Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications xxx (xxxx) xxx

stochastic processes and their applications

www.elsevier.com/locate/spa

Optimal kernel estimation of spot volatility of stochastic differential equations

José E. Figueroa-López[a,*], Cheng Li[b]

[a] *Department of Mathematics and Statistics, Washington University in St. Louis, St. Louis, MO 63130, USA*
[b] *Citadel Securities, NY, 10022, USA*

Received 6 July 2018; received in revised form 12 May 2019; accepted 19 January 2020
Available online xxxx

Abstract

A unified framework to optimally select the bandwidth and kernel function of spot volatility kernel estimators is put forward. The proposed models include not only classical Brownian motion driven dynamics but also volatility processes that are driven by long-memory fractional Brownian motions or other Gaussian processes. We characterize the leading order terms of the mean squared error, which in turn enables us to determine an explicit formula for the leading term of the optimal bandwidth. Central limit theorems for the estimation error are also obtained. A feasible plug-in type bandwidth selection procedure is then proposed, for which, as a sub-problem, a new estimator of the volatility of volatility is developed. The optimal selection of the kernel function is also investigated. For Brownian Motion type volatilities, the optimal kernel turns out to be an exponential function, while, for fractional Brownian motion type volatilities, easily implementable numerical results to compute the optimal kernels are devised. Simulation studies further confirm the good performance of the proposed methods.

MSC: primary 62M09; 62G05; secondary 60F05

Keywords: Spot volatility estimation; Kernel estimation; Bandwidth selection; Kernel function selection; Vol vol estimation

1. Introduction

The estimation of the diffusive coefficient σ_t of the dynamical stochastic system $dX_t = \mu_t dt + \sigma_t dW_t$, driven by a Brownian motion W, has received some renewed attention in the last few years. This research has partly been pushed by the advent of high-frequency data

* Corresponding author.
E-mail addresses: figueroa-lopez@wustl.edu (J.E. Figueroa-López), Cheng.Li@citadelsecurities.com (C. Li).

https://doi.org/10.1016/j.spa.2020.01.013

(HFD) in several fields but more predominantly in finance. In the latter context, σ_t is called the spot volatility of the price process $S_t = \exp(X_t)$ of a risky asset and, in addition of being a local measure of the asset's riskiness at the time t, it is also needed for many problems of finance such as option pricing and portfolio selection.

In this work, we revisit the problem of spot volatility estimation by kernel methods. Kernel estimation has a long history and extensive treatments of the method can be found in many textbooks. The selection of the bandwidth and the kernel function are of great importance for the performance of the kernel estimator in a finite sample setting. The problem has been extensively studied for density estimation and kernel regression (cf. [4,9,13]). However, in the context of spot volatility estimation, the literature related to this problem is much scarcer. In this work, we put forward a unified framework to the problem that allows us to deal not only with well studied Brownian driven volatilities but also those driven by other Gaussian processes such as fractional Brownian motions.

Literature review. Foster and Nelson [6] studied a rolling window estimator, which can be seeing as a kernel estimator with a compactly supported kernel function. Under a number of stringent conditions, they established the point-wise asymptotic normality of the estimator, and drew some conclusions about the optimal window length (i.e., bandwidth) and the optimal weight functions (kernel functions). However, in spite of the non-parametric model setting, the volatility was constrained to have a Brownian-like degree of smoothness (see Assumption A (vii) and (viii) therein) and the selection of bandwidth and kernel function was not systematically studied, since it was assumed the strict relationship[1] $h_n \asymp n^{-1/2}$ between the window's length h_n and the sample size n (see Assumption D therein). Under such a relationship, they obtained the optimal kernel weights and separately determine the optimal constant c appearing in the formula $h_n = cn^{-1/2}$, but only for the flat-weights or uniform kernel case (see Theorem 4 therein). Fan and Wang [5] also showed a point-wise asymptotic normality for a general kernel estimator under a specific constraint on the rate of convergence of the bandwidth (Condition A4 therein), which allowed them to neglect the error coming from approximating the spot volatility by a kernel weighted volatility (we refer the reader to Section 6 for details), but the achieved convergence rates are suboptimal. For a continuous Itô semimartingale with volatility driven by a Brownian motion and jumps, Alvarez et al. [2] considered the estimation of σ_t^p by taking forward finite differences of the realized power variation process of order p, which is equivalent to a forward-looking kernel estimator with uniform kernel. CLTs were also developed therein, which allowed them to argue that the best possible rate of convergence of the estimation error is $n^{-1/4}$ and that this is attained when $n^{1/2}h_n \to c \in (0, \infty)$, as $n \to \infty$. More general results along the same vein (i.e., with uniform-type kernels) have also been developed in the monograph of Jacod and Protter [7] (see Chapter 13 therein). More recently, Mancini et al. [11] have developed asymptotic normality for a more general class of spot volatility estimators, which includes kernel estimators.

Besides Foster and Nelson [6], the only work we know that studied the problem of optimal bandwidth selection of spot volatility kernel estimators is that of Kristensen [10], who also obtained asymptotic normality of the estimators. However, this work imposes a strong path-wise smoothness condition (see Remark 2.1 for details), which has several practical and theoretical drawbacks. Indeed, even for simple volatility processes, it is not possible to verify the pathwise Hölder continuity needed for a central limit theorem with *optimal rate*. Furthermore, even though an 'optimal' bandwidth formula is deduced in closed form therein,

[1] As usual, $a_n \asymp b_n$ if $ma_n \leq b_n \leq Ma_n$, for all n and some $0 < m < M < \infty$.

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

this is not well-defined if we want to attain optimal convergence rates for the estimation error (see Remark 2.1).

Our contributions. Having discussed some previous work, we now mention some motivating factors and objectives of the present work. To begin with, we wish to impose easily verifiable and general enough conditions to cover a wide range of frameworks without restricting the degree of smoothness of the volatility process. From a theoretical point of view, we also aim to provide a formal justification of the optimal convergence rate of the kernel estimator and to establish central limit theorems (CLT) and asymptotic estimates of the mean square errors with optimal rates. From the practical side, the two factors that affect the performance of the estimator, bandwidth and kernel function, ought to be optimized jointly, not separately, and meanwhile, the proposed method should remain feasible and sufficiently efficient to be implementable for HFD.

The key assumption to our unifying treatment of the problem is a mild local scaling property of the covariance structure of the volatility process. This assumption covers a wide range of frameworks including deterministic differentiable volatility processes and volatilities driven by Brownian Motion, long-memory fractional Brownian Motion, and, more generally, functions of suitable Gaussian processes. Under the referred assumption, we characterize the leading order terms of the Mean Squared Error (MSE) and, as a byproduct, we derive, in closed form, the leading order term of the optimal bandwidth. From this, the theoretical optimal convergence rate for the estimation error is identified. We then proceed to show that our optimal bandwidth formulas are feasible by proposing an iterated plug-in type algorithm for their implementation. An important intermediate step is to find an estimate of the Integrated Volatility of Volatility (IVV), for which we propose a new estimator based on the two-time scale realized variance of Zhang et al. (2005). Consistency and convergence rate of our vol vol estimator are also established. The estimation of the IVV has also been addressed in [3,15].

Equipped with an explicit formula for the asymptotically optimal MSE, we proceed to set up a well-posed problem for optimal kernel selection. Concretely, for Brownian motion driven volatilities, we prove that the optimal kernel function is the exponential kernel $K(x) = 2^{-1}\exp(-|x|)$. Such a result formalizes and extends a previous result of Foster and Nelson [6], where only kernels of bounded support were considered. We also show that, due to the nature of the data we are analyzing (namely, HFD), exponential kernel function enjoys outstanding computational advantages, as it reduces the time complexity for estimating the whole path of the volatility on all grid points from $O(n^2)$ to $O(n)$. We also consider the volatility processes driven by the long-memory fractional Brownian motion and, in such a case, we provide numerical schemes to compute the optimal kernel function.

To complement our asymptotic results based on MSE, asymptotic normality of the kernel estimators is also established for two broad types of volatility processes: Itô processes and continuous function of some Gaussian processes. In this way, our results cover volatility processes with flexible degrees of smoothness. The results are consistent with the leading order approximation of the MSE, so that CLT's with the optimal convergence rate are obtained. By contrast, as mentioned above, the CLT's of Fan and Wang [5] and Kristensen [10] have suboptimal convergence rate, while the analogous result of Foster and Nelson [6] is limited to a specific smoothness order and strong constraints on the kernel function and bandwidth. In the case of Itô volatility processes, we generalize the CLT of Alvarez et al. [2] and Jacod and Protter [7], from uniform to general forward looking kernels.

Paper Outline. The rest of the paper is organized as follows. In Section 2, we introduce the kernel estimator and our assumptions, and verify that common volatility processes satisfy our assumptions. In Section 3, we deduce the leading order approximation of the MSE of the kernel estimator and solve the optimal bandwidth selection problem. Then, in Section 4, we deal with the optimal kernel function selection problem for different types of volatility processes. A feasible implementation approach of the optimal bandwidth is discussed in Section 5, where we also introduce the two-scale estimator of the IVV. Central Limit Theorems of the kernel estimator are discussed in Section 6. Finally in Section 7, we perform Monte Carlo studies. The proofs of the main results are provided in Appendix A while the proofs of some technical lemmas and supporting propositions are deferred to the supplemental material to this article available online.

2. Kernel estimators and assumptions

In this section, we first introduce the classical kernel estimator for the spot volatility. We then discuss some needed smoothness conditions on the volatility processes and verify that most common volatility processes used in the literature indeed satisfy our assumptions. Finally, we discuss some regularity conditions on the kernel function.

Throughout the paper, we consider the following stochastic differential equation (SDE):

$$dX_t = \mu_t dt + \sigma_t dB_t, \tag{2.1}$$

where all stochastic processes ($\mu := \{\mu_t\}_{t\geq 0}, \sigma := \{\sigma_t\}_{t\geq 0}, B := \{B_t\}_{t\geq 0}$, etc.) are defined on a complete filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$, with $\mathbb{F} = \{\mathscr{F}_t\}_{t\geq 0}$. We also assume that μ and σ are adapted to the filtration $\mathbb{F}$ and $B := \{B_t\}_{t\geq 0}$ is a standard Brownian Motion (BM) adapted to $\mathbb{F}$. We assume that the process X is observed at the times $t_i := t_{i,n} := i\Delta_n$, $0 \leq i \leq n$, where $\Delta_n := T/n$. We will use $\Delta_i^n Z := \Delta Z_{t_{i-1}} := Z_{t_i} - Z_{t_{i-1}}$ to denote the increments of a process Z and $\Delta_n = T/n$ to denote the time increments. For notational simplicity, we sometimes omit the subscript n in Δ_n and the superscript in $\Delta_i^n Z$.

In this paper, we study the problem of estimating the spot volatility σ_τ, at a given time $\tau \in (0, T)$, by the kernel estimator (cf. [5,10]),

$$\hat{\sigma}^2_{\tau,n,h} := \sum_{i=1}^{n} K_h(t_{i-1} - \tau)(\Delta_i^n X)^2, \tag{2.2}$$

where $K_h(x) = K(x/h)/h$. Again, for simplicity, we sometimes omit the subscript n and/or h in the notation $\hat{\sigma}^2_{\tau,n,h}$. As is often the case with kernel estimation, the selections of the bandwidth h and kernel function K of (2.2) are of great importance in practice, especially for the finite sample settings commonly encountered in econometric applications.

We now proceed to give the required assumptions on the volatility process that allow us to examine the rate of convergence of the kernel estimator defined in (2.2). Our first assumption, which is also imposed in [10], is a non-leverage condition. For Brownian-driven volatilities and weak convergence results, this assumption *will be relaxed in Section 6*, hence allowing correlation between the Brownian motions driving the volatility and the price processes. However, for more general volatilities, including those driven by fractional Brownian motions, such an assumption would allow us to treat the bandwidth and kernel selection problems in a *unified manner under a mean-squared loss function*, which, as stated in the introduction, is one of our main objectives in this work.

Assumption 1. (μ, σ) are adapted cádlág processes independent of B.

Next, we impose some mild moment boundedness assumption on μ and σ.

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

Assumption 2. There exists $M_T > 1$ such that $\mathbb{E}[\mu_t^4 + \sigma_t^4] < M_T$, for all $0 \le t \le T$.

The following is our key assumption, which at the end of this section is shown to be satisfied by a large spectrum of volatility models.

Assumption 3. Suppose that for $\gamma > 0$ and certain functions $L : \mathbb{R}_+ \to \mathbb{R}_+$ and $C_\gamma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, such that C_γ is not identically zero and

$$C_\gamma(hr, hs) = h^\gamma C_\gamma(r, s), \qquad \text{for } r, s \in \mathbb{R}, h \in \mathbb{R}_+, \tag{2.3}$$

the variance process $V := \{V_t = \sigma_t^2 : t \ge 0\}$ satisfies

$$\mathbb{E}[(V_{t+r} - V_t)(V_{t+s} - V_t)] = L(t)C_\gamma(r, s) + o((r^2 + s^2)^{\gamma/2}), \quad r, s \to 0. \tag{2.4}$$

A function C_γ satisfying the condition (2.3) is said to be homogeneous of order γ. The index γ determines the degree of smoothness of the volatility paths $t \to \sigma_t$. It is easy to check (see details in the supplemental material to this article available online) that $C_\gamma(r, s; t) := L(t)C_\gamma(r, s)$ is unique and satisfies the following non-negative definiteness property:

$$\iint K(x)K(y)C(x, y)dxdy \ge 0. \tag{2.5}$$

We shall see in the next section that most volatility processes that are studied in the literature satisfy Assumption 3 with a function C_γ of the form:

$$C_\gamma(r, s) = \frac{1}{2}(|r|^\gamma + |s|^\gamma - |r - s|^\gamma), \tag{2.6}$$

for some $\gamma \in [1, 2]$. The case of $\gamma = 1$ covers volatility processes driven by BM, while $\gamma \in (1, 2)$ corresponds to volatility processes driven by fractional Brownian Motions (fBM) with Hurst parameter $H > 1/2$. Deterministic and sufficiently smooth volatility processes can also be incorporated by taking $\gamma = 2$. In the following section, we cover these cases and other more general models.

Remark 2.1. We now draw some connections with the work in [10]. Therein, the variance process $\{V_t\}_{t\ge 0}$ is assumed to satisfy the following pathwise condition

$$|V_{t+\delta} - V_t| \le \tilde{L}(t, |\delta|)|\delta|^\gamma + o(|\delta|^\gamma), \quad \delta \to 0, \tag{2.7}$$

where $\delta \to \tilde{L}(t, \delta)$ is a slowly varying random function. Under this condition, Kristensen [10] shows, via a central limit theorem, that the rate of convergence of the kernel estimator is $O_P(n^{-\gamma/(1+\gamma)})$. To gain some intuition about the usefulness of this approach, let us suppose that $\{V_t\}$ is a Brownian motion. In that case, the above holds for all $\gamma < 1/2$, but such choices of γ can only produce suboptimal convergence rate of the kernel estimator. Furthermore, in light of Lévy's modulus of continuity, the condition (2.7) holds for $\gamma = 1/2$, but only if $\tilde{L}(t, \delta) \to \infty$, as $\delta \to 0$. But, in that case, the optimal bandwidth selection formulas proposed by Kristensen [10] are not well defined as they require that $\lim_{\delta\to 0} \tilde{L}(t, \delta) =: \tilde{L}(t, 0)$ is finite.

Finally, we introduce the assumptions needed on the kernel function.

Assumption 4. Given $\gamma > 0$ and C_γ as defined in Assumption 3, we assume that the kernel function $K : \mathbb{R} \to \mathbb{R}$ satisfies the following conditions:

(1) $\int K(x)dx = 1$;
(2) K is Lipschitz and piecewise C^1 on its support (A, B), where $-\infty \le A < 0 < B \le \infty$;

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

(3) (i) $\int |K(x)||x|^{\gamma} dx < \infty$; (ii) $K(x)x^{\gamma+1} \to 0$, as $|x| \to \infty$; (iii) $\int |K'(x)|dx < \infty$, (iv) $V_{-\infty}^{\infty}(|K'|) < \infty$, where $V_{-\infty}^{\infty}(\cdot)$ is the total variation;

(4) $\iint K(x)K(y)C_{\gamma}(x, y)dxdy > 0$.

2.1. Common volatility processes

In this subsection, we demonstrate that many volatility processes studied in the literature satisfy Assumption 3. We consider three fundamental cases. The proofs of the results in this part are relatively simple and for the sake of space are deferred to the supplemental material to this article available online. Let us start by considering the solutions of a standard SDE driven by BM, which are widely used in practice.

Proposition 2.1. *Suppose that the process $V_t = \sigma^2(t, \omega)$ satisfies the SDE*

$$dV_t = f(t, \omega)dt + g(t, \omega)dW_t, \quad t \in [0, T], \tag{2.8}$$

where $\{W_t\}_{t\geq 0}$ is a standard Wiener process adapted to $\mathbb{F}$. Assume that $f(t, \omega)$ and $g(t, \omega)$ are adapted and progressively measurable with respect to $\mathbb{F}$, $\mathbb{E}\left[f^2(t, \omega)\right] < M$, for $t \in [0, T]$, and $\mathbb{E}\left[g^2(t, \omega)\right]$ is continuous for $t \in [0, T]$. Then, Assumption 3 is satisfied with $\gamma = 1$, $C_1(r, s) = \min\{|r|, |s|\}1_{\{rs\geq 0\}}$, and $L(t) = \mathbb{E}[g^2(t, \omega)]$. Furthermore, $C_1(r, s)$ is an integrable positive definite function; i.e., we have strict inequality in (2.5) for all $K : \mathbb{R} \to \mathbb{R}$ such that $\int |K(x)|dx > 0$.

Next, we show that some processes defined as integrals with respect to a two-sided fBM $B^{(H)} = \{B_t^{(H)} : t \in \mathbb{R}\}$ (see [14] for a detailed survey of fBM) satisfy Assumption 3. A prototypical example is the fractional Ornstein–Uhlenbeck process $Y_t^{(H)} = \sigma \int_{-\infty}^{t} e^{-\lambda(t-u)} dB_u^{(H)}$, which is frequently used to model volatility processes. It is worth mentioning that, when $H \neq 1/2$, the fBM is not a semimartingale and the problem of defining the stochastic integral with respect to fBM is more subtle. In our paper, we only focus on integrals of deterministic functions f for which the integral can be defined on a path-wise sense under the following condition (cf. [14]):

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(u)f(v)||u - v|^{2H-2} dudv < \infty. \tag{2.9}$$

Proposition 2.2. *Let $Y_t^{(H)} = \int_{-\infty}^{t} f(u)dB_u^{(H)}$ where $f(\cdot)$ is a deterministic continuous function that satisfies (2.9) and $\{B_t^{(H)}\}_{t\in\mathbb{R}}$ is a (two-sided) fBM with Hurst parameter $H \in (\frac{1}{2}, 1)$ defined on a filtered probability space $(\Omega, \mathscr{F}, \mathbb{F} = \{\mathscr{F}_t\}_{t\geq 0}, P)$. Then, the processes $Y^{(H)}$ and $\{\exp(Y_t^{(H)})\}_{t\geq 0}$ satisfy Assumption 3 with $\gamma = 2H \in (1, 2)$ and C_γ given by (2.6).*

For our final case, we show that if a Gaussian process satisfies Assumption 3, so does a suitable smooth function of the process.

Proposition 2.3. *Assume that $(Z_t)_{t\geq 0}$ is a Gaussian process that satisfies Assumption 3 uniformly over $(0, T)$,[2] with $\gamma^{(Z)} \in [1, 2)$, $L(\cdot)$, and $C_{\gamma}^{(Z)}(\cdot, \cdot)$ defined as in (2.4). For each fixed $\tau \in (0, T)$ and a function $f \in C^2(\mathbb{R})$, further assume the following:*

[2] The Assumption 3 is satisfied uniformly over $(0, T)$ if $\sup_{\tau\in(0,T)}(r^2 + s^2)^{-\gamma/2}|\mathbb{E}[(V_{\tau+r} - V_\tau)(V_{\tau+s} - V_\tau)] - L(\tau)C_\gamma(r, s)| \to 0$, as $r, s \to 0$, and, also, $\sup_{\tau\in(0,T)} |L(\tau)| < \infty$. This implies the existence of a positive constant C such that $\mathbb{E}[(Z_t - Z_s)^2] \leq C|t - s|^{\gamma}$, for all $t, s \in (0, T)$.

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

(a) $\mathbb{E}[(Z_{\tau+r} - Z_\tau)Z_\tau] = O(|r|)$, $\mathbb{E}[Z_{\tau+r}] - \mathbb{E}[Z_\tau] = O(|r|)$, *as* $r \to 0$.
(b) $\mathbb{E}[(f'(Z_\tau))^4] < \infty$, $\mathbb{E}[\sup_{t\in(\tau-\epsilon,\tau+\epsilon)}(f''(Z_t))^4] < \infty$ *for some* $\epsilon > 0$.

Then, the process $V_t := f(Z_t)$, $t \geq 0$, *satisfies Assumption 3 with* $\gamma^{(V)} = \gamma$, $L(t) = \mathbb{E}[(f'(Z_t))^2]L_Z(t)$, *and* $C_\gamma^{(V)} = C_\gamma^{(Z)}$.

3. MSE decomposition and bandwidth selection

In this section, we first deduce an explicit leading order approximation (up to $O(\frac{\Delta}{h})$ and $O(h^\gamma)$ terms) of the MSE of the estimator. In what follows, we omit n in the notations Δ_n, h_n, and $\hat{\sigma}^2_{\tau,n}$. The proof is deferred to Appendix A.

Theorem 3.1. *For the model (2.1) with* μ *and* σ *satisfying Assumptions 1–3, and a kernel function* K *satisfying Assumption 4, let*

$$MSE^a_{\tau,n,h} := 2\frac{\Delta}{h}\mathbb{E}[\sigma_\tau^4]\|K\|_2^2 + h^\gamma L(\tau)\iint K(x)K(y)C_\gamma(x,y)dxdy. \tag{3.1}$$

Then, for any $\tau \in (0,T)$ *and* $\Delta, h \to 0$ *such that* $\Delta/h \to 0$, *we have:*

$$MSE_{\tau,n,h} = \mathbb{E}[(\hat{\sigma}^2_\tau - \sigma^2_\tau)^2] = MSE^a_{\tau,n,h} + o\left(\frac{\Delta}{h}\right) + o\left(h^\gamma\right). \tag{3.2}$$

It is not hard to see from the proof of the previous result that all $o(\cdot)$ terms are uniform on $\tau \in (0,T)$ if the condition given by (2.4) is satisfied uniformly in t. Then, we readily get the following:

Corollary 3.1. *For* $0 < a < b < T$, *let*

$$MSE^a_{n,h}(a,b) := 2\frac{\Delta}{h}\int_a^b \mathbb{E}[\sigma_t^4]dt\|K\|_2^2 + h^\gamma \int_a^b L(t)dt \iint K(x)K(y)C_\gamma(x,y)dxdy. \tag{3.3}$$

Then, for the model (2.1)) with μ *and* σ *satisfying Assumptions 1–3, so that the term* $o((r^2+s^2)^{\gamma/2})$ *in Eq. (2.4) is uniform in* t, *and a kernel function* K *satisfying Assumption 4, we have*

$$IMSE_{n,h} := \int_a^b \mathbb{E}[(\hat{\sigma}^2_t - \sigma^2_t)^2]dt = MSE^a_{n,h}(a,b) + o\left(\frac{\Delta}{h}\right) + o(h^\gamma). \tag{3.4}$$

Based on the approximations above, it is natural to analyze the behavior of the approximated MSE of the kernel estimator. We focus on the integrated MSE (3.4) but an analogous analysis can be made for the local MSE (3.2). Note that, by Assumption 4, we have that $\iint K(x)K(y)C_\gamma(x,y)dxdy > 0$. We then obtain the following:

Proposition 3.1. *With the same assumptions as Corollary 3.1, the approximated optimal homogeneous bandwidth, denoted by* $\bar{h}_n^{a,opt}$, *which is defined to minimize the approximated IMSE given by (3.3), is given by*

$$\bar{h}_n^{a,opt} = n^{-1/(\gamma+1)}\left[\frac{2T\int_a^b \mathbb{E}[\sigma_t^4]dt\int K^2(x)dx}{\gamma\int_a^b L(t)dt\iint K(x)K(y)C_\gamma(x,y)dxdy}\right]^{1/(\gamma+1)}, \tag{3.5}$$

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

while the attained minimum of the approximated IMSE is given by

$$IMSE_n^{a,opt}(a,b) = n^{-\gamma/(1+\gamma)}\left(1+\frac{1}{\gamma}\right)\left(2T\int_a^b \mathbb{E}[\sigma_t^4]dt\int K^2(x)dx\right)^{\gamma/(1+\gamma)} \times\left(\gamma\int_a^b L(t)dt\iint K(x)K(y)C_\gamma(x,y)dxdy\right)^{1/(1+\gamma)}. \tag{3.6}$$

A direct consequence of the previous result is the following proposition about the optimal convergence rate.

Proposition 3.2. *With the same assumptions as those in Corollary 3.1, the optimal convergence rate of the kernel estimator is given by $n^{-\gamma/(1+\gamma)}$. This is attainable if the bandwidth is chosen as $h_n = cn^{-1/(\gamma+1)}$ for some constant $c \in (0,\infty)$.*

An important problem is to formalize the connection between the approximate optimal bandwidth $\bar{h}_n^{a,opt}$ (respectively, $h_n^{a,opt}$), which is defined as the minimizer of the MSE (3.3) (respectively, (3.1)), and the "true" optimal bandwidth, whenever it exists, which is denoted by $\bar{h}_n^*$ (respectively, h_n^*) and is defined as a value of the bandwidth that minimizes the actual IMSE (respectively, MSE) of the kernel estimator. In the supplemental material to this article available online, we show that, under a mild additional condition, they are equivalent in the sense that $\bar{h}_n^* = \bar{h}_n^{a,opt} + o(\bar{h}_n^{a,opt})$ and $h_n^* = h_n^{a,opt} + o(h_n^{a,opt})$.

4. Kernel function selection

As an important application of the optimal bandwidth selection problem defined in Section 3, we now study the problem of selecting an optimal kernel function by minimizing the optimal IMSE attained by (3.5). As shown therein, the optimal kernel function only depends on the covariance structure, $C_\gamma(\cdot,\cdot)$. There are two possible situations. The first one is when C_γ is positive definite. In such a case, we cannot improve the rate of convergence of the IMSE, but we can attempt to minimize the constant appearing before the asymptotics of the IMSE in (3.6) or, equivalently, minimize the functional:

$$I(K) = \left(\int K^2(x)dx\right)^\gamma \iint K(x)K(y)C_\gamma(x,y)dxdy. \tag{4.1}$$

Another situation is when C_γ is simply non-negative definite. In such a case, if we relax (4) of Assumption 4, it is possible to improve the rate of convergence of the IMSE by choosing a so-called "higher order" kernel function. An important instance of this case is when the volatility is deterministic and sufficiently smooth (see Remark 4.1 for more information).

In this section, we focus on the covariance function C_γ defined in Eq. (2.6) with $\gamma < 2$, which is actually positive definite. This is because C_γ admits the integral form $C_\gamma(x,y) = \int F_\gamma(x,u)F_\gamma(y,u)du$ with

$$F_\gamma(x,y) := m\left(|x-y|^{\frac{\gamma-1}{2}}\operatorname{sgn}(x-y) + |y|^{\frac{\gamma-1}{2}}\operatorname{sgn}(y)\right),$$

and a certain constant m (see [12] for details). We can then easily check that $\int\int K(x)K(y) C_\gamma(x,y)dx = \int(\int K(x)F_\gamma(x,u)dx)^2du > 0$, for an arbitrary nonzero kernel function $K : \mathbb{R}\to\mathbb{R}$. Furthermore, it also follows that its symmetrization, $K_s(x) := (K(x)+K(-x))/2$, is such that

$$\iint K(x)K(y)C_\gamma(x,y)dxdy - \iint K_s(x)K_s(y)C_\gamma(x,y)dxdy \geq 0. \tag{4.2}$$

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

The previous relation implies that in order to minimize the constant appearing before the asymptotic IMSE in (3.5), it suffices to consider symmetric kernel functions K.

Remark 4.1. In the accompanying material to this article available online, we give some new results regarding optimal kernel selection for smooth deterministic volatilities. Concretely, by using the calculus of variation with constraints, we obtain optimal kernel functions of higher orders. The second order optimal kernel is exactly that of Epanechnikov [4] kernel, while, for higher order cases, we provide ways to calculate those optimal kernel functions.

4.1. Optimal kernel selection for BM driven volatilities

Consider a BM type volatility with $\gamma = 1$ and $C_1(r,s) = 1_{\{rs>0\}} \min(|r|,|s|)$. We will show that the exponential kernel function is the optimal kernel function. Foster and Nelson [6] argued that this is the case, but their proof lacks rigor, due to their bounded support assumption on the kernel function.

From (4.1) and the relation (4.2), the objective function that we need to minimize is

$$\int_0^\infty K^2(x)dx \int_0^\infty \int_0^\infty K(x)K(y)\min(x,y)dxdy.$$

In terms of $U(x) := \int_x^\infty K(y)dy$, we can write this as

$$I^*(U) := \int_0^\infty [U'(x)]^2 dx \int_0^\infty [U(x)]^2 dx. \tag{4.3}$$

The problem is then changed to minimize $I^*(U)$ for functions U that are continuous and piecewise twice differentiable on $\mathbb{R}_+$ such that $U(0) = \frac{1}{2}$ and $\lim_{x\to+\infty} U(x) = 0$. Next, using Cauchy–Schwarz inequality, note that

$$I^*(U) \geq \left(\int_0^\infty U'(x)U(x)dx\right)^2 = \left(\int_0^\infty U(x)dU(x)\right)^2 = \left(\int_{1/2}^0 udu\right)^2 = \frac{1}{64},$$

where the first inequality becomes equality if and only if there exist non-zero constants C_1 and C_2 such that $C_1U'(x) + C_2U(x) \equiv 0$, for all $x \in \mathbb{R}_+$. We have two possible cases: (1) there exists $x_0 > 0$, such that $U(x) > 0$, for all $x \in [0, x_0)$ and $U(x_0) = 0$; (2) $U(x) > 0$, for all $x \in \mathbb{R}_+$. For the first case, we have that $U'(x)/U(x) = -C_2/C_1$, for $x \in (0, x_0)$, whose solution is $U(x) = \frac{1}{2}e^{Bx}$ and it is then impossible that $U(x_0) = 0$. Therefore, only the second case is possible and, by solving the same differential equation, we have the following.

Theorem 4.1. *For the model (2.1) with μ and σ satisfying Assumptions 1 and 3, where C_γ is given by (2.6) with $\gamma = 1$, and for a kernel function K satisfying 4, we have that the optimal kernel function that minimizes the first order approximation of the IMSE of the kernel estimator is the exponential kernel function $K^{exp}(x) = \frac{1}{2}\exp(-|x|)$.*

Remark 4.2. We can easily demonstrate to what extent the exponential kernel decreases the MSE. As seen from (3.6), $IMSE_n^{a,opt} = C\sqrt{I^*(K)}$, where the constant C does not depend on the kernel function K. Below, we show the value of $I^*(K) := I^*(U)$ for the exponential, uniform, triangular, and the Epanechnikov kernels:

$$I^*(.5\,e^{-|x|}) = \frac{1}{72} \approx 0.0138, \quad I^*(.5\,1_{\{|x|<1\}}) = \frac{1}{24} \approx 0.0416,$$
$$I^*(|1-x|1_{\{|x|<1\}}) = \frac{1}{30} \approx 0.0333, \quad I^*(.75\,(1-x^2)1_{\{|x|<1\}}) = \frac{297}{8240} \approx 0.036.$$

In Section 8 of the supplemental material to this article available online, we show some Monte Carlo experiments to illustrate the superior performance of the exponential kernel.

Let us finish by noting that the exponential kernel function not only minimizes the MSE of the kernel estimator, but also enables us to substantially reduce the computational complexity of the volatility estimation. The idea is using the decomposition

$$\hat{\sigma}^2_{\tau,exp} := \sum_{i=1}^{n} K_h^{exp}(t_{i-1}-\tau)(\Delta_i X)^2 := \hat{\sigma}^2_{\tau,+} + \hat{\sigma}^2_{\tau,*} + \hat{\sigma}^2_{\tau,-}, \tag{4.4}$$

where, fixing i_0 such that $t_{i_0-1} \le \tau < t_{i_0}$,

$$\begin{aligned} \hat{\sigma}^2_{\tau,-} &= \sum_{i<i_0} K_h^{exp}(t_{i-1}-\tau)(\Delta_i X)^2, \\ \hat{\sigma}^2_{\tau,*} &= K_h^{exp}(t_{i_0-1}-\tau)(\Delta_{i_0} X)^2, \\ \hat{\sigma}^2_{\tau,+} &= \sum_{i>i_0} K_h^{exp}(t_{i-1}-\tau)(\Delta_i X)^2. \end{aligned} \tag{4.5}$$

The computational reduction arises from the fact that $\hat{\sigma}^2_{\tau,-}$ and $\hat{\sigma}^2_{\tau,+}$ can actually be computed iteratively as follows:

$$\begin{aligned} \hat{\sigma}^2_{\tau+\Delta,-} &= e^{-\Delta/h}\left[\hat{\sigma}^2_{\tau,-} + K_h^{exp}(t_{i_0-1}-\tau)(\Delta_{i_0}X)^2\right], \\ \hat{\sigma}^2_{\tau+\Delta,+} &= e^{\Delta/h}\left[\hat{\sigma}^2_{\tau,+} - K_h^{exp}(t_{i_0}-\tau)(\Delta_{i_0+1}X)^2\right]. \end{aligned} \tag{4.6}$$

It is now clear that, in order to estimate $\{\sigma_{t_i}\}_{i=0,\ldots,n}$, using an exponential kernel, we need a time of $O(n)$, instead of the orders $O(n^2)$ or $O(n^2h)$ needed for general kernels of unbounded or bounded support, respectively.

In practice, kernel estimators suffer of biases at times closer to the boundary. As proposed in [10], this can be corrected by using the following estimator:

$$\hat{\sigma}^b_{\tau,n,h} = \frac{\sum_{i=1}^n K_h(t_{i-1}-\tau)(\Delta_i^n X)^2}{\Delta\sum_{i=1}^n K_h(t_{i-1}-\tau)}. \tag{4.7}$$

where the superscript denotes boundary effect. The denominator above can still be efficiently calculated similarly as (4.4) except that all $(\Delta_i X)^2$ are replaced by 1.

4.2. *Optimal kernel function for a fBM driven volatility*

In this section, we now consider a general fBM covariance structure, i.e. $\gamma \in (1,2)$ and C_γ given by (2.6). From (4.1) and the relation (4.2), and since $C_\gamma(x,y) + C_\gamma(x,-y) = |x|^\gamma + |y|^\gamma - \frac{1}{2}|x+y|^\gamma - \frac{1}{2}|x-y|^\gamma$ for $x, y > 0$, our goal is to minimize

$$I^*(K) = \left(\int_0^\infty K^2(x)dx\right)^\gamma \int_0^\infty \int_0^\infty K(x)K(y)A(x,y)dxdy. \tag{4.8}$$

where $A(x,y) = |x|^\gamma + |y|^\gamma - \frac{1}{2}|x+y|^\gamma - \frac{1}{2}|x-y|^\gamma$. Unfortunately, the problem of solving the calculus of variation problem associated with (4.8) and finding an explicit form for the optimal kernel function is more challenging. Therefore, we instead seek a numerical method to find the optimal kernel function, for which, we consider a two-step approximation procedure. First, since all unbounded support kernels can be approximated by a kernel with bounded support and the optimization problem is unchanged with $K(x)$ scaled by a small bandwidth, we will limit the support of $K(x)$ to be [0, 1]. Second, we approximate the kernel function K by step functions of the form $K_m(x) = \frac{1}{\sum_{i=1}^n a_i} \sum_{i=1}^m a_i 1_{[\frac{i-1}{m}, \frac{i}{m})}(x)$, with $x \in [0, 1]$, $a_i \in \mathbb{R}$ $(i = 1, \ldots, n)$, as well as approximate $\int_{(i-1)/m}^{i/m} \int_{(j-1)/m}^{j/m} A(x, y)dxdy$ with $A((i-0.5)/m, (j-0.5)/m) := A_{ij}$. Using the just described approximation, we seek to minimize:

$$f(a) = m^{\gamma} \left(\sum_{i=1}^m a_i^2 \right)^{\gamma} \left(\sum_{i=1}^m \sum_{j=1}^m a_i a_j A_{ij} \right) \left(\sum_{i=1}^m a_i \right)^{-2\gamma-2}, \tag{4.9}$$

over all valid values of $(a_1, \ldots, a_m)$, for which we use gradient descent. In spite of the high dimensionality of the optimization problem, this is still tractable, since the gradient can be calculated explicitly as

$$\begin{aligned} \frac{\partial f}{\partial a_i} &= C(\overline{a})^{2\gamma+2} \left(\overline{a^2}\right)^{\gamma-1} \left[2a_i\gamma \left(\sum_{i=1}^n \sum_{j=1}^n a_i a_j A_{ij} \right) + \overline{a^2} \left(2 \sum_{j=1}^n a_j A_{ij} \right) \right] \\ &\quad - (2\gamma+2)\, (\overline{a})^{2\gamma+1} \left(\overline{a^2}\right)^{\gamma} \left(\sum_{i=1}^n \sum_{j=1}^n a_i a_j A_{ij} \right), \end{aligned}$$

where $\overline{a} := \sum_{i=1}^n a_i/n$, $\overline{a^2} := \sum_{i=1}^n a_i^2/n$, and C is a constant that depends on n but not on the a_i's.

Fig. 1 shows the resulting optimal kernels for $\gamma = 1.0, 1.3, 1.6, 1.9$. Note that the resulting approximated optimal kernel for $\gamma = 1$ is consistent with true optimal kernel that was proved to be exponential in Section 4.1. We also observe from Fig. 1 that, as γ increases, the optimal kernel function becomes flatter and less convex. This indeed makes sense, since a higher γ indicates less chaos of the volatility, and thus more weights should be given to data farther from the estimated point.

5. Plug-in bandwidth selection methods

In this section we propose a feasible plug-in type bandwidth selection algorithm, for which, as a sub-problem, we also develop a new estimator of the volatility of volatility based on the kernel estimator of the spot volatility and a type of two-time scale realized variance estimator. We shall focus on the case of a BM type volatility as described in Proposition 2.1, while similar methods can be developed for other types of volatility structures. To implement the approximated optimal bandwidth formula (3.5), it is natural to estimate $\int_0^T \mathbb{E}[\sigma_t^4]dt$ and $\int_a^b L(t)dt = \int_a^b \mathbb{E}[g^2(t)]dt$ with the integrated quarticity of X, $IQ(X) = \int_0^T \sigma_\tau^4 d\tau$, and the quadratic variation of σ^2, $IV(\sigma^2) = \int_0^T g^2(\tau)d\tau$. A popular estimate for $\int_0^T \sigma_\tau^4 d\tau$ is the realized quarticity, which is defined by $\widehat{IQ} = (3\Delta)^{-1} \sum_{i=1}^n (\Delta_i X)^4$. The estimation of $\int_0^T g^2(\tau)d\tau$ is a more subtle problem and, below, we propose an estimator, which is termed the Two-time Scale Realized Volatility of Volatility (TSRVV) and is hereafter denoted by $\widehat{IV(\sigma^2)}_{(\text{tsrvv})}$. With these estimators, the final bandwidth can then be written as

$$h_n^{a,opt} = \left[\frac{2T\,\widehat{IQ(X)} \int K^2(x)dx}{n\widehat{IV(\sigma^2)}_{(\text{tsrvv})} \iint K(x)K(y)C_1(x, y)dxdy} \right]^{1/2}. \tag{5.1}$$

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

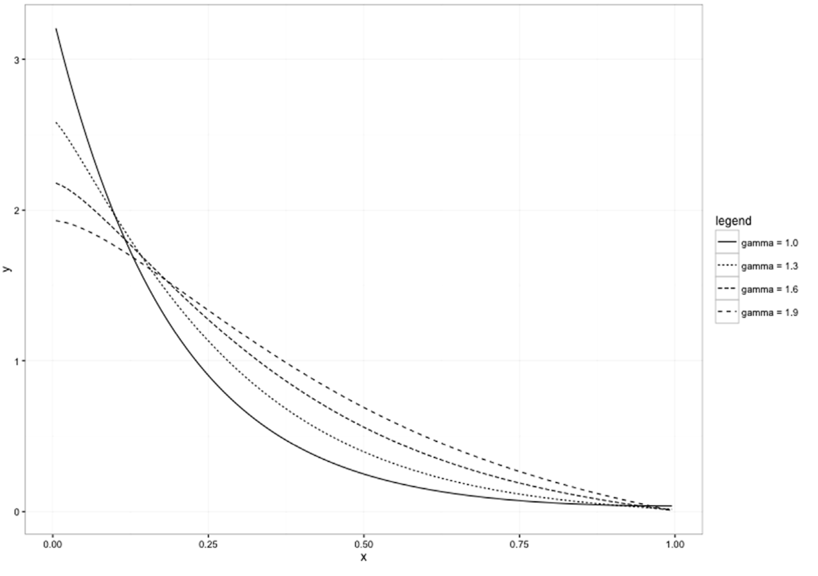

Fig. 1. Optimal Kernel Functions for Different γ.

The previous bandwidth estimator involves the spot volatility itself, through $\widehat{IV(\sigma^2)}_{(\text{tsrvv})}$, which, of course, we do not know in advance. To deal with this problem, we propose to use an iterative algorithm in the same spirit of a fixed-point type of procedure. Concretely, we start with an initial "guess" for the bandwidth such as

$$h_n^{init} = \left[\frac{2T \int K^2(x)dx}{n \iint K(x)K(y)C_1(x,y)dxdy} \right]^{1/2}. \tag{5.2}$$

With such a bandwidth, we can obtain initial estimates of the spot volatility at all the grid points. Such an initial spot volatility estimation can then be applied to compute $\widehat{IV(\sigma^2)}_{(\text{tsrvv})}$, which, in turn, can be used to obtain another estimation of the optimal bandwidth. This procedure is continued iteratively until a predetermined stopping criterion is met. Our simulations show that one or two iterations are typically enough.

We are now ready to define our estimator $\widehat{IV(\sigma^2)}_{(\text{tsrvv})}$ of $IV(\sigma^2) = \int_0^T g^2(\tau)d\tau$, which is often referred to as the Integrated Volatility of Volatility (IVV) of X. The idea is to note that, at each observation time t_i, the estimated spot volatility can be written as $\hat{\sigma}_{t_i}^2 = \sigma_{t_i}^2 + e_{t_i}$, where e_{t_i} is the estimation error. This suggests to make an analogy with the problem of estimating the realized quadratic variation of a semimartingale Y based on discrete observations of Y exposed to market microstructure. So, we can apply any of the different techniques to tackle this problem such as the Two-time Scale Realized Volatility (TSRV) estimator of Zhang et al. [16]. However, note that, unlike the problem in [16], our estimation errors are correlated and such a correlation becomes more significant when we take the difference $\Delta_i \hat{\sigma}^2 = \hat{\sigma}_{t_{i+1}}^2 - \hat{\sigma}_{t_i}^2$. To alleviate such a problem, we propose to use one-sided kernel estimators and take the difference between the right and left side estimators to find $\Delta_i \hat{\sigma}^2$. Concretely, let $\hat{\sigma}_{l,t_i}^2$ and $\hat{\sigma}_{r,t_i}^2$ be the left- and

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

right-side estimator of $\sigma^2_{t_i}$, respectively, defined as

$$\hat{\sigma}^2_{l,t_i} = \frac{\sum_{j\le i} K_h(t_{j-1}-t_i)(\Delta^n_j X)^2}{\Delta \sum_{j\le i} K_h(t_{j-1}-\tau)}, \quad \hat{\sigma}^2_{r,t_i} = \frac{\sum_{j>i} K_h(t_{j-1}-t_i)(\Delta^n_j X)^2}{\Delta \sum_{j>i} K_h(t_{j-1}-\tau)}. \tag{5.3}$$

Next, we define the following two difference terms: $\Delta_i\hat{\sigma}^2 = \hat{\sigma}^2_{r,t_{i+1}} - \hat{\sigma}^2_{l,t_i}$, $\Delta^{(k)}_i\hat{\sigma}^2 = \hat{\sigma}^2_{r,t_{i+k}} - \hat{\sigma}^2_{l,t_i}$. Finally, we can construct the estimator

$$\widehat{IV(\sigma^2)}_{(\text{tsrvv})} = \frac{1}{k}\sum_{i=b}^{n-k-b}(\Delta^{(k)}_i\hat{\sigma}^2)^2 - \frac{n-k+1}{nk}\sum_{i=b+k-1}^{n-k-b}(\Delta_i\hat{\sigma}^2)^2. \tag{5.4}$$

Here, b is a small enough integer, when compared to n. The purpose of introducing such a number b is to alleviate the boundary effect of the one sided estimators. More specifically, since we are using left- and right-side estimators of the spot volatility, we are not able to estimate, for example, $\hat{\sigma}^2_{l,t_0}$ and $\hat{\sigma}^2_{r,t_n}$. Therefore, in practice, we suggest to select an appropriate b to avoid such a problem. Theoretically, we will establish our asymptotic properties to the estimator of $\int_{t_b}^{T-t_b} g^2(\tau)d\tau$ for some small but fixed $t_b \in (0, T/2)$. Similar to Zhang et al. [16], we can take $k = n^{2/3}$ in our case. There is some work to do if one wants to optimize such a TSRVV estimator, by selecting better b and k to improve the convergence, but this is outside the scope of the present work.

The result below shows the consistency of (5.4) and shed some light on its rate of convergence. Its proof is provided in Appendix A.

Theorem 5.1. *Fix a $t_b \in (0, T/2)$. Then, for the model (2.1) with μ and σ satisfying Assumptions 1 and 2 and σ being a squared integrable Itô process as in Eq. (2.8) (thus satisfying Assumption 3), and a kernel function K satisfying Assumption 4, (5.4) is a consistent estimator of $\int_{t_b}^{T-t_b} g^2_t dt$ with $b = t_b/\Delta$. Furthermore, the convergence rate is given by $O_p(\frac{n^{1/4}}{k^{1/2}}) + O_p(\sqrt{\frac{k}{n}})$.*

Remark 5.1. Vetter [15] proposed a similar estimator for the IVV, but taking a right-sided *uniform* kernel when computing the difference $\Delta_i\hat{\sigma}^2$ of the estimated volatility and also applying a different bias correction technique from ours. It is shown therein that his estimator attains the optimal rate of convergence of $n^{-1/4}$. Simulations, that are not shown here for the sake of space, indicate that our TSRVV using the optimal exponential kernel has better performance than Vetter [15] at least for the chosen parameter choices. This suggests that there may be some room for improvement of the convergence rate stated in Theorem 5.1, which is just $O(n^{-1/8})$. On the other hand, the observed improved performance of our TSRVV may also be a consequence of the fact that we are using an exponential kernel, while the estimator in [15] uses the suboptimal uniform kernel.

To conclude, we summarize the proposed plug-in type implementation of the kernel-based spot volatility estimation. First, using $h = h^{init}_n$ as defined in (5.2), we compute the left- and right-side estimators of $\sigma^2_{t_i}$ as described in Eq. (5.3), at all grid point t_i. These are then used to estimate $IV(\sigma^2) = \int_0^T g^2(\tau)d\tau$ via (5.4). This estimator is then plugged in (5.1) to obtain an updated estimate of the bandwidth $h^{a,opt}_n$, which can again be used in (5.3)–(5.4). This procedure continues until, e.g., the value of h or $\widehat{IV(\sigma^2)}_{(\text{tsrvv})}$ do not change much. Once we reached "convergence", we use (4.7) to estimate $\hat{\sigma}_{t_i}$ with the final value of h.

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

6. Central limit theorems

In this section, we aim to characterize the limiting distribution of the estimation error of the kernel estimator by proving a Central Limit Theorem (CLT). All the proofs are given in Appendix A.

To motivate the discussion below, let us start by noting the following natural decomposition:

$$\begin{aligned}\hat{\sigma}_\tau^2 - \sigma_\tau^2 = &\left(\sum_{i=1}^{n} K_h(t_{i-1} - \tau)(\Delta_i X)^2 - \int_0^T K_h(t-\tau)\sigma_t^2 dt\right) \\ &+ \int_0^T K_h(t-\tau)(\sigma_t^2 - \sigma_\tau^2)dt + o_p\left(h^\gamma\right),\end{aligned} \tag{6.1}$$

where the last term on the right-hand side above follows from Assumption 4. Two general type of results can be found in the literature to deal with the estimation error:

(1) One approach consists of using a 'suboptimal' bandwidth so that the first error term in (6.1), which, as shown below, is of order $O_p((\Delta/h)^{1/2})$, dominates the second term, whose order is $O_p(h^{\gamma/2})$. This would be the case if, for instance, we choose $h = o(\Delta^{1/(\gamma+1)})$. Instances of this type of results can be found in [5,10,11].

(2) In the case that σ_t^2 follows an Itô process, Foster and Nelson [6] obtained a CLT for the kernel estimator $\hat{\sigma}_\tau^2$ with optimal convergence rate but under a number of stringent conditions. In particular, only kernels with bounded support were considered. More recently, under relatively mild assumptions in the Itô dynamics of X and σ, Alvarez et al. [2] obtained a CLT with optimal convergence rate but only for the forward uniform kernel function $K(x) = \mathbf{1}_{[0,1]}(x)$. Jacod and Protter [7] was able to obtain the same type of results for both forward and backward uniform kernels: $K(x) = \mathbf{1}_{[0,1]}(x)$ or $K(x) = \mathbf{1}_{[-1,0]}(x)$.

The two previous approaches have some obvious limitations. The first approach can only yield results with suboptimal convergence rates, while the second type of results only deal with one level of smoothness in the volatility process and uniform one-sided kernels. In this section, we obtain a CLT with optimal convergence rate in two broad frameworks: (i) Itô type volatilities and (ii) deterministic functions of certain Gaussian processes. These cover all the examples mentioned in Section 2.1. For the framework (i), we consider two cases: (1) A general kernel but under the no leverage Assumption 1; (2) Leverage but only forward looking kernel as in [2], even though the latter work only considers uniform kernels, while we consider here a general forward-looking kernel function. The second framework (ii) covers a wide range of models of different smoothness levels, though without leverage. In what follows, we replace Assumption 1 with the following:

Assumption 5. The processes μ and σ are adapted cádlág.

We begin with an analysis of the first error term in (6.1), which, in the nonleverage case, was already studied in [10]. Mancini et al. [11] (see Theorem 2.7 therein) also analyzed this term, but, since the proof in [11] is for a more general class of estimators and requires more technical analysis, we give a simpler proof in Appendix A.

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

Theorem 6.1. *For the model (2.1) with μ and σ satisfying Assumption 5, and a kernel function K satisfying Assumption 4, we have, for any $\tau \in (0, T)$,*

$$\left(\frac{\Delta}{h}\right)^{-1/2}\left[\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2 - \int_0^T K_h(t-\tau)\sigma_t^2 dt\right] \to_D \delta_1 N(0,1), \tag{6.2}$$

where $\delta_1^2 = 2\sigma_\tau^4 \int K^2(x)dx$.

Next, we consider the second error term in (6.1), which only involves properties of the volatility process σ and not the interaction between X and σ.

Theorem 6.2. *Let K be a kernel function satisfying Assumption 4 and fix a $\tau \in (0, T)$. Additionally, suppose that either one of the following conditions holds:*

(1) $\{\sigma_t^2\}_{t\geq 0}$ is an Itô process given by $\sigma_t^2 = \sigma_0^2 + \int_0^t f_s ds + \int_0^t g_s dW_s$ with adapted cádlág processes $\{f_t\}_{t\geq 0}$ and $\{g_t\}_{t\geq 0}$.

(2) $\sigma_t^2 := f(Z_t)$, $t \in [0, T]$, for a deterministic function $f : \mathbb{R} \to \mathbb{R}$ and a Gaussian process $\{Z_t\}_{t\geq 0}$ satisfying all requirements of Proposition 2.3.

Then, on an extension $(\bar{\Omega}, \bar{\mathscr{F}}, \bar{\mathbb{P}})$ of the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, equipped with a standard normal variable ξ independent of g_τ in (1) or Z_τ in (2), we have:

$$h^{-\gamma/2}\left(\int_0^T K_h(t-\tau)(\sigma_t^2 - \sigma_\tau^2)dt\right) \to_D \delta_2 \xi, \tag{6.3}$$

where, under the condition (1) above, $\delta_2^2 = g_\tau^2 \iint K(x)K(y)C_1(x,y)dxdy$, while, under the condition (2), $\delta_2^2 = [f'(Z_\tau)]^2 L^{(Z)}(\tau) \iint K(x)K(y)C_\gamma^{(Z)}(x,y)dxdy$.

As a byproduct of Theorems 6.1 and 6.2 and in accordance with our former Proposition 3.2, we deduce that the optimal convergence rate is $n^{-\gamma/(1+\gamma)}$ and that this would be attained if $h_n = c\Delta_n^{1/(\gamma+1)}$ for any constant $c \in (0, \infty)$. In that case, the following result shows a CLT for $\hat{\sigma}_\tau^2$ under the non-leverage Assumption 1.

Corollary 6.1. *Suppose the assumptions of Theorems 6.1 and 6.2 are satisfied as well as the nonleverage Assumption 1. Then, for the bandwidth selection $h_n = \Delta^{1/(\gamma+1)}$, we have $\Delta^{-\frac{\gamma}{2(1+\gamma)}}\left(\hat{\sigma}_\tau^2 - \sigma_\tau^2\right) \to_D \sqrt{\delta_1^2 + \delta_2^2}\,\bar{\xi}$, where δ_1 and δ_2 are defined in Theorems 6.1 and 6.2, respectively, and $\bar{\xi}$ is a standard normal random variable independent from g_τ, under the setting (1) of Theorem 6.2, or from Z_τ under the setting (2) of Theorem 6.2.*

Our final result is a CLT when $h_n = cn^{-1/(\gamma+1)}$ for general Itô volatilities (as in the setting (1) of Theorem 6.2), but only forward looking kernels. This generalizes results of Alvarez et al. [2], where only uniform forward kernels were analyzed.

Theorem 6.3. *Consider the model (2.1) with a cádlag process μ and an Itô process σ given by $\sigma_t^2 = \sigma_0^2 + \int_0^t f_s ds + \int_0^t g_s dW_s$, where W is a Brownian motion such that $\mathbb{E}(dB_t \cdot dW_t) = \rho dt$ and $\{f_t\}_{t\geq 0}$ and $\{g_t\}_{t\geq 0}$ are adapted cádlág processes. Let K be a kernel function satisfying Assumption 4 and, in addition, $K(x) = 0$ for all $x < 0$. Then, the conclusion of Corollary 6.1 holds true with $\gamma = 1$.*

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

7. Simulation results

In this section, we show some simulations to further investigate the performance of the plug-in method that we developed in Sections 3 and 5 and compare it with the cross-validation method proposed in [10]. Throughout, we will consider the Heston model:

$$dX_t = \mu_t dt + \sqrt{V_t} dB_t, \quad dV_t = \kappa(\theta - V_t)dt + \xi\sqrt{V_t} dW_t, \tag{7.1}$$

where $V_t = \sigma_t^2$ is the variance process. Regarding the parameters values, we adopt the setting used in [16]:

$$\kappa = 5, \quad \theta = 0.04, \quad \xi = 0.5, \quad \mu_t = 0.05 - V_t/2.$$

The initial values are set to be $X_0 = 1$ and $\sigma_0^2 = 0.04$. We also assume both a non-leverage setting ($\rho = 0$) and a negative leverage situation ($\rho = -0.5$) to investigate the robustness of our method against non-zero ρ values. We will consider several different sampling scenarios with 6.5 trading hours per day (the time unit is one year) and 252 trading days during the year.

In order to alleviate boundary effects, we use the estimator (4.7) throughout all the simulation. For each simulated discrete skeleton $\{X_{t_i} : 0 \le i \le n\}$, with $t_i = iT/n$, we estimate the corresponding discrete-skeleton of the variance process $\{\sigma_{t_i}^2 : 0 \le i \le n\}$, and calculate the average of the squared errors, $ASE = \frac{1}{n-2l+1} \sum_{i=l}^{n-l} (\hat{\sigma}_{t_i}^2 - \sigma_{t_i}^2)^2$, for each simulation. We use $l = [0.1n]$ to focus on evaluating the performance of the estimator without boundary effects. Then, we take the sample average of such ASE's to estimate the mean ASE, defined as $\text{MASE} = \mathbb{E}\left[\frac{1}{n-2l+1} \sum_{i=l}^{n-l} (\hat{\sigma}_{t_i}^2 - \sigma_{t_i}^2)^2\right]$.

In Table 1, we report the MASE obtained by different methods based on 2000 paths. The first column reports the performance of the plug-in method proposed in Section 5, where we use the approximated homogeneous optimal bandwidth (3.5) together with the vol vol estimator described in Eqs. (5.3)–(5.4) (we fix $b = n/10$ therein and run only two iterations after the initial initialization (5.2) of the bandwidth). In the second column, we report the results for the leave-one-out cross validation as proposed in [10]. In the third column, we give the results for an oracle plug-in method, where the true path of $\{\sigma_t\}_{t\in[0,T]}$ and ξ are used to compute $\int_0^T \sigma_t^4 dt$ and $\int_0^T g^2(t)dt = \xi^2 \int_0^T \sigma_t^2 dt$ in the formula (3.5). The final column shows a "semi-oracle" result, which only assumes the knowledge of the volatility of volatility ξ of the Heston model, but not the path of $\{\sigma_t\}_{t\in[0,T]}$, which is estimated using kernel-based estimation.

As expected, the plug-in method runs significantly faster than cross validation. As to the accuracy of the kernel estimator, simulation results show that, in almost all sampling frequencies, the plug-in method outperforms the cross-validation method. It is worth to notice that, in all cases, there is still significant loss of accuracy for the plug-in method compared to the oracle ones. From the two oracle results, it can be easily observed that such a loss of accuracy is mainly due to the estimation error of the volatility of volatility. In Section 8 of the supplemental material to this article available online, we show some Monte Carlo experiments to illustrate the performance of the vol vol estimator proposed in Section 5.

We now proceed to test the TSRVV estimator introduced in (5.3)–(5.4) with $b = n/10$. We use one month data as demonstration, and, in order to see how the estimator performs with different sampling sequence, we consider 5 min and 1 min data. Since we are considering the Heston model, we will not report the integrated volatility of volatility, but instead, we report the following estimator of IVV parameter ξ of the Heston model:

$$\widehat{\xi} := \sqrt{\frac{\widehat{IVV}^{tsrvv}}{\widehat{IV}}}.$$

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

Table 1
Comparison of Different Bandwidth Selection Methods (MASE, 2000 simulations). For the 5-days data, $T = 5/252$ (in years), while "5-min" frequency means that $\Delta = 5/(60 \cdot 6.5 \cdot 252)$ (in years), and the number of observation $n = 12 \cdot 6.5 \cdot 5 = 390$. "1 min" frequency means that $\Delta = 1/(60 \cdot 6.5 \cdot 252)$ (in years), and $n = 60 \cdot 6.5 \cdot 5 = 1950$. For 21 days (1 month) data, $T = 1/12$ (in years), in which case the number of observations is $n = 12 \cdot 6.5 \cdot 21 = 1638$ for 5-min frequency and means $\Delta = 5/(60 \cdot 6.5 \cdot 252)$ (in year), and $n = 60 \cdot 6.5 \cdot 21$.

5 days data					
Frequency	ρ	$MASE_{PI}$	$MASE_{CV}$	$MASE_{oracle}$	$MASE_{\substack{semi \\ oracle}}$
5 min	0	1.0796E−07	1.3386E−07	9.1266E−08	9.0402E−08
1 min	0	7.1439E−09	8.0542E−09	6.7286E−09	6.7074E−09
5 min	−0.5	1.0296E−07	1.4180E−07	9.2620E−08	9.2009E−08
1 min	−0.5	7.3872E−09	8.2567E−09	6.9356E−09	6.9060E−09
21 days data					
Frequency	ρ	$MASE_{PI}$	$MASE_{CV}$	$MASE_{oracle}$	$MASE_{\substack{semi \\ oracle}}$
5 min	0	1.9088E−08	2.1221E−08	1.8265E−08	1.8178E−08
1 min	0	1.7064E−09	1.6868E−09	1.5984E−09	1.5961E−09
5 min	−0.5	1.9039E−08	1.9495E−08	1.7587E−08	1.7506E−08
1 min	−0.5	1.6652E−09	1.6011E−09	1.5509E−09	1.5505E−09

Table 2
Estimation of Volatility by TSRVV (1 month data, 10000 sample paths). See the caption in Table 1 for more information about the data.

Frequency	ρ	ξ	Bias	Std	$\sqrt{MSE}$
5 min	0	0.2	−0.0006	0.0990	0.0990
5 min	0	0.5	−0.0584	0.1979	0.2063
1 min	0	0.2	−0.0122	0.0772	0.0782
1 min	0	0.5	−0.0411	0.1549	0.1603
5 min	−0.5	0.2	−0.0002	0.0987	0.0987
5 min	−0.5	0.5	−0.0571	0.1984	0.2065
1 min	−0.5	0.2	−0.0138	0.0779	0.0791
1 min	−0.5	0.5	−0.0443	0.1551	0.1613

Generally, $\xi = 0.5$ is a rule of thumb value, but we will use $\xi = 0.2$ and 0.5 to test the estimator.

The result is reported in Table 2 and as we can see, the estimator performs better when the sampling frequency increases or the value of ξ is small. However, it is also clear that estimation error is quite large, so further development of estimation of IVV should be possible.

Finally, in Table 3, we compare the estimation performance for different kernels. As shown therein, the exponential kernel performs the best in all cases. As the calculation we had in Remark 4.2, we can see that the second best kernel is the triangle kernel, since its shape is more similar to exponential kernel. Similarly, the uniform kernel performs the worst, since it is the farthest to the optimal exponential kernel.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table 3
Comparison of Different Kernel Functions (5 min data, 2000 sample paths)

Length	ρ	Exponential	Uniform	Triangle	Epanechnikov
5 days	0	2.5974E−05	2.8721E−05	2.6441E−05	2.7085E−05
5 days	−0.5	2.5233E−05	2.8252E−05	2.5759E−05	2.6490E−05
21 days	0	2.3406E−05	2.8047E−05	2.4988E−05	2.5914E−05
21 days	−0.5	2.3692E−05	2.8603E−05	2.5248E−05	2.6173E−05

Acknowledgments

The first author's research supported in part by the NSF Grants: DMS-1149692 and DMS-1613016. The authors are grateful to an anonymous referee for his/her thorough reading and multiple corrections and suggestions that improve the original manuscript significantly.

Appendix A. Proofs of main results

Proof of Theorem 3.1. Let us start by writing the MSE as follows:

$$\mathrm{MSE} = \mathbb{E}\left[(T_{1n} + T_{2n})^2\right],$$

where

$$T_{1n} := \sum_{i=1}^{n} K_h(t_{i-1} - \tau)((\Delta_i X)^2 - \Delta\sigma_\tau^2), \quad T_{2n} := \left(\sum_{i=1}^{n} K_h(t_{i-1} - \tau)\Delta - 1\right)\sigma_\tau^2.$$

Applying Lemmas 3.1 and 3.2 of the supplemental material to this article available online with $f(t) \equiv 1$, it follows that $\sum_{i=1}^{n} K_h(t_{i-1} - \tau)\Delta - 1 = O(\Delta/h) + O(h^\gamma)$ and, thus, by Assumption 2, $\mathbb{E}\left[T_{2n}^2\right] = o(\Delta/h) + o(h^\gamma)$. Furthermore, since

$$|\mathbb{E}[T_{1n}T_{2n}]| \leq \left|\sum_{i=1}^{n} K_h(t_{i-1} - \tau)\Delta - 1\right| \mathbb{E}\left[T_{1n}^2\right]^{1/2} \mathbb{E}\left[\sigma_\tau^4\right]^{1/2},$$

to conclude (3.2), it suffices to show that

$$\mathbb{E}\left[T_{1n}^2\right] = 2\frac{\Delta}{h}\mathbb{E}[\sigma_\tau^4]\|K\|_2^2 + h^\gamma L(\tau)\iint K(x)K(y)C_\gamma(x, y)dxdy + \text{h.o.t.} \tag{A.1}$$

For simplicity, in the rest of the proof, h.o.t. refers to terms of order $o(\Delta/h) + o(h^\gamma)$.

To show (A.1), let us start by applying Lemmas 3.1 and 3.2 of the supplemental material to this article available online, together with Assumptions 1 and 2, to write $\mathbb{E}\left[T_{1n}^2\right]$ as

$$\begin{aligned} &\sum_{i,j=1}^{n} K_h(t_{i-1} - \tau)K_h(t_{j-1} - \tau)\mathbb{E}[((\Delta_i X)^2 - \Delta\sigma_\tau^2)((\Delta_j X)^2 - \Delta\sigma_\tau^2)] \\ &= \sum_{i,j=1}^{n} K_h(t_{i-1} - \tau)K_h(t_{j-1} - \tau)\mathbb{E}[((\Delta_i M)^2 - \Delta\sigma_\tau^2)((\Delta_j M)^2 - \Delta\sigma_\tau^2)] + \text{h.o.t.}, \end{aligned} \tag{A.2}$$

where $M_t = \int_0^t \sigma_s dB_s$ is the martingale part of X (see Lemma 2.1 and Remark 2.1 in the supplemental material to this article for details). By Assumption 1, it follows that

$$\mathbb{E}\left[T_{1n}^2\right] = 2\sum_{i=1}^{n} K_h^2(t_{i-1} - \tau)\mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_t^2 dt\right)^2\right]$$

$$
\begin{aligned}
&+\sum_{i,j=1}^{n} K_h(t_{i-1}-\tau)K_h(t_{j-1}-\tau)\int_{t_{i-1}}^{t_i}\int_{t_{j-1}}^{t_j}\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)(\sigma_s^2-\sigma_\tau^2)]dtds+\text{h.o.t.}\\
&=: 2V_1+V_2+\text{h.o.t.}
\end{aligned}
\tag{A.3}
$$

We now proceed to analyze V_1 and V_2. Firstly, for V_1, note that

$$
\begin{aligned}
\mathbb{E}\left(\int_{t_{i-1}}^{t_i}\sigma_t^2dt\right)^2 &= \Delta^2\mathbb{E}[\sigma_\tau^4]+2\Delta\int_{t_{i-1}}^{t_i}\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)\sigma_\tau^2]dt+\mathbb{E}\left(\int_{t_{i-1}}^{t_i}(\sigma_t^2-\sigma_\tau^2)dt\right)^2\\
&=: \Delta^2\mathbb{E}[\sigma_\tau^4]+B_i+C_i.
\end{aligned}
$$

To analyze the contribution of each of the three terms above to V_1, we use Lemmas 3.1 and 3.2 of the supplemental material to this article available online with ‘kernel’ function K^2 and the following three different functions f:

$$
f(t)=1,\quad f(t)=\sqrt{\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)^2]\mathbb{E}[\sigma_\tau^4]},\quad f(t)=\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)^2],
$$

respectively. It then follows that

$$
\begin{aligned}
\Delta^2\sum_{i=1}^{n}K_h^2(t_{i-1}-\tau) &= \frac{\Delta}{h}\sum_{i=1}^{n}K^2\Big(\frac{t_{i-1}-\tau}{h}\Big)\frac{\Delta}{h}=\frac{\Delta}{h}\int K^2(x)dx+\text{h.o.t.},\\
\sum_{i=1}^{n}K_h^2(t_{i-1}-\tau)B_i &\le 2\frac{\Delta}{h}\sum_{i=1}^{n}K^2\Big(\frac{t_{i-1}-\tau}{h}\Big)\frac{1}{h}\int_{t_{i-1}}^{t_i}\sqrt{\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)^2]\mathbb{E}[\sigma_\tau^4]}dt=\text{h.o.t.},\\
\sum_{i=1}^{n}K_h^2(t_{i-1}-\tau)C_i &\le \frac{\Delta}{h}\sum_{i=1}^{n}K^2\Big(\frac{t_{i-1}-\tau}{h}\Big)\frac{1}{h}\int_{t_{i-1}}^{t_i}\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)^2]dt=\text{h.o.t.},
\end{aligned}
$$

where the second line above follows from the fact that $\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)^2]=O(|t-\tau|^\gamma)$. Putting together the previous relationships, we conclude that

$$
V_1=\sum_{i=1}^{n}K_h^2(t_{i-1}-\tau)\mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i}\sigma_t^2dt\right)^2\right]=\frac{\Delta}{h}\mathbb{E}[\sigma_\tau^4]\int K^2(x)dx+\text{h.o.t.}
$$

Next, applying directly Lemmas 3.1 and 3.2 of the supplemental material to this article available online together with Assumption 3, V_2 can be written as

$$
V_2=h^\gamma\int\int K(x)K(y)C_\gamma(x,y;\tau)dxdy+o\left(\frac{\Delta}{h}\right)+o(h^\gamma).
$$

The asymptotics for V_1 and V_2 above together with (A.3) implies (A.1), which, as argued at the beginning of the proof, leads to (3.1). ■

Proof of Theorem 5.1. The t_b here is basically to rule our boundary effects and for brevity of notation, we will write $t_b=0$ and assume we have a left side estimator near $t=0$ and a right side estimator near $T=t$, with the same convergence rate. Define the error terms from the left and right side estimators as $l_i=\hat{\sigma}_{l,t_i}^2-\sigma_{t_i}^2$ and $r_i=\hat{\sigma}_{r,t_i}^2-\sigma_{t_i}^2$, respectively. We will consider the following slightly different estimator:

$$
\widehat{IVV}_T^{(\text{tsrvv})}=\frac{1}{k}\sum_{i=0}^{n-k}(\Delta_i^{(k)}\hat{\sigma}^2)^2-\frac{1}{k}\sum_{i=0}^{n-1}(\Delta_i\hat{\sigma}^2)^2. \tag{A.4}
$$

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

In terms of the error terms r_i and l_i, this can be written as

$$\begin{aligned}\widehat{IVV}_T^{(\text{tsrvv})} = \frac{1}{k}\Bigg[&\sum_{i=0}^{n-k}(\Delta_i^{(k)}\sigma^2)^2 - \sum_{i=0}^{n-1}(\Delta_i\sigma^2)^2 - 2\sum_{i=n-k+1}^{n-1}\sigma_{t_i}^2 l_i - 2\sum_{i=1}^{k-1}\sigma_{t_i}^2 r_i \\ &+ 2\sum_{i=k+1}^{n-1}(\sigma_{t_i}^2 - \sigma_{t_{i-k+1}}^2)r_{i+1} - 2\sum_{i=0}^{n-k}(\sigma_{t_{i+k}}^2 - \sigma_{t_{i+1}}^2)l_i + 2\sum_{i=0}^{k-2}\sigma_{t_i}^2 r_{i+1} \\ &+ 2\sum_{i=n-k+1}^{n-1}\sigma_{t_{i+1}}^2 l_i - \sum_{i=n-k+1}^{n-1} l_i^2 - \sum_{i=1}^{k-1} r_i^2 - 2\sum_{i=0}^{n-k} l_i r_{i+k} + 2\sum_{i=0}^{n-1} l_i r_{i+1}\Bigg].\end{aligned}$$

Now, for each pair of similar terms, we consider the convergence rate of only one of them. The others have the same convergence rate. Indeed, from Proposition 3.1, we have that $\mathbb{E}[r_i^2] = O(n^{-1/2})$, since we are dealing with Brownian motion type volatility. Thus,

$$\mathbb{E}\left|\sum_{i=k}^{n-1}(\sigma_{t_i}^2 - \sigma_{t_{i-k+1}}^2)r_{i+1}\right| \le \sqrt{\sum_{i=0}^{n-k}\mathbb{E}[(\sigma_{t_{i+k}}^2 - \sigma_{t_{i+1}}^2)^2]\sum_{i=0}^{n-k}\mathbb{E}(r_{i+1}^2)} = O(k^{\frac{1}{2}}n^{\frac{1}{4}}),$$

$$\mathbb{E}\left|\sum_{i=n-k+1}^{n-1}\sigma_{t_i}^2 l_i\right| \le \sqrt{\sum_{i=n-k+1}^{n-1}\mathbb{E}(\sigma_{t_i}^4)\sum_{i=n-k+1}^{n-1}\mathbb{E}(l_i^2)} = O(\frac{k}{n^{1/4}}),$$

$$\mathbb{E}\sum_{i=n-k+1}^{n-1} l_i^2 = O(\frac{k}{\sqrt{n}}), \quad \mathbb{E}\left|\sum_{i=0}^{n-k} l_i r_{i+k}\right| \le \sqrt{\sum_{i=0}^{n-k}\mathbb{E}(l_i^2)\sum_{i=0}^{n-k}\mathbb{E}(r_{i+k}^2)} = O(\sqrt{n}).$$

Similarly, we can see that the difference between (5.4) and (A.4) is $O_p(\Delta)$. Putting all these together, we get

$$\text{TSRVV} - \sum_{i=0}^{n-k}(\Delta_i^{(k)}\sigma^2)^2 - \sum_{i=0}^{n-1}(\Delta_i\sigma^2)^2 = O_p\left(\frac{n^{1/4}}{k^{1/2}}\right). \tag{A.5}$$

On the other hand, with similar assumptions and proofs as Theorem 2 and 3 of Zhang et al. [16], we have the following:

$$\frac{1}{k}\left[\sum_{i=0}^{n-k}(\Delta_i^{(k)}\sigma^2)^2 - \sum_{i=0}^{n-1}(\Delta_i\sigma^2)^2\right] - \int_0^T g^2(t)dt = O_p\left(\sqrt{\frac{k}{n}}\right). \tag{A.6}$$

Therefore, we have

$$\text{TSRVV} - \int_0^T g^2(t)dt = O_p\left(\frac{n^{1/4}}{k^{1/2}}\right) + O_p\left(\sqrt{\frac{k}{n}}\right),$$

which implies the consistency and also yields that the optimal k is given by $Cn^{3/4}$, in which case the convergence rate is $n^{-1/8}$. ■

Proof of Theorem 6.1. In what follows, we are going to assume that the relevant processes (such as σ, μ, and, in the case of Brownian driven volatilities, the coefficients driving the dynamics of σ^2) are bounded. This can be justified by localization as in Section 4.4.1 in [7]

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

and Appendix A.5 in [1]. Let

$$A_n = \left(\frac{\Delta}{h}\right)^{-1/2}\left[\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2 - \int_0^T K_h(t-\tau)\sigma_t^2 dt\right].$$

Let us start with the approximations:

$$\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2 = \sum_{i=1}^{n} K_h(t_{i-1}-\tau)\left(\int_{t_{i-1}}^{t_i}\sigma_s dB_s\right)^2 + O_p(\Delta^{1/2})$$

$$\int_0^T K_h(t-\tau)\sigma_t^2 dt = \sum_{i=1}^{n} K_h(t_{i-1}-\tau)\int_{t_{i-1}}^{t_i}\sigma_t^2 dt + o_p\left((\Delta/h)^{1/2}\right).$$

The first approximation above follows from the fact that $\int_{t_{i-1}}^{t_i}\sigma_s dB_s = O_P(\Delta^{1/2})$ and $\Delta\sum_{i=1}^{n}|K_h(t_{i-1}-\tau)| \to \int_0^T |K(x)|dx$, while the second one follows from the proof of Lemma 3.1 and Remark 3.1 in the supplemental material to this article available online and the fact that σ is bounded. For an alternative proof see Lemma A.1 in [11]. We can then write:

$$A_n = \left(\frac{\Delta}{h}\right)^{-\frac{1}{2}}\sum_{i=1}^{n} K_h(t_{i-1}-\tau)\left\{\left(\int_{t_{i-1}}^{t_i}\sigma_s dB_s\right)^2 - \int_{t_{i-1}}^{t_i}\sigma_t^2 dt\right\} + o_P(1) \quad \text{(A.7)}$$

$$=: S_n + o_P(1).$$

Clearly, S_n can be written as a sum $\sum_{i=1}^{n}\alpha_{n,i}$ of martingale differences relative to $\{\mathcal{F}_{n,i} := \mathcal{F}_{t_i}\}_{i=1,\dots,n}$ with

$$\alpha_{n,i} := \left(\frac{\Delta}{h}\right)^{-1/2} K_h(t_{i-1}-\tau)\left\{\left(\int_{t_{i-1}}^{t_i}\sigma_s dB_s\right)^2 - \int_{t_{i-1}}^{t_i}\sigma_t^2 dt\right\}.$$

To obtain the CLT, we first need to show the following (see Theorem IX.7.28 in [8]):

$$B_n := \sum_{i=1}^{n}\mathbb{E}[\alpha_{n,i}^2|\mathcal{F}_{n,i-1}] \xrightarrow{P} 2\sigma_\tau^4\|K\|_2^2. \quad \text{(A.8)}$$

First note that, by Itô's lemma,

$$B_n = 4\left(\frac{\Delta}{h}\right)^{-1}\sum_{i=1}^{n} K_h^2(t_{i-1}-\tau)\int_{t_{i-1}}^{t_i}\mathbb{E}\left[\left(\int_{t_{i-1}}^{s}\sigma_u dB_u\right)^2\sigma_s^2\middle|\mathcal{F}_{n,i-1}\right]ds.$$

By the Cauchy–Schwarz and the BDG inequalities,

$$\mathbb{E}\left[\left(\int_{t_{i-1}}^{s}\sigma_u dB_u\right)^2\left(\sigma_s^2-\sigma_{t_{i-1}}^2\right)\middle|\mathcal{F}_{n,i-1}\right]^2$$

$$\leq \mathbb{E}\left[\left(\int_{t_{i-1}}^{s}\sigma_u dB_u\right)^4\middle|\mathcal{F}_{n,i-1}\right]\mathbb{E}\left[\left(\sigma_s^2-\sigma_{t_{i-1}}^2\right)^2\middle|\mathcal{F}_{n,i-1}\right]$$

$$\leq C\mathbb{E}\left[\left(\int_{t_{i-1}}^{s}\sigma_u^2 du\right)^2\middle|\mathcal{F}_{n,i-1}\right]\mathbb{E}\left[\left(\sigma_s^2-\sigma_{t_{i-1}}^2\right)^2\middle|\mathcal{F}_{n,i-1}\right] = O_P(\Delta^{2+\gamma}),$$

uniformly on i, due to Assumption 3. Therefore,

$$\left(\frac{\Delta}{h}\right)^{-1}\sum_{i=1}^{n}K_h^2(t_{i-1}-\tau)\int_{t_{i-1}}^{t_i}\mathbb{E}\left[\left(\int_{t_{i-1}}^{s}\sigma_u dB_u\right)^2(\sigma_s^2-\sigma_{t_{i-1}}^2)\middle|\mathcal{F}_{n,i-1}\right]ds$$
$$=O_P(\Delta^{(2+\gamma)/2})h\sum_{i=1}^{n}K_h^2(t_{i-1}-\tau)=O_P(\Delta^{\gamma/2}),$$

since $\Delta h\sum_{i=1}^{n}K_h^2(t_{i-1}-\tau)\to\|K\|_2^2$. We then have that:

$$\begin{aligned}B_n&=4\left(\frac{\Delta}{h}\right)^{-1}\sum_{i=1}^{n}K_h^2(t_{i-1}-\tau)\int_{t_{i-1}}^{t_i}\sigma_{t_{i-1}}^2\mathbb{E}\left[\left(\int_{t_{i-1}}^{s}\sigma_u dB_u\right)^2\middle|\mathcal{F}_{n,i-1}\right]ds+o_P(1)\\&=4\left(\frac{\Delta}{h}\right)^{-1}\sum_{i=1}^{n}K_h^2(t_{i-1}-\tau)\int_{t_{i-1}}^{t_i}\sigma_{t_{i-1}}^2\mathbb{E}\left[\int_{t_{i-1}}^{s}\sigma_u^2du\middle|\mathcal{F}_{n,i-1}\right]ds+o_P(1)\\&=4\left(\frac{\Delta}{h}\right)^{-1}\sum_{i=1}^{n}K_h^2(t_{i-1}-\tau)\sigma_{t_{i-1}}^4\int_{t_{i-1}}^{t_i}(s-t_{i-1})ds+o_P(1)\\&=2h\Delta\sum_{i=1}^{n}K_h^2(t_{i-1}-\tau)\sigma_{t_{i-1}}^4+o_P(1)\\&\xrightarrow{P}2\sigma_\tau^4\|K\|_2^2.\end{aligned}$$

The following is the final identity needed to conclude the CLT:

$$\sum_{i=1}^{n}\mathbb{E}[\alpha_{n,i}^4|\mathcal{F}_{n,i-1}]\xrightarrow{P}0,$$

for which it suffices to show that

$$T_{1n}:=\left(\frac{\Delta}{h}\right)^{-2}\sum_{i=1}^{n}K_h^4(t_{i-1}-\tau)\mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i}\sigma_s dB_s\right)^8\middle|\mathcal{F}_{n,i-1}\right]\xrightarrow{P}0,$$
$$T_{2n}:=\left(\frac{\Delta}{h}\right)^{-2}\sum_{i=1}^{n}K_h^4(t_{i-1}-\tau)\mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i}\sigma_s^2 ds\right)^4\middle|\mathcal{F}_{n,i-1}\right]\xrightarrow{P}0.$$

By BDG inequality, for some constant $C<\infty$,

$$T_{1n}\le CT_{2n}=O_P(\Delta^2)h^2\sum_{i=1}^{n}K_h^4(t_{i-1}-\tau)=O_P\left(\frac{\Delta}{h}\right),$$

since $\Delta h^3\sum_{i=1}^{n}K_h^4(t_{i-1}-\tau)\to\int K^4(x)dx$. The final ingredient to apply the CLT for martingale differences, as in Theorem IX.7.28 in [8], is to show that

$$\sum_{i=1}^{n}\mathbb{E}\left[\alpha_i\Delta_i H|\mathcal{F}_{n,i-1}\right]\xrightarrow{P}0,$$

where H is either W or an arbitrary bounded martingale orthogonal (in the martingale sense) to W. This is done in the same way as in the proof of Theorem 2.7 in [11]. ■

Proof of Theorem 6.2. **(1)** As it is standard in the literature, by virtue of localization (as in Jacod and Shiryaev, section 5.4, p.549), we assume without loss of generality that the coefficients driving the dynamics of σ are bounded on $[0, T]$. For simplicity, we will use the following notations: $V_t = \sigma_t^2 = \sigma_0^2 + \int_0^t f_s ds + \int_0^t g_s dW_s$ and $v_t = \sigma_0^2 + \int_0^t g_s dW_s$. It is easy to see from Proposition 2.1 that V and v both satisfy Assumption 3 with $\gamma^V = \gamma^v = 1$ and $C_\gamma^V = C_\gamma^v$. Now, since

$$h^{-1/2}\mathbb{E}\left|\int_0^T K_h(t-\tau)\int_\tau^t f_s ds dt\right| \leq \sup_{s\in[0,T]} |f_s| h^{-1/2}\int_0^T |K_h(t-\tau)||t-\tau| dt,$$

which is $O_p(h^{1/2}) = o_p(1)$, we can conclude that the drift term of V has a negligible contribution to the final error. Therefore, it suffices to work with the process v and only to consider the weak convergence of

$$\bar{I}_h := h^{-1/2}\left(\int_0^T K_h(t-\tau)(v_t - v_\tau)dt\right).$$

For the sake of clarity, we will first assume a right-sided kernel function (i.e., $K(x) = 0$ for all $x < 0$), so that $\bar{I}_h = h^{-1/2}\left(\int_\tau^T K_h(t-\tau)(v_t - v_\tau)dt\right) =: I_h$. Applying the integration by parts formula, we have that

$$I_h = -h^{-1/2}U\left(\frac{T-\tau}{h}\right)(v_T - v_\tau) + h^{-1/2}\int_\tau^T U\left(\frac{t-\tau}{h}\right) g_t dW_t =: R + S,$$

where $U(t) = \int_t^\infty K(u)du$ so that $\frac{d}{dt}(U((t-\tau)/h)) = -K_h(t-\tau)$. Since our assumptions on K imply that $x^{1/2}U(x) \to 0$, as $x \to \infty$, we have $R = o_P(1)$. For the other term S, let us consider the following approximation $\widetilde{S} := h^{-1/2} g_\tau \int_\tau^T U\left(\frac{t-\tau}{h}\right) dW_t$, and note that $S - \widetilde{S} = o_P(1)$ since, by Assumption 4, $\int_0^\infty U^2(x)dx < \infty$ and

$$\begin{aligned}\mathbb{E}\left[\left(S-\widetilde{S}\right)^2\right] &= \frac{1}{h}\left(\int_\tau^{\tau+\sqrt{h}} + \int_{\tau+\sqrt{h}}^T\right) U^2\left(\frac{t-\tau}{h}\right)\mathbb{E}\left[(g_t - g_\tau)^2\right] dt \\ &\leq \sup_{t\in[\tau,\tau+\sqrt{h}]} \mathbb{E}\left[(g_t-g_\tau)^2\right]\|U^2\|_1 + 4\|g^2\|_\infty \int_{1/\sqrt{h}}^\infty U^2(s)ds,\end{aligned}$$

which is clearly $o(1)$, as $h \to 0$. We also observe that conditional on $\mathcal{F}_\tau$, $\tilde{S}$ is Gaussian with mean 0 and variance:

$$g_\tau^2 h^{-1}\int_\tau^T U^2\left(\frac{t-\tau}{h}\right)dt = g_\tau^2\int_0^{\frac{t-\tau}{h}} U^2(s)ds \to g_\tau^2 \iint K(x)K(y)C_1(x,y)dxdy.$$

Therefore, $\tilde{S}|\mathcal{F}_\tau \to_D \mathcal{N}(0, \delta_2^2)$, where $\delta_2^2 = g_\tau^2 \iint K(x)K(y)C_\gamma(x,y)dxdy$. This suffices for (6.3) since, by the dominated convergence theorem,

$$\begin{aligned}\mathbb{E}\left[\exp\left(iu\tilde{S}\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(iu\tilde{S}\right)\middle|\mathcal{F}_\tau\right]\right] = \mathbb{E}\left[\exp\left(-\frac{u^2 g_\tau^2}{2h}\int_\tau^T U^2\left(\frac{t-\tau}{h}\right)dt\right)\right] \\ &\xrightarrow{h\to 0} \mathbb{E}\left[\exp\left(-\frac{u^2 g_\tau^2}{2}\int U^2(s)\,ds\right)\right],\end{aligned}$$

where recall that $\int U^2(s)\,ds = \iint K(x)K(y)C_1(x,y)dxdy$.

We now consider the general two-sided kernel case. To this end, let $\bar{U}(t)=\int_t^\infty K(u)du\mathbf{1}_{\{t>0\}}-\int_{-\infty}^t K(u)du\mathbf{1}_{\{t\le 0\}}$ and note that, by the integration by parts formula, $\bar{I}_h=h^{-1/2}\int_0^T K_h(t-\tau)(v_t-v_\tau)dt$ is such that

$$\begin{aligned}\bar{I}_h&=-h^{-1/2}\bar{U}\left(\frac{T-\tau}{h}\right)(v_0-v_\tau)+h^{-1/2}\int_0^\tau \bar{U}\left(\frac{t-\tau}{h}\right)g_t dW_t+I_h+o_P(1)\\&=:\bar{R}+\bar{S}+I_h+o_P(1).\end{aligned}$$

Same as in the one-sided kernel case, $\bar{R}=o_P(1)$ and $I_h=\widetilde{S}+o_P(1)$. For $\bar{S}$, we consider the following approximation:

$$\widetilde{\bar{S}}:=h^{-1/2}g_\tau\int_0^\tau \bar{U}\left(\frac{t-\tau}{h}\right)dW_t=h^{-1/2}g_\tau\left(\int_0^{\tau-\sqrt{h}}+\int_{\tau-\sqrt{h}}^{\tau}\right)\bar{U}\left(\frac{t-\tau}{h}\right)dW_t$$
$$=:\widetilde{\bar{S}}_1+\widetilde{\bar{S}}_2.$$

We still have $\bar{S}-\widetilde{\bar{S}}=o_P(1)$. It is also true that $\widetilde{\bar{S}}_1=o_P(1)$, as $h\to 0$, which can be justified by considering its second moment. Therefore, we have

$$\bar{I}_h=\widetilde{\bar{S}}_2+\widetilde{S}+o_P(1)=h^{-1/2}g_{\tau-\sqrt{h}}\int_{\tau-\sqrt{h}}^{T}\bar{U}\left(\frac{t-\tau}{h}\right)dW_t+o_P(1)=:\widetilde{\bar{I}}_h+o_P(1),$$

where the second equality holds since $\widetilde{\bar{S}}_2+\widetilde{S}-\widetilde{\bar{I}}_h=o_P(1)$, which again can be justified by considering the second moment and Cauchy–Schwarz' inequality. To conclude (6.3), note that, by conditioning on $\mathcal{F}_{\tau-\sqrt{h}}$,

$$\mathbb{E}\left[\exp\left(iu\widetilde{\bar{I}}_h\right)\right]=\mathbb{E}\left[\exp\left(-\frac{u^2g^2_{\tau-\sqrt{h}}}{2}\int_{-h^{-1/2}}^{\frac{T-\tau}{h}}\bar{U}^2(s)\,ds\right)\right],$$

which converges to $\mathbb{E}\left[\exp\left(-\frac{u^2g_\tau^2}{2}\iint K(x)K(y)C_1(x,y)dxdy\right)\right]$ and we conclude (6.3).

(2) In the whole proof, the superscript (Z) refers to a quantity corresponding to the process Z, while quantities without such a superscript corresponds to the process σ^2. Let us start by noting that, since Z is a Gaussian process, $h^{-\gamma/2}\left(\int_0^T K_h(t-\tau)(Z_t-Z_\tau)dt\right)\to_D \bar{\delta}_2^{1/2}N(0,1)$, where

$$\bar{\delta}_2=L^{(Z)}(\tau)\iint K(x)K(y)C_\gamma^{(Z)}(x,y)dxdy.$$

Indeed, this follows from the facts that the limit in distribution of Gaussian r.v.'s is Gaussian and that $h^{-\gamma/2}\int_0^T K_h(t-\tau)(Z_t-Z_\tau)dt$ is centered Gaussian (being the limit of Riemann sums of the form $h^{-\gamma/2}\sum_{j=0}^{m-1}K_h(t_j-\tau)(Z_{t_j}-Z_\tau)(t_{j+1}-t_j)dt$, which is Gaussian) with variance

$$\begin{aligned}&h^{-\gamma}\int_0^T\int_0^T K_h(t-\tau)K_h(s-\tau)\mathbb{E}[(Z_t-Z_\tau)(Z_s-Z_\tau)]dtds\\&=h^{-\gamma}\int_{-\tau/h}^{(T-\tau)/h}\int_{-\tau/h}^{(T-\tau)/h}K(x)K(y)\mathbb{E}[(Z_{\tau+xh}-Z_\tau)(Z_{\tau+yh}-Z_\tau)]dxdy,\end{aligned}$$

which converges to $\bar{\delta}_2$ above.

Now, for any $\epsilon\in(0,\min(\tau,T-\tau))$, and for any $t\in(\tau-\epsilon,\tau+\epsilon)$, there exists $s_t\in(\min(t,\tau),\max(t,\tau))$, such that $\sigma_t^2-\sigma_\tau^2=f'(Z_\tau)(Z_t-Z_\tau)+\frac{1}{2}f''(Z_{s_t})(Z_t-Z_\tau)^2$.

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

Then, $I := \int_0^T K_h(t-\tau)(\sigma_t^2 - \sigma_\tau^2)dt$ is such that

$$I = \int_{\tau-\epsilon}^{\tau+\epsilon} K_h(t-\tau)[f'(Z_\tau)(Z_t - Z_\tau) + \frac{1}{2} f''(Z_{s_t})(Z_t - Z_\tau)^2]dt + o(h^{\gamma/2}).$$

Indeed, to justify the term $o(h^{\gamma/2})$ above, note that, due to (3)-(ii) of Assumption 4,

$$\mathbb{E}\left|\int_{\tau+\epsilon}^{T} K_h(t-\tau)(\sigma_t^2 - \sigma_\tau^2)dt\right| \leq \int_{\tau+\epsilon}^{T} |K_h(t-\tau)|dt \left(T \int_{\tau+\epsilon}^{T} \mathbb{E}(\sigma_t^2 - \sigma_\tau^2)^2 dt\right)^{1/2}$$
$$\leq C\left(\int_{\epsilon/h}^{\infty} K(x)dx\right)^{1/2} = o(h^{\gamma/2}).$$

We can similarly deal with the integral from 0 to $\tau - \epsilon$. For the second term, once we select ϵ small enough such that

$$\mathbb{E}[(f''(Z_t))^2] < M^2, \quad \mathbb{E}[(Z_t - Z_\tau)^4] = 3\mathbb{E}[(Z_t - Z_\tau)^2]^2 \leq 3M|t-\tau|^{2\gamma},$$

for all $t \in (\tau - \epsilon, \tau + \epsilon)$ and some M, we can then apply Cauchy–Schwarz's inequality to get

$$\mathbb{E}\left|\int_{\tau-\epsilon}^{\tau+\epsilon} K_h(t-\tau) f''(Z_{s_t})(Z_t - Z_\tau)^2 dt\right| \leq 3M^2 \int_{\tau-\epsilon}^{\tau+\epsilon} |K_h(t-\tau)||t-\tau|^\gamma dt,$$

which is $O(h^\gamma) = o(h^{\gamma/2})$. Now for the first term, we have

$$h^{-\gamma/2} \int_{\tau-\epsilon}^{\tau+\epsilon} K_h(t-\tau)[f'(Z_\tau)(Z_t - Z_\tau)]dt \to_D f'(Z_\tau)\bar{\delta}_2^{1/2} N(0,1).$$

where the standard normal $N(0,1)$ appearing above is independent from Z_τ. Indeed, $(X, Y(h)) := (Z_\tau, h^{-\gamma/2} \int_{\tau-\epsilon}^{\tau+\epsilon} K_h(t-\tau)(Z_t - Z_\tau)dt)$ is bi-variate normal for all $h > 0$ and, thus, whenever the limit $(X, Y(h)) \to (X, Y)$ exists, (X, Y) is a bivariate normal variable. There exist $\alpha(h)$ and $\beta(h)$ such that $Y(h) = \alpha(h)X + \beta(h)Z(h)$, such that X is independent of $Z(h)$ and $Z(h) \stackrel{D}{=} N(0,1)$. Note that $\alpha(h)$ and $\beta(h)$ are given by

$$\alpha(h) = \frac{\mathbb{E}[XY(h)]}{\mathbb{E}[X^2]}, \quad \beta^2(h) = \mathbb{E}[Y^2(h)] - \alpha^2(h)\mathbb{E}[X^2].$$

By our assumption on Z stated in the statement of the theorem, we have $\mathbb{E}[XY(h)] = o(1)$ and, thus, $\alpha(h) = o(1)$, while

$$\beta^2(h) = L^{(Z)}(\tau) \iint K(x)K(y)C_\gamma^{(Z)}(x,y)dxdy + o(1).$$

With such representations, we have:

$$f'(X)Y(h) = \alpha(h)f'(X)X + \beta(h)f'(X)Z(h) = o_p(1) + \beta(h)f'(X)Z(h),$$

which converges to $\beta f'(X)Z$. ■

Proof of Corollary 6.1. We show the result in the first setting (1) of Theorem 6.2 (the second can be handled similarly). Let U_n and V_n be the first and second terms of the decomposition

$$\hat{\sigma}_\tau^2 - \sigma_\tau^2 = \left(\sum_{i=1}^{n} K_h(t_{i-1} - \tau)(\Delta_i X)^2 - \int_0^T K_h(t-\tau)\sigma_t^2 dt\right)$$
$$+ \int_0^T K_h(t-\tau)(\sigma_t^2 - \sigma_\tau^2)dt + o_p(h^\gamma), \tag{A.9}$$

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

Let us start by noting that

$$\mathbb{E}\left[e^{iuh^{-\gamma/2}(U_n+V_n)}\right]=\mathbb{E}\left[\mathbb{E}\left[e^{iuh^{-\gamma/2}(U_n+V_n)}\middle|\mathcal{F}(\sigma_s:s\le T)\right]\right]$$
$$=\mathbb{E}\left[e^{iuh^{-\gamma/2}V_n}\mathbb{E}\left[e^{iuh^{-\gamma/2}U_n}\middle|\mathcal{F}(\sigma_s:s\le T)\right]\right].$$

From Theorem 6.1,[3]

$$\mathbb{E}\left[e^{iuh^{-\gamma/2}U_n}\middle|\mathcal{F}(\sigma_s:s\le T)\right]\to e^{-u^2\sigma_\tau^4\int K^2(x)dx},$$

so it suffices to show that

$$\mathbb{E}\left[e^{iuh^{-\gamma/2}V_n-u^2\sigma_\tau^4\int K^2(x)dx}\right]\to\mathbb{E}\left[e^{-\frac{u^2}{2}\left(\delta_1^2+\delta_2^2\right)}\right].$$

For this, first note that, since $\sigma_{\tau-\sqrt{h}}\to\sigma_\tau$, a.s., and, σ is bounded (by virtue of localization), we have

$$\left|\mathbb{E}\left[e^{iuh^{-\gamma/2}V_n-u^2\sigma_\tau^4\int K^2(x)dx}\right]-\mathbb{E}\left[e^{iuh^{-\gamma/2}V_n-u^2\sigma_{\tau-\sqrt{h}}^4\int K^2(x)dx}\right]\right|\to 0.$$

Finally,

$$\mathbb{E}\left[e^{iuh^{-\gamma/2}V_n-u^2\sigma_{\tau-\sqrt{h}}^4\int K^2(x)dx}\right]\to\mathbb{E}\left[e^{-\frac{u^2}{2}\left(\delta_1^2+\delta_2^2\right)}\right],$$

along the same arguments as those used in the proof of Theorem 6.2. ■

Proof of Theorem 6.3. By virtue of localization (as in [8], Section 5.4, p. 549), we can (and will) assume that the relevant processes (such as σ, μ, and the coefficients driving the dynamics of σ) are bounded. We again consider the decomposition (6.1) and call the first and second terms on the right-hand side $A_{1,n}$ and $A_{2,n}$, respectively. As stated in the theorem, we take $\Delta=h^2$, in which case, the two terms attained the optimal rate $h^{1/2}$. Let us start by noting the below decomposition, which was already obtained in (A.7):

$$h^{-1/2}A_{1,n}=h^{-1/2}\sum_{i=1}^n K_h(t_{i-1}-\tau)\left\{\left(\int_{t_{i-1}}^{t_i}\sigma_s dB_s\right)^2-\int_{t_{i-1}}^{t_i}\sigma_t^2dt\right\}+o_P(1)$$
$$=:\sum_{i=1}^n\alpha_{n,i}+o_P(1).$$

For $A_{2,n}$, by similar arguments as those used in the proof of Theorem 6.2, we have

$$h^{-1/2}A_{2,n}=h^{-1/2}g_{t_{j-1}}\int_{t_{j-1}}^T U\left(\frac{t-\tau}{h}\right)dW_t+o_P(1)=:\sum_{i=1}^n\beta_{n,i}+o_P(1),$$

where $t_{j-1}=\min\{t_i:\tau\le t_i\}$ and

$$\beta_{n,i}=\begin{cases}0 & i<j\\ h^{-1/2}g_{t_{j-1}}\int_{t_{i-1}}^{t_i}U\left(\frac{t-\tau}{h}\right)dW_t, & i\ge j.\end{cases}$$

[3] Theorem 6.1 obviously holds when $\{\sigma_t\}$ is deterministic, which is the process we will get when conditioning U_n on $\mathcal{F}(\sigma_s:s\le T)$ due to the nonleverage condition.

Next, we consider the following sum of martingale differences relative to $\{\mathcal{F}_{n,i} := \mathcal{F}_{t_i}\}_{i=1,\dots,n}$:

$$S_n = \sum_{i=1}^{n} \xi_{n,i} = \sum_{i=1}^{n} \left(\alpha_{n,i} + \beta_{n,i}\right).$$

To apply the CLT for martingale differences (see Theorem IX.7.28 in Jacod and Shiryaev), we first need to show that:

$$\sum_{i=1}^{n} \mathbb{E}[\xi_{n,i}^2|\mathcal{F}_{n,i-1}] \xrightarrow{P} 2\sigma_\tau^4\|K\|_2^2 + g_\tau^2 \int\int K(x)K(y)C_1(x,y)dxdy.$$

To this end, we prove that

$$B_n := \sum_{i=1}^{n} \mathbb{E}[\alpha_{n,i}^2|\mathcal{F}_{n,i-1}] \xrightarrow{P} 2\sigma_\tau^4\|K\|_2^2 \tag{A.10}$$

$$C_n := \sum_{i=1}^{n} \mathbb{E}[\beta_{n,i}^2|\mathcal{F}_{n,i-1}] \xrightarrow{P} g_\tau^2 \int\int K(x)K(y)C_1(x,y)dxdy \tag{A.11}$$

$$D_n := \sum_{i=1}^{n} \mathbb{E}[\alpha_{n,i}\beta_{n,i}|\mathcal{F}_{n,i-1}] \xrightarrow{P} 0. \tag{A.12}$$

The proof of (A.10) is embedded in the proof of Theorem 6.1. For (A.11), note that

$$C_n = h^{-1} g_{t_{j-1}}^2 \int_{t_{j-1}}^{T} U^2\left(\frac{t-\tau}{h}\right)dt \to g_\tau^2 \int_0^\infty U^2(s)ds,$$

and it is easy to see that $\int_0^\infty U^2(s)ds = \int\int K(x)K(y)C_1(x,y)dxdy$. It remains to show (A.12). To this end, note that, in terms of $U_{is} := \int_{t_{i-1}}^{s} \sigma_u dB_u$, for $i \geq j$, $\mathbb{E}[\alpha_{n,i}\beta_{n,i}|\mathcal{F}_{n,i-1}]$ can be written as

$$\begin{aligned}
&2h^{-1}K_h(t_{i-1}-\tau)g_{t_{j-1}}\mathbb{E}\left[\int_{t_{i-1}}^{t_i} U_{is}\sigma_s dB_s \int_{t_{i-1}}^{t_i} U\left(\frac{s-\tau}{h}\right)dW_s|\mathcal{F}_{n,i-1}\right]\\
&= 2h^{-1}K_h(t_{i-1}-\tau)\mathbb{E}\left[\int_{t_{i-1}}^{t_i} U_{is}\sigma_s U\left(\frac{s-\tau}{h}\right)ds|\mathcal{F}_{n,i-1}\right]\\
&= 2h^{-1}K_h(t_{i-1}-\tau)\int_{t_{i-1}}^{t_i} \mathbb{E}\left[U_{is}(\sigma_s - \sigma_{t_{i-1}})|\mathcal{F}_{n,i-1}\right] U\left(\frac{s-\tau}{h}\right)ds.
\end{aligned}$$

By Cauchy–Schwarz inequality, the expectation inside the integral can be shown to be $O_P(\Delta)$, uniformly in i. Thus, since $\Delta\sum_{i=j}^{n}|K_h(t_{i-1}-\tau)| \to \int |K(x)|dx$, as $n\to\infty$,

$$\begin{aligned}
D_n &\leq 2h^{-1}g_{t_{j-1}}O_P(\Delta)\rho\sum_{i=j}^{n}|K_h(t_{i-1}-\tau)|\int_{t_{i-1}}^{t_i}\left|U\left(\frac{s-\tau}{h}\right)\right|ds\\
&\leq 2h^{-1}g_{t_{j-1}}O_P(\Delta^2)\rho\sum_{i=j}^{n}|K_h(t_{i-1}-\tau)| = O_P(\Delta/h) = O_P(h).
\end{aligned}$$

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.

The final identity needed to conclude the CLT is $\sum_{i=1}^{n} \mathbb{E}[\xi_{n,i}^4 | \mathcal{F}_{n,i-1}] \xrightarrow{P} 0$, for which it suffices to show that

$$T_{1n} := h^{-2} \sum_{i=1}^{n} K_h^4(t_{i-1} - \tau) \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s \right)^8 \middle| \mathcal{F}_{n,i-1} \right] \xrightarrow{P} 0,$$

$$T_{2n} := h^{-2} \sum_{i=1}^{n} K_h^4(t_{i-1} - \tau) \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s^2 ds \right)^4 \middle| \mathcal{F}_{n,i-1} \right] \xrightarrow{P} 0,$$

$$T_{3n} := h^{-2} g_{t_{j-1}}^4 \sum_{i=j}^{n} \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} U\left(\frac{t-\tau}{h} \right) dW_t \right)^4 \middle| \mathcal{F}_{n,i-1} \right] \xrightarrow{P} 0.$$

The previous limits can be shown by applying BDG inequality and using the fact that σ is bounded. ■

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spa.2020.01.013.

References

[1] Aït-Sahalia, J. Jacod, High-Frequency Financial Econometrics, Princeton University Press, 2014.

[2] A. Alvarez, F. Panloup, M. Pontier, N. Savy, Estimation of the instantaneous volatility, Stat. Inference Stoch. Process. 15 (1) (2012) 27–59.

[3] O. Barndorff-Nielsen, A. Veraart, Stochastic volatility of volatility in continuous time, CREATES Research Paper 25, 2009.

[4] V. Epanechnikov, Non-parametric estimation of a multivariate probability density, Theory Probab. Appl. 14 (1) (1969) 153–158.

[5] J. Fan, Y. Wang, Spot volatility estimation for high-frequency data, Stat. Interface 1 (2) (2008) 279–288.

[6] D. Foster, D. Nelson, Continuous record asymptotics for rolling sample variance estimators, Econometrica 64 (1) (1996) 139–174.

[7] J. Jacod, P. Protter, Discretization of Processes, Springer-Verlag Berlin Heidelberg, 2012.

[8] J. Jacod, A.N. Shiryaev, Limit Theorems for Stochastic Processes, Springer-Verlag, Berlin, 2003.

[9] M. Jones, J. Marron, S. Sheather, A brief survey of bandwidth selection for density estimation, J. Amer. Statist. Assoc. 91 (433) (1996) 401–407.

[10] D. Kristensen, Nonparametric filtering of the realized spot volatility: A kernel-based approach, Econometric Theory 26 (01) (2010) 60–93.

[11] C. Mancini, V. Mattiussi, R. Renò, Spot volatility estimation using delta sequences, Finance Stoch. 19 (2) (2015) 261–293.

[12] M. Ossiander, E. Waymire, Certain positive-definite kernels, Proc. Amer. Math. Soc. 107 (2) (1989) 487–492.

[13] B. Park, J. Marron, Comparison of data-driven bandwidth selectors, J. Amer. Statist. Assoc. 85 (409) (1990) 66–72.

[14] G. Samorodnitsky, M. Taqqu, Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance, Chapman and Hall, New York, 1994.

[15] M. Vetter, Estimation of integrated volatility of volatility with applications to goodness-of-fit testing, Bernoulli 21 (4) (2015) 2393–2418.

[16] L. Zhang, P. Mykland, Y. Aït-Sahalia, A tale of two time scales, J. Amer. Statist. Assoc. 100 (472) (2005).

Please cite this article as: J.E. Figueroa-López and C. Li, Optimal kernel estimation of spot volatility of stochastic differential equations, Stochastic Processes and their Applications (2020), https://doi.org/10.1016/j.spa.2020.01.013.