CREATES

Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2007-2

Nonparametric Filtering of the Realised Spot Volatility: A Kernel-based Approach

Dennis Kristensen

School of Economics and Management
University of Aarhus
Building 1322, DK-8000 Aarhus C
Denmark

Aarhus School of Business
University of Aarhus
Handelshøjskolen
Aarhus Universitet

UNIVERSITY OF COPENHAGEN

Nonparametric Filtering of the Realised Spot Volatility: A Kernel-based Approach*

Dennis Kristensen[†]
Columbia University and CREATES[‡]

Abstract

A kernel weighted version of the standard realised 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 realised spot volatility is obtained. We denote this the filtered spot volatility. We show consistency and asymptotic normality of the kernel smoothed realised volatility and the filtered spot volatility. The choice of bandwidth is discussed and data-driven selection methods proposed. A simulation study examines the finite sample properties of the estimators.

*I wish to thank Bruce Hansen for fruitful discussions and helpful suggestions, and participants at the Uni. of Mannheim economics seminar, the Greater NY Metropolitian Area Econometrics Colloquium 2006, the Zeuthen workshop 2006, and the conference on "Multivariate Modelling in Finance and Risk Management" for comments and suggestions.

†Economics Department, International Affairs Building, MC 3308, 420 West 118th Street, New York, NY 10027, USA. E-mail: dk2313@columbia.edu.

‡Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation.

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 2nd moment or volatility of the processes which plays a central role in the asset pricing formulas. It has long been recognized that volatility varies over time, and considerable effort has been put into modelling and forecasting this variable, see e.g. Andersen, Bollerslev and Diebold (2005) and Shepard (2005) for reviews.

The increased access to high-frequency intradaily data of asset returns within the past decade has given rise to new empirical measures of the volatility process. In particular, the so-called realised volatility measure has received considerable attention and been widely used in the empirical finance literature. The realised volatility gives a measure of the integrated volatility over a given time period and have been used 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 correct measure of the 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 et al (2002). Most of these alternatives are also integrated measures over some time window. This raises the question, 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 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 realised 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 behaviour at specific points in time, and give different weights to the volatility over the window used. In particular, we demonstrate that as the bandwidth shrinks to zero, the instantaneous (or spot) volatility can be extracted. Thus, the inclusion of a kernel and bandwidth in the calculation of the interated volatility allows us to get a better picture of the behaviour of the volatility process.

Given high-frequency data, a nonparametric estimator of the kernel smoothed in-

[1] Barndorff-Nielsen, Hansen, Lund and Shepard (2004) also employ kernels in the estimation of integrated volatility, but for a completely different reason, namely to adjust for measurement errors.

tegrated volatility is easily constructed using the same idea as for the realised 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 generalisation of already existing ones found in the literature (see e.g. Barndorff-Nielsen and Shephard, 2004a,b) 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. Wu and Chu (1993). 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. Given that the volatility process is a latent variable, not observed by the econometrician, previous work on this have based the estimation of volatility models on the raw return data itself (Andersen and Lund, 1997; Chib et al, 2002) or the realised integrated volatility (Andersen et al, 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. This new class of estimators should allow for simple numerical calculation and inference. Renò (2006) presents Monte Carlo evidence for a nonparametric estimator of a SV-model using integrated volatility.

The proposed kernel smoothed version of the realised 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 deterministic setting was taken in Mikosch and Stărica (2005). Fan et al (2003) also consider kernel estimation of deterministic functions of time in the context of term structure models.

Our estimator includes as a special case the rolling window estimator proposed by Foster and Nelson (1996); see also Andreou and Ghysels (2002) and Mykland and Zhang (2003,2006). Our theoretical results complement the ones found in these studies. Alternative approaches are pursued in Genon-Catalot et al (1992) and Malliavin and Marcino (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 derive 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 remains of the paper is organized as follows: In the next section, 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 is 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. The regularity conditions imposed on the model are given in Appendix A. All proofs and lemmas can be found in Appendix B and C respectively.

2 Kernel Smoothed Realised Volatility

Consider the Brownian semimartingale (SMG) $\{X_t\} = \{X_t : t \geq 0\}$ 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 $\{\sigma_t^2\}$ is usually denoted the (instantaneous) 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, $\{\sigma_t^2\}$ 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 realised 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 $0 = t_0 < t_2 < .... < t_m = t$ with $\Delta := \max_{i=1,...,m} |t_i - t_{i-1}|$, c.f Protter (2004, Theorem 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 subinterval of window length $h > 0$ in $[0, T]$ can be recovered perfectly,

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

E.g. if time is measured in days, IV(i), $i = 0, 1, ., , , [T] - 1$, with $h = 1$ will give us daily integrated volatility over the time span $[0, T]$.

In general however, a full trajectory is not available. The most we can hope for is a high-frequency discrete sample $\{X_{t_i} : i = 0, ..., n\}$ over the interval $[0, T]$. This can be used to obtain an estimate of the quadratic variation which in turn converges towards 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,

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

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

$$\text{RV}(\tau) = \int_{\tau-h}^{\tau} d\widehat{[X]}_s = \sum_{i=1}^{n} \mathbb{I}\{\tau - h < t_{i-1} < \tau\} \Delta X_{t_{i-1}}^2. \tag{4}$$

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 Shepard (2004a,b) and Barndorff-Nielsen et al (2004,2006) under the infill assumption: Assuming that the time distance between observations $\Delta := \max_{i=1,...,n} |t_i - t_{i-1}|$ shrinks to zero, consistency and mixed asymptotic normality of RV(τ) can be derived.

One can regard the realised volatility estimator in (4) as a histogram estimator of the instantaneous volatility where $h > 0$ is the binwidth. Here, we propose an alternative measure of the integrated volatility and construct an estimator of it in the same spirit as the realised 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 normalise 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 utilises 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. The function $\tau \mapsto \mathrm{KV}(\tau)$ can be seen as a continuous-time equivalent of the filter or rolling window examined in Foster and Nelson (1996) and Andreou and Ghysels (2002); the kernel can be chosen to satisfy the weighting schemes proposed there while the bandwidth determines the lag length. 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 generalised version of the standard integrated volatility over some fixed time window,

$$\mathrm{IV}(\tau) = \int_0^\tau g\left(t, \sigma_t^2\right) dt. \tag{5}$$

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 realised volatility estimators.

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

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

This is a Nadaraya-Watson 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 σ_t^2 is the dependent variable. By an appropriate choice of K, the realised measure $\mathrm{RV}(\tau)$ includes as special cases the standard realised volatility estimator. Observe that in contrast to these integrated volatility estimators, $\tau \mapsto \mathrm{RV}(\tau)$ 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

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

to emphasise 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, $\hat{\sigma}^2 = \sum_{i=1}^n \Delta X_{t_{i-1}}^2 / T$. Alternatively, one may use the estimator

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

The denominator will converge towards $\mathbb{I}\{0 \le \tau \le T\}$ as $n \to \infty$ and $nh \to 0$, so the two estimators in Eq. (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 an one-sided kernel. Choosing the one-sided kernel appropriately, we recover the rolling-window estimator of Foster and Nelson (1996). One-sided kernels have the advantage that it is 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 Wu and Chu (1993). We can carry over much of the theory established there to our setting.

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 generalised version of the integrated volatility measure in Eq. (5) by

$$\widehat{\mathrm{IV}}(\tau) = \int_0^\tau g\left(t, \hat{\sigma}_t^2\right) dt. \tag{8}$$

3 Asymptotics of the Volatility Estimator

We here derive the asymptotics of the volatility estimators $\mathrm{RV}(\tau)$ and $\hat{\sigma}_\tau^2$ introduced in the previous section. All our results are derived under the following standard assumptions in the econometrics literature on realised volatility:

We shall throughout work under the following set of conditions:

A.1 The process is sampled at $t_i = i\Delta$, $i = 0, 1, ..., n$, where $T = n\Delta$.

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

A.3 The volatility process $\{\sigma_t^2\}$ is locally bounded away from zero and there exists $\nu > 0$ such that:

$$\lim_{\Delta \to 0} \sqrt{\Delta} \sum_{i=1}^n \left|\sigma_{s_i}^\nu - \sigma_{t_i}^\nu\right| = 0$$

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

A.4 The mean process $\{\mu_t\}$ satisfies

$$\lim_{\Delta \to 0} \sum_{i=1}^n \frac{\left|\mu_{i\Delta} - \mu_{(i-1)\Delta}\right|}{\Delta} < \infty.$$

For a discussion of the above conditions with extensions to more realistic settings, we refer to Barndorff-Nielsen et al (2004). Note in particular that we assume that $\{\sigma_t^2\}$ is independent of $\{W_t\}$. This allows us in the following to make all arguments conditional on $\{\sigma_t^2\}$.

While the above conditions suffice to derive the asymptotics of $\mathrm{RV}(\tau)$ 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 work under the weaker assumption that $\{\sigma_t^2\}$ is smooth of order $\gamma \in (0,1]$ almost surely:

A.5 $\{\sigma_t^2\}$ satisfies
$$\left|\sigma_{t+\delta}^2 - \sigma_t^2\right| = L_t(\delta)\,\delta^\gamma + o_P(\delta^\gamma) \text{ a.s.}$$
as $\delta \to 0$, where $\delta \mapsto L_t(\delta)$ is a slowly varying (random) function at zero, $t \mapsto L_t(0)$ is continuous, and $0 < \gamma \leq 1$.

This is a rather weak assumption which are satisfied by many stochastic volatility models. In particular, it holds for any model driven by a Brownian motion for any $\gamma < 1/2$, c.f. Revuz and Yor (1998, Ch. V, Exercise 1.20). We shall also consider the case where $t \mapsto \sigma_t^2$ satisfies the stronger condition that it is $m \geq 1$ times almost surely differentiable:

A.6 $t \mapsto \sigma_t^2$ is a.s. $m \geq 1$ times continuous differentiable on $[0, T]$.

Under (A.6), the use of a so-called higher-order kernel of order m as given in (K.2) below will further reduce the bias of the kernel estimator. The smoothness condition in (A.6) was also employed in Genon-Catalot et al (1992) where σ_t^2 was assumed to be a deterministic function of time.

Finally, we need to impose regularity conditions on the kernel function:

K.1 The kernel K is continuously differentiable and satisfies $\int_{\mathbb{R}} K(z)\,dz = 1$, $\int_{\mathbb{R}} |K(z)|\,dz < \infty$; $|z|\left|K^{(s)}(z)\right| \to 0$ as $|z| \to \infty$; $\sup_z \left|K^{(s)}(z)\right| < \infty$, $s = 0, 1$.

Under (A.6), we will in addition assume that K is a higher-order kernel of order m:

K.2 $\int_{\mathbb{R}} z^i K(z)\,dz = 0$, $i = 1, ..., m-1$, and $\int_{\mathbb{R}} |z|^m K(z)\,dz < \infty$ for some $m \geq 1$.

We are now ready to derive the asymptotics of $\mathrm{RV}(\tau)$ and $\hat{\sigma}_\tau^2$. For the moment, suppose $\mu_t = 0$. Then $\Delta X_{t_{i-1}} = X_{t_i} - X_{t_{i-1}}$ can be written as
$$\Delta X_{t_i} = \int_{t_{i-1}}^{t_i} \sigma_s dW_s \overset{\text{law}}{=} \sqrt{\int_{t_{i-1}}^{t_i} \sigma_s^2 ds} U_i$$
where U_i, $i = 1, ..., n$, are i.i.d $N(0,1)$. Define
$$V_{n,i}(\tau) = K_h(t_{i-1} - \tau) \int_{t_{i-1}}^{t_i} \sigma_s^2 ds.$$
such that
$$\mathrm{RV}(\tau) - \sum_{i=1}^n V_{n,i}(\tau) \overset{\text{law}}{=} \sum_{i=1}^n V_{n,i}(\tau)\left(U_i^2 - 1\right).$$

The right hand side is a weighted sum of i.i.d. random variables, and we can therefore employ limit theorems for triangular arrays of independent variables.

Let us first consider the case of fixed $h > 0$: We easily obtain,

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

by applying the results of Barndorff-Nielsen and Shephard (2004a). Furthermore, it can be shown that

$$\sum_{i=1}^{n} V_{n,i}(\tau) \to \int_{0}^{T} K_h(s-\tau)\,\sigma_s^2 ds, \quad n\sum_{i=1}^{n} V_{n,i}^2(\tau) \to \int_{0}^{T} K_h^2(s-\tau)\,\sigma_s^4 ds,$$

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

Theorem 1 *Under (A.1)-(A.4) and (K.1),*

$$\sup_{\tau\in[0,T]} |\mathrm{RV}(\tau) - \mathrm{KV}(\tau)| = O_P\left(n^{-1/2}\right).$$

and

$$\sqrt{n}\left\{\mathrm{RV}(\cdot) - \mathrm{KV}(\cdot)\right\} \to^{d} Z(\cdot),$$

where Z is a zero mean Gaussian process with covariance

$$\mathrm{Cov}(Z(\tau), Z(t)) = 2\int_{0}^{T} K_h(s-\tau)\,K_h(s-t)\,\sigma_s^4 ds.$$

In particular,

$$\sqrt{n}\left\{\mathrm{RV}(\tau) - \mathrm{KV}(\tau)\right\} \to^{d} N\left(0, 2\int_{0}^{T} K_h^2(s-\tau)\,\sigma_s^4 ds\right).$$

A consistent estimator of $\int_0^T K_h^2(s-\tau)\,\sigma_s^4 ds$ can be obtained by the realised quarticity,

$$\mathrm{RQ}(\tau) := \frac{n}{3}\sum_{i=1}^{n} K_h^2(t_{i-1}-\tau)\,\Delta X_{t_{i-1}}^4, \tag{9}$$

c.f. Proof of Theorem 1, such that

$$\sqrt{n}\frac{\mathrm{RV}(\tau) - \mathrm{KV}(\tau)}{\sqrt{2\mathrm{RQ}(\tau)}} \to^{d} N(0,1).$$

Simulation studies found in Barndorff Nielsen and Shephard (2001) suggest that a better finite sample performance of the above asymptotic result is obtained by considering a log-tranformation of $\mathrm{RV}(\tau)$.

Next, we deal with the case $h \to 0$: First, combining standard results for kernel estimators with the above results, it holds that for any $\tau \in (0, T)$

$$E\left[\hat{\sigma}_\tau^2\right] = \sigma_\tau^2 + o(1), \quad \mathrm{Var}\left[\hat{\sigma}_\tau^2\right] = \frac{2\sigma_\tau^4}{nh}\int_{\mathbb{R}} K^2(z)\,dz + o(1/(nh)), \tag{10}$$

as $\Delta, h \to 0$. 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 show asymptotic normality, we need to refine the bias rate in Eq. (10). Under (A.5), we obtain

$$E\left[\hat{\sigma}_\tau^2\right] - \sigma_\tau^2 = \int_{\mathbb{R}} K(z)\left[\sigma_{\tau+zh}^2 - \sigma_\tau^2\right]dz = L_\tau(0)\,h^\gamma + o_{a.s.}(h^\gamma), \tag{11}$$

where $\gamma \in (0, 1]$ is the smoothness parameter. Note that the convergence rate is potentially slower than in the case of a differentiable volatility process. Under (A.6) and (K.2), the above bias expression is still valid with $\gamma = m \geq 1$ and $L_t(0) = \partial^m \sigma_t^2/\partial t^m \int z^m K(z)\,dz/m!$.

Given the above expression for the bias, we can establish asymptotic normality. We write

$$\begin{aligned}
\sqrt{nh}\frac{\hat{\sigma}_\tau^2 - \sigma_\tau^2}{\sqrt{2\sigma_\tau^4\int_{\mathbb{R}} K^2(z)\,dz}} &= \frac{\hat{\sigma}_\tau^2 - \int_0^T K_h(s-\tau)\,\sigma_s^2 ds}{\sqrt{2\int_0^T K_h^2(s-\tau)\,\sigma_s^4 ds}}\sqrt{\frac{\int_0^T K_h^2(s-\tau)\,\sigma_s^4 ds}{\sigma_\tau^4\int_{\mathbb{R}} K^2(z)\,dz/h}} \\
&\quad + \sqrt{nh}\frac{\int_0^T K_h(s-\tau)\,\sigma_s^2 ds - \sigma_\tau^2}{\sqrt{2}\sigma_\tau^2}.
\end{aligned}$$

By standard CLT results for triangular arrays, the first term converges towards weakly towards $N(0, 1)$ as $h \to 0$ and $nh \to \infty$, while the bias term goes to zero as $nh^{2\gamma+1} \to 0$.

All 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 5.

Theorem 2 *Under (A.1)-(A.5) and (K.1), for any $\varepsilon > 0$,*

$$\sup_{\tau \in [\varepsilon, T-\varepsilon]} \left|\hat{\sigma}_\tau^2 - \sigma_\tau^2\right| = O_P(h^\gamma) + O_P(1/\sqrt{nh}),$$

where $0 < \gamma \leq 1$ is the order of smoothness of σ_t^2. If in addition $nh \to \infty$ and $nh^{1+2\gamma} \to 0$,

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

with asymptotic independence across distinct points.

If (A.6) and (K.2) also hold, the above results go through with $\gamma = m \geq 1$ being the number of derivatives.

It is easily shown that the unknown component in the variance, σ_τ^4, can be consistently estimated by

$$\hat{\sigma}_\tau^4 = \frac{n}{3} \sum_{i=1}^{n} K_h \left(t_{i-1} - \tau\right) \Delta X_{t_{i-1}}^4.$$

Observe that for a given level of smoothness of σ_t^2, the highest attainable rate of convergence is $O_P\left(n^{-\gamma/(2\gamma+1)}\right)$ when the bandwidth is chosen as $h = O\left(n^{-1/(2\gamma+1)}\right)$. In the case $\gamma = 1/2$, this aligns with the convergence rate reported in Foster and Nelson (1996). Bandi and Phillips (2003, Theorem 6) consider the following spatial kernel estimator,

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

when $\sigma_t^2 = \sigma^2\left(X_t\right)$. Under regularity conditions, it satisfies,

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

where $\bar{L}_T\left(T, x\right)$ 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 $[0, T]$. The factor $\bar{L}_T\left(T, X_\tau\right)/T$ determines whether the spatial estimator is more (> 1) or less (< 1) precise than our time domain estimator (given that the two bandwidths converge with the same rate). If only (A.5) holds however, then under sufficient smoothness conditions on the function $\sigma^2\left(\cdot\right)$, the convergence rate of $\tilde{\sigma}_\tau^2$ is in general faster since Bandi and Phillips (2003) can allow for $h \to 0$ at a slower rate.

Finally, we consider the estimator of the generalised integrated volatility measure given in Eq. (8). Compared to the standard realised volatility measure, the above estimator carries an additional bias term due to the kernel smoothing inherent in $\hat{\sigma}_t^2$. By letting $h \to 0$ as $n \to \infty$, this bias term vanishes asymptotically.

Theorem 3 *Let $g\left(t, x\right)$ be continuous in t and twice continuously differentiable in x. Under (A.1)-(A.5), as $h \to 0$,*

$$\sup_{\tau \in [0,T]} |\widehat{\mathrm{IV}}\left(\tau\right) - \mathrm{IV}\left(\tau\right)| = O_P\left(h^\gamma\right) + O_P\left(1/\sqrt{n}\right).$$

If in addition $nh^{2\gamma} \to 0$,

$$\sqrt{n}\left\{\widehat{\mathrm{IV}}\left(\tau\right) - \mathrm{IV}\left(\tau\right)\right\} \to^d N\left(0, 2\int_0^\tau \left[\frac{\partial g\left(t, \sigma_t^2\right)}{\partial x}\right]^2 \sigma_t^4 dt\right).$$

If (A.6) and (K.2) also hold, the above results go through with $\gamma = m$ being the number of derivatives.

4 Boundary Effects

As demonstrated in the previous section, estimation of σ_τ^2 in the interior of the interval over which we have observed $\{X_t\}$ 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 realised 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 the so-called boundary or edge effect, well-known in the kernel estimation literature: For a given bandwidth, the symmetric kernel assigns weight outside the support of the data which 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 following local linear volatility estimator,

$$\hat{\sigma}_\tau^2 = \frac{\sum_{i=1}^n w_{t_{i-1}}(\tau)\,\Delta X_{t_{i-1}}^2}{\sum_{i=1}^n w_{t_{i-1}}(\tau)}, \tag{12}$$

where

$$w_{t_{i-1}}(\tau) = \Delta K_h(t_{i-1} - \tau)\{S_{n,2} - (t_{i-1} - \tau) S_{n,1}\}, \quad S_{n,k} = \Delta \sum_{i=1}^n K_h(t_{i-1} - \tau)(t_{i-1} - \tau)^k.$$

This is similar in spirit to the estimators of deterministic time trends considered in Fan et al (2003). The local linear estimator is known to adapt automatically to the boundaries, thereby not suffering from a boundary bias. By adapting the results of Fan and Gijbels (1992) to our setting, we can show that under (A.5), the bias and the variance in the interior of $[0, T]$ takes the same form as for the Nadaraya-Watson estimator while under (A.6) with $m = 1$, the bias expansion in Eq. (11) holds with $\gamma = 2$ and $L_\tau(0) = \frac{1}{2}\left\{\int z^2 K_0^*(z)\,dz\right\}\partial\sigma_\tau^2/\partial\tau$,where K_0^* is the so-called equivalent kernel, c.f. Fan and Gijbels (1992, Theorem 1). Choosing $\tau = T - ch$, for some constant $c > 0$, the bias expansion remains valid but now the equivalent kernel takes the form given in Fan and Gijbels (1992, Theorem 4). Thus, the estimator is not asymptotically biased at the boundary in contrast to the Nadaraya-Watson estimator with a symmetric kernel.

An alternative method is to use asymmetric kernels. For an overview of asymmetric kernel estimators, we refer to Bouezmarni and Scaillet (2005) who focus on the case of densities with non-negative support. We consider the following specific asymmetric kernel estimator based on Chen (2000),

$$\hat{\sigma}_\tau^2 = \frac{\sum_{i=1}^n K(t_{i-1}/T; \tau/T, h)\,\Delta X_{t_{i-1}}^2}{\Delta \sum_{i=1}^n K(t_{i-1}/T; \tau/T, h)}, \tag{13}$$

where $K(\cdot; y, h)$ is the Beta $(y/h + 1, (1 - y)/h + 1)$ density. This kernel adapts to where we are in the domain $[0, T]$ and in particular gives zero weight outside of this interval. By following the arguments in Chen (2000), one can show that $E[\hat{\sigma}_\tau^2] = \sigma_\tau^2 + L_\tau(0) h^\gamma + o(h^\gamma)$ uniformly over $\tau \in [0, T]$, while the variance has the standard form. If (A.6) holds with $m = 2$, then $\gamma = 1$ and $L_\tau(0) = (1 - 2\tau/T) \partial\sigma_\tau^2/\partial\tau + 1/2\tau/T(1 - \tau/T) \partial^2\sigma_\tau^2/\partial\tau^2$. So the bias is of a higher magnitude relative to the local linear estimator.

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 believe that controlling for the presence of μ_t could yield a better (finite sample) performance of our estimator of σ_t^2; this idea is for example considered in Andreou and Ghysels (2002, Eq. 1.2). Also, 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 using the kernel filtering approach taken here; further structure has to be imposed on μ_t to obtain a consistent estimator of 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 "realised 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 4 *Under (A.1)-(A.4) and (K.1),*

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

and

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

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 prevents us from getting a consistent estimator of the instantaneous drift estimator. In fact, the variance diverges as $h \to 0$.

Theorem 5 *Under (A.1)-(A.5) and (K.1), for any $\varepsilon > 0$,*

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

and as $h \to 0$,

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

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 drawback of the instantaneous volatility estimator, compared to the integrated volatity one, is its dependence on the bandwidth h. The bandwidth can be regarded as a nuisance parameter which must be chosen by the econometrician. In the previous sections, 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, a too large bandwidth will yield a dominating bias term, while a too small 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 method alike the cross-validation method.

Combining the bias and variance expressions given in Eq. (11) and (10) respectively, the approximate pointwise MSE can be written as

$$\mathrm{MSE}\left(h,\tau\right) \approx h^{2\gamma} L_{\tau}\left(0\right) + \frac{2}{nh}\sigma_{\tau}^{4}\left\|K^{2}\right\|^{2},$$

where $\left\|K^{2}\right\|^{2} = \int K^{2}\left(z\right)dz$. This is minimised by the following pointwise bandwidth choice,

$$h_{\mathrm{opt},\tau} = \left(\frac{\sigma_{\tau}^{4}\left\|K^{2}\right\|^{2}}{\gamma L_{\tau}^{2}\left(0\right)}\right)^{1/(2\gamma+1)} n^{-1/(2\gamma+1)},$$

which yields the optimal convergence rate $\mathrm{MSE}\left(h_{\mathrm{opt},\tau},\tau\right) = O\left(n^{-2\gamma/(2\gamma+1)}\right)$. Similarly, the integrated MSE, $\int_{0}^{T}\mathrm{MSE}\left(h,\tau\right)d\tau$, vanishes at the same rate when the bandwidth is chosen globally as

$$h_{\mathrm{opt}} = \left(\frac{\int_{0}^{T}\sigma_{t}^{4}dt\left\|K^{2}\right\|^{2}}{\gamma\int_{0}^{T}L_{t}^{2}\left(0\right)dt}\right)^{1/(2\gamma+1)} n^{-1/(2\gamma+1)}. \tag{14}$$

As with standard kernel methods, the optimal bandwidth here depends on unknown quantities which need to be estimated in order to make the bandwidth choice operational. In the case of the optimal bandwidth in terms of the IMSE, $\int_{0}^{T}\sigma_{t}^{4}dt$ can be estimated using the realised quarticity in Eq. (9) with a uniform kernel and $\tau = T$, $n/3\sum_{i=1}^{n}\Delta X_{t_{i-1}}^{4} \to^{P} \int_{0}^{T}\sigma_{t}^{4}dt$.

Under (A.5), the smoothness parameter γ can be estimated using the method proposed in Blanke (2002) while it appears difficult to obtain an estimator of $L_{t}\left(0\right)$. In the case of (A.6), a simple estimator of $\partial^{m}\sigma_{t}^{2}/\partial t^{m}$ is given by

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

which in turn can be used to calculate $\int_{0}^{T}\left(\partial^{m}\hat{\sigma}_{t}^{2}/\partial t^{m}\right)^{2}dt$. Unfortunately, $\partial^{m}\hat{\sigma}_{t}^{2}/\partial t^{m}$ depends on the bandwidth h itself. One way of solving the problem is to use an approximate parametric model for $\{\sigma_{t}^{2}\}$. For example, $d\sigma_{t}^{2} = \alpha\sigma_{t}^{2}dt$, $\sigma_{t}^{2} = \sigma_{0}^{2}\exp\left[\alpha t\right]$, where 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 $\partial^m \hat{\sigma}_t^2 / \partial t^m = \hat{\alpha}^m \hat{\sigma}_0^2 \exp[\hat{\alpha} t]$. One can obviously run an iterative procedure, where the previous bandwidth choice is plugged into the RHS of Eq. (14), 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,

$$\text{ISE}(h) = \int_{T_l}^{T_u} \left[\sigma_t^2 - \hat{\sigma}_t^2\right]^2 dt = \int_{T_l}^{T_u} \hat{\sigma}_t^4 dt + \int_{T_l}^{T_u} \sigma_t^4 dt - 2\int_{T_l}^{T_u} \sigma_t^2 \hat{\sigma}_t^2 dt,$$

for $0 \leq T_l < T_u \leq T$. The second term can be ignored since it is independent of h, while the first and third can be estimated by $\Delta \sum_{i=1}^n \mathbb{I}\{T_l \leq t_{i-1} \leq T_u\} \hat{\sigma}^4_{-i,t_{i-1}}$ and $2\sum_{i=1}^n \mathbb{I}\{T_l \leq t_{i-1} \leq T_u\} \Delta X^2_{t_{i-1}} \hat{\sigma}^2_{-i,t_{i-1}}$ respectively, where $\hat{\sigma}^2_{-i,t}$ is the leave-one-out estimator. Note that one cannot evaluate $\int_{T_l}^{T_u} \hat{\sigma}_t^4 dt$ using standard formulas involving the convolution of K since the integral is over the compact interval $[T_l, T_u]$. If a symmetric kernel is used, one should normally choose $T_l < T_u$ to avoid any boundary effects. With a one-sided kernel with $K(z) = 0$, $z > 0$, one can choose $T_u = T$ while T_l should still be chosen greater than zero. We define the cross-validated bandwidth as

$$h_{\text{cv}} = \arg\min_{h>0} \text{CV}(h),$$

where

$$\text{CV}(h) = \sum_{i=1}^n \mathbb{I}\{T_l \leq t_{i-1} \leq T_u\} \left[\hat{\sigma}^4_{-i,t_{i-1}} \Delta - 2\Delta X^2_{t_{i-1}} \hat{\sigma}^2_{-i,t_{i-1}}\right]. \tag{15}$$

One should be able to show that this converges towards h_{opt} by following the arguments in for example Hall (1983).

The optimal symmetric kernel weighting in terms of minimal integrated MSE for $m = 2$ proves to be the so-called Epanechnikov (1969) kernel,

$$K(z) = \begin{cases} 3/4\,(1-z)^2, & |z| \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

For one-sided kernels, Zhang and Karunamuni (1998) suggest the kernel below which minimises MSE when the kernel is restricted to have only one sign change,

$$K(z) = \begin{cases} 6\,(1 + 3z + 2z^2), & -1 \leq z \leq 0 \\ 0, & \text{otherwise} \end{cases}. \tag{16}$$

Simulation studies show however that the choice of the kernel K in kernel density and regression estimation has negligible effect, see Silverman (1986, Table 3.1). This result is supported by the simulation study carried out in the next section.

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,

$$\begin{aligned} dX_t &= \mu dt + \sigma_t dW_{1,t}, \\ d\sigma_t^2 &= \beta\left(\alpha - \sigma_t^2\right) dt + \kappa \sigma_t dW_{2,t}, \end{aligned}$$

where $W_{1,t}$ and $W_{2,t}$ are independent standard Brownian motions. This is the GARCH continuous-time limit derived in Drost and Werker (1996). The data-generating parameters are chosen to match the estimated parameter values in Andersen and Bollerslev (1998) for the Yen-USD exchange rate. We consider $\Delta = 1/(3 \times 60 \times 24)$, $1/(60 \times 24)$, $1/(12 \times 24)$, and $1/48$ corresponding to sampling every 20 sec., 1 min., 5 min. and 30 min., and set $T = 2$ (48 hrs.). In order to simulate data from the model, we employ the Milstein discretisation scheme (see Kloeden and Platten, 1999),

$$\begin{aligned} \Delta X_{i\delta} &= \mu\delta + \sigma_{(i-1)\delta}\sqrt{\delta}\varepsilon_{1,i}, \\ \Delta\sigma_{i\delta}^2 &= \beta\left(\sigma_{(i-1)\delta}^2 - \alpha\right)\delta + \kappa\sigma_{(i-1)\delta}^{2\gamma}\sqrt{\delta}\varepsilon_{1,i}, \end{aligned}$$

with $\delta = 1/100$. 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 Eq. (16), 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 CV (h) in Eq. (15).

We first only compare the performance over the interval $[1/2, 3/2]$ so we can ignore any boundary bias; this issue is investigated separately below. To evaluate the precision of the volatility estimators, the integrated mean square error IMSE $= E\int_{1/2}^{3/2}\left[\hat{\sigma}_t^2 - \sigma_t^2\right] dt$ is calculated. We do this using a discrete approximation of the integral. The results based on 400 simulations are reported in 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 recommendable 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 min. sampling, while all three estimators are rather imprecise when the sampling frequency drops to 30 min.

$1/\Delta$	Gaussian Kernel			One-sided Kernel			Beta Kernel		
	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

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

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 $t = 1.5$, 1.9, 1.95, 1.99, and 2.00 are reported for $\Delta^{-1} = 60 \times 24$. As expected the performance of the Gaussian kernel estimator quickly deteriorates as we get nearer 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 frequences.

t	Gaussian Kernel			One-sided Kernel			Beta Kernel		
	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

Table 2: Pointwise sq. bias ($\times 10^{-3}$), variance ($\times 10^{-3}$) and MSE ($\times 10^{-3}$) of kernel estimators.

The findings reported above are illustrated in Figure 1-3. We here have simulated one trajectory of $\{\sigma_t^2\}$, and keep this fixed. We then draw 400 samples of $\{X_t\}$ using this specific volatility trajectory and so analyse the behaviour of $\{\hat{\sigma}_t^2\}$ conditional on $\{\sigma_t^2\}$. In Figure 1-3, the specific trajectory is plotted together with the mean and 95%-confidence bands of $\hat{\sigma}_t^2$ using the 3 different kernels over the interval $t \in [1.5, 2.0]$. The plots support the results reported in Table 1 and 2: While the Gaussian kernel estimator performs well in the interior, as we get closer to the boundary its performance quickly deteriorates. The one-sided and Beta kernel estimator are both fairly unaffected by boundary effects. We ran this last set of simulations conditional on other volatility trajectories and obtained similar results to those reported 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 interior and at the boundary of the sampling interval.

8 Concluding Remarks

We proposed a kernel smoothed version of the standard realised volatility estimator which allows us to filter the unobserved volatility process. The filtered volatility has potential uses in jump detection, realised volatility estimation with market microstructure noise, and estimation of stochastic volatility models. Several extensions of the kernel estimator offer themselves: One could consider kernel smoothing of other transformations of $\Delta X_{t_{i-1}}$. That is, estimators on the form

$$\mathrm{RV}(\tau) = \sum_{i=1}^{n} K_h(t_{i-1} - \tau)\, g\left(\Delta X_{t_{i-1}}\right),$$

for some function g. For example, power variation $g(x) = |x|^p$, $p \geq 1$, as considered in Barndorff-Nielsen and Shephard (2004a). Also, the generalization of our results to a multivariate setting and to allow for leverage effects would be of interest. These topics are all left for future research.

References

Alizadeh, S., M.W. Brandt and F.X. Diebold (2002) Range-Based Estimation of Stochastic Volatility Models. *Journal of Finance* 57, 1047-1092.

Andersen, T.G. and T. Bollerslev (1997) Intraday Periodicity and Volatility Persistence in Financial Markets. *Journal of Empirical Finance* 4, 115-158.

Andersen, T.G. and 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 and F.X. Diebold (2004) Parametric and Nonparametric Measurement of Volatility. Forthcoming in *Handbook of Financial Econometrics* (eds. Y. Aït-Sahalia and L.P. Hansen). Amsterdam: North Holland.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001) The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96, 42-55.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003) Modeling and Forecasting Realized Volatility. *Econometrica* 71, 579-625.

Andersen, T.G. and J. Lund (1997) Estimating Continuous-Time Stochastic Volatility Models of the Short Term Interest Rate. *Journal of Econometrics* 77, 343-377.

Andreou, E. and 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. and N. Shephard (2001) Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society, Series B* 64, 253-280.

Barndorff-Nielsen, O.E. and N. Shephard (2004a) Realised Power Variation and Stochastic Volatility Models. *Bernoulli* 9, 243-265 and 1109-1111.

Barndorff-Nielsen, O.E. and N. Shephard (2004b) Econometric Analysis of Realised Covariation: High Frequency Based Covariance, Regression and Correlation in Financial Economics. *Econometrica* 72, 885-925.

Barndorff-Nielsen, O.E., S.E. Graversen and N. Shephard (2004) Power Variation & Stochastic Volatility: A Review and Some New Results. *Journal of Applied Probability* 41, 133-143.

Barndorff-Nielsen, O.E., S.E. Graversen, J. Jacod and N. Shephard (2006) Limit Theorems for Bipower Variation in Financial Econometrics. Forthcoming, *Econometric Theory.*

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lund and N. Shephard (2004) Designing Realised Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise. Working paper, University of Oxford.

Barucci, E. and R. Renò (2002) On Measuring Volatility of Diffusion Processes with High Frequency Data. *Economics Letters* 74, 371-378.

Bertsimas, D., L. Kogan, and A.W. Lo (2000) When Is Time Continuous?. *Journal of Financial Economics 55*, 173-204.

Björk, T. (2004) *Arbitrage Theory in Continuous Time*. Oxford: 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. and H. Zhou (2002) Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility. *Journal of Econometrics* 109, 33-65.

Chen, S.X. (2000) Beta Kernel Smoothers for Regression Curves. *Statistica Sinica* 10, 73-91.

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

Drost, F.C. and 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. and I. Gijbels (1992) Variable Bandwidth and Local Linear Regression Smoothers. *Annals of Statistics* 20, 2008-2036.

Fan, J. , J. Jiang, C. Zhang and 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. and D.B. Nelson (1996) Continuous Record Asymptotics for Rolling Sample Variance Estimators. *Econometrica* 64, 139-174.

Genon-Catalot, V., C. Laredo and D. Picard (1992) Non-parametric Estimation of the Diffusion Coefficient by Wavelets Methods. *Scandinavian Journal of Statistics* 19, 317-335.

Hall, P. (1983) Large Sample Optimality of Least Squares Cross-Validation in Density Estimation. *Annals of Statistics* 11, 1156-1174.

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.

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

Jones, C.S. (2003) The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets. *Journal of Econometrics* 116, 181-224.

Kloeden, P.E. and E. Platen (1999) *Numerical Solution of Stochastic Differential Equations*, 2nd Ed. Heidelberg: Springer-Verlag.

Malliavin, P. and 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. and C. Stărica (2005) Stock Market Risk-return Inference. An Unconditional Non-parametric Approach. Preprint No. 202, Dep. of Appl. Math. and Stat., University of Copenhagen.

Mykland, P.A. and L. Zhang (2003) Inference for Volatility-type Objects and Implications for Hedging. Working paper, Dep. of Statistics, Carnegie-Mellon.

Mykland, P.A. and L. Zhang (2006). ANOVA for diffusions and Ito processes. Forthcoming, *Annals of Statistics* 34(4).

Müller, H.-G. (1991) Smooth Optimum Kernel Estimators Near Endpoints. *Biometrika* 78, 521-530.

Pagan, A. and A. Ullah (1999) *Nonparametric Econometrics*. Cambridge: Cambridge University Press.

Protter, P. (2004) *Stochastic Integration and Differential Equations*, 2nd Edition. New York: Springer-Verlag.

Renò, R. (2006) Nonparametric Estimation of Stochastic Volatility Models. *Economics Letters* 90, 390-395.

Revuz, D. and M. Yor (1998) Continuous Martingales and Brownian Motion, 3rd Ed. New York: Springer-Verlag

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

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

Wu, J.S. and C.K. Chu (1993) Kernel Type Estimators of Jump Points and Values of A Regression Function. *Annals of Statistics* 21, 1545-1566.

Zhang, S. and R.J. Karunamuni (1998) On Kernel Density Estimation Near Endpoints. *Journal of Statistical Planning and Inference* 70, 301-316.

A Proofs

Proof of Theorem 1. Define

$$dX_t^* = \sigma_t dW_t,$$

and

$$\mathrm{RV}^*(\tau) = \sum_{i=1}^n (X_{t_i}^* - X_{t_{i-1}}^*)^2 K_h(t_{i-1} - \tau).$$

We have

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

where $c_{n,i}(\tau) = V_{n,i}(\tau)/\sqrt{\sum_{i=1}^n V_{n,i}^2(\tau)}$ and $\bar{U}_i^2 = (U_i^2 - 1)/\sqrt{2}$ are i.i.d. $(0,1)$. Since $\sum_{i=1}^n c_{n,i}^2(\tau) = 1$, and $\max_i c_{n,i}(\tau) \to 0$, the pointwise convergence results then follows by Barndorff-Nielsen and Shephard (2004, Corollary 3.1 and 3.2) together with Lemma 6. The extension of the weak convergence to the interval $[0,T]$ follows from, for example, Van der Vaart and Wellner (1996, Example 1.5.10) since we have pointwise weak convergence, $\mathrm{RV}^*(\tau)$ clearly is stochastically equicontinuous, and $[0,T]$ is compact.

Next, we show that the effect of a non-zero drift term is negligible. We have

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

where, c.f. Lemma 6,

$$\sum_{i=1}^n \left(\int_{t_{i-1}}^{t_i} \mu_s ds\right)^2 K_h(t_{i-1} - \tau) = \Delta \int_0^T \mu_s^2 K_h(s - \tau)\,ds + o\left(\Delta^{3/2}\right),$$

while the 2nd term has mean zero and variance

$$4\sum_{i=1}^n \int_{t_{i-1}}^{t_i} \sigma_s^2 ds \int_{t_{i-1}}^{t_i} \mu_s ds K_h(t_{i-1} - \tau) = 4\Delta \int_0^T \sigma_s^2 \mu_s K_h(s - \tau)\,ds + o\left(\Delta^{3/2}\right).$$

We conclude that $\sqrt{n}\,[\mathrm{RV}(\tau) - \mathrm{RV}^*(\tau')] = o_P(1)$.

The proof of $\mathrm{RQ}(\tau) \to^P \int_0^T K_h^2(s-\tau)\,\sigma_s^4 ds$ follows by the same arguments as above. The uniform convergence in probability follows by the fact that

$$\sup_{\tau\in[0,T]} \sqrt{n}\,|\mathrm{RV}^*(\tau) - \mathrm{IV}(\tau)| = \sqrt{n} \sup_{\tau\in[0,T]} \sqrt{2}\sum_{i=1}^n |V_{n,i}(\tau)|\,\bar{U}_i^2,$$

where the RHS has mean zero and bounded variance. ■

Proof of Theorem 2. All of the pointwise convergence results in the proof of Theorem 1 still go through with $h \to 0$ by appealing to Lemma 7 instead of 6. For any two distinct points $t \neq \tau$, due to $\mathrm{Cov}\left(\bar{U}_i, \bar{U}_j\right) = 0$, $i \neq j$,

$$\begin{aligned}
\mathrm{Cov}\left(\sqrt{nh}\hat{\sigma}_t, \sqrt{nh}\hat{\sigma}_\tau\right) &= 2nh \sum_{i=1}^{n} K_h\left(t_{i-1} - t\right) K_h\left(t_{i-1} - \tau\right) \left(\int_{t_{i-1}}^{t_i} \sigma_s^2 ds\right)^2 + o(1) \\
&= 2h \int_0^T \sigma_t^4 K_h(s - t) K_h(s - \tau) ds + o(1) \\
&= \int_{\mathbb{R}} K(z) K\left(z + \frac{\tau - t}{h}\right) \sigma_{\tau + hz}^4 dz + o(1) \\
&= o(1).
\end{aligned}$$

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

The uniform convergence is established by noting that the first term of the variance component given in Eq. (11) is of order $O_P(1/h)$ as $h \to 0$ while the bias expansion holds uniformly in $\tau \in [\varepsilon, T - \varepsilon]$. ■

Proof of Theorem 3. We have

$$\begin{aligned}
\widehat{\mathrm{IV}}(\tau) - \mathrm{IV}(\tau) &= \int_0^\tau g\left(t, \hat{\sigma}_t^2\right) - g\left(t, \sigma_t^2\right) dt \\
&= \int_0^\tau \frac{\partial g\left(t, \sigma_t^2\right)}{\partial x} \left[\hat{\sigma}_t^2 - \sigma_t^2\right] dt + \frac{1}{2} \int_0^\tau \frac{\partial g\left(t, \bar{\sigma}_t^2\right)}{\partial x} \left|\hat{\sigma}_t^2 - \sigma_t^2\right|^2 dt
\end{aligned}$$

where $\bar{\sigma}_t^2 \in \left[\hat{\sigma}_t^2, \sigma_t^2\right]$. The two terms satisfy

$$\int_0^\tau \left|\frac{\partial g\left(t, \bar{\sigma}_t^2\right)}{\partial x}\right| \left|\hat{\sigma}_t^2 - \sigma_t^2\right|^2 dt \leq C \sup_{t \in [0,T]} \left|\hat{\sigma}_t^2 - \sigma_t^2\right|^2 = O_P\left(h^{2v}\right) + O_P\left(1/(nh)\right),$$

$$\begin{aligned}
\int_0^\tau \frac{\partial g\left(t, \sigma_t^2\right)}{\partial x} \hat{\sigma}_t^2 dt &= \sum_{i=1}^{n} \Delta X_{t_{i-1}}^2 \int_0^\tau \frac{\partial g\left(t, \sigma_t^2\right)}{\partial x} K_h\left(t_{i-1} - t\right) dt dt \\
&= \sum_{i=1}^{n} \mathbb{I}\left\{t_{i-1} < \tau\right\} \Delta X_{t_{i-1}}^2 \frac{\partial g\left(t, \sigma_t^2\right)}{\partial x} + O_P\left(h^v\right)
\end{aligned}$$

where $v = \gamma$ under (A.5) and $v = m$ under (A.6). We can now employ the same techniques as in the proof of Theorem 1 to show the result. ■

Proof of Theorem 4. Define

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

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 1, and obtain from Lemma 6,

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

■

Proof of Theorem 5. The convergence results in the proof of Theorem 4 still hold as $h \to 0$ under the additional conditions stated by appealing to Lemma 7 instead of 6. ■

B Lemmas

Lemma 6 (Fixed $h > 0$) *Let the function K and the process $\{f_t\}$ satisfy:*

1. $\sup_z \left|K^{(s)}(z)\right| < \infty$, $s = 0, 1$.
2. $\int_0^T |f_s| ds < \infty$.

Then with $\Delta := \max_{i=1,\ldots,n} |t_i - t_{i-1}|$:

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

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

uniformly over $\tau \in [0, T]$.

Proof. First, for some $t_{i,s} \in [t_{i-1}, \tau]$,

$$\begin{aligned}
& \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} f_s K_h \left(t_{i-1} - \tau\right) ds - \int_0^T f_s K_h \left(s - \tau\right) ds \\
= & \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} f_s \left\{ K_h \left(t_{i-1} - \tau\right) - K_h \left(s - \tau\right) \right\} ds \\
= & \frac{1}{h^2} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} f_s K^{(1)} \left(\frac{t_{i,s} - \tau}{h} \right) \left(t_{i-1} - s\right) ds \\
= & O\left(\frac{\Delta}{h^2}\right) \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} f_s K^{(1)} \left(\frac{t_{i,s} - \tau}{h} \right) ds, \\
= & O\left(\frac{\Delta}{h^2}\right) \int_0^T f_s ds.
\end{aligned}$$

Using the same arguments as in Barndorff-Nielsen and Shephard (2004a, proof of Lemma 2),

$$\Delta^{-1} \sum_{i=1}^{n} \left(K_h \left(t_{i-1} - \tau\right) \int_{t_{i-1}}^{t_i} f_s ds \right)^2 = \sum_{i=1}^{n} K_h^2 \left(t_{i-1} - \tau\right) \int_{t_{i-1}}^{t_i} f_s^2 ds + o\left(\sqrt{\Delta}\right),$$

and the second result follows from the first part of the lemma. ■

Lemma 7 *Let the function K satisfy (A.5) and f_t is smooth of order $\gamma \in (0, 1]$. Then:*

$$\sum_{i=1}^{n} K_h \left(t_{i-1} - \tau\right) \int_{t_{i-1}}^{t_i} f_s ds = f_t \int_{\mathbb{R}} K\left(z\right) dz + \frac{\Delta}{h} f_t \int_{\mathbb{R}} K^{(1)}\left(z\right) du + h^{\gamma} L_{\tau}\left(0\right) + o_P\left(\frac{\Delta}{h}\right) + o_P\left(h^{\gamma}\right),$$

$$\Delta^{-1} \sum_{i=1}^{n} K_h^2 \left(t_{i-1} - \tau\right) \left(\int_{t_{i-1}}^{t_i} f_s ds \right)^2 = \frac{f_{\tau}^2}{h} \int_{\mathbb{R}} K^2\left(z\right) d + \frac{\Delta}{h^2} f_{\tau}^2 \int_{\mathbb{R}} 2K\left(z\right) K^{(1)}\left(z\right) du + o_P\left(\frac{\Delta}{h^2}\right),$$

where $\Delta := \max_{i=1,\ldots,n} |t_i - t_{i-1}|$. If $t \mapsto L_t(0)$ is continuous, the results hold uniformly over $\tau \in (0, T)$.

If f_t is m times differentiable and K in addition satisfies (A.6), the above results hold with $\gamma = m$ and $L_{\tau}(0) = \partial^m f_{\tau} / \partial \tau^m \int z^m K(z) dz$.

Proof. Combining

$$\begin{aligned}\frac{1}{h}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} f_s K^{(1)}\left(\frac{t_{i,s}-\tau}{h}\right) ds &= \frac{1}{h}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} f_s K^{(1)}\left(\frac{s-\tau}{h}+o\left(1\right)\right) ds \\ &= \frac{1}{h}\int_0^T f_s K^{(1)}\left(\frac{s-\tau}{h}+o\left(1\right)\right) ds \\ &= \int_{\mathbb{R}} f_{\tau+zh} K^{(1)}\left(z+o\left(1\right)\right) dz \\ &= f_\tau \int_{\mathbb{R}} K^{(1)}\left(z\right) dz + o_P\left(1\right),\end{aligned}$$

where we have applied Pagan and Ullah (1999, Lemma A.2.6.1), with the expression derived in the proof of Lemma 6, it holds that

$$\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} f_s K_h\left(t_{i-1}-\tau\right) ds = \int_0^T f_s K_h\left(s-\tau\right) ds + O\left(\frac{\Delta}{h}\right) f_\tau \int_{\mathbb{R}} K^{(1)}\left(z\right) dz.$$

Next, by standard arguments, we obtain,

$$\int_0^T f_s K_h\left(s-t\right) ds - f_\tau \int_{\mathbb{R}} K\left(z\right) dz = h^\gamma L_\tau\left(0\right) + o\left(h^\gamma\right).$$

This shows the first result. Next, observe that

$$\Delta^{-1}\sum_{i=1}^{n} K_h^2\left(t_{i-1}-t_0\right)\left(\int_{t_{i-1}}^{t_i} f_s ds\right)^2 = \sum_{i=1}^{n} K_h^2\left(t_{i-1}-t_0\right)\int_{t_{i-1}}^{t_i} f_s^2 ds + o_P\left(\sqrt{\Delta}\right),$$

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

C Figures

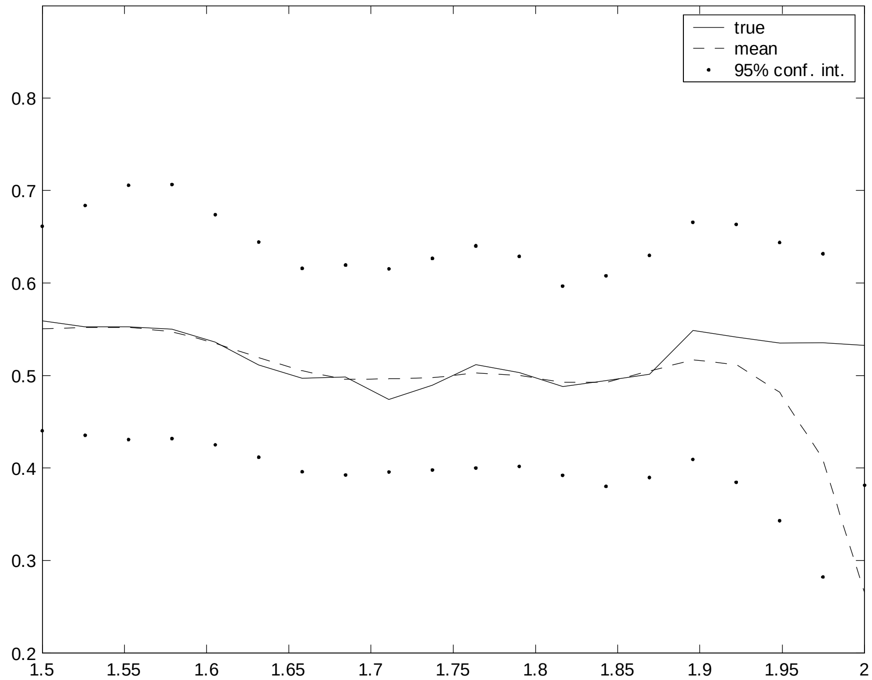

Figure 1: Gaussian Kernel estimator

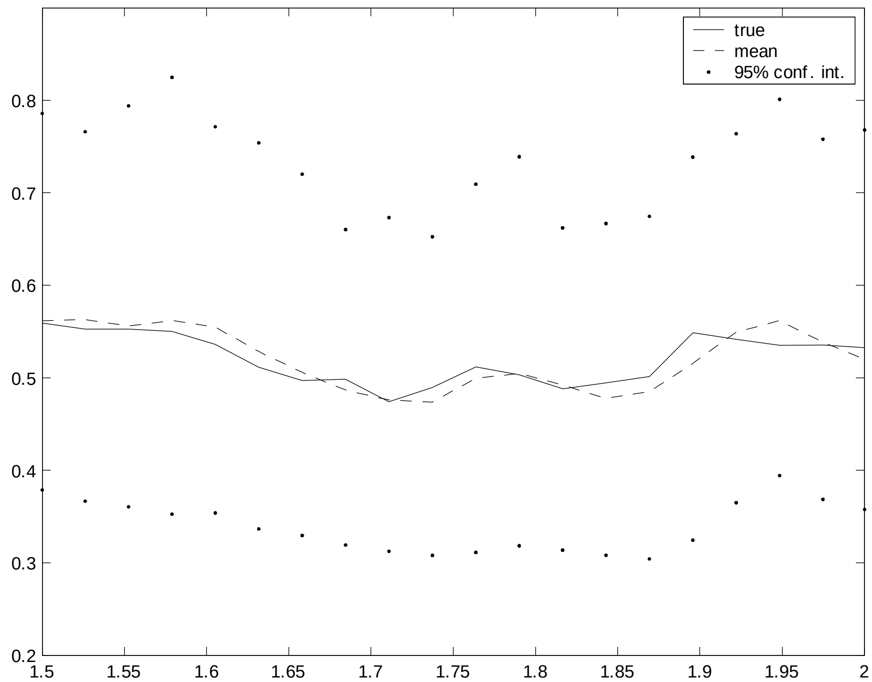

Figure 2: One-sided kernel estimator

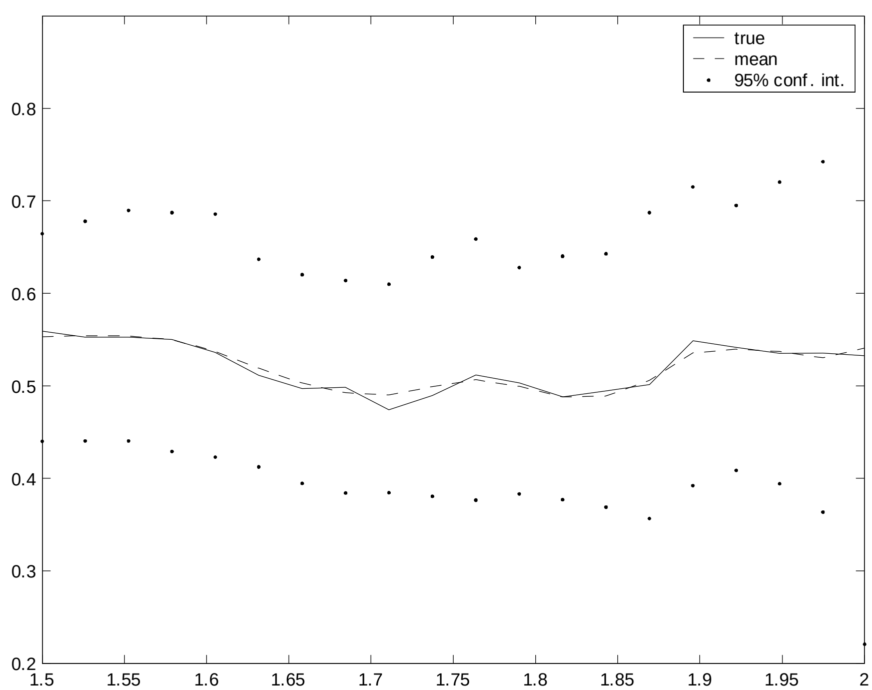

Figure 3: Beta Kernel Estimator

Research Papers 2007

CREATES
Center for Research in Econometric Analysis of Time Series

2007-1 Dennis Kristensen: Nonparametric Estimation and Misspecification Testing of Diffusion Models

2007-2 Dennis Kristensen: Nonparametric Filtering of the Realised Spot Volatility: A Kernel-based Approach