Malliavin Greeks without Malliavin Calculus

Nan Chen and Paul Glasserman
Columbia University
nc2005@columbia.edu and pg20@columbia.edu

July 2006

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 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.

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 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: differentating 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 [2], Bermin, Kohatsu-Higa, and Montero [3], Cvitanic, Ma, and Zhang [4], Davis and Johansson [5], Fournié et al. [8], Gobet and Kohatsu-Higa [10], Kohatsu-Higa and Montero [14], and many others. Using the tools of Malliavin calculus (cf. Nualart [17]), 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.

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 $[0, T]$,

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x, \tag{1}$$

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 $0 \leq t_1 \leq \cdots \leq t_m \leq T$. The expected present value of a contingent claim with this payoff is

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

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

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

When equality holds,

$$\Phi'(X_T)Y_T$$

provides an unbiased estimator of $u'(x)$, where $Y_T = dX_T/dx$ is the pathwise derivative of X_T with respect to the initial state. Under conditions in Section V.7 of Protter [18], the dynamics of Y can be obtained from (1) to get

$$dY_t = \mu'(X_t)Y_t dt + \sigma'(X_t)Y_t dW_t, \quad Y_0 = 1.$$

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

$$u(x) = \int \Phi(x_T) g(x, x_T)\, dx_T,$$

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

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

and the unbiased estimator

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

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)

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

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,

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

$i = 1, \ldots, N$, 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 $\hat{Y}_i = d\hat{X}_i/dx$,

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

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

$$\hat{u}(x) = \int \cdots \int \Phi(x_N)\hat{g}(x, x_1) \cdots \hat{g}(x_{N-1}, x_N)\, dx_N \cdots dx_1, \tag{7}$$

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

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

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 $i = 1, \ldots, N$, we may write

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

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

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

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 $i = 1, \ldots, N$. If we now average these unbiased estimators, we get

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

for small Δt. Thus, we recover the Malliavin estimator (4) as the limit of the average of combinations of pathwise and LRM estimators. Theorem 4.5 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

$$dX_t^\varepsilon = \mu(X_t^\varepsilon)\,dt + [\sigma(X_t^\varepsilon) + \varepsilon\tilde{\sigma}(X_t^\varepsilon)]\,dW_t, \quad X_0^\varepsilon = x, \tag{11}$$

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

$$dZ_t^\varepsilon = \mu'(X_t^\varepsilon)Z_t^\varepsilon\,dt + [\sigma'(X_t^\varepsilon) + \tilde{\sigma}(X_t^\varepsilon)]Z_t^\varepsilon\,dW_t, \quad Z_0^\varepsilon = 0. \tag{12}$$

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)

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

Here, D_t denotes the Malliavin derivative operator.

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

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

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

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

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 $d\hat{X}_i/d\Delta W_i = \sigma(\hat{X}_{i-1})$ to get

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

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

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

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

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

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 4.9.

3 Unbiased Estimators for Greeks

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

We define some basic notation that will be used later. We introduce two sequences of i.i.d. random variables $\{\xi_i, i \geq 1\}$ and $\{\tilde{\xi}_i, i \geq 1\}$, where ξ_i follows the normal distribution with mean 0 and variance 1, and $\tilde{\xi}_i$ the distribution of a truncated normal also with mean 0 and variance 1. The density function of the truncated normal is given by

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

where the parameters C, δ, w_L, w_R satisfy

$$\begin{aligned} & w_R = -w_L > 0, C > 0, \delta > 0; \\ & \int_{w_L}^{w_R} \frac{C}{\sqrt{2\pi\delta}} \cdot e^{-\frac{w^2}{2\delta}} dw = 1; \\ & Var[\tilde{\xi}_i] = \int_{w_L}^{w_R} w^2 \cdot \frac{C}{\sqrt{2\pi\delta}} \cdot e^{-\frac{w^2}{2\delta}} dw = 1. \end{aligned}$$

We write the Euler scheme as

$$\hat{X}_i = \hat{X}_{i-1} + \mu(\hat{X}_{i-1})\Delta t + \sigma(\hat{X}_{i-1})\sqrt{\Delta t}\xi_i \text{ (or } \tilde{\xi}_i).$$

As in Section 2, we will denote by $\hat{Y}_i$ the derivative of $\hat{X}_i$ with respect to the initial state x; and we will denote by $\hat{Z}_i$ the derivative with respect to ε in a perturbation that takes σ to $\sigma + \varepsilon\tilde{\sigma}$. We will assume that the drift coefficient μ and diffusion coefficient σ are both differentiable. Then, $\hat{Y}$ and $\hat{Z}$ obey the following recursion equations:

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

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 $\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m})$ for some $\{i_1, \cdots, i_m\} \subseteq \{1, \ldots, N\}$. We require the following:

Assumption 3.1 *There exists some positive integer* p *such that*

$$\limsup_{\|(x_1,\cdots,x_m)\| \to +\infty} \frac{|\Phi(x_1, \cdots, x_m)|}{\|(x_1, \cdots, x_m)\|^p} < +\infty.$$

Assumption 3.2 *1) μ, σ are twice differentiable;*
2)

$$\sup_x |\mu'(x)| + \sup_x |\mu''(x)| < +\infty, \quad \sup_x |\sigma'(x)| + \sup_x |\sigma''(x)| < +\infty;$$

3) σ is not degenerate, i.e., $\inf_x |\sigma(x)| > \epsilon$ for some $\epsilon > 0$;
4) $\tilde{\sigma}$ is also differentiable and

$$\sup_x |\tilde{\sigma}(x)| + \sup_x |\tilde{\sigma}'(x)| < +\infty.$$

Assumption 3.3 *γ is bounded,*

$$\sup_x |\gamma(x)| < +\infty.$$

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 underlying state at the claim maturity. For the analysis of this section, we consider the more general case of uneven weights and path-dependent claims, as in Fournié et al. [7]. With $\hat{X}_0 = x$, let $\hat{u}(x) = E[\Phi(\hat{X}_{i_1}, \ldots, \hat{X}_{i_m})]$. We consider the estimation of $\hat{u}'(x)$, the delta for the Euler approximation.

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 delta:*

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

where

$$\begin{aligned} I &= \Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \sum_{i=1}^{N} a_i \cdot \frac{\sigma'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} \cdot [(\sqrt{\Delta t} \cdot \xi_i)^2 - \Delta t]; \\ II &= \Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot (\sum_{i=1}^{N} a_i \cdot \frac{\mu'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} \sqrt{\Delta t} \cdot \xi_i) \cdot \Delta t. \end{aligned}$$

Proof. According to whether i is greater than the first exercise date i_1, we have two cases. For $i < i_1$, by the Markov property of $\hat{X}$,

$$\text{Delta} = \frac{d}{dx} E[\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) | \hat{X}_0 = x] = \frac{d}{dx} E[E[\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) | \hat{X}_i] | \hat{X}_0 = x].$$

Under Assumptions 3.1 and 3.2, we can interchange the order of differentiation and integration (cf. Lemma A.3). Furthermore, using the chain rule, we have

$$\begin{aligned}\text{Delta} &= E[\frac{d}{dx}E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i]] = E[\frac{d}{d\hat{X}_i}E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i]\cdot\frac{d\hat{X}_i}{dx}] \\ &= E[\frac{d}{d\hat{X}_i}E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i]\cdot\hat{Y}_i] \qquad (20)\end{aligned}$$

Note that given $\hat{X}_i$, $\hat{X}_{i+1}$ is normally distributed with mean $\hat{X}_i+\mu(\hat{X}_i)\Delta t$ and standard deviation $\sigma(\hat{X}_i)\sqrt{\Delta t}$. By Lemma A.2,

$$\begin{aligned}&\frac{d}{d\hat{X}_i}E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i] = \frac{d}{d\hat{X}_i}E[E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_{i+1}]|\hat{X}_i] \\ =\;& E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})\cdot\Big\{(\xi_{i+1}^2-1)\cdot\frac{\sigma'(\hat{X}_i)}{\sigma(\hat{X}_i)}+\frac{\xi_{i+1}}{\sqrt{\Delta t}}\cdot\frac{1}{\sigma(\hat{X}_i)}+\sqrt{\Delta t}\cdot\xi_{i+1}\cdot\frac{\mu'(\hat{X}_i)}{\sigma(\hat{X}_i)}\Big\}|\hat{X}_i].\end{aligned}$$

Now substitute this equation into (20) to get that

$$\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})\cdot\Big\{((\xi_{i+1})^2-1)\cdot\frac{\sigma'(\hat{X}_i)}{\sigma(\hat{X}_i)}+\frac{\xi_{i+1}}{\sqrt{\Delta t}}\cdot\frac{1}{\sigma(\hat{X}_i)}+\sqrt{\Delta t}\cdot\xi_{i+1}\cdot\frac{\mu'(\hat{X}_i)}{\sigma(\hat{X}_i)}\Big\}\cdot Y_i$$

is an unbiased estimator of delta.

For $i,k\geq i_1$ and $i_j\leq i,k<i_{j+1}$ for some j, we have

$$\begin{aligned}&E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})\cdot\Big\{(\xi_{i+1}^2-1)\cdot\frac{\sigma'(\hat{X}_i)}{\sigma(\hat{X}_i)}+\frac{\xi_{i+1}}{\sqrt{\Delta t}}\cdot\frac{1}{\sigma(\hat{X}_i)}+\sqrt{\Delta t}\cdot\xi_{i+1}\cdot\frac{\mu'(\hat{X}_i)}{\sigma(\hat{X}_i)}\Big\}\cdot\hat{Y}_i] \\ =\;& E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})\cdot\Big\{(\xi_{k+1}^2-1)\cdot\frac{\sigma'(\hat{X}_k)}{\sigma(\hat{X}_k)}+\frac{\xi_{k+1}}{\sqrt{\Delta t}}\cdot\frac{1}{\sigma(\hat{X}_k)}+\sqrt{\Delta t}\cdot\xi_{k+1}\cdot\frac{\mu'(\hat{X}_k)}{\sigma(\hat{X}_k)}\Big\}\cdot\hat{Y}_k] \\ &\qquad (21)\end{aligned}$$

Indeed, for any $i_j<i\leq i_{j+1}$, given $\mathcal{F}_{i_j}=\sigma\{\xi_1,\ldots,\xi_{i_j}\}$,

$$\begin{aligned}&E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})\cdot\Big\{(\xi_{i+1}^2-1)\cdot\frac{\sigma'(\hat{X}_i)}{\sigma(\hat{X}_i)}+\frac{\xi_{i+1}}{\sqrt{\Delta t}}\cdot\frac{1}{\sigma(\hat{X}_i)}+\sqrt{\Delta t}\cdot\xi_{i+1}\cdot\frac{\mu'(\hat{X}_i)}{\sigma(\hat{X}_i)}\Big\}\cdot\hat{Y}_i|\mathcal{F}_{i_j}] \\ =\;& E[\Phi(\hat{X}_{t_1},\cdots,\hat{X}_{t_j},\hat{X}_{t_{j+1}},\cdots,\hat{X}_{t_m})\cdot \\ &\Big\{(\xi_{i+1}^2-1)\cdot\frac{\sigma'(\hat{X}_i)}{\sigma(\hat{X}_i)}+\frac{\xi_{i+1}}{\sqrt{\Delta t}}\cdot\frac{1}{\sigma(\hat{X}_i)}+\sqrt{\Delta t}\cdot\xi_{i+1}\cdot\frac{\mu'(\hat{X}_i)}{\sigma(\hat{X}_i)}\Big\}\cdot\frac{\hat{Y}_i}{\hat{Y}_{t_j}}|\mathcal{F}_{i_j}]\cdot\hat{Y}_{t_j}\end{aligned}$$

Notice that

$$\begin{aligned}&E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})\cdot\Big\{(\xi_{i+1}^2-1)\cdot\frac{\sigma'(\hat{X}_i)}{\sigma(\hat{X}_i)}+\frac{\xi_{i+1}}{\sqrt{\Delta t}}\cdot\frac{1}{\sigma(\hat{X}_i)}+\sqrt{\Delta t}\cdot\xi_{i+1}\cdot\frac{\mu'(\hat{X}_i)}{\sigma(\hat{X}_i)}\Big\}\cdot\hat{Y}_i|\mathcal{F}_{i_j}] \\ =\;& E[\frac{d}{d\hat{X}_{t_j}}E[\Phi(\hat{X}_{t_1},\cdots,\hat{X}_{t_j},\hat{X}_{t_{j+1}},\cdots,\hat{X}_{t_m})|\mathcal{F}_{t_j}]]\cdot\hat{Y}_{t_j}\end{aligned}$$

Now we take expectations on both sides of the equation to show that (21) holds. 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 $1 \le j \le m$, the expectation of the estimator is

$$\begin{aligned}
& E[\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \sum_{i=1}^{N} a_i \sqrt{\Delta t} \xi_i \cdot \frac{\hat{Y}_{i-1}}{\sigma(\hat{X}_{i-1})} + \mathrm{I} + \mathrm{II}] \\
= & \sum_{i=1}^{N} a_i \Delta t \cdot E[\Phi \cdot \Big\{ (\xi_i^2 - 1) \cdot \frac{\sigma'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} + \frac{\xi_i}{\sqrt{\Delta t} \cdot \sigma(\hat{X}_{i-1})} + \sqrt{\Delta t} \cdot \xi_i \cdot \frac{\mu'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} \Big\} \cdot \hat{Y}_i] \\
= & \sum_{i=1}^{i_1} a_i \Delta t \cdot E[\Phi \cdot \Big\{ (\xi_i^2 - 1) \cdot \frac{\sigma'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} + \frac{\xi_i}{\sqrt{\Delta t} \cdot \sigma(\hat{X}_{i-1})} + \sqrt{\Delta t} \cdot \xi_i \cdot \frac{\mu'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} \Big\} \cdot \hat{Y}_i] \\
= & \sum_{i=1}^{i_1} a_i \Delta t \cdot \mathrm{delta} = \mathrm{delta},
\end{aligned}$$

where the third equality holds because of (21) and the identity $\sum_{i=i_j}^{i_{j+1}} a_i \Delta t = 0$. Q. E. D.

3.2 Vega

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

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

We will want to divide by values of the process $\hat{Y}$, and to do this we must restrict the process $\hat{Y}$ to be positive almost surely. This does not hold for the Euler scheme with normally distributed increments, so we use the truncated normals $\tilde{\xi}_i$ in this subsection.

Theorem 3.2 *Suppose Assumptions 3.1 and 3.2 hold and there exists a positive number w^* such that $w_R = w^*/\sqrt{\Delta t}$ and $1 - \sup_x |\sigma'(x)| w^* > 0$. Then, 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 $0 \le j \le m$,*

$$\begin{aligned}
vega \;=\; & E[\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \Big\{ \sum_{k=1}^{m} \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}} \Big] \cdot \sum_{i=i_{k-1}+1}^{i_k} a_i \frac{\hat{Y}_{i-1}}{\sigma(\hat{X}_{i-1})} \cdot \frac{\sqrt{\Delta t} \cdot \tilde{\xi}_i}{\delta} - \frac{d}{dx} \Big[\frac{\hat{Z}_{i_m}}{\hat{Y}_{i_m}} \Big] \Big\}] \\
+ \; & E[\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \{ \sum_{k=1}^{m} \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}} \Big] \cdot \sum_{i=i_{k-1}+1}^{i_k} a_i \Delta t \cdot (III + IV) \}] \\
+ \; & \frac{C}{\sqrt{2\pi\delta}} e^{-\frac{w_R^2}{2\delta}} \sum_{k=1}^{m} \sum_{i=i_{k-1}+1}^{i_k} a_i \sqrt{\Delta t} E[\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}} \Big] \cdot \frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})} \Big|_{\tilde{\xi}_i = w_R}] \\
- \; & \frac{C}{\sqrt{2\pi\delta}} e^{-\frac{w_L^2}{2\delta}} \sum_{k=1}^{m} \sum_{i=i_{k-1}+1}^{i_k} a_i \sqrt{\Delta t} E[\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}} \Big] \cdot \frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})} \Big|_{\tilde{\xi}_i = w_L}]
\end{aligned}$$

where

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

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

Proof. We first show that this result holds for Φ with compact support and a continuous derivative. Because $\partial\Phi/\partial x_{i_j}$ is continuous in a compact set, it is bounded in this set. Using Lemma A.1, we can take the derivative inside the expectation to get

$$\text{vega} \quad = \quad \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} E[\Phi(\hat{X}^{\varepsilon}_{i_1}, \cdots, \hat{X}^{\varepsilon}_{i_m})] = E[\sum_{j=1}^{m} \frac{\partial\Phi}{\partial\hat{X}_{i_j}}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \hat{Z}_{i_j}] \tag{22}$$

Under the choice of w_R, $\hat{Y}_i$ is positive almost surely for all sufficiently large N because

$$\hat{Y}_i = \prod_{j=1}^{i}(1 + \mu'(\hat{X}_{j-1})\Delta t + \sigma'(\hat{X}_{j-1})\sqrt{\Delta t}\tilde{\xi}_j) \geq \prod_{j=1}^{i}(1 - \frac{\sup_x |\mu'(x)|}{N} - \sup_x |\sigma'(x)| w^*) > 0.$$

Thus, we can divide and multiply by $\hat{Y}_i$ simultaneously in the expectation of (22) to get

$$\text{vega} = E[\sum_{j=1}^{m} \frac{\partial\Phi}{\partial\hat{X}_{i_j}}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \frac{\hat{Z}_{i_j}}{\hat{Y}_{i_j}} \cdot \hat{Y}_{i_j}]. \tag{23}$$

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

$$\frac{\hat{Z}_{i_j}}{\hat{Y}_{i_j}} \cdot \hat{Y}_{i_j} = \sum_{k=1}^{j} \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \hat{Y}_{i_j} = \sum_{k=1}^{j} \Big[\sum_{i=i_{k-1}+1}^{i_k} a_i \Delta t\Big] \cdot \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \hat{Y}_{i_j}. \tag{24}$$

Plugging (24) back into (23), we have

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

If we interchange the order of the two sums in the expectation, we get

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

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

$$\begin{aligned}
\text{vega} &= \sum_{k=1}^{m} \sum_{i=i_{k-1}+1}^{i_k} a_i \Delta t \cdot E\Big[\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \sum_{j=k}^{m} \frac{\partial \Phi}{\partial \hat{X}_{i_j}}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \frac{\hat{Y}_{i_j}}{\hat{Y}_i} \cdot \hat{Y}_i\Big] \\
&= \sum_{k=1}^{m} \sum_{i=i_{k-1}+1}^{i_k} a_i \Delta t \cdot E\Big[\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \sum_{j=k}^{m} \frac{\partial \Phi}{\partial \hat{X}_{i_j}}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \frac{d\hat{X}_{i_j}}{d\hat{X}_i} \cdot \hat{Y}_i\Big] \\
&= \sum_{k=1}^{m} \sum_{i=i_{k-1}+1}^{i_k} a_i \Delta t \cdot E\Big[\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \frac{d\Phi}{d\hat{X}_i}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \hat{Y}_i\Big]
\end{aligned}$$

where the last equality holds by the chain rule of ordinary calculus.

Let $\hat{\mathcal{F}}_i = \sigma(\tilde{\xi}_1, \cdots, \tilde{\xi}_{i-1}, \tilde{\xi}_{i+1}, \cdots, \tilde{\xi}_N)$, the σ-algebra generated by all increments except $\tilde{\xi}_i$. Note that $d\hat{X}_i/d\tilde{\xi}_i = \sigma(\hat{X}_{i-1})\sqrt{\Delta t}$, so

$$\begin{aligned}
& E\Big[\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \frac{d\Phi}{d\hat{X}_i}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \hat{Y}_i\Big] = E\Big[E\Big[\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \frac{d\Phi}{d\hat{X}_i}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \hat{Y}_i | \hat{\mathcal{F}}_i\Big]\Big] \\
= \;& E\Big[E\Big[\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \frac{d\Phi}{d\hat{X}_i}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \hat{Y}_i | \hat{\mathcal{F}}_i\Big]\Big] \\
= \;& E\Big[E\Big[\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \frac{d\Phi}{d\tilde{\xi}_i}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \frac{\hat{Y}_i}{d\hat{X}_i/d\tilde{\xi}_i} | \hat{\mathcal{F}}_i\Big]\Big] \\
= \;& E\Big[E\Big[\int_{w_L}^{w_R} \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \frac{d\Phi}{dw_i}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})\sqrt{\Delta t}} \cdot \frac{C}{\sqrt{2\pi\delta}} \cdot e^{-\frac{w_i^2}{2\delta}} dw_i | \hat{\mathcal{F}}_i\Big]\Big]. \qquad (25)
\end{aligned}$$

We can evaluate the integral as

$$\begin{aligned}
& \int_{w_L}^{w_R} \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \frac{d\Phi}{dw_i}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \hat{Y}_i \cdot e^{-\frac{w_i^2}{2\delta}} dw_i \\
= \;& \int_{w_L}^{w_R} \Phi \cdot e^{-\frac{w_i^2}{2\delta}} \cdot \Big\{\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \hat{Y}_i \cdot \frac{w_i}{\delta} - \frac{d}{dw_i}\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \hat{Y}_i - \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \frac{d\hat{Y}_i}{dw_i}\Big\} dw_i \\
+ \;& \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \Phi \cdot \hat{Y}_i \cdot e^{-\frac{w_i^2}{2\delta}}\Big|_{w_i=w_R} - \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \Phi \cdot \hat{Y}_i \cdot e^{-\frac{w_i^2}{2\delta}}\Big|_{w_i=w_L} \qquad (26)
\end{aligned}$$

Thus, combining (25) and (26), we have

$$
\begin{aligned}
& E\Big[\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \frac{d\Phi}{d\hat{X}_i}(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \hat{Y}_i\Big] \\
= \;& E\Big[\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \Big\{\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \frac{\tilde{\xi}_i}{\delta} - \frac{d}{d\tilde{\xi}_i}\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big]\Big\} \cdot \frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})\sqrt{\Delta t}}\Big] \\
- \;& E\Big[\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \hat{Y}_{i-1} \cdot \frac{\sigma'(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}\Big] \\
+ \;& \frac{C}{\sqrt{2\pi\delta}} \cdot \Big\{\Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \Phi \cdot \frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})\sqrt{\Delta t}}\Big|_{w_i=w_R} - \Big[\frac{\hat{Z}_{i_k}}{\hat{Y}_{i_k}} - \frac{\hat{Z}_{i_{k-1}}}{\hat{Y}_{i_{k-1}}}\Big] \cdot \Phi \cdot \frac{\hat{Y}_i}{\sigma(\hat{X}_{i-1})\sqrt{\Delta t}}\Big|_{w_i=w_L}\Big\} \cdot e^{-\frac{w_R^2}{2\delta}}.
\end{aligned}
$$

Because

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

and $\hat{Y}_i = \hat{Y}_{i-1} + \mu'(\hat{X}_{i-1})\hat{Y}_{i-1}\Delta t + \sigma'(\hat{X}_{i-1})\hat{Y}_{i-1}\sqrt{\Delta t} \cdot \tilde{\xi}_i$, the conclusion follows.

To finish the proof, we can apply Lemma B.1 to extend the conclusion to a general payoff function Φ satisfying Assumption 3.1. Q. E. D.

3.3 Rho

We define rho to be the derivative of $u(x)$ with respect to perturbation ε that take the drift $\mu(\cdot)$ to $\mu(\cdot) + \varepsilon\gamma(\cdot)$. In other words,

$$
\text{rho} = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} E[\Phi(\hat{X}^\varepsilon_{i_1}, \cdots, \hat{X}^\varepsilon_{i_m})]
$$

where

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

Let U^ε be

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

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, $P^\varepsilon(d\omega) = U^\varepsilon P(d\omega)$. Moreover, it is easy to see by direct calculation (cf. Lemma C.1) that this change of measure corresponds to a change of drift, in the sense that

$$
E[\Phi(\hat{X}^\varepsilon_{i_1}, \cdots, \hat{X}^\varepsilon_{i_m})] = E[\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot U^\varepsilon]. \tag{27}
$$

Theorem 3.3 *Suppose that Assumptions 3.1, 3.2 and 3.3 hold. Then,*

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

Proof. Fix a compact neighborhood of $\varepsilon = 0$, say, K. By (27) and the Cauchy-Schwarz inequality,

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

The mean-value theorem implies that, for any $\varepsilon \in K$, there exists a random variable $\theta(\varepsilon) \in K$ such that

$$\begin{aligned}
& \frac{1}{\varepsilon}(U^{\varepsilon} - 1) - \sum_{i=1}^{N} \frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} \sqrt{\Delta t} \xi_i \\
= \ & U^{\theta(\varepsilon)} \cdot [(\sum_{i=1}^{N} \frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} \sqrt{\Delta t} \xi_i - \theta(\varepsilon) \sum_{i=1}^{N} [\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}]^2 \Delta t)^2 - \sum_{i=1}^{N} [\frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})}]^2 \Delta t]
\end{aligned}$$

The function γ is bounded according to Assumption 3.3 and $1/\sigma$ is also bounded by the nondegeneracy assumption 3.2. In addition, $\theta(\varepsilon)$ is in a compact neighborhood of $\varepsilon = 0$. Thus, the right hand side will be bounded by

$$\exp(C_1 \sum_{i=1}^{N} \sqrt{\Delta t} |\xi_i| + C_2 T) \cdot (C_3 \sum_{i=1}^{N} \Delta t \cdot \xi_i^2 + C_4 \sum_{i=1}^{N} \sqrt{\Delta t} |\xi_i| + C_5 T)$$

where $C_j, 1 \leq j \leq 5$ are constants which do not depend on the realization of ξ_i and ε. It is easy to show that this quantity is L^2 integrable, which implies

$$\sup_{\varepsilon \in K} E|\frac{1}{\varepsilon}(U^{\varepsilon} - 1) - \sum_{i=1}^{N} \frac{\gamma(\hat{X}_{i-1})}{\sigma(\hat{X}_{i-1})} \sqrt{\Delta t} \xi_i|^2 < +\infty$$

Combining with the fact that Φ is L^2, we have shown that

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

Q. E. D.

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 D. For further background, see Jacod and Shiryayev [11] and Kurtz and Protter [15], [16].

In this section, we attach the index N to all processes for emphasis. 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 random variables $\{\xi_i^{(N)}, 1 \le i \le N\}$ or $\{\tilde{\xi}_i^{(N)}, 1 \le i \le N\}$, where $\mathcal{F}_t^{(N)} = \sigma(\xi_1^{(N)}, \cdots, \xi_{[Nt/T]}^{(N)})$ or $\sigma(\tilde{\xi}_1^{(N)}, \cdots, \tilde{\xi}_{[Nt/T]}^{(N)})$. The following technical result about the tightness of the processes

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

will be used repeatedly. For the definition of P-UT, see Appendix D.

Lemma 4.1 *Processes $\hat{L}^{(N)}$ and $\tilde{L}^{(N)}$ are P-UT.*

Proof. 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 the martingale inequality,

$$P(|\int_0^t H_{s-}^{(N)} d\hat{L}_s^{(N)}| > a) \le \frac{E[\hat{L}^{(N)}, \hat{L}^{(N)}]_t}{a^2} = \frac{\Delta t^{(N)}}{a^2} \sum_{j=1}^{[Nt/T]} \text{Var}[\xi_j^{(N)}].$$

Because the variance of $\xi_j^{(N)}$ is 1, the right side of the inequality above will be

$$\frac{\Delta t^{(N)}}{a^2} \sum_{j=1}^{[Nt/T]} \text{Var}[\xi_j^{(N)}] = \frac{t/T}{a^2}$$

and will converge to 0 as $a \to +\infty$. In addition, notice that the right side does not depend on H and N. Thus,

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

i.e., $\hat{L}^{(N)}$ is P-UT. Q.E.D.

4.1 Preliminary Convergence Results

First, we establish the result of weak convergence of processes $(X^{(N)}, Y^{(N)}, Z^{(N)})$ with increments $\xi^{(N)}$ or $\tilde{\xi}^{(N)}$. For this, we need the following conditions on the truncation of the normal distribution:

Assumption 4.1 $w_R^{(N)} = -w_L^{(N)} = \sqrt{N}w^*$ *where* w^* *is a positive number satisfying* $1 - \sup_x |\sigma'(x)|w^* > 0$.

Lemma 4.2 *Suppose Assumptions 3.1, 3.2 and Assumption 4.1 hold. As* $N \to +\infty$*, we have*

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

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

$$\begin{aligned} dX_t &= \mu(X_t)dt + \sigma(X_t)dW_t; \\ dY_t &= \mu'(X_t)Y_t dt + \sigma'(X_t)Y_t dW_t; \\ dZ_t &= \mu'(X_t)Z_t dt + \sigma'(X_t)Z_t dW_t + \tilde{\sigma}(X_t)dW_t. \end{aligned}$$

Proof. Introduce functions

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

Thus, $(X^{(N)}, Y^{(N)}, Z^{(N)})$ and (X, Y, Z) are the solutions to the SDEs:

$$M_t^{(N)} = \int_0^t g^{(N)}(M_{s-}^{(N)})d\hat{L}_s^{(N)} \quad (\text{or} \ \ d\tilde{L}_s^{(N)})$$

and

$$M_t = \int_0^t g(M_{s-})dW_s,$$

respectively. From Donsker's invariance principle, we know that both of $\hat{L}^{(N)}$ and $\tilde{L}^{(N)}$ weakly converge to Brownian motion W. By Lemma 4.1, the processes $\hat{L}^{(N)}$ and $\tilde{L}^{(N)}$ are P-UT. Thus, applying Lemma D.3, we have the conclusion that $(\hat{X}^{(N)}, \hat{Y}^{(N)}, \hat{Z}^{(N)}) \Rightarrow (X, Y, Z)$ in the Skorohod topology if the solution (X, Y, Z) exists globally. Q. E. D.

Remark 4.3 *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*

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

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

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

Lemma 4.4 *Under Assumptions 3.1, 3.2, and Assumptions 4.1–4.3,*

$$I^{(N)} := \Phi(\hat{X}^{(N)}_{i_1}, \cdots, \hat{X}^{(N)}_{i_m}) \cdot \sum_{i=1}^{N} a_i^{(N)} \cdot \frac{\sigma'(\hat{X}^{(N)}_{i-1})}{\sigma(\hat{X}^{(N)}_{i-1})} \cdot [(\sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)})^2 - \Delta t^{(N)}] \Rightarrow 0$$

and

$$II^{(N)} := \Phi(\hat{X}^{(N)}_{i_1}, \cdots, \hat{X}^{(N)}_{i_m}) \cdot (\sum_{i=1}^{N} a_i^{(N)} \cdot \frac{\mu'(\hat{X}^{(N)}_{i-1})}{\sigma(\hat{X}^{(N)}_{i-1})} \sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)}) \cdot \Delta t^{(N)} \Rightarrow 0.$$

Proof. Notice that $E[((\xi_i^{(N)})^2 - 1)((\xi_j^{(N)})^2 - 1)] = 0$ for all $i \neq j$ and $E[((\xi_i^{(N)})^2 - 1)^2] = 2$ because $\xi_i^{(N)}$ and $\xi_j^{(N)}$ are independent standard normal random variables. We have, by the fact that σ is not degenerate and σ' is bounded,

$$\begin{aligned} & E|\sum_{i=1}^{N} a_i^{(N)} \cdot \frac{\sigma'(\hat{X}^{(N)}_{i-1})}{\sigma(\hat{X}^{(N)}_{i-1})} \cdot [(\sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)})^2 - \Delta t^{(N)}]|^2 \\ \le\ & C_1 \Delta t^{(N)} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i^{(N)} a_j^{(N)} E[((\xi_i^{(N)})^2 - 1)((\xi_j^{(N)})^2 - 1)] \Delta t^{(N)} \\ =\ & 2C_1 \Delta t^{(N)} \cdot \sum_{i=1}^{N} (a_i^{(N)})^2 \Delta t^{(N)}. \end{aligned}$$

By Assumption 4.3 and the triangle inequality,

$$\sup_N \sum_{i=1}^{N} (a_i^{(N)})^2 \Delta t^{(N)} \le \sup_N \sup_{0 \le t \le T} |a^{(N)}_{[Nt/T]} - a_t| + \int_0^T a_t^2 dt < +\infty.$$

Thus, $\sum_{i=1}^N a_i^{(N)} \cdot \sigma'(\hat{X}_{i-1}^{(N)})/\sigma(\hat{X}_{i-1}^{(N)}) \cdot [(\sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)})^2 - \Delta t^{(N)}]$ converges to 0 in L^2 as $N \to \infty$, which also implies weak convergence in the Skorohod topology. On the other hand, using Assumption 4.2, $\Phi(\hat{X}_{i_1}^{(N)}, \cdots, \hat{X}_{i_m}^{(N)}) \Rightarrow \Phi(X_{t_1}, \cdots, X_{t_m})$. Combining both, we have $I^{(N)} \Rightarrow 0$.

For $II^{(N)}$, applying Lemma D.3 and using the fact that $(\hat{X}^{(N)}, \hat{Y}^{(N)}, \hat{Z}^{(N)})$ weakly converges, we have that $\sum_{i=1}^N a_i^{(N)} \cdot \mu'(\hat{X}_{i-1}^{(N)})/\sigma(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)}$ weakly converges. Thus,

$$(\sum_{i=1}^N a_i^{(N)} \cdot \frac{\mu'(\hat{X}_{i-1}^{(N)})}{\sigma(\hat{X}_{i-1}^{(N)})}\sqrt{\Delta t^{(N)}} \cdot \xi_i^{(N)}) \cdot \Delta t^{(N)} \Rightarrow 0.$$

Using Assumption 4.2 again, we find that $II^{(N)} \Rightarrow 0$. Q. E. D.

We now come to the main theorem of this subsection.

Theorem 4.5 *Under Assumptions 3.1, 3.2, 4.1, 4.2 and 4.3, the delta estimator (19) converges weakly to*

$$\Phi(X_{t_1}, \cdots, X_{t_m}) \cdot \int_0^T a_t \frac{Y_t}{\sigma(X_t)} dW_t.$$

Proof. Define functions

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

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

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

because $y/\sigma(x)$ is bounded when $(x, y) \in K$. Thus (32) holds. Applying Lemma D.3, we have

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

Combining this with Lemma 4.4, the theorem follows. Q. E. D.

4.3 Convergence of Vega Estimators

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

Lemma 4.6 *Under assumptions 3.1, 3.2 and Assumptions 4.1–4.3,*

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

and

$$\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot IV^{(N)} := \sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)}) \hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot [\frac{(\tilde{\xi}_i^{(N)})^2}{\delta^{(N)}} - 1] \Rightarrow 0.$$

Proof. Using arguments similar to those in Theorem 4.5, it is easy to show that

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

On the other hand, by Assumption 4.1, $|w_R| = |w_L| = \sqrt{N}w^* \to +\infty$. Thus, both $C^{(N)}$ and $\delta^{(N)}$ in the definition of the truncated normal converge to 1. Then we have

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

Using the fact that σ' is bounded and that σ is nondegenerate, and letting $\tau_b^{(N)} = \inf\{j : |\hat{Y}_j^{(N)}| > b\}$ for any $b > 0$,

$$\begin{aligned}
& E[|\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot [\frac{(\tilde{\xi}_i^{(N)})^2}{\delta^{(N)}} - 1] \cdot 1_{\{\tau_b^{(N)} \geq i\}}|^2] \\
\leq\ & \frac{2}{(\delta^{(N)})^2} E[|\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot [(\tilde{\xi}_i^{(N)})^2 - 1] \cdot 1_{\{\tau_b^{(N)} \geq i\}}|^2] \\
& +2(\frac{1}{(\delta^{(N)})^2} - 1)^2 \cdot E[|\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot \frac{\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})} \cdot 1_{\{\tau_b^{(N)} \geq i\}}|^2].
\end{aligned} \tag{28}$$

Because the variance of $\tilde{\xi}_i^{(N)}$ is 1 and $\{\tau_b^{(N)} \geq i\}$ is $\mathcal{F}_{i-1}^{(N)}$-measurable, the sum in the first term of the above inequality is a martingale. By the orthogonality of martingale differences,

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

Noting that the variance of $\tilde{\xi}_i^{(N)}$ is 1 and the fourth moment of $\tilde{\xi}_i^{(N)}$ is less than 3, we get

$$E[|(\tilde{\xi}_i^{(N)})^2 - 1|^2] = E[|\Delta\tilde{\xi}_i^{(N)}|^4] - 2E[|\Delta\tilde{\xi}_i^{(N)}|^2] + 1 \leq 2.$$

Thus the first term in (28) converges to 0. It is easy to see that the second term in (28) also converges to 0 because $\delta^{(N)} \to 1$ and $\sup_N E[|\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot (\sigma'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)})/\sigma(\hat{X}_{i-1}^{(N)}) \cdot 1_{\{\tau_b^{(N)} \geq i\}}|^2] < +\infty$. Thus, the left hand side of (28) converges to 0 as $N \to +\infty$.

To show the weak convergence of $\sum_{i=i_{k-1}+1}^{i_k} a_i^{(N)} \Delta t^{(N)} \cdot IV^{(N)}$, it remains to verify that $P^{(N)}(\tau_b^{(N)} \geq N) \to 1$. This holds in light of Remark 4.3 after Lemma 4.2. Q. E. D.

Lemma 4.7 *Under all assumptions in Section 3 and Assumptions 4.1–4.3, we can find some constant C (not depending on N) such that*

$$\sup_{1\le i\le N} E[\Phi(\hat{X}_{i_1}^{(N)}, \cdots, \hat{X}_{i_m}^{(N)}) \cdot \Big[\frac{\hat{Z}_{i_k}^{(N)}}{\hat{Y}_{i_k}^{(N)}} - \frac{\hat{Z}_{i_{k-1}}^{(N)}}{\hat{Y}_{i_{k-1}}^{(N)}}\Big] \cdot \frac{\hat{Y}_i^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})}\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}(or\ w_L^{(N)})}] \le CN^p$$

where p is the growth rate of the function Φ in Assumption 3.1.

Proof. Because the arguments are symmetric, we consider only the conditional expectation on the condition that $\tilde{\xi}_i^{(N)} = w_R^{(N)}$.

By the definition of $\hat{Y}^{(N)}$ and $\hat{Z}^{(N)}$, we know that

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

Using the arithmetic-geometric mean inequality,

$$abcd \le \frac{1}{4}(a^4 + b^4 + c^4 + d^4), \text{ for } a, b, c, d > 0,$$

the polynomial growth of Φ, the nondegeneracy of σ, and the boundedness of $\tilde{\sigma}$, we get that there exists some positive number C_1 which is not dependent on N such that

$$\begin{aligned}
& E[|\Phi(\hat{X}_{i_1}^{(N)}, \cdots, \hat{X}_{i_m}^{(N)}) \cdot \Big[\frac{\hat{Z}_{i_k}^{(N)}}{\hat{Y}_{i_k}^{(N)}} - \frac{\hat{Z}_{i_{k-1}}^{(N)}}{\hat{Y}_{i_{k-1}}^{(N)}}\Big] \cdot \frac{\hat{Y}_i^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})}|\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}] \\
\le\ & C_1 \sum_{j=i_{k-1}+1}^{i_k} E[|\Phi(\hat{X}_{i_1}^{(N)}, \cdots, \hat{X}_{i_m}^{(N)})| \cdot \frac{\sqrt{\Delta t^{(N)}}|\tilde{\xi}_j^{(N)}|}{|\hat{Y}_j^{(N)}|} \cdot |\hat{Y}_i^{(N)}|\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}] \\
\le\ & \frac{C_1}{4} \sum_{j=i_{k-1}+1}^{i_k} \{E[|\Phi(\hat{X}_{i_1}^{(N)}, \cdots, \hat{X}_{i_m}^{(N)})|^4\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}] + (\Delta t^{(N)})^2 E[|\tilde{\xi}_j^{(N)}|^4\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}] \\
& + E[|\hat{Y}_j^{(N)}|^{-4}\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}] + E[|\hat{Y}_i^{(N)}|^4\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}]\}
\end{aligned}$$

By Lemma E.1, Lemma E.2 and Lemma E.3, we know that the right side of the above inequality is bounded by

$$
\begin{aligned}
& \frac{C_1}{4}\sum_{j=i_{k-1}+1}^{i_k}\{E[\|(\hat{X}_{i_1}^{(N)},\cdots,\hat{X}_{i_m}^{(N)})\|^{4p}\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}]+(\Delta t^{(N)})^2E[|\tilde{\xi}_j^{(N)}|^4\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}] \\
& +\sup_j E[|\hat{Y}_j^{(N)}|^{-4}\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}]+\sup_j E[|\hat{Y}_j^{(N)}|^{4}\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}]\} \\
\leq\ & \frac{C_1}{4}\sum_{j=i_{k-1}+1}^{i_k}\{\sum_{k=1}^{m}E[(\hat{X}_{i_k}^{(N)})^{2p}\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}]+(\Delta t^{(N)})^2E[|\tilde{\xi}_j^{(N)}|^4\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}]+C_2+C_3\} \\
\leq\ & \frac{C_1}{4}\sum_{j=i_{k-1}+1}^{i_k}\{mC_4N^{p-1}+C_5+C_2+C_3\}
\end{aligned}
$$

where C_2, C_3, C_4 and C_5 are constants not dependent on N. Thus, we can find some $C>0$

$$
E[|\Phi(\hat{X}_{i_1}^{(N)},\cdots,\hat{X}_{i_m}^{(N)})\cdot\Big[\frac{\hat{Z}_{i_k}^{(N)}}{\hat{Y}_{i_k}^{(N)}}-\frac{\hat{Z}_{i_{k-1}}^{(N)}}{\hat{Y}_{i_{k-1}}^{(N)}}\Big]\cdot\frac{\hat{Y}_i^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})}|\Big|_{\tilde{\xi}_i^{(N)}=w_R^{(N)}}]\leq CN^p
$$

Q. E. D.

Theorem 4.8 *Under Assumptions 3.1, 3.2 and Assumptions 4.1–4.3, the estimator in Theorem 3.2 converges weakly to*

$$
\Phi(X_{t_1},\cdots,X_{t_m})\cdot\Big\{\sum_{k=1}^{m}\Big[\frac{Z_{t_k}}{Y_{t_k}}-\frac{Z_{t_{k-1}}}{Y_{t_{k-1}}}\Big]\cdot\int_{t_{k-1}}^{t_k}a_s\frac{Y_s}{\sigma(X_s)}dW_s-\frac{d}{dx}\Big[\frac{Z_{t_m}}{Y_{t_m}}\Big]\Big\}
$$

Proof. Using Lemma 4.6 and Lemma 4.7, we only need to show that

$$
\sum_{i=i_{k-1}+1}^{i_k}a_i^{(N)}\frac{\hat{Y}_{i-1}^{(N)}}{\sigma(\hat{X}_{i-1}^{(N)})}\cdot\frac{\sqrt{\Delta t^{(N)}}\cdot\tilde{\xi}_i^{(N)}}{\delta^{(N)}}\Rightarrow\int_{t_{k-1}}^{t_k}a_s\frac{Y_s}{\sigma(X_s)}dW_s \tag{29}
$$

and

$$
\frac{d}{dx}\Big[\frac{Z_{i_m}^{(N)}}{Y_{i_m}^{(N)}}\Big]\Rightarrow\frac{d}{dx}\Big[\frac{Z_{t_m}}{Y_{t_m}}\Big]. \tag{30}
$$

Equation (29) holds by arguments similar to those used in the proof of Theorem 4.5 and the fact that $\tilde{L}^{(N)}$ is P-UT (cf. Lemma 4.1).

Notice that

$$
\begin{aligned}
\frac{d\hat{Z}_i^{(N)}}{dx} &= \frac{d\hat{Z}_{i-1}^{(N)}}{dx}+[\mu''(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}\hat{Z}_{i-1}^{(N)}+\mu'(\hat{X}_{i-1}^{(N)})\frac{d\hat{Z}_{i-1}^{(N)}}{dx}]\Delta t^{(N)}+\tilde{\sigma}'(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}\tilde{\xi}_i^{(N)} \\
&\quad+[\sigma''(\hat{X}_{i-1}^{(N)})\hat{Y}_{i-1}^{(N)}\hat{Z}_{i-1}^{(N)}+\sigma'(\hat{X}_{i-1}^{(N)})\frac{d\hat{Z}_{i-1}^{(N)}}{dx}]\tilde{\xi}_i^{(N)}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{d\hat{Y}_i^{(N)}}{dx} &= \frac{d\hat{Y}_{i-1}^{(N)}}{dx} + [\mu''(\hat{X}_{i-1}^{(N)})(\hat{Y}_{i-1}^{(N)})^2 + \mu'(\hat{X}_{i-1}^{(N)})\frac{d\hat{Y}_{i-1}^{(N)}}{dx}]\Delta t^{(N)} + [\sigma''(\hat{X}_{i-1}^{(N)})(\hat{Y}_{i-1}^{(N)})^2 \\
&\quad +\sigma'(\hat{X}_{i-1}^{(N)})\frac{d\hat{Y}_{i-1}^{(N)}}{dx}]\tilde{\xi}_i^{(N)}.
\end{aligned}
$$

Using arguments similar to those in Lemma 4.2, we can show that $(d\hat{Z}^{(N)}/dx, d\hat{Y}^{(N)}/dx) \Rightarrow (dZ/dx, dY/dx)$. Thus,

$$
\frac{d}{dx}\Big[\frac{\hat{Z}_{i_m}^{(N)}}{\hat{Y}_{i_m}^{(N)}}\Big] = \frac{d\hat{Z}_{i_m}^{(N)}}{dx}\cdot\frac{1}{\hat{Y}_{i_m}^{(N)}} + \frac{\hat{Z}_{i_m}^{(N)}}{(\hat{Y}_{i_m}^{(N)})^2}\cdot\frac{d\hat{Y}_{i_m}^{(N)}}{dx} \Rightarrow \frac{dZ_{t_m}}{dx}\cdot\frac{1}{Y_{t_m}} + \frac{Z_{t_m}}{(Y_{t_m})^2}\cdot\frac{dY_{t_m}}{dx} = \frac{d}{dx}\Big[\frac{Z_{t_m}}{Y_{t_m}}\Big].
$$

Q. E. D.

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 4.5 and 4.8, we just list the theorem as follows without detailed justification:

Theorem 4.9 *Under all assumptions in Section 3 and Assumption 4.2, the estimator given in Theorem 3.3 converges weakly to*

$$
\Phi(X_{t_1},\cdots,X_{t_m})\cdot\int_0^T \frac{\gamma(X_s)}{\sigma(X_s)}dW_s.
$$

The rest of this article presents technical results in appendices.

A Differentiating Under the Integral

At several places in the paper we need to interchange the order of differentiation and integration. We justify such interchanges in this appendix.

Lemma A.1 (Theorem A.9.1, Durrett [6]) *Let $(S,\mathcal{S},\mu)$ be a probability measure space. Let f be a real valued function defined on $\mathbf{R}\times S$. Let $\epsilon > 0$ and suppose that for $x \in (y-\epsilon, y+\epsilon)$ we have*
1) $u(x) = \int_S f(x,s)\mu(ds)$ with $\int_S |f(x,s)|\mu(ds) < +\infty$;
2) for fixed s, $\partial f/\partial x$ exists and is a continuous function of x;
3) $v(x) = \int_S \frac{\partial f}{\partial x}(x,s)\mu(ds)$ is continuous at $x = y$; and
4) $\int_S\int_{-\delta}^{\delta}|\frac{\partial f}{\partial x}(y+\theta,s)|d\theta\mu(ds) < +\infty$
then $u'(y) = v(y)$.

The following two lemmas are used in the derivation of estimator delta:

Lemma A.2 *Suppose that X is normally distributed with mean $b(x)$ and variance $\sigma^2(x)$, where $b(\cdot)$ and $\sigma(\cdot)$ are both continuous and differentiable, and σ is bounded away from zero. Suppose f satisfies the following property: there exists some positive integer p such that*

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

Then,

$$\frac{d}{dx}\bigg|_{x=x_0} E[f(X)] = E[f(\sigma(x_0)\xi + b(x_0)) \cdot \{(\xi^2 - 1) \cdot \frac{\sigma'(x_0)}{\sigma(x_0)} + \xi \cdot \frac{b'(x_0)}{\sigma(x_0)}\}]$$

where ξ is normally distributed with mean 0 and variance 1.

Proof. Fix a neighborhood of x_0, say, $[x_0 - \epsilon, x_0 + \epsilon]$. Because b and σ are continuous, each attains its minimum and maximum on this interval. Thus, there are positive constants $C_1, C_2, C_3, C_4, C_5, C_6$, which are not dependent on x, such that

$$\begin{aligned}
& |\frac{d}{dx}[f(y) \cdot \frac{1}{(2\pi)^{1/2}\sigma(x)} \cdot \exp(-\frac{(y-b(x))^2}{2\sigma^2(x)})]| \\
= \ & |f(y) \cdot \frac{1}{(2\pi)^{1/2}\sigma(x)} \cdot \exp(-\frac{(y-b(x))^2}{2\sigma^2(x)}) \cdot \Big\{\Big(\frac{(y-b(x))^2}{\sigma^2(x)} - 1\Big) \cdot \frac{\sigma'(x)}{\sigma(x)} + \frac{y-b(x)}{\sigma(x)} \cdot \frac{b'(x)}{\sigma(x)}\Big\}| \\
\leq \ & C_1 \cdot |f(y)| \cdot \exp(-C_2 y^2 + C_3|y|) \cdot (C_4 y^2 + C_5|y| + C_6)
\end{aligned}$$

for all $x \in [x_0 - \epsilon, x_0 + \epsilon]$. Because f grows polynomially, the right side of this inequality is integrable. Thus Assumption 4) in Lemma A.1 holds. Now we can take derivative at $x = x_0$ under integral to get

$$\begin{aligned}
\frac{d}{dx}E[f(X)] &= \frac{d}{dx}\int f(y) \cdot \frac{1}{(2\pi)^{1/2}\sigma(x_0)} \cdot \exp(-\frac{(y-b(x_0))^2}{2\sigma^2(x_0)})dy \\
&= \int f(y) \cdot \frac{d}{dx}\frac{1}{(2\pi)^{1/2}\sigma(x_0)} \cdot \exp(-\frac{(y-b(x_0))^2}{2\sigma^2(x_0)})dy \\
&= \int f(y) \cdot \frac{1}{(2\pi)^{1/2}\sigma(x_0)} \cdot \exp(-\frac{(y-b(x_0))^2}{2\sigma^2(x_0)}) \\
&\quad \times \Big\{\Big(\frac{(y-b(x_0))^2}{\sigma^2(x_0)} - 1\Big) \cdot \frac{\sigma'(x_0)}{\sigma(x_0)} + \frac{y-b(x_0)}{\sigma(x_0)} \cdot \frac{b'(x_0)}{\sigma(x_0)}\Big\}dy
\end{aligned}$$

Let w satisfy $y = \sigma(x_0)w + b(x_0)$. Then,

$$\begin{aligned}
\frac{d}{dx}E[f(X)] &= \int f(\sigma(x_0)w + b(x_0)) \cdot \frac{1}{(2\pi)^{1/2}} \cdot \exp(-\frac{1}{2}w^2) \cdot \{(w^2 - 1) \cdot \frac{\sigma'(x_0)}{\sigma(x_0)} + w \cdot \frac{b'(x_0)}{\sigma(x_0)}\}dw \\
&= E[f(\sigma(x_0)\xi + b(x_0)) \cdot \{(\xi^2 - 1) \cdot \frac{\sigma'(x_0)}{\sigma(x_0)} + \xi \cdot \frac{b'(x_0)}{\sigma(x_0)}\}].
\end{aligned}$$

Q. E. D.

Lemma A.3 *Under Assumptions 3.1 and 3.2,*

$$Delta \;=\; E[\frac{d}{dx}E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i]]$$

Proof. Fix a neighborhood of x, say, $[x-\epsilon, x+\epsilon]$. By definition,

$$\text{Delta} \;=\; \frac{d}{dx}\int E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i = x_i]\cdot\prod_{j=1}^{i}\hat{g}(x_{j-1},x_j)dx_1\cdots dx_i$$

where $\hat{g}(x_{j-1},x_j)$ is the transition probability density given by

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

To interchange the order of differentiation and integration, we note that $E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i = x_i]$ grows polynomially with respect to x_i because of Assumption 3.1 and the exponential decay of the Gaussian transition densities. Using similar arguments as in Lemma A.2, we can apply Lemma A.1 here to obtain

$$\begin{aligned}
\text{Delta} &= \int E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i = x_i]\cdot\frac{d}{dx}g(x,x_1)\cdot\prod_{j=2}^{i}g(x_{j-1},x_j)dx_1\cdots dx_i \\
&= \int E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i = x_i]\cdot\frac{\frac{d}{dx}g(x,x_1)}{g(x,x_1)}\cdot\prod_{j=1}^{i}g(x_{j-1},x_j)dx_1\cdots dx_i \\
&= \int E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i = x_i]\cdot\Big\{\Big(\frac{(x_1 - x - \mu(x)\Delta t)^2}{\sigma^2(x)\Delta t} - 1\Big)\cdot\frac{\sigma'(x)}{\sigma(x)} + \frac{x_1 - x - \mu(x)\Delta t}{\sigma(x)\Delta t}\cdot \\
&\quad \frac{1+\mu'(x)\Delta t}{\sigma(x)}\Big\}\cdot\prod_{j=1}^{i}g(x_{j-1},x_j)dx_1\cdots dx_i \\
&= E[E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i]\cdot\Big\{\Big(\xi_1^2 - 1\Big)\cdot\frac{\sigma'(x)}{\sigma(x)} + \frac{\xi_1}{\sqrt{\Delta t}}\cdot\frac{1+\mu'(x)\Delta t}{\sigma(x)}\Big\}]
\end{aligned}$$

By Lemma A.2, if we view X_i as a function of X_1, we get

$$\begin{aligned}
&E[E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i]\cdot\Big\{\Big(\xi_1^2 - 1\Big)\cdot\frac{\sigma'(x)}{\sigma(x)} + \frac{\xi_1}{\sqrt{\Delta t}}\cdot\frac{1+\mu'(x)\Delta t}{\sigma(x)}\Big\}] \\
=\;& E[\frac{d}{dx}E[\Phi(\hat{X}_{i_1},\cdots,\hat{X}_{i_m})|\hat{X}_i]].
\end{aligned}$$

Q. E. D.

B One Lemma

Lemma B.1 *Theorem 3.2 holds for any Φ satisfying Assumption 3.1.*

Proof. It is easy to see that Φ satisying Assumption 3.1 must be an L^2 function because its growth rate is polynomial and the probability density function of X decays exponentially. Thus it can be approximated in L^2 by a sequence of functions Φ^n, each having compact support and continuous derivative,

$$\lim_{n\to\infty} E[\Phi^n(\hat{X}^\varepsilon_{i_1}, \cdots, \hat{X}^\varepsilon_{i_m})] = E[\Phi(\hat{X}^\varepsilon_{i_1}, \cdots, \hat{X}^\varepsilon_{i_m})]$$

Given a fixed neighborhood of $\varepsilon = 0$, say, K, for smooth payoff functions Φ^n, following the derivation in the proof of Theorem 3.2 for the case of $\varepsilon = 0$, we can also obtain an unbiased estimator for the case of an arbitrary $\varepsilon \in K$ of the form

$$\frac{d}{d\varepsilon} E[\Phi^n(\hat{X}^\varepsilon_{i_1}, \cdots, \hat{X}^\varepsilon_{i_m})] = E[\Phi^n(\hat{X}^\varepsilon_{i_1}, \cdots, \hat{X}^\varepsilon_{i_m}) \cdot D(\varepsilon)]$$

where D abbreviates the other terms in the estimator. By the Cauchy-Schwarz inequality,

$$|E[\Phi^n(\varepsilon)D(\varepsilon)] - E[\Phi(\varepsilon)D(\varepsilon)]| \le E[|\Phi^n(\varepsilon) - \Phi(\varepsilon)|^2]^{1/2} \cdot E[D^2(\varepsilon)]^{1/2},$$

where $\Phi^n(\varepsilon)$ and $\Phi(\varepsilon)$ abbreviate $\Phi^n(\hat{X}^\varepsilon_{i_1}, \cdots, \hat{X}^\varepsilon_{i_m})$ and $\Phi(\hat{X}^\varepsilon_{i_1}, \cdots, \hat{X}^\varepsilon_{i_m})$.

The right side of this inequality is continuous with respect to ε, so there exists an $\hat{\varepsilon}$ such that

$$\sup_{\varepsilon\in K} |E[\Phi^n(\varepsilon)D(\varepsilon)] - E[\Phi(\varepsilon)D(\varepsilon)]| \le E[|\Phi^n(\hat{\varepsilon}) - \Phi(\hat{\varepsilon})|^2]^{1/2} \cdot E[D^2(\hat{\varepsilon})]^{1/2}.$$

Thus, $E[\Phi^n(X^\varepsilon_{i_1}, \cdots, X^\varepsilon_{i_m}) \cdot D(\varepsilon)]$ converges to $E[\Phi^n(X^\varepsilon_{i_1}, \cdots, X^\varepsilon_{i_m}) \cdot D(\varepsilon)]$ uniformly on K because $E[|\Phi^n(\hat{\varepsilon}) - \Phi(\hat{\varepsilon})|^2] \to 0$ as $n \to +\infty$. We can interchange the order of differentiation and limit here to get that

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} E[\Phi(\hat{X}^\varepsilon_{i_1}, \cdots, \hat{X}^\varepsilon_{i_m})] = E[\Phi(\hat{X}_{i_1}, \cdots, \hat{X}_{i_m}) \cdot D(0)],$$

where $D(0)$ is the factor appearing that multiplies Φ in the estimator of Theorem 3.2. Q. E. D.

C Discrete Time Version of Girsanov Theorem

The following result is well-known, but we include the proof because it is short.

Lemma C.1 *Define a new probability measure $dP^\varepsilon = U^\varepsilon dP$. Then,*

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

Proof.

$$
\begin{aligned}
& P^\varepsilon[\sqrt{\Delta t}\xi_1 \in dw_1, \cdots, \sqrt{\Delta t}\xi_N \in dw_N] = U^\varepsilon P[\sqrt{\Delta t}\xi_1 \in dw_1, \cdots, \sqrt{\Delta t}\xi_N \in dw_N] \\
= \quad & \exp\Big[\varepsilon \sum_{i=1}^N \frac{\gamma(X_{i-1})}{\sigma(X_{i-1})} w_i - \frac{1}{2}\varepsilon^2 \sum_{i=1}^N [\frac{\gamma(X_{i-1})}{\sigma(X_{i-1})}]^2 \Delta t\Big] \cdot \frac{1}{(2\pi\Delta t)^{N/2}} \cdot \exp(-\frac{1}{2}\sum_{i=1}^N \frac{w_i^2}{\Delta t}) dw_1 \cdots dw_N \\
= \quad & \frac{1}{(2\pi\Delta t)^{N/2}} \cdot \exp\Big[-\frac{1}{2\Delta t}\sum_{i=1}^N (w_i - \varepsilon \frac{\gamma(X_{i-1})}{\sigma(X_{i-1})}\Delta t)^2\Big] dw_1 \cdots dw_N
\end{aligned}
$$

Q. E. D.

D Weak Convergence of Stochastic Differential Equations

This appendix presents a lemma, based on Kurtz and Protter [15], on the stability of stochastic differential equations that serves as the theoretical foundation of Section 4.

We work with a sequence of probability spaces $(\Omega^{(N)}, \mathcal{F}^{(N)}, \{\mathcal{F}_t^{(N)}\}, P^{(N)})$ and a probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$. For each N, an m-dimensional cadlag process $L^{(N)}$ is a semimartingale defined on the probability space $(\Omega^{(N)}, \mathcal{F}^{(N)}, \{\mathcal{F}_t^{(N)}\}, P^{(N)})$, which is adapted to filtration $\{\mathcal{F}_t^{(N)}\}$. Let $g^{(N)} : \mathbf{R}^k \times [0, +\infty) \to \mathbb{M}^{km}$ be a continuous function mapping a point in $\mathbf{R}^k \times [0, +\infty)$ to a $k \times m$ matrix in matrices space $\mathbb{M}^{km}$. We assume that a pair of k-dimensional processes $M^{(N)}$ and $U^{(N)}$ satisfy 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 $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, we have a triple (M, U, L) and a continuous function g such that

$$
M_t = U_t + \int_0^t g(M_{s-}, s-) dL_s. \tag{31}
$$

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

Definition D.1 (Elementary Processes) *An* elementary process *is one 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 $\mathcal{F}_{s_i}$-measurable and $|Y_i| \le 1$. The set of all elementary stochastic processes is denoted by $\mathcal{H}$.

Definition D.2 (P-UT Condition) *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 $t > 0$*

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

We now have the following lemma on the weak convergence of $M^{(N)}$:

Lemma D.3 *Suppose that $(U^{(N)}, L^{(N)})$ converge weakly to (U, L) in the Skohorod topology and that $L^{(N)}$ is P-UT. Assume that $g^{(N)}$ uniformly converges to g locally, i.e., for any compact set $K \subset \mathbf{R}^k \times [0, +\infty)$,*

$$\lim_{N\to+\infty} \sup_{(x,t)\in K} |g^{(N)}(x,t) - g(x,t)| = 0. \tag{32}$$

If there exists a global solution M of (31) and weak local uniqueness holds, then $(U^{(N)}, L^{(N)}, M^{(N)}) \Rightarrow (U, L, M)$ in the Skorohod topology.

Proof. The result follows from Theorem 5.4 in Kurtz and Protter [15] or Theorem 8.6 in Kurtz and Protter [16], provided we can verify condition C5.4 in [15] or condition 8.5 in [16].

In fact, suppose $T[0, +\infty)$ indicates the set of all nondecreasing mappings λ of $[0, +\infty)$ onto $[0, +\infty)$ such that $\lambda(t + h) - \lambda(t) \le h$ for all $t, h \ge 0$. Then, for any compact set $C \subset D_{\mathbf{R}^k}[0, +\infty) \times T[0, +\infty)$ and $t > 0$, we can find a compact set K in $\mathbf{R}^k \times [0, +\infty)$ such that $\{(y_{\lambda(s)}, \lambda(s)) : (y, \lambda) \in C, s \le t\} \subset K$ and

$$\sup_{(y,\lambda)\in C} \sup_{s\le t} |g^{(N)}(y_{\lambda(s)}, \lambda(s)) - g(y_{\lambda(s)}, \lambda(s))| \le \sup_{(x,t)\in K} |g^{(N)}(x,t) - g(x,t)|.$$

Thus, (32) implies C5.4 in [15] and condition 8.5 in [16]. Q. E. D.

E Uniform Bounds for the Moments of $\hat{X}^{(N)}$, $\hat{Y}^{(N)}$ and $(\hat{Y}^{(N)})^{-1}$

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 E.1 *For any j,*

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

Proof. By a Taylor expansion of function $1/(1+x)^4$ with Lagrange remainder (Abramowitz and Stegun [1], p.880), we have

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

for some $\theta \in (0, x)$. Furthermore, if we fix some positive number ϵ and consider the interval $[1-\sup_x |\sigma'(x)|w^* - \epsilon, 1+\sup_x |\sigma'(x)|w^* + \epsilon]$, we can get an upper bound for the function $1/(1+x)$ by letting m^* be the maximum value of $1/(1+x)^6$ over the interval:

$$\frac{1}{(1+x)^4} \leq 1 - 4x + 10m^* x^2.$$

Notice that for $i \neq j$,

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

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

$$\begin{aligned} &(1 + \mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)})^{-4} \leq \\ &\quad 1 - 4[\mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}] + 10m^*[\mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)}]^2. \end{aligned} \tag{33}$$

Consider any $1 \leq i \leq N$, $i \neq j$. Without loss of generality, we assume that $i > j$. By (33),

$$\begin{aligned} & E[(\hat{Y}_i^{(N)})^{-4} | \tilde{\xi}_j^{(N)} = w_R^{(N)}] \\ = \;& E[(\hat{Y}_{i-1}^{(N)})^{-4} \cdot E[(1 + \mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)})^{-4} | \hat{\mathcal{F}}_j] | \tilde{\xi}_j^{(N)} = w_R^{(N)}] \\ \leq \;& (1 + (4\sup_x |\mu'(x)| + 10m^* \sup_x |\sigma'(x)|^2)\Delta t^{(N)} + \sup_x |\mu'(x)|^2 (\Delta t^{(N)})^2) \cdot E[(Y_{i-1}^{(N)})^{-4} | \tilde{\xi}_j^{(N)} = w_R^{(N)}] \end{aligned}$$

Induction leads to

$$\sup_{1 \leq i \leq N} E[(\hat{Y}_i^{(N)})^{-4}] \leq \frac{(1 + (4\sup_x |\mu'(x)| + 10m^* \sup_x |\sigma'(x)|_2)\Delta t^{(N)} + \sup_x |\mu'(x)|^2 (\Delta t^{(N)})^2)^{N-1}}{1 - \sup_x |\sigma'(x)| w^* - \epsilon}.$$

because the jth increment is given by $1+\mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)}+\sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}} w_R^{(N)}$ which is bounded by $1/(1-\sup_x |\sigma'(x)|w^* - \epsilon)$. As $N \to +\infty$, the limit of the right hand side exists. Accordingly, $\sup_N \sup_{1 \leq i \leq N} E[(\hat{Y}_i^{(N)})^{-4}] < +\infty$. Q. E. D.

Lemma E.2 *Suppose that p is a positive integer. Then, there exist some constant C which does not depend on N such that*

$$\sup_{1 \leq i \leq N} E[(\hat{X}_i^{(N)})^{2p} | \tilde{\xi}_j^{(N)} = w_R^{(N)}] \leq C N^{p-1}.$$

Proof. Without loss of generality, we only consider the case $i > j$.

$$
\begin{aligned}
E[(\hat{X}_i^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] &= E[(\sum_{k=1}^{i}(\mu(\hat{X}_{k-1}^{(N)})\Delta t^{(N)} + \sigma(\hat{X}_{k-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_k^{(N)}))^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] \\
&\leq 2^{2p-1}E[(\sum_{k\neq j}(\mu(\hat{X}_{k-1}^{(N)})\Delta t^{(N)} + \sigma(\hat{X}_{k-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_k^{(N)})^{2p}|\tilde{\xi}_j^{(N)} = w_R^{(N)}] \\
&\quad +2^{2p-1}E[(\mu(\hat{X}_{j-1}^{(N)})\Delta t^{(N)} + \sigma(\hat{X}_{j-1}^{(N)})w^*)^{2p}] \\
&\leq 2^{2p-1}N^{2p-1}\sum_{k\neq j}E[(\mu(\hat{X}_{k-1}^{(N)})\Delta t^{(N)} + \sigma(\hat{X}_{k-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_k^{(N)})^{2p}] \\
&\quad +2^{2p-1}E[(\mu(\hat{X}_{j-1}^{(N)})\Delta t^{(N)} + \sigma(\hat{X}_{j-1}^{(N)})w^*)^{2p}]. \qquad (34)
\end{aligned}
$$

Note that $E[(\tilde{\xi}_k^{(N)})^{2p}] \leq E[(\xi^{(N)})^{2p}]$ where $\xi^{(N)}$ is a standard normal with mean 0 and variance 1, and recall that $\Delta t^{(N)} = T/N$. Then, there exists some constant C_1 and C_2 such that

$$
E[(\mu(\hat{X}_{k-1}^{(N)})\Delta t^{(N)} + \sigma(\hat{X}_{k-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_k^{(N)})^{2p}] \leq 2^{2p-1}C_1 \cdot ((\Delta t^{(N)})^{2p} + (\Delta t^{(N)})^p) \leq \frac{C_2}{N^p}. \quad (35)
$$

Together, (34) and (35) imply the statement of the lemma. Q. E. D.

Lemma E.3 *For any* j,

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

Proof. Without loss of generality, we only consider the case of $i > j$.

$$
\begin{aligned}
E[(\hat{Y}_i^{(N)})^4|\Delta\hat{W}_j^{(N)} = w_R^{(N)}] &= E[\prod_{k=1}^{i}(1 + \mu'(\hat{X}_{k-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{k-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_k^{(N)})^4|\tilde{\xi}_j^{(N)} = w_R^{(N)}] \\
&= E[\prod_{k=1}^{i-1}(1 + \mu'(\hat{X}_{k-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{k-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_k^{(N)})^4 \cdot \\
&\quad E[(1 + \mu'(\hat{X}_{i-1}^{(N)})\Delta t^{(N)} + \sigma'(\hat{X}_{i-1}^{(N)})\sqrt{\Delta t^{(N)}}\tilde{\xi}_i^{(N)})^4|\hat{\mathcal{F}}_i, \tilde{\xi}_j^{(N)} = w_R^{(N)}]]
\end{aligned}
$$

Because the functions μ' and σ' are bounded, there exists some constant C such that

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

By induction, we know that

$E[(\hat{Y}_i^{(N)})^4|\tilde{\xi}_j^{(N)} = w_R^{(N)}] \leq (1 + C\Delta t^{(N)} + o(\Delta t^{(N)}))^{i-1} \cdot (1 + \sup_x |\mu'(x)|\Delta t^{(N)} + \sup_x |\sigma'(x)|w^*)$.

Thus,

$\sup_{1\leq i\leq N} E[(\hat{Y}_i^{(N)})^4|\tilde{\xi}_j^{(N)} = w_R^{(N)}] \leq (1 + C\Delta t^{(N)} + o(\Delta t^{(N)}))^N \cdot (1 + \sup_x |\mu'(x)|\Delta t^{(N)} + \sup_x |\sigma'(x)|w^*)$.

As $N \to +\infty$, the right hand side of the inequality approaches a finite limit, and this proves the lemma. Q. E. D.

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

References

[1] Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* 9th Printing. Dover Publications, New York.

[2] Benhamou, E. (2003). Optimal Malliavin Weighting Function for the Computation of Greeks. *Mathematical Finance* **13**, pp.37–53.

[3] Bermin, H.-P., Kohatsu-Higa, A., and M. Montero. (2003). Local Vega Index and Variance Reduction Methods. *Mathematical Finance* **13**, 85–97.

[4] Cvitanic, J., J. Ma, and J. Zhang. (2003) Efficient Computation of Hedging Portfolios for Options with Discontinuous Payoffs. *Mathematical Finance* **13**, 135–151.

[5] Davis, M., and Johansson, M. (2006). Malliavin Monte Carlo Greeks for Jump Diffusions. *Stochastic Processes and their Applications* **116**, pp.101–129.

[6] Durrett, R. (1996). *Probability: Theory and Examples.* Second Edition. Duxbury Press.

[7] Fournié, E., J-M Lasry, J. Lebuchoux, P-L Lions, and N. Touzi. (1999). Applications of Malliavin Calculus to Monte Carlo Methods in Finance. *Finance and Stochastics*, Vol. **3**, pp. 391-412.

[8] Fournié, E., J-M Lasry, J. Lebuchoux, and P-L Lions. (2001). Applications of Malliavin Calculus to Monte Carlo Methods in Finance. II *Finance and Stochastics*, Vol. **5**, pp.201–236.

[9] Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering.* Springer-Verlag, New York.

[10] E. Gobet and A. Kohatsu-Higa. (2003). Computation of Greeks for Barrier and Lookback Options using Malliavin Calculus. *Electronic Communications in Probability*, **8**, 51–62.

[11] Jacod, J. and A. N. Shiryaev. (2003). *Limit Theorems for Stochastic Processes.* Second Edition. Springer-Verlag, Berlin, Heidelberg.

[12] Karatzas , I. and S. E. Shreve. (1991). *Brownian Motion and Stochastic Calculus* Springer-Verlag, New York, Berlin, Heidelberg.

[13] Kloeden , P. E. and E. Platen. (1992). *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin, Heidelberg.

[14] Kohatsu-Higa, A., and M. Montero. (2004). Malliavin Calculus in Finance. *Handbook of Computational and Numerical Methods in Finance*, Birkhauser, pp.111–174.

[15] Kurtz, T. G. and P. Protter. (1991). Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations. *Annals of Probability*, Vol. **19**, No. 3, pp. 1035-1070.

[16] Kurtz, T. G. and P. Protter. (1995). Weak Convergence of Stochastic Integrals and Differential Equations. *Lecture Notes in Mathematics*, Vol. **1627**, pp. 1-41.

[17] Nualart, D. (1995). *Malliavin Calculus and Related Topics*, Springer-Verlag.

[18] Protter, P. (1990). *Stochastic Integration and Differential Equations: A New Approach.* Springer-Verlag, Berlin, Heidelberg.