Nonparametric Filtering of the Realized Spot Volatility: A Kernel-Based Approach

Author(s): Dennis Kristensen

Source: *Econometric Theory*, Feb., 2010, Vol. 26, No. 1 (Feb., 2010), pp. 60-93

Published by: Cambridge University Press

Stable URL: https://www.jstor.org/stable/40388620

Econometric Theory, **26**, 2010, 60–93.
doi:10.1017/S0266466609090616

NONPARAMETRIC FILTERING OF THE REALIZED SPOT VOLATILITY: A KERNEL-BASED APPROACH

DENNIS KRISTENSEN
Columbia University

and

CREATES

A kernel weighted version of the standard realized integrated volatility estimator is proposed. By different choices of the kernel and bandwidth, the measure allows us to focus on specific characteristics of the volatility process. In particular, as the bandwidth vanishes, an estimator of the realized spot volatility is obtained. We denote this the filtered spot volatility. We show consistency and asymptotic normality of the kernel smoothed realized volatility and the filtered spot volatility. We consider boundary issues and propose two methods to handle these. The choice of bandwidth is discussed and data-driven selection methods are proposed. A simulation study examines the finite sample properties of the estimators.

1. INTRODUCTION

Continuous-time models for the dynamics of asset returns have widespread use in financial economics. They lead to simple, yet elegant pricing formulas of financial instruments, and are the primary building blocks in the literature on portfolio management and risk analysis; see, for example, Björk (2004). One of the main components in these models is the conditional second moment or volatility of the processes, which plays a central role in asset pricing formulas. It has long been recognized that volatility varies over time, and considerable effort has been put into modeling and forecasting this variable; see, e.g., Andersen, Bollerslev, and Diebold (2005) and Shephard (2005) for reviews.

The increased access to high-frequency intradaily data of asset returns has given rise to new empirical measures of the volatility process. In particular, the so-called realized volatility measure has received considerable attention and been widely used in the empirical finance literature. The realized volatility gives a measure of the integrated volatility over a given time period and has been used

I wish to thank Bruce Hansen for fruitful discussions and helpful suggestions, and participants at the University of Mannheim economics seminar, the Greater NY Metropolitan Area Econometrics Colloquium 2006, the Zeuthen workshop 2006, and the conference on "Multivariate Modelling in Finance and Risk Management" for comments and suggestions. CREATES, the Center for Research in Econometric Analysis of Time Series, is funded by the Danish National Research Foundation. Address correspondence to Dennis Kristensen, Economics Department, International Affairs Building, MC 3308, 420 West 118th Street, New York, NY 10027, USA; e-mail: dk2313@columbia.edu.

  0266-4666/10 $15.00

in a wide range of applications from forecasting of daily volatility to detection of jump components.

It is, however, not obvious that the integrated volatility is the only relevant measure of volatility in asset returns. This has led to other measures being proposed such as the range-based volatility and power volatility; see, e.g., Alizadeh, Brandt, and Diebold (2002). Most of these alternatives are also integrated measures over some time window. This raises the question of what an appropriate choice of the window is. In empirical applications, one day is the standard time window length over which the (transformed) volatility is integrated. In many cases, one would expect that the optimal solution would be to try to recover the actual instantaneous (or spot) volatility instead; if nothing else, one would then be able to calculate any time integral over (a transformation of) the volatility, and thereby recover any of the integrated measures in the literature.

We here propose to estimate the instantaneous volatility by kernel methods. The estimator is a kernel-weighted version of the standard integrated volatility estimator, which depends on a kernel function and a time window/bandwidth chosen by the user.[1] For a fixed bandwidth and a uniform kernel, it collapses to the standard realized volatility measure, but in general it can be seen as a continuous-time weighted moving average of the instantaneous volatility. The bandwidth choice allows the user to focus on the volatility behavior at specific points in time, and to give different weights to the volatility over the window used. In particular, we demonstrate that as the bandwidth shrinks to zero, the spot volatility can be extracted. Thus, the inclusion of a kernel and bandwidth in the calculation of the integrated volatility allows us to get a better picture of the behavior of the volatility process.

Given high-frequency data, a nonparametric estimator of the kernel-smoothed integrated volatility is easily constructed using the same idea as for the realized volatility estimator: We simply take a kernel-weighted average of the squared increments of data. We derive its asymptotic properties, showing consistency and mixed asymptotic normality. We do this for both fixed and shrinking bandwidths. In the former case, our results are a generalization of already existing ones found in the literature to include weighting. In the latter case, the limit is the instantaneous volatility process, and we denote our estimator the filtered volatility in this case.

As already pointed out, a nice feature of the filtered volatility is that it can be used to estimate any functional of the instantaneous volatility, including the standard integrated volatility. We demonstrate that for a broad class of nonlinear integrated volatility measures, we obtain $\sqrt{n}$ -consistent estimators by substituting in the filtered volatility.

The filtered version of the instantaneous volatility has direct use in financial markets. For example, a trader will able to measure the most recent volatility of the market he trades in. Also, in option pricing with stochastic volatility, the current volatility is needed as an input in the option pricing formulas. Furthermore, the filtered volatility has several interesting applications in financial econometrics:

The filtered volatility can be used in the analysis of periodic components in intradaily volatility as found in, e.g., Andersen and Bollerslev (1997). By an appropriate choice of the kernel, it has potential usage in detecting jumps in the volatility process, see, e.g., Gijbels, Lambert, and Qiu (2007). The local nature of the filtered volatility means that the presence of market microstructure noise potentially can be dealt with.

Finally, the filtered volatility allows for a new estimation strategy of stochastic volatility models as originally proposed in Renò (2006); see also Renò (2008) and Kanaya and Kristensen (2008). Given that the volatility process is a latent variable, not observed by the econometrician, previous works on this have based the estimation of volatility models on the raw return data itself (Andersen and Lund, 1997; Chib, Nardari, and Shephard, 2002) or the realized integrated volatility (Andersen, Bollerslev, Diebold, and Labys, 2003; Bollerslev and Zhou, 2002). Our filtered version of the instantaneous volatility opens up for a new class of estimators, where one can directly estimate the stochastic volatility model by substituting the filtered version for actual observations of the volatility. One thereby circumvents the problem of latent variables/missing data.

The proposed kernel-smoothed version of the realized volatility can be regarded as a kernel regression estimator in the time domain. A similar approach to the estimation of the instantaneous volatility but in a discrete-time setting was taken in Mikosch and Stărica (2005). Fan, Jiang, Zhang, and Zhou (2003) also consider kernel estimation of deterministic functions of time in the context of term structure diffusion models.

Our estimator includes as a special case the rolling window estimator proposed by Foster and Nelson (1996); see also, Andreou and Ghysels (2002), Fan, Fan, and Lv (2007), and Mykland and Zhang (2003, 2006). Our theoretical results complement the ones found in these studies. Alternative approaches are pursued in Genon-Catalot, Laredo, and Picard (1992) and Malliavin and Mancino (2006). The former study considers a deterministic, smooth volatility process and use wavelets methods to estimate the instantaneous volatility. The latter study obtains an expression of the volatility in the frequency domain and derives an estimator of it in terms of Fourier transforms; see also, Barucci and Renò (2002) and Høg and Lund (2003) for related work.

While our kernel estimator is established in the time domain, spatial kernel estimators drift and diffusion estimators of a fully observed Markov process are proposed in Bandi and Phillips (2003); see also, Florens-Zmirou (1993). In their setting, the volatility process is a function of the observed process only, and they estimate the instantaneous volatility by kernel smoothing over the spatial domain of the observed process. We smooth over the time domain instead, which enables us to estimate the volatility process without imposing any Markov restrictions. We also consider a drift estimator in the time domain, and demonstrate that one cannot recover the drift process; this holds even as the time span over which the estimation is performed diverges. This is in contrast to Bandi and Phillips (2003), whose spatial estimator of the drift is consistent as the time span goes to infinity.

The remainder of the paper is organized as follows: In Section 2, we introduce the kernel smoothed measure of volatility and discuss its relationship to already existing measures. In Section 3, the asymptotic properties of the volatility estimator are established, while Section 4 deals with the equivalent drift estimator. The choice of bandwidth is discussed in Section 5, while the results of a simulation study are presented in Section 6. We conclude in Section 7 by describing how the estimator can be extended to a multivariate setting and allow for jumps and market microstructure noise. All proofs and lemmas can be found in Appendixes A and B, respectively.

2. KERNEL-SMOOTHED REALIZED VOLATILITY

Consider the Brownian semimartingale $\{X_t\} = \{X_t : t \geq 0\}$ solving

$$dX_t = \mu_t dt + \sigma_t dW_t, \tag{1}$$

where $\{W_t\}$ is a standard Brownian motion, while $\{\mu_t\}$ and $\{\sigma_t\}$ are adapted stochastic processes. The process $\left\{\sigma_t^2\right\}$ is usually denoted the (instantaneous/spot) volatility process, while $\{\mu_t\}$ is the drift process. In the finance literature, this is a commonly used model for log-asset prices, and the focus is normally on the volatility, since this is the essential ingredient in asset pricing. However, $\left\{\sigma_t^2\right\}$ is not observed and one instead has to rely on observations of the process $\{X_t\}$ to draw inference on the volatility.

One very fruitful approach to extracting information regarding the volatility is the so-called realized volatility, which has received considerable attention over the past decade. This estimator centers around the concept of the quadratic variation of $\{X_t\}$, which gives a convenient representation of the integrated volatility. The quadratic variation at time $t > 0$, $[X]_t$, can be defined as

$$[X]_t = \lim_{\Delta \to 0} \sum_{i=1}^{m} \left(X_{t_i} - X_{t_{i-1}}\right)^2 \tag{2}$$

for any partition $0 = t_0 < t_2 < \cdots < t_m = t$ with $\Delta := \max_{i=1,\dots,m} |t_i - t_{i-1}|$, cf. Protter (2004, Thm. II.22). The quadratic variation gives an alternative representation of the time integral of the volatility as follows:

$$[X]_t = \int_0^t \sigma_s^2 \, ds.$$

The right-hand side is normally referred to as the integrated volatility (IV). Thus, given a continuously observed trajectory $\{X_t : 1 \leq t \leq T\}$, the integrated volatility over any time window of length $h > 0$ in $[0, T]$ can be recovered perfectly,

$$\mathrm{IV}(\tau) := \int_{\tau-h}^{\tau} \sigma_s^2 \, ds = [X]_\tau - [X]_{\tau-h} = \int_{\tau-h}^{\tau} d[X]_s. \tag{3}$$

For example, if time is measured in days, $\mathrm{IV}(\tau)$, $\tau = 0, 1, \ldots, [T]-1$, with $h = 1$ will give us daily integrated volatility over the time span $[0, T]$. Here, $[T]$ denotes the integer part of T.

In general, however, a full trajectory is not available. The most we can hope for is a high-frequency discrete sample $\{X_{t_i} : i = 0, \ldots, n\}$ over the interval $[0, T]$. This can be used to obtain an estimate of the quadratic variation that in turn converges toward the integrated volatility as the time distance between observations shrinks to zero. A natural estimator of the quadratic variation is to simply take a sample average of the squared increments over the time interval of interest,

$$\widehat{[X]}_t = \sum_{i=1}^{n} \mathbb{I}\{t_{i-1} < t\} \left(\Delta X_{t_{i-1}}\right)^2,$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function, and $\Delta X_{t_{i-1}} = X_{t_i} - X_{t_{i-1}}$, $i = 1, \ldots, n$. This leads to the so-called realized volatility estimator of $\mathrm{IV}(\tau)$ given by

$$\widehat{\mathrm{IV}}(\tau) = \int_{\tau-h}^{\tau} d\widehat{[X]}_s = \sum_{i=1}^{n} \mathbb{I}\{\tau - h < t_{i-1} < \tau\} \left(\Delta X_{t_{i-1}}\right)^2. \quad \textbf{(4)}$$

This estimator has received widespread attention in empirical finance within the past decade (see, e.g., Andersen et al., 2003). Its theoretical properties have been studied in detail in, amongst others, Barndorff-Nielsen and Shephard (2004a,b, 2006) and Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006) under the infill assumption: Assuming that the time distance between observations $\Delta := \max_{i=1,\ldots,n} |t_i - t_{i-1}|$ shrinks to zero, consistency and mixed asymptotic normality of $\mathrm{RV}(\tau)$ can be derived.

One can regard the realized volatility estimator in (4) as a histogram estimator of the instantaneous volatility, where $h > 0$ is the bin width. Here, we propose an alternative measure of the integrated volatility and construct an estimator of it in the same spirit as the realized volatility. The measure we will consider is

$$\mathrm{KV}(\tau) = \int_0^T K_h(s-\tau)\sigma_s^2\, ds = \int_0^T K_h(s-\tau)\, d[X]_s,$$

where $K_h(z) := K(z/h)/h$, $K : \mathbb{R} \mapsto \mathbb{R}$ is a kernel which we normalize to $\int_{\mathbb{R}} K(z)\, dz = 1$, and $h > 0$ is the bandwidth. $\mathrm{KV}(\tau)$ delivers a kernel weighted average of the quadratic variation. Note that with

$$K(z) = \mathbb{I}\{-1 < z < 0\} : \mathrm{KV}(\tau) h = \mathrm{IV}(\tau).$$

So with the above uniform kernel, the standard integrated volatility can be recovered for a fixed band/window width. This is a member of the class of so-called one-sided kernels where $K(z) = 0$, $z > 0$. For this class of kernels, $\tau \mapsto \mathrm{KV}(\tau)$ is an adaptive process, since it only utilizes information available up to time τ of the observed process,

$$K(z) = 0, z > 0 : \mathrm{KV}(\tau) = \int_0^{\tau} K_h(s-\tau)\, d[X]_s.$$

The fact that with a one-sided kernel/filter we can calculate $\mathrm{KV}(\tau)$ from a continuous record of $\{X_t\}$ up to time τ makes it suitable for forecasting purposes. In general, with two-sided kernels, $\mathrm{KV}(\tau)$ takes a weighted average of the instantaneous volatility over the whole sample period relative to the point in time $0 < \tau < T$. The weighting scheme is jointly determined by the choice of K and h.

As demonstrated above, for fixed $h > 0$, $\mathrm{KV}(\tau)$ gives a weighted measure of the integrated volatility. However, as $h \to 0$, we are able to recover the instantaneous volatility at any point of continuity τ of $t \mapsto \sigma_t^2$. Using standard results for kernel estimators, one can easily show that

$$\sigma_\tau^2 = \lim_{h\to 0} \mathrm{KV}(\tau).$$

This holds irrespective of the kernel being one- or two-sided. So if our object of interest is the instantaneous volatility, $\mathrm{KV}(\tau)$ for $h > 0$ gives us a biased estimate of this. By letting the bandwidth shrink to zero however, we recover σ_τ^2. So while the integrated volatility gives us a measure of the volatility over a given window in time, we are also able to obtain a snapshot of the volatility at any given point in time by letting that window shrink to zero.

From a theoretical point of view, a measure of the instantaneous volatility should always be preferable to the integrated volatility, since the latter can always be calculated given the former. In fact, we shall consider the following generalized version of the standard integrated volatility over some fixed time window,

$$\mathrm{IV}_g = \int_0^T g\left(s, \sigma_s^2\right) ds, \tag{5}$$

where $g : [0, T] \times \mathbb{R} \mapsto \mathbb{R}$ is a time-dependent transformation of the volatility. This includes most standard measures. For example, with $g(t, x) = x$, we obtain the standard integrated volatility. But for general functions g, this measure cannot be estimated using standard realized volatility estimators.

A natural estimator of $\mathrm{KV}(\tau)$ is to extend $\widehat{\mathrm{IV}}(\tau)$ to include kernel weights. In the following we shall write $\widehat{\mathrm{KV}}(\tau)$ for the following kernel-smoothed sample average of the squared increments:

$$\widehat{\mathrm{KV}}(\tau) = \int_0^T K_h(s-\tau)\, d\widehat{[X]}_s = \sum_{i=1}^n K_h(t_{i-1}-\tau)\left(\Delta X_{t_{i-1}}\right)^2.$$

This is a Nadaraya-Watson (NW) type kernel estimator. Kernel smoothing is a familiar technique in nonparametric econometrics where kernels are used to recover objects such as densities and regression functions; an overview can be found in Silverman (1986). Here, we smooth over the time domain, which can be considered as the regressor while $\Delta X_{t_{i-1}}^2$ is the dependent variable. By an appropriate choice of K, the realized measure $\widehat{\mathrm{KV}}(\tau)$ includes as special cases the standard realized volatility estimator. Observe, however, that in contrast to the standard integrated volatility estimator, $\tau \mapsto \widehat{\mathrm{KV}}(\tau)$ here can be made continuous and differentiable by choosing K to have these properties.

When we consider shrinking bandwidth sequences, $h \to 0$, we shall write

$$\hat{\sigma}_\tau^2 = \sum_{i=1}^{n} K_h(t_{i-1} - \tau)\left(\Delta X_{t_{i-1}}\right)^2 \tag{6}$$

to emphasize that we are working with an estimator of the instantaneous volatility at time τ. An interpretation of $\hat{\sigma}_\tau^2$ is as a local version of the standard parametric estimator in the Black-Scholes model, $\hat{\sigma}^2 = \sum_{i=1}^{n}\left(\Delta X_{t_{i-1}}\right)^2/T$. Alternatively, one may use the estimator

$$\hat{\sigma}_\tau^2 = \frac{\sum_{i=1}^{n} K_h(t_{i-1} - \tau)\left(\Delta X_{t_{i-1}}\right)^2}{\sum_{i=1}^{n} K_h(t_{i-1} - \tau)(t_i - t_{i-1})}. \tag{7}$$

This is particularly useful when $t_1, \ldots, t_n$ are random such that we do not observe $\{X_t\}$ at equidistant time points. If the time points, on the other hand, are equidistant, then the denominator will converge toward $\mathbb{I}\{0 < \tau < T\}$ as $n \to \infty$ and $nh \to 0$, so the two estimators in (6) and (7) are asymptotically equivalent.

One usually uses symmetric kernels such as the Gaussian one in kernel smoothing, but as discussed above, one may prefer for some applications to use a one-sided kernel. One-sided kernels have the advantage that they are adapted to the observed process, and will in general lead to a more precise estimate when τ is close to T. In a standard regression framework, one-sided kernels are used in the estimation of end and jump points; see, for example, Zhang and Karunamuni (1998) and Gijbels, Pope, and Wand (2007). We can carry over much of the theory established there to our setting.

Finally, we note that many other choices of weighting functions could be used instead of $K_h(t_{i-1} - \tau)$. Most nonparametric estimators can be written on the form

$$\mathrm{RV}_w(\tau) \equiv \sum_{i=1}^{n} w(t_{i-1}, \tau)\left(\Delta X_{t_{i-1}}\right)^2 \tag{8}$$

for some weighting function $(s, \tau) \mapsto w(s, \tau)$. In particular, with $w(s, \tau) = K_h(s - \tau)$ we recover $\widehat{\mathrm{KV}}(\tau)$, but other choices are available such as the local linear and the rolling sample estimator. The local linear estimator is another kernel-based estimator where (after normalizing K to $\int z^2 K(z)\,dz = 1$) the weight is given by

$$w(s, \tau) = K_h(s - \tau)\left\{S_{n,2}(\tau) - \frac{s - \tau}{h} S_{n,1}(\tau)\right\}, \tag{9}$$

and $S_{n,k}(\tau) = \Delta h^{-k} \sum_{j=1}^{n} K_h\left(t_{j-1} - \tau\right)\left(t_{j-1} - \tau\right)^k$; see Fan and Gijbels (1996) for an introduction and Fan et al. (2003) for an application to the estimation of deterministic time trends in diffusion models. Foster and Nelson (1996) consider rolling sample estimators where $w_n(t_{i-1} - \tau)$ is a sequence of weighting functions satisfying $\Delta \sum_{i=1}^{n} w_n(t_{i-1} - \tau) = 1$ and other restrictions; see Foster and Nelson (1996) for more details. Most popular filters can be written on the form $w_n(s, \tau) = w((s - \tau)/(M\Delta))/(M\Delta)$, where $M\Delta = M_n \Delta_n$ determines the lag-length. Choosing $h = M\Delta$, we see that this class of rolling sample estimators takes the same form as the kernel estimator. Another popular filter is the one-sided exponential filter where $\hat{\sigma}_{t_k}^2 = (1 - \lambda) \sum_{i=1}^{k-1} \lambda^i (\Delta X_{t_{k-i}})^2$ as considered in, e.g., Fan et al. (2007). By defining $K_{e,1}(z) = \exp(-|z|) \mathbb{I}\{z \le 0\}$ and choosing $h = -T/t_k \log(\lambda)$, Gijbels et al. (1999) show that the exponential filter is a special case of the NW estimator. One can alternatively use the two-sided exponential filter that corresponds to $K_{e,2}(z) = \exp(-|z|)/2$. Suitably modified versions of the wavelets estimators considered in Genon-Catalot et al. (1992) and Malliavin and Mancino (2006) can also be written on the form (8).

Once $\hat{\sigma}_\tau^2$ has been obtained, it can be used for various purposes, as discussed in the introduction. For example, we can easily obtain an estimator of the following generalized version of the integrated volatility measure in (5) by simply plugging in $\hat{\sigma}_s^2$.

3. ASYMPTOTICS OF THE VOLATILITY ESTIMATOR

We here derive the asymptotics of the volatility estimators $\widehat{\mathrm{KV}}(\tau)$ and $\hat{\sigma}_\tau^2$ introduced in the Section 2. We shall throughout work under the following set of conditions:

Assumption A.1. The process is sampled at $t_i = i\Delta$, $i = 0, 1, \ldots, n$, such that $T = n\Delta$.

Assumption A.2. The processes $\{\mu_t\}$ and $\{\sigma_t^2\}$ are jointly independent of $\{W_t\}$.

Assumption A.3. The volatility processes $\{\sigma_t^2\}$ are locally bounded away from zero and the mean and volatility satisfy:

$$\lim_{\Delta \to 0} \Delta \sum_{i=1}^{n} \left| \mu_{s_i}^2 - \mu_{t_i}^2 \right| = 0, \quad \lim_{\Delta \to 0} \Delta \sum_{i=1}^{n} \left| \sigma_{s_i}^4 - \sigma_{t_i}^4 \right| = 0$$

for any sequences $(i-1)\Delta \le s_i \le t_i \le i\Delta$, $i = 1, \ldots, n$.

The assumption in (A.2) that the mean and volatility processes are independent of $\{W_t\}$ allows us in the following to make all arguments conditional on $\{\mu_t\}$ and $\{\sigma_t^2\}$, which greatly facilitates the proofs. In particular, (A.3) is stated as a deterministic assumption, since we condition on the given sample paths of the drift and volatility that generated the observations. On the other hand, (A.2) rules out leverage effects where σ_t^2 is correlated with W_t. This assumption does not appear to be strictly necessary: As demonstrated in Kanaya and Kristensen (2008), the spot volatility estimator remains consistent when allowing for leverage effects,[2] but the proofs become more complicated.

Assumption (A.3) imposes restrictions on the local behavior of the mean and volatility processes. It allows for deterministic patterns, jumps, and nonstationarity, and is automatically satisfied when the mean and volatility processes have continuous trajectories. In particular, it allows for standard diffusion models such as the Heston (1993), Hull and White (1987), and Stein and Stein (1991) models. We conjecture that (A.3) can be weakened by following the proof strategy of, e.g., Barndorff-Nielsen and Shephard (2006), but this will lead to longer and more complicated proofs, and we therefore maintain (A.3) here.

While the above conditions suffice to derive the asymptotics of $\widehat{\mathrm{KV}}(\tau)$ for fixed bandwidth $h > 0$, we need to impose smoothness assumptions on the volatility process for $h \to 0$ in order to control the bias component. A standard approach to bias reduction is to assume the object of interest is differentiable up to a certain order. This assumption is, however, violated by standard stochastic volatility models. Instead, we here introduce a more general smoothness condition that allows for $\{\sigma_t^2\}$ to have nondifferentiable trajectories as long as they are smooth of order $0 < \gamma < 1$ almost surely. To this end, we introduce the following function space:

DEFINITION 1. *The space* $C^{m,\gamma}[0,T]$, $m \geq 0$ *and* $0 < \gamma < 1$, *consists of functions* $f : [0,T] \mapsto \mathbb{R}$ *that are m times differentiable with the mth derivative,* $f^{(m)}(t)$, *satisfying*

$$\left|f^{(m)}(t+\delta) - f^{(m)}(t)\right| \leq L_f(t,|\delta|)\,|\delta|^{\gamma} + o\left(|\delta|^{\gamma}\right), \quad \delta \to 0,$$

where $\delta \mapsto L_f(t,\delta)$ *is a slowly varying (random) function at zero and* $t \mapsto L_f(t,0)$ *is continuous.*

We then require that the trajectories of $\{\sigma_t^2\}$ lie in this space:

Assumption A.4. $t \mapsto \sigma_t^2$ lies in $C^{m,\gamma}[0,T]$ for some $m \geq 0$ and $\gamma \geq 0$, and $\int_0^T |\mu_t|\,dt < \infty$.

This condition is a strengthening of (A.3); if (A.4) holds, the conditions imposed on $\{\sigma_t^2\}$ in (A.3) are satisfied. A similar smoothness condition with $m \geq 2$ was imposed in Genon-Catalot et al. (1992), where σ_t^2 was assumed to be a deterministic function of time. We, however, only require $m \geq 0$ in order to include standard stochastic volatility models. In particular, for any model driven by a Brownian motion, (A.4) holds with $m = 0$ and any $\gamma < 1/2$, cf. Revuz and Yor (1998, Ch. V, Exercise 1.20). It does, however, rule out jumps in the volatility; in Section 8, we discuss how jumps can be allowed for within our framework. The condition that $\int_0^T |\mu_t| \, dt < \infty$ can be substituted for local boundedness of the drift is assumed in, amongst others, Barndorff-Nielsen and Shephard (2006).[3]

Finally, we impose regularity conditions on the kernel function:

Kernel K.1. The kernel $K : \mathbb{R} \mapsto \mathbb{R}$ is continuously differentiable and satisfies:

1. $\int_{\mathbb{R}} K(z) \, dz = 1$; $|z| \left| K^{(i)}(z) \right| \to 0$ as $|z| \to \infty$, $i = 0, 1$; there exists $\Lambda, L < \infty$ such that $\left| K^{(i)}(u) \right| \leq \Lambda_1$ and, for some $\nu > 1$, $\left| K^{(i)}(u) \right| \leq \Lambda_1 \|u\|^{-\nu}$ for $\|u\| \geq L$, $i = 0, 1$.
2. $\int_{\mathbb{R}} z^i K(z) \, dz = 0$, $i = 1, \ldots, r-1$, and $\int_{\mathbb{R}} |z|^r |K(z)| \, dz < \infty$, for some $r \geq 0$.

The assumptions imposed on the kernel are satisfied by most standard kernels for $r \leq 2$. For $r > 2$, K is a so-called higher-order kernel which leads to a further reduction of the bias of the kernel estimator if $m > 2$ as well. On the other hand, $r > 2$ requires $K(z) < 0$ for some values of z, which in turn potentially could lead to $\hat{\sigma}_\tau^2 < 0$ in small samples. Higher-order kernels are traditionally used to reduce the bias in the estimation of functions that are more than twice differentiable. Since in the leading case $m = 0$, one could argue that higher-order kernels are not of any use here. However, as demonstrated in, e.g., Cline and Hart (1991), higher-order kernels can potentially reduce bias even when the object of interest is nonsmooth and has kinks and jumps.

We are now ready to derive the asymptotics of $\widehat{\mathrm{KV}}(\tau)$ and $\hat{\sigma}_\tau^2$. We first derive the asymptotics of the more general estimator $\mathrm{RV}_w(\tau)$ as given in (8) and then apply the results to the kernel estimator. For the moment, suppose $\mu_t = 0$. Then $\Delta X_{t_{i-1}}$ can be written as

$$\Delta X_{t_{i-1}} = \int_{t_{i-1}}^{t_i} \sigma_s \, dW_s \overset{\text{law}}{=} \sqrt{\int_{t_{i-1}}^{t_i} \sigma_s^2 \, ds} \times U_i,$$

where U_i, $i = 1, \ldots, n$, are i.i.d. $N(0, 1)$. Define

$$v_{n,i}(\tau) = w(t_{i-1}, \tau) \int_{t_{i-1}}^{t_i} \sigma_s^2 \, ds \qquad \textbf{(10)}$$

such that

$$\mathrm{RV}_w(\tau) - \sum_{i=1}^n v_{n,i}(\tau) \overset{\text{law}}{=} \sum_{i=1}^n v_{n,i}(\tau) \left(U_i^2 - 1 \right).$$

The right-hand side is a weighted sum of i.i.d. random variables, and we can therefore employ standard limit theorems for triangular arrays of independent variables. We therefore easily obtain as $n \to \infty$,

$$\mathrm{RV}_w(\tau) - \sum_{i=1}^n v_{n,i}(\tau) \to^P 0, \quad \frac{\mathrm{RV}_w(\tau) - \sum_{i=1}^n v_{n,i}(\tau)}{\sqrt{2 \sum_{i=1}^n v_{n,i}^2(\tau)}} \to^d N(0, 1), \qquad \textbf{(11)}$$

by a generalization of results of Barndorff-Nielsen and Shephard (2006). Furthermore, it can be shown that under smoothness conditions on w,

$$\sum_{i=1}^{n} v_{n,i}(\tau) = \int_0^T w(s,\tau)\sigma_s^2\,ds + o(\sqrt{\Delta}),$$

$$\Delta^{-1}\sum_{i=1}^{n} v_{n,i}^2(\tau) = \int_0^T w^2(s,\tau)\sigma_s^4\,ds + o(1). \tag{12}$$

The above results can be extended to allow for $\mu_t \neq 0$ as described in the full proof found in Appendix A.

THEOREM 2. *Assume (A.1)–(A.3) hold and* $(s,\tau) \mapsto w(s,\tau)$ *is continuously differentiable on* $[0,T]\times[0,T]$. *Then* $\mathrm{RV}_w(\tau)$ *given in (8) satisfies*

$$\sqrt{\Delta^{-1}}\left\{\mathrm{RV}_w(\cdot) - \int_0^T w(s,\cdot)\sigma_s^2\,ds\right\} \to^d Z_w(\cdot),$$

where $Z_w(\cdot)$ *is a zero mean Gaussian process with covariance*

$$\mathrm{Cov}(Z_w(\tau_1), Z_w(\tau_2)) = 2\int_0^T w(s,\tau_1)\,w(s,\tau_2)\sigma_s^4\,ds.$$

In particular, for any $\tau \in [0,T]$:

$$\sqrt{\Delta^{-1}}\left\{\mathrm{RV}_w(\tau) - \int_0^T w(s,\tau)\sigma_s^2\,ds\right\} \to^d N\left(0, 2\int_0^T w^2(s,\tau)\sigma_s^4\,ds\right).$$

A consistent estimator of the weighted integrated quarticity, $\mathrm{IQ}_w(\tau) = \int_0^T w^2(s,\tau)\sigma_s^4\,ds$, is given by

$$\widehat{\mathrm{IQ}}_w(\tau) := \frac{n}{3}\sum_{i=1}^{n} w^2(t_{i-1},\tau)\left(\Delta X_{t_{i-1}}\right)^4, \tag{13}$$

cf. Proof of Theorem 2, such that

$$\sqrt{n}\frac{\mathrm{RV}_w(\tau) - \int_0^T w_.(s,\tau)\sigma_s^2\,ds}{\sqrt{2\widehat{\mathrm{IQ}}_w(\tau)}} \to^d N(0,1).$$

The above theorem allows for a number of different weighting schemes, and could potentially be used to derive the asymptotics of local polynomial estimators and rolling window estimators. In this section, we focus on the case where $w(s,\tau) = K_h(s-\tau)$ with $h \to 0$, and then in Section 4 derive asymptotic results for the local linear estimator. First, by Lemma 7, we obtain for any $\tau \in (0,T)$ that the leading bias and variance terms are:

$$\mathrm{E}\left[\hat{\sigma}_\tau^2\right] - \sigma_\tau^2 = \frac{\partial^m \sigma_\tau^2}{\partial \tau^m} L_{\sigma^2}(\tau, 0) \frac{\kappa_q}{m!} h^q + o\left(h^q\right), \tag{14}$$

$$\mathrm{Var}\left[\hat{\sigma}_\tau^2\right] = \frac{2T\sigma_\tau^4}{nh} \left\|K^2\right\|^2 + o(1/(nh)), \tag{15}$$

as $\Delta, h \to 0$, where $\kappa_q = \int_{\mathbb{R}} K(z) z^q dz$, $\left\|K^2\right\|^2 = \int K^2(z)\, dz$ and $q = m + \gamma$. Note that for nondifferentiable trajectories, the bias is of order h^γ, which disappears at a relatively slow rate.

Using the standard bias-variance argument, we conclude that as $h \to 0$ and $nh \to \infty$, $\hat{\sigma}_\tau^2 \to^P \sigma_\tau^2$. To establish asymptotic normality, write:

$$\sqrt{\Delta^{-1}h} \frac{\hat{\sigma}_\tau^2 - \sigma_\tau^2}{\sqrt{2\sigma_\tau^4 \int_{\mathbb{R}} K^2(z)\, dz}} = \frac{\hat{\sigma}_\tau^2 - \sum_{i=1}^n v_{n,i}(\tau)}{\sqrt{2\sum_{i=1}^n v_{n,i}^2(\tau)}} \sqrt{\frac{\Delta^{-1}h \sum_{i=1}^n v_{n,i}^2(\tau)}{\sigma_\tau^4 \int_{\mathbb{R}} K^2(z)\, dz}}$$

$$+ \sqrt{\Delta^{-1}h} \frac{\sum_{i=1}^n v_{n,i}(\tau) - \sigma_\tau^2}{\sqrt{2\sigma_\tau^4 \int_{\mathbb{R}} K^2(z)\, dz}}.$$

By the same arguments as for fixed $h > 0$, the first term converges weakly toward $N(0,1)$ as $h \to 0$ and $nh \to \infty$, while the bias term goes to zero as $nh^{2(m+\gamma)+1} \to 0$.

All of the above results only go through for $\tau \in (0, T)$ for general kernels satisfying (K.1). In order to obtain results for $\tau = 0$ and T, we need to use either boundary kernels or local polynomial estimators. This is discussed in further detail in Section 4.

THEOREM 3. *Assume that (A.1)–(A.4) hold and K satisfies (K.1) with $r \geq m + \gamma$. Then, for any $a \to 0$ with $a/h \to 0$,*

$$\sup_{\tau \in [a, T-a]} \left|\hat{\sigma}_\tau^2 - \sigma_\tau^2\right| = O_P\left(h^{m+\gamma}\right) + O_P\left(\log(n)/\sqrt{nh}\right).$$

As $nh \to \infty$ and $nh^{2(m+\gamma)+1} \to 0$,

$$\sqrt{\Delta^{-1}h}\left\{\hat{\sigma}_\tau^2 - \sigma_\tau^2\right\} \to^d N\left(0, 2\sigma_\tau^4 \int_{\mathbb{R}} K^2(z)\, dz\right)$$

for any $\tau \in (0, T)$ with asymptotic independence across distinct points.

A simple way of estimating the unknown component in the variance, σ_τ^4, is by $\left(\hat{\sigma}_\tau^2\right)^2$. Alternative, one can use the weighted realized quarticity estimator,

$$\frac{\Delta}{3} \sum_{i=1}^n K_h\left(t_{i-1} - \tau\right)\left(\Delta X_{t_{i-1}}\right)^4 \to^P \sigma_\tau^4, \quad \text{as } h \to 0 \text{ and } nh \to \infty.$$

Observe that for a given level of smoothness of σ_t^2, the highest attainable pointwise rate of convergence is $O_P\left(n^{-q/(2q+1)}\right)$ when the bandwidth is chosen as

$h = O\left(n^{-1/(2q+1)}\right)$, where $q = m + \gamma$. As a special case, we recover Foster and Nelson (1996, Thm. 2). As pointed out earlier, the rolling window estimator $\bar{\sigma}_\tau^2 = \mathrm{RV}_w(\tau)$, where $w(t,\tau) = w_n(t_{i-1} - \tau)$ satisfies $\Delta \sum_{i=1}^n w_n(t_{i-1} - \tau) = 1$ and other restrictions, can in general be written as a kernel estimator. Foster and Nelson (1996) show that $\Delta^{-1/4}\left(\bar{\sigma}_\tau^2 - \sigma_\tau^2\right) \to^d N\left(0, 2\sigma_\tau^4 S_{ww}\right)$, where $S_{ww} = \lim_{\Delta \to 0} \Delta \sum_i w^2(t_{i-1}) = \int K^2(z)\,dz$. In the case $m = 0$ and $\gamma = 1/2$, we obtain the same result for our estimator when $h = O\left(n^{-1/2}\right)$. Foster and Nelson (1996) also discuss optimal choices of the weights; in Section 6, we derive some optimal kernels and compare these with their results.

In the case where $\sigma_t^2 = \sigma^2(X_t)$, Bandi and Phillips (2003, Thm. 6) consider the following spatial kernel estimator,

$$\tilde{\sigma}_\tau^2 = \frac{\sum_{i=1}^n \mathcal{K}_b\left(X_{t_{i-1}} - X_\tau\right)\left(\Delta X_{t_{i-1}}\right)^2}{\Delta \sum_{i=1}^n \mathcal{K}_b\left(X_{t_{i-1}} - X_\tau\right)},$$

where $\mathcal{K}$ is a kernel and $b > 0$ a bandwidth (potentially different from K and h). Under regularity conditions, it satisfies

$$\sqrt{\Delta^{-1} b}\left\{\tilde{\sigma}_\tau^2 - \sigma_\tau^2\right\} \to^d N\left(0, \frac{2\sigma_\tau^4 \int_{\mathbb{R}} \mathcal{K}^2(z)\,dz}{\bar{L}_T(T, X_\tau)}\right),$$

where $\bar{L}_T(T,x)$ is the chronological local time of $\{X_t\}$, a random measure of the time that the process has spent in the vicinity of x over $[0,T]$. When $\mathcal{K} = K$ and $b = h$, the factor $\bar{L}_T(T, X_\tau)$ determines whether the spatial estimator is more (> 1) or less (< 1) precise than our time domain estimator. However, since (A.4) only holds with $m = 0$ in the case where $\sigma_t^2 = \sigma^2(X_t)$, then the convergence rate of $\tilde{\sigma}_\tau^2$ is in general faster under sufficient smoothness conditions on the function $\sigma^2(\cdot)$ since Bandi and Phillips (2003) can allow for $b \to 0$ at a slower rate.

In the case where $\sigma_t^2 = \sigma^2(f_t)$ for a set of observed factors $f_t \in \mathbb{R}^d$, the kernel estimator of Bandi and Phillips (2003, Thm. 6) has been extended by Fan et al. (2007). In this case, the convergence rate of $\tilde{\sigma}_\tau^2 = \tilde{\sigma}^2(f_\tau)$ is of order $\Delta^{-1} b^d$ and is potentially slower than $\Delta^{-1} h$. This lead Fan et al. (2007) to combine $\tilde{\sigma}_\tau^2$ with the rolling-window estimator of Foster and Nelson (1996) into a more precise volatility estimator.

Finally, we consider an estimator of the generalized integrated volatility measure given in (5). Compared to the standard realized volatility measure, the above estimator carries an additional bias term due to the kernel smoothing inherent in $\hat{\sigma}_\tau^2$ for $\tau \in (0,T)$. At the end points of the sample, there is an additional boundary bias that we control for by introducing a trimming parameter $a > 0$. The resulting estimator is then defined as

$$\widehat{\mathrm{IV}}_g = \int_a^{T-a} g\left(s, \hat{\sigma}_s^2\right) ds.$$

By letting $a, h \to 0$ as $n \to \infty$, the two bias terms vanish asymptotically.

THEOREM 4. *Assume that (A.1)–(A.4) hold and K satisfies (K.1) with $r \geq m+\gamma$. Let $g(t,x)$ be continuous in t and twice continuously differentiable in x. As $nh^{2(m+\gamma)} \to 0$, $a/h \to 0$, and $a/\sqrt{\Delta} \to 0$,*

$$\sqrt{\Delta^{-1}}\left\{\widehat{\mathrm{IV}}_g - \mathrm{IV}_g\right\} \to^d N\left(0, 2\int_0^T \left[\frac{\partial g\left(s,\sigma_s^2\right)}{\partial \sigma_s^2}\right]^2 \sigma_s^4 ds\right).$$

4. BOUNDARY EFFECTS AND LOCAL LINEAR ESTIMATORS

As demonstrated in Section 3, estimation of σ_τ^2 for τ the interior of $[0,T]$ can in principle be done using standard symmetric kernels. However, it may be of interest to obtain estimates near or at the boundaries of the interval. In particular, the point $\tau = T$ may be important, since this will yield an estimate of the most recent realized volatility. Using a standard symmetric kernel to estimate σ_τ^2 at $\tau = T$ will lead to $E\left[\hat{\sigma}_T^2\right] = \frac{1}{2}\sigma_T^2 + o(1)$ as $h \to 0$; this can be shown by, for example, following Müller (1991). This is caused by so-called boundary or edge effects, as is well known in the kernel estimation literature: For a given bandwidth, the symmetric kernel assigns weight outside the support of the data that causes distortion. A number of different solutions to this problem have been suggested; see Zhang and Karunamuni (1998) for an overview. We here focus on two specific solutions in turn: local linear kernel regression methods, and asymmetric kernels.

We first consider the local linear volatility estimator as given in (8)–(9), since it is known to adapt automatically to the boundaries, thereby not suffering from any boundary biases. Let $\hat{\sigma}_{\mathrm{LL},\tau}^2$ denote the local linear estimator. For any fixed $\tau \in (0,T)$, it can be represented as:

$$\tau \in (0,T) : \hat{\sigma}_{\mathrm{LL},\tau}^2 = \sum_{i=1}^n \frac{1}{h} K^*\left(\frac{t_{i-1}-\tau}{h}\right)\left[\left(\Delta X_{t_{i-1}}\right)^2 + o(1)\right].$$

At the vicinity of the two boundaries, $\hat{\sigma}_\tau^{\mathrm{LL}}$ has the following representation for any constant $c \geq 0$:

$$\tau = ch \quad \text{and} \quad \tau = T - ch : \hat{\sigma}_{\mathrm{LL},\tau}^2 = \sum_{i=1}^n \frac{1}{h} K_c^*\left(\frac{t_{i-1}-\tau}{h}\right)\left[\left(\Delta X_{t_{i-1}}\right)^2 + o(1)\right].$$

Here, K^* and K_c^* denote the so-called equivalent kernels, cf. Fan and Gijbels (1996, Ch. 3), where expressions of these can be found. It now follows from Lemma 7 that under (A.4) the bias of the local linear estimator can be represented as in (14) for $\tau \in [0,T]$ when setting $\kappa_q = \kappa_q^* = \int_{\mathbb{R}} z^q K^*(z)\,dz$ for $\tau \in (0,T)$ and $\kappa_q = \kappa_{c,q}^* = \int_{\mathbb{R}} z^q K_c^*(z)\,dz$ for $\tau = ch$ and $\tau = T - ch$. Moreover, the variance of $\hat{\sigma}_{\mathrm{LL},\tau}^2$ takes the form given in (15) when choosing K as the appropriate equivalent kernel. In conclusion, the local linear estimator is not asymptotically biased at the boundaries in contrast to the NW estimator with a symmetric kernel. By combining these results with Theorem 3, we obtain as a by-product that the local linear estimator has the same pointwise asymptotic distribution as the NW estimator but that the result now holds for any $\tau \in [0, T]$.

An alternative way of removing boundary effects is to use boundary kernels; for an overview of these, we refer to Bouezmarni and Scaillet (2005). We here consider one specific kernel estimator as proposed in Chen (2000) in a different context:

$$\hat{\sigma}^2_{\mathrm{B},\tau} = \frac{\sum_{i=1}^n K(t_{i-1}/T; \tau/T, b)\left(\Delta X_{t_{i-1}}\right)^2}{\Delta \sum_{i=1}^n K(t_{i-1}/T; \tau/T, b)}, \tag{16}$$

where $K(\cdot; y, b)$ is the $\mathrm{Beta}(y/b+1, (1-y)/b+1)$ density, and $b > 0$ is the smoothing parameter. Note that we use b instead of h to denote the smoothing parameter. This kernel adapts to where we are in the domain $[0, T]$ and in particular gives zero weight outside of this interval. By combining the arguments in Chen (2000) with Lemma 6, one can show that

$$E\left[\hat{\sigma}^2_{\mathrm{B},\tau}\right] = \sigma^2_\tau + B_q(\tau) b^{q/2} + o\left(b^{q/2}\right) + O(1/n)$$

uniformly over $\tau \in [0, T]$ for some function $B_q(\tau)$. In the case $m = 2$ and $\gamma = 0$, it takes the form

$$B_2(\tau) = (T - 2\tau)\frac{\partial \sigma^2_\tau}{\partial \tau} + \frac{1}{2}\tau(T-\tau)\frac{\partial^2 \sigma^2_\tau}{\partial^2 \tau}.$$

The variance can be shown to satisfy

$$\mathrm{Var}(\hat{\sigma}^2_{\mathrm{B},\tau}) = \begin{cases} \frac{1}{n\sqrt{b}} V(\tau)\{1 + O(1/n)\}, & \tau \in (0, T) \\ \frac{1}{nb} V_c(\tau)\{1 + O(1/n)\}, & \tau = cb \text{ and } \tau = T - cb \end{cases},$$

where

$$V(\tau) = \frac{\sigma^4_\tau}{\sqrt{\tau(T-\tau)}}\frac{1}{\sqrt{\pi}}, \qquad V_c(\tau) = \sigma^4_\tau \frac{\Gamma(2c+1)}{\sqrt{\pi} 2^{2c} \Gamma^2(c+1)}.$$

So here the bias remains of the same order uniformly over $[0, T]$, while the variance is of larger order, $1/(nb)$, near the boundaries compared to the interior, $1/(n\sqrt{b})$. But the estimator remains consistent for all values of $\tau \in [0, T]$ by suitable choices of $b \to 0$.

5. A NONPARAMETRIC DRIFT ESTIMATOR

As mentioned earlier, the drift and diffusion term can be interpreted as the instantaneous conditional mean and variance. In a standard regression framework, an estimator of the conditional mean is first obtained and the variance is then estimated from the residuals; see, e.g., Mikosch and Stărica (2005). In contrast, in the diffusion setting considered here, one does not need to take into account the presence of μ_t when estimating σ_t^2 as demonstrated in the previous section. But one might try to control for the presence of μ_t to improve on the (finite sample) performance of our estimator of σ_t^2; this idea is, for example, considered in Foster and Nelson (1996, Eq. 5) and Andreou and Ghysels (2002, Eq. 1.2). Additionally, one might be interested in μ_t itself; for example, in bond pricing (see, e.g., Fan et al., 2003). However, as demonstrated in the following, one cannot estimate μ_t consistently without imposing further structure on it.

We start out by considering an estimator of the integrated mean (or drift),

$$\mathrm{IM}(\tau) = \int_0^T K_h(s-\tau)\,\mu_s\,ds.$$

A natural choice for this is the corresponding "realized mean" estimator,

$$\mathrm{RM}(\tau) = \sum_{i=1}^{n} K_h(t_{i-1}-\tau)\,\Delta X_{t_{i-1}}.$$

When we consider the case $h \to 0$, we will write

$$\hat{\mu}_\tau = \sum_{i=1}^{n} K_h(t_{i-1}-\tau)\,\Delta X_{t_{i-1}}.$$

THEOREM 5. *Under (A.1)–(A.3) and (K.1),*

$$\sup_{\tau\in[0,T]} |\mathrm{RM}(\tau) - \mathrm{IM}(\tau)| = O_P(1).$$

and

$$\mathrm{RM}(\tau) - \mathrm{IM}(\tau) \to^d N\left(0, \int_0^T K_h^2(s-\tau)\,\sigma_s^2\,ds\right).$$

If additionally $\mu_\tau \in C^{m,\gamma}[0,T]$, then for any $\varepsilon > 0$,

$$\sup_{\tau\in[\varepsilon,T-\varepsilon]} \left|\hat{\mu}_\tau - \mu_\tau\right| = o_P\left(h^{m+\gamma}\right) + O_P(1/h),$$

and as $h \to 0$,

$$\sqrt{h}\left\{\hat{\mu}_\tau - \mu_\tau\right\} \to^d N\left(0, \sigma_\tau^2 \int_{\mathbb{R}} K^2(z)\,d\tau\right).$$

So our integrated drift estimator is inconsistent: The estimator is unbiased but the variance does not decrease with sample size. This result carries through to the filtered estimator of the instantaneous drift: As $h \to 0$, the bias disappears, but the variance term diverges, which prevents us from getting a consistent estimator of the instantaneous drift estimator.

The inconsistency result of drift estimators given a fixed time span was already noted by Merton (1980) for the case of a geometric Brownian motion; see also Bertsimas, Kogan, and Lo (1996). Similar findings are established in Bandi and Phillips (2003): Their spatial kernel estimator of the drift term for Markov processes only converges as $T \to \infty$ and at a slower rate than the diffusion estimator. However, while the spatial estimator is consistent as $T \to \infty$, an increasing time span will not change the asymptotic properties of our time domain estimator. The failure of recovering the drift in our setting owes to the fact that the observed process only visits a given point in the time domain once whether the time span grows or not. The observations around a given point in time carries enough information to extract the diffusion term, but not the drift term. On the other hand, if the process is recurrent, it visits any given point in the spatial domain infinitely often as time goes to infinity, which allows Bandi and Phillips (2003) to recover the drift by smoothing over the spatial domain.

The results obtained in the previous section for the boundary $\tau = T$ can easily be adapted to the drift estimator.

6. CHOICE OF BANDWIDTH AND KERNEL

One of the main drawbacks of the instantaneous volatility estimator, compared to the integrated volatility one, is its dependence on the bandwidth h. The bandwidth can be regarded as a nuisance parameter that must be chosen by the econometrician. In Sections 3–4, we derived a set of permissible bandwidth sequences yielding consistency and asymptotic normality of the filtered volatility process. These are asymptotic results, however, and do not give much guidance in choosing the bandwidth for a given finite sample. This problem is equivalent to the lag length choice in the rolling window estimators considered in Foster and Nelson (1996) and Andreou and Ghysels (2002). In particular, too large a bandwidth will yield a dominating bias term, while too small a bandwidth choice will lead to an excessive variance of the estimator. So in practice, great care has to be shown when choosing the bandwidth, and data-driven methods for doing so will be useful. We here first derive the optimal bandwidth choice in terms of the mean square error (MSE) criterion, which in turn allows us to obtain operational devices for the bandwidth choice. We also propose a data-driven cross-validation method.

Combining the bias and variance expressions given in (14)–(15), the approximate pointwise MSE can be written as

$$\mathrm{MSE}(h,\tau) \approx h^{2q}\kappa_q^2 B_{m,\tau} + \frac{2T}{nh}\sigma_\tau^4 \left\| K^2 \right\|^2,$$

where $B_{m,\tau} = \frac{\partial^m \sigma_\tau^2}{\partial \tau^m} L_{\sigma^2}^2(\tau,0)/m!$. This is minimized by the following pointwise bandwidth choice,

$$h_{\text{opt},\tau} = \left(\frac{\sigma_\tau^4 T \left\| K^2 \right\|^2}{q \kappa_q^2 B_{m,\tau}} \right)^{1/(2q+1)} n^{-1/(2q+1)},$$

which yields the optimal convergence rate $\text{MSE}\left(h_{\text{opt},\tau}, \tau\right) = O\left(n^{-2q/(2q+1)}\right)$. Similarly, the integrated MSE, $\int_0^T \text{MSE}(h,s)\, ds = O\left(n^{-2q/(2q+1)}\right)$ when the bandwidth is chosen globally as

$$h_{\text{opt}} = \left(\frac{T \int_0^T \sigma_s^4 \, ds \left\| K^2 \right\|^2}{q \kappa_q^2 \int_0^T B_{m,s} \, ds} \right)^{1/(2q+1)} n^{-1/(2q+1)}. \tag{17}$$

As with standard kernel methods, the optimal bandwidth here depends on unknown quantities that need to be estimated in order to make the bandwidth choice operational. In the case of the optimal bandwidth in terms of the IMSE, $\int_0^T \sigma_s^4 \, ds$ can be estimated using the realized quarticity in (13) with a uniform kernel and $\tau = T$, $\Delta/3 \sum_{i=1}^n \left(\Delta X_{t_{i-1}}\right)^4 \to^P \int_0^T \sigma_s^4 \, ds$.

When $m = 0$, the smoothness parameter γ can be estimated using the method proposed in Blanke (2002), while it appears difficult to obtain an estimator of $L_{\sigma^2}(\tau, 0)$. For $m > 0$, a simple estimator of $\partial^m \sigma_t^2 / \partial t^m$ is given by

$$\frac{\partial^m \hat{\sigma}_t^2}{\partial t^m} = \frac{1}{h^{m+1}} \sum_{i=1}^n K^{(m)}\left(\frac{t_{i-1} - t}{h} \right) \left(\Delta X_{t_{i-1}}\right)^2,$$

which in turn can be used to calculate $\int_0^T \left(\partial^m \hat{\sigma}_s^2 / \partial s^m\right)^2 ds$. Unfortunately, $\partial^m \hat{\sigma}_s^2 / \partial s^m$ depends on the bandwidth h itself. One way of solving the problem is to use an approximate parametric model for $\left\{\sigma_t^2\right\}$. For example, $d\sigma_t^2 = \alpha \sigma_t^2 \, dt$, in which case we can estimate σ_0^2 and α as

$$\begin{bmatrix} \log \hat{\sigma}_0^2 \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} 1 & \tau_1 \\ 1 & \tau_2 \end{bmatrix}^{-1} \begin{bmatrix} \log \widehat{[X]}_{\tau_1} \\ \log \widehat{[X]}_{\tau_2} \end{bmatrix}.$$

We then have $\partial^m \hat{\sigma}_s^2 / \partial s^m = \hat{\alpha}^m \hat{\sigma}_0^2 \exp\left[\hat{\alpha} s\right]$. One can obviously run an iterative procedure, where the previous bandwidth choice is plugged into the right-hand side of (17), and a new bandwidth is obtained.

Another way of estimating the optimal bandwidth h_{opt} is by cross-validation which is a data-driven selection method. Consider the integrated square error $\text{ISE}(h) = \int_{T_l}^{T_u} \left[\sigma_s^2 - \hat{\sigma}_s^2\right]^2 ds$ for $0 \le T_l < T_u \le T$. If a symmetric kernel is used, one should normally choose $0 < T_l < T_u < T$ to avoid boundary effects. Noting that $\left(\Delta X_{t_i}\right)^2 / \Delta \to^P \sigma_{t_i}^2$, we can use the following estimator of $\text{ISE}(h)$:

$$\widehat{\text{ISE}}(h) = \sum_{i=1}^n \mathbb{I}\{T_l \le t_{i-1} \le T_u\} \left[\left(\Delta X_{t_i}\right)^2 / \Delta - \hat{\sigma}_{-i,t_i}^2 \right]. \tag{18}$$

We then define the cross-validated bandwidth as $h_{\mathrm{cv}} = \arg\min_{h>0} \widehat{\mathrm{ISE}}(h)$. One should be able to show that this converges toward h_{opt} by following the arguments in, for example, Härdle, Hall, and Marron (1988) and van Es (1992).

Next, we discuss how to choose the kernel K in order to minimize the integrated MSE. For the case $m = 2$, standard results of the literature apply: The optimal symmetric kernel is the so-called Epanechnikov (1969) kernel, but this will however have problems at the boundaries, as demonstrated in Section 4. Instead, one may want to consider one-sided kernels. Within the class of backward-looking kernels, Zhang and Karunamuni (1998) show that the following kernel minimizes MSE when the kernel is restricted to have only one sign change:

$$K(z) = \begin{cases} 6\left(1+3z+2z^2\right), & -1 \le z \le 0 \\ 0, & \text{otherwise} \end{cases}. \tag{19}$$

In the case where $m = 0$, few results exist. However, van Eeden (1986) shows that optimal symmetric kernel in the case where $t \mapsto \sigma_t^2$ has discontinuities is the double exponential density. This is exactly the same result obtained in Nelson and Foster (1996), who show that the optimal two-sided filter is given by the exponential filter presented at the end of Section 2.

Simulation studies have demonstrated that for sufficiently smooth regressions and no boundary problems, the choice of the kernel K has negligible effects on the performance; see, e.g., Silverman (1986, Table 3.1). This result is partially supported by the simulation study carried out in Section 7: When τ is not too close to the boundaries, different choices of K lead to very similar performance. However, when τ is close to either of the two boundaries, boundary kernels perform significantly better.

7. A SIMULATION STUDY

We here examine the performance of the filtered volatility process. In particular, we wish to investigate how it performs relative to the time distance between observations, and how the bandwidth selection rules proposed in the previous section work. We consider the following stochastic volatility model,

$$dX_t = \mu dt + \sigma_t dW_{1,t},$$

$$d\sigma_t^2 = \beta\left(\alpha - \sigma_t^2\right) dt + \rho dW_{1,t} + \kappa\sigma_t\sqrt{1-\rho^2} dW_{2,t},$$

where $W_{1,t}$ and $W_{2,t}$ are standard Brownian motions. This is the continuous-time limit version of the GARCH model as derived in Drost and Werker (1996), and satisfies (A.3)–(A.4). The parameter ρ measures the leverage effect/dependence between X_t and σ_t^2. Initially, we set $\rho = 0$, such that (A.2) is satisfied. The remaining data-generating parameters are chosen to match the estimated parameter values in Andersen and Bollerslev (1998) for the Yen-USD exchange rate. We consider $\Delta^{-1} = 3 \times 60 \times 24$, 60×24, 12×24, and 48 corresponding to sampling every 20 seconds, 1 minute, 5 minutes, and 30 minutes, and set $T = 2$ (48 hours). In order to simulate data from the model, we employ the Euler discretization scheme (see Kloeden and Platten, 1999),

$$\Delta X_{i\delta} = \mu\delta + \sigma_{(i-1)\delta}\sqrt{\delta}\varepsilon_{1,i},$$
$$\Delta\sigma^2_{i\delta} = \beta(\alpha - \sigma^2_{(i-1)\delta})\delta + \rho\sqrt{\delta}\varepsilon_{1,i} + \kappa\sigma_{(i-1)\delta}\sqrt{1-\rho^2}\sqrt{\delta}\varepsilon_{2,i},$$

where $(\varepsilon_{1,i}, \varepsilon_{2,i})$ are i.i.d. normal variables with variance 1. Here, $\delta > 0$ is the length of the discretization step; it is chosen as $\delta = \underline{\Delta}/100$, where $\underline{\Delta}^{-1} = 3 \times 60 \times 24$ corresponds to the highest sampling frequency used in the simulation study.[4] We implement four different estimators of the instantaneous volatility: the Nadaraya-Watson estimator with either (i) the Gaussian kernel or (ii) the one-sided kernel given in (19), and (iii) the asymmetric Beta kernel estimator. For all three estimators, cross-validation was used to choose the bandwidth; this was done by minimizing the criterion function $\widehat{\mathrm{ISE}}(h)$ in (18).

We first only compare the performance over the interval $[1/2, 3/2]$ so we can ignore any boundary biases; this issue is investigated separately below. To evaluate the precision of the volatility estimators, the integrated mean square error $\mathrm{IMSE} = \int_{1/2}^{3/2} \mathrm{E}[(\hat{\sigma}^2_s - \sigma^2_s)^2]ds$ is calculated. We do this using a discrete approximation of the integral. The results based on 400 simulations are reported in Table 1. The Gaussian and the Beta kernel estimator have similar bias, variance, and MSE for high-frequencies, while the Beta estimator is superior for lower ones. The one-sided kernel estimator performs significantly worse for all frequencies, so it is not recommended to use one-sided kernels for estimation in the interior of the sampling interval. As predicted by the theory, the bias and variance increases as the sampling frequency $\Delta^{-1} = T^{-1}n$ shrinks. The Gaussian and Beta kernel estimator still performs reasonably well for 5-minute sampling, while all three estimators are rather imprecise when the sampling frequency drops to 30 minutes.

Next, we investigate the performance of the three kernel estimators near the boundary $T = 2$. In Table 2, the pointwise bias, variance, and MSE at $\tau = 1.50$, 1.90, 1.95, 1.99, and 2.00 are reported when $\Delta^{-1} = 60 \times 24$. As expected, the performance of the Gaussian kernel estimator quickly deteriorates as we get nearer

TABLE 1. Performance in interior

	Gaussian kernel			One-sided kernel			Beta kernel		
$1/\Delta$	Bias^2	Var.	MSE	Bias^2	Var.	MSE	Bias^2	Var.	MSE
$3 \times 60 \times 24$	0.33	0.32	0.65	0.59	0.98	1.57	0.44	0.27	0.71
60×24	0.35	1.30	1.65	1.04	2.72	4.11	0.60	0.74	1.33
12×24	0.85	6.12	6.97	2.87	12.74	15.60	1.08	2.60	3.68
48	3.00	41.82	44.83	8.15	95.27	103.41	2.06	13.40	15.45

Note: Integrated sq. bias ($\times 10^{-4}$), variance ($\times 10^{-4}$), and MSE ($\times 10^{-4}$) of kernel estimators over $\tau \in [1/2, 3/2]$.

TABLE 2. Performance at the boundary

	Gaussian kernel			One-sided kernel			Beta kernel		
τ	Bias^2	Var.	MSE	Bias^2	Var.	MSE	Bias^2	Var.	MSE
1.50	0.01	4.16	4.17	0.04	1.97	2.01	0.03	1.19	1.22
1.90	0.23	3.61	3.84	0.02	2.02	2.04	0.05	1.07	1.12
1.95	3.18	6.30	9.48	0.03	2.20	2.23	0.03	1.47	1.50
1.99	34.34	6.52	40.86	0.09	2.68	2.77	0.07	2.15	2.22
2.00	59.79	2.26	62.05	0.11	2.62	2.73	0.06	2.49	2.55

Note: Pointwise sq. bias ($\times 10^{-3}$), variance ($\times 10^{-3}$), and MSE ($\times 10^{-3}$) of kernel estimators with $\Delta^{-1} = 60 \times 24$.

the boundary, while the one-sided kernel estimator is fairly stable but suffers consistently from a higher variance. The Beta kernel estimator experiences a small increase in bias as we get closer to the boundary, but is superior in terms of variance, and therefore has the smallest MSE of the three estimators. Similar results were found at other sampling frequencies.

The findings reported above are illustrated in Figures 1–3. We here simulated one trajectory of $\{\sigma_t^2\}$ and kept this fixed. We then drew 400 samples of $\{X_t\}$

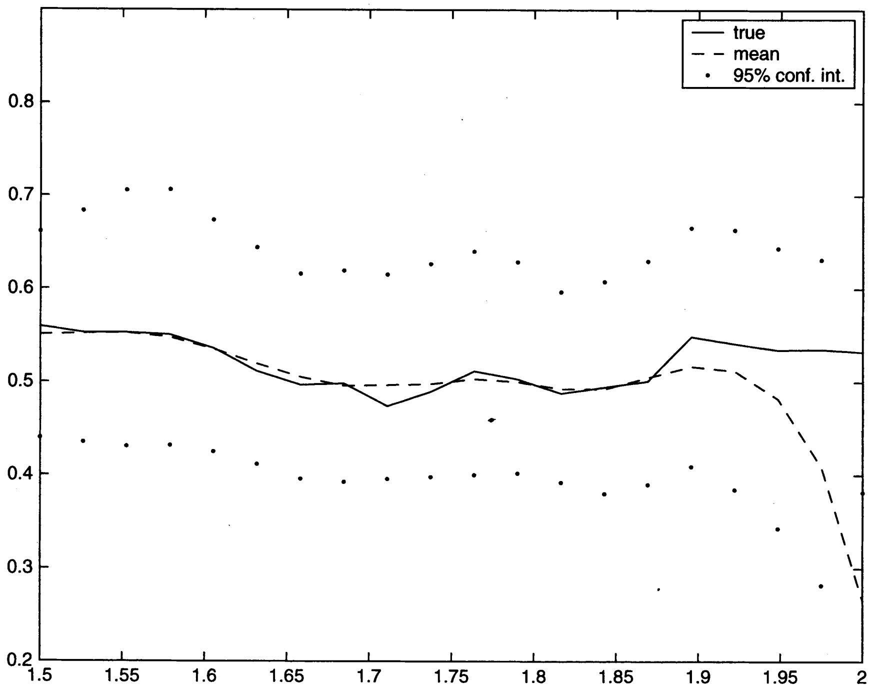

FIGURE 1. Gaussian kernel estimator.

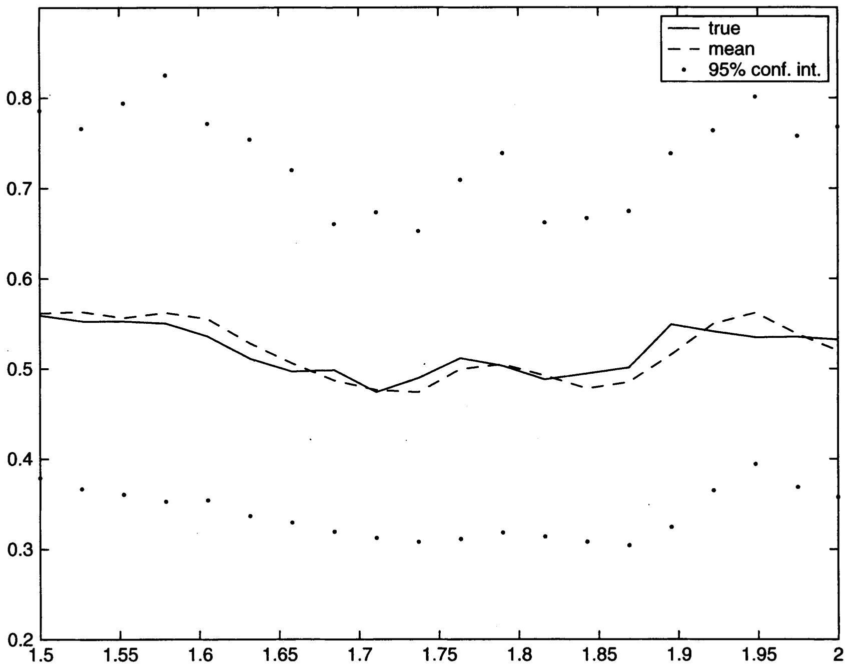

FIGURE 2. One-sided kernel estimator.

using this specific volatility trajectory and so analyzed the behavior of $\{\hat{\sigma}_t^2\}$ conditional on $\{\sigma_t^2\}$. In Figures 1–3, the specific trajectory is plotted together with the mean and 95%-confidence bands of $\hat{\sigma}_\tau^2$ using the 3 different kernels over the interval $\tau \in [1.5, 2.0]$: the Gaussian, the Beta, and the uniform one-sided one. The plots support the results reported in Tables 1 and 2: While the Gaussian kernel estimator performs well in the interior, as we get closer to the boundary, its performance quickly deteriorates. In contrast, the one-sided and the Beta kernel estimator are both fairly unaffected by boundary effects. We ran this last set of simulations conditional on other volatility trajectories and obtained results similar to those reported here.

Finally, we consider the situation where $\rho \neq 0$ in order to investigate how robust our estimator is to leverage effects. In Table 3, the performance of the kernel estimator using the Beta kernel on data sampled at frequency $\Delta^{-1} = 60 \times 24$ is reported for different values of ρ. There is a substantial deterioration of the precision as the leverage effect becomes more pronounced, but for weak leverage effects the estimator performs reasonably well. Similar results were obtained for the other kernels and at other sampling frequencies, and we therefore do not report these results here.

The overall conclusion is that the Beta kernel estimator is superior to the Gaussian and the one-sided one in terms of MSE, performing equally well in the

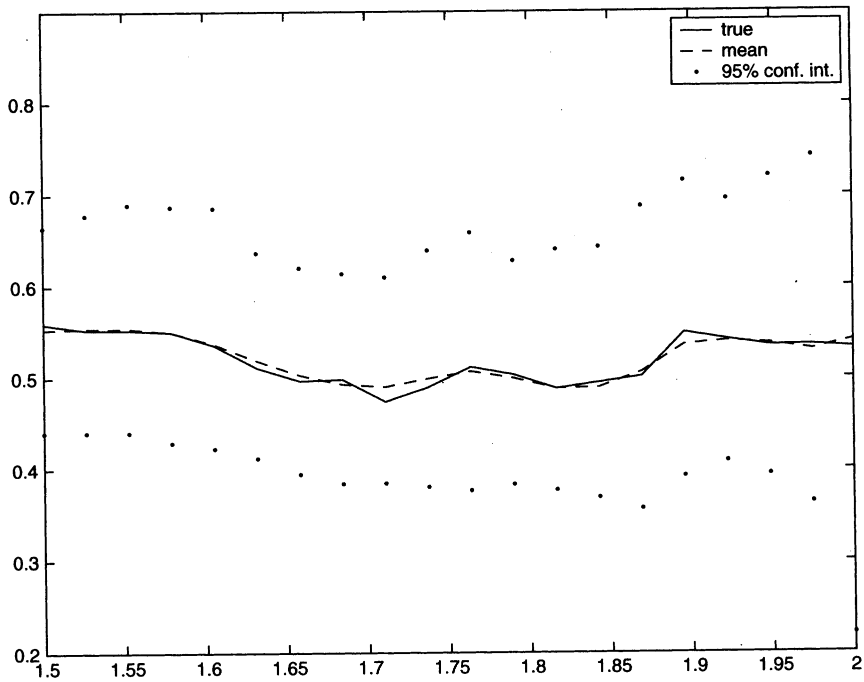

FIGURE 3. Beta kernel estimator.

interior and at the boundary of the sampling interval. The estimators are not very robust toward leverage effects.

8. CONCLUDING REMARKS

We have here proposed a kernel-smoothed version of the standard realized volatility estimator that allows us to filter the unobserved volatility process. Several extensions of the kernel estimator offer themselves, which we discuss below in turn. The derivation of their theoretical properties are left for future research.

TABLE 3. Performance of beta kernel estimator with leverage effects

ρ	−1.00	−0.80	−0.60	−0.40	−0.20	0.00
Bias^2	0.0023	0.0016	0.0010	4.80×10^{-4}	1.20×10^{-4}	0.60×10^{-4}
Var.	0.2626	0.1690	0.0956	0.0426	0.0109	0.74×10^{-4}
MSE	0.2649	0.1706	0.0966	0.0431	0.0110	1.33×10^{-4}

Note: Integrated sq. bias, variance, and MSE over $\tau \in [1/2, 3/2]$ with $\Delta^{-1} = 60 \times 24$.

Kernel-Weighted Power Variation. An interesting extension of the proposed estimator would be to consider estimators of the form

$$v_\tau^{[r]} = \sum_{i=1}^{n} K_h(t_{i-1} - \tau) \left|\Delta X_{t_{i-1}}\right|^r, \quad r > 0,$$

which includes as a special case the spot volatility estimator considered here ($r = 2$). We conjecture that by extending the results of, e.g., Barndorff-Nielsen and Shephard (2006) to include weights, we can obtain asymptotic results for the above estimator.

Presence of Jumps. Another important extension would be to allow for jumps in X_t,

$$X_t = \int_0^t \mu_s\, ds + \int_0^t \sigma_s\, dW_s + \sum_{j=1}^{N_t} C_j, \tag{20}$$

where $\{N_t\}$ is a counting process and $C_1, C_2, \ldots$ are associated jumps. To estimate σ_τ^2 in this setting, one could use kernel-weighted bipower variation:

$$\hat{\sigma}_\tau^2 = \kappa_1^{-1} \sum_{i=1}^{n} K_h(t_{i-1} - \tau) \left|\Delta X_{t_i}\right| \left|\Delta X_{t_{i-1}}\right|, \tag{21}$$

with $\kappa_1 = E[|U|] = \sqrt{2/\pi}$, $U \sim N(0,1)$. The motivation for this estimator is the same as without jumps: We know that $\kappa_1^{-1} \sum_{i=2}^{n} \left|\Delta X_{t_i}\right| \left|\Delta X_{t_{i-1}}\right| \to^P \int_0^T \sigma_s^2 ds$, cf. Barndorff-Nielsen and Shephard (2004b), so we would therefore expect that $\hat{\sigma}_\tau^2$ defined in (21) satisfies $\hat{\sigma}_\tau^2 \to^P \int_0^T K_h(s-\tau)\sigma_s^2\, ds$ for fixed $h > 0$, and we can now proceed as we did without jumps.

If there is a jump component (with finite activity) in $\{\sigma_t^2\}$, the trajectories will only be piece-wise continuous and so the general theory developed in Sections 3–4 is only valid at the points of continuity. To estimate σ_τ^2 at points of discontinuity, we here follow a strategy similar to that of Gijbels et al. (2007). We introduce the following two estimators:

$$\hat{\sigma}_{1,\tau}^2 = \sum_{i=1}^{n} K_{1,h}(t_{i-1} - \tau)\left(\Delta X_{t_{i-1}}\right)^2, \quad \hat{\sigma}_{2,\tau}^2 = \sum_{i=1}^{n} K_{2,h}(t_{i-1} - \tau)\left(\Delta X_{t_{i-1}}\right)^2,$$

where the two kernels, $K_1(z)$ and $K_2(z)$, have support $[-1, 0]$ and $[0, 1]$, respectively. For example, $K_1(z) = \mathbb{I}\{-1 \le z \le 0\}$ and $K_1(z) = \mathbb{I}\{0 \le z \le 1\}$. Thus, the two estimators make use of a backward and forward-looking filter/kernel, respectively. Clearly, they are both consistent estimators of σ_τ^2 at any continuity point $\tau \in (0, T)$, i.e., where no jump has occurred. Now, let τ_0 denote a point in time where a jump occurs of size J_{τ_0}. Then $\sigma_{\tau_0+}^2 = \sigma_{\tau_0-}^2 + J_{\tau_0}$, and we obtain $\hat{\sigma}_{1,\tau_0}^2 \to^P \sigma_{\tau_0-}^2$ and $\hat{\sigma}_{2,\tau_0}^2 \to^P \sigma_{\tau_0+}^2$. In order to discriminate between continuity points and jump points, we introduce the local residual sum of squares for each of the two estimators, $RSS_k(\tau)$ for $k = 1, 2$, and then define our final estimator as

$$\hat{\sigma}_\tau^2 = \begin{cases} \hat{\sigma}_{1,\tau_0}^2, & RSS_1(\tau) < RSS_2(\tau) \\ \frac{1}{2}(\hat{\sigma}_{1,\tau_0}^2 + \hat{\sigma}_{2,\tau_0}^2), & RSS_1(\tau) = RSS_2(\tau) \\ \hat{\sigma}_{2,\tau_0}^2, & RSS_1(\tau) > RSS_2(\tau) \end{cases}. \quad (22)$$

Also note that at the same time we get estimates of the jump sizes, $\hat{J}_\tau := \hat{\sigma}_{2,\tau}^2 - \hat{\sigma}_{1,\tau}^2$ for all τ. The two estimators in (21)–(22) can easily be combined to allow for both jumps in $\{X_t\}$ and $\{\sigma_t^2\}$.

Multivariate Models. Suppose that $X_t \in \mathbb{R}^d$ solves a multivariate version of (1) where $\mu_t, W_t \in \mathbb{R}^d$ and $\sigma_t \in \mathbb{R}^{d\times d}$. We can then follow Barndorff-Nielsen and Shephard (2004a), and define

$$\hat{\sigma}_\tau^2 = \sum_{i=1}^{n} K_h(t_{i-1} - \tau)\,\Delta X_{t_i} \Delta X_{t_i}' \in \mathbb{R}^{d\times d}. \quad (23)$$

Again, we expect that the results of Barndorff-Nielsen and Shephard (2004a) can be extended to show that $\hat{\sigma}_\tau^2 \to^P \sigma_\tau^2$.

Market Microstructure Noise. Suppose that X_t solves the univariate version of (1), but instead of X_t we only observe

$$Y_{t_i} = X_{t_i} + \varepsilon_i,$$

where ε_i are unobserved i.i.d. $(0, \sigma_\varepsilon^2)$ errors. In this case, $\hat{\sigma}_\tau^2 = \sum_{i=1}^n K_h(t_{i-1} - \tau) (\Delta Y_{t_i})^2$ will be biased, $\mathrm{E}\left[\hat{\sigma}_\tau^2\right] \approx \sigma_\tau^2 + 2n\sigma_\varepsilon^2$ as $\Delta, h \to 0$. However, as above, we can take already existing estimators of the integrated volatility that are robust to market microstructure noise and introduce kernel smoothing in those estimators to obtain a spot volatility estimator. As an example, we could take the estimator proposed in Barndorff-Nielsen, Hansen, Lund, and Shephard (2008) and include kernel smoothing:

$$\hat{\sigma}_\tau^2 = \hat{\sigma}_\tau^2(0) + \sum_{m=1}^{M} w\left(\frac{m-1}{M}\right)\left\{\hat{\sigma}_\tau^2(m) + \hat{\sigma}_\tau^2(-m)\right\},$$

where $w(\cdot)$ is a weighting function and

$$\hat{\sigma}_\tau^2(m) = \sum_{i=m}^{n} K_h(t_{i-m} - \tau)\,\Delta Y_{t_i} \Delta Y_{t_{i-m}}.$$

NOTES

1. Barndorff-Nielsen et al. (2008) also employ kernels in the estimation of integrated volatility, but for a completely different reason, namely, to adjust for measurement errors. Their estimator has very little in common with the one proposed here.
2. No distributional theory is derived though.
3. I thank an anonymous referee for pointing this out.
4. As a referee pointed out, given that the transition density of the volatility model is known, one can in fact simulate the exact distribution; see Broadie and Kaya (2006).

REFERENCES

Alizadeh, S., M.W. Brandt, & F.X. Diebold (2002) Range-based estimation of stochastic volatility models. *Journal of Finance* 57, 1047–1092.

Andersen, T.G. & T. Bollerslev (1997) Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 4, 115–158.

Andersen, T.G. & T. Bollerslev (1998) Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.

Andersen, T.G., T. Bollerslev, & F.X. Diebold (2009) Parametric and nonparametric measurement of volatility. In Y. Aït-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*, in press. North Holland.

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

Andersen, T.G. & J. Lund (1997) Estimating continuous-time stochastic volatility models of the short term interest rate. *Journal of Econometrics* 77, 343–377.

Andreou, E. & E. Ghysels (2002) Rolling-sample volatility estimators: some new theoretical, simulation and empirical results. *Journal of Business & Economic Statistics* 20, 363–375.

Bandi, F.M. & P.C.B. Phillips (2003) Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 241–283.

Barndorff-Nielsen, O.E. & N. Shephard (2004a) Econometric analysis of realised covariation: High frequency based covariance, regression and correlation in financial economics. *Econometrica* 72, 885–925.

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

Barndorff-Nielsen, O.E. & N. Shephard (2006) Power variation and time change. *Theory of Probability and Its Applications* 50, 1–15.

Barndorff-Nielsen, O.E., S.E. Graversen, J. Jacod, & N. Shephard (2006) Limit theorems for bipower variation in financial econometrics. *Econometric Theory* 22, 677–719.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, & N. Shephard (2008) Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.

Barucci, E. & R. Renò (2002) On measuring volatility of diffusion processes with high frequency data. *Economics Letters* 74, 371–378.

Bertsimas, D., L. Kogan, & A.W. Lo (2000) When is time continuous?. *Journal of Financial Economics 55*, 173–204.

Björk, T. (2004) *Arbitrage Theory in Continuous Time*. Oxford University Press.

Blanke, D. (2002) Estimation of local smoothness coefficients for continuous time processes. *Statistical Inference for Stochastic Processes* 5, 65–93.

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

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

Bouezmarni, T. & O. Scaillet (2005) Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory* 21, 390–412.

Broadie, M. & Ö Kaya (2006) Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research* 54, 217–231.

Chen, S.X. (2000) Beta kernel smoothers for regression curves. *Statistica Sinica* 10, 73–91.

Chib, S., F. Nardari, & N. Shephard (2002) Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108, 281–316.

Cline, D.B.H. & J.D. Hart (1991) Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics* 22, 69–84.

Drost, F.C. & B.J.M. Werker (1996) Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics* 74, 31–57.

Epanechnikov, V.A. (1969) Nonparametric estimates of a multivariate probability density. *Theory of Probability and Applications* 14, 153–158.

Fan, J., Y. Fan, & J. Lv (2007) Aggregation of nonparametric estimators for volatility matrix. *Journal of Financial Econometrics* 5, 321–357.

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

Fan, J., J. Jiang, C. Zhang, & Z. Zhou (2003) Time-dependent diffusion models for term structure dynamics. *Statistica Sinica* 13, 965–992.

Florens-Zmirou, D. (1993) On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability* 30, 790–804.

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

Genon-Catalot, V., C. Laredo, & D. Picard (1992) Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statistics* 19, 317–335.

Gijbels, I., A. Lambert, & P. Qiu (2007) Jump-preserving regression and smoothing using local linear fitting: A compromise. *Annals of the Institute of Statistical Mathematics* 59, 235–272.

Gijbels, I., A. Pope, & M.P. Wand (1999) Understanding exponential smoothing via kernel regression. *Journal of the Royal Statistical Society (B)* 61, 39–50.

Härdle, W., P. Hall, & J.S. Marron (1988) How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association* 83, 86–101.

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

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

Høg, E. & A. Lund (2003) Wavelet Estimation of Integrated Volatility. Manuscript, Aarhus School of Business.

Kanaya, S. & D. Kristensen (2008) Estimation of Stochastic Volatility Models by NonParametric Filtering. Manuscript, Department of Economics, Columbia University.

Kloeden, P.E. & E. Platen (1999) *Numerical Solution of Stochastic Differential Equations*, 2nd ed. Springer-Verlag.

Kristensen, D. (2009) Uniform convergence rates of kernel estimators with heterogenous dependent data. *Econometric Theory* 25, 1428–1440.

Malliavin, P. & M.E. Mancino (2006) A Fourier Transform Method for Nonparametric Estimation of Volatility. Manuscript, Università di Firenze.

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

Mikosch, T. & C. Stărica (2005) Stock Market Risk-return Inference. An Unconditional Non-parametric Approach. Preprint no. 202, Department of Applied Mathematcis and Statistics, University of Copenhagen.

Müller, H.-G. (1991) Smooth optimum kernel estimators near endpoints. *Biometrika* 78, 521–530.

Mykland, P.A. & L. Zhang (2003) Inference for Volatility-Type Objects and Implications for Hedging. Technical report 506, Department of Statistics, Carnegie-Mellon.

Mykland, P.A. & L. Zhang (2006). ANOVA for diffusions and Ito processes. *Annals of Statistics* 34 1931–1963.

Protter, P. (2004) *Stochastic Integration and Differential Equations*, 2nd ed. Springer-Verlag.

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

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

Revuz, D. & M. Yor (1998) *Continuous Martingales and Brownian Motion*, 3rd ed. Springer-Verlag

Shephard, N. (2005) *Stochastic Volatility: Selected Readings*. Oxford University Press.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Stein, E.M. & J.E. Stein (1991) Stock price distributions with stochastic volatility. *Review of Financial Studies* 4, 727–752.

van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer-Verlag.

van Eeden, C. (1985) Mean integrated squared error of kernel estimators when the density and its derivative are not necessarily continuous. *Annals of the Institute of Statistical Mathematics* 37, 461–472

van Es, B. (1992) Asymptotics for least squares cross-validation bandwidths in nonsmooth cases. *Annals of Statistics* 20, 1647–1657.

Zhang, S. & R.J. Karunamuni (1998) On kernel density estimation near endpoints. *Journal of Statistical Planning and Inference* 70, 301–316.

APPENDIX A: Proofs

Proof of Theorem 2. Define

$$\mathrm{RV}_w^*(\tau) = \sum_{i=1}^{n} w\left(t_{i-1}, \tau\right)\left(X_{t_i}^* - X_{t_{i-1}}^*\right)^2,$$

where $\{X_t^*\}$ solves $dX_t^* = \sigma_t dW_t$. Then, with $v_{n,i}(\tau)$ defined in (10),

$$\sqrt{n}\frac{\mathrm{RV}_w^*(\tau) - \int_0^T w(s,\tau)\sigma^2(s)ds}{\sqrt{2\int_0^T w^2(s,\tau)\sigma^4(s)ds}} = \frac{\mathrm{RV}_w^*(\tau) - \sum_{i=1}^n v_{n,i}(\tau)}{\sqrt{2\sum_{i=1}^n v_{n,i}^2(\tau)}}\sqrt{\frac{n\sum_{i=1}^n v_{n,i}^2(\tau)}{\int_0^T w^2(s,\tau)\sigma^4(s)ds}}$$
$$+\sqrt{n}\frac{\sum_{i=1}^n v_{n,i}(\tau) - \int_0^T w(s,\tau)\sigma^2(s)ds}{\sqrt{2\int_0^T w^2(s,\tau)\sigma^4(s)ds}}.$$

We then wish to show that (11) and (12) hold. First, we write

$$\frac{\mathrm{RV}_w^*(\tau) - \sum_{i=1}^n v_{n,i}(\tau)}{\sqrt{2\sum_{i=1}^n v_{n,i}^2(\tau)}} = \sum_{i=1}^n c_{n,i}(\tau)\bar{U}_i^2,$$

where $c_{n,i}(\tau) = v_{n,i}(\tau)/\sqrt{\sum_{i=1}^n v_{n,i}^2(\tau)}$ and $\bar{U}_i^2 = \left(U_i^2 - 1\right)/\sqrt{2}$ are i.i.d. (0, 1). Since $\sum_{i=1}^n c_{n,i}^2(\tau) = 1$, and $\max_{i=1,\dots,n}\left|c_{n,i}(\tau)\right| \to 0$, it follows by Barndorff-Nielsen and

Shephard (2006, Cor. 3.1) that the convergence results in (11) holds. Next, (12) follows by Lemma 6(i)–(ii):

$$\sum_{i=1}^{n} v_{n,i}(\tau) = \sum_{i=1}^{n} w(t_{i-1},\tau) \int_{t_{i-1}}^{t_i} \sigma_s^2 \, ds = \int_0^T w(s,\tau)\sigma_s^2 \, ds + O(\Delta),$$

$$n\sum_{i=1}^{n} v_{n,i}^2(\tau) = \Delta^{-1} \sum_{i=1}^{n} w^2(t_{i-1},\tau) \left(\int_{t_{i-1}}^{t_i} \sigma_s^2 \, ds \right)^2 = \int_0^T w^2(s,\tau)\sigma^4(s)\, ds + o(1).$$

The above convergence result can easily be extended to hold for any collection of points $(\tau_1, \ldots, \tau_m)$: With $\mathrm{IV}_w(\tau) = \int_0^T w(s,\tau)\sigma_s^2 \, ds$,

$$\sqrt{n}\left[\left(\mathrm{RV}_w^*(\tau_1), \ldots, \mathrm{RV}_w^*(\tau_m) \right) - \left(\mathrm{IV}_w(\tau_1), \ldots, \mathrm{IV}_w(\tau_m) \right) \right] \to N_m(0, V_m),$$

where $V_m = \left[V_{m,ij} \right]_{ij} \in \mathbb{R}^{m \times m}$ has elements $V_{m,ij} = 2\int_0^T w(s,\tau_i) w(s,\tau_j)\sigma^4(s)\, ds$. The extension of the weak convergence of the random process $\tau \mapsto \mathrm{RV}_w^*(\tau)$ to take place in the function space $C([0,T])$ equipped with the sup-norm now follows from, for example, van der Vaart and Wellner (1996, Ex. 1.5.10) if we can verify that $\mathrm{RV}_w^*(\tau)$ is stochastically equicontinuous. This follows from the fact that:

$$\begin{aligned} \left| \mathrm{RV}_w^*(\tau) - \mathrm{RV}_w^*(\tau') \right| &\le \sum_{i=1}^{n} \left| w(t_{i-1},\tau) - w(t_{i-1},\tau') \right| \left(X_{t_i}^* - X_{t_{i-1}}^* \right)^2 \\ &\le \left| \tau - \tau' \right| \times \sup_{s,t} \left| w^{(1)}(s,t) \right| \sum_{i=1}^{n} \left(X_{t_i}^* - X_{t_{i-1}}^* \right)^2 \\ &= \left| \tau - \tau' \right| \times O_P(1). \end{aligned}$$

Next, we show that the effect of a nonzero drift term is negligible, $\sqrt{n}\left[\mathrm{RV}_w(\tau) - \mathrm{RV}_w^*(\tau) \right] = o_P(1)$:

$$\begin{aligned} \mathrm{RV}_w(\tau) - \mathrm{RV}_w^*(\tau) &= \sum_{i=1}^{n} w(t_{i-1},\tau) \left(\int_{t_{i-1}}^{t_i} \mu_s \, ds \right)^2 \\ &\quad + 2\sum_{i=1}^{n} w(t_{i-1},\tau) \int_{t_{i-1}}^{t_i} \sigma_s \, dW_s \int_{t_{i-1}}^{t_i} \mu_s \, ds, \end{aligned}$$

where, cf. Lemma 6(ii),

$$\sum_{i=1}^{n} w(t_{i-1},\tau) \left(\int_{t_{i-1}}^{t_i} \mu_s \, ds \right)^2 = \Delta \int_0^T w(s,\tau)\mu_s^2 ds + o(\Delta),$$

while the second term has mean zero and variance

$$\begin{aligned} 4\sum_{i=1}^{n} w^2(t_{i-1},\tau) \int_{t_{i-1}}^{t_i} \sigma_s^2 ds \left(\int_{t_{i-1}}^{t_i} \mu_s \, ds \right)^2 &\le 4\Delta \sup_s \sigma_s^2 \times \sum_{i=1}^{n} w^2(t_{i-1},\tau) \left(\int_{t_{i-1}}^{t_i} \mu_s \, ds \right)^2 \\ &= 4\Delta^2 \sup_s \sigma_s^2 \times \left[\int_0^T w^2(s,\tau)\mu_s^2 \, ds + o(1) \right], \end{aligned}$$

where we again have applied Lemma 6(ii). Finally, the proof of $\widehat{\mathrm{IQ}}_w(\tau) \to^P \mathrm{IQ}_w(\tau)$ follows by the same arguments as above. ■

Proof of Theorem 3. All of the pointwise convergence results in the proof of Theorem 2 still go through with $h \to 0$: It still holds that $\sum_{i=1}^n c_{n,i}(\tau)\bar{U}_i^2 \to^d N(0,1)$ with $w(s,t) = K_h(s-t)$ and $h \to 0$, and by appealing to Lemma 7 instead of 6, we obtain

$$\sqrt{\Delta^{-1}h}\left[\sum_{i=1}^n v_{n,i}(\tau) - \sigma_\tau^2\right] = O\left(\sqrt{\Delta^{-1}h^{2(m+\gamma)+1}}\right) + O\left(1/\sqrt{\Delta^{-1}h}\right),$$

$$\frac{h}{\Delta}\sum_{i=1}^n v_{n,i}^2(\tau) - \sigma_\tau^4\left\|K^2\right\|^2 = O(\Delta^\gamma) + O\left(1/\left(\Delta^{-1}h\right)\right),$$

where both right-hand-side terms go to zero.

For any two distinct points $\tau \neq \tau'$, due to $\mathrm{Cov}\left(\bar{U}_i, \bar{U}_j\right) = 0$, $i \neq j$,

$$\begin{aligned}\mathrm{Cov}\left(\sqrt{nh}\hat{\sigma}_\tau, \sqrt{nh}\hat{\sigma}_{\tau'}\right) &= 2nh\sum_{i=1}^n K_h\left(t_{i-1}-\tau\right)K_h\left(t_{i-1}-\tau'\right)\left(\int_{t_{i-1}}^{t_i}\sigma_s^2\,ds\right)^2 + o(1)\\ &= 2h\int_0^T \sigma_t^4 K_h(s-\tau)K_h\left(s-\tau'\right)ds + o(1)\\ &= \int_{\mathbb{R}} K(z)K\left(z+\frac{\tau-\tau'}{h}\right)\sigma_{\tau+hz}^4\,dz + o(1)\\ &= o(1).\end{aligned}$$

One can now show the asymptotic independence result by the Cramer-Wold device.

To show uniform convergence, we first note that the bias is of order $O\left(h^{m+\gamma}\right)$ uniformly over $\tau \in [a, T-a]$, cf. Lemma 7. We then need to show that the variance component is uniformly of order $O\left(\log(n)/\sqrt{nh}\right)$; this is established by applying the general result of Kristensen (2009, Thm. 1): In his notation, $X_{n,i} = t_{i-1}$ and $Y_{n,i} = \int_{t_{i-1}}^{t_i}\sigma_s^2\,ds \times U_i^2$, and we then proceed to verify his Assumptions A.1–A.5. Kristensen (2009, Assump. A.1) holds with $\beta = +\infty$ since $\left(X_{n,i}, Y_{n,i}\right)$, $i = 1,\ldots,n$, are independent (conditional on $\{\sigma_t\}$). Kristensen (2009, Assump. A.2) is satisfied with $\lambda = 0$ and any $s > 2$ since U_i has all moments. The densities of $X_{n,i}$ and $\left(X_{n,i}, X_{n,i-1}\right)$ are given by $f_{n,i}(x) = \mathbb{I}\{0 \le x \le 1\}$ and $f_{n,ij}(x,y) = \mathbb{I}\{0 \le x \le 1, 0 \le y \le 1\}$, respectively, and $\mathrm{E}\left[\left|Y_{n,i}\right| | X_{n,i} = x\right] = \int_x^{t_i}\sigma_s^2\,ds \times \mathrm{E}\left[U_i^2\right]$, which are both bounded such that Kristensen (2009, Assump. A.4–A.5) hold. Finally, the kernel K is easily seen to satisfy Kristensen (2009, Assump. A.6). ■

Proof of Theorem 4. We have

$$\widehat{\mathrm{IV}}_g - \mathrm{IV}_g = \int_a^{T-a}\left[g\left(s,\hat{\sigma}_s^2\right) - g\left(s,\sigma_s^2\right)\right]ds + \int_0^a g\left(s,\sigma_s^2\right)ds + \int_{T-a}^T g\left(s,\sigma_s^2\right)ds.$$

The first term can be written as

$$\int_a^{T-a}\left[g\left(s,\hat{\sigma}_s^2\right)-g\left(s,\sigma_s^2\right)\right]ds=\int_a^{T-a}\frac{\partial g\left(s,\sigma_t^2\right)}{\partial x}\left[\hat{\sigma}_s^2-\sigma_s^2\right]ds$$
$$+\frac{1}{2}\int_a^{T-a}\frac{\partial^2 g\left(s,\bar{\sigma}_t^2\right)}{\partial x^2}\left|\hat{\sigma}_s^2-\sigma_s^2\right|^2 ds,$$

where $\bar{\sigma}_{ts}^2\in\left[\hat{\sigma}_s^2,\sigma_s^2\right]$. By Theorem 3,

$$\int_a^{T-a}\left|\frac{\partial^2 g\left(t,\bar{\sigma}_s^2\right)}{\partial x^2}\right|\left|\hat{\sigma}_s^2-\sigma_s^2\right|^2 ds\le C\sup_{s\in[a,T-a]}\left|\hat{\sigma}_s^2-\sigma_s^2\right|^2$$
$$=O_P\left(h^{2(m+\gamma)}\right)+O_P\left(\log(n)/(nh)\right),$$

while

$$\int_0^T\frac{\partial g\left(t,\sigma_t^2\right)}{\partial x}\hat{\sigma}_t^2dt=\sum_{i=1}^n\Delta X_{t_{i-1}}^2\int_0^T\frac{\partial g\left(t,\sigma_t^2\right)}{\partial x}K_h\left(t_{i-1}-t\right)dt\,dt$$
$$=\sum_{i=1}^n\Delta X_{t_{i-1}}^2\frac{\partial g\left(t_{i-1},\sigma_{t_{i-1}}^2\right)}{\partial x}+O_P\left(h^{m+\gamma}\right).$$

By Theorem 2, with $w(s,\tau)=\frac{\partial^2 g(s,\sigma_s^2)}{\partial x^2}$, the claimed weak convergence limit is obtained. Finally,

$$\left|\int_0^a g\left(s,\sigma_s^2\right)ds\right|\le a\sup_s\left|g\left(s,\sigma_s^2\right)\right|=O(a)$$

and similarly for the other term. ■

Proof of Theorem 5. Define

$$M_i(\tau)=K_h\left(t_{i-1}-\tau\right)\int_{t_{i-1}}^{t_i}\mu_s\,ds,\qquad V_i(\tau)=K_h\left(t_{i-1}-\tau\right)\sqrt{\int_{t_{i-1}}^{t_i}\sigma_s^2\,ds},$$

such that

$$\frac{\mathrm{RM}(\tau)-\sum_{i=1}^n M_i(\tau)}{\sqrt{\sum_{i=1}^n V_{n,i}^2(\tau)}}=\sum_{i=1}^n c_{n,i}(\tau)U_i,$$

where $c_{n,i}(\tau)=V_{n,i}(\tau)/\sqrt{\sum_{i=1}^n V_{n,i}^2((\tau))}$. We proceed as in the proof of Theorem 2, and obtain from Lemma 6,

$$\sum_{i=1}^n M_i(\tau,h)\to\int_0^T K_h(s-\tau)\mu_s\,ds,\qquad\sum_{i=1}^n V_i^2(\tau,h)\to\int_0^T K_h^2(s-\tau)\sigma_s^2\,ds.$$

The above convergence results still hold as $h \to 0$ under the additional conditions stated by appealing to Lemma 7 instead of 6. ■

APPENDIX B: Lemmas

LEMMA 6. *Assume the following conditions:*

C1. $t \mapsto f(t)$ *satisfies* $\int_0^T |f(s)|ds < \infty$ *and* $\Delta \sum_{i=1}^n \left|f^2(s_i) - f^2(t_i)\right| = o(1)$ *for any sequences* $(i-1)\Delta \le s_i \le t_i \le i\Delta$, $i = 1, \ldots, n$.

C2. $t \mapsto w(t)$ *satisfies* $w \in C_{1,0}$ *such that* $\bar{w}_k \equiv \sup_{0 \le u \le T} \left|w^{(k)}(u)\right| < \infty$, $k = 0, 1$.

Then uniformly over $\tau \in [0, T]$, *as* $\Delta \to 0$*:*

(i) $\displaystyle\sum_{i=1}^n w(t_{i-1}) \int_{t_{i-1}}^{t_i} f(s)\,ds = \int_0^T w(s) f(s)\,ds + O(\Delta) \times \bar{w}_1,$

(ii) $\displaystyle\Delta^{-1} \sum_{i=1}^n w(t_{i-1}) \left(\int_{t_{i-1}}^{t_i} f_s\,ds\right)^2 = \int_0^T w(s) f^2(s)\,ds + o(1) \times \bar{w}_0 + O(\Delta) \times \bar{w}_1.$

Proof. By the mean value theorem, there exists $t_i(s) \in [t_{i-1}, s]$, $i = 1, \ldots n$, such that

$$\sum_{i=1}^n w(t_{i-1}) \int_{t_{i-1}}^{t_i} f(s)\,ds - \int_0^T w(s) f(s)\,ds = \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \left\{w(t_{i-1}) - w(s)\right\} f(s)\,ds$$
$$= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} w^{(1)}(t_i(s)) \left(t_{i-1} - s\right) f(s)\,ds,$$

where

$$\left|\int_{t_{i-1}}^{t_i} w^{(1)}(t_i(s)) \left(t_{i-1} - s\right) f(s)\,ds\right| \le \Delta \int_{t_{i-1}}^{t_i} \left|w^{(1)}(t_i(s))\right| |f(s)|\,ds$$
$$\le \Delta \bar{w}_1 \int_{t_{i-1}}^{t_i} |f(s)|\,ds.$$

To show (ii), first apply the mean value theorem to obtain for $\bar{t}_i, \tilde{t}_i \in [t_{i-1}, t_i]$,

$$\left|\Delta^{-1} \sum_{i=1}^n w(t_{i-1}) \left(\int_{t_{i-1}}^{t_i} f(s)\,ds\right)^2 - \sum_{i=1}^n w(t_{i-1}) \int_{t_{i-1}}^{t_i} f^2(s)\,ds\right|$$
$$\le \sum_{i=1}^n \left|w(t_{i-1})\right| \left|\Delta^{-1} \left(\int_{t_{i-1}}^{t_i} f(s)\,ds\right)^2 - \int_{t_{i-1}}^{t_i} f^2(s)\,ds\right|$$
$$= \Delta \sum_{i=1}^n \left|w(t_{i-1})\right| \left|f^2(\bar{t}_i) - f^2(\tilde{t}_i)\right|$$
$$= \bar{w}_0 \times \Delta \sum_{i=1}^n \left|f^2(\bar{t}_i) - f^2(\tilde{t}_i)\right|$$
$$= o(1),$$

and (ii) now follows by an application of (i) to $\sum_{i=1}^n w(t_{i-1}) \int_{t_{i-1}}^{t_i} f^2(s)\,ds$. ■

LEMMA 7. *Let the function K satisfy (K.1) and $f \in C^{m,\gamma}[0,T]$. Then uniformly over $\tau \in [a, T-a]$, as $\Delta, h, a/h \to 0$:*

$$\sum_{i=1}^{n} K_h\left(t_{i-1}-\tau\right)\int_{t_{i-1}}^{t_i} f_s\,ds = f(\tau)+h^{m+\gamma}L_f(\tau,0)\int_{\mathbb{R}} K(z)\,z^{m+\gamma}\,dz$$

$$+O\left(\frac{\Delta}{h}\right)+o\left(h^{m+\gamma}\right),$$

$$\sum_{i=1}^{n} K_h^2\left(t_{i-1}-\tau\right)\left(\int_{t_{i-1}}^{t_i} f_s\,ds\right)^2 = \frac{\Delta}{h}f^2(\tau)\int_{\mathbb{R}} K^2(z)\,dz+O\left(\frac{\Delta^{1+\gamma}}{h}\right)+O\left(\frac{\Delta^2}{h^2}\right).$$

Proof. By Lemma 6(i),

$$\sum_{i=1}^{n} K_h\left(t_{i-1}-\tau\right)\int_{t_{i-1}}^{t_i} f_s\,ds = \int_0^T K_h(s-\tau)\,f(s)\,ds+O(\Delta)\times\sup_z\frac{1}{h^2}\left|K'\left(\frac{z-\tau}{h}\right)\right|,$$

where $\sup_z \frac{1}{h^2}\left|K'\left(\frac{z-\tau}{h}\right)\right| = O(1/h)$ by (K.1). Next, using an mth order Taylor expansion in conjunction with (K.1),

$$\int_0^T K_h(s-\tau)\,f(s)\,ds = \sum_{k=0}^{m-1} f^{(k)}(\tau)\int_0^T K_h(s-\tau)(s-\tau)^k\,ds$$

$$+\int_0^T K_h(s-\tau)(s-\tau)^m f^{(m)}(\bar{s})\,ds, \tag{B.1}$$

for some $\bar{s} \in [s,\tau]$, where $\int_0^T K_h(s-\tau)(s-\tau)^k\,ds = h^i\int_{-\tau/h}^{(T-\tau)/h} K(z)\,z^k\,dz$. Each of the integrals on the right-hand side is bounded by

$$\sup_{\tau\in[a,1-a]}\left|\int_{-\tau/h}^{(1-\tau)/h} K(z)\,z^k\,du\right| \le \sup_{\tau\in[a,1-a]}\left|\int_{-\infty}^{-\tau/h} K(u)\,u^k\,du\right|$$

$$+\sup_{\tau\in[a,1-a]}\left|\int_{(1-\tau)/h}^{+\infty} K(u)\,u^k\,du\right|$$

$$\le \int_{-\infty}^{-a/h} |K(u)|\,|u|^k\,du+\int_{(1-a)/h}^{+\infty} |K(u)|\,|u|^k\,du$$

$$= o\left(h^{m-k}\right).$$

The last equality is a consequence of $\int_{-\infty}^{+\infty} |z|^{m+1}|K(z)|\,dz < \infty$, since this implies $|z|^{m+1}$ $K(z) \to 0$ as $|z| \to \infty$, which in turn can be used with l'Hôpital's Rule to conclude that

$$\frac{\int_{-\infty}^{-a/h} K(z)\,z^k\,dz}{h^{m-k}} \propto (-1)^k\,\frac{K(-a/h)(a/h)^{k-2}}{h^{m-k-1}}$$

$$= a^{m+i-1}(-1)^i\,\frac{K(-a/h)}{(a/h)^{m+1}} \to 0, \quad a/h \to 0,$$

and, similarly, $\int_{(T-a)/h}^{\infty} K(z) z^a dz = o\left(h^{m-a}\right)$. With $L_f(s,\Delta)$ denoting the "Lipschitz coefficient" of $f(s)$, the final term in (B.1) satisfies

$$\int_0^T K_h(s-\tau)(s-\tau)^m \left[f^{(m)}(\bar{s}) - f^{(m)}(\tau)\right] ds$$

$$= \int_0^T K_h(s-\tau)(s-\tau)^m \left[L_f(\tau, |\bar{s}-\tau|)(\bar{s}-\tau)^\gamma + o(1)\right] ds$$

$$= h^{m+\gamma} \int_0^T K_h(s-\tau) L_f(\tau, |s-\tau| + o(1))(s-\tau)^{m+\gamma} ds + o\left(h^{m+\gamma}\right)$$

$$= h^{m+\gamma} \int_{-\infty}^{+\infty} K(z) L_f(\tau + zh, zh + o(1)) z^{m+\gamma} dz + o\left(h^{m+\gamma}\right)$$

$$= h^{m+\gamma} L_f(\tau, 0) \int_{-\infty}^{+\infty} K(z) z^{m+\gamma} dz + o\left(h^{m+\gamma}\right).$$

Next, we refine Lemma 6(ii): For $\bar{t}_i, \tilde{t}_i \in \left[t_{i-1}, t_i\right]$,

$$\Delta^{-1} \sum_{i=1}^n K_h^2(t_{i-1}-\tau) \left| \left(\int_{t_{i-1}}^{t_i} f(s) ds\right)^2 - \int_{t_{i-1}}^{t_i} f^2(s) ds \right|$$

$$= \Delta \sum_{i=1}^n K_h^2(t_{i-1}-\tau) \left| f^2(\bar{t}_i) - f^2(\tilde{t}_i) \right|$$

$$\leq C \frac{\Delta}{h} \sum_{i=1}^n \left| f(\bar{t}_i) + f(\tilde{t}_i) \right| \left| f(\bar{t}_i) - f^2(\tilde{t}_i) \right|$$

$$\leq 2C \frac{\Delta^\delta}{h} \times \sup_{0 \leq s \leq T} |f(s)| \times \sup_{0 \leq s \leq T} L_f(s, \Delta),$$

and, using the the same arguments as before, we obtain the second result. ■