J. Appl. Prob. **46,** 791–811 (2009)
Printed in England
© *Applied Probability Trust* 2009

DOUBLE KERNEL ESTIMATION OF SENSITIVITIES

ROMUALD ELIE,* *Université Paris-Dauphine and CREST*

Abstract

In this paper we address the general issue of estimating the sensitivity of the expectation of a random variable with respect to a parameter characterizing its evolution. In finance, for example, the sensitivities of the price of a contingent claim are called the *Greeks*. A new way of estimating the Greeks has recently been introduced in Elie, Fermanian and Touzi (2007) through a randomization of the parameter of interest combined with nonparametric estimation techniques. In this paper we study another type of estimator that turns out to be closely related to the score function, which is well known to be the optimal Greek weight. This estimator relies on the use of two distinct kernel functions and the main interest of this paper is to provide its asymptotic properties. Under a slightly more stringent condition, its rate of convergence is the same as the one of the estimator introduced in Elie, Fermanian and Touzi (2007) and outperforms the finite differences estimator. In addition to the technical interest of the proofs, this result is very encouraging in the dynamic of creating new types of estimator for the sensitivities.

Keywords: Sensitivity estimation; Monte Carlo simulation; nonparametric regression

2000 Mathematics Subject Classification: Primary 62G08
Secondary 11K45

1. Introduction

This paper is closely related to the work of Elie *et al.* [5], and so we will try to follow their notation. Let λ be some given parameter in $\mathbb{R}^d$, and define the function

$$V^{\phi}(\lambda) := \mathrm{E}[\phi(Z(\lambda))],$$

where $Z(\cdot)$ is a parameterized random variable with values in $\mathbb{R}^n$ and $\phi\colon \mathbb{R}^n \to \mathbb{R}$ is a measurable function. A well-understood issue is the numerical computation of the function $V^{\phi}(\lambda)$ by means of a Monte Carlo procedure for example. A more difficult problem consists in approximating the sensitivity of V^{ϕ} with respect to the parameter λ. For some given parameter λ^0, we denote by β^0 the expression of interest defined by

$$\beta^0 := \nabla_\lambda V^{\phi}(\lambda^0) = \nabla_\lambda \mathrm{E}[\phi(Z(\lambda))]_{|\lambda=\lambda^0}. \tag{1.1}$$

In financial applications, V^{ϕ} interprets as the no-arbitrage price of a contingent claim, defined by the payoff $\phi(Z(\lambda))$, in the context of a complete market with prices measured in terms of the price of the nonrisky asset. The sensitivities of V^{ϕ} with respect to the parameter λ are often called *Greeks*, and their interest to practitioners is now well established.

Received 3 April 2007; revision received 8 July 2009.
* Postal address: CEREMADE, CNRS UMR 7534, Université Paris-Dauphine, Place du Marechal de Lattre de Tassigny, 75775 Paris Cedex 16, France. Email address: elie@ensae.fr

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

To the author's knowledge, three methods are considered mainly for the computation of the sensitivities of V^ϕ. They are compared in detail in the survey paper of Kohatsu-Higa and Montero [9], and so we just briefly present their construction and main properties here.

First, the finite differences method consists in approximating the derivative of the price by its variation in response to a small perturbation ε of the parameter λ of interest:

$$\beta^0 \sim \frac{V^\phi(\lambda^0+\varepsilon) - V^\phi(\lambda^0)}{\varepsilon}. \tag{1.2}$$

Given a number of Monte Carlo simulations for the prices, the choice of ε is related to an equilibrium between the bias and the variance of the estimator. For discontinuous payoff functions ϕ, this method appears inefficient, owing to the poor precision of approximation (1.2). A theoretical study of these estimators is reported in [4], [10], or [12].

Second, we can invert the differentiation and the expectation operators to obtain the pathwise estimator given by a Monte Carlo estimation based on the representation

$$\beta^0 = \mathrm{E}[\phi'(Z(\lambda^0))\nabla_\lambda Z(\lambda^0)].$$

This method, introduced in [3], therefore requires a lot of regularity on the payoff function ϕ as well as the computation of the tangent process $\nabla_\lambda Z$ of the underlying. Efficient numerical schemes for the implementation of this method can be found in [7].

Finally, we can compute β^0 by reporting the differentiation operator on the regular distribution of the underlying $Z(\lambda)$. Whenever this random variable admits a density $f(\lambda,\cdot)$ with respect to the Lebesgue measure, we obtain the so-called likelihood ratio estimator based on

$$\beta^0 = \mathrm{E}[\phi(Z(\lambda^0))s(\lambda^0, Z(\lambda^0))] \quad \text{where} \quad s := \frac{\nabla_\lambda f}{f}. \tag{1.3}$$

The application of this trick in finance has also been introduced in [3]. This type of representation has been generalized by Fournié *et al.* [6], who studied the properties of the random variables π satisfying

$$\mathrm{E}[\phi(Z(\lambda^0))\pi] \quad \text{for any function } \phi \in \mathcal{L}^\infty(\mathbb{R}^n, \mathbb{R}).$$

By means of a Malliavin integration by parts argument, they characterized the set of so-called Greek weights π. After tedious computations, this characterization leads in some cases to explicit weights. Nevertheless, beyond all those Greek weight-based estimators, the one related to the score function s and given by (1.3) leads to the smallest variance.

As in [5], the main purpose of this paper is to study estimators of the Greek β^0 whenever the payoff function lacks regularity and the density f of the underlying is unknown. As detailed in the next section, a randomization of the parameter λ of interest allows us to rewrite the sensitivity β^0 given by (1.3) as a conditional expectation. Combining a nonparametric estimation of this conditional expectation with a truncation argument and a kernel estimation of the unknown score function s leads to our estimator $\tilde\beta_n$. A slightly different form of $\tilde\beta_n$, without the useful truncation modification, is presented in [5], where it serves as a basis to introduce other ones through an integration by parts argument. The main contribution of this paper is the presentation of the rather demanding derivation of its asymptotic properties suggested in [5]. The use of a truncated version of the classical kernel estimator allows us to reduce the induced required assumptions on the coefficients. We provide the asymptotic mean-square error and distribution of the proposed estimator, leading to the common calibration of the different parameters of simulation.

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

Despite the more general form of $\tilde{\beta}_n$, it surprisingly achieves the same rate of convergence as the estimator introduced in [5]. From a practical perspective, we have to admit that, as argued in [5], its numerical implementation is more demanding. Nevertheless, the choice of the two distinct kernel functions significantly increases the class of possible sensitivity estimators. From a technical point of view, the asymptotics of the estimator require a precise derivation of the properties of a kernel estimator of the score function, which appear to be of great interest in themselves. Therefore, this paper offers a new contribution to the current literature on the combination of several nonparametric estimators, and its particular application to the computation of the Greeks is furthermore promising in the development of competitive numerical computations of sensitivities.

The paper is organized as follows. In Section 2 we present in detail the construction of this estimator. Its asymptotic properties as well as its practical implementation are discussed in Section 3. Finally, for ease of presentation, the proofs are reported in Section 4.

2. Construction of the estimator

Throughout this paper, we consider a complete probability space $(\Omega, \mathcal{F}, \mathrm{P})$ supporting a Brownian motion W valued in $\mathbb{R}^m$. We assume that $\mathcal{F}$ is the P-completion of the σ-algebra generated by W. Let $Z(\lambda)$ be a given random variable valued in $\mathbb{R}^n$ and parameterized by $\lambda \in \mathbb{R}^d$, and let $\phi \in \mathcal{L}^\infty(\mathbb{R}^n, \mathbb{R})$ be a payoff function. The purpose of this paper is to construct an estimator of β^0 defined in (1.1) as the sensitivity of V^ϕ with respect to λ at a given point λ^0.

We shall demonstrate in this section the intuition behind the construction of the suggested estimator. We first identify the score function s defined in (1.3) as the optimal Greek weight in the sense of [6]. Considering the realistic case where the score function is unknown, we propose to approximate it through a kernel estimation procedure. Combining Monte Carlo simulations with the randomization of the parameter λ, we are able to construct a nonparametric estimator of the score function, leading naturally to the estimation of β^0. The reader interested in the asymptotic properties of the estimator should progress directly to Section 3.

2.1. The score function as the optimal Greek weight

We assume that the distribution of $Z(\lambda)$ is absolutely continuous with respect to the Lebesgue measure, and denote by $f(\lambda, \cdot)$ the associated density. As mentioned in the introduction, under mild smoothness assumptions on the density f, we directly compute that

$$\beta^0 = \mathrm{E}[\phi(Z(\lambda^0))s(\lambda^0, Z(\lambda^0))], \quad \text{where} \quad s := \frac{\nabla_\lambda f}{f} = \nabla_\lambda \ln f.$$

In the context of the Black–Scholes model, Broadie and Glasserman [3] noticed that this representation allows β^0 to be computed by a direct Monte Carlo procedure. It is important to note that the score function s depends only on the distribution of the underlying $Z(\lambda^0)$. In a more general framework, Fournie *et al.* [6] considered the set

$$\mathcal{W} := \{\pi \in \mathcal{L}^2(\Omega, \mathbb{R}^d) \colon \nabla_\lambda V^\phi(\lambda^0) = \mathrm{E}[\phi(Z^0)\pi] \text{ for all } \phi \in \mathcal{L}^\infty(\mathbb{R}^n, \mathbb{R})\}.$$

Assuming that $\mathrm{E}[|s(\lambda^0, Z(\lambda^0))|^2] < \infty$, we note that $s(\lambda^0, Z(\lambda^0)) \in \mathcal{W}$. In [6], the authors constructed a new characterization of the set $\mathcal{W}$ by means of a Malliavin integration by parts argument. After rather tedious computations, this representation sometimes allowed some alternative Greek weights π to the score $s(\lambda^0, Z(\lambda^0))$ to be produced. When the density f and, therefore, the score function s of the underlying are unknown, these alternative weights appear to be very helpful.

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

Nevertheless, their obtention is unfortunately still limited to particular cases and the following argument demonstrates that the estimator based on the score $s(\lambda^0, Z(\lambda^0))$ is of minimal variance beyond the class of Greek weight-based estimators. Indeed, from the arbitrariness of $\phi \in \mathcal{L}^\infty(\mathbb{R}^n, \mathbb{R})$, we rewrite

$$\mathcal{W} = \{\pi \in \mathcal{L}^2(\Omega, \mathbb{R}^d)\colon \mathrm{E}[\pi \mid Z(\lambda^0)] = s(\lambda^0, Z(\lambda^0))\}.$$

We then deduce that, for any $\pi \in \mathcal{W}$,

$$\begin{aligned}
\mathrm{var}[\phi[Z(\lambda^0)]\pi] &= \mathrm{E}[\phi(Z(\lambda^0))^2\, \mathrm{E}[\pi\pi^\top \mid Z(\lambda^0)]] - \nabla V^\phi(\lambda^0)\nabla V^\phi(\lambda^0)^\top \\
&\geq \mathrm{E}[\phi(Z(\lambda^0))^2\, \mathrm{E}[\pi \mid Z(\lambda^0)]\, \mathrm{E}[\pi \mid Z(\lambda^0)]^\top] - \nabla V^\phi(\lambda^0)\nabla V^\phi(\lambda^0)^\top \\
&= \mathrm{E}[\phi(Z(\lambda^0))^2 s(\lambda^0, Z(\lambda^0)) s(\lambda^0, Z(\lambda^0))^\top] - \nabla V^\phi(\lambda^0)\nabla V^\phi(\lambda^0)^\top \\
&= \mathrm{var}[\phi(Z(\lambda^0)) s(\lambda^0, Z(\lambda^0))],
\end{aligned}$$

where '$\top$' denotes the transposition operator. Hence,

$$s(\lambda^0, Z(\lambda^0)) \in \mathcal{W} \quad \text{is a minimizer of} \quad \mathrm{var}[\phi(Z(\lambda^0))\pi], \qquad \pi \in \mathcal{W}.$$

As in [5], in this paper we intend to construct a nonparametric estimator based on the approximation of the optimal Greek weight given by the unknown score $s(\lambda^0, Z(\lambda^0))$.

2.2. Randomization of the parameter

In order to be able to estimate the unknown score function s, the idea is to create an artificial density around the parameter λ^0, on which we can report the differentiation operation. This well-known technique in the nonparametric statistics literature (see, e.g. [1]) is based on the randomization of the parameter λ of interest. We may, for example, interpret the classical finite difference operator (1.2) as a particular case of a randomizing distribution of λ with two dirac masses at points λ^0 and $\lambda^0 + \varepsilon$.

We then introduce $\ell\colon \mathbb{R}^d \to \mathbb{R}$, some given probability density function, with support containing the origin in its interior and set

$$\varphi(\lambda, z) := \ell(\lambda^0 - \lambda) f(\lambda, z) \quad \text{for } \lambda \in \mathbb{R}^d \text{and} z \in \mathbb{R}^n.$$

Considering a couple of random variables (Λ, Z) with density φ, we therefore rewrite β^0 as

$$\beta^0 = \mathrm{E}[\phi(Z) s(\Lambda, Z) \mid \Lambda = \lambda^0]. \tag{2.1}$$

Although we restrict to the case where the density f of the underlying $Z(\lambda)$ is unknown, we still consider that we can simulate $Z(\lambda)$. This is not a limitation in practice since $Z(\lambda)$ is typically characterized by a stochastic differential equation, which can be classically discretized. Hence, we introduce a sequence

$$(\Lambda_i, Z_i)_{1\leq i\leq N} \text{ of } N \text{ independent random variables with distribution } \varphi,$$

so that, for any $i \leq N$, $\ell(\lambda^0 - \cdot)$ is the density of Λ^i and $f(\Lambda^i, \cdot)$ is the conditional density of Z^i given Λ^i.

We now introduce a kernel function $K : \mathbb{R}^d \to \mathbb{R}$, i.e. such that $\int_{\mathbb{R}^d} K\, \mathrm{d}l = 1$. Given the N observations $(\Lambda_i, Z_i)_{1\leq i\leq N}$, the conditional expectation given by (2.1) can be approximated by the classical kernel estimator

$$\bar{\beta}_N := \frac{1}{\ell(0) N h^d} \sum_{i=1}^N \phi(Z_i) s(\Lambda_i, Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h}\right), \tag{2.2}$$

where the bandwidth $h > 0$ of the estimator is a small parameter.

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

This estimator is of course not implementable since the score function s is unknown. Nevertheless, as detailed in the next subsection, the extra regular source of randomness introduced by ℓ allows us to approximate s and leads to a computable estimator of β^0.

2.3. The double kernel-based estimator

In order to approximate the score function s, we shall first estimate the unknown density φ of (Λ, Z). For this purpose, we introduce a second kernel function $H\colon \mathbb{R}^n \to \mathbb{R}$. Given $N-1$ observations $(\Lambda_j, Z_j)_{1\le j\le N,\, j\neq i}$, we define $\hat{\varphi}^{-i}$ to be the classical nonparametric estimator of the density φ, given by

$$\hat{\varphi}^{-i}(\lambda, z) := \frac{h^{-d-n}}{N-1} \sum_{j=1,\, j\neq i}^{N} K\left(\frac{\lambda-\Lambda_j}{h}\right) H\left(\frac{z-Z_j}{h}\right). \tag{2.3}$$

We denote by $\hat{\varphi_\lambda}^{-i}(\lambda, z)$ the derivative of this estimator with respect to λ and we deduce that

$$\hat{\varphi_\lambda}^{-i}(\lambda, z) := \nabla_\lambda \hat{\varphi}^{-i}(\lambda, z) = \frac{h^{-d-n-1}}{N-1} \sum_{j=1,\, j\neq i}^{N} \nabla K\left(\frac{\lambda-\Lambda_j}{h}\right) H\left(\frac{z-Z_j}{h}\right).$$

Observe now that s and φ are closely related since we easily compute

$$s(\lambda, z) = \frac{\nabla_\lambda f(\lambda, z)}{f(\lambda, z)} = \frac{\nabla_\lambda \varphi(\lambda, z)}{\varphi(\lambda, z)} - \frac{\nabla \ell(\lambda^0-\lambda)}{\ell(\lambda^0-\lambda)} \quad \text{for } \lambda \in \mathbb{R}^d \text{ and } z \in \mathbb{R}^n.$$

Given the observations $(\Lambda_j, Z_j)_{1\le j\le N,\, j\neq i}$, this naturally leads to the following estimator of the score function s:

$$\hat{s}_N^{-i}(\lambda, z) := \frac{\hat{\varphi_\lambda}^{-i}(\lambda, z)}{\hat{\varphi}^{-i}(\lambda, z) + (\delta/3 - \hat{\varphi}^{-i}(\lambda, z))\,\mathbf{1}_{\{|\hat{\varphi}^{-i}(\lambda,z)|<\delta/3\}}} + \frac{\nabla \ell(\lambda^0-\lambda)}{\ell(\lambda^0-\lambda)}, \tag{2.4}$$

where δ is some small fixed parameter ensuring that the estimator $\hat{\varphi}^{-i}$ stays away from 0. This technical truncation will simply ensure the nonexplosion of the estimator, and the convergence of the estimator will necessitate some control on the small values of the true density φ detailed in Assumption 3.2, below.

In order to construct an estimator of β^0, we now replace each score $s(\Lambda_i, Z_i)$ in (2.2) by the approximation $\hat{s}_N^{-i}(\Lambda_i, Z_i)$ based on the $N-1$ remaining observations. Our estimator is thus defined by

$$\tilde{\beta}_N := \frac{1}{\ell(0) N h^d} \sum_{i=1}^{N} \phi(Z_i) \hat{s}_N^{-i}(\Lambda_i, Z_i) K\left(\frac{\lambda^0-\Lambda_i}{h}\right). \tag{2.5}$$

Based on this type of representation, Elie *et al.* [5] introduced two other estimators by means of an integration by parts argument. Even if the representations proposed in [5] appear more simple, we surprisingly show in the next section that our estimator (2.5) achieves a similar rate of convergence, under a few more stringent conditions. Even if the practical implementation and computation of $\tilde{\beta}_N$ is more time consuming, the general form of (2.4) offers a large class of possible estimators, related to different kernel functions K and H. Since the rate of convergence of these estimators is similar, we sincerely believe that this result is very encouraging in the dynamic of creating new types of estimator for the sensitivities. Moreover, the technical proof for the convergence of the estimator appears to be of great interest in itself.

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

3. Asymptotic properties

In this section we present the main results of the paper. We first provide the asymptotic properties of the estimator $\tilde{\beta}_N$ defined in (2.5). In particular, the obtention of the asymptotic mean-square error of the estimator leads to the common optimal choice of the number of simulations N and the bandwidth h of the two kernel functions K and H.

3.1. Notation

Before stating our results, we recall that the order of a kernel function $K\colon \mathbb{R}^d \to \mathbb{R}$ is defined as the smallest nonzero integer p such that there exist some integers $(j_1, \dots, j_p)$, with $j_k \in \{1, \dots, d\}$, satisfying

$$\int_{\mathbb{R}^d} l_{\alpha_1} \cdots l_{\alpha_r} K(l)\,\mathrm{d}l = 0 \quad \text{for } 0 < r < p,\ \alpha_k \in \{1, \dots, d\},$$
$$\text{and} \quad \int_{\mathbb{R}^d} l_{j_1} \cdots l_{j_p} K(l)\,\mathrm{d}l \neq 0.$$

Typically, if K is the product of d even univariate kernels then it is (at least) of order $p = 2$.

In the subsequent subsections, the kernel functions K and H will be respectively of order p and q, and we shall use the notation

$$\xi_K^p[\psi](\lambda, z) := \frac{(-1)^p}{p!} \sum_{j_1,\dots,j_p=1}^{d} \biggl(\int_{\mathbb{R}^d} l_{j_1} \cdots l_{j_p} K(l)\,\mathrm{d}l\biggr) \nabla^p_{\lambda_{j_1}\cdots\lambda_{j_p}} \psi(\lambda, z),$$
$$\xi_H^q[\psi](\lambda, z) := \frac{(-1)^q}{q!} \sum_{j_1,\dots,j_q=1}^{d} \biggl(\int_{\mathbb{R}^d} v_{j_1} \cdots v_{j_q} H(v)\,\mathrm{d}v\biggr) \nabla^q_{z_{j_1}\cdots z_{j_q}} \psi(\lambda, z),$$

for every smooth function ψ defined on $\mathbb{R}^d \times \mathbb{R}^n$. We shall also define $A^{\otimes} := AA^{\top}$ for every matrix A, and let C denote a constant whose value may change from line to line.

3.2. Asymptotic moments and distribution of the estimator

We shall work under the following three assumptions respectively concerning the kernels K and H, the payoff function ϕ, and the unknown density function f.

Assumption 3.1. *The kernels K and H are the product of some univariate, compactly supported Lipschitz kernels with orders p and q, respectively, and ∇K has bounded variation.*

Assumption 3.2. *The payoff function ϕ is continuous and has compact support. Moreover, there exists $\delta > 0$ such that, for every $z \in \mathbb{R}^n$, $\inf\{\varphi(\lambda, z)\colon (\lambda, z) \in \mathcal{V}(\lambda^0) \times C_\phi\} > \delta$ for some neighborhood $\mathcal{V}(\lambda^0)$ of λ^0, and some compact subset C_ϕ of $\mathbb{R}^n$ with* $\operatorname{supp}(\phi) \subset \operatorname{int}(C_\phi)$.

Assumption 3.3. *For every λ, the function $\nabla_\lambda f(\lambda, \cdot)$ is q times differentiable and, for every integer $j \leq q$, the function $\lambda \mapsto \nabla_z^j \nabla_\lambda \varphi(\lambda, z)$ is continuous at $\lambda = \lambda^0$ uniformly with respect to $z \in S$ for some subset S such that* $\operatorname{supp}(\phi) \subset \operatorname{int}(S)$.

For every z, the functions $f(\cdot, z)$ and ℓ are $p + 1$ times differentiable and, for every integer $i \leq p + 1$, the function $\lambda \mapsto \nabla_\lambda^i f(\lambda, z)$ is continuous at λ^0 uniformly with respects to $z \in S$ for some subset S such that $\operatorname{supp}(\phi) \subset \operatorname{int}(S)$.

Remark 3.1. We have to admit that Assumption 3.2 is at first glance rather restrictive on the class of possible payoff functions for financial applications. Nevertheless, we observe that most

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

of the classical payoff functions can be included. In particular, the call option can be considered here even if the payoff does not have compact support. We just need to approximate the Greeks associated to the corresponding put option and use the correspondence provided by the call-put parity relation satisfied in any arbitrage-free market.

We first present the asymptotic bias and variance of the estimator.

Proposition 3.1. *Under Assumptions 3.1, 3.2, and 3.3, choose N and h so that*

$$h \to 0 \quad \textit{and} \quad \frac{(\ln N)^4}{N h^{d+n+n\vee 2}} \to 0 \quad \textit{as } N \to \infty. \tag{3.1}$$

Then, the bias and the variance of $\tilde{\beta}_N$ satisfy

$$\mathrm{E}[\tilde{\beta}_N] - \beta^0 \sim C_1 h^p + C_2 h^q + \frac{C_3}{N h^{d+n+1}} \quad \textit{and} \quad \mathrm{var}[\tilde{\beta}_N] \sim \frac{\tilde{\Sigma}}{N h^{d+2}}, \tag{3.2}$$

where

$$C_1 := \frac{1}{\ell(0)} \int_{\mathbb{R}^d} \bigg(\xi_K^p[\ell(\lambda^0 - \cdot) f_\lambda + \varphi_\lambda] - \frac{\varphi_\lambda}{\varphi} \xi_K^p[\varphi] \bigg)(\lambda^0, z) \phi(z)\, \mathrm{d}z,$$

$$C_2 := \frac{1}{\ell(0)} \int_{\mathbb{R}^d} \bigg(\xi_H^q[\varphi_\lambda] - \frac{\varphi_\lambda}{\varphi} \xi_H^q[\varphi] \bigg)(\lambda^0, z) \phi(z)\, \mathrm{d}z,$$

$$C_3 := \frac{1}{\ell(0)} \int_{(\mathbb{R}^d)^4} \frac{\phi(z)}{\varphi(\lambda^0, z)} K(l_2 - l_1) K(l_1) \nabla K(l_1) H^2(v)\, \mathrm{d}l_1\, \mathrm{d}l_2\, \mathrm{d}v\, \mathrm{d}z,$$

$$\tilde{\Sigma} := \frac{\mathrm{E}[\phi^2(Z^0)]}{\ell(0)} \int_{\mathbb{R}^d} \bigg(\int_{\mathbb{R}^d} K(l_2 - l_1) \nabla K(l_1)\, \mathrm{d}l_1 \bigg\}^{\otimes} \mathrm{d}l_2.$$

We now turn to the asymptotic distribution of the estimator.

Theorem 3.1. (i) *Under the conditions of Proposition 3.1, we have*

$$\sqrt{N h^{d+2}} (\tilde{\beta}_N - \mathrm{E}[\tilde{\beta}_N]) \xrightarrow{\text{LAW}} \mathcal{N}(0, \tilde{\Sigma}) \quad \textit{as } N \to \infty.$$

(ii) *If, in addition, $N h^{d+2+2(p\wedge q)} \to 0$ then the bias vanishes and*

$$\sqrt{N h^{d+2}} (\tilde{\beta}_N - \beta^0) \xrightarrow{\text{LAW}} \mathcal{N}(0, \tilde{\Sigma}) \quad \textit{as } N \to \infty.$$

The technical proofs of Proposition 3.1 and Theorem 3.1 are reported in Section 4.

Remark 3.2. Note that the condition $n < (p \wedge q) + 1$ is necessary in order to satisfy (3.1) and the condition of Theorem 3.1(ii). Thus, for basket derivatives or Bermudan options in finance, it is necessary to consider high-order kernels, which is not a limitation in practice.

3.3. Dependence with respect to the price process dynamics

We should typically imagine the random variable Z as the terminal value of a price process X^λ, whose dynamics are given by a parametrized stochastic differential equation of the form:

$$X_0^\lambda = x(\lambda), \qquad \mathrm{d}X_u^\lambda = \mu(u, \lambda, X_u^\lambda)\, \mathrm{d}u + \sigma(u, \lambda, X_u^\lambda)\, \mathrm{d}W_u, \tag{3.3}$$

where $x \colon \mathbb{R}^d \to \mathbb{R}^n$, $\mu \colon [0, T] \times \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$, and $\sigma \colon [0, T] \times \mathbb{R}^d \times \mathbb{R}^n \to \mathcal{M}_{\mathbb{R}}^{n,m}$ are deterministic Lipschitz functions. In this case, $Z = X_T^\lambda$ can be simulated easily via any time discretization scheme, even if its density f is unknown.

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

We detail in this paragraph how the regularity of f required in Assumption 3.3 can be induced from conditions on the coefficients x, μ, and σ. First, the absolute continuity of X_T^λ is ensured by the classical uniform ellipticity condition: suppose that the matrix $\sigma\sigma^\top$ is symmetric and positive, and that there exists a constant $c_\sigma > 1$ such that

$$\frac{1}{c_\sigma} I_d(x) \le \sigma\sigma^\top(t, \lambda, x) \le c_\sigma I_d(x) \quad \text{for all } (t, \lambda, x) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^n. \tag{3.4}$$

Second, the density f of X_T^λ inherits the regularity of the coefficients x, μ, and σ through the properties of the corresponding transition densities. Following the arguments of Theorem A.2.2 of [2, p. 478] (see also Proposition 5.1 of [8]), Assumption 3.3 is satisfied whenever (3.4) holds, ℓ is of class C^1, x is of class C^{q+2}, and the coefficients μ and σ are of class C^1 in (t, λ, x), C^{p+2} in λ, as well as C^{q+2} in x.

It is worth noting that this analysis gives rise to more tractable assumptions for Proposition 3.1 and Theorem 3.1 in the realistic framework where Z is the terminal value of a price process with dynamics of the form (3.3).

3.4. Optimal choices of N and h

We investigate in this section the optimal balance between the number of simulations N and the bandwidth h. As mentioned in Remark 3.2, we suppose that $n < (p \wedge q) + 1$. Under this condition and the assumptions of Proposition 3.1, we obtain a simplification in the asymptotic expression of the bias, and the mean-square error (MSE) of the estimator becomes

$$\mathrm{MSE}(\tilde\beta_N) := \mathrm{E}[|\tilde\beta_N - \beta^0|^2] \sim \frac{\mathrm{tr}(\tilde\Sigma)}{Nh^{d+2}} + |C_1|^2 h^{2p} + |C_2|^2 h^{2q}.$$

Minimizing the MSE in h, we obtain the asymptotically optimal bandwidth selector:

$$\tilde h = \left(\frac{(d+2)\mathrm{tr}(\tilde\Sigma)}{2(p\wedge q)|C_1 \mathbf{1}_{\{p\le q\}} + C_2 \mathbf{1}_{\{q \le p\}}|^2 N} \right)^{1/(d+2(p\wedge q)+2)}. \tag{3.5}$$

Therefore, $\tilde h$ is of order $N^{-1/(d+2(p\wedge q)+2)}$, leading to an MSE of order $N^{-2(p\wedge q)/(d+2(p\wedge q)+2)}$. Consequently, despite its more complicated form, the double kernel estimator achieves the same rate of convergence as the one introduced in [5]. The only constraint is the use of kernel functions of order sufficiently large, i.e. satisfying $p \wedge q > n - 1$. Since, given a large number of simulations, we should always use a kernel function of high order, this constraint is not relevant in practice.

3.5. Remarks and extensions

In this section we regroup some remarks and possible extensions of the method, which unfortunately go beyond the scope of the paper.

Considering a randomizing distribution ℓ with radius equal to the bandwidth h, we can improve the rate of convergence of the estimator. Indeed, the asymptotic variance of the estimator then reduces to a term of order $1/\sqrt{Nh^2}$, leading to an MSE of order $N^{-(p\wedge q)/(p\wedge q)+1}$. Remarkably, the speed of convergence of the estimator does not depend in this case on the dimension of the underlying X. For a continuous payoff function, the best finite differences estimator achieves an MSE of order $N^{-4/5}$; see [4]. Therefore, this estimator outperforms the finite differences one as soon as $p \wedge q > 4 \vee (n-1)$. We choose to omit the proof of this result, which is technically rather demanding.

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

With no doubt, the choice of the randomizing function ℓ is crucial for the precision of the estimator presented here. In the particular case of a uniform randomizing distribution ℓ, the analytical form of the estimator simplifies and, after tedious asymptotic developments, we can see that the optimal choice for the radius of the distribution ℓ is the bandwidth h of the kernel function K, i.e. the particular case discussed above. From an empirical point of view, the optimal choice of the randomizing density ℓ should be intimately related to the choice of the kernel function K. A simple example where these two density functions are identical can naturally be considered.

As for the practical calibration of the optimal bandwidth $\tilde{h}$ given by (3.5), we need to estimate the constants C_1, C_2, and $\tilde{\Sigma}$. As for the choice of the bumping parameter of the finite differences estimator, it can be approximated by a preliminary Monte Carlo procedure with very few simulations. For example, the procedure proposed in [5] can be directly adapted to this setting.

Finally, a generalization of the above estimator could be considered by taking two different bandwidths. Intuitively, the bandwidth for the estimation of the score function introduced in (2.3) should be smaller than the one considered for the approximation of the conditional expectation in (2.2). Indeed, the signification of these two parameters is rather different, but this question is left for further research.

4. Proofs

This section is dedicated to the proofs of Proposition 3.1 and Theorem 3.1, characterizing the asymptotic behavior of $\tilde{\beta}_N$. In this section we shall always work under the assumptions of Proposition 3.1.

4.1. Preliminaries

Recall that

$$\tilde{\beta}_N := \frac{1}{\ell(0)Nh^d} \sum_{i=1}^{N} \phi(Z_i)\hat{s}_N^{-i}(\Lambda_i, Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h}\right), \tag{4.1}$$

where

$$\hat{s}_N^{-i}(\lambda, z) := \frac{\hat{\varphi_\lambda}^{-i}(\lambda, z)}{\hat{\varphi}^{-i,\delta}(\lambda, z)} + \frac{\nabla\ell(\lambda^0 - \lambda)}{\ell(\lambda^0 - \lambda)},$$

where $\hat{\varphi}^{-i,\delta} := \hat{\varphi}^{-i} + (\delta/3 - \hat{\varphi}^{-i})\, \mathbf{1}_{\{|\hat{\varphi}^{-i}| \leq \delta/3\}}$ is a truncated version of $\hat{\varphi}^{-i}(\lambda, z)$ defined by

$$\hat{\varphi}^{-i}(\lambda, z) := \frac{h^{-d-n}}{N-1} \sum_{j=1,\, j\neq i}^{N} K\left(\frac{\lambda - \Lambda_j}{h}\right) H\left(\frac{z - Z_j}{h}\right) \quad \text{and} \quad \hat{\varphi_\lambda}^{-i} = \nabla_\lambda \hat{\varphi}^{-i}.$$

For every λ and z, we set

$$\bar{\varphi}(\lambda, z) := \mathrm{E}[\hat{\varphi}^{-1}(\lambda, z)] = \int_{\mathbb{R}^d \times \mathbb{R}^d} K(l)H(v)\varphi(\lambda - hl, z - hv)\, \mathrm{d}l\, \mathrm{d}v,$$

and its derivative is given by

$$\bar{\varphi}_\lambda(\lambda, z) = h^{-1} \int_{\mathbb{R}^d \times \mathbb{R}^d} \nabla K(l)H(v)\varphi(\lambda - hl, z - hv)\, \mathrm{d}l\, \mathrm{d}v.$$

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

Arguing as in the proof of Proposition 4.1 of [5], a Taylor expansion combined with a classical change of variable leads to

$$\bar{\varphi}(\lambda, z) - \varphi(\lambda, z) = \xi_K^p[\varphi](\lambda, z)h^p + \xi_H^q[\varphi](\lambda, z)h^q + o(h^{p\wedge q}). \tag{4.2}$$

Similarly, we obtain

$$\bar{\varphi}_\lambda(\lambda, z) - \varphi_\lambda(\lambda, z) = \xi_K^p[\varphi_\lambda](\lambda, z)h^p + \xi_H^q[\varphi_\lambda](\lambda, z)h^q + o(h^{p\wedge q}). \tag{4.3}$$

Remark 4.1. Since ϕ and K have compact support by Assumption 3.2, it follows that, for sufficiently small h, the sum in (4.1) is restricted to pairs (Λ_i, Z_i) with values in $C_K \times C_\phi$, where $C_K \subset \mathcal{V}(\lambda^0)$ is defined in Assumption 3.2 and C_ϕ is a compact subset of $\mathbb{R}^n$ such that $\operatorname{supp}\phi \subset C_\phi$.

For any function ψ defined on $C_K \times C_\phi$, we set

$$||\psi||_\infty := \sup_{(\lambda,z)\in C_K\times C_\phi} |\psi(\lambda, z)|,$$

and, in the following, $||\cdot||_r$ refers to the $\mathcal{L}_r(\Omega)$-norm.

Remark 4.2. By Assumption 3.3, since (λ, z) vary in a compact subset of $\mathbb{R}^d \times \mathbb{R}^n$, the remainder terms in (4.2) and (4.3) are uniformly bounded in (λ, z). By the same argument, we also see that $\xi_K^p[\varphi]$, $\xi_H^q[\varphi]$, $\xi_K^p[\varphi_\lambda]$, and $\xi_H^q[\varphi_\lambda]$ are uniformly bounded so that

$$\|\bar{\varphi} - \varphi\|_\infty = O(h^{p\wedge q}) \quad \text{and} \quad \|\bar{\varphi}_\lambda - \varphi_\lambda\|_\infty = O(h^{p\wedge q}). \tag{4.4}$$

We now study further the tails of the estimators $\hat{\varphi}^{-i}$ and we obtain the following estimates.

Lemma 4.1. *There exist α_1 and α_2 such that*

$$\sup_{i\le N} \mathrm{P}[|\hat{\varphi}^{-i} - \bar{\varphi}|(\lambda, z) > t] \le 2\exp\biggl(-\frac{t^2}{\alpha_1 + \alpha_2 t} N h^{d+n}\biggr), \qquad (\lambda, z) \in C_K \times C_\phi. \tag{4.5}$$

Furthermore, for any $t > 0$, there exist $C_t > 0$ and $c_t > 0$ satisfying

$$\mathrm{P}\Bigl[\sup_{i\le N} \|\hat{\varphi}^{-i} - \bar{\varphi}\|_\infty > t\Bigr] \le C_t N^3 \exp(-c_t N h^{d+n}). \tag{4.6}$$

Finally, for any integer $r \ge 1$, we have

$$\Bigl\| \sup_{1\le i\le N} \|\hat{\varphi}^{-i} - \bar{\varphi}\|_\infty \Bigr\|_{2r} = O\biggl(\frac{\ln(N)}{\sqrt{N h^{d+n}}}\biggr). \tag{4.7}$$

Proof. Observe first that there exist α_1 and α_2 such that, for any $(\lambda, z) \in C_K \times C_\phi$, the random variables $K[(\lambda - \Lambda^i)/h]H[(z - z^i)/h]$ are bounded by $3\alpha_2/2$ and, by the usual change of variable, their variance is bounded from above by $\alpha_1 h^{d+n}/2$. Therefore, (4.5) follows directly from the Bernstein inequality.

We now turn to the proof of the second estimate and first observe that

$$\mathrm{P}\Bigl[\sup_{i\le N} \|\hat{\varphi}^{-i} - \bar{\varphi}\|_\infty > t\Bigr] \le N\,\mathrm{P}[\|\hat{\varphi} - \bar{\varphi}\|_\infty > t], \tag{4.8}$$

where, for ease of notation in this proof, we introduce $\hat{\varphi} := \hat{\varphi}^{-1}$. Applying the Liebscher strategy (see [11]), we recover the compact set $C_K \times C_\phi$ by $C_0(R_{N,h})^{-d-n}$ balls $B_j := B((\lambda_j, z_j), R_{N,h})$, where C_0 is a constant chosen large enough. On each ball B_j, we have

$$\sup_{B_j} |\hat{\varphi} - \bar{\varphi}| \leq |\hat{\varphi} - \bar{\varphi}|(\lambda_j, z_j) + \sup_{(\lambda,z)\in B_j} |\hat{\varphi}(\lambda, z) - \hat{\varphi}(\lambda_j, z_j)| + \sup_{(\lambda,z)\in B_j} |\bar{\varphi}(\lambda, z) - \bar{\varphi}(\lambda_j, z_j)|. \tag{4.9}$$

According to Assumption 3.1, the kernel functions K and H are Lipschitz and compactly supported. Therefore, there exists an $M > 0$ such that

$$\sup_{(\lambda,z)\in B_j} |\hat{\varphi}(\lambda, z) - \hat{\varphi}(\lambda_j, z_j)| \leq C \frac{R_{N,h}}{h} \hat{\psi}(\lambda_j, z_j),$$

where $\hat{\psi}$ is the classical histogram kernel estimator of the density φ defined by

$$\hat{\psi}(\lambda, z) := \frac{1}{4M^2 N h^{d+n}} \sum_{i=1}^{N} \mathbf{1}_{\{|\Lambda_i - \lambda| \leq Mh\}} \mathbf{1}_{\{|Z_i - z| \leq Mh\}}.$$

Introducing the notation $\bar{\psi} := \mathrm{E}[\hat{\psi}]$ and choosing $R_{N,h}$ such that $R_{N,h} = o(h)$, we then deduce from (4.9) that

$$\sup_{B_j} |\hat{\varphi} - \bar{\varphi}| \leq |\hat{\varphi} - \bar{\varphi}|(\lambda_j, z_j) + |\hat{\psi} - \bar{\psi}|(\lambda_j, z_j) + 2C \frac{R_{N,h}}{h} \bar{\psi}(\lambda_j, z_j).$$

Summing over all the balls B_j, we obtain

$$\mathrm{P}[\|\hat{\varphi} - \bar{\varphi}\|_\infty > t] \leq C_0 R_{N,h}^{-(d+n)} \left(\mathrm{P}\left[|\hat{\varphi} - \bar{\varphi}|(\lambda_j, z_j) > \frac{t}{3} \right] + \mathrm{P}\left[|\hat{\psi} - \bar{\psi}|(\lambda_j, z_j) > \frac{t}{3} \right] \right) + C_0 R_{N,h}^{-(d+n)} \, \mathrm{P}\left[2Ch^{-1} R_{N,h} |\bar{\psi}|(\lambda_j, z_j) > \frac{t}{3} \right].$$

Therefore, applying estimate (4.5) to both kernel estimators $\hat{\varphi}$ and $\hat{\psi}$, we deduce the existence of γ_1 and γ_2 satisfying

$$\mathrm{P}[\|\hat{\varphi} - \bar{\varphi}\|_\infty > t] \leq C R_{N,h}^{-(d+n)} \left(\exp\left(-\frac{t^2}{\gamma_1 + \gamma_2 t} N h^{d+n} \right) + \mathrm{P}\left[2C \frac{R_{N,h}}{h} |\bar{\psi}|(\lambda_j, z_j) > \frac{t}{3} \right] \right). \tag{4.10}$$

But $\bar{\psi}$ is bounded so that for any given t, the last term on the right-hand side equals 0 for small enough h. Since $Nh^{d+n} \to \infty$ according to (3.1), choosing $R_{N,h} = h^2$, we deduce (4.6) from (4.8).

We now turn to the moment inequalities and introduce the notation

$$Y_N := \frac{\sqrt{Nh^{d+n}}}{\ln(N)} \sup_{i \leq N} \|\hat{\varphi}^{-i} - \bar{\varphi}\|_\infty,$$

so that we simply need to prove that $\|Y_N\|_{2r} < \infty$ for all integers $r \geq 1$. Fix $r \in \mathbb{N}^*$, and observe that

$$\mathrm{E}[Y_N^{2r}] = \int_0^\infty 2rs^{2r-1} \, \mathrm{P}[Y_N > s] \, \mathrm{d}s \leq C_a + \int_a^\infty 2rs^{2r-1} \, \mathrm{P}[Y_N > s] \, \mathrm{d}s \tag{4.11}$$

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

for any $a > 0$. We now fix s large enough and take $R_{N,h} = h\ln(N)/\sqrt{Nh^{d+n}}$ in (4.10) and (4.8), so that we obtain, for large enough N, the existence of δ_1 and δ_2 satisfying

$$\mathrm{P}[Y_N > s] \le CN\biggl(\frac{\sqrt{Nh^{d+n}}}{h\ln(N)}\biggr)^{d+n}\exp\biggl(-\frac{s\ln(N)^2}{\delta_1+\delta_2 s\ln(N)/\sqrt{Nh^{d+n}}}\biggr).$$

Since $\ln(N)/\sqrt{Nh^{d+n}} \to 0$ and $h \to 0$, we deduce that, for large enough N,

$$\begin{aligned}\mathrm{P}[Y_N > s] &\le CN^{d+n}\exp\biggl(-\frac{s\ln(N)^2}{\delta_1+\delta_2 s\ln(N)/\sqrt{Nh^{d+n}}}\biggr)\\ &\le C\exp((d+n)\ln(N)-s(\ln N)^{3/2})\\ &\le C\mathrm{e}^{-s}.\end{aligned}$$

Substituting this estimate into (4.11) completes the proof.

Since ∇K has bounded variation, the exact same reasoning can apply to the estimators $\hat{\varphi}_\lambda^{-i}$, and we similarly derive

$$\Bigl\|\sup_{1\le i\le N}\|\hat{\varphi_\lambda}^{-i}-\bar{\varphi_\lambda}\|_\infty\Bigr\|_{2r} = O\biggl(\frac{\ln N}{h\sqrt{Nh^{d+n}}}\biggr),\qquad r\in\mathbb{N}^*. \tag{4.12}$$

The estimates of the previous lemma also allow us to control the error due to the truncation of $\hat{\varphi}^{-i}$. Indeed, since the function φ admits δ as a lower bound according to Assumption 3.2, it follows from (4.4) that $\bar{\varphi} > 2\delta/3$ for small enough h, and (4.5) leads to

$$\mathrm{P}\biggl[|\hat{\varphi}^{-1}(\lambda,z)|<\frac{\delta}{3}\biggr]\le\mathrm{P}\biggl[|\hat{\varphi}^{-1}-\bar{\varphi}|(\lambda,z)>\frac{\delta}{3}\biggr]\le 2\exp(-CNh^{d+n}). \tag{4.13}$$

Introducing $\bar{\varphi}^\delta := \mathrm{E}[\hat{\varphi}^{-1,\delta}]$, we derive

$$\|\bar{\varphi}^\delta-\bar{\varphi}\|_\infty\le\frac{\delta}{3}\sup_{C_K\times C_\phi}\mathrm{P}\biggl[|\hat{\varphi}^{-1}|(\lambda,z)<\frac{\delta}{3}\biggr]\le\frac{2\delta}{3}\exp(-CNh^{d+n}), \tag{4.14}$$

and combining (3.1) and (4.4), we deduce that

$$\|\bar{\varphi}^\delta-\varphi\|_\infty = O(h^{p\wedge q}). \tag{4.15}$$

Similarly, applying (4.6), we obtain

$$\begin{aligned}\Bigl\|\sup_{1\le i\le N}\|\hat{\varphi}^{-i,\delta}-\hat{\varphi}^{-i}\|_\infty\Bigr\|_{2r}&\le\delta\,\mathrm{P}\biggl[\sup_{i\le N}\|\hat{\varphi}^{-i}-\bar{\varphi}\|_\infty>\frac{\delta}{3}\biggr]\\ &\le C\delta N^3\exp(-CNh^{d+n}),\qquad r\in\mathbb{N}.\end{aligned} \tag{4.16}$$

Observe also that (4.14) and (4.16) combined with (3.1) allows us to derive

$$\Bigl\|\sup_{1\le i\le N}\|\hat{\varphi}^{-,\delta}-\bar{\varphi}^\delta\|_\infty\Bigr\|_{2r} = O\biggl(\frac{\ln N}{\sqrt{Nh^{d+n}}}\biggr)\qquad\text{for any } r\in\mathbb{N}^*. \tag{4.17}$$

Finally, since (λ, z) vary in a compact subset, Assumptions 3.2 and 3.3 imply that

$$\|\varphi\|_\infty+\|\varphi_\lambda\|_\infty+\Bigl\|\frac{1}{\varphi}\Bigr\|_\infty<\infty. \tag{4.18}$$

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

It then follows from (4.4), (4.15), and the truncation procedure that

$$\|\bar{\varphi}\|_\infty + \|\bar{\varphi}^\delta\|_\infty + \|\bar{\varphi_\lambda}\|_\infty + \left\|\frac{1}{\bar{\varphi}}\right\|_\infty + \left\|\frac{1}{\bar{\varphi}^\delta}\right\|_\infty + \sup_{1\le i\le N}\left\|\frac{1}{\hat{\varphi}^{-i,\delta}}\right\|_\infty < \infty. \tag{4.19}$$

4.2. A suitable decomposition

For any $N \in \mathbb{N}$ and $i \le N$, we define the following functions on $\mathbb{R}^d \times \mathbb{R}^n \times \Omega$:

$$t^1_{i,N} := s, \qquad t^2_{i,N} := \frac{\bar{\varphi_\lambda} - \varphi_\lambda}{\varphi}, \qquad t^3_{i,N} := \frac{(\varphi - \bar{\varphi}^\delta)\varphi_\lambda}{\varphi^2},$$

$$t^4_{i,N} := \frac{(\varphi - \bar{\varphi}^\delta(\bar{\varphi_\lambda}\varphi - \bar{\varphi}^\delta\varphi_\lambda)}{\varphi^2\bar{\varphi}^\delta}, \qquad t^5_{i,N} := \frac{\hat{\varphi_\lambda}^{-i} - \bar{\varphi_\lambda}}{\varphi},$$

$$t^6_{i,N} := \frac{(\bar{\varphi}^\delta - \hat{\varphi}^{-i,\delta})\bar{\varphi_\lambda}}{(\bar{\varphi}^\delta)^2}, \qquad t^7_{i,N} := \frac{(\hat{\varphi_\lambda}^{-i} - \bar{\varphi_\lambda})(\varphi^\delta - \bar{\varphi}^\delta)}{\varphi^\delta\bar{\varphi}^\delta},$$

$$t^8_{i,N} := \frac{(\bar{\varphi}^\delta - \hat{\varphi}^{-i,\delta})(\hat{\varphi_\lambda}^{-i} - \bar{\varphi_\lambda})}{\hat{\varphi}^{-i,\delta}\bar{\varphi}^\delta}, \qquad t^9_{i,N} := \frac{(\bar{\varphi}^\delta - \hat{\varphi}^{-i,\delta})^2\bar{\varphi_\lambda}}{\hat{\varphi}^{-i,\delta}(\bar{\varphi}^\delta)^2},$$

so that

$$\hat{s}_N^{-i}(\Lambda_i, Z_i) = \sum_{j=1}^{9} t^j_{i,N}(\Lambda_i, Z_i).$$

This implies the following decomposition of the estimator $\tilde{\beta}_N$:

$$\tilde{\beta}_N = \sum_{j=1}^{9} T_N^j, \quad \text{where} \quad T_N^j := \frac{1}{\ell(0)Nh^d}\sum_{i=1}^{N}\phi(Z_i)t^j_{i,N}(\Lambda_i, Z_i)K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \tag{4.20}$$

for every $j = 1, \dots, 9$. By (4.18) and (4.19), we observe that

$$\|t^j_{i,N}\|_\infty < \infty \quad \text{for all } j = 1, \dots, 4.$$

Lemma 4.2. *For any $j = 1, \dots, 4$, we have* $\mathrm{E}[T_N^j] = O(\|t^j_{1,N}\|_\infty)$.

Proof. The result is derived from the following inequality:

$$\begin{aligned}|\mathrm{E}[T_N^j]| &\le \frac{1}{\ell(0)h^d}\left|\mathrm{E}\left[\phi(Z_1)t^j_{1,N}(\Lambda_1, Z_1)K\left(\frac{\lambda^0 - \Lambda_1}{h}\right)\right]\right| \\ &\le \frac{1}{\ell(0)}\left|\int_{\mathbb{R}^d\times\mathbb{R}^d}\phi(z)t^j_{1,N}(\lambda^0 - hl, z)K(l)\,\mathrm{d}l\,\mathrm{d}v\right| \\ &\le C\|t^j_{1,N}\|_\infty.\end{aligned}$$

Lemma 4.3. *For every $j = 1, \dots, 4$,* $\mathrm{var}[T_N^j] = O(N^{-1}h^{-d}\|t^j_{1,N}\|^2_\infty)$.

Proof. For any $j = 1, \dots, 4$, the N random variables $T_N^j(\Lambda_i, Z_i)$ are independent and

$$\begin{aligned}\mathrm{var}[T_N^j] &= \frac{1}{\ell(0)^2Nh^{2d}}\,\mathrm{var}\left[\phi(Z_1)t^j_{1,N}(\Lambda_1, Z_1)K\left(\frac{\lambda^0 - \Lambda_1}{h}\right)\right] \\ &\le \frac{1}{\ell(0)^2Nh^{2d}}\,\mathrm{E}\left[\phi^2(Z_1)t^j_{1,N}(\Lambda_1, Z_1)^2K^2\left(\frac{\lambda^0 - \Lambda_1}{h}\right)\right] \\ &\le \frac{\|t^j_{1,N}\|^2_\infty}{\ell(0)^2Nh^d}\int_{\mathbb{R}^d\times\mathbb{R}^d}\phi^2(z)K^2(l)\,\mathrm{d}l\,\mathrm{d}v.\end{aligned}$$

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

The analysis of T_N^j for $j > 4$ requires more effort because of the dependence between the random variables $t_{i,N}^j(\Lambda_i, Z_i)$.

Lemma 4.4. *We have* $\mathrm{E}[T_N^5] = 0$ *and* $\mathrm{var}[T_N^5] \sim \tilde{\Sigma}/(Nh^{d+2})$, *where* $\tilde{\Sigma}$ *is defined in Proposition 3.1.*

Proof. We introduce, for any $i = 1, \dots, N$ and $j = 1, \dots, N$,

$$\mathcal{T}_{ij} := \frac{\phi(Z_i)}{\varphi(\Lambda_i, Z_i)} K\left(\frac{\lambda^0 - \Lambda_i}{h}\right)\left\{\nabla_\lambda K\left(\frac{\Lambda_i - \Lambda_j}{h}\right) H\left(\frac{Z_i - Z_j}{h}\right) - h^{d+n+1}\bar{\varphi_\lambda}(\Lambda_i, Z_i)\right\},$$

so that T_N^5 can be rewritten as

$$T_N^5 = \frac{h^{-2d-n-1}}{\ell(0)N(N-1)} \sum_{i<j} (\mathcal{T}_{ij} + \mathcal{T}_{ji}).$$

By definition, for any $i = 1, \dots, N$ and $j = 1, \dots, N$ with $i \neq j$, we have

$$\bar{\varphi_\lambda}(\Lambda_i, Z_i) = \frac{1}{h^{d+n+1}} \mathrm{E}\left[\nabla_\lambda K\left(\frac{\Lambda_i - \Lambda_j}{h}\right) H\left(\frac{Z_i - Z_j}{h}\right) \,\middle|\, \Lambda_i, Z_i\right].$$

Therefore, $\mathrm{E}[\mathcal{T}_{ij}] = 0$ whenever $i \neq j$, leading to $\mathrm{E}[T_N^5] = 0$.

Since the $\mathcal{T}_{ij}$ are not independent, the computation of the variance requires decomposing T_N^5:

$$T_N^5 = T_N^{5,1} + T_N^{5,2}, \tag{4.21}$$

where

$$T_N^{5,1} := \frac{h^{-2d-n-1}}{\ell(0)N(N-1)} \sum_{i<j} (\mathcal{T}_{ij} + \mathcal{T}_{ji} - b(\Lambda_i, Z_i) - b(\Lambda_j, Z_j)),$$

$$T_N^{5,2} := \frac{h^{-2d-n-1}}{\ell(0)N(N-1)} \sum_{i<j} (b(\Lambda_i, Z_i) + b(\Lambda_j, Z_j)),$$

and $b(\lambda, z) := \mathrm{E}[\mathcal{T}_{12} \mid \Lambda_2 = \lambda,\ Z_2 = z]$.

Let us first study the $T_N^{5,1}$ term. Setting $\Upsilon_{ij} := \mathcal{T}_{ij} + \mathcal{T}_{ji} - b(\Lambda_i, Z_i) - b(\Lambda_j, Z_j)$, we derive the key property

$$\mathrm{E}[\Upsilon_{ij} \mid \Lambda_i, Z_i] = \mathrm{E}[\Upsilon_{ij} \mid \Lambda_j, Z_j] = 0. \tag{4.22}$$

Therefore, $T_N^{5,1}$ has zero mean and we derive

$$\mathrm{var}[T_N^{5,1}] = \frac{h^{-4d-2n-2}}{\ell(0)^2 N^2 (N-1)^2} \sum_{i<j} \mathrm{E}[\Upsilon_{ij}\Upsilon_{ij}^\top] = \frac{h^{-4d-2n-2}}{2\ell(0)^2 N(N-1)} \mathrm{E}[\Upsilon_{12}\Upsilon_{12}^\top].$$

By (4.22), we compute

$$\mathrm{E}[\Upsilon_{12}\Upsilon_{12}^\top] = 2\,\mathrm{E}[\mathcal{T}_{12}\mathcal{T}_{12}^\top] + 2\,\mathrm{E}[\mathcal{T}_{12}\mathcal{T}_{21}^\top] - 2\,\mathrm{E}[b^2(\Lambda_1, Z_1)].$$

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

Next we estimate that $|\mathrm{E}[\mathcal{T}_{12}\mathcal{T}_{12}^{\top}]|$ is dominated by

$$\mathrm{E}\bigg[\frac{\phi^2(Z_1)}{\varphi^2(\Lambda_1, Z_1)}K^2\bigg(\frac{\lambda^0-\Lambda_1}{h}\bigg)|\nabla_\lambda K|^2\bigg(\frac{\Lambda_1-\Lambda_2}{h}\bigg)H^2\bigg(\frac{Z_1-Z_2}{h}\bigg)\bigg]$$
$$+\, h^{2d+n}\int_{(\mathbb{R}^d)^4}\phi^2(z)K^2(l_1)|\nabla_\lambda K|^2(l_2)H^2(v)\frac{\varphi(\lambda^0-hl_1-hl_2, z-hv)}{\varphi(\lambda^0-hl_1, z)}\,\mathrm{d}l_1\,\mathrm{d}l_2\,\mathrm{d}z\,\mathrm{d}v,$$

by the usual change of variables. Clearly, the first term on the right-hand side is of order $O(h^{2d+n})$, while the second term is of order $O(h^{3d+2n+2})$ by (4.19). Similarly, we have $\mathrm{E}[\mathcal{T}_{12}\mathcal{T}_{21}^{\top}] = O(h^{2d+n})$. Moreover, $\mathrm{E}[b^2(\Lambda_1, Z_1)] = O(N^{-2}h^{-d-2})$. We deduce that

$$\operatorname{var}[T_N^{5,1}] = O\bigg(\frac{1}{N^2h^{2d+n+2}}\bigg) = o\bigg(\frac{1}{Nh^{2+d}}\bigg), \tag{4.23}$$

using the relations between N and h given by (3.1).

Next we rewrite $T_N^{5,2}$ as

$$T_N^{5,2} = \frac{h^{-2d-n-1}}{\ell(0)N}\sum_i b(\Lambda_i, Z_i).$$

By the usual change of variables,

$$b(\lambda, z) = h^{d+n}\int_{\mathbb{R}^d\times\mathbb{R}^d}\phi(z+hv)K\bigg(\frac{\lambda^0-\lambda}{h}-l\bigg)\nabla K(l)H(v)\,\mathrm{d}l\,\mathrm{d}v$$
$$-\, h^{n+1}\int_{\mathbb{R}^d}\phi(z)\bar{\varphi_\lambda}(\lambda^0-hl, z)K(l)\,\mathrm{d}l.$$

By direct calculation, it is easily checked that the second term is negligible. Then, by the usual change of variables, it follows that

$$\mathrm{E}[b(\Lambda_i, Z_i)b(\Lambda_i, Z_i)^{\top}]$$
$$\sim h^{3d+2n}\int_{\mathbb{R}^d\times\mathbb{R}^d}\bigg(\int_{\mathbb{R}^d\times\mathbb{R}^d}\phi(z+hv)K(l_2-l_1)\nabla K(l_1)H(v)\,\mathrm{d}l_1\,\mathrm{d}v\bigg)^{\otimes}$$
$$\times\, \varphi(\lambda^0-hl_2, z)\,\mathrm{d}l_2\,\mathrm{d}z.$$

By Assumptions 3.2 and 3.3, we deduce from the dominated convergence theorem, together with the fact that $\mathrm{E}[b(\Lambda_i, Z_i)] = 0$, that

$$\operatorname{var}[T_N^{5,2}] \sim \frac{1}{Nh^{d+2}}\int_{\mathbb{R}^d\times\mathbb{R}^d}\phi^2(z)\bigg(\int_{\mathbb{R}^d}K(l_2-l_1)\nabla K(l_1)\,\mathrm{d}l_1\bigg)^{\otimes}\varphi(\lambda^0, z)\,\mathrm{d}l_2\,\mathrm{d}z. \tag{4.24}$$

The proof is completed by collecting estimates (4.23) and (4.24) into (4.21).

Lemma 4.5. *We have* $\mathrm{E}[T_N^6] = o(h^{p\wedge q})$ *and* $\operatorname{var}[T_N^6] = o(N^{-1}h^{-d-2})$.

Proof. We decompose $t_{i,N}^6$ into the sum of

$$t_{i,N}^{6,1} := \frac{(\bar\varphi-\hat\varphi^{-i})\bar{\varphi_\lambda}}{(\bar\varphi^\delta)^2}, \qquad t_{i,N}^{6,2} := \frac{(\hat\varphi^{-i}-\hat\varphi^{-i,\delta})\bar{\varphi_\lambda}}{(\bar\varphi^\delta)^2}, \quad \text{and} \quad t_{i,N}^{6,3} := \frac{(\bar\varphi^\delta-\bar\varphi)\bar{\varphi_\lambda}}{(\bar\varphi^\delta)^2},$$

and we study the corresponding $T_N^{6,1}$, $T_N^{6,2}$, and $T_N^{6,3}$ separately.

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

It can easily be checked that $T_N^{6,1}$ can be dealt with as T_N^5. By the same calculation we obtain $\mathrm{E}[T_N^{6,1}] = 0$ and

$$\operatorname{var}[T_N^{6,1}] \sim \frac{h^{-4d-2n}}{\ell(0)^2 N^2} \sum_i \operatorname{var}[\tilde{b}(\Lambda_i, Z_i)],$$

where $\tilde{b}(\lambda, z)$ is given by

$$\mathrm{E}\left[\frac{\phi(Z_i)\varphi_\lambda(\Lambda_i, Z_i)}{\varphi(\Lambda_i, Z_i)^2} K\left(\frac{\lambda^0 - \Lambda_i}{h}\right)\left\{K\left(\frac{\Lambda_i - \lambda}{h}\right) H\left(\frac{Z_i - z}{h}\right) - h^{d+n}\bar{\varphi}(\Lambda_i, Z_i)\right\}\right].$$

The variables $\tilde{b}(\Lambda_i, Z_i)$ also have zero mean and, as in the proof of Lemma 4.4, the usual change of variables implies that

$$h^{-3d-2n} \operatorname{var}[\tilde{b}(\Lambda_i, Z_i)] \sim \int_{\mathbb{R}^d \times \mathbb{R}^d} [G_6(l_2, z)]^{\otimes} \varphi(\lambda^0 - hl_2, z)\, \mathrm{d}l_2\, \mathrm{d}z,$$

with

$$G_6(l_2, z) := \int_{\mathbb{R}^d \times \mathbb{R}^d} \phi(z + hv) \frac{\varphi_\lambda}{\varphi}(\lambda^0 + hl_1 - hl_2, z + hv) K(l_2 - l_1) K(l_1) H(v)\, \mathrm{d}l_1\, \mathrm{d}v.$$

By the continuity and the uniform boundedness of ϕ and φ_λ/φ implied by Assumptions 3.2 and 3.3, we derive

$$\operatorname{var}[T_n^{6,1}] = O\left(\frac{1}{Nh^d}\right) = o\left(\frac{1}{Nh^{d+2}}\right).$$

We now turn to $T_N^{6,2}$ and compute

$$|T_N^{6,2}| \le C \sup_{i \le N} \|\hat{\varphi}^{-i,\delta} - \hat{\varphi}^{-i}\|_\infty \left(\frac{1}{Nh^d} \sum_{i=1}^N \left|\phi(Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h}\right)\right|\right).$$

Therefore, we deduce from the Cauchy–Schwarz inequality that

$$|\mathrm{E}[T_N^{6,2}]| \le C \left\| \sup_{i \le N} \|\hat{\varphi}^{-i,\delta} - \hat{\varphi}^{-i}\|_\infty \right\|_2 \mathrm{E}\left[\left(\frac{1}{Nh^d} \sum_{i=1}^N \left|\phi(Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h}\right)\right|\right)^2\right]^{1/2},$$

and (3.1) combined with (4.16) leads to $\mathrm{E}[T_N^{6,2}] = o(h^{p \wedge q})$. Similarly, we obtain

$$\operatorname{var}[T_N^{6,2}] \le C \left\| \sup_{i \le N} \|\hat{\varphi}^{-i,\delta} - \hat{\varphi}^{-i}\|_\infty \right\|_4 \mathrm{E}\left[\left(\frac{1}{Nh^d} \sum_{i=1}^N \left|\phi(Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h}\right)\right|\right)^4\right]^{1/4},$$

which leads to $\operatorname{var}[T_n^{6,2}] = o(N^{-1}h^{-d-2})$.

We finally observe that $T_N^{6,3}$ is treated similarly thanks to (4.14).

Lemma 4.6. *We have* $\mathrm{E}[T_N^7] = 0$ *and* $\operatorname{var}[T_N^7] = o(N^{-1}h^{-d-2})$.

Proof. Observe that

$$t_N^7(\lambda, z) = t_N^5(\lambda, z)\psi(\lambda, z), \quad \text{where} \quad \psi := \frac{\varphi - \bar{\varphi}^\delta}{\bar{\varphi}^\delta}.$$

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

Following the lines of the proof of Lemma 4.4, we see that $\mathrm{E}[T_N^7] = 0$, and we estimate

$$Nh^{d+2}\,\mathrm{var}[T_N^7] \sim \int_{\mathbb{R}^d\times\mathbb{R}^d} [G_7(u,z)]^{\otimes}\varphi(\lambda^0 - hu, z)\,\mathrm{d}u\,\mathrm{d}z,$$

with

$$G_7(u,z) := \int \phi(z+hv)\psi(\lambda^0 + hl - hu, z+hv)K(u-l)\nabla K(l)H(v)\,\mathrm{d}l\,\mathrm{d}v.$$

By (4.15) and (4.19), it follows that $\|\psi\|_\infty = O(h^{p\wedge q})$ and, since φ and ϕ are uniformly bounded, we deduce that

$$\mathrm{var}[T_N^7] = O\biggl(\frac{h^{p\wedge q}}{Nh^{d+2}}\biggr) = o\biggl(\frac{1}{Nh^{d+2}}\biggr).$$

Lemma 4.7. *We have*

$$\mathrm{E}[T_N^8] \sim \frac{h^{-d-n-1}}{\ell(0)N}\biggl(\int \phi(z)\,\mathrm{d}z\biggr)\biggl(\int H^2(v)\,\mathrm{d}v\biggr)\int_{\mathbb{R}^d\times\mathbb{R}^d} K(l_1 - l_2)K(l_2)\nabla K(l_2)\,\mathrm{d}l_1\,\mathrm{d}l_2$$

and $\mathrm{var}[T_N^8] = o(N^{-1}h^{-d-2})$.

Proof. We split the proof it two steps.

Step 1. We first estimate $\mathrm{E}[T_N^8]$. We rewrite $t_N^8(\lambda,z)$ as $t_N^{8,1}(\lambda,z) + t_N^{8,2}(\lambda,z) + t_N^{8,3}(\lambda,z)$ with

$$t_{i,N}^{8,1} = \frac{(\bar{\varphi} - \hat{\varphi}^{-i})(\hat{\varphi_\lambda}^{-i} - \bar{\varphi_\lambda})}{\varphi^2},$$

$$t_{i,N}^{8,2} = \frac{(\bar{\varphi}^\delta - \bar{\varphi})(\hat{\varphi_\lambda}^{-i} - \bar{\varphi_\lambda})}{\varphi^2} + \frac{(\hat{\varphi}^{-i} - \hat{\varphi}^{-i,\delta})(\hat{\varphi_\lambda}^{-i} - \bar{\varphi_\lambda})}{\varphi^2},$$

$$t_{i,N}^{8,3} = \frac{(\bar{\varphi}^\delta - \hat{\varphi}^{-i,\delta})^2(\hat{\varphi_\lambda}^{-i} - \bar{\varphi_\lambda})}{\hat{\varphi}^{-i,\delta}(\bar{\varphi}^\delta)^2} + \frac{(\bar{\varphi}^\delta - \hat{\varphi}^{-i,\delta})(\hat{\varphi_\lambda}^{-i} - \bar{\varphi_\lambda})(\varphi^2 - (\bar{\varphi}^\delta)^2)}{\varphi^2(\bar{\varphi}^\delta)^2}.$$

Then $T_N^8 = T_N^{8,1} + T_N^{8,2} + T_N^{8,3}$, where

$$T_N^{8,k} := \frac{1}{\ell(0)Nh^d}\sum_{i=1}^N \phi(Z_i)t_{i,N}^{8,k}(\Lambda_i, Z_i)K\biggl(\frac{\lambda^0 - \Lambda_i}{h}\biggr) \quad \text{for } k = 1,2,3.$$

We now introduce

$$U_{ij} := \nabla_\lambda K\biggl(\frac{\Lambda_i - \Lambda_j}{h}\biggr)H\biggl(\frac{Z_i - Z_j}{h}\biggr) - \mathrm{E}\biggl[\nabla_\lambda K\biggl(\frac{\Lambda_i - \Lambda_j}{h}\biggr)H\biggl(\frac{Z_i - Z_j}{h}\biggr)\biggm| \Lambda_i, Z_i\biggr],$$

$$V_{ij} := K\biggl(\frac{\Lambda_i - \Lambda_j}{h}\biggr)H\biggl(\frac{Z_i - Z_j}{h}\biggr) - \mathrm{E}\biggl[K\biggl(\frac{\Lambda_i - \Lambda_j}{h}\biggr)H\biggl(\frac{Z_i - Z_j}{h}\biggr)\biggm| \Lambda_i, Z_i\biggr],$$

so that

$$\mathrm{E}[U_{ij}V_{ik} \mid \Lambda_i, Z_i] = \mathrm{E}[U_{ij} \mid \Lambda_i, Z_i]\,\mathrm{E}[V_{ik} \mid \Lambda_i, Z_i] = 0 \quad \text{whenever } j \neq k.$$

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

Using this property, we directly compute that

$$\mathrm{E}[t_N^{8,1}(\Lambda_1, Z_1) \mid \Lambda_1, Z_1] = \frac{h^{-2d-2n-1}}{(N-1)^2\varphi^2(\Lambda_1, Z_1)} \mathrm{E}\Bigg[\sum_{j\neq 1}\sum_{k\neq 1} U_{1j}V_{1k} \,\Bigg|\, \Lambda_1, Z_1\Bigg]$$

$$= \frac{h^{-2d-2n-1}}{(N-1)\varphi^2(\Lambda_1, Z_1)} \mathrm{E}[U_{12}V_{12} \mid \Lambda_1, Z_1].$$

Since the expectation of $T_N^{8,1}$ is given by

$$\mathrm{E}[T_N^{8,1}] = \frac{h^{-d}}{\ell(0)} \mathrm{E}\bigg[\phi(Z_1)K\bigg(\frac{\lambda^0 - \Lambda_1}{h}\bigg) \mathrm{E}[t_{1,N}^{8,1}(\Lambda_1, Z_1) \mid \Lambda_1, Z_1]\bigg],$$

we derive, by the usual change of variables,

$$\ell(0)Nh^{d+n+1}\,\mathrm{E}[T_N^{8,1}] \sim \int_{\mathbb{R}^d\times\mathbb{R}^d} G_8(l_2, z)\varphi(\lambda^0 - hl_2, z)\,\mathrm{d}l_2\,\mathrm{d}z,$$

with

$$G_8(l_2, z) := \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{\phi(z+hv)}{\varphi(\lambda^0 + hl_1 - hl_2, z+hv)} K(l_2 - l_1)K(l_1)\nabla K(l_1)H^2(v)\,\mathrm{d}l_1\,\mathrm{d}v.$$

Finally, by the continuity and the uniform boundedness of φ and ϕ, we derive

$$\mathrm{E}[T_N^{8,1}] \sim \frac{h^{-d-n-1}}{\ell(0)N} \int_{(\mathbb{R}^d)^4} \phi(z)K(l_2 - l_1)K(l_1)\nabla K(l_1)H^2(v)\,\mathrm{d}l_1\,\mathrm{d}v\,\mathrm{d}l_2\,\mathrm{d}z. \tag{4.25}$$

Furthermore, by the Cauchy–Schwarz inequality and (3.1), we have

$$|\,\mathrm{E}[T_N^{8,k}]| \le \Big\|\sup_{i\le N}\|t_{i,N}^{8,k}\|_\infty\Big\|_2\, \mathrm{E}\bigg[\bigg(\frac{1}{Nh^d}\sum_{i=1}^N\bigg|\phi(Z_i)K\bigg(\frac{\lambda^0 - \Lambda_i}{h}\bigg)\bigg|\bigg)^2\bigg]^{1/2} \tag{4.26}$$

$$\le C\Big\|\sup_{i\le N}\|t_{i,N}^{8,k}\|_\infty\Big\|_2, \qquad k = 2, 3. \tag{4.27}$$

Finally, combining relations (4.4)–(4.19), the Cauchy–Schwarz inequality, and (3.1), we obtain

$$\Big\|\sup_{i\le N}\|t_{i,N}^{8,2}\|_\infty\Big\|_2 = o\bigg(\frac{1}{Nh^{d+n+1}}\bigg)$$

and

$$\Big\|\sup_{i\le N}\|t_{i,N}^{8,3}\|_\infty\Big\|_2 = O\bigg(\frac{(\ln N)^3}{Nh^{d+n+1}\sqrt{Nh^{d+n}}}\bigg) = o\bigg(\frac{1}{Nh^{d+n+1}}\bigg).$$

Therefore, (4.25) and (4.26) lead to the expected equivalent for $\mathrm{E}[T_N^8]$.

Step 2. We now study the variance of T_N^8. We first note that the Cauchy–Schwarz inequality and (3.1) lead to

$$\mathrm{var}[T_N^8] \le C\Big\|\sup_{i\le N}\|t_{i,N}^8\|_\infty^4\Big\|_4^2.$$

But, again using the Cauchy–Schwarz inequality and relations (3.1), (4.4), (4.19), and (4.17), we deduce that

$$\mathrm{var}[T_N^8] = O\bigg(\frac{\ln^4 N}{N^2h^{2d+2n+2}}\bigg) = o\bigg(\frac{1}{Nh^{d+2}}\bigg).$$

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

Lemma 4.8. *We have* $\mathrm{E}[T_N^9] = O(N^{-1}h^{-d-n})$ *and* $\mathrm{var}[T_N^9] = o(N^{-1}h^{-d-2})$.

Proof. It can be easily checked that T_N^9 can be dealt with as T_N^8 and, following the lines of the proof of Lemma 4.7, we obtain the announced result.

4.3. Asymptotic bias and variance

This section is devoted to the proof of Proposition 3.1 characterizing the asymptotic bias and variance of the double kernel-based estimator $\tilde{\beta}_N$.

Proof of Proposition 3.1. We split the proof in two steps.

Step 1. We first derive the expectation of $\tilde{\beta}_N$. Note that $T_N^1 = \bar{\beta}_N$ as defined in (2.2), which satisfies

$$\mathrm{E}[\bar{\beta}_N] = \frac{1}{\ell(0)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \phi(z) K(l) s(\lambda^0 - hl, z) \varphi(\lambda^0 - hl, z)\, \mathrm{d}t\, \mathrm{d}z.$$

The regularity of the function $s\varphi$ given by Assumption 3.3 enables us to derive

$$\mathrm{E}[T_N^1] - \beta \sim \frac{h^p}{\ell(0)} \int_{\mathbb{R}^d} \xi_K^p[\ell f_\lambda](\lambda^0, z)\phi(z)\, \mathrm{d}z.$$

Using Remark 4.2, we deduce from (4.3) that we have

$$\mathrm{E}[T_N^2] = \frac{h^p}{\ell(0)} \int_{\mathbb{R}^d} \xi_K^p[\varphi_\lambda](\lambda^0, z)\phi(z)\, \mathrm{d}z + \frac{h^q}{\ell(0)} \int_{\mathbb{R}^d} \xi_H^q[\varphi_\lambda](\lambda^0, z)\phi(z)\, \mathrm{d}z + o(h^{p\wedge q}).$$

We now rewrite $t_{i,N}^3$ as the sum of

$$t_{i,N}^{3,1} := \frac{(\varphi - \bar{\varphi})\varphi_\lambda}{\varphi^2} \quad \text{and} \quad t_{i,N}^{3,2} := \frac{(\bar{\varphi}^\delta - \bar{\varphi})\varphi_\lambda}{\varphi^2},$$

and study separately the corresponding $T_N^{3,1}$ and $T_N^{3,2}$. From (4.2) we derive

$$\mathrm{E}[T_N^{3,1}] = -\frac{h^p}{\ell(0)} \int_{\mathbb{R}^d} \frac{\varphi_\lambda \xi_K^p[\varphi]}{\varphi}(\lambda^0, z)\phi(z)\, \mathrm{d}z - \frac{h^q}{\ell(0)} \int_{\mathbb{R}^d} \frac{\varphi_\lambda \xi_H^q[\varphi]}{\varphi}(\lambda^0, z)\phi(z)\, \mathrm{d}z + o(h^{p\wedge q}),$$

and we directly deduce from (3.1) and (4.14) that $\mathrm{E}[T_N^{3,2}] = o(h^{p\wedge q})$.

Note that

$$t_{i,N}^4 = \frac{(\varphi - \bar{\varphi}^\delta)^2 \varphi_\lambda}{\varphi^2 \bar{\varphi}^\delta} + \frac{(\bar{\varphi_\lambda} - \varphi_\lambda)(\varphi - \bar{\varphi}^\delta)}{\varphi \bar{\varphi}^\delta}.$$

Then, using (4.4), (4.15), (4.18), and (4.19), we derive $||t_{i,N}^4||_\infty = o(h^{p\wedge q})$, and Lemma 4.2 leads to $\mathrm{E}[T_N^4] = o(h^{p\wedge q})$.

From Lemmas 4.4, 4.5, and 4.6, we have $\mathrm{E}[T_N^j] = 0$ for $j = 5, \dots, 7$, and Lemma 4.7 gives

$$\mathrm{E}[T_N^8] \sim \frac{h^{-d-n-1}}{\ell(0)N} \int_{(\mathbb{R}^d)^4} \frac{\phi(z)}{\varphi(\lambda^0, z)} K(l_2 - l_1) K(l_1) \nabla K(l_1) H^2(v)\, \mathrm{d}l_1\, \mathrm{d}v\, \mathrm{d}l_2\, \mathrm{d}z.$$

Finally, Lemma 4.8 tells us that $\mathrm{E}[T_N^9] = o(N^{-1}h^{-d-n-1})$.

We then obtain $\mathrm{E}[\tilde{\beta}_N]$ by summing up the $\mathrm{E}[T_N^j]$ for $j = 1, \dots, 9$.

Step 2. We now analyze the variance of $\tilde{\beta}_N$. For any $j = 1, \dots, 4$, expressions (4.4), (4.15), (4.18), and (4.19) imply that $||t_N^j||_\infty = O(1)$. Then Lemma 4.3 leads to

$$\mathrm{var}[T_N^j] = o(N^{-1}h^{-d-2}) \quad \text{for every } j = 1, \dots, 4.$$

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

From Lemma 4.4 we obtain

$$\operatorname{var}[T_N^5] \sim \frac{1}{\ell(0)Nh^{d+2}} \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi^2(z)\bigg\{\int_{\mathbb{R}^d} K(l_2-l_1)\nabla K(l_1)\,\mathrm{d}l_1\bigg\}^{\otimes} f(\lambda^0,z)\,\mathrm{d}z\,\mathrm{d}l_2. \quad (4.28)$$

Indeed, Lemmas 4.5 to 4.8 also imply that

$$\operatorname{var}[T_N^j] = o(N^{-1}h^{-d-2}) \quad \text{for every } j = 5,\dots,9.$$

Hence, $\operatorname{cov}(T_N^j, T_N^k) = o(N^{-1}h^{-d-2})$ unless $j = k = 5$ and $\operatorname{var}[\tilde{\beta}_N]$ is given by expression (4.28).

4.4. Central limit theorem

This section is devoted to the proof of Theorem 3.1, which provides a central limit theorem for the double kernel-based estimator $\tilde{\beta}_N$.

Proof of Proposition 3.1. As in the proof of Proposition 3.1, the variance of $\tilde{\beta}_N$ is given by the variance of

$$T_N^{5,2} = \frac{h^{-2d-n-1}}{\ell(0)N} \sum_i b(\Lambda_i, Z_i),$$

where

$$b(\lambda,z) := h^{d+n} \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi(z+hv)K\bigg(\frac{\lambda^0-\lambda}{h} - l\bigg)\nabla K(l)H(v)\,\mathrm{d}l\,\mathrm{d}v$$
$$- h^{n+1}\int_{\mathbb{R}^d} \phi(z)\bar{\varphi}_\lambda(\lambda^0-hl,z)K(l)\,\mathrm{d}l.$$

As in the proofs of Theorems 4.1 or 4.2 of [5], using Kolmogorov's condition with the fourth moment of b and the Cramer–Wold device, we find that $T_N^{5,2}$ is asymptotically normal. We then finally deduce that

$$\sqrt{Nh^{d+2}}(\tilde{\beta}_N - \mathrm{E}[\tilde{\beta}_N]) \xrightarrow{\text{LAW}} \mathcal{N}(0,\tilde{\Sigma}) \quad \text{as } N\to\infty.$$

Under the additional condition $Nh^{d+2+2(p\wedge q)} \to 0$, we conclude the proof, denoting that the bias vanishes in the previous expression.

References

[1] Ait-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica* **64,** 527–560.
[2] Azencott, R. (1982). Densité des diffusions en temps petit: développements asymptotiques. I. In *Séminaire de Probabilités XVIII* (Lecture Notes Math. **1059**), Springer, Berlin, pp. 402–498.
[3] Broadie, M. and Glasserman, P. (1996). Estimating security prices using simulation. *Manag. Sci.* **42,** 269–285.
[4] Detemple, J., Garcia, R. and Rindisbacher, M. (2005). Asymptotic properties of Monte Carlo estimators of derivatives. *Manag. Sci.* **51,** 1657–1675.
[5] Elie, R., Fermanian, J. D. and Touzi, N. (2007). Kernel estimation of Greek weights by parameter randomization. *Ann. Appl. Prob.* **17,** 1399–1423.
[6] Fournié *et al.* (1999). Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance Stoch.* **3,** 391–412.
[7] Giles, M. and Glasserman, P. (2006). Smoking adjoints: fast Monte Carlo Greeks. *Risk* **19,** 92–96.
[8] Gobet, E. (2001). Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli* **7,** 899–912.
[9] Kohatsu-Higa, A. and Montero, M. (2004). Malliavin calculus in finance. In *Handbook of Computational and Numerical Methods in Finance*, Birkhäuser, Boston, MA, pp. 111–174.

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press

[10] L'Ecuyer, P. and Perron, G. (1994). On the convergence rates of IPA and FDC derivative estimators. *Operat. Res.* **42,** 643–656.

[11] Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stoch. Proc. Appl.* **65,** 69–80.

[12] Milstein, G. and Tretyakov, M. (2005). Numerical analysis of Monte Carlo evaluation of Greeks by finite differences. *J. Comput. Finance* **8,** 1–34.

https://doi.org/10.1239/jap/1253279852 Published online by Cambridge University Press