}(x)\hat{\mu}}.
\end{align}
Alternative (possibly superior) identification methods can of course be employed. Here we lay out the main ideas by considering the simplest, and most intuitive, identification approach.

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

\begin{align*}
\theta^1(x) &= m_{\log \sigma^2}(x), \\
\theta^2(x) &= \Lambda^2_{\log \sigma^2}(x) + \sigma_\xi^2 \lambda_{\log \sigma^2}(x), \\
\theta^4(x) &= 3\sigma_\xi^2 \lambda_{\log \sigma^2}(x), \\
\theta^6(x) &= 15\sigma_\xi^6 \lambda_{\log \sigma^2}(x), \\
\vdots
\end{align*}

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

\begin{align*}
\tilde{\sigma}_\xi^2 &= \frac{1}{n} \sum_{i=1}^{n} \frac{\tilde{\theta}^6(\tilde{\sigma}_{iT/n}^2)}{5 \tilde{\theta}^4(\tilde{\sigma}_{iT/n}^2)}, \\
\tilde{\lambda}_{\log \sigma^2}(x) &= \frac{\tilde{\theta}^4(x)}{3 \tilde{\sigma}_\xi^4}, \\
\tilde{\Lambda}^2_{\log \sigma^2}(x) &= \tilde{\theta}^2(x) - \tilde{\lambda}_{\log \sigma^2}(x) \tilde{\sigma}_\xi^2, \\
\tilde{m}_{\log \sigma^2}(x) &= \tilde{\theta}^1(x).
\end{align*}

This identification procedure has proved successful in the analysis of the temporal dynamics of spot interest rate series in continuous time (Johannes, 2004).

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 $\tilde{\theta}^j(x) \xrightarrow{p} \theta^j(x)$ for all $j$, and by Slutsky’s theorem, consistency (in probability) of the relevant functions and jump size moments.
Theorem 4. (The infinitesimal moments.) If \( k, n, T \to \infty \) and \( h_{n,T}, \phi_{n,T} \to 0 \) so that

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

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

\[
\hat{\theta}^j(x) \xrightarrow{p} \theta^j(x) \quad j \geq 1,
\]

where \( \Sigma_0^{\alpha^2}(T, x) \propto v(T) \). If

\[
\lim_{n,T \to \infty} h_{n,T} v(T) = \infty, \\
\lim_{n,T \to \infty} h_{n,T}^5 v(T) = C_3, \\
\lim_{n,T \to \infty} \frac{v(T)}{h_{n,T}} \left( \Delta_{n,T} \log \frac{1}{\Delta_{n,T}} \right)^{1/2} = 0, \\
\lim_{k,n,T \to \infty} \frac{T^{3/2} v(T)^{-1}}{\Delta_{n,T} \Delta_{n,T}^{k} n^{3/2} \phi_{n,T}^{\beta}} + \frac{T^{3/2} v(T)^{-1}}{\Delta_{n,T} \Delta_{n,T}^{k} n^{3/2} \phi_{n,T}^{\beta}} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} = 0,
\]

then,

\[ \sqrt{h_{n,T} \Sigma_0^{\alpha^2}(T, x)} \left\{ \hat{\theta}^j(x) - \theta^j(x) - \Gamma_{\theta^j}(x) \right\} \Rightarrow N (0, K_2 \theta^{2j}(x)) , \quad \forall j \geq 1 \]

with

\[
\Gamma_{\theta^j}(x) = h_{n,T}^2 K_1 \left[ \theta^{\prime\prime}(x) \frac{s^\prime(x)}{s(x)} + \frac{1}{2} \theta^{\prime\prime\prime}(x) \right],
\]

where \( s(dx) \) is the process’ invariant measure and \( C_3 \) is a constant.

**Proof.** See Appendix B.

**Remark 9.** Contrary to the no jump case, all infinitesimal moments converge at the same rate. In particular, an enlarging span of data \( (T \to \infty) \) is necessary to guarantee \( h_{n,T} \Sigma_0^{\alpha^2}(T, x) \xrightarrow{a.s.} \infty \) and, hence, consistency of all moments. As earlier, selection of the optimal bandwidth rate
\[
\left(\frac{1}{n(T)}\right)^{1/5}
\]
yields an asymptotic bias term which depends on the process’ invariant measure and may be eliminated by slight undersmoothing.

We now discuss, for both models presented earlier, 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 asymptotic bias terms.

We initially assume that the moments of the jump sizes are estimated by averaging higher-order infinitesimal moments over a fixed time period \(T\). In other words, \(\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \frac{\theta_{i}(T/\bar{n})}{4\theta_{i}(\sigma_{T,\bar{n}})}\) and \(\hat{\sigma}_{\xi}^{2} = \frac{1}{n} \sum_{i=1}^{n} \frac{\theta_{i}^{2}(\sigma_{T,\bar{n}})}{5\theta_{i}(\sigma_{T,\bar{n}})}\) with \(\bar{T}/\bar{n} \to 0\) as \(\bar{n} \to \infty\) over a fixed \(\bar{T} < T\). Importantly, for consistency (see Remark 9), the higher-order moments \(\hat{\theta}^{j}\) in \(\hat{\mu}\) and \(\hat{\sigma}_{\xi}^{2}\) continue to be estimated over an asymptotically expanding \(T\). As we will show, the fixed \(T\) case is theoretically interesting when dealing with generic (stationary and nonstationary) recurrent processes. We will relax it (and let \(T\) diverge with \(T\)) when focusing on ergodic (or strictly stationary) systems (Remark 10).

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

Assume

\[
\lim_{n,T \to \infty} h_{n,T} v(T) = \infty,
\]

\[
\lim_{n,T \to \infty} h_{n,T}^{5} v(T) = 0,
\]

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

\[
\lim_{k,n,T \to \infty} \frac{T^{3/2} v(T)^{-1}}{\Delta_{n,T}^{1/2} k^{1/2} \phi_{n,T}^{\beta}} + \frac{T^{3/2} v(T)^{-1}}{\Delta_{n,T}^{1/2} \phi_{n,T}^{\beta}} \left(\frac{1}{\Delta_{n,T}}\right)^{1/2} = 0,
\]

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

**Generalized Duffle et al.’s model**

**Expected jump size:**

\[
(\bar{Y}(T))^{-1/2} T \{\hat{\mu} - \mu\} \Rightarrow \mathcal{N}(0, 1),
\]

where
\[ Y(T) = \int_{-\infty}^{\infty} \lambda_{\sigma^2}(x) \left( \frac{\mathbb{T}_{\sigma^2}(T, x)}{L_{\sigma^2}(T, x)} \right) \mathbb{E} \left( \left( \frac{1}{4\theta^3(x)} (\xi^\sigma)_4 - \frac{\theta^4(x)}{4(\theta^3(x))^2} (\xi^\sigma)_3 \right)^2 \right) dx. \]

**Jump intensity:**
\[ \sqrt{h_{n,T} \frac{\hat{T}_{\sigma^2}(T, x)}{L_{\sigma^2}(T, x)}} \left\{ \hat{\lambda}_{\sigma^2}(x) - \lambda_{\sigma^2}(x) \right\} \Rightarrow \mathbb{N} \left( 0, K_2 \frac{\lambda_{\sigma^2}(x) \mathbb{E} \left( (\xi^\sigma)_8 \right)}{(24)^2 \mu^8} \right). \]

**Diffusive function:**
\[ \sqrt{h_{n,T} \frac{\hat{T}_{\sigma^2}(T, x)}{L_{\sigma^2}(T, x)}} \left\{ \hat{\Lambda}_{\sigma^2}(x) - \Lambda_{\sigma^2}(x) \right\} \Rightarrow \mathbb{N} \left( 0, K_2 \lambda_{\sigma^2}(x) \mathbb{E} \left( (\xi^\sigma)_2 - \frac{1}{12\mu^2} (\xi^\sigma)_4 \right)^2 \right) \).

**Drift function:**
\[ \sqrt{h_{n,T} \frac{\hat{T}_{\sigma^2}(T, x)}{L_{\sigma^2}(T, x)}} \left\{ \hat{m}_{\sigma^2}(x) - m_{\sigma^2}(x) \right\} \Rightarrow \mathbb{N} \left( 0, K_2 \left( \Lambda_{\sigma^2}(x) + \lambda_{\sigma^2}(x) \mathbb{E} \left( \xi^\sigma - \frac{1}{24\mu^3} (\xi^\sigma)_4 \right)^2 \right) \right) \).

**Log-variance model**

**Jump standard deviation:**
\[ (Y(T))^{-1/2} 2T \sigma_\xi \left\{ \hat{\sigma}_\xi - \sigma_\xi \right\} \Rightarrow \mathbb{N}(0, 1), \]

where
\[ Y(T) = \int_{-\infty}^{\infty} \lambda_{\log \sigma^2}(x) \left( \frac{\mathbb{T}_{\sigma^2}(T, x)}{L_{\sigma^2}(T, x)} \right) \mathbb{E} \left( \left( \frac{1}{5\theta^4(x)} (\xi^\sigma)_6 - \frac{\theta^6(x)}{5(\theta^4(x))^2} (\xi^\sigma)_4 \right)^2 \right) dx. \]

**Jump intensity:**
\[ \sqrt{h_{n,T} \frac{\hat{T}_{\sigma^2}(T, x)}{L_{\sigma^2}(T, x)}} \left\{ \hat{\lambda}_{\log \sigma^2}(x) - \lambda_{\log \sigma^2}(x) \right\} \Rightarrow \mathbb{N} \left( 0, K_2 \frac{\lambda_{\log \sigma^2}(x) \mathbb{E} \left( (\xi^\sigma)_8 \right)}{(3)^2 \sigma_\xi^3} \right). \]

**Diffusive function:**

19
\[ \sqrt{n,T} \tilde{L}_{\sigma^2}(T, x) \left\{ \hat{\Lambda}^2_{\log \sigma^2}(x) - \Lambda^2_{\log \sigma^2}(x) \right\} \Rightarrow \mathcal{N} \left( 0, K_2 \lambda_{\log \sigma^2}(x) E \left[ \left( \left( \xi^\sigma \right)^2 - \frac{1}{3 \sigma^2_\xi} \left( \xi^\sigma \right)^4 \right)^2 \right] \right). \]

**Drift function:**

\[ \sqrt{n,T} \tilde{L}_{\sigma^2}(T, x) \left\{ \hat{m}_{\log \sigma^2}(x) - m_{\log \sigma^2}(x) \right\} \Rightarrow \mathcal{N} \left( 0, K_2 \left( \Lambda^2_{\log \sigma^2}(x) + \lambda_{\log \sigma^2}(x) E \left( \left( \xi^\sigma \right)^2 \right) \right) \right). \]

**Proof.** See Appendix B.

**Remark 10 (The ergodic case.)** In the positive recurrent and strictly stationary case, \( v(T) = T \) and \( \frac{\tilde{L}_{\sigma^2}(T, x)}{T} \xrightarrow{p} p(x) \), where \( p(x) \) is the stationary density of the spot volatility 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 \( \sqrt{n,T} \tilde{T} \). The later depends on the volatility process’ time-invariant probability distribution, i.e., \( p(x) \).

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

\[ \sqrt{T} \left\{ \mu - \mu \right\} \Rightarrow \mathcal{N} \left( 0, \int_{-\infty}^{\infty} \lambda_{\sigma^2}(x) E \left( \frac{1}{4 \theta^4(x)} \left( \xi^\sigma \right)^4 - \frac{\theta^4(x)}{4 \left( \theta^3(x) \right)^2} \left( \xi^\sigma \right)^3 \right)^2 p(x) dx \right) \]

and

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

**Remark 11 (Covariance estimation.)** Statistical inference is straightforward given estimates of the relevant asymptotic variances. To this extent, in the case of the generalized Duffie et al.’s model, we notice that:
Furthermore,\[\frac{h_{n,T}}{h_{n,T}} \left(\frac{\Delta_{n,T}}{\Delta_{n,T}}\right)^2 \sum_{i=1}^{n} \left(\frac{\hat{\theta}_i}{16 \left(\theta_i^3\right)^2} - \frac{2\hat{\theta}_i^4}{16 \left(\theta_i^3\right)^3} + \frac{(\hat{\theta}_i^4)^2}{16 \left(\theta_i^3\right)^4}\right) \frac{\sum_{j=1}^{n} K \left(\frac{\partial_j \theta / \partial x - \partial T / \partial x}{h_{n,T}}\right)}{\sum_{j=1}^{n} K \left(\frac{\partial_j \theta / \partial x - \partial T / \partial x}{h_{n,T}}\right)}\]

\[\sim \int_{-\infty}^{\infty} \lambda_{\sigma^2}(x) \left(\frac{T_{\sigma^2}(T, x)}{T_{\sigma^2}(T, x)}\right) E \left(\left(\frac{1}{4\theta^3(x)} (\xi^\sigma)^4 - \frac{\theta^4(x)}{4 (\theta^3(x))^2} (\xi^\sigma)^3\right)^2\right) dx,\]

\[\frac{\hat{\theta}_i^8(x)}{(24)^2 \mu^8} \sim \frac{\lambda_{\sigma^2}(x) E \left((\xi^\sigma)^8\right)}{(24)^2 \mu^8},\]

and

\[\hat{\theta}_i^2(x) - \frac{2\hat{\theta}_i^5(x)}{24\mu^3} + \frac{\hat{\theta}_i^8(x)}{(24)^2 \mu^4} \sim \lambda_{\sigma^2}(x) + \lambda_{\sigma^2}(x) E \left(\left(\xi^\sigma - \frac{1}{24\mu^4} (\xi^\sigma)^4\right)^2\right).\]

Furthermore,

\[\frac{h_{n,T}}{h_{n,T}} \left(\frac{\Delta_{n,T}}{\Delta_{n,T}}\right)^2 \sum_{i=1}^{n} \left(\frac{\hat{\theta}_i^{12}}{25 (\hat{\theta}_i^4)^2} - \frac{2\hat{\theta}_i^{6-10}}{25 (\hat{\theta}_i^4)^3} + \frac{(\hat{\theta}_i^4)^2}{25 (\hat{\theta}_i^4)^4}\right) \frac{\sum_{j=1}^{n} K \left(\frac{\partial_j \theta / \partial x - \partial T / \partial x}{h_{n,T}}\right)}{\sum_{j=1}^{n} K \left(\frac{\partial_j \theta / \partial x - \partial T / \partial x}{h_{n,T}}\right)}\]

\[\sim \int_{-\infty}^{\infty} \lambda_{\log \sigma^2}(x) \left(\frac{T_{\log \sigma^2}(T, x)}{T_{\log \sigma^2}(T, x)}\right) E \left(\left(\frac{1}{5\theta^4(x)} (\xi^\sigma)^6 - \frac{\theta^6(x)}{5 (\theta^4(x))^2} (\xi^\sigma)^4\right)^2\right) dx,\]

\[\frac{\hat{\theta}_i^8(x)}{(3)^2 \hat{\sigma}_\xi^8} \sim \frac{\lambda_{\log \sigma^2}(x) E \left((\xi^\sigma)^8\right)}{(3)^2 \hat{\sigma}_\xi^8},\]

and, of course,

\[\hat{\theta}_i^2(x) \sim \lambda_{\log \sigma^2}(x) + \lambda_{\log \sigma^2}(x) E \left((\xi^\sigma)^2\right)\]

in the log-variance case.
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, plus the sum of the squared jumps. We therefore need to consider estimators which solely identify integrated variance. The realized bypower variation measure of Barndorff-Nielsen and Shephard (2004, 2005), which we formally define in Section 8 below, achieves, among other procedures, this goal.

**Remark 12 (Bypower variation in the no noise case)** Assumption 2 holds with $\alpha = \frac{1}{2}$, $\beta = 0$, $a \approx 2.6$, $b = 0$, and $\eta = 1$.

Even though our theoretical results go through unchanged when using this alternative estimator while abstracting from market microstructure noise, care should be exercised in practice to account for the presence of noise in the price process. Below we consider a modification of the classical bypower variation estimator which, albeit technically inconsistent in the presence of noise, has been shown to be fairly robust to it and perform satisfactorily in empirical work.

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 bypower variation or alternative identification methods robusts to jumps in returns) as well as infinitesimal moment estimates for the return process (e.g., $\tilde{\theta}_j^j(\sigma^2)$ with $j = 1, 2, \ldots$), the relevant functions and the features of the return jump distribution can be identified by using a scheme similar to those in Section 5. Specifically, we will assume Gaussian mean zero jumps, i.e., $dJ_t^r = \varphi dN_t^r$ with $\varphi \sim \mathcal{N}(0, \sigma_\varphi^2)$, and employ

$$
\tilde{\sigma}_\varphi^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{\tilde{\theta}_6^6(\sigma_{iT/n}^2)}{5\tilde{\theta}_4^4(\sigma_{iT/n}^2)}, \quad (24)
$$

$$
\tilde{\lambda}_r(\sigma^2) = \frac{\tilde{\theta}_4^4(\sigma^2)}{3\tilde{\theta}_4^4}, \quad (25)
$$

$$
\tilde{\mu}(\sigma^2) = \tilde{\theta}_1^1(\sigma^2). \quad (26)
$$

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

22
The leverage function can be identified as follows:

\[
\hat{\rho}(\sigma^2) = \left( \frac{1}{\Delta_{n,T}} \sum_{i=1}^{n-1} K \left( \frac{\bar{\sigma}^2_{iT/n} - \sigma^2}{h_{n,T}} \right) \left( \log(p_{i+1}/T/n) - \log(p_i/T/n) \right) \left( \bar{\sigma}^2_{iT/n} - \bar{\sigma}^2_{iT/n} \right) \right)
\]

\[
+ \frac{1}{\Delta_{n,T}} \sum_{i=1}^{n} K \left( \frac{\bar{\sigma}^2_{iT/n} - \sigma^2}{h_{n,T}} \right) \sqrt{\sigma^2 \tilde{\Lambda}^2_{f(.)}(\sigma^2)}
\]

where \(\tilde{\Lambda}^2_{f(.)}(\sigma^2)\) may be estimated by virtue of Eq. (14) or Eq. (22) (depending on the assumed variance model). Our empirical work will use Eq. (14). In light of the independence of the jumps in returns and volatility and the independence between jumps and Brownian shocks, \(\hat{\rho}(\sigma^2)\) is expected to identify \(\rho(\sigma^2)\) consistently as we discuss in Theorem 7 below.

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

\[
\lim_{n,T \to \infty} h_{n,T} v(T) = \infty,
\]

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

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

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

\[
\hat{\mu}(\sigma^2) \overset{p}{\to} \mu(\sigma^2),
\]

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

\[
\lim_{n,T \to \infty} h_{n,T} v(T) = \infty,
\]

\[
\lim_{n,T \to \infty} h_{n,T}^5 v(T) = C_4,
\]

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

\[
\lim_{k,n,T \to \infty} \frac{T^{3/2} v(T)^{-1}}{\Delta_{n,T} h_{n,T} k^{\alpha} \phi_{n,T}^{1/2}} + \frac{T^{3/2} v(T)^{-1}}{\Delta_{n,T} h_{n,T} \phi_{n,T}^{1/2}} \left( \phi_{n,T} \log \frac{1}{\phi_{n,T}} \right)^{1/2} = 0,
\]
then,

$$\sqrt{h_{n,T} L_{a^2}(T, \sigma^2) \{ \tilde{\mu}(\sigma^2) - \mu(\sigma^2) - \Gamma_{\tilde{\mu}(\sigma^2)}(\sigma^2) \}} \Rightarrow N(0, K_2 \theta^2(\sigma^2)),$$

with

$$\Gamma_{\tilde{\mu}(\sigma^2)}(\sigma^2) = h_{n,T}^2 K_1 \left[ \mu'(\sigma^2) \frac{s'(\sigma^2)}{s(\sigma^2)} + \frac{1}{2} \mu''(\sigma^2) \right],$$

where

$$\theta^2(\sigma^2) = \sigma^2 + \lambda_r(\sigma^2) E(\varphi^2),$$

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

We now turn to leverage.

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

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

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

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

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

$$\tilde{\rho}(\sigma^2) \overset{p}{\to} \rho(\sigma^2),$$

where $L_{a^2}(T, \sigma^2) \propto v(T)$.

**Weak convergence:** to be added.

## 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 bypower 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 \([i\Delta n, T, i\Delta n + \phi n, T]\) with \( i = 1, ..., n \). The intra-daily price formation mechanism is defined as:

\[
\log(p_j^*) = \log(p_j) + \eta_j \quad j = 1, ..., k
\]

or, in terms of continuously-compounded returns,

\[
\log(p_j^*) - \log(p_{j-1}) = \log(p_j) - \log(p_{j-1}) + \eta_j - \eta_{j-1},
\]

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 \( \mathbb{E}(\eta^2) = \sigma^2 \) (or, equivalently, \( \mathbb{E}(\varepsilon^2) = 2\sigma^2 \)). 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 \( i = 1, ..., q \). The first sub-grid starts at \( t_0 \) and takes every \( q \)-th arrival time, i.e., \( \Psi^{(1)} = (t_0, t_0+q, t_0+2q, ..., ) \), the second sub-grid starts at \( t_1 \) and also takes every \( q \)-th arrival time, i.e., \( \Psi^{(2)} = (t_1, t_1+q, t_1+2q, ..., ) \), and so on. Given the generic \( i \)-th sub-grid of arrival times, define
\[ \hat{V}^{(i)} = \sum_{t_j, t_{j+} \in \Psi^{(i)}} (\log(p_{t_j+})^* - \log(p_{t_j})^*)^2, \]

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

\[ \hat{V}^{ZMA} = \frac{\sum_{i=1}^{q} \hat{V}^{(i)}}{q} - \bar{k} \hat{E}(\varepsilon^2), \]

where \( \bar{k} = \frac{k - q + 1}{q} \) and \( \hat{E}(\varepsilon^2) = \sum_{i=1}^{k} (\log(p_{t_i+})^* - \log(p_{t_i})^*)^2 \).

Under the previous price formation mechanism, but in the absence of jumps in the return process, the estimator is consistent for \( \int \sigma_d^2 ds \) at speed \( k^{1/6} \) provided \( q, k \to \infty \) with \( \frac{q}{k} \to 0 \) and \( \frac{k^2}{q} \to \infty \) (Zhang et al., 2005).

The estimator is biased in finite samples. We optimize its finite sample properties by (1) bias-correcting it and (2) optimizing its finite sample variance as suggested by Bandi and Russell (2006) and Bandi et al. (2008). Define \( \hat{V}^{ZMAadj} = c(q, k) \hat{V}^{ZMA} \), where

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

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

\[ \text{var} \left( \hat{V}^{ZMAadj} \right) = (c(q, k))^2 \text{var} \left( \hat{V}^{ZMA} \right), \]

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

\[
\begin{align*}
\text{var} \left( \hat{V}^{ZMA} \right) &= \Theta^{ZMA} - \frac{1}{3} (Q + V^2) \left( \frac{q}{k} \right)^2 + \left( -\frac{1}{3} V^2 - \frac{1}{k^2} - 4V^2 - \frac{1}{k^2} + \frac{4}{3} Q \right) \frac{q}{k} \\
&+ \left( -\frac{4}{k^4} (Q + V^2) + \left( \frac{8\sigma_4^4 + 16\sigma_2^2 V - 8Q - \frac{56}{3} V^2}{k^3} \right) + \left( \frac{24\sigma_2^2 V - \frac{10}{3} Q + 8\sigma_4^4}{k^2} \right) \right) \frac{k}{q} + \left( \frac{2}{k^5} Q + \left( -\frac{4\sigma_4^4 - 8\sigma_2^2 V + 4Q - 8V^2}{k^4} \right) + \left( -\frac{4\sigma_4^4 - 16\sigma_2^2 V + 2Q}{k^3} \right) \right) \\
&+ \left( \frac{8\sigma_4^4 - 8\sigma_2^2 V}{k^2} \right) + \frac{8}{k^4} \frac{k^2}{q^2},
\end{align*}
\]

and

\[ \Theta^{ZMA} = (-4\sigma_4^4 - 8V\sigma_2^2)^{1/k} + \left( -4\sigma_4^4 - 8\sigma_2^2 V + \frac{13}{3} Q + \frac{79}{3} V^2 \right)^{1/k^2} + \frac{1}{k^3} (2Q + 8V^2) \]
with $\sigma^2_\eta = \mathbf{E}(\eta^2)$, $V = \int \sigma^2_s ds$ and $Q = \int \sigma^4_s ds$ (Bandi and Russell, 2005). Given preliminary estimates of $V$, the quarticity term $Q$, and $\mathbf{E}(\eta^2)$, the optimal $q$ is defined as the minimizer of $\text{var} \left( \widehat{V}_{ZMAadj} \right)$ (Bandi et al., 2008). Thus,

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

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

**Kernel estimators.** Write

$$\widehat{V}_{BNHLS} = \widehat{\gamma}_0 + \sum_{s=1}^{q} w_s (\widehat{\gamma}_s + \widehat{\gamma}_{-s}),$$

where $\widehat{\gamma}_s = \sum_{j=1}^{k} r_j^s r_{j-s}$ with $s = -q, \ldots, q$, $w_s = g \left( \frac{s-1}{q} \right)$, and $g(.)$ is a kernel function on $[0, 1]$ satisfying $g(0) = 1$ and $g(1) = 0$.

Under the previous price formation mechanism, but in the absence of jumps in return process, this family of estimators is consistent at rate $k^{1/6}$ if $q = ck^{2/3}$. Provided $g(.)$ is also chosen in such a way to guarantee that $g'(0) = 0$ and $g'(1) = 0$, then the number of autocovariances can be selected as $q = ck^{1/2}$ and the corresponding estimator is consistent at rate $k^{1/4}$ (Barndorff-Nielsen et al., 2006). 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 $\int \sigma^2_s ds$ in finite samples. Their finite sample variance, however, can be optimized (Bandi and Russell, 2005). Write $\varsigma = \frac{q}{k}$. Hence,

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

$$+ 4 \left( \mathbf{E}(\eta^2) \right)^2 (w^T \Omega_3 w) + (2\mathbf{E}(\eta^2)V)4(w^T \Omega_4 w),$$

with

$$w = \left( 1, 1, g \left( \frac{1}{\varsigma k} \right), \ldots, g \left( \frac{\varsigma k - 1}{\varsigma k} \right) \right)^T,$$

and $\Omega_a a = 1, \ldots, 4$ are $(\varsigma k + 1, \varsigma k + 1)$ square matrices. For $j \leq \varsigma k$, the matrices $\Omega_1$ and $\Omega_4$ are defined as follows:
\[
\begin{align*}
\Omega_1[1, 1] &= 2, \quad \Omega_1[1 + j, 1 + j] = 4, \\
\Omega_4[1, 1] &= 1, \quad \Omega_4[2, 1] = -1, \quad \Omega_4[1, 2] = -1, \quad \Omega_4[2, 2] = 2, \quad \Omega_4[1 + j, 1 + j] = 2, \\
\Omega_4[1 + j, j] &= -1, \quad \Omega_4[j, j + 1] = -1.
\end{align*}
\]

For \( j \leq \zeta M - 1 \), the matrices \( \Omega_2 \) and \( \Omega_3 \) are defined as follows:

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

Thus, using preliminary estimates of \( \sigma^2_n, V, \) and \( Q \),

\[
\frac{\hat{\sigma}^2_{iT/n}}{\phi_{n,T}} = \frac{\hat{V}^{BNHLS}_{iT/n}(\zeta^*)}{\hat{V}^{BNHLS}_{iT/n}(\zeta^*)},
\]

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

**Bypower variation.** Write

\[
\hat{V}^{BNS}_{iT/n} = \mu^{-2} \sum_{j=2}^{k} |r_j||r_{j-1}|
\]

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 \( \hat{V}^{BNS}_{iT/n} \) is consistent for \( \int \sigma^2_s ds \) 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

\[
\hat{V}^{BNS(stag.)}_{iT/n} = \mu^{-2} \left( \frac{k - 2}{k} \right)^{-1} \sum_{j=3}^{k} |r_j||r_{j-2}|
\]
This correction does not yield consistency in the presence of noise but has been shown to perform well in practice. When estimating models with discontinuities in the return process, we employ

\[ \delta^2 \sigma_{iT/n}^4 = \frac{\hat{\gamma}^{BNS(\text{stag.})}_{iT/n}}{\phi_{n,T}}. \]

### 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.

The noise-return second moment \( E(\varepsilon^2) \), \( V \), 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 \( E(\varepsilon^2) \) 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 follow common practice in the literature and express the variance estimates in daily terms. Since the original estimates are for an intra-daily 6-hour period, we multiply them by a constant factor \( \delta \) defined as \( \delta = \frac{\sum_{i=1}^{n} \left( r_{S&P500}^{iT/n} \right)^2 / \sum_{i=1}^{n} \hat{V}_{iT/n} }{ \sum_{i=1}^{n} r_{S&P500}^{iT/n} } \), where \( r_{S&P500}^{iT/n} \) is the return on the S&P500 index over day \( i \). This procedure ensures that the average of the transformed variances, i.e., \( \delta \hat{V} \), 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 $\delta$.\textsuperscript{3}

Table 2 provides descriptive statistics about $r_{iT/n}^{SK\&P500}$ and $\sigma_{iT/n}^2$ 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 bypower variation estimator. Because in the presence of jumps in the return process 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 bypower 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. As said, we allow for a nonlinear drift, diffusion, and jump intensity. We choose a straightforward identification scheme as laid out in Eq. (12) through Eq. (15).\textsuperscript{4} Only for the time being, we assume absence of jumps in the return process. Fig. 2 and Fig. 3 present the results pertaining to 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 estimators are similar. In light of its high persistence, the variance process is only mildly mean-reverting. A zero drift specification for variance can hardly be rejected. The diffusive function is nonlinear. Below, we will show that it conforms more naturally with a CEV specification than with a linear specification, as introduced by Heston (1993) and adopted by many others. The point estimates suggests about

\textsuperscript{3}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.

\textsuperscript{4}The infinitesimal moments’ bandwidths are set equal to $c_j \times \text{stdc} (\sigma^2) \times n^{-1/5}$, 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 bandwidths are larger than the second moment bandwidth). Our selection procedure is extremely reasonable but admittedly simple. Automated bandwidth selection in the context of continuous-time models is a largely underexplored research area.
4 volatility jumps per year. The estimated expected size of the jumps is about 5.

We now turn to the empirically more compelling case of jumps in the return process and use staggered bypower variation to identify integrated variance. The corresponding results are in Fig. 4. As expected based on the temporal dynamics of the relevant series (Fig. 1) and the economic interpretation of bypower 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). Despite small differences in the point estimates, their estimated linear drift, constant jump intensity, and average jump size are statistically supported by our data. While the average jump sizes are very close, our point estimates suggest a slightly larger number of jumps per year (about 4 versus 1.5). We will come back to this issue in Subsection 9.2 below. Importantly, our diffusion estimates differ from those in Eraker et al. (2003). We find more volatility associated with the process’ continuous component. We also find that the variance’s diffusion function is better represented by a CEV specification than by a linear one (i.e., \( \Lambda^2(x) \propto x^{3/2} \)). Using models without jumps in variance, Chacko and Viceira (2001) and Jones (2002) also show nonlinear structures in the variance of variance. In Chacko and Viceira (2001) the need for nonlinearities in variance diminishes with the addition of jumps in returns.

To further illustrate the need for an alternative variance structure, Fig. 5(a) represents the ratio \( \frac{(\Delta \sigma^2_t - m(\sigma^2_t)\Delta t)}{\Lambda(\sigma^2_t)\sqrt{\Delta t}} \) with a functional specification for \( m(\sigma^2_t) \) and a linear parametric specification for \( \Lambda(\sigma^2_t) \) as in Eraker et al. (2003). Fig. 5(b) displays the same ratio with a functional drift and a parametric CEV diffusion consistent with our data \( (\Lambda^2(x) = 0.1x^{3/2}) \).\(^5\) The ratio’s dynamics appear to be better behaved in the latter case with occasional (positive) jumps as implied by the assumed exponential jump distribution. Importantly, descriptive statistic from the data nicely conform with descriptive statistics from simulated data\(^6\) from a CEV stochastic variance model as described in Section 10 below.

It is now of interest to verify 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.

---

\(^5\)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.

\(^6\)Not reported, but available on request.
9.1 The joint volatility/return dynamics

We estimate a nonlinear model for returns with Gaussian jumps (Fig. 6). The identification scheme is therefore consistent with Eq. (20) and Eq. (23) applied to returns rather than to log-variances as discussed in Section 7. 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).

The return dynamics suggest the presence of a statistically-insignificant nonlinear risk-return trade-off. The insignificance of the trade-off is of course not surprising and fully consistent with much empirical work on the evaluation of the relation between conditional mean returns and conditional variance at low (daily, here) frequencies (see, e.g., Bandi and Perron, 2008, for references). Importantly, in our case the use of high-frequency data does not yield a stronger dependence between conditional mean and conditional variance as suggested in some recent work (see, e.g., Bali and Peng, 2006). Similarly, Eraker et al. (2003) stress that experimentation with a linear risk-return model did not deliver significant estimates using their filtering methods and, therefore, estimate a specification with constant mean (whose numerical value is reported in Fig. 6(a)).

We find hump-shaped (in the variance level) leverage effects around –0.5. We also find smaller jump sizes (implying about 95% jumps between 3.4% and -3.4%) and more jumps (about 65 per year at the mean variance level) than in Eraker et al. (2003). While these results are somewhat unusual in the context of stochastic volatility models (in particular, the reported number of jumps is usually lower and around 1 or 2 per year), it is somewhat consistent with recent jump testing procedures making use of high-frequency price data (see, e.g., Andersen et al., 2003).

This said, given Eq. (20) through Eq. (21), downward biased jump size estimates may yield upward biased jump intensities. The question we now ask is then: can a simple identification procedure, such as the one we purposely adopted, combined with reasonable bandwidth choices, deliver our estimates if the true process is, in fact, the affine process of Eraker et al. (2003)? In Section 10 we address this question by simulation. The goal of the simulations is therefore twofold. On the one hand, immediately, we wish to verify the finite sample accuracy of our functional estimates, bandwidth choices, and identification method. On the other hand, we wish to show that, provided meaningful identification schemes and bandwidths are adopted, even potentially biased (in finite samples) estimates can be informative about the set of likely

\footnote{The bandwidths are set as described in Footnote 4 above.}
underlying data-generating processes. Generally speaking, the set of data generating processes which delivers finite sample "biases" which are compatible with the "biases" observed in the data should be considered viable. While this intuition can be made rigorous in the context of simulation-based estimation procedures selecting parametric models on the basis of the similarity between their induced estimates and the estimates obtained from data (much in the same spirit as indirect inference, Gourieroux et al., 1993), in this paper we apply it to our problem in a less formal fashion. 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 jumps 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:

\[ r_{t,t+\Delta t} = b\Delta t + \sqrt{\sigma_t^2 \Delta t} \varepsilon_t^r + \xi_t^r J^r_t, \]
\[ \sigma_{t+\Delta t}^2 - \sigma_t^2 = \kappa(\theta - \sigma_t^2) \Delta t + \sigma_i \sqrt{\sigma_t^2 \Delta t} \varepsilon_t^i + \xi_t^i J^i_t, \]

where \( \{J^r_t, J^i_t\} \) are Bernoulli random variables with constant intensities \( \lambda^r \Delta t \) and \( \lambda^i \Delta t \), \( \{\varepsilon_t^r, \varepsilon_t^i\} \) are standard Gaussian random variables with correlation \( \rho \), \( \xi_t^r \) is a mean zero Gaussian random variable with standard deviation \( \sigma_r \), \( \xi_t^i \) is an exponential random variable with mean \( \mu_i \), and \( \Delta t \) is a time-discretization (one day). The parameters are those in Table III, Column 5, of Eraker et al. (2003). We simulate the model under the null of no jumps, i.e., \( J^r_t = 0 \) and \( J^i_t = 0 \). 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 band of the estimated conditional kurtosis 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. (27) and Eq. (28)), it is immediate to show that the alternative is consistent with our empirical statistics (Fig. 8). Importantly, this
coherence increases when replacing $\sigma_v \sqrt{\sigma^2}$ with a nonlinear (diffusive) 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 them.

10 Simulations

The simulations rely again on the bi-variate system in Eq. (27) and Eq. (28)). We again generate 2,053 observations for every sample path and 1,000 paths. However, in agreement with data, the diffusive variance-of-variance is specified as being equal to $\sigma_v \sqrt{(\sigma^2_t)^{1.5} \Delta_t}$. As earlier, the parameters are those in Table III, Column 5, of Eraker et al. (2003) with the exception of $\sigma_v = 0.31$. The specification $\Lambda^2(\sigma^2_t) = 0.1(\sigma^2_t)^{1.5}$ provides superior fit for our data, as illustrated previously.

Figs. 8 and 9 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 results. Interestingly, the expected jump size is only slightly downward biased, while the intensity of the jumps is somewhat upward biased and tends to increase when moving away from the bulk of the data. We confirm that the variance jump parameters in Eraker et al. (2003) could yield our jump estimates and are consistent with them. As in the simulations, our jump intensity estimates are slightly larger than those in Eraker et al. (about 4 versus 1.5 annual jumps) and increase when moving to regions with sparser data. Again, in agreement with the simulations, our estimated expected jump size is very close to that reported by Eraker et al. (2003).

We now turn to the return and joint dynamics. The return drift and the leverage parameter are fairly accurately estimated. However, the standard deviation of the return Gaussian jumps is fairly obviously downward biased whereas the return jump intensities are upward biased with, again, an increasing nonlinear trend when moving to values away from the center of the simulated data. Similar patterns are observed in the data. The estimated jump standard deviation is lower than in Eraker et al. (2003) (about 1.7 rather than about 3). Equivalently, the estimated intensity of the jumps is nonlinearly increasing and considerably larger than about 1.5. While, in light of these results, our estimated jump probabilities are likely upward biased, their size is arguably hard to justify given a model implying only 1.5 jumps per year. This simulation evidence provides some support for recent studies (relying on high-frequency
data) indicating the presence of a larger (than 1 or 2) number of jumps in daily returns per year.

We conclude with three observations. First, aside from nonlinearities in leverage, we find that the single most important deviation from affine stochastic volatility models with Gaussian jumps in returns and exponential jumps in volatility is the nonlinear shape of the variance’s diffusion function. Simulations show that this function is generally estimated accurately. Second, the features of the jumps (their probability and jump distribution) appear to be more easily identifiable for variance than for the return process. In particular, the size of the return jumps can be downward biased, whereas the probability of jumps in returns has the potential to be substantially upward biased. This said, while our jump intensities may be upward biased, they are hard to reconcile with a model with only 1.5 jumps per year. We confirm, as in previous work, that detection and measurement of jumps in returns is a hard empirical problem. Third, we emphasize that, while very informative, simple 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. We leave these issues for future work.

11 Conclusions

To be added.

12 Appendix A

This Appendix adapts our theory to a variety of integrated variance estimators recently proposed in the literature. Specifically, we demonstrate the coherence of Assumption 2 with the properties of these estimators when re-defined in their spot variance version, i.e. $\hat{\sigma}^2_t = \frac{\hat{V}_t}{\hat{\sigma}_n^2}$ with $\phi_{n,T} \to 0$, as proposed in this work. We do so in the absence of return jumps and microstructure noise as well as in the presence of either return jumps or market microstructure noise.

Case 1. $dJ^r = 0$, no 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. Realized range (Christensen and Podolskij, 2007): $\alpha = \frac{1}{2}$, $\beta = 0$, $a \approx 0.4$, $b = 0$, and $\eta = 1$.

3. Fourier estimator (Malliavin and Mancino, 2002): same as realized variance.

Case 2. $dJ^r \neq 0$, no microstructure noise.
1. **Bypower variation** (Barndorff-Nielsen and Shephard, 2004, 2005): \( \alpha = \frac{1}{2}, \beta = 0, a \approx 2.6, b = 0, \) and \( \eta = 1. \)

2. **Threshold estimator** (Mancini, 2007): \( \alpha = \frac{1}{2}, \beta = 0, a \approx 2, b = 0, \eta = 1. \)

3. **Threshold bipower variation** (Corsi et al., 2008): same as bypower variation.

**Case 3.** \( dJ^r = 0, \) with microstructure noise.

1. **Two-scale estimator** (Zhang et al., 2005): In what follows we consider alternative cases leading to \( b = 0, \) and \( \eta = 1. \)

From Bandi and Russell (2006), Theorem 2, when \( \phi_{n,T} \to 0 \) and \( k \to \infty, \) the dominating terms of the estimator's variance decomposition, i.e., \( \mathbf{V}_{\sigma^2} \left( \widehat{V}^{ZMA} - \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds \right), \) are

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

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

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

If \( K = \tau k^{2/3}, \) then \( v = \frac{K}{k} = \tau \frac{1}{k^{2/3}} \) and

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

or

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

Notice, also, that

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

\[
\to 0.
\]

If \( K = \tau \phi_{n,T} k^{2/3} \) with \( \tau \phi_{n,T} = \left( \frac{12 (E(\varepsilon^2))^2}{\phi_{n,T} (\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^4 ds)^{1/3}} \right) \) and \( \phi_{n,T} \) is so that \( \frac{K}{k} \to 0, \) then
\[ V_{\sigma^2} \left\{ k^{1/6} \left( \frac{\hat{V}_{ZMA}}{\phi_{n,T}} - \frac{\int_{T/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \]

\[
\sim \frac{4}{3} \tau^{1/3} \phi_{n,T} \frac{\int_{T/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}^2} \left[ \frac{8}{\tau^2 \phi_{n,T}} \left( \mathbb{E}(\varepsilon^2) \right)^2 \right]^{2/3} \phi_{n,T}^4 \]

and

\[ V_{\sigma^2} \left\{ \left( \phi_{n,T} \right)^{1/3} k^{1/6} \left( \frac{\hat{V}_{ZMA}}{\phi_{n,T}} - \frac{\int_{T/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \sim 2 \left( 12 \left( \mathbb{E}(\varepsilon^2) \right)^2 \right)^{1/3} \left( \sigma_{T/n}^2 \right)^{2/3} . \quad (29) \]

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

\[ E_{\sigma^2} \left\{ \left( \phi_{n,T} \right)^{1/3} k^{1/6} \left( \frac{\hat{V}_{ZMA}}{\phi_{n,T}} - \frac{\int_{T/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \sim -k^{1/6} \int_{T/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds \]

\[
\sim -k^{1/6} \int_{T/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds \quad \frac{k}{\phi_{n,T}} \left( \frac{k}{\phi_{n,T}} \right)^{2/3} \phi_{n,T} \]

\[
\sim - \frac{k^{1/6}}{k^{2/3}} \int_{T/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds \quad \phi_{n,T}^{2/3} \phi_{n,T} \]

\[
\sim - \frac{k^{1/6}}{k^{2/3}} \int_{T/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds \quad \phi_{n,T}^{2/3} \phi_{n,T} \]

\[
\sim - \frac{1}{k^{1/6} \phi_{n,T}^{1/3}} \sigma_{T/n}^2 \quad \phi_{n,T} \]

\[
\to 0,
\]

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

\[ \hat{V}_{ZMA} - \int_{T/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds = O_p \left( \frac{K}{k} \phi_{n,T} \right)^2 \]

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

\[ V_{\sigma^2} \left\{ \left( \frac{\hat{V}_{ZMA}}{\phi_{n,T}} - \frac{\int_{T/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \sim \frac{4}{3} \tau^{2/3} \phi_{n,T} \left( \frac{k}{\phi_{n,T}} \right)^{2/3} \left( \frac{8}{\tau^2 \phi_{n,T}} \left( \mathbb{E}(\varepsilon^2) \right)^2 \right)^{4/3} \phi_{n,T}^4 \]

37
and

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

The optimal \( \tau \) is therefore equal to \( \frac{12 (E(\varepsilon^2))^2}{\sigma_{i-T/n}^4} \)^{1/3} which leads to

\[ V_{\sigma^2} \left\{ \left( \phi_{n,T} \right)^{1/3} \frac{1}{k^{1/6}} \left( \frac{\hat{V}_{ZMA} - \int_{i-T/n}^{i-T/n + \phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \sim \frac{4}{3} \tau \sigma_{i-T/n}^4 + 8 \frac{(E(\varepsilon^2))^2}{\tau^2} \]

\[ \sim 2 \left( \frac{12 (E(\varepsilon^2))^2}{\tau^2} \right)^{1/3} \left( \sigma_{i-T/n}^4 \right)^{2/3} . \]

The final expression is the same as that in Eq. (29) above since \( K = \tau_{\phi_{n,T}} k^{2/3} \) with \( \tau_{\phi_{n,T}} = \frac{12 (E(\varepsilon^2))^2}{\phi_{n,T} \int_{i-T/n}^{i-T/n + \phi_{n,T}} \sigma_s^4 ds} \) can be re-defined as \( K = \tau_{\phi_{n,T}} \left( \frac{k}{\phi_{n,T}} \right)^{2/3} \left( \phi_{n,T} \right)^{2/3} \sim K^0 \).

3. Kernel estimators (Barndorff-Nielsen et al., 2006). We again consider alternative cases yielding \( b \neq 0 \), \( \beta \neq 0 \), and \( \eta \neq 1 \).

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

\[ V_{\sigma^2} \left\{ \left( \phi_{n,T} \right)^{1/3} \frac{1}{k^{1/6}} \left( \frac{\hat{V}_{BNHLS} - \int_{i-T/n}^{i-T/n + \phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \]

\[ \sim 4g_{\beta,0} \left( \phi_{n,T} \int_{i-T/n}^{i-T/n + \phi_{n,T}} \sigma_s^4 ds \right) v \]

\[ -4 \frac{1}{\nu k} \{ g'(0) + g_{\beta,0} \} \left\{ 2E(\varepsilon^2) \int_{i-T/n}^{i-T/n + \phi_{n,T}} \sigma_s^2 ds + E(\varepsilon^2)^2 \right\} \]

\[ +4E(\varepsilon^2)^2 \left\{ \frac{1}{\nu^2 k^2} \{ g'(0)^2 + g'(1)^2 \} + \frac{1}{\nu^3 k^3} \{ g''(0)^2 + g_{\beta,0} \} \right\} \]

\[ -4E(\varepsilon^2)^2 \frac{1}{\nu k} \frac{1}{2} \{ g'(0)^2 \} \],

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

\[ V_{\sigma^2} \left\{ \left( \phi_{n,T} \right)^{1/3} \frac{1}{k^{1/6}} \left( \frac{\hat{V}_{BNHLS} - \int_{i-T/n}^{i-T/n + \phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \]

\[ \sim 4g_{\beta,0} \left( \phi_{n,T} \int_{i-T/n}^{i-T/n + \phi_{n,T}} \sigma_s^4 ds \right) v + 4E(\varepsilon^2)^2 \frac{1}{\nu^2 k^2} \{ g'(0)^2 + g'(1)^2 \} . \]

38
Similar expressions as in the two-scale case arise. In particular, if $K = \tau k^{2/3}$, then

$$V_{\sigma^2} \left\{ k^{1/6} \left( \frac{\hat{V}_{BNHLS}}{\phi_{n,T}} - \frac{\int_{t/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \sim 4\tau g_{0,0}^* \left( \frac{\phi_{n,T}}{\tau^{1/2}} \right)^2 \left( \frac{\sigma_{\tau/n}^4}{\phi_{n,T}} \right)^{1/3} \left( \frac{\hat{V}_{BNHLS}}{\phi_{n,T}} - \frac{\int_{t/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right)$$

$$+ \frac{1}{(\phi_{n,T})^2} E(\varepsilon^2) \left( \frac{\phi_{n,T}}{\tau^{1/2}} \right)^2 \left( \frac{\sigma_{\tau/n}^4}{\phi_{n,T}} \right)^{1/3} \left( \frac{\hat{V}_{BNHLS}}{\phi_{n,T}} - \frac{\int_{t/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right)$$

and

$$V_{\sigma^2} \left\{ \phi_{n,T} k^{1/6} \left( \frac{\hat{V}_{ZMA}}{\phi_{n,T}} - \frac{\int_{t/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \sim 4\tau g_{0,0}^* \left( \frac{\phi_{n,T}}{\tau^{1/2}} \right)^{2/3} \left( \frac{\sigma_{\tau/n}^4}{\phi_{n,T}} \right)^{1/3} \left( \frac{\hat{V}_{ZMA}}{\phi_{n,T}} - \frac{\int_{t/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right)$$

$$+ \frac{1}{(\phi_{n,T})^2} E(\varepsilon^2) \left( \frac{\phi_{n,T}}{\tau^{1/2}} \right)^{2/3} \left( \frac{\sigma_{\tau/n}^4}{\phi_{n,T}} \right)^{1/3} \left( \frac{\hat{V}_{ZMA}}{\phi_{n,T}} - \frac{\int_{t/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right)$$

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

$$V_{\sigma^2} \left\{ \phi_{n,T} k^{1/6} \left( \frac{\hat{V}_{ZMA}}{\phi_{n,T}} - \frac{\int_{t/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \sim 4\tau g_{0,0}^* \left( \frac{\phi_{n,T}}{\tau^{1/2}} \right)^{2/3} \left( \frac{\sigma_{\tau/n}^4}{\phi_{n,T}} \right)^{1/3} \left( \frac{\hat{V}_{ZMA}}{\phi_{n,T}} - \frac{\int_{t/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right)$$

which is an adaptation to spot volatility estimation of the optimal choice of Barndorff-Nielsen et al. (2006). This choice implies

$$V_{\sigma^2} \left\{ \phi_{n,T} k^{1/6} \left( \frac{\hat{V}_{ZMA}}{\phi_{n,T}} - \frac{\int_{t/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\} \sim 4\tau g_{0,0}^* \left( \frac{\phi_{n,T}}{\tau^{1/2}} \right)^{2/3} \left( \frac{\sigma_{\tau/n}^4}{\phi_{n,T}} \right)^{1/3} \left( \frac{\hat{V}_{ZMA}}{\phi_{n,T}} - \frac{\int_{t/n}^{T/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right)$$

The optimal $\tau$ is now

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

which is an adaptation to spot volatility estimation of the optimal choice of Barndorff-Nielsen et al. (2006). This choice implies

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

Now notice that if $g'(0) = 0$ and $g'(1) = 0$, then
\[
V_{\sigma^2} \left\{ \left( \hat{V}_{\text{BNHLS}} - \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds \right) \right\}
\]
\[
\sim 4g_{0,0}^\bullet \left( \frac{\phi_{n,T}}{\phi_{n,T}} \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds \right) v
\]
\[-4 \frac{1}{v^2} \{g_0^0 \} \{E(\varepsilon)^2\} \]
\[+4E(\varepsilon)^2 \left[ \frac{1}{v^3 k^2} \{g_0^0(0)^2 + g_0^4 \} \right].
\]

If \( K = \tau k^{1/2} \), then

\[
V_{\sigma^2} \left\{ \left( \frac{\hat{V}_{\text{BNHLS}}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\}
\]
\[
\sim 4\tau g_{0,0}^\bullet \left( \frac{\phi_{n,T}}{\phi_{n,T}} \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds \right) \frac{1}{(\phi_{n,T})^2} \frac{1}{k^{1/2}} - 4 \frac{1}{\tau^{1/2} (\phi_{n,T})^2} \{g_0^0 \} \{E(\varepsilon)^2\}
\]
\[+4 \left( \frac{E(\varepsilon)^2}{\phi_{n,T}} \right)^2 \left[ \frac{1}{v^3 k^2} \{g_0^0(0)^2 + g_0^4 \} \right]
\]
and

\[
V_{\sigma^2} \left\{ \frac{\phi_{n,T} k^{1/4}}{} \left( \frac{\hat{V}_{\text{BNHLS}}}{\phi_{n,T}} - \frac{\int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds}{\phi_{n,T}} \right) \right\}
\]
\[
\sim 4 \left[ \frac{1}{v^3} \{g_0^0(0)^2 + g_0^4 \} - \frac{1}{\tau} \{g_0^0 \} \right] E(\varepsilon)^2.
\]

13 Appendix B

We begin with useful preliminary lemmas.

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

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

If, in addition, \( h_{n,T}L_{\sigma^2(T,x)} = O_{u.s.}(1) \), then
\[
\sqrt{h_{n,T}\tilde{L}_{\sigma^2}(T, x)} \{m_1(x) - m(x) - \Gamma_m(x)\} \Rightarrow N(0, K_2\Lambda^2(x)),
\]

with
\[
\Gamma_m(x) = h_{n,T}^2K_1 \left[ m'(x) \frac{s'(x)}{s(x)} + \frac{1}{2} m''(x) \right],
\]

where \(s(x)\) is the diffusion’s speed measure.

**Lemma A.2. (Bandi and Phillips, 2003)** Assume \(dJ_t^x = 0 \forall t\). Also, assume \(\Delta_{n,T} = \frac{T}{n} \to 0\) and \(h_{n,T} \to 0\) (as \(n, T \to \infty\)) in such a way as to guarantee that \(\frac{L_{\sigma^2}(T, x)}{h_{n,T}} \left(\Delta_{n,T} \log \left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} \overset{a.s.}{\to} 0\), where \(L_{\sigma^2}(T, x)\) is the chronological local time of \(\sigma^2\). Then,
\[
\Lambda^2_{11}(x) = 1 \sum_{i=0}^{n-1} K \left( \frac{\sigma_{i+1/n}^2 - x}{h_{n,T}} \right) \left( f(\sigma_{i+1/T/n}^2) - f(\sigma_{i/T/n}^2) \right)^2 \overset{a.s.}{\to} \Lambda^2(x).
\]

If, in addition, \(\frac{h_{n,T}^5 \tilde{L}_{\sigma^2}(T, x)}{\Delta_{n,T}} = O_{a.s.}(1)\), then
\[
\sqrt{h_{n,T}\tilde{L}_{\sigma^2}(T, x)} \frac{\{\Lambda^2_{11}(x) - \Lambda^2(x) - \Gamma_{\Lambda}(x)\}}{\Delta_{n,T}} \Rightarrow N(0, 2K_2\Lambda^4(x)),
\]

with
\[
\Gamma_{\Lambda}(x) = h_{n,T}^2K_1 \left[ \Lambda'^2(x) \frac{s'(x)}{s(x)} + \frac{1}{2} \Lambda''^2(x) \right],
\]

where \(s(x)\) is the diffusion’s speed measure.

**Lemma A.3. (Bandi and Nguyen, 2003)** Assume \(\Delta_{n,T} = \frac{T}{n} \to 0\) and \(h_{n,T} \to 0\) (as \(n, T \to \infty\)) in such a way as to guarantee that \(\frac{L_{\sigma^2}(T, x)}{h_{n,T}} \left(\Delta_{n,T} \log \left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2} \overset{a.s.}{\to} 0\), where \(L_{\sigma^2}(T, x)\) is the chronological local time of \(\sigma^2\). Then,
\[
\theta^j_1(x) = 1 \sum_{i=0}^{n-1} K \left( \frac{\sigma_{i+1/n}^2 - x}{h_{n,T}} \right) \left( f(\sigma_{i+1/T/n}^2) - f(\sigma_{i/T/n}^2) \right)^j \overset{a.s.}{\to} \theta^j(x) \quad \forall j \geq 1.
\]

If, in addition, \(h_{n,T}^5 \tilde{L}_{\sigma^2}(T, x) = O_{a.s.}(1)\), then
\[
\sqrt{h_{n,T}\tilde{L}_{\sigma^2}(T, x)} \\{\theta^j_1(x) - \theta^j(x) - \Gamma_{\theta^j}(x)\} \Rightarrow N(0, K_2\theta^2j(x)),
\]

with
\[
\Gamma_{\theta}(x) = h_{n,T}^2 K_1 \left[ \theta'(x) \frac{s'(x)}{s(x)} + \frac{1}{2} \theta''(x) \right],
\]

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

**Proof of Theorem 1.** We wish to show that

\[
\tilde{L}_n(x) = \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} K \left( \frac{\sigma_{i\Delta_{n,T}}^2 - x}{h_{n,T}} \right) - \int_0^T \frac{1}{h_{n,T}} K \left( \frac{\sigma_s^2 - x}{h_{n,T}} \right) ds = o_p(1).
\]

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

\[
\tilde{L}_n(x) \leq \frac{1}{h_{n,T}} \sum_{i=0}^{n-1} \int_{iT/n}^{(i+1)T/n} \left| K \left( \frac{\sigma_{i\Delta_{n,T}}^2 - x}{h_{n,T}} \right) - K \left( \frac{\sigma_s^2 - x}{h_{n,T}} \right) \right| ds
\]

\[
+ \frac{\Delta_{n,T}}{h_{n,T}} K \left( \frac{\sigma_0^2 - x}{h_{n,T}} \right) + \frac{\Delta_{n,T}}{h_{n,T}} K \left( \frac{\sigma^2_{n\Delta_{n,T}} - x}{h_{n,T}} \right)
\]

\[
\leq \frac{1}{h_{n,T}} \sum_{i=0}^{n-1} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left| K' \left( \frac{\sigma_{i\Delta_{n,T}}^2 - x}{h_{n,T}} \right) \right| \left| \left( \frac{\sigma_s^2 - \sigma_{i\Delta_{n,T}}^2}{h_{n,T}} \right) \right| ds + 2C_1 O_{a.s.} \left( \frac{\Delta_{n,T}}{h_{n,T}} \right),
\]

where \( \sigma_{i\Delta_{n,T}}^2 \) is a value on the line segment connecting \( \sigma_{i\Delta_{n,T}}^2 \) with \( \sigma_s^2 \). Now notice that

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

by the diffusion’s modulus of continuity. Also, given Assumption 2,

\[
\sigma_{i\Delta_{n,T}}^2 - \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma_s^2 ds = O_p \left( \frac{1}{k^{\alpha} \phi_{n,T}} \right)
\]

uniformly over \( i = 1, \ldots, n \). Finally,

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

This implies that

\[
\sup_{i \leq n} \left| \frac{\sigma_s^2 - \sigma_{i\Delta_{n,T}}^2}{h_{n,T}} \right| = O_p \left( \frac{1}{h_{n,T}} \left( \phi_{n,T} \log \frac{1}{\phi_{n,T}} \right)^{1/2} + \frac{1}{h_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} + \frac{1}{k^{\alpha} \phi_{n,T} h_{n,T}} \right)
\]

and, neglecting the negligible (under our assumptions on the bandwidth \( h_{n,T} \) term \( O_{a.s.} \left( \frac{\Delta_{n,T}}{h_{n,T}} \right) \),
\[
\bar{L}_n(x) \leq O_p(g(T, n, k, \phi_{n,T})) \sum_{i=0}^{n-1} \int_{\Delta_{n,T}} \left| K_i \left( \frac{\sigma^2_{\Delta_{n,T}} - x}{h_{n,T}} \right) \right| ds
\]
\[
= O_p(g(T, n, k, \phi_{n,T})) \int_0^T \left| K_i \left( \frac{\sigma^2 - x}{h_{n,T}} + o_p(1) \right) \right| ds
\]
\[
= O_p(g(T, n, k, \phi_{n,T})) \int_{-\infty}^\infty \left| K_i' (v + o_p(1)) \right| \bar{L}_{\sigma^2} (T, v h_{n,T} + x) dv
\]
\[
= C_2 O_p(g(T, n, k, \phi_{n,T}) \bar{L}_{\sigma^2} (T, x))
\]

by the occupation time formula for diffusions (see, e.g., ...) and the integrability of \( K' (\cdot) \) from Assumption 1.

**Proof of Theorem 2.**

Write
\[
\bar{K}_i = \Phi \left( \frac{\sigma^2_{\Delta_{n,T}} - x}{h_{n,T}} \right), \quad K_i = \Phi \left( \frac{\sigma^2_{\Delta_{n,T}} - x}{h_{n,T}} \right), \quad \bar{K}_i = \Phi \left( \frac{\sigma^2_{\Delta_{n,T}} - x}{h_{n,T}} \right) + o_p(1)
\]

Immediately, \( \bar{K}_i - K_i = O_p \left( \frac{1}{h_{n,T} k_t^a \phi_{n,T}^2} \right) \), \( K_i^* - K_i = O_p \left( \frac{1}{h_{n,T}} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} \right) \),

\[
\bar{K}_i - K_i = O_p \left( \frac{1}{h_{n,T} k_t^a \phi_{n,T}^2} + \frac{1}{h_{n,T}} \left( \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right) \right)^{1/2} \right) = O_p(q(T, n, k, \phi_{n,T}))
\]

and

\[
f(\sigma^2_{\Delta_{n,T}}) - f(\bar{\sigma}^2_{\Delta_{n,T}}) = f(\sigma^2_{\Delta_{n,T}}) - f(\bar{\sigma}^2_{\Delta_{n,T}}) + O_p \left( h_{n,T} q(T, n, k, \phi_{n,T}) \right).
\]

Now, write
\[
\tilde{\theta}^1(x) = \sum_{i=1}^n \left[ K_i + O_p(q(T, n, k, \phi_{n,T})) \right] \left[ f(\sigma^2_{\Delta_{n,T}}) - f(\bar{\sigma}^2_{\Delta_{n,T}}) + O_p \left( h_{n,T} q(T, n, k, \phi_{n,T}) \right) \right]
\]

and, since,
\[
\frac{1}{\Delta_{n,T} \sum_{i=1}^n K_i + O_p(q(T, n, k, \phi_{n,T}))}
\]
\[
= \frac{1}{\Delta_{n,T} \sum_{i=1}^n K_i} - \frac{TO_p(q(T, n, k, \phi_{n,T}))}{(\Delta_{n,T} \sum_{i=1}^n K_i)^2} + o_p(q(T, n, k, \phi_{n,T}))
\]

then, neglecting higher-order terms in \( q(T, n, k, \phi_{n,T}) \),
\[
\hat{\theta}^1(x) = m(x)
\]
\[
= TO_p(q(T, n, k, \phi_{n, T}) \sum_{i=1}^{n} [K_i + O_p(q(T, n, k, \phi_{n, T}))] \left[ f(\sigma_{(i+1)\Delta_{n,T}}^2) - f(\sigma_{i\Delta_{n,T}}^2) + O_p(h_{n,T}q(T, n, k, \phi_{n, T})) \right]
\]
\[
+ \left( \Delta_{n,T} \sum_{i=1}^{n} K_i \right)^2
\]
\[
= \hat{m}(x) + R_1 + R_2 + R_3.
\]

Since \( \hat{m}(x) = m(x) + o_p(1) \) from Lemma 1, we only need to show that the remainder term \( R_1 + R_2 + R_3 = o_p(1) \). Write

\[
R_3 = O_p \left( \frac{h_{n,T}q(T, n, k, \phi_{n, T})}{\Delta_{n,T}} \right)
\]
\[
= O_p \left( \frac{1}{\Delta_{n,T}k^2\phi_{n, T}^2} + \frac{1}{\Delta_{n,T}} \left( \phi_{n, T} \log \left( \frac{1}{\phi_{n, T}} \right) \right)^{1/2} \right)
\]
\[
\to 0
\]

under our assumptions. Now write

\[
R_1 = TO_p(q(T, n, k, \phi_{n, T})) \sum_{i=1}^{n} K_i [f(\sigma_{(i+1)\Delta_{n,T}}^2) - f(\sigma_{i\Delta_{n,T}}^2) + O_p(h_{n,T}q(T, n, k, \phi_{n, T}))]
\]
\[
+ \left( \Delta_{n,T} \sum_{i=1}^{n} K_i \right)^2
\]
\[
= R_4 + R_5.
\]

Hence,

\[
R_4 = O_p \left( \frac{Tq(T, n, k, \phi_{n, T})}{v(T)h_{n,T}\Delta_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} + \frac{Tq^2(T, n, k, \phi_{n, T})}{v(T)\Delta_{n,T}} \right) \to 0,
\]

\[
R_5 = O_p \left( \frac{Tq^2(T, n, k, \phi_{n, T})}{v(T)h_{n,T}^2\Delta_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} + \frac{Tq^3(T, n, k, \phi_{n, T})h_{n,T}}{v(T)h_{n,T}^2} \right)
\]
\[
= O_p \left( \frac{T^2q^2(T, n, k, \phi_{n, T})}{v(T)h_{n,T}^2\Delta_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} + \frac{T^2q^3(T, n, k, \phi_{n, T})h_{n,T}}{v(T)h_{n,T}^2\Delta_{n,T}} \right)
\]
\[
= O_p \left( \frac{T^2q^2(T, n, k, \phi_{n, T})\Delta_{n,T}}{h_{n,T}^2\Delta_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} + \frac{T^2q(T, n, k, \phi_{n, T})}{h_{n,T}} \right)
\]
\[
\to 0,
\]
and

\[
R_2 = O_p \left( \frac{q(T, n, k, \phi_{n,T})}{h_n T v(T)} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} + \frac{nq^2(T, n, k, \phi_{n,T}) h_n T}{h_n T v(T)} \right) \\
= O_p \left( \frac{q(T, n, k, \phi_{n,T})}{h_n T v(T) \Delta_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} + \frac{Tq^2(T, n, k, \phi_{n,T}) h_n T}{h_n T v(T) \Delta_{n,T}} \right) \\
= O_p \left( \frac{Tv(T)^{-1} q(T, n, k, \phi_{n,T})}{h_n T \Delta_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} + \frac{Tv(T)^{-1} q^2(T, n, k, \phi_{n,T})}{\Delta_{n,T}} \right) \\
\to 0,
\]

since

\[
\frac{Tv(T)^{-1} q(T, n, k, \phi_{n,T})}{\Delta_{n,T}} = \frac{Tv(T)^{-1} + Tv(T)^{-1} \phi_{n,T} \log \left( \frac{1}{\phi_{n,T}} \right)}{\Delta_{n,T} h_n T k_n^d} \to 0.
\]

The weak convergence result derives immediately from Lemma A.1 under our conditions on \( h_n T \) and \( \phi_{n,T} \).

**Proof of Theorem 2.** Following the same lines as for the proof of Theorem 1, write

\[
\hat{\theta}^2(x) = \sum_{i=1}^{n} \left[ K_i + O_p(q(T, n, k, \phi_{n,T})) \right] \left[ f(\sigma_{(i+1)\Delta_{n,T}}^2) - f(\sigma_{i\Delta_{n,T}}^2) \right] + O_p \left( h_n T q(T, n, k, \phi_{n,T}) \right)^2 \\
= \sum_{i=1}^{n} \left[ K_i + O_p(q(T, n, k, \phi_{n,T})) \right] \left[ f(\sigma_{(i+1)\Delta_{n,T}}^2) - f(\sigma_{i\Delta_{n,T}}^2) \right]^2 \\
\frac{\Delta_{n,T} \sum_{i=1}^{n} K_i}{\sum_{i=1}^{n} K_i} + \sum_{i=1}^{n} \left[ K_i + O_p(q(T, n, k, \phi_{n,T})) \right] \left[ O_p \left( h_n T q(T, n, k, \phi_{n,T}) \right)^2 \right] \\
\frac{\Delta_{n,T} \sum_{i=1}^{n} K_i}{\sum_{i=1}^{n} K_i} + 2 \sum_{i=1}^{n} \left[ K_i + O_p(q(T, n, k, \phi_{n,T})) \right] \left[ f(\sigma_{(i+1)\Delta_{n,T}}^2) - f(\sigma_{i\Delta_{n,T}}^2) \right] \left[ O_p \left( h_n T q(T, n, k, \phi_{n,T}) \right)^2 \right] \\
\frac{\Delta_{n,T} \sum_{i=1}^{n} K_i}{\sum_{i=1}^{n} K_i} - \sum_{i=1}^{n} \left[ K_i + O_p(q(T, n, k, \phi_{n,T})) \right] \left[ f(\sigma_{(i+1)\Delta_{n,T}}^2) - f(\sigma_{i\Delta_{n,T}}^2) \right] \left[ O_p \left( h_n T q(T, n, k, \phi_{n,T}) \right)^2 \right] \\
\frac{\Delta_{n,T} \sum_{i=1}^{n} K_i}{\sum_{i=1}^{n} K_i} + O_p(q(T, n, k, \phi_{n,T}) \sum_{i=1}^{n} \left[ K_i + O_p(q(T, n, k, \phi_{n,T})) \right] \left[ f(\sigma_{(i+1)\Delta_{n,T}}^2) - f(\sigma_{i\Delta_{n,T}}^2) \right] + O_p \left( h_n T q(T, n, k, \phi_{n,T}) \right)^2 \\
\frac{\Delta_{n,T} \sum_{i=1}^{n} K_i}{\sum_{i=1}^{n} K_i}^2
\]

or
\[ \hat{\theta}^1(x) \]

\[ = \Lambda_1^2(x) + \frac{\sum_{i=1}^{n} O_p(q(T, n, k, \phi_{n,T})) [f(\sigma_{i+1}^2) - f(\sigma_{i}^2)]^2}{\Delta_{n,T} \sum_{i=1}^{n} K_i} \]

\[ + \frac{\sum_{i=1}^{n} [K_i + O_p(q(T, n, k, \phi_{n,T}))] [O_p(h_{n,T}q(T, n, k, \phi_{n,T}))]^2}{\Delta_{n,T} \sum_{i=1}^{n} K_i} \]

\[ + 2 \sum_{i=1}^{n} [K_i + O_p(q(T, n, k, \phi_{n,T}))] [f(\sigma_{i+1}^2) - f(\sigma_{i}^2)] [O_p(h_{n,T}q(T, n, k, \phi_{n,T}))]^2 \]

\[ \frac{TO_p(q(T, n, k, \phi_{n,T})) \sum_{i=1}^{n} [K_i + O_p(q(T, n, k, \phi_{n,T}))] [f(\sigma_{i+1}^2) - f(\sigma_{i}^2)]^2}{(\Delta_{n,T} \sum_{i=1}^{n} K_i)} \]

\[ \frac{TO_p(q(T, n, k, \phi_{n,T})) \sum_{i=1}^{n} [K_i + O_p(q(T, n, k, \phi_{n,T}))] [O_p(h_{n,T}q(T, n, k, \phi_{n,T}))]^2}{(\Delta_{n,T} \sum_{i=1}^{n} K_i)} \]

\[ 2TO_p(q(T, n, k, \phi_{n,T})) \sum_{i=1}^{n} [K_i + O_p(q(T, n, k, \phi_{n,T}))] [f(\sigma_{i+1}^2) - f(\sigma_{i}^2)] [O_p(h_{n,T}q(T, n, k, \phi_{n,T}))] \]

\[ \frac{TO_p(q(T, n, k, \phi_{n,T})) \sum_{i=1}^{n} [K_i + O_p(q(T, n, k, \phi_{n,T}))] [f(\sigma_{i+1}^2) - f(\sigma_{i}^2)]^2}{(\Delta_{n,T} \sum_{i=1}^{n} K_i)} \]

\[ = \Lambda_1^2(x) + R_1 + R_2 + R_3 + R_4 + R_5 + R_6. \]

Hence,

\[ R_1 = O_p \left( \frac{\Delta_{n,T} q(T, n, k, \phi_{n,T})}{v(T) h_{n,T}} \right) \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right) \]

\[ = O_p \left( \frac{TV(T)^{-1} q(T, n, k, \phi_{n,T})}{\Delta_{n,T} h_{n,T}} \right) \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right) \]

\[ \rightarrow 0, \]

\[ R_2 = O_p \left( \frac{\Delta_{n,T} q^2(T, n, k, \phi_{n,T})}{h_{n,T}} \right) + O_p \left( \frac{\Delta_{n,T} q^3(T, n, k, \phi_{n,T})}{v(T) h_{n,T}} \right) \]

\[ = O_p \left( \frac{\Delta_{n,T} q^2(T, n, k, \phi_{n,T})}{h_{n,T}} \right) + O_p \left( \frac{TV(T)^{-1} h_{n,T} q^3(T, n, k, \phi_{n,T})}{\Delta_{n,T}} \right) \]

\[ \rightarrow 0, \]
\[ R_3 = O_p \left( \frac{h_n T q(T, n, k, \phi_{n,T})}{\Delta_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} \right) + O_p \left( \frac{n h_n T q^2(T, n, k, \phi_{n,T})}{v(T) h_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} \right) = O_p \left( \frac{h_n T q(T, n, k, \phi_{n,T})}{\Delta_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} \right) + O_p \left( \frac{T v(T)^{-1} q^2(T, n, k, \phi_{n,T})}{\Delta_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} \right) \rightarrow 0, \]

\[ R_4 = O_p \left( \frac{T q(T, n, k, \phi_{n,T})}{\Delta_{n,T} v(T) h_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right) + \frac{T n q^2(T, n, k, \phi_{n,T})}{v(T)^{2} h_{n,T}^2} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right) \right) = O_p \left( \frac{T v(T)^{-1} q(T, n, k, \phi_{n,T})}{\Delta_{n,T} h_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right) + \frac{T^2 v(T)^{-2} q^2(T, n, k, \phi_{n,T})}{\Delta_{n,T} h_{n,T}^2} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right) \right) \rightarrow 0, \]

\[ R_5 = O_p \left( \frac{T h_{n,T}^2 q^3(T, n, k, \phi_{n,T})}{\Delta_{n,T} v(T) h_{n,T}} + \frac{T h_{n,T}^2 q^4(T, n, k, \phi_{n,T})}{v(T)^2 h_{n,T}^2} \right) = O_p \left( \frac{T v(T)^{-1} h_{n,T} q^3(T, n, k, \phi_{n,T})}{\Delta_{n,T}} + \frac{T^2 v(T)^{-2} q^4(T, n, k, \phi_{n,T})}{\Delta_{n,T}} \right) \rightarrow 0, \]

and

\[ R_6 = O_p \left( \frac{T h_{n,T}^2 q^2(T, n, k, \phi_{n,T})}{\Delta_{n,T} v(T) h_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} \right) + O_p \left( \frac{T h_{n,T}^2 q^3(T, n, k, \phi_{n,T})}{v(T)^2 h_{n,T}^2} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} \right) = O_p \left( \frac{T v(T)^{-1} q^2(T, n, k, \phi_{n,T})}{\Delta_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} \right) + O_p \left( \frac{T^2 v(T)^{-2} q^3(T, n, k, \phi_{n,T})}{\Delta_{n,T} h_{n,T}} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} \right) \rightarrow 0. \]

Weak convergence readily derives from Lemma A.2 under our conditions on \( h_{n,T} \) and \( \phi_{n,T} \).
Proof of Theorem 4. The proof follows the same line as that of Theorem 1 with Lemma A.3 replacing Lemma A.1.

Proof of Theorem 5.
Write
\begin{align*}
U_{n,T}(r) &= \sum_{j=1}^{[nr]-1} w(\sigma^2_{rt/n}) \int_{j\Delta_n,T}^{(j+1)\Delta_n,T} c(\xi) f_2(\sigma^2_{rT/n}) \pi_\sigma(ds, d\xi) \\
&= \frac{T}{\pi} \sum_{i=1}^{\pi} \frac{1}{h_{n,T}} \frac{\sum_{j=1}^{[nr]-1} \Delta_n,T}{h_{n,T}} \Delta_n,T k \left( \frac{\sigma^2_{rT/n} - \sigma^2_{T/n}}{h_{n,T}} \right) f_1(\sigma^2_{T/n}).
\end{align*}

Hence, the quadratic variation is
\[ \sum_{j=1}^{[nr]-1} w^2(\sigma^2_{rT/n}) f_2^2(\sigma^2_{rT/n}) \int_{j\Delta_n,T}^{(j+1)\Delta_n,T} \lambda(X_s) E(c^2(\xi)) ds \]
\[ = \Delta_n,T \sum_{j=1}^{[nr]-1} w^2(\sigma^2_{rT/n}) f_2^2(\sigma^2_{rT/n}) \lambda(\sigma^2_{rT/n}) E(c^2(\xi)) \]
\[ = \Delta_n,T \sum_{j=1}^{[nr]-1} \left( \frac{T}{\pi} \sum_{i=1}^{\pi} \frac{1}{h_{n,T}} \frac{\sum_{j=1}^{[nr]-1} \Delta_n,T}{h_{n,T}} \Delta_n,T k \left( \frac{\sigma^2_{rT/n} - \sigma^2_{T/n}}{h_{n,T}} \right) f_1(\sigma^2_{T/n}) \right)^2 f_2^2(\sigma^2_{rT/n}) \lambda(\sigma^2_{rT/n}) E(c^2(\xi)^2) \]
\[ = \int_0^T \left( \int_0^T \frac{1}{h_{n,T}} \frac{k \left( \frac{b-a}{h_{n,T}} \right)}{d c} f_1(a) \right)^2 f_2^2(b) \lambda(b) E(c^2(\xi)^2) db \]
\[ = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \frac{1}{h_{n,T}} \frac{k \left( \frac{b-a}{h_{n,T}} \right)}{L(T, c) d c} f_1(a) L(T, a) \right)^2 f_2^2(b) \lambda(b) E(c^2(\xi)^2) L(T, b) db \]
\[ = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \frac{k \left( f \right)}{L(T, uh+b-fh) d u} f_1(b-fh) L(T, b-fh) d f \right)^2 g^2(b) \lambda(b) E(c^2(\xi)^2) L(T, b) db \]
\[ = \int_{-\infty}^{\infty} \left( \frac{f_2^2(b) L^2(T, b)}{L(T, b)} \right) f_2^2(b) \lambda(b) E(c^2(\xi)^2) db \]

Remaining derivations to be added.

Proof of Theorem 6. As in the case of Theorem 1, write
\[ \tilde{\mu}(\sigma^2) = \frac{\sum_{i=1}^n [K_i + O_p(q(T, n, k, \phi_n, T))] [\log(p_{i+1}T/n) - \log(p_{i+1}T/n)]}{\Delta_n,T \sum_{i=1}^n [K_i + O_p(q(T, n, k, \phi_n, T))]} \]
and, neglecting higher order terms,
\[ \hat{\mu}(\sigma^2) = \frac{\sum_{i=1}^{n} K_i \{ \log(p_{i+1}T/n) - \log(p_{i}T/n) \}}{\Delta_n T \sum_{i=1}^{n} K_i} - \frac{TO_p(q(T, n, k, \phi_{n,T}) \sum_{i=1}^{n} [K_i + O_p(q(T, n, k, \phi_{n,T}))] \{ \log(p_{i+1}T/n) - \log(p_{i}T/n) \}}{(\Delta_n T \sum_{i=1}^{n} K_i)^2} + \frac{\sum_{i=1}^{n} O_p(q(T, n, k, \phi_{n,T})) \log(p_{i+1}T/n) - \log(p_{i}T/n))}{\Delta_n T \sum_{i=1}^{n} K_i} \]

This implies \( R_1 = R_3 + R_4 \), where

\[ R_3 = - \frac{TO_p(q(T, n, k, \phi_{n,T}) \sum_{i=1}^{n} K_i \{ \log(p_{i+1}T/n) - \log(p_{i}T/n) \}}{(\Delta_n T \sum_{i=1}^{n} K_i)^2} \]

\[ = O_p \left( \frac{T v(T)^{-1} q(T, n, k, \phi_{n,T})}{\Delta_n T h_{n,T}} \left( \Delta_n T \log \left( \frac{1}{\Delta_n T} \right) \right)^{1/2} \right) \to 0, \]

and

\[ R_4 = - \frac{TO_p(q(T, n, k, \phi_{n,T}) \sum_{i=1}^{n} O_p(q(T, n, k, \phi_{n,T})) \log(p_{i+1}T/n) - \log(p_{i}T/n))}{(\Delta_n T \sum_{i=1}^{n} K_i)^2} \]

\[ = O_p \left( \frac{T^2 v(T)^{-2} q^2(T, n, k, \phi_{n,T})}{\Delta_n T h_{n,T}^2} \left( \Delta_n T \log \left( \frac{1}{\Delta_n T} \right) \right)^{1/2} \right) \to 0, \]

Also,

\[ R_2 = O_p \left( \frac{T v(T)^{-1} q(T, n, k, \phi_{n,T})}{h_{n,T} \Delta_n T} \left( \Delta_n T \log \left( \frac{1}{\Delta_n T} \right) \right)^{1/2} \right) \to 0. \]

Now write

\[ \tilde{\mu}(\sigma^2) = \frac{\sum_{i=1}^{n} K_i \int_{T/n}^{(i+1)T/n} \mu(\sigma_{s-}^2) ds}{\Delta_n T \sum_{i=1}^{n} K_i} + \frac{\sum_{i=1}^{n} K_i \int_{T/n}^{(i+1)T/n} \sigma_{s-}^2 dW_s}{\Delta_n T \sum_{i=1}^{n} K_i} \]

\[ + \frac{\sum_{i=1}^{n} K_i \int_{T/n}^{(i+1)T/n} \int \varphi(y) ds, dy}{\Delta_n T \sum_{i=1}^{n} K_i} \]

\[ = \alpha_{n,T} + \beta_{n,T} + \gamma_{n,T}. \]

Noting that \( \int_{T/n}^{(i+1)T/n} \sigma_{s-}^2 dW_s \) and \( \int_{T/n}^{(i+1)T/n} \int \varphi(y) ds, dy \) are martingale differences with finite variance, the methods in Bandi and Nguyen (2003) can be readily applied to show that \( \alpha_{n,T} \to 0 \), \( \mu(\sigma^2), \beta_{n,T} = o_p(1), \) and \( \gamma_{n,T} = o_p(1). \) The same methods yield the weak convergence result in the statement of the theorem.
Proof of Theorem 7. Using the method of proof of Theorem 1, write

\[ \tilde{C}(\sigma^2) = \frac{1}{\Delta_n,T} \sum_{i=1}^{n-1} K \left( \frac{\sigma_{(i+1)T/n}^2 - \sigma_{iT/n}^2}{h_{n,T}} \right) \left( \log(p_{(i+1)T/n}) - \log(p_{iT/n}) \right) \left( \sigma_{(i+1)T/n}^2 - \sigma_{iT/n}^2 \right) \]

\[ = \frac{1}{\Delta_n,T} \sum_{i=1}^{n-1} K \left( \frac{\sigma_{(i+1)T/n}^2 - \sigma_{iT/n}^2}{h_{n,T}} \right) \left( \log(p_{(i+1)T/n}) - \log(p_{iT/n}) \right) \left( \sigma_{(i+1)T/n}^2 - \sigma_{iT/n}^2 \right) + R_0, \]

where \( R_0 \to 0 \) under our assumptions on the bandwidths. Ito’s Lemma for Levy processes (see, e.g., Cont and Tankov, 2004) yields

\[ d \log(p) \bullet \sigma^2 = \log(p) d\sigma^2 + \sigma^2 d \log(p) + d \log(p) \bullet d\sigma^2. \]

Hence,

\[ \log \left( \frac{p_{(i+1)T/n}}{p_{iT/n}} \right) \sigma_{(i+1)T/n}^2 - \log \left( \frac{p_{iT/n}}{p_{iT/n}} \right) \sigma_{iT/n}^2 \]

\[ = \int_{iT/n}^{(i+1)T/n} \log(p_s) ds + \int_{iT/n}^{(i+1)T/n} \sigma_{s}^2 d \log(p_s) + \int_{iT/n}^{(i+1)T/n} d \log(p_s) \bullet d\sigma_s^2. \]

Now notice that

\[ \left( \log \left( \frac{p_{(i+1)T/n}}{p_{iT/n}} \right) \right) \left( \sigma_{(i+1)T/n}^2 - \sigma_{iT/n}^2 \right) \]

\[ = \log \left( \frac{p_{(i+1)T/n}}{p_{iT/n}} \right) \sigma_{(i+1)T/n}^2 - \log \left( \frac{p_{iT/n}}{p_{iT/n}} \right) \sigma_{iT/n}^2 \]

\[ + 2 \log \left( \frac{p_{iT/n}}{p_{iT/n}} \right) \sigma_{iT/n}^2 - \log \left( \frac{p_{(i+1)T/n}}{p_{iT/n}} \right) \sigma_{(i+1)T/n}^2 - \log \left( \frac{p_{iT/n}}{p_{iT/n}} \right) \sigma_{iT/n}^2 \]

\[ = \log \left( \frac{p_{(i+1)T/n}}{p_{iT/n}} \right) \sigma_{(i+1)T/n}^2 - \log \left( \frac{p_{iT/n}}{p_{iT/n}} \right) \sigma_{iT/n}^2 \]

\[ - \left( \log \left( \frac{p_{(i+1)T/n}}{p_{iT/n}} \right) \right) \sigma_{iT/n}^2 - \left( \sigma_{(i+1)T/n}^2 - \sigma_{iT/n}^2 \right) \log \left( \frac{p_{iT/n}}{p_{iT/n}} \right). \]

Thus,

\[ \left( \log \left( \frac{p_{(i+1)T/n}}{p_{iT/n}} \right) \right) \left( \sigma_{(i+1)T/n}^2 - \sigma_{iT/n}^2 \right) \]

\[ = \int_{iT/n}^{(i+1)T/n} \left( \sigma_{s}^2 - \sigma_{iT/n}^2 \right) \mu(s) ds + \int_{iT/n}^{(i+1)T/n} \left( \sigma_{s}^2 - \sigma_{iT/n}^2 \right) \sigma_{s} dW_s^\rho \]

\[ + \int_{iT/n}^{(i+1)T/n} \left( \sigma_{s}^2 - \sigma_{iT/n}^2 \right) \varphi^\rho_{s} ds, d\varphi \]

\[ + \int_{iT/n}^{(i+1)T/n} \left( \log(p_{s}) - \log(p_{iT/n}) \right) \mu(s) ds + \int_{iT/n}^{(i+1)T/n} \left( \log(p_{s}) - \log(p_{iT/n}) \right) \Lambda(s) \sigma_s dW_s^\sigma \]

\[ + \int_{iT/n}^{(i+1)T/n} \left( \log(p_{s}) - \log(p_{iT/n}) \right) \xi \sigma ds, d\xi \]

\[ + \sum_{iT/n \leq s \leq (i+1)T/n} \left( \Delta \log(p_{s}) \Delta \sigma_s^2 \right) + \int_{iT/n}^{(i+1)T/n} \rho(s) \sigma_s \Lambda(s) \sigma_s ds, \]
where $\Delta \log p_s = \log p_s - \log p_s^-$ and $\Delta \sigma^2_s = \sigma^2_s - \sigma^2_{s-}$. For convenience, in what follows we compensate the random measure $v_t(ds, d\xi)$ and, consequently, re-write $m(\sigma^2_{s-})$ as $m(\sigma^2_{s-}) + \lambda(\sigma^2_{s-})\mu$. In other words, we add and subtract the conditional first moment of the exponential jump size $\lambda(\sigma^2_{s-})\mu$ to make the jump component a martingale. Write

$$ R_1 = \frac{1}{\Delta_{n,T}} \sum_{i=1}^{n-1} K \left( \frac{\sigma^2_{i+1/T} - \sigma^2_i}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} \left( \sigma^2_{s-} - \sigma^2_{i+1/T} \right) \sigma_{s-} dW^r_s \right] + \frac{1}{\Delta_{n,T}} \sum_{i=1}^{n-1} K \left( \frac{\sigma^2_{i+1/T} - \sigma^2_i}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} \left( \log(p_s^-) - \log(p_{i+1/T/n}) \right) \Lambda(\sigma^2_{s-}) dW^\sigma_s \right] $$

$$ + O.a.s. \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} \sum_{i=1}^{n-1} K \left( \frac{\sigma^2_{i+1/T} - \sigma^2_i}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} \Lambda(\sigma^2_{s-}) dW^\sigma_s \right] $$

$$ + O.a.s. \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} \sum_{i=1}^{n-1} K \left( \frac{\sigma^2_{i+1/T} - \sigma^2_i}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} \Lambda(\sigma^2_{s-}) dW^\sigma_s \right] $$

$$ = O.a.s. \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} (\beta^1_{n,T} + \beta^2_{n,T}) $$

and
The terms $\beta_{n,T}^1, \beta_{n,T}^2, \gamma_{n,T}^1$ and $\gamma_{n,T}^2$ are averages of martingales difference sequences converging to zero (in probability) at rate $\sqrt{v(T)h_n,T}$ (see, e.g., Bandi and Nguyen, 2003). Thus, $R_1 + R_2 \to 0$. Now write

$$R_3 = \frac{1}{\Delta_{n,T}} \sum_{i=1}^{n-1} K \left( \frac{\sigma_{iT/n}^2 - \sigma_{T/n}^2}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} \left( \sigma_{s-}^2 - \sigma_{iT/n}^2 \right) \mu(ds) \right]$$

$$\leq O_{a.s.} \left( \Delta_{n,T} \log \left( \frac{1}{\Delta_{n,T}} \right) \right)^{1/2} \sum_{i=1}^{n-1} K \left( \frac{\sigma_{iT/n}^2 - \sigma_{T/n}^2}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} \mu(ds) \right]$$

Notice that,

$$R_4 = \frac{1}{\Delta_{n,T}} \sum_{i=1}^{n-1} K \left( \frac{\sigma_{iT/n}^2 - \sigma_{T/n}^2}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} \Delta \log p_{s-} \Delta \sigma_{s-}^2 \right] \to 0,$$
using $\Delta_{n,T} \sum_{i=1}^{n} K \left( \frac{\sigma^2_{T/n} - \sigma^2}{h_{n,T}} \right) = O_{a.s.} (h_{n,T} v(T)) \to \infty$ and the fact that the number of jumps in every trajectory is finite. Finally,

\[
\tilde{C}(\sigma^2) = R_0 + R_1 + R_2 + R_3 + R_4 \\
+ \frac{1}{\Delta_{n,T}} \sum_{i=1}^{n-1} K \left( \frac{\sigma^2_{T/n} - \sigma^2}{h_{n,T}} \right) \left[ \int_{iT/n}^{(i+1)T/n} \rho(\sigma^2_s) \sigma_s^2 \Lambda(\sigma^2_s) ds \right]
\]

\[
\begin{align*}
\Delta_{n,T} \sum_{i=1}^{n} K \left( \frac{\sigma^2_{T/n} - \sigma^2}{h_{n,T}} \right) \\
&\xrightarrow{p} \rho(\sigma^2) \sqrt{\sigma^2 \Lambda(\sigma^2)} ds.
\end{align*}
\]

Hence, $\tilde{\rho}(\sigma^2) = \frac{\tilde{C}(\sigma^2)}{\sqrt{\sigma^2 \Lambda(\sigma^2)}} \xrightarrow{p} \rho(\sigma^2)$ if $\Lambda(\sigma^2)$ is a consistent estimate.
References


Table 1.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>11.53</td>
<td>0.0015</td>
<td>117.27</td>
<td>3648</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$q$</th>
<th>ZMAadj</th>
<th>BNHLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Std.</td>
<td>11.18</td>
<td>9.19</td>
</tr>
<tr>
<td>Max</td>
<td>216</td>
<td>189</td>
</tr>
<tr>
<td>Min</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

We report summary statistics about the data (mid-quotes on SPIDERS between January 2, 1998 and March 31, 2006): average duration between mid-quote updates, average spread, average price, and average number of daily observations. The remaining statistics summarize the empirical distribution of the optimal number of auto-covariances of the bias-corrected two-scale estimators (ZMAadj) and the flat-top Tukey-Hanning kernel estimator.

Table 2.

<table>
<thead>
<tr>
<th>Returns</th>
<th>ZMAadj</th>
<th>BNHLS</th>
<th>BNS (stag.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.021</td>
<td>1.417</td>
<td>1.417</td>
</tr>
<tr>
<td>Std.</td>
<td>1.190</td>
<td>1.690</td>
<td>1.618</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.069</td>
<td>6.125</td>
<td>6.011</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.570</td>
<td>66.629</td>
<td>64.093</td>
</tr>
<tr>
<td>Max</td>
<td>5.867</td>
<td>27.110</td>
<td>25.789</td>
</tr>
<tr>
<td>Min</td>
<td>-7.015</td>
<td>-0.286</td>
<td>0.002</td>
</tr>
<tr>
<td>1st autocorr.</td>
<td>-0.026</td>
<td>0.752</td>
<td>0.762</td>
</tr>
</tbody>
</table>

We report summary statistics about the S&P500 return data and the variance estimates used in the main text, namely the bias-corrected two-scale estimator (ZMAadj), the modified Tukey-Hanning kernel estimator, and the staggered by-power variation estimator. The sample period is January 2, 1998 - March 31, 2006.
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 8(a)  
Price conditional kurtosis

Figure 8(b)  
Variance conditional kurtosis

[Graphs showing price and variance conditional kurtosis with different lines representing mean estimates, 2.5th percentile, 97.5th percentile, and kurtosis from data.]