Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Stochastic Processes and their Applications 117 (2007) 1689–1723
stochastic processes and their applications
www.elsevier.com/locate/spa

Malliavin Greeks without Malliavin calculus

Nan Chen[a,*], Paul Glasserman[b,1]

[a] *Chinese University of Hong Kong, Hong Kong*
[b] *Columbia Business School, Columbia University, New York, NY 10027, USA*

Received 2 August 2006; received in revised form 13 March 2007; accepted 25 March 2007
Available online 5 June 2007

Abstract

We derive and analyze Monte Carlo estimators of price sensitivities ("Greeks") for contingent claims priced in a diffusion model. There have traditionally been two categories of methods for estimating sensitivities: methods that differentiate paths and methods that differentiate densities. A more recent line of work derives estimators through Malliavin calculus. The purpose of this article is to investigate connections between Malliavin estimators and the more traditional and elementary pathwise method and likelihood ratio method. Malliavin estimators have been derived directly for diffusion processes, but implementation typically requires simulation of a discrete-time approximation. This raises the question of whether one should discretize first and then differentiate, or differentiate first and then discretize. We show that in several important cases the first route leads to the same estimators as are found through Malliavin calculus, but using only elementary techniques. Time-averaging of multiple estimators emerges as a key feature in achieving convergence to the continuous-time limit.

Keywords: Monte Carlo simulation; Likelihood ratio method; Pathwise derivative method; Malliavin calculus; Weak convergence

1. Introduction

The calculation of price sensitivities is a central modeling and computational problem for derivative securities. The prices of derivative securities are, to varying degrees, observable in

* Corresponding address: Chinese University of Hong Kong, System Engineering, Shatin NT, Hong Kong. Tel.: +852 26098237.

E-mail addresses: nchen@se.cuhk.edu.hk (N. Chen), pg20@columbia.edu (P. Glasserman).

[1] Tel.: +1 212 854 4102.

0304-4149/$ - see front matter
doi:10.1016/j.spa.2007.03.012

the market; but the hedging of derivative securities is based on price sensitivities, and these sensitivities – which are not observable – require models and computational tools.

The computational effort required for the accurate calculation of price sensitivities (or "Greeks") is often substantially greater than that required for the calculation of the prices themselves. This is particularly true of Monte Carlo simulation, for which the computing time required for sensitivities can easily be 10–100 times greater than the computing time required to estimate prices to the same level of precision.

The simplest and crudest approach to the Monte Carlo estimation of sensitivities to a parameter simulates at two or more values of the underlying parameter and produces a finite-difference approximation to the price sensitivity. In the case of "delta", this means simulating from different initial states; in the case of "vega", this means simulating at different values of a volatility parameter; and in the case of "rho", this means simulating at different values of a drift parameter. Finite-difference estimators are easy to implement, but are prone to large bias, large variance, and added computational requirements.

Alternative methods seek to produce better estimators through some analysis of the underlying model. These methods evaluate derivatives directly, without finite-difference approximations. *Pathwise* methods treat the parameter of differentiation as a parameter of the evolution of the underlying model and differentiate this evolution. At the other extreme, the *likelihood ratio method* puts the parameter in the measure describing the underlying model and differentiates this measure. Because a price calculated by Monte Carlo is an expectation – the integral of a discounted payoff evaluated on each path, integrated against a probability measure – all estimators of sensitivities must involve some combination of these basic ideas: differentiating the evolution of the path, or differentiating the measure. For general background on estimating sensitivities and many references to the literature on this problem, see, e.g., Chapter 7 of Glasserman [9].

More recently, a fairly large and growing literature has developed around the derivation of sensitivity estimators using Malliavin calculus. This line of work originated in Fournié, Lasry, Lebuchoux, and Touzi [7] and includes Benhamou [1], Bermin, Kohatsu-Higa, and Montero [2], Cvitanic, Ma, and Zhang [3], Davis and Johansson [4], Fournié et al. [8], Gobet and Kohatsu-Higa [10], Kohatsu-Higa and Montero [12], and many others. Using the tools of Malliavin calculus (cf. Nualart [15]), this approach derives estimators in continuous time, though their implementation typically requires some form of time-discretization. The purpose of this article is to investigate the connection between Malliavin estimators and estimators derived using the more elementary ideas of the pathwise and likelihood ratio methods (LRM).

Our approach is as follows. We begin with a model specified through a stochastic differential equation. Whereas the application of Malliavin calculus would, in effect, first differentiate and then discretize, we discretize first. For simplicity, we use an Euler scheme. In the time-discrete approximation, it is easy to derive pathwise and LRM estimators. Our main contribution is to show how to combine these methods and then pass to the continuous-time limit in a way that *produces the Malliavin estimators*. To put this another way, discretizing the Malliavin estimators yields estimators that are equivalent (up to terms that vanish in the continuous-time limit) to estimators derived using the more elementary methods. We carry this out for three important cases of the Malliavin approach considered in Fournié et al. [7].

An insight that emerges from this analysis is the critical role played by time-averaging of multiple unbiased sensitivity estimators in passing to the continuous-time limit. This becomes particularly evident in the case of delta. A straightforward application of LRM to an Euler scheme produces a delta estimator that explodes as the time increment decreases to zero. To obtain a meaningful limit, we associate a separate unbiased estimator with each step along a (time-discretized) path and average these estimators. The average converges to the Malliavin estimator, though none of the individual estimators does. This observation sheds light on the flexible weights that often appear in Malliavin estimators, and indicates that a virtue of the Malliavin derivation is that it implicitly undertakes the necessary averaging.

We do not see the derivations in this article as inherently better or worse than those using Malliavin calculus. Working directly in continuous time often permits the use of powerful and efficient tools for analysis; working in discrete time allows more elementary arguments and can produce estimators that can be implemented without further approximation. Both approaches have advantages, and the purpose of this article is to illustrate connections between them. We do this for three important cases — sensitivities to an initial state, a drift parameter, and a diffusion parameter. Because our objective is to provide insight, we restrict our analysis to one-dimensional problems.

The rest of this paper is organized as follows. Section 2 outlines the main steps in our derivations. In Section 3, we verify that the estimators we derive are unbiased for the discrete-time approximations with which we work. In Section 4, we show that these estimators converge weakly as the time step decreases. Several technical results are collected in Appendices A–C.

2. Preview of main results

To prevent technical considerations from obscuring the simplicity of our main results, in this section we outline our derivations without discussing the conditions required for their validity. Subsequent sections are devoted to justifying the approach we sketch here.

We suppose that the underlying model dynamics are given by a stochastic differential equation on $[0, T]$,

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \qquad X_0 = x, \tag{1}$$

where W is a standard Brownian motion. For simplicity, we restrict attention to scalar X. Consider a (discounted) payoff function Φ that depends on the values of the underlying asset at times $0 \le t_1 < \cdots < t_m \le T$. The expected present value of a contingent claim with this payoff is

$$u(x) = E[\Phi(X_{t_1}, \ldots, X_{t_m})], \tag{2}$$

the expectation taken with $X_0 = x$. In this section, we focus on the case $m = 1$, in which $u(x) = E[\Phi(X_T)]$.

2.1. Delta

We begin by considering delta, the sensitivity of $u(x)$ to the initial state x. When applicable, the pathwise method brings the derivative with respect to x inside the expectation to get

$$u'(x) = E\left[\Phi'(X_T)\frac{\mathrm{d}X_T}{\mathrm{d}x}\right].$$

When equality holds,

$$\Phi'(X_T)Y_T$$

provides an unbiased estimator of $u'(x)$, where $Y_T = \mathrm{d}X_T/\mathrm{d}x$ is the pathwise derivative of X_T with respect to the initial state. Under conditions in Section V.7 of Protter [16], the dynamics of Y can be obtained from (1) to get

$$\mathrm{d}Y_t = \mu'(X_t)Y_t\mathrm{d}t + \sigma'(X_t)Y_t\mathrm{d}W_t, \qquad Y_0 = 1.$$

The likelihood ratio method (LRM) estimator starts from the transition density $g(x,\cdot)$ describing the distribution of X_T given $X_0 = x$. The price $u(x)$ is given by

$$u(x) = \int \Phi(x_T)g(x,x_T)\,\mathrm{d}x_T,$$

so bringing the derivative inside the integral and then multiplying and dividing by $g(x,x_T)$ yields

$$\begin{aligned} u'(x) &= \int \Phi(x_T)\frac{\mathrm{d}}{\mathrm{d}x}g(x,x_T)\,\mathrm{d}x_T \\ &= \int \Phi(x_T)\left[\frac{\mathrm{d}}{\mathrm{d}x}\log g(x,x_T)\right]g(x,x_T)\,\mathrm{d}x_T \\ &= E\left[\Phi(X_T)\frac{\mathrm{d}}{\mathrm{d}x}\log g(x,X_T)\right] \end{aligned} \tag{3}$$

and the unbiased estimator

$$\Phi(X_T)\frac{\mathrm{d}}{\mathrm{d}x}\log g(x,X_T).$$

By differentiating the density, the LRM method avoids imposing any smoothness conditions on Φ. However, it requires existence and knowledge of g.

The (or rather, *a*) Malliavin estimator for this problem is (cf. Fournié et al. [7], p. 399)

$$\Phi(X_T)\frac{1}{T}\int_0^T \frac{Y_t}{\sigma(X_t)}\,\mathrm{d}W_t. \tag{4}$$

Like the LRM estimator, this estimator multiplies the payoff $\Phi(X_T)$ by a random weight to estimate the derivative. In contrast to the LRM estimator, it does not involve the transition density g.

Consider, now, an Euler approximation,

$$\hat{X}_i = \hat{X}_{i-1} + \mu(\hat{X}_{i-1})\Delta t + \sigma(\hat{X}_{i-1})\Delta W_i, \qquad \hat{X}_0 = x, \tag{5}$$

$i = 1,\dots,N$, with time step $\Delta t = T/N$ and $\Delta W_i = W(i\Delta t) - W((i-1)\Delta t)$. Let $\hat{u}(x) = E[\Phi(\hat{X}_N)]$ and let $\hat{Y}_i = \mathrm{d}\hat{X}_i/\mathrm{d}x$,

$$\hat{Y}_i = \hat{Y}_{i-1} + \mu'(\hat{X}_{i-1})\hat{Y}_{i-1}\Delta t + \sigma'(\hat{X}_{i-1})\hat{Y}_{i-1}\Delta W_i, \qquad \hat{Y}_0 = 1. \tag{6}$$

The pathwise estimator of $\hat{u}'(x)$, the delta for the Euler scheme, is

$$\Phi'(\hat{X}_N)\hat{Y}_N.$$

For the LRM estimator, we may write

$$\hat{u}(x) = \int\cdots\int \Phi(x_N)\hat{g}(x,x_1)\cdots\hat{g}(x_{N-1},x_N)\,\mathrm{d}x_N\cdots\mathrm{d}x_1, \tag{7}$$

where $\hat{g}(x_{i-1},x_i)$ is the transition density from $\hat{X}_{i-1} = x_{i-1}$ to $\hat{X}_i = x_i$. Proceeding as before, we arrive at the estimator

$$\Phi(\hat{X}_N)\sum_{i=1}^N \frac{\mathrm{d}}{\mathrm{d}x}\log\hat{g}(\hat{X}_{i-1},\hat{X}_i) = \Phi(\hat{X}_N)\frac{\mathrm{d}}{\mathrm{d}x}\log\hat{g}(x,\hat{X}_1), \tag{8}$$

noting that only the first of the transition densities depends on the initial state x.

Whereas the transition density of the continuous-time process X is often unknown, the transition density for the Euler scheme is Gaussian. In particular, $\hat{X}_1$ is normally distributed with mean $x + \mu(x)\Delta t$ and variance $\sigma^2(x)\Delta t$. As a consequence, we can differentiate the log density and, after some simplification, write the estimator (8) as

$$\Phi(\hat{X}_N)\left(\frac{\Delta W_1}{\sigma(x)\Delta t} + o_p(1)\right),$$

where $o_p(1)$ converges weakly to zero as Δt approaches zero. While this estimator is, under mild conditions, unbiased for $\hat{u}'(x)$ for all Δt, it clearly behaves badly as Δt approaches zero.

But we have more flexibility than (8) initially indicates. For any $i = 1, \dots, N$, we may write

$$\hat{u}(x) = E\left[\int \cdots \int \Phi(x_N)\hat{g}(\hat{X}_{i-1}(x), x_i) \cdots \hat{g}(x_{N-1}, x_N)\,\mathrm{d}x_N \cdots \mathrm{d}x_i\right].$$

Here, we have written $\hat{X}_{i-1}$ as $\hat{X}_{i-1}(x)$ to stress that $\hat{X}_{i-1}$ now has a functional dependence on the initial state x through the Euler recursion (5). Differentiating inside the expectation and integral and proceeding as before, we get the estimator

$$\Phi(\hat{X}_N)\frac{\mathrm{d}}{\mathrm{d}\hat{X}_{i-1}}\log \hat{g}(\hat{X}_{i-1}, \hat{X}_i)\frac{\mathrm{d}\hat{X}_{i-1}}{\mathrm{d}x}. \tag{9}$$

The new factor (which we will write as $\hat{Y}_{i-1}$) enters through the chain rule of ordinary calculus. This estimator puts the dependence on x in the path up to the $(i-1)$st step (as in the pathwise method), and then treats $\hat{X}_{i-1}$ as a parameter of the conditional distribution of $\hat{X}_i$ (as in the LRM method).

Again using the fact that $\hat{g}$ is Gaussian, we can write (9) as

$$\Phi(\hat{X}_N)\left(\frac{\Delta W_i}{\sigma(\hat{X}_{i-1})\Delta t} + o_p(1)\right)\hat{Y}_{i-1}.$$

Under mild conditions, this is unbiased for $\hat{u}'(x)$ for all Δt, for all $i = 1, \dots, N$. If we now average these unbiased estimators, we get

$$\Phi(\hat{X}_N)\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\frac{\Delta W_i}{\sigma(\hat{X}_{i-1})\Delta t} + o_p(1)\right)\hat{Y}_{i-1} \approx \Phi(X_T)\frac{1}{T}\int_0^T \frac{Y_t}{\sigma(X_t)}\,\mathrm{d}W_t, \tag{10}$$

for small Δt. Thus, we recover the Malliavin estimator (4) as the limit of the average of combinations of pathwise and LRM estimators. Theorem 4.6 makes this limit precise.

2.2. Vega

Next, we turn to the estimation of vega, or sensitivity to changes in the diffusion coefficient. Suppose X^ε satisfies

$$\mathrm{d}X_t^\varepsilon = \mu(X_t^\varepsilon)\,\mathrm{d}t + [\sigma(X_t^\varepsilon) + \varepsilon\tilde{\sigma}(X_t^\varepsilon)]\,\mathrm{d}W_t, \qquad X_0^\varepsilon = x, \tag{11}$$

for some $\tilde{\sigma}$, and $Z_t^\varepsilon = \mathrm{d}X_t^\varepsilon/\mathrm{d}\varepsilon$ satisfies

$$\mathrm{d}Z_t^\varepsilon = \mu'(X_t^\varepsilon)Z_t^\varepsilon\,\mathrm{d}t + \tilde{\sigma}(X_t^\varepsilon)\,\mathrm{d}W_t + [\sigma'(X_t^\varepsilon) + \varepsilon\tilde{\sigma}(X_t^\varepsilon)]Z_t^\varepsilon\,\mathrm{d}W_t, \qquad Z_0^\varepsilon = 0. \tag{12}$$

Let Z_t denote Z_t^ε at $\varepsilon = 0$. The Malliavin estimator for the sensitivity of $E[\Phi(X_T)]$ with respect to ε at $\varepsilon = 0$ is (cf. Fournié et al. [7], p. 403)

$$\Phi(X_T)\left\{\frac{Z_T}{Y_T}\frac{1}{T}\int_0^T \frac{Y_t}{\sigma(X_t)}\,\mathrm{d}W_t - \frac{1}{T}\int_0^T D_t\left(\frac{Z_T}{Y_T}\right)\frac{Y_t}{\sigma(X_t)}\,\mathrm{d}t\right\}. \tag{13}$$

Here, D_t denotes the Malliavin derivative operator.

For the Euler scheme, let $\hat{Z}_i = \mathrm{d}\hat{X}_i/\mathrm{d}\varepsilon$ at $\varepsilon = 0$, for $i = 1, \ldots, N$. The pathwise estimator of $\mathrm{d}E[\Phi(\hat{X}_N)]/\mathrm{d}\varepsilon$ is

$$\Phi'(\hat{X}_N)\hat{Z}_N. \tag{14}$$

The LRM estimator has the form in (18), but ε now affects the transition densities through their variances rather than their means. Straightforward calculation shows that (18) becomes

$$\Phi(\hat{X}_N)\sum_{i=1}^N \frac{\tilde{\sigma}(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\left(\frac{\Delta W_i^2}{\Delta t} - 1\right). \tag{15}$$

This estimator fails to converge as Δt approaches zero; so, as we did for delta, we combine the ideas of the pathwise and LRM techniques with averaging along the path. Observe that, for small ε, the effect on $\hat{X}_N$ of perturbing σ by $\varepsilon\tilde{\sigma}$ is the same as the effect of perturbing the initial state x by $\varepsilon\hat{Z}_N/\hat{Y}_N$, provided $\hat{Y}_N \neq 0$. Thus, we may write the pathwise derivative in (14) as

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Phi(\hat{X}_N) = \left(\frac{\mathrm{d}}{\mathrm{d}x}\Phi(\hat{X}_N)\right)\frac{\hat{Z}_N}{\hat{Y}_N}.$$

By converting the sensitivity to ε to a sensitivity to x, we can take advantage of the derivation in Section 2.1. In order not to rely on differentiability of Φ, we can take the derivative with respect to x by multiplying by the LRM factor appearing on the left side of (10) to get

$$\left(\frac{1}{N}\sum_{i=1}^N \frac{\hat{Y}_{i-1}\Delta W_i}{\sigma(\hat{X}_{i-1})\Delta t} + o_p(1)\right)\Phi(\hat{X}_N)\frac{\hat{Z}_N}{\hat{Y}_N}.$$

But multiplying by the LRM factor has the effect of differentiating the product $\Phi(\hat{X}_N)(\hat{Z}_N/\hat{Y}_N)$, though what we want is the derivative of the first factor. To compensate, we subtract the derivative of the second factor and (recalling that $N\Delta t = T$) get

$$\Phi(\hat{X}_N)\left\{\left(\frac{1}{T}\sum_{i=1}^N \frac{\hat{Y}_{i-1}\Delta W_i}{\sigma(\hat{X}_{i-1})} + o_p(1)\right)\frac{\hat{Z}_N}{\hat{Y}_N} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\hat{Z}_N}{\hat{Y}_N}\right)\right\}. \tag{16}$$

This estimator subtracts a pathwise derivative from an LRM derivative.

The new term in (16) can be evaluated directly by recursively differentiating the Euler approximation (5). To make its connection to the Malliavin estimator more evident, we again average over the path and then use the fact that $\mathrm{d}\hat{X}_i/\mathrm{d}\Delta W_i = \sigma(\hat{X}_{i-1})$ to get

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\hat{Z}_N}{\hat{Y}_N}\right) = \frac{1}{N}\sum_{i=1}^N \frac{\mathrm{d}}{\mathrm{d}\hat{X}_i}\left(\frac{\hat{Z}_N}{\hat{Y}_N}\right)\hat{Y}_i = \frac{1}{T}\sum_{i=1}^N \frac{\mathrm{d}}{\mathrm{d}\Delta W_i}\left(\frac{\hat{Z}_N}{\hat{Y}_N}\right)\frac{\hat{Y}_i\Delta t}{\sigma(\hat{X}_{i-1})}.$$

Substituting this expression in (16) then suggests the convergence of (16) to (13). We will show that this approach does indeed produce estimators that are unbiased for all Δt and that converge as $\Delta t \to 0$. Some care will be required to handle division by $\hat{Y}_N$.

2.3. Rho

To consider sensitivities with respect to changes in drift, we consider a family of processes X^ε satisfying

$$dX_t^\varepsilon = [\mu(X_t^\varepsilon) + \varepsilon\gamma(X_t^\varepsilon)]\,dt + \sigma(X_t^\varepsilon)\,dW_t, \qquad X_0^\varepsilon = x,$$

for some γ, and we consider the derivative with respect to ε at $\varepsilon = 0$. The Malliavin estimator for this problem is (cf. Fournié et al. [7], p. 398)

$$\Phi(X_T)\int_0^T \frac{\gamma(X_t)}{\sigma(X_t)}\,dW_t. \tag{17}$$

The Euler approximation for the perturbed process is

$$\hat{X}_i^\varepsilon = [\mu(\hat{X}_{i-1}^\varepsilon) + \varepsilon\gamma(\hat{X}_{i-1}^\varepsilon)]\,\Delta t + \sigma(\hat{X}_{i-1}^\varepsilon)\,\Delta W_i, \qquad \hat{X}_0^\varepsilon = x.$$

Letting $\hat{g}_\varepsilon$ denote the transition density for this Euler approximation, the LRM estimator of the sensitivity is

$$\Phi(\hat{X}_N)\sum_{i=1}^N \frac{d}{d\varepsilon}\log \hat{g}_\varepsilon(\hat{X}_{i-1}, \hat{X}_i)\bigg|_{\varepsilon=0}. \tag{18}$$

The fact that the transition density is Gaussian simplifies this to

$$\Phi(\hat{X}_N)\sum_{i=1}^N \frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\Delta W_i.$$

The convergence of this estimator to the Malliavin estimator (17) now seems evident, and is stated precisely in Theorem 4.10.

3. Unbiased estimators for Greeks

In this section, we derive estimators of delta, vega and rho that are unbiased for the Euler approximation.

Introduce a sequence of i.i.d. random variables $\{\xi_i, i \geq 1\}$, where ξ_i follows the normal distribution with mean 0 and variance 1. Write the Euler scheme as

$$\hat{X}_i = \hat{X}_{i-1} + \mu(\hat{X}_{i-1})\Delta t + \sigma(\hat{X}_{i-1})\sqrt{\Delta t}\xi_i, \qquad \hat{X}_0 = x.$$

We also need the following technical conditions on the payoff function and the drift and volatility functions. The payoff Φ is a function of m variables, $\Phi : \mathbf{R}^m \to \mathbf{R}$, given by $\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m})$ for some $\{i_1, \ldots, i_m\} \subseteq \{1, \ldots, N\}$. We require the following:

Assumption 3.1. The growth rate of the function Φ is polynomial, i.e., there exists a positive integer p such that

$$\limsup_{\|x\|\to+\infty} \frac{|\Phi(x)|}{\|x\|^p} < +\infty.$$

Assumption 3.2. (1) The drift μ and the volatility σ are both twice differentiable with bounded first and second derivatives;
(2) σ is not degenerate, i.e., $\inf_x \sigma(x) > \epsilon$ for some $\epsilon > 0$;
(3) The function $\tilde{\sigma}$ preserves non-degeneracy under a small perturbation, i.e., there exists a neighborhood of 0, K, and some $\epsilon > 0$ such that for any $\varepsilon \in K$,

$$\inf_x [\sigma(x) + \varepsilon\tilde{\sigma}(x)] > \epsilon.$$

(4) $\tilde{\sigma}$ is differentiable and its derivative is bounded.

Assumption 3.3. The drift perturbation function γ is bounded.

In Sections 3.1–3.3, we consider the sensitivity of $u(x)$ with respect to a perturbation of the initial point x (delta), the volatility σ (vega) and the drift μ (rho), respectively. As in (6), define $\hat{Y}_i$ to be the derivative of $\hat{X}_i$ with respect to x, which satisfies the following recursion:

$$\hat{Y}_i = \hat{Y}_{i-1} + \mu'(\hat{X}_{i-1})\hat{Y}_{i-1}\Delta t + \sigma'(\hat{X}_{i-1})\hat{Y}_{i-1}\sqrt{\Delta t}\xi_i, \qquad \hat{Y}_0 = 1.$$

We will want to divide by values of the process $\hat{Y}$ in Section 3.2 and therefore we need to restrict the process $\hat{Y}$ to be positive almost surely to avoid the complication that $1/\hat{Y}$ could be non-integrable around 0. But this does not hold with normally distributed increments, so we will use truncated normals $\tilde{\xi}_i$ where necessary.

3.1. Delta

In the outline of Section 2.1, we averaged N unbiased estimators of delta with even weights and considered payoff functions depending only on the value of the underlying asset at the claim maturity. For the analysis of this section, we consider a more general case of uneven weights and path-dependent claims, as in Fournié et al. [7].

Theorem 3.1. *Suppose the weights $\{a_i : 0 \le i \le N\}$ satisfy $\sum_{i=1}^{i_j} a_i \Delta t = 1$ for all $1 \le j \le m$, and suppose Assumptions 3.1 and 3.2 hold. Then, the following is an unbiased estimator for $\hat{u}'(x)$:*

$$\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m}) \cdot \sum_{i=1}^{N} a_i \cdot \frac{\hat{Y}_{i-1}}{\sigma(\hat{X}_{i-1})} \cdot \sqrt{\Delta t}\xi_i + \mathrm{I} + \mathrm{II}, \tag{19}$$

where

$$\mathrm{I} = \Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m}) \cdot \sum_{i=1}^{N} a_i \cdot \frac{\sigma'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} \cdot [(\sqrt{\Delta t} \cdot \xi_i)^2 - \Delta t];$$

$$\mathrm{II} = \Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m}) \cdot \left(\sum_{i=1}^{N} a_i \cdot \frac{\mu'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} \sqrt{\Delta t} \cdot \xi_i \right) \cdot \Delta t.$$

Proof. By the Markov property of $\hat{X}$, we may write

$$\hat{u}(x) = \int \cdots \int \Phi(x_{i_1}, \ldots, x_{i_m}) \hat{g}(x, x_1) \cdots \hat{g}(x_{N-1}, x_N)\, \mathrm{d}x_N \ldots \mathrm{d}x_1,$$

where $\hat{g}(x_{i-1}, x_i)$ is the transition density function of $\hat{X}_i$, given the condition that $\hat{X}_{i-1} = x_{i-1}$.

$$\hat{g}(x_{i-1}, x_i) = \frac{1}{\sqrt{2\pi \Delta t}\sigma(x_{i-1})} \exp\left(-\frac{(x_i - x_{i-1} - \mu(x_{i-1})\Delta t)^2}{2\sigma^2(x_{i-1})\Delta t}\right)$$

because of the assumption of normal increments ξ. Under Assumptions 3.1 and 3.2, the growth rate of function Φ is polynomial and the transition density $\hat{g}$ decays exponentially. So we can interchange the order of differentiation and integration to get

$$\begin{aligned}\hat{u}'(x) &= \int \cdots \int \Phi(x_{i_1}, \ldots, x_{i_m}) \frac{\mathrm{d}\hat{g}(x, x_1)}{\mathrm{d}x} \hat{g}(x_1, x_2) \cdots \hat{g}(x_{N-1}, x_N)\, \mathrm{d}x_N \ldots \mathrm{d}x_1 \\ &= \int \cdots \int \Phi(x_{i_1}, \ldots, x_{i_m}) \frac{\mathrm{d}\ln \hat{g}(x, x_1)}{\mathrm{d}x} \\ &\quad \times \hat{g}(x, x_1)\hat{g}(x_1, x_2) \cdots \hat{g}(x_{N-1}, x_N)\, \mathrm{d}x_N \ldots \mathrm{d}x_1.\end{aligned}$$

After some algebra using the form of $\hat{g}$ given above, we have

$$\begin{aligned}\hat{u}'(x) &= \int \cdots \int \Phi(x_{i_1}, \ldots, x_{i_m}) \left\{ \frac{(x_1 - x - \mu(x)\Delta t)/(\sigma(x)\sqrt{\Delta t})}{\sigma(x)\sqrt{\Delta t}} \right. \\ &\quad \left. + \frac{x_1 - x - \mu(x)\Delta t}{\sigma(x)} \frac{\mu'(x)}{\sigma(x)} + \frac{\sigma'(x)}{\sigma(x)} \left[\left(\frac{x_1 - x - \mu(x)\Delta t}{\sigma(x)\sqrt{\Delta t}} \right)^2 - 1 \right] \right\} \\ &\quad \cdot \hat{g}(x, x_1) \cdots \hat{g}(x_{N-1}, x_N)\, \mathrm{d}x_N \cdots \mathrm{d}x_1 \\ &= E\left[\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m}) \left\{ \frac{(\hat{X}_1 - x - \mu(x)\Delta t)/(\sigma(x)\sqrt{\Delta t})}{\sigma(x)\sqrt{\Delta t}} \right. \right. \\ &\quad \left. \left. + \frac{\hat{X}_1 - x - \mu(x)\Delta t}{\sigma(x)} \frac{\mu'(x)}{\sigma(x)} + \frac{\sigma'(x)}{\sigma(x)} \left[\left(\frac{\hat{X}_1 - x - \mu(x)\Delta t}{\sigma(x)\sqrt{\Delta t}} \right)^2 - 1 \right] \right\} \right] \\ &= E\left[\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m}) \left\{ \frac{\xi_1}{\sigma(x)\sqrt{\Delta t}} + \xi_1 \sqrt{\Delta t} \frac{\mu'(x)}{\sigma(x)} + \frac{\sigma'(x)}{\sigma(x)} (\xi_1^2 - 1) \right\} \hat{Y}_0 \right]\end{aligned}$$

where we use the iteration $\hat{X}_1 = x + \mu(x)\Delta t + \sigma(x)\sqrt{\Delta t}\xi_1$ and $\hat{Y}_0 = 1$.

For the case that $i \le i_1$, the first reference date, by the Markovian property of $\hat{X}$,

$$\hat{u}(x) = E[E[\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m}) | \hat{X}_{i-1}]].$$

Interchanging the order of differentiation and expectation, and applying the chain rule, we have

$$\begin{aligned}\hat{u}'(x) &= \frac{\mathrm{d}}{\mathrm{d}x} E[E[\Phi | \hat{X}_{i-1}]] = E\left[\frac{\mathrm{d}}{\mathrm{d}\hat{X}_{i-1}} E[\Phi | \hat{X}_{i-1}] \cdot \frac{\mathrm{d}\hat{X}_{i-1}}{\mathrm{d}x} \right] \\ &= E\left[\frac{\mathrm{d}}{\mathrm{d}\hat{X}_{i-1}} E[\Phi | \hat{X}_{i-1}] \cdot \hat{Y}_{i-1} \right].\end{aligned}$$

Treating $\hat{X}_{i-1}$ as a new initial point and applying the same arguments as above here,

$$\frac{\mathrm{d}}{\mathrm{d}\hat{X}_{i-1}} E[\Phi | \hat{X}_{i-1}] = E\left[\Phi \cdot \left\{ \frac{\xi_i}{\sqrt{\Delta t}\sigma(\hat{X}_{i-1})} \right. \right.$$

$$+ \quad \sqrt{\Delta t}\xi_i \frac{\mu'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} + (\xi_i^2 - 1)\frac{\sigma'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\Bigg\}\Bigg| \hat{X}_{i-1}\Bigg].$$

Thus,

$$\hat{u}'(x) = E\left[\Phi \cdot \left\{\frac{\xi_i}{\sqrt{\Delta t}\sigma(\hat{X}_{i-1})} + \sqrt{\Delta t}\xi_i \frac{\mu'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} + (\xi_i^2 - 1)\frac{\sigma'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right\} \cdot \hat{Y}_{i-1}\right]. \quad (20)$$

For the case that $i > i_1$,

$$J_i := \Phi \cdot \left\{\frac{\xi_i}{\sqrt{\Delta t}\sigma(\hat{X}_{i-1})} + \sqrt{\Delta t}\xi_i \frac{\mu'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} + (\xi_i^2 - 1)\frac{\sigma'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right\} \cdot \hat{Y}_{i-1}$$

is not necessarily an unbiased estimator for $\hat{u}'(x)$. But we can show that $E[J_l] = E[J_k]$ for any $i_j < l, k \le i_{j+1}$. Therefore, given any set of weights $\{a_i : 0 \le i \le N\}$ such that $\sum_{i=1}^{i_j} a_i \Delta t = 1$ for all $1 \le j \le m$,

$$\begin{aligned}\sum_{i=1}^{N} a_i \Delta t \cdot E[J_i] &= \sum_{i=1}^{i_1} a_i \Delta t \cdot E[J_i] + \sum_{j=1}^{m} \sum_{i=i_j+1}^{i_{j+1}} a_i \Delta t \cdot E[J_i] \\ &= \hat{u}(x) \cdot \sum_{i=1}^{i_1} a_i \Delta t + \sum_{j=1}^{m} E[J_{i_j+1}] \sum_{i=i_j+1}^{i_{j+1}} a_i \Delta t.\end{aligned}$$

The right hand side of the above equality is $\hat{u}(x)$ because $\sum_{i=1}^{i_1} a_i \Delta t = 1$ and $\sum_{i=i_j+1}^{i_{j+1}} a_i \Delta t = 0$, and the left hand side is exactly

$$E\left[\sum_{i=1}^{N} a_i \Delta t \cdot J_i\right] = E\left[\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m}) \cdot \sum_{i=1}^{N} a_i \cdot \frac{\hat{Y}_{i-1}}{\sigma(\hat{X}_{i-1})} \cdot \sqrt{\Delta t}\xi_i + \mathrm{I} + \mathrm{II}\right].$$

Indeed, given $\mathcal{F}_{i_j} = \sigma\{\xi_1, \ldots, \xi_{i_j}\}$, we can see that if $\hat{Y}_{i_j} = 0$, then $\hat{Y}_l = 0$ by the dynamics of $\hat{Y}$ and thus $J_l = 0$ for all $i_j < l \le i_{j+1}$. We only need to show that when $\hat{Y}_{i_j} \ne 0$, $E[J_l|\mathcal{F}_{i_j}]$ is the same for all $i_j < l \le i_{j+1}$:

$$\begin{aligned}E[J_l|\mathcal{F}_{i_j}] = E\Bigg[&\Phi \cdot \left\{\frac{\xi_l}{\sqrt{\Delta t}\sigma(\hat{X}_{l-1})} + \sqrt{\Delta t}\xi_l \frac{\mu'(\hat{X}_{l-1})}{\sigma(\hat{X}_{l-1})} + (\xi_l^2 - 1)\frac{\sigma'(\hat{X}_{l-1})}{\sigma(\hat{X}_{l-1})}\right\} \\ &\cdot \frac{\hat{Y}_{l-1}}{\hat{Y}_{i_j}}\Bigg|\mathcal{F}_{i_j}\Bigg] \cdot \hat{Y}_{i_j}.\end{aligned} \quad (21)$$

Notice that $\hat{Y}_{l-1}/\hat{Y}_{i_j} = \mathrm{d}\hat{X}_{l-1}/\mathrm{d}\hat{X}_{i_j}$, given $\mathcal{F}_{i_j}$. Thus, the conditional expectation in the right hand side of (21) should be equal to

$$\frac{\mathrm{d}}{\mathrm{d}\hat{X}_{i_j}} E[\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_j}, \hat{X}_{i_{j+1}}, \ldots, \hat{X}_{i_m})|\mathcal{F}_{i_j}]$$

if we apply the arguments leading to (20) here. It does not depend on l. So does the left hand side of (21). □

3.2. Vega

We take vega to be the derivative of $u(x)$ with respect to a perturbation ε that takes $\sigma(\cdot)$ to $\sigma(\cdot)+\varepsilon\tilde{\sigma}(\cdot)$. In other words,

$$\text{vega} = \left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} E[\Phi(\hat{X}^{\varepsilon}_{i_1},\ldots,\hat{X}^{\varepsilon}_{i_m})],$$

where

$$\hat{X}^{\varepsilon}_i = \hat{X}^{\varepsilon}_{i-1} + \mu(\hat{X}^{\varepsilon}_{i-1})\Delta t + [\sigma(\hat{X}^{\varepsilon}_{i-1}) + \varepsilon\tilde{\sigma}(\hat{X}^{\varepsilon}_{i-1})]\sqrt{\Delta t}\tilde{\xi}_i, \qquad \hat{X}^{\varepsilon}_0 = x,$$

and $\tilde{\xi}_i$ is a truncated normal distributed random variable whose density function is given by

$$f(w) = \begin{cases} 0, & w < w_L; \\ \dfrac{C}{\sqrt{2\pi\delta}}\cdot \mathrm{e}^{-\frac{w^2}{2\delta}}, & w_L < w < w_R; \\ 0, & w > w_R. \end{cases}$$

The parameters C, δ, w_L, w_R satisfy (cf. Lemma A.1 for the existence of such parameters):

$$w_R = -w_L > 0, \qquad C > 0, \qquad \delta > 0; \tag{22}$$

$$\int_{w_L}^{w_R} \frac{C}{\sqrt{2\pi\delta}}\cdot \mathrm{e}^{-\frac{w^2}{2\delta}}\mathrm{d}w = 1; \tag{23}$$

$$\mathrm{Var}[\tilde{\xi}_i] = \int_{w_L}^{w_R} w^2\cdot\frac{C}{\sqrt{2\pi\delta}}\cdot \mathrm{e}^{-\frac{w^2}{2\delta}}\mathrm{d}w = 1. \tag{24}$$

(22) implies $E[\tilde{\xi}_i]=0$, (23) makes sure the function f is a probability density function.

Define $\hat{Z}^{\varepsilon} := \mathrm{d}\hat{X}^{\varepsilon}/\mathrm{d}\varepsilon$, which satisfies the following recursion

$$\begin{aligned}\hat{Z}^{\varepsilon}_i &= \hat{Z}^{\varepsilon}_{i-1} + \mu'(\hat{X}^{\varepsilon}_{i-1})\hat{Z}^{\varepsilon}_{i-1}\Delta t + [\sigma'(\hat{X}^{\varepsilon}_{i-1}) \\ &\quad + \varepsilon\tilde{\sigma}'(\hat{X}^{\varepsilon}_{i-1})]\hat{Z}^{\varepsilon}_{i-1}\sqrt{\Delta t}\tilde{\xi}_i + \tilde{\sigma}(\hat{X}^{\varepsilon}_{i-1})\sqrt{\Delta t}\tilde{\xi}_i, \quad \hat{Z}^{\varepsilon}_0 = 0.\end{aligned}$$

Let $\hat{Z}$ denote $\hat{Z}^{\varepsilon}$ at $\varepsilon=0$. In addition, we define $\hat{Y}^{\varepsilon} = \mathrm{d}\hat{X}^{\varepsilon}/\mathrm{d}x$, which satisfies the recursion:

$$\hat{Y}^{\varepsilon}_i = \hat{Y}^{\varepsilon}_{i-1} + \mu'(\hat{X}^{\varepsilon}_{i-1})\hat{Y}^{\varepsilon}_{i-1}\Delta t + [\sigma'(\hat{X}^{\varepsilon}_{i-1}) + \varepsilon\tilde{\sigma}'(\hat{X}^{\varepsilon}_{i-1})]\hat{Y}^{\varepsilon}_{i-1}\sqrt{\Delta t}\tilde{\xi}_i, \qquad \hat{Y}^{\varepsilon}_0 = 1.$$

It is easy to see that $\hat{Y}$ is equal to $\hat{Y}^{\varepsilon}$ at $\varepsilon=0$.

The following is the main theorem in the subsection:

Theorem 3.2. *Suppose Assumptions 3.1 and 3.2 hold. Then, for any sufficiently large integer N, and for any weight set $\{a_i : 0 \le i \le N\}$ such that $\sum_{i=i_j+1}^{i_j} a_i\Delta t = 1$ for all $0 \le j \le m$,*

$$\begin{aligned}\text{vega} = E\Bigg[&\Phi(\hat{X}_{i_1},\ldots,\hat{X}_{i_m})\cdot\Bigg\{\sum_{k=1}^{m}\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right] \\ &\cdot\sum_{i=i_{k-1}+1}^{i_k} a_i\frac{\hat{Y}_{i-1}}{\sigma(\hat{X}_{i-1})}\cdot\frac{\sqrt{\Delta t}\cdot\tilde{\xi}_i}{\delta} - \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\hat{Z}_{i_m}}{\hat{Y}_{i_m}}\right]\Bigg\}\Bigg]\end{aligned}$$

$$+E\left[\Phi(\hat{X}_{i_1},\ldots,\hat{X}_{i_m})\cdot\left\{\sum_{k=1}^{m}\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\sum_{i=i_{k-1}+1}^{i_k}a_i\Delta t\cdot(\mathrm{III}+\mathrm{IV})\right\}\right]$$

$$+\frac{C}{\sqrt{2\pi\delta}}\mathrm{e}^{-\frac{w_R^2}{2\delta}}\sum_{k=1}^{m}\sum_{i=i_{k-1}+1}^{i_k}a_i\sqrt{\Delta t}E\left[\Phi(\hat{X}_{i_1},\ldots,\hat{X}_{i_m})\cdot\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})}\bigg|_{\tilde{\xi}_i=w_R}\right]$$

$$-\frac{C}{\sqrt{2\pi\delta}}\mathrm{e}^{-\frac{w_L^2}{2\delta}}\sum_{k=1}^{m}\sum_{i=i_{k-1}+1}^{i_k}a_i\sqrt{\Delta t}E\left[\Phi(\hat{X}_{i_1},\ldots,\hat{X}_{i_m})\cdot\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})}\bigg|_{\tilde{\xi}_i=w_L}\right],$$

where

$$\mathrm{III}=\frac{\mu'(\hat{X}_{i-1})\hat{Y}_{i-1}}{\sigma(\hat{X}_{i-1})}\cdot\frac{\sqrt{\Delta t}\tilde{\xi}_i}{\delta};\qquad \mathrm{IV}=\frac{\sigma'(\hat{X}_{i-1})\hat{Y}_{i-1}}{\sigma(\hat{X}_{i-1})}\cdot\left[\frac{\tilde{\xi}_i^2}{\delta}-1\right]$$

and notation $f(x_1,\ldots,x_n)|_{x_j=w}$ denotes the value of function f when x_j is fixed as w.

Proof. We first show that this result holds for Φ with a continuous derivative. By the truncated normal assumption, the range of all possible values of $\hat{X}$ and $\hat{Z}$ must be within a compact set. Over this range, $\partial\Phi/\partial x$ is bounded because it is continuous. Therefore, we can take the derivative inside the expectation to get, by the bounded convergence theorem,

$$\text{vega}=\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\bigg|_{\varepsilon=0}E[\Phi(\hat{X}_{i_1}^{\varepsilon},\ldots,\hat{X}_{i_m}^{\varepsilon})]=E\left[\sum_{j=1}^{m}\frac{\partial\Phi}{\partial\hat{X}_{i_j}}(\hat{X}_{i_1},\ldots,\hat{X}_{i_m})\cdot\hat{Z}_{i_j}\right].\tag{25}$$

For any given (w_R, w_L, C, δ) satisfying (22)–(24), we always can find a big enough N such that

$$1-\sup_x|\mu'(x)|\Delta t-\sup_x|\sigma'(x)|\sqrt{\Delta t}w_R=1-\frac{\sup_x|\mu'(x)|T}{N}-\frac{\sup_x|\sigma'(x)|\sqrt{T}w_R}{\sqrt{N}}>0.$$

Under such choice of N, $\hat{Y}$ is positive almost surely because, for any i,

$$\begin{aligned}\hat{Y}_i&=\prod_{j=1}^{i}(1+\mu'(\hat{X}_{j-1})\Delta t+\sigma'(\hat{X}_{j-1})\sqrt{\Delta t}\tilde{\xi}_j)\\&\geq\prod_{j=1}^{i}(1-\sup_x|\mu'(x)|\Delta t-\sup_x|\sigma'(x)|\sqrt{\Delta t}w_R)>0.\end{aligned}$$

Now divide and multiply by $\hat{Y}$ simultaneously in the expectation of (25) to get

$$\text{vega} = E\left[\sum_{j=1}^{m} \frac{\partial \Phi}{\partial \hat{X}_{i_j}}(\hat{X}_{i_1}, \dots, \hat{X}_{i_m}) \cdot \frac{\hat{Z}_{i_j}}{\hat{Y}_{i_j}} \cdot \hat{Y}_{i_j}\right]. \tag{26}$$

For any weight set $\{a_i : 0 \le i \le N\}$ such that $\sum_{i=i_j+1}^{i_j} a_i \Delta t = 1$, notice that $\hat{Z}_0/\hat{Y}_0 = 0$,

$$\frac{\hat{Z}_{i_j}}{\hat{Y}_{i_j}} \cdot \hat{Y}_{i_j} = \sum_{k=1}^{j}\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right] \cdot \hat{Y}_{i_j} = \sum_{k=1}^{j}\left[\sum_{i=i_{k-1}+1}^{i_k} a_i \Delta t\right] \cdot \left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right] \cdot \hat{Y}_{i_j}. \tag{27}$$

Plugging (27) back into (26) and interchanging the order of the sum over j and the sums over k and i, we get

$$\text{vega} = \sum_{k=1}^{m} \sum_{i=i_{k-1}+1}^{i_k} a_i \Delta t \cdot E\left[\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right] \cdot \sum_{j=k}^{m} \frac{\partial \Phi}{\partial \hat{X}_{i_j}}(\hat{X}_{i_1}, \dots, \hat{X}_{i_m}) \cdot \frac{\hat{Y}_{i_j}}{\hat{Y}_i} \cdot \hat{Y}_i\right].$$

Notice that $\hat{Y}_{i_j}/\hat{Y}_i = (\mathrm{d}\hat{X}_{i_j}/\mathrm{d}x)/(\mathrm{d}\hat{X}_i/\mathrm{d}x) = \mathrm{d}\hat{X}_{i_j}/\mathrm{d}\hat{X}_i$, so

$$\text{vega} = \sum_{k=1}^{m} \sum_{i=i_{k-1}+1}^{i_k} a_i \Delta t \cdot E\left[\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right] \cdot \sum_{j=k}^{m} \frac{\partial \Phi}{\partial \hat{X}_{i_j}}(\hat{X}_{i_1}, \dots, \hat{X}_{i_m}) \cdot \frac{\mathrm{d}\hat{X}_{i_j}}{\mathrm{d}\hat{X}_i} \cdot \hat{Y}_i\right]. \tag{28}$$

Given a sample path of $\tilde{\xi}_1, \dots, \tilde{\xi}_{i-1}, \tilde{\xi}_{i+1}, \dots, \tilde{\xi}_N$ (all increments but $\tilde{\xi}_i$), $\hat{X}_{i_j}$ and $\hat{X}_i$ become functions of $\tilde{\xi}_i$. The implicit function theorem leads to

$$\frac{\mathrm{d}\hat{X}_{i_j}}{\mathrm{d}\hat{X}_i} = \frac{\mathrm{d}\hat{X}_{i_j}/\mathrm{d}\tilde{\xi}_i}{\mathrm{d}\hat{X}_i/\mathrm{d}\tilde{\xi}_i} = \frac{\mathrm{d}\hat{X}_{i_j}/\mathrm{d}\tilde{\xi}_i}{\sigma(\hat{X}_{i-1})\sqrt{\Delta t}}$$

and since $i_{k-1} < i \le i_k$,

$$\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right] \cdot \sum_{j=k}^{m} \frac{\partial \Phi}{\partial \hat{X}_{i_j}} \cdot \frac{\mathrm{d}\hat{X}_{i_j}}{\mathrm{d}\hat{X}_i} \cdot \hat{Y}_i = \left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right] \cdot \sum_{j=k}^{m} \frac{\partial \Phi}{\partial \hat{X}_{i_j}} \cdot \frac{\mathrm{d}\hat{X}_{i_j}}{\mathrm{d}\tilde{\xi}_i} \frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})\sqrt{\Delta t}}$$
$$= \left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right] \cdot \frac{\mathrm{d}\Phi}{\mathrm{d}\tilde{\xi}_i} \cdot \frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})\sqrt{\Delta t}}$$

by the chain rule $\mathrm{d}\Phi/\mathrm{d}\tilde{\xi}_i = \sum_{j=k}^{m}(\partial \Phi/\partial \hat{X}_{i_j}) \cdot (\mathrm{d}\hat{X}_{i_j}/\mathrm{d}\tilde{\xi}_i)$. Taking integrals on both sides of the above equality with respect to the density of $\tilde{\xi}_i$, we have

$$\int_{w_L}^{w_R} \frac{C}{\sqrt{2\pi\delta}} \mathrm{e}^{-\frac{w_i^2}{2\delta}} \cdot \left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right] \cdot \frac{\mathrm{d}\Phi}{\mathrm{d}w_i} \cdot \frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})\sqrt{\Delta t}} \mathrm{d}w_i$$
$$= E\left[\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right] \cdot \sum_{j=k}^{m} \frac{\partial \Phi}{\partial \hat{X}_{i_j}} \cdot \frac{\mathrm{d}\hat{X}_{i_j}}{\mathrm{d}\hat{X}_i} \cdot \hat{Y}_i \,|\, \tilde{\xi}_1, \dots, \tilde{\xi}_{i-1}, \tilde{\xi}_{i+1}, \dots, \tilde{\xi}_N\right].$$

Using the integration by parts formula, the left hand side integral is

$$\frac{C}{\sqrt{2\pi\delta}}\int_{w_L}^{w_R}\Phi e^{-\frac{w_i^2}{2\delta}}\cdot\left\{\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\frac{\hat{Y}_i}{\sqrt{\Delta t}\sigma(\hat{X}_{i-1})}\cdot\frac{w_i}{\delta}\right.$$
$$-\frac{d}{dw_i}\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\frac{\hat{Y}_i}{\sqrt{\Delta t}\sigma(\hat{X}_{i-1})}-\left.\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\hat{Y}_{i-1}\cdot\frac{\sigma'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right\}dw_i$$
$$+\frac{Ce^{-\frac{w_R^2}{2\delta}}}{\sqrt{2\pi\delta}}\left\{\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\Phi\cdot\frac{\hat{Y}_i}{\sqrt{\Delta t}\sigma(\hat{X}_{i-1})}\right|_{w_i=w_R}$$
$$\left.-\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\Phi\cdot\frac{\hat{Y}_i}{\sqrt{\Delta t}\sigma(\hat{X}_{i-1})}\right|_{w_i=w_L}\Bigg\}$$

where we view all of $\hat{X}, \hat{Y}$ and $\hat{Z}$ as functions of w_i and note that $d\hat{Y}_i/dw_i = \hat{Y}_{i-1}\cdot\sigma'(\hat{X}_{i-1})/\sigma(\hat{X}_{i-1})$. In addition, $d\hat{X}_i/dw_i = \sqrt{\Delta t}\sigma(\hat{X}_{i-1})$ and $\hat{Y}_i = d\hat{X}_i/dx$. So,

$$\frac{d}{dw_i}\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\frac{\hat{Y}_i}{\sqrt{\Delta t}\sigma(\hat{X}_{i-1})}=\frac{d}{d\hat{X}_i}\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\frac{d\hat{X}_i}{dx}=\frac{d}{dx}\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right].$$

In summary,

$$E\left[\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\sum_{j=k}^{m}\frac{\partial\Phi}{\partial\hat{X}_{i_j}}\cdot\frac{d\hat{X}_{i_j}}{d\hat{X}_i}\cdot\hat{Y}_i\right]$$
$$=E\left[E\left[\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\sum_{j=k}^{m}\frac{\partial\Phi}{\partial\hat{X}_{i_j}}\cdot\frac{d\hat{X}_{i_j}}{d\hat{X}_i}\cdot\hat{Y}_i\middle|\tilde{\xi}_1,\ldots,\tilde{\xi}_{i-1},\tilde{\xi}_{i+1},\ldots,\tilde{\xi}_N\right]\right]$$
$$=E\left[\Phi\cdot\left\{\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\frac{\tilde{\xi}_i}{\delta}\cdot\frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})\sqrt{\Delta t}}-\frac{d}{dx}\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\right\}\right]$$
$$-E\left[\Phi\cdot\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\hat{Y}_{i-1}\cdot\frac{\sigma'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right]+\frac{C}{\sqrt{2\pi\delta}}\cdot E\left[\left\{\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\right.\right.$$
$$\left.\left.\cdot\Phi\cdot\frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})\sqrt{\Delta t}}\right|_{\tilde{\xi}_i=w_R}-\left[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}}-\frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\right]\cdot\Phi\cdot\left.\frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})\sqrt{\Delta t}}\right|_{\tilde{\xi}_i=w_L}\right\}\cdot e^{-\frac{w_R^2}{2\delta}}\Bigg].$$

We get the result of the theorem by plugging the above equality back into (28).

To finish the proof, we can apply Lemma A.2 to extend the conclusion to a general payoff function Φ satisfying Assumption 3.1. □

3.3. Rho

In this subsection, we go back to normal increments ξ_i. Define rho to be the derivative of $u(x)$ with respect to a perturbation ε that takes the drift $\mu(\cdot)$ to $\mu(\cdot)+\varepsilon\gamma(\cdot)$. In other words,

$$\text{rho} = \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} E[\Phi(\hat{X}^{\varepsilon}_{i_1},\ldots,\hat{X}^{\varepsilon}_{i_m})]$$

where

$$\hat{X}_i^\varepsilon = \hat{X}_{i-1}^\varepsilon + [\mu(\hat{X}_{i-1}^\varepsilon) + \varepsilon\gamma(\hat{X}_{i-1}^\varepsilon)]\Delta t + \sigma(\hat{X}_{i-1}^\varepsilon)\sqrt{\Delta t}\cdot\xi_i.$$

Let U^ε be

$$U^\varepsilon = \exp\left[\varepsilon\sum_{i=1}^N \frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i - \frac{1}{2}\varepsilon^2\sum_{i=1}^N\left[\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right]^2\Delta t\right].$$

It is easy to check that $U^\varepsilon > 0$ and $E[U^\varepsilon] = 1$. Thus, we can define a new probability measure using U^ε, $P^\varepsilon(\mathrm{d}\omega) = U^\varepsilon P(\mathrm{d}\omega)$. Moreover, it is easy to see (cf. Lemma A.3) that this change of measure corresponds to a change of drift, in the sense that

$$E[\Phi(\hat{X}_{i_1}^\varepsilon, \ldots, \hat{X}_{i_m}^\varepsilon)] = E[\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m})\cdot U^\varepsilon]. \tag{29}$$

Theorem 3.3. *Suppose that Assumptions 3.1–3.3 hold. Then,*

$$\text{rho} = E\left[\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m})\cdot\sum_{i=1}^N \frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i\right].$$

Proof. Fix a compact neighborhood of $\varepsilon = 0$, say, $[-l, l]$. By (29) and the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\left|\frac{1}{\varepsilon}(E[\Phi(\hat{X}_{i_1}^\varepsilon, \ldots, \hat{X}_{i_m}^\varepsilon)] - u(x)) - E\left[\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m})\cdot\sum_{i=1}^N \frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i\right]\right| \\
&\quad = \left|E\left[\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m})\cdot\frac{1}{\varepsilon}(U^\varepsilon - 1)\right] - E\left[\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m})\cdot\sum_{i=1}^N \frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i\right]\right| \\
&\quad \le \left[E\left|\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m})\right|^2\right]^{1/2}\cdot\left[E\left|\frac{1}{\varepsilon}(U^\varepsilon - 1) - \sum_{i=1}^N \frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i\right|^2\right]^{1/2}.
\end{aligned}$$

By Taylor expansion with Cauchy's remainder and the mean-value theorem for integration, we can find $\theta(\varepsilon) \in [-l, l]$ (which may also depend on the values of ξ's) such that

$$\begin{aligned}
U^\varepsilon &= U^0 + \left.\frac{\mathrm{d}U^\varepsilon}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0}\cdot\varepsilon + \frac{1}{2}\int_0^\varepsilon \frac{\mathrm{d}^2U^x}{\mathrm{d}x^2}(\varepsilon - x)^2\mathrm{d}x \\
&= 1 + \varepsilon\left[\sum_{i=1}^N \frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i\right] + U^{\theta(\varepsilon)}\cdot(\varepsilon - \theta(\varepsilon)) \\
&\quad \cdot\left[\left(\sum_{i=1}^N \frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i - \theta(\varepsilon)\sum_{i=1}^N\left[\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right]^2\Delta t\right)^2\right. \\
&\quad \left. - \sum_{i=1}^N\left[\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right]^2\Delta t\right]\cdot\varepsilon.
\end{aligned}$$

Under Assumptions 3.2 and 3.3, γ and $1/\sigma$ are bounded, and we clearly have $\theta(\varepsilon) \in [-l, l]$. It follows that there are positive constants C_1 and C_2 (not dependent on ξ) such that

$$U^{\theta(\varepsilon)} = \exp\left[\theta(\varepsilon)\sum_{i=1}^{N}\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i - \frac{1}{2}\theta(\varepsilon)^2\sum_{i=1}^{N}\left[\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right]^2\Delta t\right]$$
$$\leq \exp\left[C_1\sqrt{\Delta t}\sum_{i=1}^{N}|\xi_i| + C_2\sum_{i=1}^{N}\Delta t\right]. \tag{30}$$

Furthermore, there also exist positive constants C_3, C_4 (not dependent on ξ) such that

$$\left(\sum_{i=1}^{N}\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i - \theta(\varepsilon)\sum_{i=1}^{N}\left[\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right]^2\Delta t\right)^2 - \sum_{i=1}^{N}\left[\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right]^2\Delta t$$
$$\leq C_3\Delta t\sum_{i=1}^{N}\xi_i^2 + C_4\sum_{i=1}^{N}\Delta t. \tag{31}$$

With (30) and (31), using the exponential decay of the density of the ξs, it is easy to show that

$$\sup_{\varepsilon\in[-l,l]} E\left|U^{\theta(\varepsilon)}\cdot\left[\left(\sum_{i=1}^{N}\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i - \theta(\varepsilon)\sum_{i=1}^{N}\left[\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right]^2\Delta t\right)^2 - \sum_{i=1}^{N}\left[\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right]^2\Delta t\right]\right|^2 < +\infty,$$

which implies

$$E\left|\frac{1}{\varepsilon}(U^{\varepsilon}-1) - \sum_{i=1}^{N}\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i\right|^2$$
$$\leq \varepsilon\cdot\sup_{\varepsilon\in[-l,l]} E\left|U^{\theta(\varepsilon)}\cdot\left[\left(\sum_{i=1}^{N}\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i - \theta(\varepsilon)\sum_{i=1}^{N}\left[\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right]^2\Delta t\right)^2 - \sum_{i=1}^{N}\left[\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\right]^2\Delta t\right]\right|^2 \to 0.$$

Combining this with the fact that $E[\Phi^2] < \infty$, we have shown that

$$\text{rho} = \left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} E[\Phi(\hat{X}^{\varepsilon}_{i_1},\ldots,\hat{X}^{\varepsilon}_{i_m})] = E\left[\Phi(\hat{X}_{i_1},\ldots,\hat{X}_{i_m})\cdot\sum_{i=1}^{N}\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\sqrt{\Delta t}\xi_i\right].$$

□

4. Convergence results

In this section, we show the consistency between the estimators obtained in Section 3 and the Malliavin estimators. We prove that each of the unbiased estimators converges in distribution to the corresponding Malliavin estimator as $N \to +\infty$. The theoretical cornerstone of this analysis is the theory of weak convergence for stochastic differential equations. We review the necessary results in Appendix B. For further background, see Jacod and Shiryayev [11] and Kurtz and Protter [13,14]. In the section, we attach index N to such notations as $\hat{X}$, $\hat{Y}$ and $\hat{Z}$ to stress that they are under N-step Euler approximation.

For each positive integer N, there is a probability space $(\Omega^{(N)}, \mathcal{F}^{(N)}, \{\mathcal{F}_t^{(N)}\}, P^{(N)})$ and on it defined a sequence of i.i.d. random variables $\{\xi_i^{(N)}, 1 \le i \le N\}$ or $\{\tilde{\xi}_i^{(N)}, 1 \le i \le N\}$, where $\mathcal{F}_t^{(N)} = \sigma(\xi_1^{(N)}, \ldots, \xi_{[Nt/T]}^{(N)})$ or $\sigma(\tilde{\xi}_1^{(N)}, \ldots, \tilde{\xi}_{[Nt/T]}^{(N)})$. The distributions of $\xi_1^{(N)}$ and $\tilde{\xi}_1^{(N)}$ are standard normal or truncated normal, respectively. Define

$$\hat{L}_t^{(N)} = \sum_{i=1}^{[Nt/T]} \sqrt{\Delta t^{(N)}}\xi_i^{(N)} \quad \text{and} \quad \tilde{L}_t^{(N)} = \sum_{i=1}^{[Nt/T]} \sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}.$$

Thus, $\hat{L}^{(N)}$ and $\tilde{L}^{(N)}$ induce two different sequences of probability measures on the space $\mathcal{D}^1[0, T]$ (cf. Appendix B for the definition of the space $\mathcal{D}^1[0, T]$).

To show the estimator of vega will converge to the Malliavin estimator, we need to get rid of the last two terms in the expression in Theorem 3.2. For this purpose, we choose the parameters of the truncated normal as follows:

Assumption 4.1. $(w_L^{(N)}, w_R^{(N)}, C^{(N)}, \delta^{(N)})$ satisfy

$$C^{(N)} = \frac{1}{2\psi(\sqrt{N}x^*) - 1}, \qquad \delta^{(N)} = \frac{1}{1 - F(\sqrt{N}x^*)},$$
$$w_R^{(N)} = -w_L^{(N)} = \sqrt{N}x^*\delta^{(N)},$$

where $x^* > 0$ is a fixed number such that

$$1 - \sup_x |\sigma'(x)|\sqrt{T}x^* > 0,$$

ψ is the cumulative probability function of standard normal and F is defined as in Lemma A.1. Thus, as $N \to +\infty$, $C^{(N)}$ and $\delta^{(N)}$ converge to 1 and $w_R^{(N)}$ goes to $+\infty$ in the order of $\sqrt{N}$.

Remark 4.1. By Lemma A.1, it is easy to show that $(w_L^{(N)}, w_R^{(N)}, C^{(N)}, \delta^{(N)})$ must satisfy (22)–(24). Meanwhile, for all sufficiently large N,

$$1 - \sup_x |\mu'(x)|\Delta t^{(N)} - \sup_x |\sigma'(x)|\sqrt{\Delta t^{(N)}}w_R^{(N)}$$
$$= 1 - \frac{\sup_x |\mu'(x)|T}{N} - \sup_x |\sigma'(x)|\sqrt{T}x^*\delta^{(N)} > 0.$$

Thus, all $Y^{(N)}$ are indeed positive, almost surely, and the proof of Theorem 3.2 goes through for all sufficiently large N.

4.1. Preliminary convergence results

As the first step, we establish the weak convergence result of processes $(\hat{X}^{(N)}, \hat{Y}^{(N)}, \hat{Z}^{(N)})$ with increments $\xi^{(N)}$ or $\tilde{\xi}^{(N)}$ in this subsection. For the purpose, we introduce functions

$$g^{(N)}(x, y, z) = g(x, y, z) = \begin{bmatrix} \mu(x) & \sigma(x) \\ \mu'(x)y & \sigma'(x)y \\ \mu'(x)z & \sigma'(x)z + \tilde{\sigma}(x) \end{bmatrix}.$$

One can easily see that $(\hat{X}^{(N)}, \hat{Y}^{(N)}, \hat{Z}^{(N)})$ is the solution to the following SDEs:

$$\hat{M}_t^{(N)} = \int_0^t g^{(N)}(\hat{M}_{s-}^{(N)})\mathrm{d}\hat{L}_s^{(N)} \quad \text{or} \quad \tilde{M}_t^{(N)} = \int_0^t g^{(N)}(\tilde{M}_{s-}^{(N)})\mathrm{d}\tilde{L}_s^{(N)}, \tag{32}$$

dependent on which increments we are using. The sample paths of both $\hat{M}^{(N)}$ and $\tilde{M}^{(N)}$ are in $\mathcal{D}^3[0, T]$.

The following lemma is needed to establish the tightness of processes $\hat{L}^{(N)}$ and $\tilde{L}^{(N)}$ (see Appendix B for the definition of P-UT):

Lemma 4.2. *Processes $\hat{L}^{(N)}$ and $\tilde{L}^{(N)}$ are P-UT.*

Proof. In the following proof, we do not need any special properties of the normal (or truncated normal) distribution except that $E[\xi] = 0$ and $\mathrm{Var}[\xi] = 1$. So, without loss of generality, we only consider the case of $\hat{L}^{(N)}$. For any elementary predictable process

$$H_t^{(N)} = Y_0^{(N)} 1_{\{0\}} + \sum_{i=1}^{k} Y_i^{(N)} 1_{(s_i, s_{i+1}]}(t)$$

defined on the probability space $(\Omega^{(N)}, \mathcal{F}^{(N)}, \{\mathcal{F}_t^{(N)}\}, P^{(N)})$ and satisfying $|Y_i^{(N)}| \le 1$ for all $1 \le i \le k$ and for any $a > 0$, by Chebyshev's inequality,

$$P\left(\left|\int_0^t H_{s-}^{(N)}\mathrm{d}\hat{L}_s^{(N)}\right| > a\right) \le \frac{E[\int_0^t H_{s-}^{(N)}\mathrm{d}\hat{L}_s^{(N)}]^2}{a^2} = \frac{\sum_{i=1}^{k} E[|Y_i^{(N)}|^2 \cdot |\hat{L}_{s_{i+1}}^{(N)} - \hat{L}_{s_i}^{(N)}|^2]}{a^2}.$$

By the assumption that $|Y_i^{(N)}| \le 1$, the right side is bounded above by

$$\frac{\sum_{i=1}^{k} E[|\hat{L}_{s_{i+1}}^{(N)} - \hat{L}_{s_i}^{(N)}|^2]}{a^2} = \frac{\Delta t^{(N)}}{a^2} \sum_{j=1}^{[Nt/T]} \mathrm{Var}[\xi_j^{(N)}].$$

Because the variance of $\xi_j^{(N)}$ is 1, the right side of the inequality above will be $\Delta t^{(N)}[Nt/T]/a^2$. This upper bound does not depend on H and N, and it converges to 0 as $a \to +\infty$. Thus,

$$\lim_{a \to +\infty} \sup_{H^{(N)} \in \mathcal{H}^{(N)}, N} P\left(\left|\int_0^t H_{s-}^{(N)}\mathrm{d}\hat{L}_s^{(N)}\right| > a\right) \to 0.$$

where $\mathcal{H}^{(N)}$ is the set of all elementary processes on $(\Omega^{(N)}, \mathcal{F}^{(N)}, \{\mathcal{F}_t^{(N)}\}, P^{(N)})$, i.e., $\hat{L}^{(N)}$ is P-UT. □

Now we are able to show the weak convergence of $(\hat{X}^{(N)}, \hat{Y}^{(N)}, \hat{Z}^{(N)})$ in the Skorokhod topology:

Lemma 4.3. *Suppose Assumptions 3.1, 3.2 and 4.1 hold and we have $(\hat{X}^{(N)}, \hat{Y}^{(N)}, \hat{Z}^{(N)})$ defined as the solution to the SDE (32). As $N \to +\infty$,*

$$(\hat{X}^{(N)}, \hat{Y}^{(N)}, \hat{Z}^{(N)}) \Rightarrow (X, Y, Z),$$

where (X, Y, Z) is a global solution of the following SDE:

$$M_t = \int_0^t g(M_{s-})\mathrm{d}W_s.$$

Proof. Suppose that $(\hat{X}^{(N)}, \hat{Y}^{(N)}, \hat{Z}^{(N)})$ is the solution to the first SDE in (32). Using Donsker's invariance principle for $\hat{L}^{(N)}$, we know that $\hat{L}^{(N)}$ weakly converge to Brownian motion W. Furthermore, $\hat{L}^{(N)}$ are P-UT by Lemma 4.2. Thus, applying Lemma B.3, we have the conclusion that $(\hat{X}^{(N)}, \hat{Y}^{(N)}, \hat{Z}^{(N)}) \Rightarrow (X, Y, Z)$ in the Skorokhod topology if the solution (X, Y, Z) exists globally.

As for the case that $(\hat{X}^{(N)}, \hat{Y}^{(N)}, \hat{Z}^{(N)})$ is the solution for the second SDE in (32). All of the above arguments works except that now we can not use Donsker's invariance principle to show $\tilde{L}^{(N)}$ converges to Brownian motion because $\tilde{L}^{(N)}$ is the sum of a triangular array (the distribution of $\tilde{\xi}_1^{(N)}$ depends on N by Assumption 4.1). Here we cite the Lindeberg–Feller theorem for martingales (cf. Durrett [5], Theorem 7.7.3, p. 414) to establish such convergence. Indeed, for any $t \in [0, T]$, as $N \to +\infty$,

$$\sum_{j=1}^{[Nt/T]} E[\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}]^2 = \Delta t^{(N)} \sum_{j=1}^{[Nt/T]} E[\tilde{\xi}_i^{(N)}]^2 = \Delta t^{(N)} \cdot [Nt/T] \to t$$

and for any $\epsilon > 0$,

$$\begin{aligned}\sum_{j=1}^{N} E[(\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)})^2 \cdot 1_{\{|\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}|>\epsilon\}}] &\le \frac{1}{\epsilon^2}\sum_{j=1}^{N} E[(\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)})^4] \\ &= \frac{N}{\epsilon^2}(\Delta t^{(N)})^2 E[(\tilde{\xi}_1^{(N)})^4].\end{aligned}$$

Note that $E[(\tilde{\xi}_1^{(N)})^4]$ is uniformly bounded for all N. The right hand side of the above inequality converges to 0. In summary, both (i) and (ii) of Theorem 7.7.3 are satisfied. We have the conclusion that $\tilde{L}^{(N)}$ weakly converges to Brownian motion W. □

Remark 4.4. This lemma implies that $\hat{Y}^{(N)}$ is a tight process. In other words, we have the following limit holds (cf. [11], p. 350): for any $K > 0$,

$$\lim_N P^{(N)}(\sup_{1\le i\le N} |\hat{Y}_i^{(N)}| > K) = 0.$$

4.2. Convergence of delta estimators

To show the convergence of estimators for delta, vega and rho, we need another two assumptions:

Assumption 4.2. Φ is a.s. continuous under the measure of X_t, i.e., there exists a set $E \subset \mathbf{R}^m$ such that Φ is continuous outside of E and

$$P((X_{t_1}, \ldots, X_{t_m}) \in E) = 0.$$

Assumption 4.3. There exists an L^2 function $\{a_t : 1 \le t \le T\}$ for which the sequence of weights $\{a_i^{(N)} : 1 \le i \le N\}$ satisfies

$$\lim_{N\to+\infty} \sup_{0\le t\le T} |a^{(N)}_{[Nt/T]} - a_t| = 0.$$

Lemma 4.5. *Under Assumptions* 3.1, 3.2, 4.2 *and* 4.3,

$$\mathrm{I}^{(N)} := \Phi(\hat{X}^{(N)}_{i_1}, \ldots, \hat{X}^{(N)}_{i_m}) \cdot \sum_{i=1}^{N} a_i^{(N)} \cdot \frac{\sigma'(\hat{X}^{(N)}_{i-1})}{\sigma(\hat{X}^{(N)}_{i-1})} \cdot [(\sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)})^2 - \Delta t^{(N)}]$$

and

$$\mathrm{II}^{(N)} := \Phi(\hat{X}^{(N)}_{i_1}, \ldots, \hat{X}^{(N)}_{i_m}) \cdot \left(\sum_{i=1}^{N} a_i^{(N)} \cdot \frac{\mu'(\hat{X}^{(N)}_{i-1})}{\sigma(\hat{X}^{(N)}_{i-1})} \sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)} \right) \cdot \Delta t^{(N)}$$

converge weakly to 0.

Proof. Notice that

$$E\left|\sum_{i=1}^{N} a_i^{(N)} \cdot \frac{\sigma'(\hat{X}^{(N)}_{i-1})}{\sigma(\hat{X}^{(N)}_{i-1})} \cdot [(\sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)})^2 - \Delta t^{(N)}]\right|^2$$
$$= 2\sum_{i=1}^{N}\sum_{j=1}^{i-1} a_i^{(N)} a_j^{(N)} (\Delta t^{(N)})^2 E\left[\frac{\sigma'(\hat{X}^{(N)}_{i-1})}{\sigma(\hat{X}^{(N)}_{i-1})}\frac{\sigma'(\hat{X}^{(N)}_{j-1})}{\sigma(\hat{X}^{(N)}_{j-1})}((\xi_i^{(N)})^2 - 1)((\xi_j^{(N)})^2 - 1)\right]$$
$$+ \sum_{i=1}^{N} (a_i^{(N)} \Delta t^{(N)})^2 E\left[\left(\frac{\sigma'(\hat{X}^{(N)}_{i-1})}{\sigma(\hat{X}^{(N)}_{i-1})}\right)^2 ((\xi_i^{(N)})^2 - 1)^2\right]. \tag{33}$$

The expectation in the first summand in the right hand side of (33) is 0 because, conditional on $\mathcal{F}_{i-1}$,

$$E\left[\frac{\sigma'(\hat{X}^{(N)}_{i-1})}{\sigma(\hat{X}^{(N)}_{i-1})}\frac{\sigma'(\hat{X}^{(N)}_{j-1})}{\sigma(\hat{X}^{(N)}_{j-1})}((\xi_i^{(N)})^2 - 1)((\xi_j^{(N)})^2 - 1)\middle| \mathcal{F}_{i-1}\right]$$
$$= \frac{\sigma'(\hat{X}^{(N)}_{i-1})}{\sigma(\hat{X}^{(N)}_{i-1})}\frac{\sigma'(\hat{X}^{(N)}_{j-1})}{\sigma(\hat{X}^{(N)}_{j-1})}((\xi_j^{(N)})^2 - 1)E[((\xi_i^{(N)})^2 - 1)|\mathcal{F}_{i-1}] = 0$$

for all $i, j, i > j$. And the second summand in the right hand side of (33) is bounded by:

$$\sup_x \left(\frac{\sigma'(x)}{\sigma(x)}\right)^2 \sum_{i=1}^{N} (a_i^{(N)} \Delta t^{(N)})^2 E[((\xi_i^{(N)})^2 - 1)^2]$$
$$= 2\left(\frac{\sigma'(x)}{\sigma(x)}\right)^2 \sum_{i=1}^{N} (a_i^{(N)})^2 \Delta t^{(N)} \cdot \Delta t^{(N)},$$

because the assumption that ξ has a standard normal distribution leads to $E[(\xi_i^{(N)})^2 - 1]^2 = 2$. On the other hand, according to Assumption 4.3, $\sup_N \sum_{i=1}^N (a_i^{(N)})^2 \Delta t^{(N)} < +\infty$. Thus,

$$\sum_{i=1}^N a_i^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)})}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot [(\sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)})^2 - \Delta t^{(N)}]$$

converges to 0 in L^2 as $N \to \infty$, which also implies its weak convergence to 0. In addition, we know that process $\hat{X}^{(N)}$ weakly converge to X in the Skorohod topology (cf. Lemma 4.3). Combining this with Assumption 4.2 shows that $\Phi(\hat{X}_{i_1}^{(N)}, \ldots, \hat{X}_{i_m}^{(N)})$ converges to $\Phi(X_{t_1}, \ldots, X_{t_m})$ weakly. Therefore $\mathrm{I}^{(N)}$ converges weakly to 0.

As for $\mathrm{II}^{(N)}$, by applying Lemma B.3, we can show that

$$\sum_{i=1}^N a_i^{(N)} \cdot \frac{\mu'(\hat{X}_{i-1}^{(N)})}{\sigma(\hat{X}_{i-1}^{(N)})} \sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)}$$

converges weakly because we know that the processes $\hat{X}^{(N)}$ converge weakly by Lemma 4.3. Thus, $(\sum_{i=1}^N a_i^{(N)} \cdot (\mu'(\hat{X}_{i-1}^{(N)})/\sigma(\hat{X}_{i-1}^{(N)}))\sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)}) \cdot \Delta t^{(N)}$ converges to 0 weakly. Using Assumption 4.2 again, we find that $\mathrm{II}^{(N)}$ converge weakly to 0. □

We now come to the main theorem of this subsection.

Theorem 4.6. *Under Assumptions 3.1, 3.2, 4.1 and 4.2, the delta estimator (19) converges weakly to*

$$\Phi(X_{t_1}, \ldots, X_{t_m}) \cdot \int_0^T a_t \frac{Y_t}{\sigma(X_t)} \mathrm{d}W_t.$$

Proof. Define functions

$$g^{(N)}(x, y; s) = \frac{y}{\sigma(x)} \cdot a_{[Ns/T]}^{(N)}, \qquad g(x, y; s) = \frac{y}{\sigma(x)} \cdot a_s.$$

It is easy to see that for any compact set K, we have

$$\begin{aligned} &\lim_{N\to+\infty} \sup_{(x,y,s)\in K} |g^{(N)}(x, y; s) - g(x, y; s)| \\ &\quad = \sup_{(x,y)\in K} \frac{y}{\sigma(x)} \cdot \lim_{N\to+\infty} \sup_{s\in K} |a_{[Ns/T]}^{(N)} - a_s| = 0 \end{aligned}$$

because $y/\sigma(x)$ is bounded when $(x, y) \in K$. Thus (41) holds. Applying Lemma B.3, we have

$$\sum_{i=1}^N a_i^{(N)} \cdot \frac{\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot \sqrt{\Delta t^{(N)}} \xi_i^{(N)} \Rightarrow \int_0^T a_t \frac{Y_t}{\sigma(X_t)} \mathrm{d}W_t.$$

Combining this with Lemma 4.5, the theorem follows. □

4.3. Convergence of vega estimators

This section establishes the convergence of the vega estimators we derived in Section 3.

Lemma 4.7. *Under Assumptions 3.1, 3.2 and 4.1–4.3,*

$$\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \mathrm{III}^{(N)} := \sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \frac{\mu'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot \frac{\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}}{\delta^{(N)}}$$

and

$$\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \mathrm{IV}^{(N)} := \sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot \left[\frac{(\tilde{\xi}_i^{(N)})^2}{\delta^{(N)}} - 1\right]$$

converge to 0 *weakly.*

Proof. Using similar arguments as those in Theorem 4.6, it is easy to show that

$$\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \cdot \frac{\mu'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot \sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)} \Rightarrow \int_{t_{k-1}}^{t_k} a_s \frac{\mu'(X_s)Y_s}{\sigma(X_s)} \mathrm{d}W_s.$$

On the other hand, by Assumption 4.1, $\delta^{(N)}$ converge to 1. Then we have

$$\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \mathrm{III}^{(N)} = \frac{\Delta t^{(N)}}{\delta^{(N)}} \left[\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \cdot \frac{\mu'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot \sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}\right]. \Rightarrow 0.$$

As for the second convergence in the statement of the lemma, we first show

$$\lim_{N\to+\infty} E\left[\left|\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot \left[\frac{(\tilde{\xi}_i^{(N)})^2}{\delta^{(N)}} - 1\right] \cdot 1_{\{\tau_b^{(N)} \geq i\}}\right|^2\right] = 0,$$

for any fixed $b > 0$, where $\tau_b^{(N)} := \inf\{j : |\hat{Y}_j^{(N)}| > b\}$. Once we establish this limit, we can conclude that $\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \mathrm{IV}^{(N)} \Rightarrow 0$, in light of Remark 4.4 (following Lemma 4.3) noting that $P^{(N)}(\tau_b^{(N)} \geq N) \to 1$.

Indeed, the left hand side of the above limit is bounded by, using the inequality $(a + b)^2 \leq 2a^2 + 2b^2$,

$$\frac{2}{(\delta^{(N)})^2} E\left[\left|\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot [(\tilde{\xi}_i^{(N)})^2 - 1] \cdot 1_{\{\tau_b^{(N)} \geq i\}}\right|^2\right] + 2\left(\frac{1}{(\delta^{(N)})^2} - 1\right)^2 \cdot E\left[\left|\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot 1_{\{\tau_b^{(N)} \geq i\}}\right|^2\right]. \quad (34)$$

It is easy to see that the second term in (34) converges to 0 because $\delta^{(N)} \to 1$ and

$$E\left[\left|\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot 1_{\{\tau_b^{(N)} \geq i\}}\right|^2\right]$$

$$\leq N \cdot \sum_{i=i_{k-1}+1}^{i_k} E\left[(a_i^{(N)})^2(\Delta t^{(N)})^2 \cdot \left[\frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})}\right]^2 \cdot 1_{\{\tau_b^{(N)} \geq i\}}\right]$$

$$\leq \sup_x \left(\frac{\sigma'(x)b}{\sigma(x)}\right)^2 \sum_{i=i_{k-1}+1}^{i_k} (a_i^{(N)})^2(\Delta t^{(N)}) \cdot T$$

is bounded uniformly for all N. For the first term in (34),

$$a_i^{(N)}\Delta t^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot [(\tilde{\xi}_i^{(N)})^2 - 1] \cdot 1_{\{\tau_b^{(N)} \geq i\}}$$

is a martingale difference sequence because the variance of $\tilde{\xi}_i^{(N)}$ is 1 and $\Delta t^{(N)} \cdot \sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}/\sigma(\hat{X}_{i-1}^{(N)}) \cdot 1_{\{\tau_b^{(N)} \geq i\}}$ is $\mathcal{F}_{i-1}^{(N)}$-measurable. By the orthogonality of martingale differences and the assumption that σ' is bounded and σ is non-degenerate,

$$E\left[\left|\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)}\Delta t^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot [(\tilde{\xi}_i^{(N)})^2 - 1] \cdot 1_{\{\tau_b^{(N)} \geq i\}}\right|^2\right]$$

$$= \Delta t^{(N)} \cdot \sum_{i=i_{k-1}+1}^{i_k} |a_i^{(N)}|^2 E\left[\left|\frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot [(\tilde{\xi}_i^{(N)})^2 - 1] \cdot 1_{\{\tau_b^{(N)} \geq i\}}\right|^2 \cdot \Delta t^{(N)}\right]$$

$$\leq Cb^2\Delta t^{(N)} \cdot \sum_{i=i_{k-1}+1}^{i_k} |a_i^{(N)}|^2 E[|(\tilde{\xi}_i^{(N)})^2 - 1|^2] \cdot \Delta t^{(N)}. \tag{35}$$

Noting that the variance of $\tilde{\xi}_i^{(N)}$ is 1 and the fourth moment of $\tilde{\xi}_i^{(N)}$ is also bounded by some constant independent of N, we find that $E[|(\tilde{\xi}_i^{(N)})^2 - 1|^2]$ is bounded uniformly in N. Thus the right hand side of (35) converges to 0, which implies the convergence of the first term of (34) to 0. □

Lemma 4.8. *Under all assumptions in Section 3 and Assumptions 4.1–4.3, there exists a constant C which does not depend on N such that, for every $1 \leq k \leq m$,*

$$\sup_{i_{k-1}+1 \leq i \leq i_k} E\left[\Phi(\hat{X}_{i_1}^{(N)}, \ldots, \hat{X}_{i_m}^{(N)}) \cdot \left[\frac{\hat{Z}_{i_k}^{(N)}}{\hat{Y}_{i_k}^{(N)}} - \frac{\hat{Z}_{i_{k-1}}^{(N)}}{\hat{Y}_{i_{k-1}}^{(N)}}\right] \cdot \frac{\hat{Y}_i^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})}\bigg|_{\tilde{\xi}_i^{(N)} = w_R^{(N)} (or\ w_L^{(N)})}\right]$$
$$< CN.$$

Therefore,

$$\frac{C^{(N)}}{\sqrt{2\pi\delta^{(N)}}} \exp\left(-\frac{(w_R^{(N)})^2}{2\delta^{(N)}}\right) \sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)}\sqrt{\Delta t^{(N)}} E$$
$$\times \left[\Phi^{(N)} \cdot \left[\frac{\hat{Z}_{i_k}^{(N)}}{\hat{Y}_{i_k}^{(N)}} - \frac{\hat{Z}_{i_{k-1}}^{(N)}}{\hat{Y}_{i_{k-1}}^{(N)}}\right] \cdot \frac{\hat{Y}_i^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})}\bigg|_{\tilde{\xi}_i^{(N)} = w_R^{(N)} (or\ w_L^{(N)})}\right]$$

is bounded by $C_1 N^{3/2} \cdot \exp(-C_2 N)$ *for some constant* C_1 *and* C_2 *which do not depend on* N *either, and thus it converges to* 0 *as* $N \to +\infty$. *(Here,* $\Phi^{(N)} = \Phi(\hat{X}_{i_1}^{(N)}, \ldots, \hat{X}_{i_m}^{(N)})$.*)*

Proof. We consider only the conditional expectation on the condition that $\tilde{\xi}_i^{(N)} = w_R^{(N)}$ because the arguments are the same for both cases of $w_R^{(N)}$ and $w_L^{(N)}$.

By the definition of $\hat{Y}^{(N)}$ and $\hat{Z}^{(N)}$ (cf. the part before Theorem 3.2 in Section 3.2), we know that

$$\begin{aligned}\frac{\hat{Z}_j^{(N)}}{\hat{Y}_j^{(N)}} &= \frac{\hat{Z}_{j-1}^{(N)}}{\hat{Y}_j^{(N)}} \cdot (1 + \mu'(\hat{X}_{j-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{j-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_j^{(N)}) + \frac{\tilde{\sigma}(\hat{X}_{j-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_j^{(N)}}{Y_j^{(N)}} \\ &= \frac{\hat{Z}_{j-1}^{(N)}}{\hat{Y}_j^{(N)}} \cdot \frac{\hat{Y}_j^{(N)}}{\hat{Y}_{j-1}^{(N)}} + \frac{\tilde{\sigma}(\hat{X}_{j-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_j^{(N)}}{Y_j^{(N)}} = \frac{\hat{Z}_{j-1}^{(N)}}{\hat{Y}_{j-1}^{(N)}} + \frac{\tilde{\sigma}(\hat{X}_{j-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_j^{(N)}}{Y_j^{(N)}}.\end{aligned}$$

Thus,

$$\frac{\hat{Z}_{i_k}^{(N)}}{\hat{Y}_{i_k}^{(N)}} - \frac{\hat{Z}_{i_{k-1}}^{(N)}}{\hat{Y}_{i_{k-1}}^{(N)}} = \sum_{j=i_{k-1}+1}^{i_k} \left[\frac{\hat{Z}_j^{(N)}}{\hat{Y}_j^{(N)}} - \frac{\hat{Z}_{j-1}^{(N)}}{\hat{Y}_{j-1}^{(N)}}\right] = \sum_{j=i_{k-1}+1}^{i_k} \frac{\tilde{\sigma}(\hat{X}_{j-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_j^{(N)}}{Y_j^{(N)}}.$$

Because σ is non-degenerate and $|\tilde{\sigma}(x)| \le |\tilde{\sigma}(0)| + \sup_x |\tilde{\sigma}'(x)| \cdot |x|$, there must exists two constants C_3 and C_4 which does not depend on N such that

$$\begin{aligned}&E\left[\left|\Phi(\hat{X}_{i_1}^{(N)}, \ldots, \hat{X}_{i_m}^{(N)}) \cdot \left[\frac{\hat{Z}_{i_k}^{(N)}}{\hat{Y}_{i_k}^{(N)}} - \frac{\hat{Z}_{i_{k-1}}^{(N)}}{\hat{Y}_{i_{k-1}}^{(N)}}\right] \cdot \frac{\hat{Y}_i^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})}\right|\Bigg|_{\tilde{\xi}_i^{(N)} = w_R^{(N)}}\right] \\ &= \sum_{j=i_{k-1}+1}^{i_k} E\left[\left|\Phi(\hat{X}_{i_1}^{(N)}, \ldots, \hat{X}_{i_m}^{(N)}) \cdot \frac{\tilde{\sigma}(\hat{X}_{j-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_j^{(N)}}{Y_j^{(N)}} \cdot \frac{\hat{Y}_i^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})}\right|\Bigg|_{\tilde{\xi}_i^{(N)} = w_R^{(N)}}\right] \\ &\le \sum_{j=i_{k-1}+1}^{i_k} E\left[|\Phi(\hat{X}_{i_1}^{(N)}, \ldots, \hat{X}_{i_m}^{(N)})| \cdot \frac{\sqrt{\Delta t^{(N)}}(C_3 + C_4|\hat{X}_{j-1}^{(N)}|)|\tilde{\xi}_j^{(N)}|}{|\hat{Y}_j^{(N)}|}\right. \\ &\qquad \left. \cdot |\hat{Y}_i^{(N)}|\Bigg|_{\tilde{\xi}_i^{(N)} = w_R^{(N)}}\right].\end{aligned}$$

Using the arithmetic–geometric mean inequality, for any positive numbers a, b, c, d, we know that $abcd \le (a^4 + b^4 + c^4 + d^4)/4$. So the right hand side of the above inequality is less than

$$\begin{aligned}\frac{1}{4}\sum_{j=i_{k-1}+1}^{i_k} &\Big\{E[|\Phi(\hat{X}_{i_1}^{(N)}, \ldots, \hat{X}_{i_m}^{(N)})|^4|_{\tilde{\xi}_i^{(N)} = w_R^{(N)}}] \\ &+ (\Delta t^{(N)})^2 E[|C_3 + C_4|\hat{X}_{j-1}^{(N)}||^4 |\tilde{\xi}_j^{(N)}|^4|_{\tilde{\xi}_i^{(N)} = w_R^{(N)}}] \\ &+ E[|\hat{Y}_j^{(N)}|^{-4}|_{\tilde{\xi}_i^{(N)} = w_R^{(N)}}] + E[|\hat{Y}_i^{(N)}|^4|_{\tilde{\xi}_i^{(N)} = w_R^{(N)}}]\Big\}. \end{aligned} \tag{36}$$

By Lemmas C.3 and C.2, the last two terms in (36) are both bounded by constants independent of N. The polynomial growth of Φ implies that

$$E[|\Phi(\hat{X}_{i_1}^{(N)}, \dots, \hat{X}_{i_m}^{(N)})|^4|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}] \leq E\left[\left.\sum_{k=1}^{m}(\hat{X}_{i_k}^{(N)})^2\right|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}\right]^{2p}$$
$$\leq m^{2p}\sum_{k=1}^{m} E[(\hat{X}_{i_k}^{(N)})^{4p}|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}].$$

By Lemma C.1, we know that it is also bounded by a constant independent of N.

For

$$(\Delta t^{(N)})^2 E[|\hat{X}_{j-1}^{(N)}|^4|\tilde{\xi}_j^{(N)}|^4|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}],$$

we consider two cases. If $j \neq i$, noting that $\hat{X}_{j-1}^{(N)}$ and $\tilde{\xi}_j^{(N)}$ are independent, this equals

$$(\Delta t^{(N)})^2 E[|\hat{X}_{j-1}^{(N)}|^4|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}] \cdot E[|\tilde{\xi}_j^{(N)}|^4]$$
$$\leq \sup_{1\leq j\leq N} E|\tilde{\xi}_j^{(N)}|^4 \cdot (\Delta t^{(N)})^2 \cdot E[|\hat{X}_{j-1}^{(N)}|^4|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}],$$

which is bounded by a constant by Lemma C.1 and the fact that $\sup_{1\leq j\leq N} E|\tilde{\xi}_j^{(N)}|^4$ is uniformly bounded for all N; if $j = i$, it equals

$$E[|\hat{X}_{j-1}^{(N)}|^4 \cdot (\sqrt{\Delta t^{(N)}}w_R^{(N)})^4] = (x^*\sqrt{T}\delta^{(N)})^4 \cdot \sup_j E|\hat{X}_{j-1}^{(N)}|^4,$$

which is also bounded by a constant because $\delta^{(N)} \to 1$ and $\sup_j E|\hat{X}_{j-1}^{(N)}|^4$ bounded.

In summary, all summands in (36) are bounded by constants independent of N. Therefore, there exists a constant C such that

$$E\left[\left|\Phi(\hat{X}_{i_1}^{(N)}, \dots, \hat{X}_{i_m}^{(N)}) \cdot \left[\frac{\hat{Z}_{i_k}^{(N)}}{\hat{Y}_{i_k}^{(N)}} - \frac{\hat{Z}_{i_{k-1}}^{(N)}}{\hat{Y}_{i_{k-1}}^{(N)}}\right] \cdot \frac{\hat{Y}_i^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})}\right|\Bigg|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}\right] < CN. \quad \square$$

Theorem 4.9. *Under Assumptions 3.1, 3.2 and 4.1–4.3, the estimator in Theorem 3.2 converges weakly to*

$$\Phi(X_{t_1}, \dots, X_{t_m}) \cdot \left\{\sum_{k=1}^{m}\left[\frac{Z_{t_k}}{Y_{t_k}} - \frac{Z_{t_{k-1}}}{Y_{t_{k-1}}}\right] \cdot \int_{t_{k-1}}^{t_k} a_s \frac{Y_s}{\sigma(X_s)} \mathrm{d}W_s - \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{Z_{t_m}}{Y_{t_m}}\right]\right\}.$$

Proof. Using Lemmas 4.7 and 4.8, we only need to show that

$$\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \frac{\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot \frac{\sqrt{\Delta t^{(N)}} \cdot \tilde{\xi}_i^{(N)}}{\delta^{(N)}} \Rightarrow \int_{t_{k-1}}^{t_k} a_s \frac{Y_s}{\sigma(X_s)} \mathrm{d}W_s \tag{37}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{Z_{i_m}^{(N)}}{Y_{i_m}^{(N)}}\right] \Rightarrow \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{Z_{t_m}}{Y_{t_m}}\right]. \tag{38}$$

Eq. (37) holds by arguments similar to those used in the proof of Theorem 4.6 and the fact that $\tilde{L}^{(N)}$ is P-UT (cf. Lemma 4.2).

Now turn to the proof of (38). Notice that

$$\frac{\mathrm{d}\hat{Z}_i^{(N)}}{\mathrm{d}x} = \frac{\mathrm{d}\hat{Z}_{i-1}^{(N)}}{\mathrm{d}x} + \left[\mu''(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}\hat{Z}_{i-1}^{(N)} + \mu'(\hat{X}_{i-1}^{(N)})\frac{\mathrm{d}\hat{Z}_{i-1}^{(N)}}{\mathrm{d}x}\right]\Delta t^{(N)} + \tilde{\sigma}'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}\tilde{\xi}_i^{(N)}$$
$$+ \left[\sigma''(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}\hat{Z}_{i-1}^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\frac{\mathrm{d}\hat{Z}_{i-1}^{(N)}}{\mathrm{d}x}\right]\xi_i^{(N)}\sqrt{\Delta t^{(N)}}$$

and

$$\frac{\mathrm{d}\hat{Y}_i^{(N)}}{\mathrm{d}x} = \frac{\mathrm{d}\hat{Y}_{i-1}^{(N)}}{\mathrm{d}x} + \left[\mu''(\hat{X}_{i-1}^{(N)})(\hat{Y}_{i-1}^{(N)})^2 + \mu'(\hat{X}_{i-1}^{(N)})\frac{\mathrm{d}\hat{Y}_{i-1}^{(N)}}{\mathrm{d}x}\right]\Delta t^{(N)} + \left[\sigma''(\hat{X}_{i-1}^{(N)})(\hat{Y}_{i-1}^{(N)})^2\right.$$
$$\left. + \sigma'(\hat{X}_{i-1}^{(N)})\frac{\mathrm{d}\hat{Y}_{i-1}^{(N)}}{\mathrm{d}x}\right]\xi_i^{(N)}\sqrt{\Delta t^{(N)}}.$$

Using arguments similar to those in Lemma 4.3, we can show that $(\mathrm{d}\hat{Z}^{(N)}/\mathrm{d}x, \mathrm{d}\hat{Y}^{(N)}/\mathrm{d}x)$ weakly converges to $(\mathrm{d}Z/\mathrm{d}x, \mathrm{d}Y/\mathrm{d}x)$ in the Skorohod topology. Thus,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\hat{Z}_{i_m}^{(N)}}{\hat{Y}_{i_m}^{(N)}}\right] = \frac{\mathrm{d}\hat{Z}_{i_m}^{(N)}}{\mathrm{d}x}\cdot\frac{1}{\hat{Y}_{i_m}^{(N)}} - \frac{\hat{Z}_{i_m}^{(N)}}{(\hat{Y}_{i_m}^{(N)})^2}\cdot\frac{\mathrm{d}\hat{Y}_{i_m}^{(N)}}{\mathrm{d}x}$$
$$\Rightarrow \frac{\mathrm{d}Z_{t_m}}{\mathrm{d}x}\cdot\frac{1}{Y_{t_m}} - \frac{Z_{t_m}}{(Y_{t_m})^2}\cdot\frac{\mathrm{d}Y_{t_m}}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{Z_{t_m}}{Y_{t_m}}\right]. \quad \square$$

4.4. Convergence of rho estimators

In this section, we turn to the convergence of the rho estimators. Because the necessary arguments are very similar to (and simpler than) what we used in Theorems 4.6 and 4.9, we just list the theorem as follows without detailed justification:

Theorem 4.10. *Under all assumptions in Section 3 and Assumption 4.2, the estimator given in Theorem 3.3 converges weakly to*

$$\Phi(X_{t_1}, \ldots, X_{t_m}) \cdot \int_0^T \frac{\gamma(X_s)}{\sigma(X_s)}\mathrm{d}W_s.$$

The rest of this article presents technical results in Appendices A–C.

Acknowledgment

This work is supported, in part, by NSF grants DMI0300044 and DMS0410234.

Appendix A. Some technical lemmas

Lemma A.1. *There exist parameters satisfying (22)–(24).*

Proof. Notice that

$$\lim_{x\to+\infty}\frac{x\mathrm{e}^{-x^2/2}}{2\psi(x)-1}=0,$$

where $\psi(\cdot)$ is the cumulative normal distribution function. Thus, we can choose an x such that

$$F(x):=\frac{2}{\sqrt{2\pi}}\cdot\frac{x\mathrm{e}^{-x^2/2}}{2\psi(x)-1}<1.$$

Let $C=1/(2\psi(x)-1)$, $\delta=1/(1-F(x))$ and $w_R=-w_L=x\sqrt{\delta}$. Then,

$$\int_{w_L}^{w_R}\frac{C}{\sqrt{2\pi\delta}}\cdot\mathrm{e}^{-\frac{w^2}{2\delta}}\mathrm{d}w=C\cdot\int_{w_L/\sqrt{\delta}}^{w_R/\sqrt{\delta}}\frac{1}{\sqrt{2\pi}}\cdot\mathrm{e}^{-\frac{w^2}{2}}\mathrm{d}w=C\cdot(2\psi(x)-1)=1;$$

$$\begin{aligned}\int_{w_L}^{w_R}w^2\frac{C}{\sqrt{2\pi\delta}}\cdot\mathrm{e}^{-\frac{w^2}{2\delta}}\mathrm{d}w &= C\delta\cdot\int_{w_L/\sqrt{\delta}}^{w_R/\sqrt{\delta}}\frac{w^2}{\sqrt{2\pi}}\cdot\mathrm{e}^{-\frac{w^2}{2}}\mathrm{d}w\\ &=\delta\left(1-\frac{C\cdot 2x\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\right)=\delta(1-F(x))=1.\quad\square\end{aligned}$$

Lemma A.2. *Theorem 3.2 holds for any Φ satisfying Assumption 3.1.*

Proof. Consider any Φ satisfying Assumption 3.1, not necessarily differentiable. Let φ be a C^∞ function

$$\varphi(x_1,\ldots,x_m)=\begin{cases}c\exp\left(\dfrac{1}{\sum_{i=1}^m x_i^2-1}\right); & \sum_{i=1}^m x_i^2<1;\\ 0; & \text{otherwise,}\end{cases}$$

where c is selected so that $\int_{\mathbf{R}^m}\varphi(x)dx=1$ and let $\varphi^n(x)=n^m\varphi(nx)$. We employ the mollifiers $\varphi^n(x)$ to obtain differentiable approximations to Φ by convolution:

$$\phi^n(x)=\int_{R^m}\varphi^n(x-y)\Phi(y)\mathrm{d}y.$$

Then, the function ϕ^n is C^∞ for every n and $\phi^n\to\Phi$ as $n\to+\infty$ (cf. Evans [6], Theorem C.6 (i) and (ii), p. 630).

Pick a neighborhood K of $\varepsilon=0$ as in Assumption 3.2.

$$\begin{aligned}&\sup_{\varepsilon\in K}E[|\phi^n(\hat{X}^\varepsilon_{i_1},\ldots,\hat{X}^\varepsilon_{i_m})-\Phi(\hat{X}^\varepsilon_{i_1},\ldots,\hat{X}^\varepsilon_{i_m})|^2]\\ &\quad=\sup_{\varepsilon\in K}\int_{[w_L,w_R]^N}|\phi^n(\hat{x}^\varepsilon_{i_1},\ldots,\hat{x}^\varepsilon_{i_m})-\Phi(\hat{x}^\varepsilon_{i_1},\ldots,\hat{x}^\varepsilon_{i_m})|^2\\ &\quad\cdot\prod_{j=1}^N\left(\frac{C}{\sqrt{2\pi\delta}}\exp(-\frac{w_j^2}{2\delta})\right)\mathrm{d}w_1\cdots\mathrm{d}w_N\end{aligned}$$

where the production is the joint density function of $(\tilde{\xi}_1,\ldots,\tilde{\xi}_N)$ and x^ε_i is given by the recursion: $\hat{x}^\varepsilon_i=\hat{x}^\varepsilon_{i-1}+\mu(\hat{x}^\varepsilon_{i-1})\Delta t+[\sigma(\hat{x}^\varepsilon_{i-1})+\varepsilon\tilde{\sigma}(\hat{x}^\varepsilon_{i-1})]\sqrt{\Delta t}w_i$. We can see that the right hand side of the above equality should be less than

$$\sup_{\varepsilon\in K}\int_{\mathbf{R}^N}|\phi^n(\hat{x}^{\varepsilon}_{i_1},\ldots,\hat{x}^{\varepsilon}_{i_m})-\Phi(\hat{x}^{\varepsilon}_{i_1},\ldots,\hat{x}^{\varepsilon}_{i_m})|^2\cdot\prod_{j=1}^{N}\left(\frac{C}{\sqrt{2\pi\delta}}\exp\left(-\frac{w_j^2}{2\delta}\right)\right)\mathrm{d}w_1\cdots\mathrm{d}w_N$$

if we enlarge the whole integration region from $[w_L, w_R]^N$ to $\mathbf{R}^N$. Now applying change of variable here to switch integration variables from $(w_1,\ldots,w_N)$ to $(\hat{x}^{\varepsilon}_1,\ldots,\hat{x}^{\varepsilon}_N)$, we have the above should equal to

$$\left(\frac{C}{\sqrt{2\pi\delta}}\right)^N\sup_{\varepsilon\in K}\int_{\mathbf{R}^N}|\phi^n-\Phi|^2$$
$$\cdot\prod_{j=1}^{N}\left(\frac{1}{\sqrt{\Delta t}(\sigma(x_{j-1})+\varepsilon\tilde{\sigma}(x_{j-1}))}\exp\left(-\frac{(x_i-x_{i-1}-\mu(x_{i-1})\Delta t)^2}{2\delta(\sigma(x_{j-1})+\varepsilon\tilde{\sigma}(x_{j-1}))^2\Delta t}\right)\right)\mathrm{d}x.$$

By part (3) of Assumption 3.2, we know that $\sigma+\varepsilon\tilde{\sigma}$ is bounded below by a positive constant ϵ for all $\varepsilon\in K$. Thus the above should be less than

$$C^N\int_{\mathbf{R}^N}|\phi^n-\Phi|^2\cdot\prod_{j=1}^{N}\left(\frac{1}{\sqrt{2\pi\delta\Delta t\epsilon}}\exp\left(-\frac{(x_i-x_{i-1}-\mu(x_{i-1})\Delta t)^2}{2\delta\epsilon^2\Delta t}\right)\right)\mathrm{d}x_1\cdots\mathrm{d}x_N.$$

Notice that the production is the joint density function of $(\hat{X}^{\epsilon}_1,\ldots,\hat{X}^{\epsilon}_N)$ where $\hat{X}^{\epsilon}_i$ is defined by the recursion: $\hat{X}^{\epsilon}_i=\hat{X}^{\epsilon}_{i-1}+\mu(\hat{X}^{\epsilon}_{i-1})\Delta t+\sqrt{\delta}\epsilon\sqrt{\Delta t}\xi_i$. One can easily show that the above converges to 0 as $n\to+\infty$ if using the argument of Theorem C.6(iv) in Evans [6] (cf. p. 630. The proof there is under the Lebesgue measure. But one can easily change it to cover our case under the measure induced by $(\hat{X}^{\epsilon}_1,\ldots,\hat{X}^{\epsilon}_N)$). This implies,

$$\lim_{n\to\infty}\sup_{\varepsilon\in K}E[|\phi^n(\hat{X}^{\varepsilon}_{i_1},\ldots,\hat{X}^{\varepsilon}_{i_m})-\Phi(\hat{X}^{\varepsilon}_{i_1},\ldots,\hat{X}^{\varepsilon}_{i_m})|^2]=0.$$

Next, applying the arguments which lead to Theorem 3.2 and noticing that ϕ^n is differentiable, for any $\varepsilon\in K$,

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}E[\phi^n(\hat{X}^{\varepsilon}_{i_1},\ldots,\hat{X}^{\varepsilon}_{i_m})]=E[\phi^n(\hat{X}^{\varepsilon}_{i_1},\ldots,\hat{X}^{\varepsilon}_{i_m})\cdot D(\varepsilon)] \tag{39}$$

where

$$D(\varepsilon)=\left\{\left[\frac{\hat{Z}^{\varepsilon}_{i_k}}{\hat{Y}^{\varepsilon}_{i_k}}-\frac{\hat{Z}^{\varepsilon}_{i_{k-1}}}{\hat{Y}^{\varepsilon}_{i_{k-1}}}\right]\cdot\sum_{i=i_{k-1}+1}^{i_k}a_i\frac{\hat{Y}^{\varepsilon}_{i-1}}{\sigma(\hat{X}^{\varepsilon}_{i-1})}\cdot\frac{\sqrt{\Delta t}\cdot\tilde{\xi}_i}{\delta}-\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\hat{Z}^{\varepsilon}_{i_m}}{\hat{Y}^{\varepsilon}_{i_m}}\right]\right\}$$
$$+\left\{\sum_{k=1}^{m}\left[\frac{\hat{Z}^{\varepsilon}_{i_k}}{\hat{Y}^{\varepsilon}_{i_k}}-\frac{\hat{Z}^{\varepsilon}_{i_{k-1}}}{\hat{Y}^{\varepsilon}_{i_{k-1}}}\right]\cdot\sum_{i=i_{k-1}+1}^{i_k}a_i\Delta t\right.$$
$$\left.\cdot\left(\frac{\mu'(\hat{X}^{\varepsilon}_{i-1})\hat{Y}^{\varepsilon}_{i-1}}{\sigma(\hat{X}^{\varepsilon}_{i-1})}\cdot\frac{\sqrt{\Delta t}\tilde{\xi}_i}{\delta}+\frac{\sigma'(\hat{X}^{\varepsilon}_{i-1})\hat{Y}^{\varepsilon}_{i-1}}{\sigma(\hat{X}^{\varepsilon}_{i-1})}\cdot\left[\frac{\tilde{\xi}_i^2}{\delta}-1\right]\right)\right\}$$

$$+ \frac{C}{\sqrt{2\pi\delta}} e^{-\frac{w_R^2}{2\delta}} \sum_{k=1}^{m} \sum_{i=i_{k-1}+1}^{i_k} a_i \sqrt{\Delta t} \cdot \left[\frac{\hat{Z}_{i_k}^{\varepsilon}}{\hat{Y}_{i_k}^{\varepsilon}} - \frac{\hat{Z}_{i_{k-1}}^{\varepsilon}}{\hat{Y}_{i_{k-1}}^{\varepsilon}} \right] \cdot \frac{\hat{Y}_i^{\varepsilon}}{\sigma(\hat{X}_{i-1}^{\varepsilon})} \Bigg|_{\tilde{\xi}_i = w_R}$$

$$- \frac{C}{\sqrt{2\pi\delta}} e^{-\frac{w_L^2}{2\delta}} \sum_{k=1}^{m} \sum_{i=i_{k-1}+1}^{i_k} a_i \sqrt{\Delta t} \cdot \left[\frac{\hat{Z}_{i_k}^{\varepsilon}}{\hat{Y}_{i_k}^{\varepsilon}} - \frac{\hat{Z}_{i_{k-1}}^{\varepsilon}}{\hat{Y}_{i_{k-1}}^{\varepsilon}} \right] \cdot \frac{\hat{Y}_i^{\varepsilon}}{\sigma(\hat{X}_{i-1}^{\varepsilon})} \Bigg|_{\tilde{\xi}_i = w_L}.$$

By the Cauchy–Schwarz inequality,

$$\sup_{\varepsilon\in K} E[|\phi^n(\hat{X}_{i_1}^{\varepsilon}, \ldots, \hat{X}_{i_m}^{\varepsilon}) - \Phi(\hat{X}_{i_1}^{\varepsilon}, \ldots, \hat{X}_{i_m}^{\varepsilon})| \cdot |D(\varepsilon)|]$$
$$\leq \sup_{\varepsilon\in K} E[|\phi^n(\hat{X}_{i_1}^{\varepsilon}, \ldots, \hat{X}_{i_m}^{\varepsilon}) - \Phi(\hat{X}_{i_1}^{\varepsilon}, \ldots, \hat{X}_{i_m}^{\varepsilon})|^2] \cdot \sup_{\varepsilon\in K} E[|D(\varepsilon)|^2].$$

Note that $D(\varepsilon)$ is a continuous function with respect to ε and realization of $(\tilde{\xi}_1, \ldots, \tilde{\xi}_N)$. □

The following result is a discrete version of the well-known Girsanov theorem. We state it without proof.

Lemma A.3. *Define a new probability measure* $dP^{\varepsilon} = U^{\varepsilon} dP$ *with* U^{ε} *defined as in Section* 3.3. *Then,*

$$P^{\varepsilon}[\sqrt{\Delta t}\xi_1 \in dw_1, \ldots, \sqrt{\Delta t}\xi_N \in dw_N] = \frac{1}{(2\pi\Delta t)^{N/2}}$$
$$\cdot \exp\left[-\frac{1}{2} \sum_{i=1}^{N} \frac{(w_i - \varepsilon \frac{\gamma(X_{i-1})}{\sigma(X_{i-1})} \Delta t)^2}{\Delta t} \right] dw_1 \cdots dw_N.$$

Appendix B. Weak convergence of stochastic differential equations

This appendix presents a lemma, based on Kurtz and Protter [13] or Kurtz and Protter [14] on the stability of stochastic differential equations that serves as the theoretical foundation of Section 4.

Suppose that the space of all cadlag functions: $[0, T] \to \mathbf{R}^m$ (that is, functions with right continuity and left limits) by $\mathcal{D}^m[0, T]$. Equip this space with the Skorokhod topology: a sequence of processes $\alpha_n \in \mathcal{D}^m[0, T]$ converges to a process α if and only if there exists a sequence of functions λ_n, which are all strictly increasing with $\lambda_n(0) = 0$ and $\lambda_n(T) = T$, such that

$$\lim_{n\to+\infty} \sup_{0\leq t\leq T} |\lambda_n(s) - s| = 0 \quad \text{and} \quad \lim_{n\to+\infty} \sup_{0\leq t\leq T} |\alpha_n(\lambda_n(s)) - \alpha(s)| = 0.$$

We work with a sequence of probability spaces $\mathcal{B}^{(N)} := (\Omega^{(N)}, \mathcal{F}^{(N)}, \{\mathcal{F}_t^{(N)}\}, P^{(N)})$ and a probability space $\mathcal{B} := (\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$. For each N, an m-dimensional process $L^{(N)}$ is a semimartingale defined on the probability space $\mathcal{B}^{(N)}$, which is adapted to the filtration $\{\mathcal{F}_t^{(N)}\}$. The sample paths of $L^{(N)}$ is in the space $\mathcal{D}^m[0, T]$. Let $g^{(N)} : \mathbf{R}^k \times [0, +\infty) \to \mathbb{M}^{km}$ be a continuous function mapping, with $\mathbb{M}^{km}$ the space of $k \times m$ matrices. Assume that a pair of k-dimensional processes $M^{(N)}$ and $U^{(N)}$ satisfies the following SDE:

$$M_t^{(N)} = U_t^{(N)} + \int_0^t g^{(N)}(M_{s-}^{(N)}, s-) dL_s^{(N)}.$$

Similarly, on the space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, a triple (M, U, L) and a continuous function g satisfy

$$M_t = U_t + \int_0^t g(M_{s-}, s-)\mathrm{d}L_s, \tag{40}$$

where the sample paths of L is in $\mathcal{D}^m[0, T]$ too.

In addition, we call a sequence of probability measures on $\mathcal{D}^m[0, T]$, η_n, *weakly converges in the Skorokhod topology* to another probability measure η if for any bounded continuous functional $f\colon \mathcal{D}^m[0, T] \to \mathbf{R}$, we have

$$\lim_{n\to+\infty} \int_{\mathcal{D}^m[0,T]} f(\omega)\mathrm{d}\eta_m(\omega) = \int_{\mathcal{D}^m[0,T]} f(\omega)\mathrm{d}\eta(\omega).$$

Then we denote that $\eta_m \Rightarrow \eta$.

We would like to present a sufficient condition under which processes $M^{(N)}$ converge weakly to M. But before doing that, several definitions are needed.

Definition B.1 (*Elementary Processes*). An *elementary process* on $\mathcal{B}^{(N)}$ is a predictable process of the form

$$H_t = Y_0 1_{\{0\}} + \sum_{i=1}^{k} Y_i 1_{(s_i, s_{i+1}]}(t),$$

for some positive integer k and $0 < s_1 < \cdots < s_{k+1}$, with Y_i being $\mathcal{F}_{s_i}^{(N)}$-measurable and $|Y_i| \le 1$. The set of all elementary stochastic processes on $\mathcal{B}^{(N)}$ is denoted by $\mathcal{H}^{(N)}$ and the set of all elementary processes on $\mathcal{B}$ is denoted by $\mathcal{H}$.

Definition B.2 (*P-UT Condition (Definition 6.1, Jacod and Shiryaev [11])*). A sequence $L^{(N)}$ of adapted cadlag k-dimensional processes defined on probability spaces $(\Omega^{(N)}, \mathcal{F}^{(N)}, \{\mathcal{F}_t^{(N)}\}, P^{(N)})$ is said to be *predictably uniformly tight (or P-UT)* if for every $t > 0$

$$\lim_{a\to+\infty} \sup_{H^{(N)}\in\mathcal{H}^{(N)}, N} P^{(N)}\left(\left|\int_0^t H_{s-}^{(N)}\mathrm{d}L_s^{(N)}\right| > a\right) = 0.$$

We now have the following lemma on the weak convergence of $M^{(N)}$. It is a simple corollary of Theorem 5.4 in Kurtz and Protter [13] or Theorem 8.6 in Kurtz and Protter [14]:

Lemma B.3. *Suppose that $(U^{(N)}, L^{(N)})$ converge weakly to (U, L) in the Skorokhod topology and that $L^{(N)}$ is P-UT. Assume that $g^{(N)}$ uniformly converges to g locally,*

$$\lim_{N\to+\infty} \sup_{(x,t)\in K} |g^{(N)}(x, t) - g(x, t)| = 0. \tag{41}$$

If there exists a global solution M of (40) and weak local uniqueness holds, then $(U^{(N)}, L^{(N)}, M^{(N)}) \Rightarrow (U, L, M)$ in the Skorokhod topology.

Appendix C. Uniform bounds on moments

In this appendix, we derive various upper bounds for the moments of $\hat{X}^{(N)}$, $\hat{Y}^{(N)}$ and $(\hat{Y}^{(N)})^{-1}$ used in the proof of weak convergence of the vega estimators. Throughout this appendix, these processes are assumed to be constructed from the truncated normal increments $\tilde{\xi}_i^{(N)}$.

Lemma C.1. *Suppose that p is a positive integer. Then,*

$$\sup_N \sup_{1\le i\le N} E[(\hat{X}_i^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] < +\infty.$$

Proof. Using the multinomial theorem, we know that

$$(a+b+c)^{2p} = a^{2p} + \sum_{k_1<2p, k_1+k_2=2p-1} \binom{2p}{k_1, k_2, 1} a^{k_1} b^{k_2} c + \sum_{k_1<2p, k_3\ne 1} \binom{2p}{k_1, k_2, k_3} a^{k_1} b^{k_2} c^{k_3}. \tag{42}$$

For any index $i \ne j$, substituting a, b, c in the above equation by $\hat{X}_{i-1}^{(N)}$, $\mu(\hat{X}_{i-1}^{(N)})\Delta t^{(N)}$ and $\sigma(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}$ and taking expectations of both sides,

$$E[(\hat{X}_i^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] = E\Bigg[(\hat{X}_{i-1}^{(N)})^{2p} + \sum_{k_1<2p, k_3\ne 1} \binom{2p}{k_1, k_2, k_3} \times (\Delta t^{(N)})^{k_2+\frac{k_3}{2}} (\hat{X}_{i-1}^{(N)})^{k_1} (\mu(\hat{X}_{i-1}^{(N)}))^{k_2} (\sigma(\hat{X}_{i-1}^{(N)})\tilde{\xi}_i^{(N)})^{k_3} |\tilde{\xi}_j^{(N)} = w_R^{(N)}\Bigg], \tag{43}$$

where the middle term in (42) is absent because

$$E[\sigma(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] = E[\sigma(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}} E[\tilde{\xi}_i^{(N)}|\mathcal{F}_{i-1}]|\tilde{\xi}_j^{(N)} = w_R^{(N)}] = 0.$$

In addition, the assumption of bounded derivatives of μ and σ implies that there exists a constant C_1, independent of N and index i, such that $|\mu(x)| \le C_1(1+|x|)$ and $|\sigma(x)| \le C_1(1+|x|)$ for all x. Then, the right hand side of (43) is bounded by

$$E\Bigg[|\hat{X}_{i-1}^{(N)}|^{2p} + \sum_{k_1<2p, k_3\ne 1} (\Delta t^{(N)})^{k_2+\frac{k_3}{2}} C_1^{k_2+k_3} (1+|\hat{X}_{i-1}^{(N)}|)^{k_1+k_2+k_3} |\tilde{\xi}_i^{(N)}|^{k_3} |\tilde{\xi}_j^{(N)} = w_R^{(N)}\Bigg]$$
$$\le E\Bigg[|\hat{X}_{i-1}^{(N)}|^{2p} + C_1^{2p}(1+|\hat{X}_{i-1}^{(N)}|)^{2p} \sum_{k_1<2p, k_3\ne 1} (\Delta t^{(N)})^{k_2+\frac{k_3}{2}} |\tilde{\xi}_i^{(N)}|^{k_3} |\tilde{\xi}_j^{(N)} = w_R^{(N)}\Bigg]$$

where $k_1+k_2+k_3 = 2p$. Notice that $\hat{X}_{i-1}^{(N)}$ and $\tilde{\xi}_i^{(N)}$ are independent, that $\max_{1\le k\le N} E[|\tilde{\xi}_i^{(N)}|^k]$ is bounded uniformly for all N, and that $k_2+k_3/2 \ge 1$ when $k_1+k_2+k_3 = 2p$, $k_1 < 2p$, $k_3 \ne 1$. We can find a constant C_2, independent of N and index i, such that the above is less than

$$E[|\hat{X}_{i-1}^{(N)}|^{2p} + C_2\Delta t^{(N)}(1+|\hat{X}_{i-1}^{(N)}|)^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}].$$

Using the inequality $(a+b)^{2p} \le 2^{2p}(a^{2p}+b^{2p})$, the above is bounded by

$$(1+2^{2p}C_2\Delta t^{(N)})E[|\hat{X}_{i-1}^{(N)}|^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] + 2^{2p}C_2\Delta t^{(N)}. \tag{44}$$

For index j,

$$E[(\hat{X}_j^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] = E[(\hat{X}_{j-1}^{(N)} + \mu(\hat{X}_{j-1}^{(N)})\Delta t^{(N)} + \sigma(\hat{X}_{j-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_j^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}]$$
$$= E[(\hat{X}_{j-1}^{(N)} + \mu(\hat{X}_{j-1}^{(N)})\Delta t^{(N)} + \sigma(\hat{X}_{j-1}^{(N)})\sqrt{\Delta t^{(N)}}w_R^{(N)})^{2p}].$$

By Assumption 4.1, $\sqrt{\Delta t^{(N)}}w_R^{(N)} = \sqrt{T}x^*\delta^{(N)}$, which is bounded by a constant C_3 for all N because $\delta^{(N)} \to 1$. Using the vector inequality $(a+b+c)^{2p} \le 3^{2p-1}(a^{2p}+b^{2p}+c^{2p})$, and $|\mu(x)| \le C_1(1+|x|)$ and $|\sigma(x)| \le C_1(1+|x|)$, we have

$$E[(\hat{X}_j^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] \le 3^{2p-1}E[|\hat{X}_{j-1}^{(N)}|^{2p} + |\mu(\hat{X}_{j-1}^{(N)})|^{2p}|\Delta t^{(N)}|^{2p} + C_3^{2p}|\sigma(\hat{X}_{j-1}^{(N)})|^{2p}]$$
$$\le 3^{2p-1}E[1+|\hat{X}_{j-1}^{(N)}|]^{2p}\cdot(1+C_1^{2p}|\Delta t^{(N)}|^{2p} + C_3^{2p}C_1^{2p}).$$

By the vector inequality $(a+b)^{2p} \le 2^{2p}(a^{2p}+b^{2p})$ again, we can find constant C_4 such that the right hand side of the above is less than $C_4(1+E[|\hat{X}_{j-1}^{(N)}|^{2p}])$.

In summary, for a general i, we can find a common constant C, which is independent of N, such that

$$E[(\hat{X}_i^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] \le (1+C\Delta t^{(N)})\cdot E[(\hat{X}_{i-1}^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] + C\Delta t^{(N)}$$

if $i \ne j$ or

$$E[(\hat{X}_i^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] \le CE[(\hat{X}_{i-1}^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] + C$$

if $i = j$. Accordingly, by induction, we can easily get that

$$E[(\hat{X}_i^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] \le (1+C+2C\Delta t^{(N)})(1+C\Delta t^{(N)})^{N-1}.$$

The limit of the right hand side exists when $N \to +\infty$. Therefore,

$$\sup_N \sup_{1\le i\le N} E[(\hat{X}_i^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] < +\infty. \quad \square$$

Lemma C.2. *For any j,*

$$\sup_N \sup_{1\le i\le N} E[(\hat{Y}_i^{(N)})^4|\tilde{\xi}_j^{(N)} = w_R^{(N)}] < +\infty.$$

Proof. For $i \ne j$,

$$E[(\hat{Y}_i^{(N)})^4|\tilde{\xi}_j^{(N)} = w_R^{(N)}] = E[(\hat{Y}_{i-1}^{(N)})^4\cdot(1+\mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)})^4|\tilde{\xi}_j^{(N)} = w_R^{(N)}]$$
$$= E[(\hat{Y}_{i-1}^{(N)})^4\cdot E[(1+\mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)})^4]|\tilde{\xi}_j^{(N)} = w_R^{(N)}].$$

Using the multinomial theorem and similar argument as in the proof of Lemma C.1, and the fact that μ' and σ' are both bounded, the right hand side of the above equation is bounded by

$$(1 + C_5 \Delta t^{(N)}) \cdot E[(\hat{Y}_{i-1}^{(N)})^4 | \tilde{\xi}_j^{(N)} = w_R^{(N)}]$$

where C_5 is a constant which does not depend on N and index i.

For index j,

$$\begin{aligned} E[(\hat{Y}_j^{(N)})^4 | \tilde{\xi}_j^{(N)} = w_R^{(N)}] &= E[(\hat{Y}_{j-1}^{(N)})^4 \cdot (1 + \mu'(\hat{X}_{j-1}^{(N)}) \Delta t^{(N)} \\ &\quad + \sigma'(\hat{X}_{j-1}^{(N)}) \sqrt{\Delta t^{(N)}} \tilde{\xi}_j^{(N)})^4 | \tilde{\xi}_j^{(N)} = w_R^{(N)}] \\ &= E[(\hat{Y}_{j-1}^{(N)})^4 \cdot (1 + \mu'(\hat{X}_{j-1}^{(N)}) \Delta t^{(N)} \\ &\quad + \sigma'(\hat{X}_{j-1}^{(N)}) \sqrt{\Delta t^{(N)}} w_R^{(N)})^4 | \tilde{\xi}_j^{(N)} = w_R^{(N)}]. \end{aligned}$$

We know from the proof of Lemma C.1 that $\sqrt{\Delta t^{(N)}} w_R^{(N)} \leq C_3$. Thus, by the boundedness of μ' and σ' and similar arguments as those in the proof of Lemma C.1,

$$E[(\hat{Y}_j^{(N)})^4 | \tilde{\xi}_j^{(N)} = w_R^{(N)}] \leq 3^3 E[|\hat{Y}_{j-1}^{(N)}|^4] \cdot (1 + C_6^4 |\Delta t^{(N)}|^4 + C_6^4 C_3^4).$$

where $\sup_x |\mu'| < C_6$ and $\sup_x |\sigma'| < C_6$.

By induction,

$$\begin{aligned} E[(\hat{Y}_i^{(N)})^{2p} | \tilde{\xi}_j^{(N)} = w_R^{(N)}] &\leq (1 + C_5 \Delta t^{(N)})^{N-1} \cdot 3^3 \cdot (1 + C_6^4 |\Delta t^{(N)}|^4 + C_6^4 C_3^4) \\ &\to \mathrm{e}^{C_5 T} 3^3 \cdot (1 + C_6^4 C_3^4). \quad \square \end{aligned}$$

Lemma C.3. *For any j,*

$$\sup_N \sup_{1 \leq i \leq N} E[(Y_i^{(N)})^{-4} | \tilde{\xi}_j^{(N)} = w_R^{(N)}] < +\infty.$$

Proof. Consider the function $G(x) := 1/(1+x)^4$. If we expand it at $x = 0$ in a Taylor series with Lagrange remainder, we get

$$\frac{1}{(1+x)^4} = G(0) + G'(0)x + \frac{1}{2} G''(\theta) x^2 = 1 - 4x + \frac{10x^2}{(1+\theta)^6}$$

for some θ between 0 and x. Furthermore, if we choose some positive number ϵ such that $1 - \sup_x |\sigma'(x)| \sqrt{T} x^* - \epsilon > 0$ (by Assumption 4.1, such ϵ exists because $1 - \sup_x |\sigma'(x)| \sqrt{T} x^* > 0$) and suppose x lies in an interval $[-\sup_x |\sigma'(x)| \sqrt{T} x^* - \epsilon, \sup_x |\sigma'(x)| \sqrt{T} x^* + \epsilon]$, then we can bound the function $G(x)$ by $1 - 4x + 10 m^* x^2$, where m^* is the maximum value of $1/(1+\theta)^6$ over the interval $[-\sup_x |\sigma'(x)| \sqrt{T} x^* - \epsilon, \sup_x |\sigma'(x)| \sqrt{T} x^* + \epsilon]$.

Notice that for any index $i \neq j$, by Assumption 4.1,

$$\begin{aligned} |\mu'(\hat{X}_{i-1}^{(N)}) \Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)}) \sqrt{\Delta t^{(N)}} \tilde{\xi}_i^{(N)}| &\leq \sup_x |\mu'(x)| \Delta t^{(N)} + \sup_x |\sigma'(x)| \sqrt{\Delta t^{(N)}} w_R^{(N)} \\ &= \sup_x |\mu'(x)| \Delta t^{(N)} + \sup_x |\sigma'(x)| \sqrt{T} x^* \delta^{(N)}. \end{aligned}$$

In other words, $\mu'(\hat{X}_{i-1}^{(N)}) \Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)}) \sqrt{\Delta t^{(N)}} \tilde{\xi}_i^{(N)}$ falls in the interval $[-\sup_x |\sigma'(x)| \sqrt{T} x^* - \epsilon, \sup_x |\sigma'(x)| \sqrt{T} x^* + \epsilon]$ almost surely when N is big enough. Therefore,

$$(1 + \mu'(\hat{X}_{i-1}^{(N)}) \Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)}) \sqrt{\Delta t^{(N)}} \tilde{\xi}_i^{(N)})^{-4}$$

$$\leq 1 - 4[\mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}]$$
$$+ 10m^*[\mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}]^2$$
$$= 1 + [-4\mu'\Delta t^{(N)} + 10m^*(\mu'\Delta t^{(N)})^2] + [-4\sigma' + 20m^*\mu'\sqrt{\Delta t^{(N)}}\sigma']\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}$$
$$+ 10m^*(\sigma')^2\Delta t^{(N)}(\tilde{\xi}_i^{(N)})^2.$$

Taking expectations on both sides of the above inequality, we have

$$E[(1 + \mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)})^{-4}|\tilde{\xi}_j^{(N)} = w_R^{(N)}]$$
$$\leq 1 + [-4\mu'\Delta t^{(N)} + 10m^*(\mu'\Delta t^{(N)})^2] + 10m^*(\sigma')^2\Delta t^{(N)}E[\tilde{\xi}_i^{(N)}]^2.$$

Because $E[\tilde{\xi}_i^{(N)}]^2$ is bounded uniformly for all N and μ' and σ' are both bounded, there exists a constant C_7 such that

$$E[(1 + \mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)})^{-4}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] \leq 1 + C_7\Delta t^{(N)}.$$

For index j,

$$E[(1 + \mu'(\hat{X}_{j-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{j-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_j^{(N)})^{-4}|\tilde{\xi}_j^{(N)} = w_R^{(N)}]$$
$$= E[(1 + \mu'(\hat{X}_{j-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{j-1}^{(N)})\sqrt{\Delta t^{(N)}}w_R^{(N)})^{-4}|\tilde{\xi}_j^{(N)} = w_R^{(N)}]$$
$$= E[(1 + \mu'(\hat{X}_{j-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{j-1}^{(N)})\sqrt{T}x^*\delta^{(N)})^{-4}|\tilde{\xi}_j^{(N)} = w_R^{(N)}]$$

where $w_R^{(N)} = \sqrt{N}x^*\delta^{(N)}$ by Assumption 4.1. Using the fact that μ' and σ' are bounded and that $\delta^{(N)} \to 1$, we find that the right hand side of the above equality is bounded by $(1 - \sup_x |\sigma'(x)|\sqrt{T}x^* - \epsilon)^{-4}$ if N is big enough.

Accordingly,

$$E[(\hat{Y}_i^{(N)})^{-4}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] = E[(\hat{Y}_{i-1}^{(N)})^{-4} \cdot (1 + \mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)}$$
$$+ \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)})^{-4}|\tilde{\xi}_j^{(N)} = w_R^{(N)}]$$

will be less than

$$(1 + C_7\Delta t^{(N)})E[(\hat{Y}_{i-1}^{(N)})^{-4}|\tilde{\xi}_j^{(N)} = w_R^{(N)}]$$

if $i \neq j$, and will be less than

$$(1 - \sup_x |\sigma'(x)|\sqrt{T}x^* - \epsilon)^{-4}E[(\hat{Y}_{i-1}^{(N)})^{-4}|\tilde{\xi}_j^{(N)} = w_R^{(N)}]$$

if $i = j$. By induction, we get

$$E[(\hat{Y}_i^{(N)})^{-4}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] \leq \frac{(1 + C_7\Delta t^{(N)})^{N-1}}{(1 - \sup_x |\sigma'(x)|\sqrt{T}x^* - \epsilon)^4}$$
$$\to \frac{e^{C_7 T}}{(1 - \sup_x |\sigma'(x)|\sqrt{T}x^* - \epsilon)^4}. \quad \square$$

References

[1] E. Benhamou, Optimal Malliavin weighting function for the computation of Greeks, Mathematical Finance 13 (2003) 37–53.
[2] H.-P. Bermin, A. Kohatsu-Higa, M. Montero, Local vega index and variance reduction methods, Mathematical Finance 13 (2003) 85–97.
[3] J. Cvitanic, J. Ma, J. Zhang, Efficient computation of hedging portfolios for options with discontinuous payoffs, Mathematical Finance 13 (2003) 135–151.
[4] M. Davis, M. Johansson, Malliavin Monte Carlo Greeks for jump diffusions, Stochastic Processes and their Applications 116 (2006) 101–129.
[5] R. Durrett, Probability: Theory and Examples, 2nd ed., Duxbury Press, 1996.
[6] L. Evans, Partial Differential Equations, in: Graduate Studies in Mathematics, vol. 19, American Mathematical Society, 1998.
[7] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, N. Touzi, Applications of Malliavin calculus to Monte Carlo methods in finance, Finance and Stochastics 3 (1999) 391–412.
[8] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, Applications of Malliavin calculus to Monte Carlo methods in finance. II, Finance and Stochastics 5 (2001) 201–236.
[9] P. Glasserman, Monte Carlo Methods in Financial Engineering, Springer-Verlag, New York, 2004.
[10] E. Gobet, A. Kohatsu-Higa, Computation of Greeks for barrier and lookback options using Malliavin calculus, Electronic Communications in Probability 8 (2003) 51–62.
[11] J. Jacod, A.N. Shiryayev, Limit Theorems for Stochastic Processes, 2nd ed., Springer-Verlag, Berlin, Heidelberg, 2003.
[12] A. Kohatsu-Higa, M. Montero, Malliavin calculus in finance, in: Handbook of Computational and Numerical Methods in Finance, Birkhauser, 2004, pp. 111–174.
[13] T.G. Kurtz, P. Protter, Weak limit theorems for stochastic integrals and stochastic differential equations, Annals of Probability 19 (3) (1991) 1035–1070.
[14] T.G. Kurtz, P. Protter, Weak convergence of stochastic integrals and differential equations, Lecture Notes in Mathematics 1627 (1995) 1–41.
[15] D. Nualart, Malliavin Calculus and Related Topics, Springer-Verlag, 1995.
[16] P. Protter, Stochastic Integration and Differential Equations, Springer-Verlag, Berlin, Heidelberg, 1990.