INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2004-26

Optimal Greek Weights by Kernel Estimation

R. ELIE[1]
J.-D. FERMANIAN[2]
N. TOUZI[3]

Les documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.

Working papers do not reflect the position of INSEE but only the views of the authors.

[1] ENSAE & CREST. Mail : elie@ensae.fr
[2] Ixis CIB & CREST. Mail : fermania@ensae.fr
[3] CREST & CEREMADE. Mail : touzi@ensae.fr

Optimal Greek weights by Kernel estimation

R. Elie
ENSAE & CREST
elie@ensae.fr

J.-D. Fermanian
Ixis CIB & CREST
fermania@ensae.fr

N. Touzi
CREST & CEREMADE
touzi@ensae.fr

Abstract

A Greek weight associated to a parameterized random variable $Z(\lambda)$ is a random variable π such that $\nabla_\lambda E[\phi(Z(\lambda))] = E[\phi(Z(\lambda))\pi]$. The importance of the set of Greek weights for the purpose of Monte Carlo simulations has been highlighted in the recent literature. Our main concern in this paper is to device methods which produce the optimal weight in a general context where the density of $Z(\lambda)$ is not explicitly known. Given an a priori distribution on the parameter λ, the first natural idea is to introduce an estimator of the score function by the classical Kernel estimation technique of the conditional density. By an integration by parts argument on the limit of this first Kernel estimator, we define an alternative simpler Kernel-based estimator which turns out to be closely related to the partial gradient of the Kernel-based estimator of $\mathbb{E}[\phi(Z(\lambda))]$. We provide an asymptotic analysis of the mean squared error of these estimators, as well as their asymptotic distributions and compare them to those of more classical estimators: the finite differences one. We next investigate further the short maturity properties of these estimators. We finally present numerical results obtained with these estimators.

Key words: Greek weights, Monte Carlo simulation, Non-parametric regression.

MSC 1991 subject classifications: Primary 11K45, 62G08; secondary 60G07.

Résumé

Un poids associé à une variable aléatoire $Z(\lambda)$ est une variable aléatoire π telle que $\nabla_\lambda E[\phi(Z(\lambda))] = E[\phi(Z(\lambda))\pi]$. Ces poids sont utiles pour simuler par Monte Carlo les sensibilité de prix d'options et de manière plus générale des dérivées d'espérance. Nous proposons une méthode permettant d'estimer le poids optimal dans le cas où la densité de $Z(\lambda)$ n'est pas explicite. Etant donnée une distribution a-priori sur la variable λ, il est naturel d'utiliser un estimateur non paramétrique de la fonction score. A l'aide d'intégration par partie sur la limite de cet estimateur, nous proposons un nouvel estimateur fortement lié à la dérivée de l'estimateur non paramétrique de $\mathbb{E}[\phi(Z(\lambda))]$. Les propriétés asymptotiques de ces estimateurs sont comparées à celles d'estimateurs plus classiques, puis nous présentons des résultats numériques.

1 Introduction

Let λ be some given parameter in $\mathbb{R}^d$, and define the function

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

where $Z(.)$ is a parameterized random variable with values in $\mathbb{R}^n$ and $\phi : \mathbb{R}^n \to \mathbb{R}$ is a measurable function. In many applications, we are interested in the numerical computation of the function $V^\phi(\lambda)$ for some parameter λ^0, together with the sensitivities of V^ϕ with respect to the parameter λ.

In particular, in the financial literature, V^ϕ represents 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 non-risky asset (so that the model is reduced to the zero-interest rate situation). The sensitivities of V^ϕ with respect to the parameter λ are called *Greeks*, and are widely used by the practitioners in their *profit and loss* analysis. In the context of the Black-Scholes model, the derivative of the option price with respect to the current underlying asset price is the so-called *Delta*, and represents the number of shares of risky asset to be held at each time in order to realize a dynamic perfect hedge of the option. The *Gamma* is the second derivative of the option price, with respect to the underlying asset price. It is an indicator of the variation of the hedging portfolio. Another important *Greek* is the so-called *Vega* (although not a Greek letter !) which is the derivative of the option price with respect to the *volatility* coefficient (see e.g. Hull (2002), for more details).

Given a numerical scheme for the computation of the function V^ϕ, the first natural idea for the numerical computation of the *Greeks* is the finite differences approximation of the corresponding derivative. In addition to the generic standard error on the numerical computation of the expectation, this approximation leads to a biased estimator at a finite distance and appears to be inefficient for discontinuous payoff functions ϕ. We refer to L'Ecuyer and Perron (1994) and Detemple, Garcia and Rindisbacher (2003) for a theoretical analysis of the rate of convergence of this estimator. Two direct methods for computing the *Greeks* have been presented by Broadie and Glasserman (1996) : (i) the pathwise method, which consists in differentiating the random variable $\phi(Z(\lambda))$ inside the expectation operator, and (ii) the likelihood ratio method which reports the differentiation on the distribution of $Z(\lambda)$. The first method requires the computation of the gradient of the payoff function ϕ, which is a serious limitation in practice as ϕ is typically highly complicated or even not differentiable. As for the second method (ii), it was (apparently) restricted to the very special cases where the distribution of $Z(\lambda)$ is known explicitly. This difficulty was overcome by Fournié, Lasry, Lebuchoux, Lions and Touzi (1999) who showed that, for *smooth* random variables $Z(.)$,

$$\nabla_\lambda \mathbb{E}[\phi(Z(\lambda))] \;=\; \mathbb{E}[\phi(Z(\lambda))\pi], \tag{1.1}$$

where π, the so-called *Greek weight*, is a random variable independent of the pay-off function ϕ. The key technique in order to obtain such a representation is the use of the integration-by-parts formula from Malliavin calculus. Further developments of this result were obtained by Gobet and Kohatsu-Higa (2003). Section 2 reports a quick overview of the notion of Greek weights together with the connection with the Malliavin calculus. The comparison of the above different methods is available in the survey paper of Kohatsu-Higa and Montero (2004). Notice that other representations of the Greeks in term of expectations were also derived by Gobet (2004) by means of martingale property arguments.

An important observation is that the set of Greek weights which satisfy (1.1) is a non-finite convex set of random variables. By an easy variance reduction argument, it is easily seen that the *score* $\pi^* := \nabla_\lambda \ln f\left(\lambda^0, Z(\lambda^0)\right)$ minimizes $\mathbb{V}\mathrm{ar}\left[\phi(Z(\lambda))\pi\right]$, whenever the density $f(\lambda, z)$ of the random variable $Z(\lambda)$ exists. In general, the use of the Malliavin calculus does not lead to this optimal Greek weight, except in trivial cases where the density $f(\lambda, z)$ is explicitly known, as covered by Broadie and Glasserman (1996).

The main purpose of this paper is to focus on the use of the optimal Greek weight in order to estimate the corresponding Greek by the Monte Carlo method. To do this, we assume that our observations consist of a family $\{(\Lambda_i, Z_i),\ N \geq i \geq 1\}$ of independent pairs (Λ_i, Z_i) drawn in the density $\ell(\lambda) f(\lambda, z)$, where ℓ is some given a-priori distribution on the parameter λ. We define various Kernel estimators of the Greek

$$\nabla_\lambda \mathbb{E}[\phi(Z(\lambda))]_{|\lambda=\lambda_0} \quad = \quad \mathbb{E}[\phi\left(Z(\lambda_0)\right) s\left(\lambda_0, Z(\lambda_0)\right)]\,, \tag{1.2}$$

where $s(\lambda, z) := \nabla_\lambda \ln f\left(\lambda, z\right)$ is the score function. The first natural idea is to notice that

$$\mathbb{E}[\phi\left(Z(\lambda_0)\right) s\left(\lambda_0, Z(\lambda_0)\right)] \quad = \quad \mathbb{E}[\phi\left(Z(\Lambda)\right) s\left(\Lambda, Z(\Lambda)\right) \mid \Lambda = \lambda_0]\,, \tag{1.3}$$

which is a usual regression function. Thus, a two-steps estimation method is proposed : we first perform a Kernel-based estimator $\hat{s}$ of the score function, and then we define a Kernel regression estimator of the Greek by substituting $\hat{s}$ to s. In the sequel, the resulting estimator is referred to as the double Kernel-based estimator and is denoted by $\tilde{\beta}$.

Our next Kernel estimator of the Greek is based on a convenient integration-by-parts in (1.2). This leads to a much simpler estimator $\hat{\beta}$ which turns out to be closely related to the estimator $\acute{\beta}$, obtained by direct differentiation of the classical Kernel regression estimator of $V^\phi(\lambda) = \mathbb{E}[\phi\left(Z(\Lambda)\right) \mid \Lambda = \lambda_0]$. These two estimators will be referred to as the single Kernel-based estimators.

These three estimators are defined precisely in Section 3, and their asymptotic properties are discussed in Section 4. We show that $\hat{\beta}$ and $\acute{\beta}$ are asymptotically equivalent. The asymptotic properties of $\tilde{\beta}$ are derived under stronger conditions on the pay-off function ϕ and the Kernels. The simultaneous choice of the bandwidth, and the number of observations is also more restrictive in the latter case. We study further the particular case where

the a priori density ℓ is uniform and derive its optimal bandwidth. The improved rate of convergence of our single Kernel based estimators is then compared to the finite differences one. When the payoff function ϕ is not regular, our estimator converges asymptotically faster than the finite differences one as soon as the order of the Kernel function is large enough.

Finally, Section 5 compares the short time performance of the single-Kernel estimator $\hat{\beta}$ with the Malliavin-based estimator, whose Greek weight is well-known to suffer from a singularity for short maturity problems. We shall derive the asymptotic properties of $\hat{\beta}$ in the situation where the bandwidth of the Kernel and the maturity shrink to zero, and the number of observations goes to infinity. This allows to fix the (theoretical) relative orders for these three parameters in order to obtain the optimal rate of convergence.

Some numerical results are reported in Section 6. We estimate the delta of an European and an Asian digital call option. Our experiments show that the Malliavin-based estimators, when they exist, are the most efficient. The single-Kernel based estimators still outperform the finite differences one when the number of simulations is sufficiently large.

2 The Greek weights set

Throughout this paper we consider a complete probability space $(\Omega, \mathcal{F}, P)$ supporting a Brownian motion W with values in $\mathbb{R}^m$. We assume that $\mathcal{F}$ is the $P-$completion of the $\sigma-$algebra generated by W.

Let $Z(\lambda)$ be some random variable, valued in $\mathbb{R}^n$, depending on some finite dimensional parameter $\lambda \in \mathbb{R}^d$, and set

$$V^{\phi}(\lambda) \;:=\; E\left[\phi\left(Z(\lambda)\right)\right] \quad \text{for} \quad \phi \in \mathbb{L}^{\infty}(\mathbb{R}^n, \mathbb{R})\,.$$

In order to simplify the presentation, we shall focus our attention on some fixed particular value λ^0 of λ, and we denote

$$Z^0 \;:=\; Z(\lambda^0)\,.$$

The chief goal of this paper is to device efficient methods for the computation of the sensitivity parameter

$$\beta^0 \;:=\; \nabla_\lambda V^{\phi}(\lambda^0),$$

for arbitrary functions ϕ chosen from a suitable large class. In this section, we provide a quick review of Malliavin-based approach introduced in Fournié et al. (1999, 2000).

2.1 Definition

We assume that the distribution of $Z(\lambda)$ is absolutely continuous with respect to the Lebesgue measure, and we denote by $f(\lambda, z)$ the associated density, i.e.

$$\mathbb{E}[\phi(Z(\lambda))] \;=\; \int \phi(z) f(\lambda, z) dz \quad \text{for all} \quad \phi \in \mathbb{L}^\infty(\mathbb{R}^n, \mathbb{R}) .$$

Under mild smoothness assumptions on the density f, we directly compute that

$$\nabla V^\phi(\lambda^0) \;:=\; \frac{\partial V^\phi}{\partial \lambda}(\lambda^0) \;=\; E\left[\phi(Z^0) S^0\right] , \qquad S^0 \;:=\; s(\lambda^0, Z^0) ,$$

where the function s is independent of ϕ and is explicitly given by

$$s(\lambda, z) \;:=\; \nabla_\lambda \left\{ \ln f(\lambda, z) \right\} .$$

Hence the set

$$\mathcal{W} \;:=\; \left\{ \pi \in \mathbb{L}^1(\Omega, \mathbb{R}^d) \;:\; \nabla V^\phi(\lambda^0) = E\left[\phi(Z^0)\pi\right] \text{ for all } \phi \in \mathbb{L}^\infty(\mathbb{R}^n, \mathbb{R}) \right\}$$

is not empty. From the arbitrariness of $\phi \in \mathbb{L}^\infty(\mathbb{R}^n, \mathbb{R})$, it is immediately seen that

$$\mathcal{W} \;=\; \left\{ \pi \in \mathbb{L}^1(\Omega, \mathbb{R}^d) \;:\; E[\pi | Z^0] = S^0 \right\} .$$

We finally observe that

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\phi(Z^0)\pi\right] &= E\left[\phi(Z^0)^2 \pi\pi'\right] - E\left[\phi(Z^0)\pi\right] E\left[\phi(Z^0)\pi\right]' \\
&= E\left[\phi(Z^0)^2 \pi\pi'\right] - \nabla V^\phi(\lambda^0) \nabla V^\phi(\lambda^0)' \\
&= E\left[\phi(Z^0)^2 E[\pi\pi' | Z^0]\right] - \nabla V^\phi(\lambda^0) \nabla V^\phi(\lambda^0)' \\
&\geq E\left[\phi(Z^0)^2 E[\pi | Z^0] E[\pi | Z^0]'\right] - \nabla V^\phi(\lambda^0) \nabla V^\phi(\lambda^0)' \\
&= E\left[\phi(Z^0)^2 S^0 S^{0\prime}\right] - \nabla V^\phi(\lambda^0) \nabla V^\phi(\lambda^0)' \;=\; \mathbb{V}\mathrm{ar}\left[\phi(Z^0) S^0\right]
\end{aligned}$$

so that

$$S^0 \in \mathcal{W} \quad \text{is a minimizer of} \quad \mathbb{V}\mathrm{ar}\left[\phi(Z^0)\pi\right] \;,\; \pi \in \mathcal{W} .$$

Throughout this paper, we call S^0 the optimal Greek weight. Our main objective is to define Monte Carlo estimators of the Greek value β^0, which asymptotically achieve the minimum variance.

2.2 Malliavin Greek weights

We first recall the definition of the Malliavin gradient operator. Let

$$\mathcal{S} := \left\{ F = f\left(\int_{\mathbb{R}_+} h_t^1 \cdot dW_t, \ldots, \int_{\mathbb{R}_+} h_t^n \cdot dW_t \right) ,\; n \in \mathbb{N},\; f \in \mathbb{C}_p^\infty(\mathbb{R}^n),\; h^i \in L^2\left(\mathbb{R}_+, \mathbb{R}^m\right) \right\} ,$$

where $\mathbf{C}_p^\infty(\mathbb{R}^n)$ is the set of all infinitely continuously differentiable functions $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that f and all of its partial derivatives have polynomial growth. The Malliavin derivative of any random variable F in $\mathcal{S}$ is defined by :

$$D_t F \;:=\; \sum_{i=1}^n \nabla_{x_i} f\left(\int_{\mathbb{R}_+} h_t^1 \cdot dW_t, \cdots, \int_{\mathbb{R}_+} h_t^n \cdot dW_t\right) h_t^i .$$

This operator is then extended to $\mathbb{L}^2(\Omega, \mathbb{R}^d)$, by taking the closure $\mathcal{S}$ with respect to the semi norm $\|F\| := \left(\mathbb{E}|F|^2 + \mathbb{E}\int_{\mathbb{R}_+} |D_t F|^2 dt\right)^{1/2}$ (see e.g. Nualart (1995)). This produces the domain $\mathbb{D}^{1,2}$ of the Malliavin operator D, as a dense subset of $\mathbb{L}^2(\Omega, \mathbb{R}^d)$. The Malliavin derivative of functions valued in $\mathbb{R}^d$ is defined componentwise.

We next show how the operator D allows to derive Greek weights in $\mathcal{W}$, without appealing to the explicit knowledge of the density $f(\lambda, z)$. Observe that, for every $\pi \in \mathcal{W}$:

$$\mathbb{E}[\pi] \;=\; \mathbb{E}[S_0] \;=\; \frac{\partial}{\partial \lambda}\int f(\lambda, t) dt \;=\; 0 .$$

If, in addition, $\pi \in \mathbb{L}^2(\Omega, \mathbb{R}^d)$, then it follows from the representation theorem that

$$\pi \;=\; \int_0^\infty u_s \, dW_s$$

for some $u \in \mathbb{L}_{\mathbb{a}}^2(\mathbb{R}_+ \times \Omega, \mathbb{M}_{\mathbb{R}}(d, m))$ with $\mathbb{E}\left[\int_0^\infty |u_s|^2 ds\right] < \infty$. Here, $\mathbb{M}_{\mathbb{R}}(d, m)$ is the collection of all real matrices with d rows and m columns, and $\mathbb{L}_{\mathbb{a}}^2(\mathbb{R}_+ \times \Omega, \mathbb{M}_{\mathbb{R}}(d, m))$ is the set of all adapted processes with values in $\mathbb{M}_{\mathbb{R}}(d, m)$.

Assume that

$$Z_0 \;\in\; \mathbb{D}^{1,2} , \tag{2.1}$$

and let ϕ a $\mathbf{C}_b^1(\mathbb{R}^n, \mathbb{R})$ function and $\pi \in \mathcal{W}$. Then, it follows from the Malliavin integration by parts formula that

$$\begin{aligned} \nabla V^\phi(\lambda^0) \;=\; \mathbb{E}\left[\phi(Z^0)\int_0^\infty u_s \, dW_s\right] \;&=\; \mathbb{E}\left[\int_0^\infty u_s D_s \phi(Z^0) \, ds\right] \\ &=\; \mathbb{E}\left[\int_0^\infty u_s (D_s Z^0)' ds \, \nabla\phi(Z^0)\right] , \end{aligned} \tag{2.2}$$

where $\left(D_s Z^0\right)_{ij} = \left(D_s Z_i^0\right)_j$ and Z_i^0 is the i-th entry of Z^0, $i = 1, \ldots, n$, $j = 1, \ldots, m$. On the other hand,

$$\nabla V^\phi(\lambda_0) \;=\; \mathbb{E}\left[\nabla Z^0 \nabla\phi(Z^0)\right] \quad \text{where} \quad \nabla Z^0 \;:=\; \frac{\partial Z'}{\partial \lambda}(\lambda^0) . \tag{2.3}$$

By arbitrariness of $\phi \in \mathbf{C}_b^1(\mathbb{R}^n, \mathbb{R})$, we deduce from (2.2) and (2.3) that

$$\mathbb{E}\left[\int_0^\infty u_s (D_s Z^0)' ds \,\middle|\, Z^0\right] \;=\; \mathbb{E}\left[\nabla Z^0 \,\middle|\, Z^0\right] . \tag{2.4}$$

Conversely, Let u be an arbitrary process in $\mathbb{L}^2(\mathbb{R}_+ \times \Omega, \mathbb{M}_{\mathbb{R}}(d,m))$ with (2.4). Observe that u does not need to be adapted. Then $\pi := \int_0^\infty u_s dW_s$ satisfies $\nabla V^\phi(\lambda_0) = \mathbb{E}[\phi(Z_0)\pi]$ for every $\phi \in \mathbf{C}_b^1(\mathbb{R}^n, \mathbb{R})$. By a density argument, this property is easily seen to hold for every $\phi \in L^\infty(\mathbb{R}^n, \mathbb{R})$. Hence $\pi \in \mathcal{W} \cap \mathbb{L}^2(\Omega, \mathbb{R}^d)$. We have then proved the following result :

Proposition 2.1 *Assume that $Z_0 \in \mathbb{D}^{1,2}$. Then*

$$\mathcal{W} \cap \mathbb{L}^2(\Omega, \mathbb{R}^d) \;=\; \left\{ \int_0^\infty u_s \, dW_s \;:\; u \in \mathbb{L}^2(\mathbb{R}_+ \times \Omega, \mathbb{M}_{\mathbb{R}}(d,m)) \;\; \textit{and (2.4) holds} \right\}.$$

This result allows to obtain a family of Greek weights without any knowledge of the density distribution of the random variable Z^0. However there is no guarantee for the weight defined by some process $u \in \mathbb{L}^2(\mathbb{R}_+ \times \Omega, \mathbb{M}_{\mathbb{R}}(d,m))$ satisfying (2.4) to produce the optimal Greek weight: see the last two examples of the subsequent Subsection 2.3.

The chief goal of this paper is to introduce Kernel-based estimators which focus on the optimal weight S^0. Of course, our estimators do not have the parametric rate of convergence, but we believe that this critic does not exclude our estimators in finite samples. The main advantage of our estimators remains their simplicity of computation in comparison to Malliavin-based estimators.

Note that the Malliavin Greek weights also lead to estimators of the Greeks which do not have the parametric rate of convergence. Indeed, except the trivial gaussian case, the Malliavin weight is a stochastic integral which needs to be approximated on some given time grid. This leads to a loss of the parametric rate.

2.3 Examples of Malliavin Greek weights

We now provide some examples in the context of the Black-Scholes model. In the first two examples, we derive the optimal Greek weight by the Malliavin integration by parts technique. The last examples show the limitation of this technique as the optimal Greek weight can not be derived. The results of this section are not needed everywhere else in the paper. The reader interested in our statistical results can move straight away to the next section.

Let $T > 0$ be some given finite maturity, and define

$$S_T^{s,\mu,\sigma} \;:=\; s \exp\left[\left(\mu - (\sigma^2/2)\right) T + \sigma W_T\right].$$

In this simple example, the Malliavin derivative process is given by

$$D_r S_T \;=\; \sigma S_T \, 1_{r \le T} \quad \text{for all} \quad r \ge 0.$$

Example 2.1 (Delta of a European option, Black-Scholes model) With $Z^0(s) := S_T^{s,\mu,\sigma}$, we directly compute that $\int_0^\infty D_r Z^0 u_r dr = \sigma S_T \int_0^T u_r dr$ for every $u \in L^2(\mathbb{R}_+ \times \Omega, \mathbb{R})$. Clearly the constant process $u_r^0 := (\sigma s T)^{-1} 1_{r \le T}$ satisfies Condition (2.4), and the associated Greek weight is

$$\pi^0 \;=\; \int_0^T u_r^0 dW_r \;=\; (\sigma s T)^{-1} W_T .$$

Since π^0 is a deterministic function of S_T, we see that π^0 is the optimal Greek weight.

Example 2.2 (Vega of a European option, Black-Scholes model) We now consider the case $Z^0(\sigma) := S_T^{s,\mu,\sigma}$. It is easily checked that the constant process $u_r^0 := [(\sigma T)^{-1} W_T - 1] 1_{r \le T}$ satisfies Condition (2.4), and the associated Greek weight is

$$\pi^0 \;=\; \int_0^T u_r^0 dW_r \;=\; (\sigma T)^{-1} \left[-\sigma T W_T + W_T^2 - T \right] .$$

Since π^0 is a deterministic function of S_T, we see that π^0 is the optimal Greek weight.

Example 2.3 (Delta of an Asian option, Black-Scholes model) We now set $Z^0(s) := \int_0^T S_t^{s,\mu,\sigma} \, dt$. We directly compute that $D_r Z^0 = \sigma \int_r^T S_t dt \; 1_{r \le T}$ for all $r \ge 0$, so that Condition (2.4) reduces to

$$\sigma s \, \mathbb{E}\left[\int_0^T \int_r^T S_t u_r \, dt \, dr \,\middle|\, Z^0 \right] \;=\; \int_0^T S_t dt .$$

Direct computation shows that the process $u_r^0 := \left(\sigma s \int_0^T S_t dt \right)^{-1} 2 S_r$ satisfies Condition (2.4), and the associated Greek weight is

$$\pi^0 \;=\; \int_0^T u_r^0 \, dW_r \;=\; \frac{2}{\sigma^2 s} \left[-\mu + \frac{\sigma^2}{2} + \frac{S_T - s}{\int_0^T S_t dt} \right] .$$

Observe that π^0 is not $\sigma(Z^0)-$measurable. Hence π^0 is not the optimal Greek weight.

Example 2.4 (Delta of an Euro-Asian option, Black-Scholes model) We now set $Z^0(s) := \left(S_T^{s,\mu,\sigma} , \int_0^T S_t^{s,\mu,\sigma} \, dt \right)$. We directly compute that $D_r Z^0 = \sigma \left(S_T , \int_r^T S_t dt \right) 1_{r \le T}$ for all $r \ge 0$, so that Condition (2.4) reduces to

$$\sigma s \, \mathbb{E}\left[\int_0^T u_r dr \,\middle|\, Z^0 \right] \;=\; 1 \quad \text{and} \quad \sigma s \, \mathbb{E}\left[\int_0^T \int_r^T S_t u_r dt dr \,\middle|\, Z^0 \right] \;=\; \int_0^T S_t dt .$$

By direct computation, we see that this condition is satisfied by the process

$$u_r^0 \;:=\; \frac{2}{\sigma s} \left[\frac{-S_r}{\int_0^T S_t dt} + 3 \, \frac{S_r \int_r^T S_t dt}{\left(\int_0^T S_t dt \right)^2} \right] ,$$

and the associated Greek weight is given by

$$\pi^0 = \int_0^T u_r^0 \, dW_r = \frac{1}{\sigma^2 s} \left[-\mu + 3\sigma^2 - \frac{2S_T + 4s}{\int_0^T S_t dt} + 6 \frac{\int_0^T S_t^2 dt}{\left(\int_0^T S_t dt\right)^2} \right].$$

Observe that π^0 is not $\sigma(Z^0)-$measurable. Hence it is not the optimal Greek weight.

2.4 Short maturity singularity

Section 5 of this paper analyzes the case where the random variable $Z(\lambda)$ is the time t realization of a Markov process defined by a stochastic differential equation. We prove a Central Limit Theorem for our Greek estimators when $t \to 0$. In this subsection, we justify the importance of this short time analysis for the purpose of financial applications.

1. *European options with short maturity.* The first two examples of the previous section, exhibit Greek weights which explode when the maturity T shrinks to zero. This feature is by no means specific to the gaussian case. It is shown in Fournié et al. (1999) that this is the rule for any continuous-time process defined as the solution of a (smooth) stochastic differential equation.

2. *path dependent options with fixed maturity.* It was also observed in Fournié et al. (1999) that this problem arises more generally for path dependent payoffs. Indeed, let π : $0 = T_0 < T_1 < \ldots < T_s = 1$ be a partition of the interval $[0, 1]$, and let

$$Z(\lambda) = \phi(X_{T_1}(\lambda), \ldots, X_{T_s}(\lambda))$$

where $\{X_t(\lambda), t \in [0, 1]\}$ is some given continuous-time Markov process parameterized by λ. The partition π is typically a time-grid on which the continuous-time process is discretized. So one should think about the mesh $\max\{|T_i - T_{i-1}| \; : \; 1 \leq i \leq s\}$ to be small.

In this context, by a conditional expectation argument, the Malliavin Greek weights derived by Fournié et al. (1999) are the same as those derived for the random variable

$$\begin{aligned} \bar{Z}(\lambda) := \mathbb{E}[\phi(X_{T_1}(\lambda), \ldots, X_{T_s}(\lambda)) | \mathcal{F}_{T_1}] &= \mathbb{E}[\phi(X_{T_1}(\lambda), \ldots, X_{T_s}(\lambda)) | X_{T_1}(\lambda)] \\ &=: \bar{\phi}(X_{T_1}(\lambda)) , \end{aligned}$$

so that we are reduced to the short maturity T_1 which induces singular Greek weights.

3. *American option / optimal stopping problems with fixed maturity.* Consider the Bermudean approximation V_0 of an American style option with fixed maturity, i.e. the optimal stopping problem with stopping possibilities restricted to the partition π defined above. Then,

by the so-called dynamic programming principle, the value of the Bermudean option can be computed by the backward scheme :

$$V_d(\lambda) \ := \ \phi(X_{T_d}(\lambda)) \quad \text{and} \quad V_{i-1}(\lambda) \ := \ \max\left\{\phi\left(X_{T_{i-1}}(\lambda)\right) \ , \ \mathbb{E}\left[V_i(\lambda)|\mathcal{F}_{T_{i-1}}\right]\right\} .$$

From the Markov property of the process $X(\lambda)$, we see that $V_0(\lambda) = \mathbb{E}[\psi\left(X_{T_1}(\lambda)\right)|X_0 = x]$ in the *continuation region* $\{x \ : \ \phi(x) < V_0(\lambda)\}$, where $\psi\left(X_{T_1}(\lambda)\right) = V_1(\lambda)$, and we are reduced again to a short maturity context, implying the singularity of the Greek weights.

3 Kernel estimation and optimal Greek weight

3.1 Notations

Let $\ell : \mathbb{R}^d \longrightarrow \mathbb{R}$ be some given probability density function, with support containing the parameter of interest λ^0 in its interior, and set

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

We consider a sequence

$$(\Lambda_i, Z_i)_{1 \leq i \leq N} \quad \text{of independent r.v. with distribution} \quad \varphi(\lambda, z) . \tag{3.1}$$

In our application to the estimation of the Greek

$$\beta \ := \ \nabla V^{\phi}(\lambda^0) , \qquad \phi \in \mathbb{L}^{\infty} ,$$

the random variables $(\Lambda_i, Z_i)_{1 \leq i \leq N}$ are generated independently by simulation.

Remark 3.1 Notice that the simulation of $(\Lambda_i, Z_i)_{i \geq 1}$ can be performed easily even in cases where the density φ can not be written explicitly. This applies typically to the case where $Z(\lambda) = X_T(\lambda)$, for some integer T, where $\{X_t(\lambda), t \in \mathbb{N}\}$ is a Markov chain with given transition density. Then, for a given value of λ, the simulation of Z is easily feasible by usual methods. However the marginal distribution of $Z(\lambda)$ is typically very complicated so that it is useless for the numerical computation of the score function $s(\lambda, z)$.

For any function $\psi(\lambda, z)$ defined on $\mathbb{R}^d \times \mathbb{R}^n$, we denote $\psi_\lambda \ := \ \nabla_\lambda \psi$ its partial gradient with respect to the variable λ. For any matrix A, we denote $A^{\otimes} \ := \ AA'$.

In this section, we provide various estimation methods of β based on the Kernel non-parametric kernel method. We then introduce the Kernel functions

$$H \ : \ \mathbb{R}^n \ \longrightarrow \mathbb{R} \quad \text{and} \quad K \ : \ \mathbb{R}^d \ \longrightarrow \mathbb{R},$$

that are symmetric componentwise, compactly supported, C^1, and satisfy $\int H \ = \ \int K \ = 1$.

3.2 A first Kernel estimator of the Greek

The main idea is that the optimal weight S^0 requires a priori the knowledge of the probability density function $f(\lambda, z)$ and the associated *score* function $s(\lambda, z)$. Indeed, if these functions were explicitly known, then a natural non-parametric estimator of the Greek β using the observations (3.1) is :

$$\bar{\beta}_N \;:=\; \frac{1}{\ell(\lambda^0)\, N h^d} \sum_{i=1}^{N} \phi(Z_i)\; s(\Lambda_i, Z_i)\; K\left(\frac{\lambda^0 - \Lambda_i}{h}\right). \tag{3.2}$$

Since s is not explicitly known in our applications of interest, we introduce our first Kernel-based estimator of β :

$$\tilde{\beta}_N \;:=\; \frac{1}{\ell(\lambda^0)\, N h^d} \sum_{i=1}^{N} \phi(Z_i)\; \hat{s}_N^{-i}(\Lambda_i, Z_i)\; K\left(\frac{\lambda^0 - \Lambda_i}{h}\right), \tag{3.3}$$

where

$$\begin{aligned} \hat{s}_N^{-i}(\lambda, z) \;&:=\; \frac{\partial}{\partial \lambda} \ln \left\{ \frac{1}{\ell(\lambda)\, N h^{d+n}} \sum_{j=1, j\neq i}^{N} K\left(\frac{\lambda - \Lambda_j}{h}\right) H\left(\frac{z - Z_j}{h}\right) \right\} \\ &=\; \frac{\hat{\varphi_\lambda}^{-i}}{\hat{\varphi}^{-i}}(\lambda, z) \;-\; \frac{\nabla \ell(\lambda)}{\ell(\lambda)}, \end{aligned} \tag{3.4}$$

and

$$\hat{\varphi}^{-i}(\lambda, z) \;:=\; \frac{h^{-d-n}}{N-1} \sum_{j=1, j\neq i}^{N} K\left(\frac{\lambda - \Lambda_j}{h}\right) H\left(\frac{z - Z_j}{h}\right), \tag{3.5}$$

$$\hat{\varphi_\lambda}^{-i}(\lambda, z) \;:=\; \frac{h^{-d-n-1}}{N-1} \sum_{j=1, j\neq i}^{N} \nabla K\left(\frac{\lambda - \Lambda_j}{h}\right) H\left(\frac{z - Z_j}{h}\right). \tag{3.6}$$

The asymptotic properties of this estimator will be provided in Section 4. Before discussing the rate of convergence of $\tilde{\beta}_N$, let us highlight two drawbacks of this estimator. First, its expression involves a product of two (possibly multidimensional) Kernels K and H. Thus, it suffers from the so-called "curse of dimensionality". Moreover, its calculation is time-consuming. In the subsequent subsections, we introduce two alternative Kernel estimators of β, which involve a single Kernel function and a single summation.

3.3 A simpler Kernel estimator of the Greek

For fixed $h > 0$, it follows from the Law of large numbers that

$$\bar{\beta}_N \underset{N\to\infty}{\longrightarrow} \mathbb{E}\left[\bar{\beta}_N\right] \;=\; \frac{1}{\ell(\lambda^0) h^d}\, \mathbb{E}\left[\phi(Z) s(\Lambda, Z)\, K\left(\frac{\lambda^0 - \Lambda}{h}\right)\right], \quad P-\text{a.s.} \tag{3.7}$$

where (Λ, Z) is a random variable with distribution $\varphi(\lambda, z)$. Recalling the definition of s, and integrating by parts with respect to the variables $\lambda_1, \ldots, \lambda_d$, we see that for $h > 0$ sufficiently small :

$$\begin{aligned}
\mathbb{E}\left[\bar{\beta}_N\right] &= \frac{1}{\ell(\lambda^0)h^d} \int \phi(z) K\left(\frac{\lambda^0 - \lambda}{h}\right) \ell(\lambda) \nabla_\lambda f(\lambda, z)\, d\lambda dz \\
&= \frac{1}{\ell(\lambda^0)h^{d+1}} \int \phi(z) \left(\nabla K\left(\frac{\lambda^0 - \lambda}{h}\right) - hK\left(\frac{\lambda^0 - \lambda}{h}\right) \frac{\nabla \ell}{\ell}(\lambda)\right) \varphi(\lambda, z)\, d\lambda dz \\
&= \frac{1}{\ell(\lambda^0)h^{d+1}} \mathbb{E}\left[\phi(Z) \left(\nabla K\left(\frac{\lambda^0 - \Lambda}{h}\right) - hK\left(\frac{\lambda^0 - \Lambda}{h}\right) \frac{\nabla \ell}{\ell}(\Lambda)\right)\right],
\end{aligned}$$

since the Kernels are compactly supported. This suggests the following simpler Kernel estimator of the Greek β :

$$\hat{\beta}_N := \frac{1}{\ell(\lambda^0)Nh^{d+1}} \sum_{i=1}^{N} \phi(Z_i) \left(\nabla K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) - hK\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \frac{\nabla \ell}{\ell}(\Lambda_i)\right). \quad (3.8)$$

The asymptotic properties of $\hat{\beta}_N$ will be provided in Section 4.

3.4 Differentiating the Kernel estimator of the price

We next start out from the natural Kernel estimator of $V^\phi(\lambda)$:

$$\hat{V}_N^\phi(\lambda) := \frac{1}{Nh^d\, \ell(\lambda)} \sum_{i=1}^{N} \phi(Z_i) K\left(\frac{\lambda - \Lambda_i}{h}\right).$$

Differentiating $\hat{V}_N^\phi(\lambda)$ with respect to λ, we obtain our final Kernel estimator of the Greek:

$$\check{\beta}_N := \frac{1}{\ell(\lambda^0)Nh^{d+1}} \sum_{i=1}^{N} \phi(Z_i) \left(\nabla K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) - hK\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \frac{\nabla \ell}{\ell}(\lambda^0)\right). \quad (3.9)$$

Remark 3.2 Observe that our two estimators $\hat{\beta}_N$ and $\check{\beta}_N$ are closely related by :

$$\check{\beta}_N = \hat{\beta}_N + \frac{1}{\ell(\lambda^0)Nh^d} \sum_{i=1}^{N} \phi(Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \left(\frac{\nabla \ell}{\ell}(\Lambda_i) - \frac{\nabla \ell}{\ell}(\lambda^0)\right).$$

In particular,

$$\check{\beta}_N = \hat{\beta}_N \quad \text{whenever} \quad \ell \text{ is a uniform probability density.}$$

The asymptotic properties of this third estimator will also be provided in Section 4.

4 Asymptotic results

In this section, we compare the estimators defined in the previous section from the viewpoint of their asymptotic distributions. Observe that the estimators $\hat{\beta}_N$ and $\check{\beta}_N$ involve the Kernel function K, while the first estimator $\tilde{\beta}_N$ involves both Kernel functions K and H. From the practical viewpoint, the *single Kernel-based* estimators are clearly easier to implement. The main results of this section imply that there is no advantage from using the cumbersome *double Kernel-based* estimator.

Technically speaking, our derivation of the asymptotic properties of the double Kernel-based estimator requires more stringent conditions, and involves heavy calculations. We then start by presenting the asymptotic results for the estimators $\hat{\beta}_N$ and $\check{\beta}_N$.

Before stating our results, we recall that the order of the Kernel function K is defined as the smallest non zero integer p such that there exist some integers $(j_1, \ldots, j_p)$, $j_k \in \{1, \ldots, d\}$, such that

$$\int l_{\alpha_1} \ldots l_{\alpha_r} K(l) dl \;=\; 0 \quad \text{when } 0 < r < p,\ \alpha_k \in \{1, \ldots, d\}, \quad \text{and}$$
$$\int l_{j_1} \ldots l_{j_p} K(l) dl \neq 0.$$

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

Finally, to simplify the notations and for every smooth function ψ defined on $\mathbb{R}^d \times \mathbb{R}^n$, we introduce the following map on $\mathbb{R}^d \times \mathbb{R}^n$:

$$\xi_K^p[\psi](\lambda, z) \;:=\; \frac{(-1)^p}{p!} \sum_{j_1, \ldots, j_p = 1}^{d} \left(\int l_{j_1} \ldots l_{j_p} K(l) dl \right) \nabla^p_{\lambda_{j_1} \ldots \lambda_{j_p}} \psi(\lambda, z) \,. \tag{4.1}$$

4.1 Asymptotic results for the single Kernel-based estimators

Our first result requires some regularity conditions on the density functions f and ℓ.

Assumption R1 *For every z, the functions $f(\cdot, z)$ and ℓ are $p+1$ times differentiable, and for every integer $i \leq p$, the function $\lambda \longmapsto \nabla^i_\lambda \{\ell(\lambda) \nabla_\lambda f(\lambda, z)\}$ is continuous at $\lambda = \lambda^0$ uniformly with respects to $z \in S$, for some subset S s.t.* $\mathrm{Supp}(\phi) \subset int(S)$.

Proposition 4.1 *Let the Kernel function K be of order $p > 0$, and let Assumption* R1 *hold. Then, as $N \to \infty$ and $h \to 0$, the bias and the variance of $\hat{\beta}_N$ satisfy*

$$\mathbb{E}\left[\hat{\beta}_N\right] - \beta \;\sim\; C_1 h^p \quad \textit{and} \quad \mathbb{V}\mathrm{ar}\left[\hat{\beta}_N\right] \;\sim\; N^{-1} h^{-d-2}\, \Sigma \,, \tag{4.2}$$

where

$$C_1 \;:=\; \frac{1}{\ell(\lambda^0)}\int \xi_K^p\left[\ell\, f_\lambda\right](\lambda^0,z)\,\phi(z)\,dz\;,\quad \textit{and}\quad \Sigma \;:=\; \frac{1}{\ell(\lambda^0)}\mathbb{E}[\phi^2(Z^0)]\int \nabla K^{\otimes}. \tag{4.3}$$

Proof. By definition of the estimator $\hat\beta_N$, we have $\mathbb{E}\left[\hat\beta_N\right] = \mathbb{E}\left[\bar\beta_N\right]$. By (3.7), this provides

$$\begin{aligned}\psi(h) \;:=\; \mathbb{E}\left[\hat\beta_N\right] &= \frac{1}{\ell(\lambda^0)h^d}\int \phi(z)\ell(\lambda)\nabla_\lambda f(\lambda,z)K\left(\frac{\lambda^0-\lambda}{h}\right)d\lambda dz\\ &= \frac{1}{\ell(\lambda^0)}\int \phi(z)\ell(\lambda^0-hl)\,f_\lambda(\lambda^0-hl,z)K(l)dl\,dz.\end{aligned}$$

Clearly,

$$\psi(0) \;=\; \int \phi(z)f_\lambda(\lambda^0,z)dz \;=\; \beta.$$

Moreover, since K has compact support, it follows from Assumption R1 that the function ψ is q times differentiable at zero, with derivatives obtained by differentiating inside the integral sign, i.e.

$$\begin{aligned}\psi^{(i)}(0) \;&:=\; \frac{\partial^i\psi}{\partial h^i}(0)\\ &=\; \frac{(-1)^i}{\ell(\lambda^0)}\sum_{j_1,\dots,j_i=1}^{d}\left(\int l_{j_1}\dots l_{j_i}K(l)\,dl\right)\left(\int \phi(z)\left[\nabla^i_{\lambda_{j_1},\dots,\lambda_{j_i}}\{\ell\, f_\lambda\}\right](\lambda^0,z)\,dz\right)\end{aligned}$$

for every $1\le i\le p$. Since p is the order of K, observe that $\psi^{(i)}(0)=0$ for every $1\le i<p$, so that a Taylor expansion of ψ provides the first part of the proposition.

As for the mean, we directly compute that

$$\mathrm{Var}\left[\hat\beta_N\right] \;=\; \frac{(v_1-v_2^{\otimes})}{Nh^{2d+2}\ell(\lambda^0)^2}\;,$$

where

$$\begin{aligned}v_1 \;&:=\; \mathbb{E}\left[\phi(Z)^2\left(\nabla K\left(\frac{\lambda^0-\Lambda}{h}\right)-hK\left(\frac{\lambda^0-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\Lambda)\right)^{\otimes}\right]\;,\\ v_2 \;&:=\; \mathbb{E}\left[\phi(Z)\left(\nabla K\left(\frac{\lambda^0-\Lambda}{h}\right)-hK\left(\frac{\lambda^0-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\Lambda)\right)\right].\end{aligned}$$

By a similar argument as in the first part of this proof, we compute that

$$\begin{aligned}v_1 \;&=\; h^d\int \phi^2(z)\left(\nabla K\left(l\right)-hK\left(l\right)\frac{\nabla\ell}{\ell}(\lambda^0-hl)\right)^{\otimes}\ell(\lambda^0-hl)f(\lambda^0-hl,z)dl\,dz\\ &\sim\; h^d\ell(\lambda^0)\left(\int \nabla K(l)^{\otimes}dl\right)\mathbb{E}\left[\phi^2(Z^0)\right]\;.\end{aligned}$$

The required result follows by observing that $v_2 = O\left(h^{d+1}\right)$. □

We are now ready for our first main result.

Theorem 4.1 (i) *Let the conditions of Proposition 4.1 hold, and assume further that*

$$h \longrightarrow 0 \quad \text{and} \quad N\,h^{d+2} \longrightarrow \infty \quad \text{as} \quad N \to \infty. \tag{4.4}$$

Then, with Σ as in (4.3), the asymptotic distribution of $\hat{\beta}_N$ is given by

$$\sqrt{Nh^{d+2}}\left(\hat{\beta}_N - \mathbb{E}[\hat{\beta}_N]\right) \quad \underset{N\to\infty}{\overset{law}{\longrightarrow}} \quad \mathcal{N}(0,\Sigma)\ .$$

(ii) *In addition to the above conditions, assume that*

$$N\,h^{d+2+2p} \longrightarrow 0 \quad \text{as} \quad N \to \infty. \tag{4.5}$$

Then the bias vanishes and

$$\sqrt{Nh^{d+2}}\left(\hat{\beta}_N - \beta\right) \quad \underset{N\to\infty}{\overset{law}{\longrightarrow}} \quad \mathcal{N}(0,\Sigma)\ .$$

Proof. We shall prove this result by verifying the Lyapounov conditions (see e.g. Billingsley (1968), p.44). Let a be an arbitrary vector in $\mathbb{R}^d$, and define, for every $i = 1,\ldots,N$,

$$X_i^N \ := \ a'\ \left(Y_i^N - \mathbb{E}[Y_i^N]\right)\ ,$$

where

$$Y_i^N \ := \ \frac{1}{Nh^{d+1}\ell(\lambda^0)}\phi(Z_i)\left(\nabla K\left(\frac{\lambda^0-\Lambda_i}{h}\right) - hK\left(\frac{\lambda^0-\Lambda_i}{h}\right)\frac{\nabla\ell}{\ell}(\Lambda_i)\right)\ .$$

In view of Proposition 4.1, the only condition which remains to check in order to verify the Lyapounov conditions is

$$\sup_N \frac{1}{\sigma_N^\delta}\sum_{i=1}^N \mathbb{E}[|X_i^N|^\delta] \ < \ +\infty \text{ for some } \delta > 2, \quad \text{where} \quad \sigma_N^2 \ := \ \mathbb{V}\text{ar}\left[\sum_{i=1}^N X_i^N\right]\ . \tag{4.6}$$

In order to prove (4.6), we start by observing that

$$\sigma_N^2 \ \sim \ \frac{1}{Nh^{d+2}\ell(\lambda^0)}\ \mathbb{E}[\phi^2(Z^0)]\int |a'\nabla K(l)|^2\,dl$$

as a consequence of (4.2). We next estimate by the Minkowski inequality and (4.2) that

$$\begin{aligned}
\left\|X_i^N\right\|_\delta \ &\leq \ \left\|a'Y_i^N\right\|_\delta \ + \ \left|a'\mathbb{E}[Y_i^N]\right| \\
&= \ \left\|a'Y_i^N\right\|_\delta \ + \ \frac{1}{N}\ \left|a'\mathbb{E}[\hat{\beta}_N]\right| \\
&\leq \ \frac{\sum_{i=1}^d \left\|\phi(Z)a_i\left(\nabla_i K\left(\frac{\lambda^0-\Lambda}{h}\right) - hK\left(\frac{\lambda^0-\Lambda}{h}\right)\frac{\nabla_i\ell}{\ell}(\Lambda)\right)\right\|_\delta}{Nh^{d+1}\ell(\lambda^0)} \ + \ O\left(\frac{1}{N}\right) \\
&\leq \ Const\ \frac{h^{d/\delta}}{Nh^{d+1}} + \frac{1}{N}\,,
\end{aligned}$$

where the last estimate is obtained by a Taylor expansion with respect to the h variable, in the neighborhood of the origin, following the method used in the proof of Proposition 4.1. Hence,

$$\frac{1}{\sigma_N^\delta}\sum_{i=1}^N \mathbb{E}\left[|X_i^N|^\delta\right] \;\leq\; Const\, N\,\frac{h^d}{(Nh^{d+1})^\delta}\,(Nh^{d+2})^{\delta/2} \;\leq\; \frac{Const}{(Nh^d)^{(\delta-2)/2}}\,,$$

and condition (4.6) is satisfied when $Nh^d \to \infty$, as assumed in (4.4). Therefore, $\sum_{i=1}^N X_i^N$ is asymptotically gaussian, with a variance matrix given by σ_N^2. By the arbitrariness of $a \in \mathbb{R}^d$, the required result follows from the Cramer-Wold device. □

We next turn to the estimator $\check{\beta}$ which was defined as the gradient, with respect to λ, of the Kernel based estimator $\tilde{V}_N^\phi(\lambda)$ of the function $V_N^\phi(\lambda)$. The asymptotic properties of this estimator are obtained by following the techniques of the previous proofs and require the following regularity condition on the densities f and ℓ.

Assumption R2 *For every z, the functions $f(\cdot,z)$ and ℓ are $p+1$ times differentiable, and for every integer integer $i \leq p$, the function $\lambda \longmapsto \nabla_\lambda^i\{\ell(\lambda)f(\lambda,z)\}$ is continuous at $\lambda = \lambda^0$ uniformly with respects to $z \in S$, for some subset S s.t.* $\mathrm{Supp}(\phi) \subset int(S)$.

Proposition 4.2 *Let the Kernel function K be of order $p > 0$, and let Assumption* R2 *hold. Then, the bias and the variance of $\check{\beta}_N$ satisfy*

$$\mathbb{E}[\check{\beta}_N] - \beta \sim C_2 h^p \qquad and \qquad \mathbb{V}\mathrm{ar}[\check{\beta}_N] \sim \frac{\Sigma}{Nh^{d+2}}\,,$$

where Σ is given by (4.3), *and*

$$C_2 \;:=\; \frac{1}{\ell(\lambda^0)}\int\left(\xi_K^p[\varphi_\lambda] - \frac{\nabla\ell}{\ell}(\lambda^0)\,\xi_K^p[\varphi]\right)(\lambda^0,z)\,\phi(z)\,dz\,.$$

Proof. The proof is essentially similar to that of Proposition 4.1. Recall that the estimators $\check{\beta}_N$ and $\hat{\beta}_N$ are related by :

$$\check{\beta}_N \;=\; \hat{\beta}_N \;+\; \frac{1}{\ell(\lambda^0)Nh^d}\sum_{i=1}^N \phi(Z_i)K\left(\frac{\lambda^0-\Lambda_i}{h}\right)\left(\frac{\nabla\ell}{\ell}(\Lambda_i) - \frac{\nabla\ell}{\ell}(\lambda^0)\right). \qquad (4.7)$$

We start by analyzing the bias term. Recall from the proof of Proposition 4.1 that :

$$\mathbb{E}\left[\hat{\beta}_N\right] \;=\; \frac{1}{\ell(\lambda^0)}\int \phi(z)\ell(\lambda^0-hl)f_\lambda(\lambda^0-hl,z)K(l)\,dl\,dz\,.$$

We then deduce from (4.7) that :

$$\mathbb{E}\left[\check{\beta}_N\right] \;=\; \frac{1}{\ell(\lambda^0)}\int \phi(z)\left(\varphi_\lambda(\lambda^0-hl,z) - \frac{\nabla\ell}{\ell}(\lambda_0)\varphi(\lambda^0-hl,z)\right)K(l)\,dl\,dz\,.$$

We now observe that Assumption R2 allows to derive an expansion in the h variable of the above expression, near the origin, up to the order p. The coefficients of the expansion are obtained by simple differentiation inside the integral sign. Finally, since p is the order of the Kernel K, it is easily seen that the coefficients of h^i, $i < p$, in this expansion vanish, and the only non-zero coefficient is that of h^p.

The variance of $\check{\beta}_N$ is also treated by the same argument as in the proof of Proposition 4.1, and the dominant term in the expansion of the variance is easily seen to be the same as in that proof. □

Proposition 4.2 says that $\tilde{\beta}_N$ and $\check{\beta}_N$ have the same asymptotic variance, and the orders of their asymptotic biases are the same. Our next result states that these two estimators have exactly the same asymptotic distribution.

Theorem 4.2 (i) *Let the conditions of Proposition 4.2 hold, and assume further that* (4.4) *holds. Then, with* Σ *as in* (4.3), *the asymptotic distribution of* $\check{\beta}_N$ *is given by*

$$\sqrt{Nh^{d+2}}\left(\check{\beta}_N - \mathbb{E}[\check{\beta}_N]\right) \quad \underset{N\to\infty}{\overset{law}{\longrightarrow}} \quad \mathcal{N}(0,\Sigma) \ .$$

(ii) *If we assume that* (4.5) *holds, in addition to the above conditions, then the bias vanishes and :*

$$\sqrt{Nh^{d+2}}\left(\check{\beta}_N - \beta\right) \quad \underset{N\to\infty}{\overset{law}{\longrightarrow}} \quad \mathcal{N}(0,\Sigma) \ .$$

Proof. Define the sequence

$$Y_i^N \ := \ \frac{1}{Nh^{d+1}\ell(\lambda^0)}\,\phi(Z_i)\left(\nabla K\left(\frac{\lambda^0-\Lambda^i}{h}\right) - hK\left(\frac{\lambda^0-\Lambda^i}{h}\right)\frac{\nabla\ell}{\ell}(\lambda^0)\right) ,$$

and follow the lines of the proof of Theorem 4.1. □

4.2 Asymptotic properties of the double Kernel-based estimator

As in the previous section, we start by analyzing the asymptotics of the bias and the variance of $\tilde{\beta}_N$. We first introduce some additional conditions which will be needed in our subsequent analysis.

Assumption K *K and H are the product of some univariate Kernels with orders respectively p and q, and ∇K has bounded variation.*

Assumption S *ϕ is continuous and has compact support. Moreover, for every $z \in \mathbb{R}^n$,*

$$\inf\left\{\varphi(\lambda,z) \ : \ (\lambda,z)\in\mathcal{V}(\lambda^0)\times C_\phi\right\} \ > \ 0,$$

for some neighborhood $\mathcal{V}(\lambda^0)$ of λ^0, and some compact subset C_ϕ of $\mathbb{R}^n$ with $\mathrm{Supp}(\phi) \subset \mathrm{int}(C_\phi)$.

Assumption R3 *For every λ, the function $\nabla_\lambda f(\lambda, \cdot)$ is q times differentiable, and for every integer $i \leq q$, the function $\lambda \longmapsto \nabla_z^i \nabla_\lambda \varphi(\lambda, z)$ is continuous at $\lambda = \lambda^0$ uniformly with respect to $z \in S$, for some subset S s.t.* $\mathrm{Supp}(\phi) \subset int(S)$.

Notice that Assumption S is in practice unfortunately very restrictive concerning the choice of ϕ. Following the notation (4.1), we define for any function ψ on $\mathbb{R}^d \times \mathbb{R}^n$:

$$\xi_H^q[\psi](\lambda, z) \;:=\; \frac{(-1)^q}{q!} \sum_{j_1,\ldots,j_q=1}^{d} \left(\int v_{j_1} \ldots v_{j_q} H(v) dv \right) \nabla^q_{z_{j_1}\ldots z_{j_q}} \psi(\lambda, z) \,. \tag{4.8}$$

Proposition 4.3 *Let Assumptions* K, S, R1, R2 *and* R3 *hold. Choose further N and h so that*

$$h \;\longrightarrow\; 0 \quad \textit{and} \quad \frac{ln^2 N}{N\, h^{d+2n}} \;\longrightarrow\; 0 \quad \textit{as} \;\; N \;\to\; \infty \,. \tag{4.9}$$

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

$$\mathbb{E}\left[\tilde{\beta}_N\right] - \beta \;\sim\; C_3 h^p + C_4 h^q + \frac{C_5}{N h^{d+n+1}} \quad \textit{and} \quad \mathbb{V}\mathrm{ar}\left[\tilde{\beta}_N\right] \;\sim\; \frac{\tilde{\Sigma}}{N h^{d+2}} \,,$$

where

$$\begin{aligned}
C_3 \;&:=\; \frac{1}{\ell(\lambda^0)} \int \left[\xi_K^p\left[\ell f_\lambda + \varphi_\lambda\right] - \frac{\varphi_\lambda}{\varphi} \xi_K^p\left[\varphi\right] \right] (\lambda^0, z) \phi(z)\, dz \\
C_4 \;&:=\; \frac{1}{\ell(\lambda^0)} \int \left[\xi_H^q\left[\varphi_\lambda\right] - \frac{\varphi_\lambda}{\varphi} \xi_H^q\left[\varphi\right] \right] (\lambda^0, z)\, \phi(z)\, dz \\
C_5 \;&:=\; \frac{1}{\ell(\lambda^0)} \int \frac{\phi(z)}{\varphi(\lambda^0, z)} K(l_2 - l_1) K(l_1) \nabla K(l_1) H^2(v)\, dl_1\, dl_2\, dv\, dz
\end{aligned}$$

and

$$\tilde{\Sigma} \;:=\; \frac{1}{\ell(\lambda^0)} \int \phi^2(z) \left\{ \int K(l_2 - l_1) \nabla K(l_1)\, dl_1 \right\}^{\otimes} f(\lambda^0, z)\, dz\, dl_2 \,.$$

The proof of this result involves heavy calculations, and is reported in the Appendix section.

Theorem 4.3 (i) *Let the conditions of Proposition 4.3 hold. Then, the asymptotic distribution of $\tilde{\beta}_N$ is given by*

$$\sqrt{N h^{d+2}} \left(\tilde{\beta}_N - \mathbb{E}[\tilde{\beta}_N] \right) \;\underset{N\to\infty}{\overset{law}{\longrightarrow}}\; \mathcal{N}\left(0, \tilde{\Sigma}\right) \,.$$

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

$$\sqrt{N h^{d+2}} \left(\tilde{\beta}_N - \beta \right) \;\underset{N\to\infty}{\overset{law}{\longrightarrow}}\; \mathcal{N}\left(0, \tilde{\Sigma}\right) \,.$$

The proof is also reported in the Appendix Section. Note that it is necessary to have $n < (p \wedge q)+1$, in order to satisfy (4.9) and the previous conditions of (ii). Thus, for basket derivatives or bermudean options, il would be necessary to consider high-order kernels.

4.3 Optimal choice of N and h

The two single Kernel-based estimators $\hat{\beta}_N$ and $\check{\beta}_N$ have similar asymptotic properties. They both have a bias of order h^p, a variance of order $1/(Nh^{d+2})$ and a convergence in distribution at the rate $\sqrt{Nh^{d+2}}$. Therefore, the determination methods of the optimal N and h will be similar for both of them.

Typically, considering $\hat{\beta}_N$, under some assumptions, it follows from Proposition 4.1 that

$$\mathbb{E}\left[\hat{\beta}_N\right] - \beta \sim C_1 h^p \quad \text{and} \quad \mathbb{V}\text{ar}\left[\hat{\beta}_N\right] \sim \frac{\Sigma}{Nh^{d+2}}.$$

We deduce from this result an equivalent for the mean squared error between $\hat{\beta}_N$ and β, i.e.

$$MSE(\hat{\beta}_N) \;:=\; \mathbb{E}\left[|\hat{\beta}_N - \beta|^2\right] \sim \frac{\mathrm{Tr}(\Sigma)}{Nh^{d+2}} + h^{2p}|C_1|^2 .$$

Differentiating with respect to h, we get the following asymptotically optimal bandwidth selector :

$$\hat{h} \;=\; \left(\frac{(d+2)Tr(\Sigma)}{2p|C_1|^2 N}\right)^{1/(d+2p+2)} . \tag{4.10}$$

Note that the optimal $\hat{h}$ is of order $N^{-1/(d+2p+2)}$, leading to an MSE of order $N^{-2p/(d+2p+2)}$. The same calculations with the estimator $\check{\beta}_N$ leads to the optimal bandwidth

$$\check{h} \;=\; \left(\frac{(d+2)Tr(\Sigma)}{2p|C_2|^2 N}\right)^{1/(d+2p+2)} . \tag{4.11}$$

The study of the double Kernel-based estimator is a little more intricate. Recall that, under the condition $ln^2 N/(N\, h^{d+2n}) \to 0$,

$$\mathbb{E}\left[\tilde{\beta}_N\right] - \beta \sim C_3 h^p + C_4 h^q + \frac{C_5}{Nh^{d+n+1}} \quad \text{and} \quad \mathbb{V}\text{ar}\left[\tilde{\beta}_N\right] \sim \frac{\tilde{\Sigma}}{Nh^{d+2}}.$$

After some calculation, the minimization of the MSE of $\tilde{\beta}_N$ under the above constraint leads also to an optimal $\tilde{h}$ of order $N^{-1/(d+2p(\wedge q)+2)}$ and an associated MSE of order $N^{-2(p\wedge q)/(d+2(p\wedge q)+2)}$. And therefore, the optimal bandwidth is given by:

$$\tilde{h} \;=\; \left(\frac{(d+2)Tr(\tilde{\Sigma})}{2(p\wedge q)\,|C_3 1_{p\leq q} + C_4 1_{q\leq p}|^2\, N}\right)^{1/(d+2(p\wedge q)+2)} . \tag{4.12}$$

Note that the latter $\tilde{h}$ satisfies $N\tilde{h}^{d+2n} >> \ln^2 N$ assuming $n < 1 + p \wedge q$. These results imply an asymptotic theoretical choices for h relative to N, but we may still encounter difficulties in the numerical calculation of h. Even if the optimal order α of h were known, we still need to evaluate the associated constant coefficients. From our empirical experiments, we observed that the accuracy of our estimators depends heavily on the choice of the bandwidth h, as usual in kernel estimation.

4.4 The case of an uniform a priori distribution on A

In this subsection we study further the case where ℓ is the most natural density in $\mathbb{R}^d$: a density centered at λ^0, the product of independent uniform densities on $\mathbb{R}$ with width $2a$. Then , we investigate the optimal choice of the width a of the density ℓ given by :

$$\ell(x) \;=\; \frac{1}{(2a)^d} 1_{|x-\lambda_0|_\infty \leq a} \;,$$

As pointed out in Remark 3.2, in this case, the single kernel density estimators are equal:

$$\hat{\beta}_N \;=\; \check{\beta}_N \;=\; \frac{(2a)^d}{Nh^{d+1}} \sum_{i=1}^{N} \phi(Z_i) \nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right).$$

In order to simplify the presentation and without loss of generality, we will consider Kernel functions K with support on $[-1,1]^d$.

For a given a, by proposition 4.1, when N goes to infinity and h goes to zero, our estimator $\hat{\beta}_N$ has the following properties :

$$\mathbb{E}\left[\hat{\beta}_N\right] - \beta \sim \left(\int \xi_K^p \left[f_\lambda\right](\lambda^0, z) \phi(z) dz \right) h^p \quad \text{and} \quad \mathbb{V}\text{ar}\left[\hat{\beta}_N\right] \sim \frac{(2a)^d}{Nh^{d+2}} \mathbb{E}[\phi^2(Z^0)] \int \nabla K^{\otimes}.$$

Notice here that the asymptotic bias is independent of a, and the variance is an increasing function of a. This leads to study the asymptotic properties of our estimator, when the parameter a goes to zero. To do this, we first rewrite Assumption R1 in the case where ℓ is uniform:

Assumption R4 *For every z, the function $f(\cdot, z)$ is $p+1$ times differentiable, and for every integer $i \leq p+1$, the function $\lambda \longmapsto \nabla_\lambda^i f(\lambda, z)$ is continuous at $\lambda = \lambda^0$ uniformly with respects to $z \in S$, for some subset S s.t.* $\text{Supp}(\phi) \subset int(S)$.

Proposition 4.4 *Let the Kernel function K be of order $p > 0$ and let Assumption* R4 *hold. Then, as $N \to \infty$, $h \to 0$ and $a \to 0$ with $a \geq h$, the bias and the variance of $\hat{\beta}_N$ satisfy*

$$\mathbb{E}\left[\hat{\beta}_N\right] - \beta \;\sim\; C_a h^p \quad \textit{and} \quad \mathbb{V}\text{ar}\left[\hat{\beta}_N\right] \;\sim\; N^{-1} h^{-d-2} a^d \, \Sigma_a \;, \tag{4.13}$$

where

$$C_u := \int \xi_K^p [f_\lambda](\lambda^0, z)\, \phi(z)\, dz\ , \quad and \quad \Sigma_u := 2^d\, \mathbb{E}[\phi^2(Z^0)] \int \nabla K^{\otimes 2}. \tag{4.14}$$

Proof. The proof is similar to the one of proposition 4.1. Denoting by $\mathbf{1_d}$ the element of $\mathbb{R}^d$ with each component equal to 1, we rewrite

$$\begin{aligned}
\mathbb{E}\left[\hat{\beta}_N\right] &= \frac{1}{h^{d+1}} \int_{\mathbb{R}^n} \phi(z) \left(\int_{\lambda^0 - a\mathbf{1_d}}^{\lambda^0 + a\mathbf{1_d}} \nabla K \left(\frac{\lambda^0 - \lambda}{h} \right) f(\lambda, z)\, d\lambda \right) dz \\
&= \frac{1}{h} \int_{\mathbb{R}^n} \phi(z) \left(\int_{[-\frac{a}{h}, \frac{a}{h}]^d} \nabla K(u) f(\lambda^0 - uh, z)\, du \right) dz.
\end{aligned}$$

The only noticeable part of the proof is to see why we need a to be greater than h. Since K is supported on $[-1,1]^d$, it allows us to replace in our last term the integration on $[-\frac{a}{h}, \frac{a}{h}]^d$ by an integration on $\mathbb{R}^d$ which is necessary to get the convergence of our estimator to β. Then, as in the proof of proposition 4.1, an integration by parts followed by Taylor expansions gives us the expected equivalent of the bias. The exact same argument applies for the computation of the variance of $\hat{\beta}_N$. □

Letting a go to zero, we obtained here the same asymptotic properties as the one found in proposition 4.1, as long as a is bigger than h. Therefore, the asymptotic optimal a is simply the bandwidth h. The Kernel-based estimator $\hat{\beta}_N^*$, associated to this optimal uniform density ℓ with width $2h$, simply rewrites

$$\hat{\beta}_N^* := \frac{2^d}{Nh} \sum_{i=1}^N \phi(Z_i) \nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right)\ , \tag{4.15}$$

and satisfies

$$\mathbb{E}\left[\hat{\beta}_N^*\right] - \beta \sim C_u h^p \quad \text{and} \quad \mathbb{V}\text{ar}\left[\hat{\beta}_N^*\right] \sim N^{-1} h^{-2}\, \Sigma_u\ , \tag{4.16}$$

with C_u and Σ_u defined in 4.14.

Minimizing the mean square error associated to this estimator, we obtain the optimal bandwidth

$$h^* = \left(\frac{Tr\Sigma_u}{p|C_u|^2\, N} \right)^{\frac{1}{2p+2}}$$

leading to the following MSE:

$$MSE\left(\hat{\beta}_N^*\right) = 2(p+1) p^{-\frac{p}{p+1}} \left[|C_u|^2\, (Tr\Sigma_u)^p \right]^{\frac{1}{p+1}} \ N^{-\frac{p}{p+1}}.$$

As in the study of the previous estimators, we also obtain a central limit theorem for the estimator $\hat{\beta}_N^*$.

Theorem 4.4 (i) *Let the conditions of Proposition 4.4 hold in the particular case where $a = h$, and assume further that*

$$h \longrightarrow 0 \quad \text{and} \quad N h^2 \longrightarrow \infty \quad \text{as} \quad N \to \infty . \tag{4.17}$$

Then, with Σ_a as in (4.14), the asymptotic distribution of $\hat{\beta}_N^*$ is given by

$$\sqrt{Nh^2}\left(\hat{\beta}_N^* - \mathbb{E}[\hat{\beta}_N^*]\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0, \Sigma_a) .$$

(ii) *If in addition $Nh^{2p+2} \to 0$, then the bias vanishes and :*

$$\sqrt{Nh^2}\left(\hat{\beta}_N^* - \beta\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0, \Sigma_a) .$$

A remarkable feature of the above asymptotic result is that the rate of convergence is independent of the dimension d of the parameter λ^0.

Remark 4.1 Notice that the optimal choice of the bandwidth parameter a depends on the Kernel function K only through its support. For instance, if $\text{supp}(K) = [-M, M]^d$, then the optimal bandwidth of the uniform density ℓ is $a = Mh$.

4.5 Comparison with the finite differences estimators

We first start by recalling the finite differences estimators. For ease of presentation, we let $d = 1$. The finite differences estimator of the parameter $\beta := \nabla_\lambda \mathbb{E}[\phi(Z(\lambda^0))]$ is based on the finite differences approximation of the gradient

$$\nabla_\lambda \mathbb{E}[\phi(Z(\lambda^0))] \sim \frac{\mathbb{E}[\phi(Z(\lambda^0 + \alpha\varepsilon))] - \mathbb{E}[\phi(Z(\lambda^0 - (1-\alpha)\varepsilon))]}{\varepsilon},$$

where $\varepsilon > 0$ is a "small" parameter, and $\alpha \in [0, 1]$. The values $\alpha = 0$, 0.5 and 1 correspond respectively to the backward, centered and forward finite difference. The above finite difference approximation suggests the following finite differences estimator of β :

$$\hat{\beta}_N^{FD} = \frac{1}{N\varepsilon}\sum_{i=1}^{N}\left(\phi\left[Z^i(\lambda^0 + \alpha\varepsilon)\right] - \phi\left[Z^i(\lambda^0 - (1-\alpha)\varepsilon)\right]\right) .$$

The asymptotic properties of these estimators were first studied by L'Ecuyer and Perron (1994). In the case where $\lambda \mapsto \phi[Z_\lambda] \in C_b^3(\mathbb{R}^d)$, when $N \to \infty$ and $\varepsilon \to 0$ with $N^{1/4}\varepsilon \to 0$, they obtained a parametric rate of convergence :

$$\sqrt{N}\left(\hat{\beta}_N^{FD} - \beta\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0, \Sigma_\alpha) , \qquad \text{for } \alpha = 0 , \frac{1}{2} \text{ and } 1 .$$

When the payoff function ϕ has a countable number of discontinuities, Detemple, Garcia and Rindisbacher (2003) obtained the following central limit theorems :

$$\text{For } \alpha = \frac{1}{2}, \quad \text{when } N^{1/5}\varepsilon \to 0 , \quad N^{2/5}\left(\hat{\beta}_N^{FD} - \beta\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0, \Sigma_\alpha) .$$

$$\text{For } \alpha = 0, 1 , \quad \text{when } N^{1/3}\varepsilon \to 0 , \quad N^{1/3}\left(\hat{\beta}_N^{FD} - \beta\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0, \Sigma_\alpha) .$$

In the general case $d \geq 1$, the finite differences estimators are defined componentwise, and therefore, the rate of convergence is not affected by the dimension d of the parameter λ^0.

The main objective of this paragraph is to provide an asymptotic comparison of the single-Kernel based estimator with the finite differences one.

The key point of our single-Kernel based estimators is that the derivation with respect to the parameter λ is reported on the density of $Z(\lambda)$ so that our asymptotic results do not involve the regularity of the pay-off function ϕ. For any pay-off function ϕ, we derived in Theorems 4.1 and 4.2 that our single-Kernel based estimators $\hat{\beta}_N$ and $\check{\beta}_N$ satisfy:

$$\text{when } N\,h^{d+2p+2} \to 0\,, \qquad \sqrt{Nh^{d+2}}\,\left(\hat{\beta}_N - \beta\right) \xrightarrow[N\to\infty]{law} \mathcal{N}\left(0,\Sigma\right)\,,$$

where p is the order of the Kernel function. Minimizing the Mean Square Error of the estimator, we obtained in section 4.3 an optimal h of order $N^{-1/(d+2p+2)}$ which, of course, almost satisfies the condition required in the convergence in distribution. Therefore, taking a bandwidth h of order $N^{-1/(d+2p+2)-2\delta/(d+2)}$ with $\delta > 0$ leads to a convergence in distribution at rate N^r with $r := p/(d+2p+2) - \delta$. Therefore, the single-Kernel based estimators, with Kernel of order $p > 2d+4$ and δ sufficiently small, achieves a convergence rate of order $r > 2/5$. Hence, they outperform all the finite differences estimators in the case of discontinuous payoffs.

Notice that, by taking Kernel functions of order p sufficiently large, we can obtain a convergence rate in distribution as close as desired to the parametric rate $\sqrt{N}$.

Remark 4.2 If we consider the optimized Kernel estimator $\hat{\beta}^*_n$, based on a uniform density ℓ with width $2h$, derived in section 4.4, we obtain a rate of convergence $\sqrt{Nh^2}$ under condition $Nh^{2p+2} \to 0$. Therefore, in order to outperform the finite differences estimators of a greek associated to a discontinuous payoff function ϕ, one just need to use a Kernel function of order $p > 4$.

5 Short time asymptotics

In this section, we are interested in the behaviour of the estimator $\hat{\beta}_N$ when $Z(\lambda)$ is a process of the form X_t^λ with a small time period t. We refer to Subsection 2.4 for some instances where this problem appears naturally. Since t is now an important variable, we shall emphasize more the dependence of V^ϕ on t by denoting

$$V^\phi(\lambda,t) \;:=\; \mathbb{E}\left[\phi\left(X_t^\lambda\right)\right]$$

The main objective of our analysis is to device an optimal choice of the number of simulations N and the bandwidth h for $\hat{\beta}_N$ given a short maturity t, i.e. given $t \longrightarrow 0$.

Since $\nabla_\lambda V^\phi(\lambda^0, t)$ converges to $\beta^0 := \nabla_\lambda \phi\left(X_0^{\lambda^0}\right)$, the present context requires further smoothness conditions on the function ϕ.

5.1 Parameterized stochastic differential equation

Let $\{X_u^\lambda, u \geq 0\}$ be a process with values in $\mathbb{R}^n$ defined by the stochastic differential equation

$$X_0^\lambda = x(\lambda), \qquad dX_u^\lambda = \mu(u, \lambda, X_u^\lambda)du + \sigma(u, \lambda, X_u^\lambda)dW_u, \tag{5.1}$$

where W is a Brownian motion with values in $\mathbb{R}^n$, and the functions x, μ and σ satisfy the following assumption:

Assumption SDE *The function $x(.)$ belongs to $\mathbf{C}^{p+2}(\mathbb{R}^d, \mathbb{R}^n)$, and the coefficients μ, σ are continuous with $\mu(u, ., .) \in \mathbf{C}_b^{p+2}(\mathbb{R}^d \times \mathbb{R}^n, \mathbb{R}^n)$ and $\sigma(u, ., .) \in \mathbf{C}_b^{p+2}(\mathbb{R}^d \times \mathbb{R}^n, \mathcal{M}_{\mathbb{R}}^n)$ for every $u \in \mathbb{R}_+$.*

Lemma 5.1 *Under Assumption* SDE, *the solution X^λ of the stochastic differential equation* (5.1) *is $p+2$ times differentiable in λ and each of the derivatives $\nabla_\lambda^i X_t^\lambda$ is locally (α, β)-Hölder continuous in (t, λ) for any $\alpha < 1$ and $\beta < \frac{1}{2}$. Furthermore, for any compact sets $\mathcal{K} \subset \mathbb{R}^k$ and $\mathcal{L} \subset \mathbb{R}$, we can find $M \in \mathbb{R}$ such that for any $\lambda_1, \lambda_2 \in \mathcal{K}$, $t_1, t_2 \in \mathcal{L}$:*

$$\mathbb{E}\left[\left|\nabla_\lambda^i X_{t_1}^{\lambda_1} - \nabla_\lambda^i X_{t_2}^{\lambda_2}\right|\right] \leq M\left(|\lambda_1 - \lambda_2| + |t_1 - t_2|^{\frac{1}{2}}\right), \tag{5.2}$$

and

$$\sup_{\lambda \in \mathcal{K}, t \in \mathcal{L}} \mathbb{E}\left[\left|\nabla_\lambda^i X_t^\lambda\right|^k\right] < \infty \quad \textit{for all} \quad k \in \mathbb{N}. \tag{5.3}$$

Proof. Define the maps $\bar{\mu}(u, x, \lambda) := (\mu(u, \lambda, x)', 0) \in \mathbb{R}^{n+d}$ and $\bar{\sigma}(u, x, \lambda) := (\sigma(u, \lambda, x)', 0) \in \mathcal{M}_{\mathbb{R}}^{n+d}$, and consider the process Y defined by the stochastic differential equation

$$Y_0 = y \quad \text{and} \quad dY_u = \bar{\mu}(u, Y_u)du + \bar{\sigma}(u, Y_u)dW_u,$$

so that the parameterized process X^λ coincides with the first n components of the process Y with initial condition $y = (x(\lambda), \lambda)$.

Under Assumption R3, the coefficients of the stochastic differential equation defining Y are in $\mathbf{C}_b^{p+2}(\mathbb{R}^d, \mathbb{R}^n)$. From Theorem 3.3 p 223 in Kunita (1984) , we conclude that the flow $Y_t(y)$ is $q+2$ times differentiable with respect to its initial value y, and every derivative $\nabla_y^k Y_t$ is locally (α, β)-Hölder continuous in (t, y) for any $\alpha < 1$ and $\beta < \frac{1}{2}$. Now, since the function x is smooth, this property is inherited by the process X^λ by composition.

We next turn to the proof of (5.2). It is shown in Kunita, proof of Theorem 3.3 p 223 in Kunita (1984), that, given two solutions Y_t^1 and Y_t^2 starting respectively at y_1 and y_2, there is a constant C such that, for any $p > 2$,

$$\mathbb{E}\left[|\nabla^k Y_t^1 - \nabla^k Y_s^2|^p\right] \leq C\left(|y_1 - y_2|^p + (1 + |y_1| + |y_2|)^p |s - t|^{\frac{p}{2}}\right) \quad \text{for} \quad s, t \geq 0 \quad (5.4)$$

Since the $\mathbb{L}^1$ norm is dominated by the $\mathbb{L}^p$ norm, and the function $x(.)$ is locally Lipschitz, this implies (5.2). Finally, (5.3) is a direct consequence of (5.4). □

5.2 The main results

Let

$$\mathcal{L}_t^\lambda g(x) := \mu(t, \lambda, x) \cdot \nabla g(x) + \frac{1}{2} \operatorname{Tr}\left[\sigma\sigma'(t, \lambda, x) \nabla^2 g(x)\right], \qquad g \in C^2(\mathbb{R}^n, \mathbb{R})$$

denote the infinitesimal generator of the process X^λ.

As in the previous section, we consider a sequence (Λ_i, X_t^i) of independent pairs of random variables where the distribution density of Λ_i is ℓ, and X_t^i is the solution of the stochastic differential equation (5.1) with parameter λ fixed to Λ_i. In view of the results of the previous section, we shall only consider the estimator $\hat{\beta}_N$ of the Greek defined by

$$\hat{\beta}_N = \frac{1}{\ell(\lambda^0) N h^{d+1}} \sum_{i=1}^N \phi(X_t^i) \left(\nabla K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) - h K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \frac{\nabla \ell}{\ell}(\Lambda_i)\right).$$

Theorem 5.1 *Let the Kernel function K be of order $p > 0$, and let assumption* SDE *hold. Assume further that the density function ℓ is in $C^{p+1}(\mathbb{R}^d, \mathbb{R})$ and ϕ is in $C^{p+3}(\mathbb{R}^d, \mathbb{R})$. If (4.4) holds and $t \to 0$, then the bias and the variance of $\hat{\beta}_N$ satisfy*

$$\mathbb{E}[\hat{\beta}_N] - \beta^0 \sim C_6 h^p + C_7 t, \quad \mathbb{E}[\hat{\beta}_N] - \nabla_\lambda V^\phi(\lambda^0, t) \sim C_6 h^p \quad \text{and} \quad \mathbb{V}\text{ar}[\hat{\beta}_N] \sim \frac{\Sigma_0}{N h^{d+2}},$$

where

$$\begin{aligned}
C_6 &:= \frac{(-1)^p}{p! \ell(\lambda^0)} \sum_{j_1, \ldots, j_p = 1}^d \nabla^p_{\lambda_{j_1}, \ldots, \lambda_{j_p}} [\ell(\lambda^0) \nabla(\phi \circ x)(\lambda^0)] \int l_{j_1} \ldots l_{j_p} K(l)\, dl, \\
C_7 &:= \nabla_\lambda \left[\mathcal{L}_0^{\lambda^0} \phi\left(x(\lambda^0)\right)\right], \\
\Sigma_0 &:= \frac{1}{\ell(\lambda^0)} \phi^2(x(\lambda^0)) \int \nabla K(l) \nabla K(l)' dl,
\end{aligned}$$

and the asymptotic distribution of $\hat{\beta}_N$ is given by

$$\sqrt{N h^{d+2}} \left(\hat{\beta}_N - \mathbb{E}\left[\hat{\beta}_N\right]\right) \xrightarrow[N \to \infty]{law} \mathcal{N}(0, \Sigma_0). \qquad (5.5)$$

If, in addition, (4.5) *holds, we get :*

$$\sqrt{Nh^{d+2}}\left(\hat{\beta}_N - \nabla_\lambda V^\phi(\lambda^0,t)\right) \quad \underset{N\to\infty}{\overset{law}{\longrightarrow}} \quad \mathcal{N}(0,\Sigma_0)\ . \tag{5.6}$$

And the addition of condition $Nh^{d+2}t^2 \longrightarrow 0$ *gives us :*

$$\sqrt{Nh^{d+2}}\left(\hat{\beta}_N - \beta^0\right) \quad \underset{N\to\infty}{\overset{law}{\longrightarrow}} \quad \mathcal{N}(0,\Sigma_0)\ . \tag{5.7}$$

Before proceeding to the proof of this result, let us comment on the optimal choice of N and h given a short time t. Since, we are trying to estimate $\nabla_\lambda V^\phi(\lambda^0,t)$ using $\hat{\beta}_N$, we have to minimize

$$\mathbb{E}\left[\left|\hat{\beta}_N - \nabla_\lambda V^\phi(\lambda^0,t)\right|^2\right] \ \sim\ \frac{Tr(\Sigma_0)}{Nh^{d+2}} + |C_6|^2 h^{2p}\,.$$

Then, as in the fixed time study, the optimal bandwidth h^* is given by :

$$h^* \ = \ \left(\frac{(d+2)Tr(\Sigma^0)}{2p|C_6|^2 N}\right)^{1/(d+2p+2)}\ . \tag{5.8}$$

Indeed, theorem 5.1 says that, considering a process X evaluated at a short time t, the asymptotic equivalents of the bias and of the variance are obtained by sending t to zero in the expressions of the fixed maturity case. From a practical point of view, the interest is that the constants C_6 and Σ^0 do not depend on time t and are much easier to evaluate than the corresponding Σ and C_1.

Proof of Theorem 5.1 We split the proof in three steps.

1. We first study the bias term. Using the same technique as in the proof of Proposition 4.1, we obtain

$$\begin{aligned}
\mathbb{E}\left[\hat{\beta}_N\right] &= \frac{1}{\ell(\lambda^0)h^{d+1}}\mathbb{E}\left[\phi\left(X_t^\Lambda\right)\left(\nabla K\left(\frac{\lambda^0-\Lambda}{h}\right) - hK\left(\frac{\lambda^0-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\Lambda)\right)\right] \\
&= \frac{1}{\ell(\lambda^0)h^{d+1}}\mathbb{E}\left[V^\phi(\Lambda,t)\left(\nabla K\left(\frac{\lambda^0-\Lambda}{h}\right) - hK\left(\frac{\lambda^0-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\Lambda)\right)\right] \\
&= \frac{(-1)}{\ell(\lambda^0)h^d}\int V^\phi(\lambda,t)\nabla\left(K\left(\frac{\lambda^0-\lambda}{h}\right)\ell(\lambda)\right)d\lambda \\
&= \frac{1}{\ell(\lambda^0)h^d}\int \nabla_\lambda V^\phi(\lambda,t)K\left(\frac{\lambda^0-\lambda}{h}\right)\ell(\lambda)\,d\lambda \\
&= \frac{1}{\ell(\lambda^0)}\int \nabla_\lambda V^\phi(\lambda^0-hl,t)\ell(\lambda^0-hl)K(l)\,dl\,.
\end{aligned} \tag{5.9}$$

We will use the latter expression in order to derive an expansion of the bias with respect to the pair (h,t) near the origin. Before this, let us derive a suitable representation of $\nabla_\lambda V^\phi(\lambda^0-hl,t)$. Since $\nabla\phi$ and σ are bounded, it follows from Itô's lemma that :

$$V^\phi(\lambda^0-hl,t) \ = \ \phi\left(x(\lambda^0-hl)\right) + \mathbb{E}\left[\int_0^t \mathcal{L}_s^{\lambda^0-hl}\phi\left(X_s^{\lambda^0-hl}\right)ds\right]\ .$$

By (5.3) of Lemma 5.1, the above expression is differentiable with respect to λ^0, and :

$$\nabla_\lambda V^\phi(\lambda^0 - hl, t) \;=\; \nabla_\lambda(\phi \circ x)(\lambda^0 - hl) + \mathbb{E}\left[\int_0^t \nabla_\lambda[\mathcal{L}_s^{\lambda^0 - hl}\phi\left(X_s^{\lambda^0 - hl}\right)]ds\right].$$

In view of this, we decompose the expression of $\mathbb{E}[\hat{\beta}_N] - \beta^0$ given by expression (5.9) into three pieces :

$$\mathbb{E}[\hat{\beta}_N] - \beta^0 \;=\; A + B + C\,,$$

(i) The first term is given by

$$A \;:=\; \frac{1}{\ell(\lambda^0)}\int \nabla_\lambda(\phi \circ x)(\lambda^0 - hl)\ell(\lambda^0 - hl)K(l)dl \,-\, \beta^0 \;\sim\; C_6\, h^p$$

where C_6 is defined in the statement of the theorem, and the latter equivalence follows from the fact that p is the order of K by the same argument as in Proposition 4.1 .

(ii) The second term is given by

$$\begin{aligned} B \;&:=\; \mathbb{E}\left[\int_0^t \nabla_\lambda[\mathcal{L}_s^{\lambda^0}\phi\left(X_s^{\lambda^0}\right)]ds\right] \\ &\sim\; t\,\nabla_\lambda[\mathcal{L}_0^{\lambda^0}\phi\left(x(\lambda^0)\right)] \;=\; C_7\, t, \end{aligned}$$

as a consequence of the a.s. continuity at the origin of the map $s \longmapsto \nabla_\lambda \mathcal{L}_s\phi(X_s^{\lambda^0})$, and the dominated convergence theorem together with (5.2) of Lemma 5.1.

(iii) We now show that the remaining term

$$\begin{aligned} C \;&:=\; \frac{1}{\ell(\lambda^0)}\int \mathbb{E}\left[\int_0^t \nabla_\lambda[\mathcal{L}_s^{\lambda^0 - hl}\phi]\left(X_s^{\lambda^0 - hl}\right)ds\right]\ell(\lambda^0 - hl)K(l)dl - \mathbb{E}\left[\int_0^t \nabla_\lambda[\mathcal{L}_s^{\lambda^0}\phi]\left(X_s^{\lambda^0}\right)ds\right] \\ &=\; \frac{1}{\ell(\lambda^0)}\int \mathbb{E}\left[\int_0^t \left(\nabla_\lambda[\mathcal{L}_s^{\lambda^0 - hl}\phi]\left(X_s^{\lambda^0 - hl}\right)\ell(\lambda^0 - hl) - \nabla_\lambda[\mathcal{L}_s^{\lambda^0}\phi\left(X_s^{\lambda^0}\right)]\ell(\lambda^0)\right)ds\right]K(l)dl, \end{aligned}$$

is dominated by A and B. To see this, observe that the first $p-1$ terms of the order p Taylor expansion of the integrand disappear, by the fact that p is the order of the Kernel K. Using (5.3) of Lemma 5.1 and the regularity of the derivatives of μ, σ and ϕ, the expectation of the remainder term in the expansion can be bounded uniformly in s and l. Therefore, $|C| = O(th^p)$ and C is negligible with respect to A and B.

Thus $E\left[\hat{\beta}_N\right] - \beta^0 \sim C_6 h^p + C_7 t$ as announced in the statement of theorem. And, noticing simply that $\mathbb{E}[\hat{\beta}_N] - \nabla_\lambda V^\phi(\lambda^0, t) = A + C$, we get the second announced result $\mathbb{E}[\hat{\beta}_N] - \nabla_\lambda V^\phi(\lambda^0, t) \sim C_6 h^p$.

2. We now compute the variance of $\hat{\beta}_N$. As in Proposition 4.1, we can rewrite :

$$\mathrm{Var}\left[\phi\left(X_t^\Lambda\right)\left(\nabla K\left(\frac{\lambda^0 - \Lambda}{h}\right) - hK\left(\frac{\lambda^0 - \Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\Lambda)\right)\right] = v_1 - v_2^{\otimes}\,,$$

where

$$v_2 \;:=\; \mathbb{E}\left[\phi(X_t^{\Lambda})\left(\nabla K\left(\frac{\lambda^0-\Lambda}{h}\right)-hK\left(\frac{\lambda^0-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\Lambda)\right)\right] \;=\; \mathrm{O}\left(h^{d+1}\right),$$

and,

$$\begin{aligned} v_1 \;&:=\; \mathbb{E}\left[\phi^2\left(X_t^{\Lambda}\right)\left(\nabla K\left(\frac{\lambda^0-\Lambda}{h}\right)-hK\left(\frac{\lambda^0-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\Lambda)\right)^{\otimes}\right] \\ &=\; \mathbb{E}\left[V^{\phi^2}(\Lambda,t)\left(\nabla K\left(\frac{\lambda^0-\Lambda}{h}\right)-hK\left(\frac{\lambda^0-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\Lambda)\right)^{\otimes}\right] \\ &=\; h^d\int V^{\phi^2}(\lambda^0-hl,t)\left(\nabla K(l)-hK(l)\frac{\nabla\ell}{\ell}(\lambda^0-hl)\right)^{\otimes}\ell(\lambda^0-hl)\,dl\,. \end{aligned}$$

Now observe that the following equivalence

$$\left(\nabla K(l)-hK(l)\frac{\nabla\ell}{\ell}(\lambda^0-hl)\right)^{\otimes}\ell(\lambda^0-hl) \;\sim\; \nabla K(l)^{\otimes}\ell(\lambda^0)+M_1.lh+hK(l)\frac{\nabla\ell}{\ell}(\lambda^0),$$

holds uniformly in l for some constant M_1. Also, from the first step of this proof, we have

$$V^{\phi^2}(\lambda^0-hl,t) \;=\; \phi^2\left(x(\lambda^0)\right)+O(t),$$

uniformly w.r.t. l in a compact subset. Then

$$v_1 \;\sim\; h^d\ell(\lambda^0)\phi^2(x(\lambda^0))\int\nabla K^{\otimes}.$$

Hence $v_2^{\otimes}$ is dominated by v_1, and we get the expression of the variance reported in the statement of the theorem follows from the last equivalence.

3. We now turn to the derivation of the asymptotic distribution of $\tilde{\beta}_N$. The proof is again similar to that of Theorem 4.1, and consists in verifying the Lyapounov conditions (Billingsley, 1968, p.44). Let a be a d-dimensional vector and let us define, for every $i=1,\ldots,N$,

$$\begin{aligned} U_i^N \;&:=\; \frac{1}{Nh^{d+1}\ell(\lambda^0)}\phi\left(X_t^i\right)\left(\nabla K\left(\frac{\lambda^0-\Lambda_i}{h}\right)-hK\left(\frac{\lambda^0-\Lambda_i}{h}\right)\frac{\nabla\ell}{\ell}(\Lambda_i)\right), \\ V_i^N \;&:=\; a'U_i^N-\mathbb{E}\left[a'U_i^N\right]. \end{aligned}$$

It is sufficient to show that, for some $\delta>2$, we have

$$\sup_N\frac{1}{\sigma_N^{\delta}}\sum_{i=1}^N\mathbb{E}\left[|V_i^N|^{\delta}\right] \;<\; \infty \quad\text{where}\quad \sigma_N^2 \;:=\; \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^N V_i^N\right]. \tag{5.10}$$

To check this, we directly estimate by the Minkowski inequality that

$$\begin{aligned}
\left\|V_i^N\right\|_\delta &\leq \frac{\sum_{i=1}^d \left\|a_i\phi\left(X_t^\Lambda\right)\left(\nabla_i K\left(\frac{\lambda^0-\Lambda}{h}\right) - hK\left(\frac{\lambda^0-\Lambda}{h}\right)\frac{\nabla_i \ell}{\ell}(\Lambda)\right)\right\|_\delta}{Nh^{d+1}\ell(\lambda^0)} + O\left(\frac{1}{N}\right) \\
&\leq \frac{|a|_\infty \sum_{i=1}^d \left\|\{V^{\phi^\delta}(\Lambda,t)\}^{1/\delta}\left(\nabla_i K\left(\frac{\lambda^0-\lambda}{h}\right) - hK\left(\frac{\lambda^0-\lambda}{h}\right)\frac{\nabla_i \ell}{\ell}(\lambda)\right)\right\|_\delta}{Nh^{d+1}\ell(\lambda^0)} + O\left(\frac{1}{N}\right) \\
&\leq Const\left(\frac{h^{d/\delta}}{Nh^{d+1}} + \frac{1}{N}\right),
\end{aligned}$$

by the usual change of variable and Taylor expansion. On the other hand, it follows from the equivalent of the variance derived in the previous step of this proof that

$$\sigma_N^2 \sim \frac{\phi^2\left(x(\lambda^0)\right)}{Nh^{d+2}\ell(\lambda^0)}\int a'\nabla K(t)\nabla K(t)'a\,.$$

The two last estimates imply that Condition (5.10) holds under Condition (4.4). Therefore, $\sum_{i=1}^N V_i^N$ is asymptotically gaussian for any $a \in \mathbb{R}^d$, and the Cramer-Wold device concludes the proof. □

6 Numerical results

In this section, we present some numerical results obtained in the Black-Scholes model. We consider an underlying asset whose dynamics is given by:

$$S_t^x = x + r\int_0^t S_u^x\,du + \sigma\int_0^t S_u^x\,dW_u\,,$$

and we focus on the estimation of the so-called Delta :

$$\beta = \nabla_x \mathbb{E}[\phi(Z^x)]\,,$$

where $Z^x = S_T^x$ for an European option and $Z^x = \int_0^T S_t^x dt$ for an Asian option, for instance. We naturally denote by $f(x,.)$ the density of Z^x.

We simulate observations X_i with a uniform density ℓ centered at $S_0 = x$. We use the optimal uniform density found in subsection 4.4, i.e. a uniform density with bandwidth $2h$, where h is the bandwidth of the Kernel. The single-kernel based estimator $\hat{\beta}_N^*$ is therefore given by (4.15).

6.1 Computation of the optimal bandwidth

The choice of the bandwidth is crucial in Kernel estimation. The asymptotic results of Section 4 provide the expression of the asymptotic optimal bandwidth

$$h^* = \left(\frac{\Sigma_u}{pC_u^2 N}\right)^{\frac{1}{2p+2}},$$

where

$$C_u = \frac{(-1)^p}{p!}\left(\int u^p K(u)du\right)\mathbb{E}\left[\phi(Z^x)\frac{\nabla_x^{p+1} f(x,Z^x)}{f(x,Z^x)}\right] \quad \text{and} \quad \Sigma_u = 2\mathbb{E}[\phi^2(Z^x)]\int(\nabla K)^2 .$$

Given a Kernel function K, the coefficient Σ_u can be estimated by a straightforward Monte Carlo procedure. In this subsection, we focus on the estimation of the parameter

$$\mathcal{E} \;:=\; \mathbb{E}\left[\phi(Z^x)\frac{\nabla_x^{p+1} f(x,Z^x)}{f(x,Z^x)}\right],$$

in order to deduce an approximation of the coefficient C_u.

6.1.1 The Gaussian distribution case

Let $Z^x = S_T^x = xe^Y$, where Y has a normal distribution with mean $m := (r - \frac{\sigma^2}{2})T$ and variance $\Sigma := \sigma^2 T$. Then, it is easily checked that :

$$\nabla_x^p f(x,z) \;=\; \left[\sum_{i=0}^{p} a_i^p\, d(x,z)^i\right]\frac{f(x,z)}{x^p} \quad \text{where} \quad d(x,z) \;:=\; \frac{\ln z - \ln x - m}{\Sigma}$$

and the coefficients $(a_i^j)_{(i,j)\in\{0,\dots,p\}^2}$ are given by

$$a_i^0 = 1_{\{i=0\}}, \qquad a_i^{j+1} \;=\; a_{i-1}^j \;-\; j\,a_i^j \;-\; \frac{i+1}{\Sigma}\,a_{i+1}^j,$$

with the convention $a_i^j = 0$ for $i < 0$ and $i > p$. Hence :

$$\mathcal{E} \;=\; \frac{1}{x^p}\mathbb{E}\left[\phi(Z^x)\left(\sum_{i=0}^{p} a_i^p\, d(x,Z^x)^i\right)\right],$$

and this parameter can be estimated by a straightforward Monte Carlo procedure.

6.1.2 An empirical rule for the general case

In practice, the distribution function is unknown, and the calculation of the previous paragraph can not be used to estimate $\mathcal{E}$. We suggest to mimic the same principle as the usual Silverman's rule-of-thumb in Kernel estimation (see Scott (1992) e.g.).

Assume we have some estimates of the mean $\hat{m}$ and the variance $\hat{\Sigma}$ of the random variable $\ln(Z^x/x)$. We then approximate the coefficient $\mathcal{E}$ by

$$\hat{\mathcal{E}} \;=\; \frac{1}{x^p}\mathbb{E}\left[\phi(Z^x)\left(\sum_{i=0}^{p} \hat{a}_i^p\, \hat{d}(x,Z^x)^i\right)\right],$$

where $\hat{a}_i^p$ and $\hat{\Sigma}$ are computed as in the previous paragraph, substituting $\hat{m}$ and $\hat{\Sigma}$ to m and Σ.

6.2 Numerical comparison of the estimators

We present here numerical results obtained for the estimation of the delta of an European and an Asian at-the-money digital calls, i.e. with a payoff of the form $\phi(s) = 1_{s>K}$. Since this payoff function is discontinuous, the results of Section 4.5 show that the single-Kernel based estimator achieves a better rate of convergence than the finite differences estimators, whenever the Kernel has order $p > 4$. The main object of this section is to verify the empirical validity of these asymptotic results.

Our numerical experiments are performed with the following values of the parameters :

$$S_0 = 120, \quad r = 0, \quad \sigma = 0.2, \quad T = 1, \quad \text{and} \quad K = 120.$$

We use the following polynomial Kernel functions of order 2, 4 and 6, respectively, with support on $[-1, 1]$:

$$\begin{aligned} K_2(u) &= \frac{3}{4}(1-u^2)\,, \\ K_4(u) &= \frac{15}{32}(1-u^2)(3-7u^2)\,, \\ K_6(u) &= \frac{105}{256}(1-u^2)(33u^4-30u^2+5)\,. \end{aligned}$$

The numerical procedure is divided in two steps: first, we estimate the optimal theoretical bandwidth by Monte Carlo procedures with few simulations, as described in the previous subsection. Then we estimate the delta of the option by means of a single-Kernel based estimator with the estimated bandwidth.

European Digital Call

In the context of the Black-Scholes model, it was observed by Fournié et al. (2000) that the optimal weight for European options can be obtained by means of the Malliavin integration by part formula (see Section 2.3). Therefore, we are not hoping to compete with the the Malliavin-based Monte Carlo estimator.

The distribution of the different estimators based on $N = 10^6$ simulations are reported in Figure 1. In order to isolate the performance of the estimators, we use no variance reduction techniques. The good performance of the Malliavin estimator is confirmed by our numerical experiments. However, we observe surprisingly that the three Kernel based estimators are less accurate than the centered finite differences one. Recall that the asymptotic results of Section 4.5 show that the Kernel of order 6 performs asymptotically better than the finite differences one. We find the explanation of these empirical facts by looking at the constant terms C_u and Σ_u appearing in the asymptotic decomposition of the bias and the variance of the Kernel-based estimators $\hat{\beta}^*_N$. First, these constants are such that the

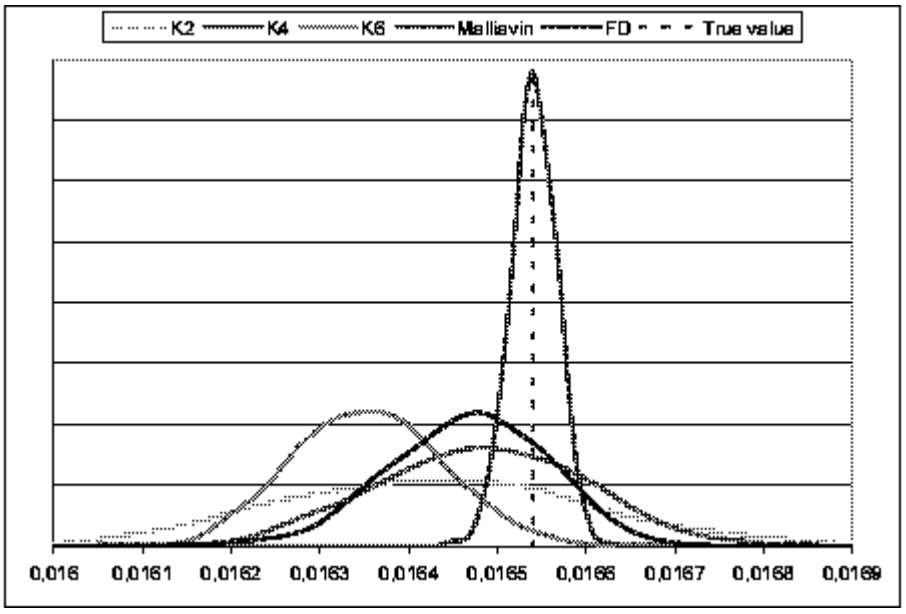

Figure 1: Delta of an European Digital Call, N = 1 Million

Mean Square Error of the Kernel estimator of order 4 is ten times the MSE of the finite differences estimator. This explains why the Kernel estimator of order 4 can not beat the finite differences estimator since they achieve the same convergence rate. Secondly, we observe that, when the order of the Kernel increases, C_u decreases and Σ_u increases. Therefore the variance of the estimators decreases slowly and the optimal bandwidth h^* increases so that the hypothesis that h goes to zero becomes less valid. This explains why the theoretical asymptotic bias differs from the empirical one and, more precisely in our experiment, why the estimator based on the Kernel of order 6 is much more biased than the finite differences one. To overcome this difficulty, one should simply increase the number of simulations. Figure 2 shows the distribution of the finite differences estimator and the Kernel based estimator of order 6 based on $N = 10^9$ simulations. In this case, we observe that the Kernel based estimator of order 6 outperforms the finite differences one. This confirms the theoretical asymptotic results obtained in section 4.5.

Asian Digital Call Option

In this paragraph, we investigate the case of an Asian option, where the Malliavin integration by part formula does not lead to the optimal weight, see Section 2.3. The distribution of the different estimators based on $N = 10^6$ simulations are reported in Figure 3. Even if the Malliavin weight is not optimal, the Malliavin estimator still outperforms the other estimators. As for the European digital call, the finite differences estimator outperforms

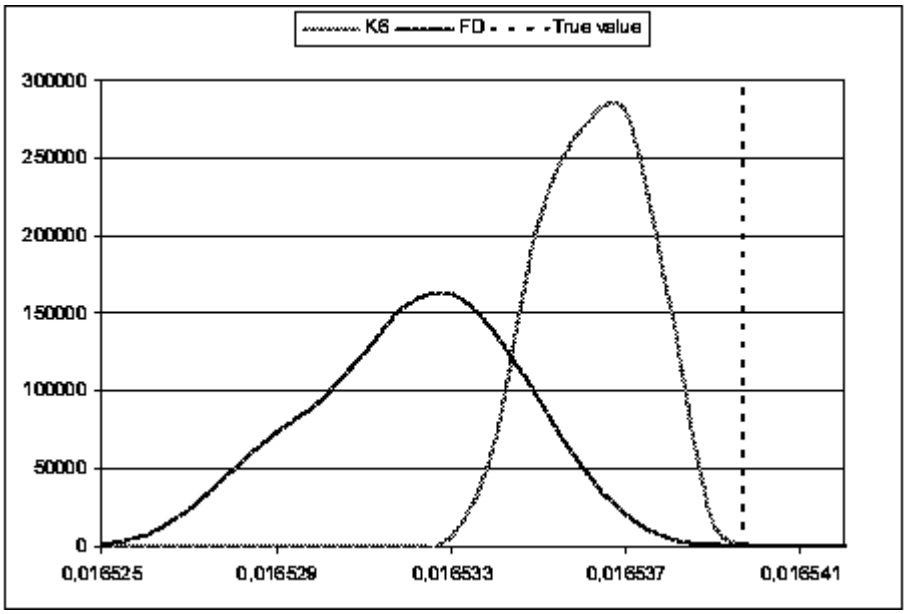

Figure 2: Delta of an European Digital Call, N = 1 Billion

the Kernel based estimators but one simply requires more simulations in order to make the kernel estimator of order 6 more efficient than the finite differences one.

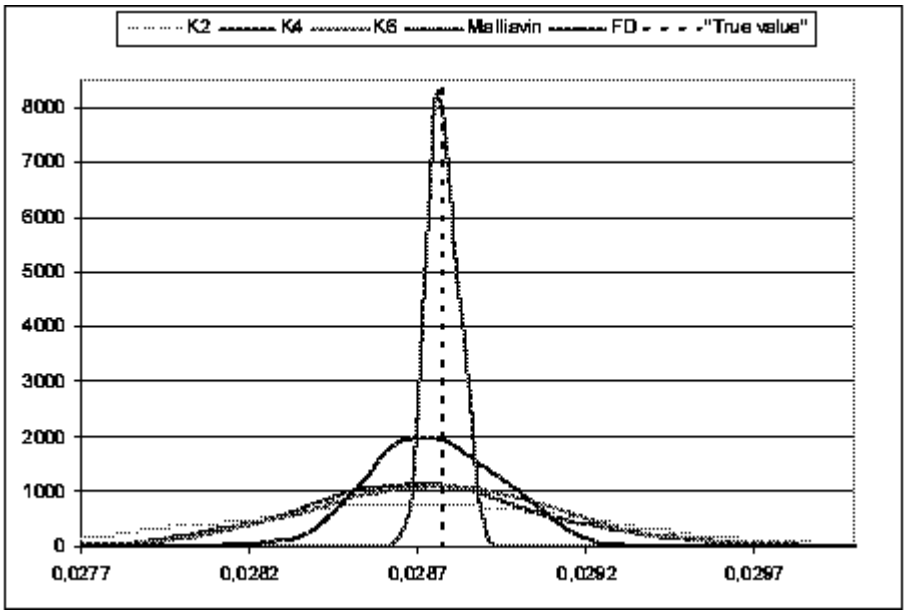

Figure 3: Delta of an Asian Digital Call, N = 1 Million

Conclusion

Our Kernel based estimator with order $p > 4$ of the delta of a digital option outperforms asymptotically the finite differences one, but one requires a large number of simulation to verify this fact empirically. When the density of the underlying is unknown and the pay-off function is irregular, the Malliavin based estimator is still more efficient than the others. Nevertheless, in general, Malliavin weigths are difficult to derive although the other estimators are very easy to construct.

7 APPENDIX : Proof of Proposition 4.3 and Theorem 4.3

7.1 Preliminaries

Recall that

$$\tilde{\beta}_N \;:=\; \frac{1}{\ell(\lambda^0)\,Nh^d}\sum_{i=1}^{N}\phi(Z_i)\,\hat{s}_N^{-i}(\Lambda_i,Z_i)\,K\left(\frac{\lambda^0-\Lambda_i}{h}\right), \tag{7.1}$$

where

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

with

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

For every λ, z, we set

$$\bar{\varphi}(\lambda,z) \;:=\; E[\hat{\varphi}^{-1}(\lambda,z)] = \int K(l)H(v)\varphi(\lambda-hl,z-hv)\,dt\,dv,$$

and its derivative is given by

$$\bar{\varphi}_\lambda(\lambda,z) \;=\; h^{-1}\int\nabla K(l)H(v)\varphi(\lambda-hl,z-hv)\,dt\,dv$$

Arguing as in the proof of Proposition 4.1, we next compute that

$$\bar{\varphi}(\lambda,z)-\varphi(\lambda,z) \;=\; \xi_K^p[\varphi](\lambda,z)\,h^p \;+\; \xi_H^q[\varphi](\lambda,z)\,h^q \;+\; o(h^{p\wedge q}). \tag{7.2}$$

Similarly, we get

$$\bar{\varphi}_\lambda(\lambda,z)-\varphi_\lambda(\lambda,z) \;=\; \xi_K^p[\varphi_\lambda](\lambda,z)\,h^p \;+\; \xi_H^q[\varphi_\lambda](\lambda,z)\,h^q \;+\; o(h^{p\wedge q}). \tag{7.3}$$

Remark 7.1 Since ϕ and K have compact support by Assumption S, it follows that, for sufficiently small h, the sum in (7.1) is restricted to pairs (Λ_i, Z_i) with values in $\mathcal{V}(\lambda^0) \times C_\phi$ where $\mathcal{V}(\lambda^0)$ is a neighborhood of λ^0, and C_ϕ is a compact subset of $\mathbb{R}^n$ such that $\mathrm{Supp}\,\phi \subset C_\phi$.

For any function ψ defined on $\mathcal{V}(\lambda^0) \times C_\phi$, set

$$\|\psi\|_\infty := \sup_{(\lambda,z)\in\mathcal{V}(\lambda^0)\times C_\phi} |\psi(\lambda, z)|.$$

Remark 7.2 By Assumptions R2 and R3, since (λ, z) vary in a compact subset of $\mathbb{R}^d \times \mathbb{R}^n$, the remainder terms in (7.2) and (7.3) 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 :

$$\|\bar{\varphi} - \varphi\|_\infty = O\left(h^{p\wedge q}\right) \quad \textit{and} \quad \|\bar{\varphi}_\lambda - \varphi_\lambda\|_\infty = O\left(h^{p\wedge q}\right). \tag{7.4}$$

From Assumption K, ∇K has bounded variation and we report the following result from Pollard (1984), Example 2.38 and Exercise 2.27 : almost everywhere,

$$\sup_{1\le i\le N} \left\|\hat{\varphi}^{-i} - \bar{\varphi}\right\|_\infty = O\left(\frac{\ln N}{\sqrt{Nh^{d+n}}}\right) \text{ and}$$
$$\sup_{1\le i\le N} \left\|\hat{\varphi}_\lambda{}^{-i} - \bar{\varphi}_\lambda\right\|_\infty = O\left(\frac{\ln N}{h\sqrt{Nh^{d+n}}}\right). \tag{7.5}$$

Notice that this is the only place where we need ∇K to have bounded variation.

Since (λ, z) vary in a compact subset, assumption R3 implies that

$$\|\varphi\|_\infty < \infty, \quad \text{and} \quad \|\varphi_\lambda\|_\infty < \infty. \tag{7.6}$$

It then follows from equations (7.2), (7.3) and (7.5) that a.e.

$$\|\bar{\varphi}\|_\infty + \|\bar{\varphi}_\lambda\|_\infty + \sup_{1\le i\le N} \left\|\hat{\varphi}^{-i}\right\|_\infty + \sup_{1\le i\le N} \left\|\hat{\varphi}_\lambda{}^{-i}\right\|_\infty < \infty. \tag{7.7}$$

By Assumption S, we also have

$$\|1/\varphi\|_\infty < \infty \quad \text{and} \quad \|1/\varphi_\lambda\|_\infty < \infty, \tag{7.8}$$

so that, by (7.2), (7.3) and (7.5), for h small enough and N large enough, we have a.s.

$$\|1/\bar{\varphi}\|_\infty + \|1/\bar{\varphi}_\lambda\|_\infty + \sup_{1\le i\le N} \left\|1/\hat{\varphi}^{-i}\right\|_\infty + \sup_{1\le i\le N} \left\|1/\hat{\varphi}^{-i}\right\|_\infty < \infty. \tag{7.9}$$

7.2 A suitable decomposition of $\bar{\beta}_N$

We define the following functions $t_N^1, \ldots, t_N^9$ on $\mathbb{R}^d \times \mathbb{R}^n$:

$$t_N^1 := s, \qquad t_N^2 := \frac{\bar{\varphi}\lambda - \varphi\lambda}{\varphi}, \quad t_N^3 := \frac{(\varphi - \bar{\varphi})\,\varphi\lambda}{\varphi^2}, \quad t_N^4 := \frac{(\varphi - \bar{\varphi})\,(\bar{\varphi}\lambda\,\varphi - \bar{\varphi}\,\varphi\lambda)}{\varphi^2\,\bar{\varphi}},$$

$$t_N^5 := \frac{\hat{\varphi}\lambda^{-i} - \bar{\varphi}\lambda}{\varphi}, \quad t_N^6 := \frac{(\bar{\varphi} - \hat{\varphi}^{-i})\,\bar{\varphi}\lambda}{\bar{\varphi}^2}, \quad t_N^7 := \frac{(\hat{\varphi}\lambda^{-i} - \bar{\varphi}\lambda)\,(\varphi - \bar{\varphi})}{\varphi\,\bar{\varphi}},$$

$$t_N^8 := \frac{(\bar{\varphi} - \hat{\varphi}^{-i})(\hat{\varphi}\lambda^{-i} - \bar{\varphi}\lambda)}{\hat{\varphi}^{-i}\,\bar{\varphi}} \quad \text{and} \quad t_N^9 := \frac{(\bar{\varphi} - \hat{\varphi}^{-i})^2\bar{\varphi}\lambda}{\hat{\varphi}^{-i}\,\bar{\varphi}^2},$$

$$\text{so that} \qquad \hat{s}_N^{-i}(\Lambda_i, Z_i) = \sum_{j=1}^{9} t_N^j(\Lambda_i, Z_i)\,.$$

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

$$\tilde{\beta}_N = \sum_{j=1}^{9} T_N^j, \quad \text{where} \quad T_N^j := \frac{1}{\ell(\lambda^0)\,Nh^d} \sum_{i=1}^{N} \phi(Z_i)\; t_N^j(\Lambda_i, Z_i)\; K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \tag{7.10}$$

for every $j = 1, \ldots, 9$. By (7.6), (7.7), (7.8), and (7.9), we observe that

$$\left\| t_N^j \right\|_\infty < \infty \quad \text{for all} \quad j = 1, \ldots, 9\,.$$

Throughout this proof, C is a constant whose value may change from line to line.

Lemma 7.1 *For any $j = 1, \ldots, 9$, we have* $\mathbb{E}\left[T_N^j\right] = \mathrm{O}\left(\left\| t_N^j \right\|_\infty\right)$.

Proof. The result is derived from the following inequality:

$$\begin{aligned}
\left|\mathbb{E}[T_N^j]\right| &\leq \frac{1}{\ell(\lambda^0)\,h^d}\left|\mathbb{E}\left[\phi(Z_1)\; t_N^j(\Lambda_1, Z_1)\; K\left(\frac{\lambda^0 - \Lambda_1}{h}\right)\right]\right| \\
&\leq \frac{1}{\ell(\lambda^0)}\left|\int \phi(z)\; t_N^j(\lambda^0 - hl, z)\; K(l)\, dl\, dv\right| \\
&\leq C\, \|t_N^j\|_\infty\,.
\end{aligned}$$

Lemma 7.2 *For every $j = 1 \ldots 4$,* $\mathbb{V}\mathrm{ar}(T_N^j) = \mathrm{O}\left(N^{-1}h^{-d}\| t_N^j \|_\infty^2\right)$.

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

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[T_N^j] &= \frac{1}{\ell(\lambda^0)^2\,Nh^{2d}}\mathbb{V}\mathrm{ar}\left[\phi(Z_1)\; t_N^j(\Lambda_1, Z_1)\; K\left(\frac{\lambda^0 - \Lambda_1}{h}\right)\right] \\
&\leq \frac{1}{\ell(\lambda^0)^2\,Nh^{2d}}\mathbb{E}\left[\phi^2(Z_1)\; t_N^j(\Lambda_1, Z_1)^2\; K^2\left(\frac{\lambda^0 - \Lambda_1}{h}\right)\right] \\
&\leq \frac{\|t_N^j\|_\infty^2}{\ell(\lambda^0)^2\,Nh^d}\int \phi^2(z)\; K^2(l)\, dl\, dv\,.
\end{aligned}$$

□

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

Lemma 7.3 $\mathbb{E}[T_N^5] = 0$ *and* $\mathbb{V}\mathrm{ar}(T_N^5) \sim \tilde{\Sigma}/(Nh^{d+2})$ *where*

$$\tilde{\Sigma} := \frac{1}{\ell(\lambda^0)} \int \phi^2(z) \left\{ \int K(l_2 - l_1) \nabla K(l_1)\, dl_1 \right\}^{\otimes} f(\lambda^0, z)\, dl_2\, dz\,.$$

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 re-written in

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

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

$$\bar{\varphi}_\lambda(\Lambda_i, Z_i) = \frac{1}{h^{d+n+1}} \mathbb{E}\left[\nabla_\lambda K\left(\frac{\Lambda_i - \Lambda_j}{h}\right) H\left(\frac{Z_i - Z_j}{h}\right) \mid \Lambda_i, Z_i \right].$$

Therefore, $\mathbb{E}[\mathcal{T}_{ij}] = 0$ for any $i = 1 \dots N$ and $j = 1 \dots N$ with $i \neq j$, leading to $\mathbb{E}[T_N^5] = 0$.

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

$$T_N^5 = T_N^{5,1} + T_N^{5,2}\,, \tag{7.11}$$

where

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

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

and

$$b(\lambda, z) := E\left[\mathcal{T}_{12} | \Lambda_2 = \lambda, Z_2 = z\right],$$

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

$$\mathbb{E}[\Upsilon_{ij} | \Lambda_i, Z_i] = \mathbb{E}[\Upsilon_{ij} | \Lambda_j, Z_j] = 0\,. \tag{7.12}$$

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

$$\mathbb{V}\mathrm{ar}[T_N^{5,1}] = \frac{h^{-4d-2n-2}}{\ell(\lambda^0)^2\, N^2(N-1)^2} \sum_{i<j} \mathbb{E}[\Upsilon_{ij}\Upsilon'_{ij}] = \frac{h^{-4d-2n-2}}{2\ell(\lambda^0)^2\, N(N-1)} \mathbb{E}[\Upsilon_{12}\Upsilon'_{12}].$$

By (7.12), we compute :

$$\mathbb{E}[\Upsilon_{12}\,\Upsilon'_{12}] \;=\; 2\mathbb{E}[\mathcal{T}_{12}\mathcal{T}'_{12}] + 2\mathbb{E}[\mathcal{T}_{12}\mathcal{T}'_{21}] - 2\mathbb{E}[b^2(\Lambda_1, Z_1)].$$

We next estimate that :

$$\begin{aligned} &\left|\mathbb{E}[\mathcal{T}_{12}\mathcal{T}'_{12}]\right| \\ \leq\; & \mathbb{E}\left[\frac{\phi^2(Z_1)}{\varphi^2(\Lambda_1, Z_1)}\, K^2\left(\frac{\lambda^0-\Lambda_1}{h}\right) |\nabla_\lambda K|^2 \left(\frac{\Lambda_1-\Lambda_2}{h}\right) H^2\left(\frac{Z_1-Z_2}{h}\right)\right] \\ & + h^{2d+n} \int \phi^2(z)\; K^2(l_1)|\nabla_\lambda K|^2(l_2) H^2(v)\, \frac{\varphi(\lambda_0 - hl_1 - hl_2, z - hv)}{\varphi(\lambda_0 - hl_1, z)}\, dl_1\, dl_2\, dz\, dv\,, \end{aligned}$$

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 one is a $O(h^{3d+2n+2})$ by (7.7). Similarly, $\mathbb{E}[\mathcal{T}_{12}\mathcal{T}'_{21}] = O(h^{2d+n})$. Moreover, $\mathbb{E}[b^2(\Lambda_1, Z_1)] = O(N^{-2}h^{-d-2})$. We deduce that

$$\mathbb{V}\mathrm{ar}(T_N^{5,1}) \;=\; O\left(\frac{1}{N^2 h^{2d+n+2}}\right) = o\left(\frac{1}{Nh^{2+d}}\right), \tag{7.13}$$

using the relations between N and h given by (4.9).

2. We next rewrite $T_N^{5,2}$ in :

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

By the usual change of variables,

$$\begin{aligned} b(\lambda, z) \;&=\; h^{d+n} \int \phi(z+hv)\; K\left(\frac{\lambda^0-\lambda}{h} - l\right) \nabla K(l) H(v)\, dl\, dv \\ &\quad - h^{n+1} \int \phi(z)\; \bar{\varphi}_\lambda(\lambda^0 - hl, z) K(l)\, dl. \end{aligned}$$

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

$$\begin{aligned} &\mathbb{E}[b(\Lambda_i, Z_i) b(\Lambda_i, Z_i)'] \\ \sim\; & h^{3d+2n} \int \left\{ \int \phi(z+hv) K(l_2 - l_1) \nabla K(l_1) H(v)\, dl_1\, dv \right\}^{\otimes} \varphi(\lambda^0 - hl_2, z)\, dl_2\, dz\,. \end{aligned}$$

By Assumptions S and R3, we deduce from the dominated convergence theorem together with the fact that $\mathbb{E}[b(\Lambda_i, Z_i)] = 0$ that

$$\mathbb{V}\mathrm{ar}[T_N^{5,2}] \;\sim\; \frac{1}{Nh^{d+2}} \int \phi^2(z) \left\{ \int K(l_2 - l_1) \nabla K(l_1)\, dl_1 \right\}^{\otimes} \varphi(\lambda^0, z)\, dl_2\, dz\,. \tag{7.14}$$

The proof is completed by collecting the estimates (7.13) and (7.14) into (7.11). □

Lemma 7.4 $\mathbb{E}[T_N^6] = 0$ *and* $\mathbb{V}\mathrm{ar}(T_N^6) = o(N^{-1}h^{-d-2})$.

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

$$\mathbb{V}\mathrm{ar}(T_N^6) \sim \frac{h^{-4d-2n}}{\ell(\lambda^0)^2 N^2} \sum_i \mathbb{V}\mathrm{ar}(\tilde{b}(\Lambda_i, Z_i))$$

where $\tilde{b}$ is defined by :

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

The variables $\tilde{b}(\Lambda_i, Z_i)$ have also zero mean and, as in the proof of lemma 7.3, the usual change of variables implies that

$$h^{-3d-2n}\, \mathbb{V}\mathrm{ar}(\tilde{b}(\Lambda_i, Z_i)) \sim$$
$$\int \left\{\int \phi(z + hv)\frac{\varphi_\lambda}{\varphi}(\lambda^0 + hl_1 - hl_2, z + hv) K(l_2 - l_1) K(l_1) H(v)\, dl_1\, dv\right\}^{\otimes} \varphi(\lambda^0 - hl_2, z)\, dl_2\, dz\,.$$

By the continuity and the uniform boundedness of ϕ and φ_λ/φ implied by Assumptions S and R3, we derive

$$\mathbb{V}\mathrm{ar}(T_n^6) = O\left(\frac{1}{Nh^d}\right) = o\left(\frac{1}{Nh^{d+2}}\right).$$

□

Lemma 7.5 $\mathbb{E}[T_N^7] = 0$ *and* $\mathbb{V}\mathrm{ar}(T_N^7) = o(N^{-1}h^{-d-2})$.

Proof. Observe that

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

Following the lines of the proof of lemma 7.3, we see that $\mathbb{E}[T_N^7] = 0$, and we estimate

$$Nh^{d+2}\mathbb{V}\mathrm{ar}(T_N^7) \sim$$
$$\int \left\{\int \phi(z + hv)\psi(\lambda^0 + hl - hu, z + hv) K(u - l) \nabla K(l) H(v)\, dl\, dv\right\}^{\otimes} \varphi(\lambda^0 - hu, z)\, du\, dz\,.$$

By (7.4) and (7.9) it follows that $\|\psi\|_\infty = O(h^{p\wedge q})$ and, since φ and ϕ are uniformly bounded, we deduce that

$$\mathbb{V}\mathrm{ar}(T_N^7) = O\left(\frac{h^{p\wedge q}}{Nh^{d+2}}\right) = o\left(\frac{1}{Nh^{d+2}}\right).$$

□

Lemma 7.6 $\mathbb{E}\left[T_N^8\right] \sim \frac{h^{-d-n-1}}{\ell(\lambda^0)N}\left(\int \phi\right)\left(\int H^2\right)\int K(l_1-l_2)K(l_2)\nabla K(l_2)dl_1 dl_2$ *and* $\mathbb{V}\mathrm{ar}(T_N^8) = o(N^{-1}h^{-d-2})$.

Proof. 1. We first estimate $\mathbb{E}\left[T_N^8\right]$. Write $t_N^8(\lambda,z) = t_N^{8,1}(\lambda,z) + t_N^{8,2}(\lambda,z)$, with

$$\begin{aligned} t_N^{8,1} &= \frac{(\bar{\varphi}-\hat{\varphi}^{-i})(\hat{\varphi_\lambda}^{-i}-\bar{\varphi_\lambda})}{\varphi^2}, \\ t_N^{8,2} &= \frac{(\bar{\varphi}-\hat{\varphi}^{-i})^2(\hat{\varphi_\lambda}^{-i}-\bar{\varphi_\lambda})}{\hat{\varphi}^{-i}\,\bar{\varphi}^2} + \frac{(\bar{\varphi}-\hat{\varphi}^{-i})(\hat{\varphi_\lambda}^{-i}-\bar{\varphi_\lambda})(\varphi^2-\bar{\varphi}^2)}{\varphi^2\,\bar{\varphi}^2}. \end{aligned}$$

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

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

Set :

$$\begin{aligned} U_{ij} &:= \nabla_\lambda K\left(\frac{\Lambda_i-\Lambda_j}{h}\right)H\left(\frac{Z_i-Z_j}{h}\right) - \mathbb{E}\left[\nabla_\lambda K\left(\frac{\Lambda_i-\Lambda_j}{h}\right)H\left(\frac{Z_i-Z_j}{h}\right)|\Lambda_i,Z_i\right], \\ V_{ij} &:= K\left(\frac{\Lambda_i-\Lambda_j}{h}\right)H\left(\frac{Z_i-Z_j}{h}\right) - \mathbb{E}\left[K\left(\frac{\Lambda_i-\Lambda_j}{h}\right)H\left(\frac{Z_i-Z_j}{h}\right)|\Lambda_i,Z_i\right], \end{aligned}$$

so that

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

Using this property, we compute directly that

$$\begin{aligned} \mathbb{E}\left[t_N^{8,1}(\Lambda_1,Z_1)|\Lambda_1,Z_1\right] &= \frac{h^{-2d-2n-1}}{(N-1)^2\varphi^2(\Lambda_1,Z_1)}\mathbb{E}\left[\sum_{j\neq 1}\sum_{k\neq 1}U_{1j}\,V_{1k}|\Lambda_1,Z_1\right] \\ &= \frac{h^{-2d-2n-1}}{(N-1)\varphi^2(\Lambda_1,Z_1)}\mathbb{E}\left[U_{12}\,V_{12}|\Lambda_1,Z_1\right]. \end{aligned}$$

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

$$\mathbb{E}\left[T_N^{8,1}\right] = \frac{h^{-d}}{\ell(\lambda^0)}\mathbb{E}\left[\phi(Z_1)K\left(\frac{\lambda^0-\Lambda_1}{h}\right)\mathbb{E}\left[t_N^{8,1}(\Lambda_1,Z_1)|\Lambda_1,Z_1\right]\right],$$

we derive by the usual change of variables,

$$\begin{aligned} &\ell(\lambda^0)Nh^{d+n+1}\,\mathbb{E}\left[T_N^{8,1}\right] \sim \\ &\int \frac{\phi(z+hv)}{\varphi(\lambda^0+hl_1-hl_2,z+hv)}K(l_2-l_1)K(l_1)\nabla K(l_1)H^2(v)\varphi(\lambda^0-hl_2,z)\,dl_1\,dv\,dl_2\,dz. \end{aligned}$$

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

$$\mathbb{E}\left[T_N^{8,1}\right] \sim \frac{h^{-d-n-1}}{\ell(\lambda^0)N}\int \phi(z)K(l_2-l_1)K(l_1)\nabla K(l_1)H^2(v)\,dl_1\,dv\,dl_2\,dz.$$

Furthermore, by relations (7.4) to (7.9), we derive $\|t_N^{8,2}\|_\infty = o\left(N^{-1}h^{-d-n-1}\right)$. From lemma 7.1, we deduce $\mathbb{E}\left[T_N^{8,2}\right] = o\left(N^{-1}h^{-d-n-1}\right)$ and therefore get the expected equivalent for $\mathbb{E}\left[T_N^8\right]$.

2. We now study the variance of T_N^8. Using relations (7.5) and (7.9), we derive:

$$\left|T_N^8\right| \leq \frac{C}{Nh^d}\sum_{i=1}^N \frac{\ln^2 N}{Nh^{d+n+1}}\left|\phi(Z_i)K\left(\frac{\lambda^0-\Lambda_i}{h}\right)\right|,$$

so that

$$\mathbb{V}\mathrm{ar}\left(T_N^8\right) \leq \frac{C^2\ln^4 N}{N^2h^{2d+2n+2}}\mathbb{E}\left[\left(\frac{1}{Nh^d}\sum_{i=1}^N\left|\phi(Z_i)K\left(\frac{\lambda^0-\Lambda_i}{h}\right)\right|\right)^2\right].$$

By the usual change of variables, we finally deduce

$$\mathbb{V}\mathrm{ar}\left(T_N^8\right) = O\left(\frac{\ln^4 N}{N^2h^{2d+2n+2}}\right) = o\left(\frac{1}{Nh^{d+2}}\right),$$

by relation (4.9). □

Lemma 7.7 $\mathbb{E}[T_N^9] = O(N^{-1}h^{-d-n})$ *and* $\mathbb{V}\mathrm{ar}(T_N^9) = o(N^{-1}h^{-d-2})$

Proof. It can be easily checked that T_N^9 can be dealt as T_N^8 and, following the lines of the proof of lemma 7.6, we obtain the announced result.

7.3 Proof of Proposition 4.3

1. We first derive the expectation of $\tilde{\beta}_N$.

Notice that $T_N^1 = \bar{\beta}_N$ as defined in (3.2) which satisfies

$$\mathbb{E}\left[\bar{\beta}_N\right] = \frac{1}{\ell(\lambda^0)}\int \phi(z)K(l)s(\lambda^0-hl,z)\varphi(\lambda^0-hl,z)\,dt\,dz.$$

The regularity of function $s\varphi$ given by assumption R1 enables us to derive :

$$\mathbb{E}[T_N^1]-\beta \sim \frac{h^p}{\ell(\lambda^0)}\int \xi_K^p\left[\ell f_\lambda\right](\lambda^0,z)\phi(z)\,dz. \tag{7.15}$$

Using remark 7.2, we derive from (7.3) :

$$\mathbb{E}[T_N^2] = \frac{h^p}{\ell(\lambda^0)}\int \xi_K^p\left[\varphi_\lambda\right](\lambda^0,z)\phi(z)\,dz + \frac{h^q}{\ell(\lambda^0)}\int \xi_H^q\left[\varphi_\lambda\right](\lambda^0,z)\phi(z)\,dz + o(h^{p\wedge q}).$$

Similarly, we compute from (7.2) :

$$\mathbb{E}[T_N^3] = -\frac{h^p}{\ell(\lambda^0)}\int \frac{\varphi_\lambda\xi_K^p\left[\varphi\right]}{\varphi}(\lambda^0,z)\phi(z)\,dz - \frac{h^q}{\ell(\lambda^0)}\int \frac{\varphi_\lambda\xi_H^q\left[\varphi\right]}{\varphi}(\lambda^0,z)\phi(z)\,dz + o(h^{p\wedge q}).$$

Note that

$$t_N^1 = \frac{(\varphi - \bar{\varphi})^2 \varphi_\lambda}{\varphi^2 \bar{\varphi}} + \frac{(\bar{\varphi}_\lambda - \varphi_\lambda)(\varphi - \bar{\varphi})}{\varphi \bar{\varphi}}.$$

Then, using (7.4), (7.6), (7.8) and (7.9), we derive $\|t_N^1\|_\infty = o\left(h^{p \wedge q}\right)$ and lemma 7.1 leads to $\mathbb{E}(T_N^1) = o(h^{p \wedge q})$.

From lemmas 7.3, 7.4 and 7.5, we have $\mathbb{E}(T_N^j) = 0$ for $j = 5 \ldots 7$ and lemma 7.6 gives

$$\mathbb{E}\left[T_N^8\right] \sim \frac{h^{-d-n-1}}{\ell(\lambda^0) N} \int \frac{\phi(z)}{\varphi(\lambda^0, z)} K(l_2 - l_1) K(l_1) \nabla K(l_1) H^2(v)\, dl_1\, dv\, dl_2\, dz\,.$$

Finally, lemma 7.7 tells us $\mathbb{E}[T_N^9] = o(N^{-1} h^{-d-n-1})$.

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

2. We then analyze the variance of $\tilde{\beta}_N$. For any $j = 1 \ldots 4$, expressions (7.4), (7.6), (7.8), (7.9) and the uniform boundedness of $\nabla \ln \ell$ imply $\|t_N^j\|_\infty = O(1)$. Then lemma 7.2 leads to

$$\mathbb{V}\mathrm{ar}(T_N^j) = o(N^{-1} h^{-d-2}) \quad \text{for every } j = 1 \ldots 4.$$

From lemma 7.3, we get

$$\mathbb{V}\mathrm{ar}(T_N^5) \sim \frac{1}{\ell(\lambda^o) N h^{d+2}} \int \phi^2(z) \left\{ \int K(l_2 - l_1) \nabla K(l_1)\, dl_1 \right\}^{\otimes} f(\lambda^0, z)\, dz\, dl_2\,. \quad (7.16)$$

Indeed lemmas 7.4 to 7.7 imply also

$$\mathbb{V}\mathrm{ar}(T_N^j) = o(N^{-1} h^{-d-2}) \quad \text{for every } j = 5 \ldots 9.$$

Hence, $\mathbb{C}\mathrm{ov}(T_N^j, T_N^k) = o(N^{-1} h^{-d-2})$ unless $j = k = 5$ and $\mathbb{V}\mathrm{ar}(\tilde{\beta}_N)$ is given by expression (7.16). □

7.4 Proof of Theorem 4.3

As we saw in the proof of proposition 4.3, the variance of $\tilde{\beta}_N$ is given by the variance of

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

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

$$h^{d+n} \int \phi(z + hv)\, K\left(\frac{\lambda^0 - \lambda}{h} - l\right) \nabla K(l) H(v)\, dl\, dv \; - \; h^{n+1} \int \phi(z)\, \bar{\varphi}_\lambda(\lambda^0 - hl, z) K(l)\, dl.$$

As in the proofs of theorems 4.1 and 4.2, using Kolmogorov's condition with the fourth moment of b and the Cramer-Wold device, we derive that $T_N^{5,2}$ is asymptotically normal. We then finally deduce that :

$$\sqrt{N h^{d+2}} \left(\tilde{\beta}_N - \mathbb{E}[\tilde{\beta}_N]\right) \xrightarrow[N \to \infty]{law} \mathcal{N}\left(0, \tilde{\Sigma}\right)\,.$$

□

References

[1] BILLINGSLEY, P. (1968). *Convergence of probability measures*, Wiley.

[2] BROADIE, M. AND GLASSERMAN, P. (1996). Estimating security prices using simulation. *Management Science* **42** 269-285.

[3] DETEMPLE, J., GARCIA, R. AND RINDISBACHER, M. (2003). Asymptotic Properties of Monte Carlo Estimators of Derivatives. *Working paper CIRANO*, Montréal.

[4] L'ECUYER, P., PERRON, G. (1994). On the Convergence Rates of IPA and FDC derivative Estimators. *Operations Research* **42** 643-656.

[5] FOURNIÉ, E., LASRY, J.M., LEBUCHOUX, J., LIONS, P.L., AND TOUZI, N. (1999). Applications of Malliavin Calculus to Monte Carlo Methods in Finance. *Finance and Stochastics* **3** 391-412.

[6] FOURNIÉ, E., LASRY, J.M., LEBUCHOUX, J. AND LIONS, P.L. (2000). Applications of Malliavin Calculus to Monte Carlo Methods in Finance. II. *Finance and Stochastics* **5** 201-236.

[7] GOBET, E. (2004). Revisiting the Greeks for European and American Options. *Proceedings of the "International Symposium on Stochastic Processes and Mathematical Finance"*.

[8] GOBET, E., KOHATSU-HIGA, A. (2003). Computation of Greeks for barrier and Lookback options using Malliavin Calculus. *Electronic Communications in Probability* **8** 51-62.

[9] HULL, J. (2002). *Options, futures, and other derivatives*. Prentice Hall.

[10] KOHATSU-HIGA, A. AND MONTERO, M. (2004). Malliavin Calculus in Finance. To appear in *Handbook of Computational Finance*, Birkhauser.

[11] KUNITA, H. (1984). *Ecole d'été de Probabilité de Saint Flour XII - 1982, Stochastic differential equations and stochastic flow of diffeomorphisms*. Springer-Verlag.

[12] NUALART, D. (1995). *Malliavin Calculus and Related Topics*. Probability and its applications, Springer Verlag.

[13] POLLARD, D. (1984). *Convergence of stochastic processes*. Springer.

[14] SCOTT, D.W. (1992). *Multivariate Density estimation*. Wiley.