Asymptotic Properties of Monte Carlo Estimators of Derivatives

by

Jérôme Detemple, René Garcia and Marcel Rindisbacher.*

First draft: March, 2001.

This draft: February, 2005.

Abstract

We study the convergence of Monte Carlo estimators of derivatives when the transition density of the underlying state variables is unknown. Three types of estimators are compared. These are respectively based on Malliavin derivatives, on the covariation with the driving Wiener process, and on finite difference approximations of the derivative. We analyze two different estimators based on Malliavin derivatives. The first one, the Malliavin path estimator, extends the path derivative estimator of Broadie and Glasserman (1996) to general diffusion models. The second one, the Malliavin weight estimator, proposed by Fournié et. al. (1999), is based on an integration by parts argument and generalizes the likelihood ratio derivative estimator. It is shown that for discontinuous payoff functions only the estimators based on Malliavin derivatives attain the optimal convergence rate for Monte Carlo schemes. Estimators based on the covariation or on finite difference approximations are found to converge at slower rates. Their asymptotic distributions are shown to depend on additional second order biases even for smooth payoff functions.

Detemple is affiliated with Boston University School of Management and CIRANO, Garcia with the Economics Department, Université de Montréal, CIREQ and CIRANO and Rindisbacher with the Rotman School of Management, University of Toronto and CIRANO. Financial support from MITACS is gratefully acknowledged. The second author also thanks Hydro-Québec and the Bank of Canada for financial support. Address for correspondence: Jérôme Detemple, Boston University School of Management, 595 Commonwealth Avenue, Boston, MA, 02215.

1 Introduction

Hedging policies for contingent claims, optimal portfolios in asset allocation models, stock volatility calculations, as well as other problems in financial economics can be described in terms of the derivatives of a function $f(t,x)$, that solves the partial differential equation

$$\mathcal{L}_t f(t,x) + k(t,x) = 0; \quad \text{subject to} \quad f(T,x) = u(T,x), \tag{1}$$

where $\mathcal{L}_t f(t,x) = \partial_t f(t,x) + \partial_x f(t,x)A + \frac{1}{2} trace\left(\partial_{xx} f(t,x)BB'\right)$ is the infinitesimal generator of the diffusion

$$dX_v = A(X_v)dv + B(X_v)dW_v. \tag{2}$$

Explicit expressions for the derivatives of $f(t,x)$ are, in general, unknown and one must resort to numerical techniques for computation. Lattice methods, such as finite difference schemes, finite element schemes or finite Markov chain approximations of the diffusion, have been widely used for that purpose. The problem with these approaches is that their computational complexity (i.e. the number of arithmetic operations) grows exponentially as the number of state variables increases. In addition, these estimators for derivatives converge at a slower rate than the estimators for the function $f(t,x)$.

Malliavin calculus, i.e. the stochastic calculus of variations (see section 2 for an introduction to this calculus), can be used to overcome this curse of dimensionality and loss in speed of convergence.[1] With its help, derivatives of f can be written in the form

$$\partial_x f(t,x) = \mathbf{E}_{t,x}[g(X_T)] \tag{3}$$

where g is a function of the terminal value of the state variables. Naturally, this expression suggests a Monte Carlo (MC) estimator computed by averaging over independent replications $X_T^i, i = 1, ..., M$ of the terminal value X_T of (2). This estimate is attractive because its computational complexity grows only linearly with the dimensionality of the problem. Furthermore, as the estimator based on (3) is of the same form as the estimator for f, it converges at the same speed.

The optimal convergence rate for Monte Carlo simulation is $M^{-1/2}$.[2] This rate follows from the Central Limit Theorem applied to the sample average over independent replications $g(X_T^i)$ that

[1] Malliavin calculus is applied in Fournié et al. (1999), (2001) to compute the Greeks of option prices and in Detemple, Garcia and Rindisbacher (2003) to calculate dynamic asset allocation rules.

[2] This convergence rate is lower than the rate of lattice methods. However, when convergence rates are expressed in terms of work load, they are the same. For example, a finite difference approximation of a univariate partial differential equation (PDE) converges at the rate N^{-1} where N is the number of discretization points. As the number of arithmetic operations (work load W) is proportional to the number of discretization points squared ($W = const. \times N^2$) the convergence rate can be restated as $W^{-1/2}$. In contrast, a Monte Carlo estimator based on M replications converges, by the Central Limit theorem, at the rate $M^{-1/2}$. As the work is proportional to the number of replications ($W = const. \times M$) this convergence rate is also $W^{-1/2}$.

estimates (3). The optimal convergence rate is attained when (2) has an explicit transition density so that X_T^i can be drawn from the true distribution.

Unfortunately, in most cases the transition density is unknown, precluding sampling from the true distribution. One must then resort to a numerical scheme to approximate the terminal value of the diffusion. The typical scheme discretizes time into a finite number of intervals N, each of length $h \equiv (T-t)/N$, and approximates the stochastic differential equation (SDE) for X by its discretized version. It then draws innovations over each interval and constructs the corresponding values of the discretized SDE. Many discretization schemes have been proposed (see Kloeden and Platen (1997)). The simplest one, the Euler scheme, is a local linearization of the SDE given by

$$X_{t+(k+1)h}^{i,N} = X_{t+kh}^{i,N} + A\left(X_{t+kh}^{i,N}\right)h + B\left(X_{t+kh}^{i,N}\right)\left(W_{t+(k+1)h}^{i} - W_{t+kh}^{i}\right) \tag{4}$$

for $k = 0, \ldots, N-1$, subject to the initial condition $X_t^{i,N} = x$. The estimator of the derivative (3) is constructed by taking the average over a sample of M terminal values of (4). This gives

$$[\partial_x f]^{M,N} = \frac{1}{M}\sum_{i=1}^{M} g(X_T^{i,N}). \tag{5}$$

As the function g is obtained using Malliavin calculus, (5) is called a *Monte Carlo Malliavin Derivative* (MCMD) estimator. MCMD estimators for derivatives $\partial_x f$ are of the same form as MC estimators for the function f. As we show in this paper, they are the only Monte Carlo estimators for derivatives that attain the optimal convergence rate for MC schemes, $M^{-1/2}$.

The question of estimating the expected value of the derivative of a diffusion process has already been studied in the literature. For settings where the transition probability of the diffusion is known, Broadie and Glasserman (1996) have proposed two classes of estimators, the path derivative and the likelihood ratio, or score, estimators. Results from Malliavin calculus enable us to extend these notions to general diffusion settings. A path derivative estimator, in Broadie and Glasserman (1996), is based on the partial derivative of the deterministic function which maps the initial value into the terminal value of the diffusion, with respect to the initial value. For general diffusion processes such a function does not exist. The functional which associates the terminal value to the initial value, i.e. the stochastic flow, depends on the initial value through the entire trajectory of the state variables. A rigorous definition of a "path" derivative can nevertheless be provided in terms of the tangent or first variation process of the stochastic flow (see Section 2.3 for definitions and details). As the Malliavin derivative is proportional to the tangent process, it can be used to generalize the concept of a path derivative estimator to general diffusions. We call the corresponding estimator an *MCMD-path* estimator (MCMD-P). On the other hand, the likelihood ratio estimator, in Broadie and Glasserman (1996), is obtained using an integration-by-parts argument for the Riemann integral involving the known transition density. As we show in Appendix C, Malliavin calculus provides an abstract integration-by-parts formula on Wiener space. We show how this permits a generalization of the likelihood ratio estimator to situations where the transition density is unknown. An MCMD estimator obtained by this integration-by-parts formula, is called an *MCMD-weight* estimator (MCMD-W). It is important to underline that, in contrast to MCMD estimators, the implementation of path derivative and likelihood ratio estimators requires explicit knowledge of the transition density of the diffusion. The numerical implementation of MCMD estimators requires the application of a discretization scheme to simulate the state variables, their Malliavin derivatives or the Malliavin weights. It should also be noted that another extension of the likelihood ratio method, to situations where the transition density is unknown, can be achieved by discretizing the model first and then applying the likelihood ratio approach. The resulting estimator is based on a convergent approximation of the true score by the score of the discretized model. This approximation constitutes an additional source of error that may affect the asymptotic properties of the estimator.

Cvitanic, Goukasian and Zapatero (2002, 2003) have proposed an alternative approach, based on an approximation of the covariation between the function $f(X_T)$ and the Brownian motion W. This *Monte Carlo Covariation* estimator (MCC), can be seen as a convergent approximation of the MCMD-W estimator (see Section 3.2. for details).

Finally, by perturbing the initial value of the diffusion and forming finite differences, one can construct *Monte Carlo finite difference* (MCFD) schemes that estimate the derivative $\partial_x f$. Forward, backward and central differencing schemes are available. The corresponding estimators are the *Monte Carlo forward finite difference* (MCFFD), the *Monte Carlo backward finite difference* (MCBFD) and the *Monte Carlo central finite difference* (MCCFD) estimators. These schemes were introduced in a non-diffusion setting by Glynn (1989) and their asymptotic properties studied by L'Ecuyer and Perron (1994). Both papers assume that transition densities are known. In our more general setting, the transition density is unknown and a numerical discretization procedure must be used to approximate the diffusion. As discussed in Section 3.3, MCFD estimators are approximate MCMD-P estimators.

When a discretization scheme is combined with Monte Carlo averaging to estimate an expectation involving the solution of an SDE two errors have to be dealt with. The first one is the discretization error due to the finite number of discretization points, N, in the approximation (4) of (2). The second one is the Monte Carlo error associated with the computation of an expectation by averaging over a finite number of replications, M, of the relevant random variables. Both errors affect the estimator constructed and influence its asymptotic properties. This paper studies the error behavior for MCMD, MCC and the three finite difference schemes MCFD described above.

For discontinuous payoff functions we show that the convergence rates of MCC, MCFFD and MCBFD estimators equal $M^{-1/3}$ and that MCCFD estimators converge at the rate $M^{-2/5}$. An

MCMD estimator, on the other hand, converges at the rate $M^{-1/2}$. It is therefore the only estimator, among those studied, that preserves the convergence rate attained by the MC estimator of the function itself. It is also the only MC estimator for derivatives that attains the maximal convergence rate, $M^{-1/2}$, for MC schemes. The reason is that all other estimators are numerical approximations of convergent approximations, instead of explicit representations, of the derivative. Because MCFD and MCC estimators are based on convergent approximations their implementation depends on additional perturbation parameters. These correspond to a spatial shift of the initial condition for MCFD and a time shift of the initial increment of the Wiener process for MCC. The additional parameter implies a slower convergence rate for discontinuous payoffs. In contrast, for continuous payoff functions the rate $M^{-1/2}$ is also attained by MCFD estimators: the variance of these estimators is of second order in the convergence parameters, speeding up convergence.

The asymptotic efficiency of each estimator depends on the length of the asymptotic confidence intervals implied by the limit distribution. For each method we show that the weak limit along the efficient path of M, N and any other parameter controlling the approximation, has a non-zero mean representing a *second order bias*. Ignoring this bias leads to incorrect efficiency assessments. Indeed, in this case, asymptotic confidence intervals cover the true value with probability lower than the nominal size of the interval. Corresponding confidence sets are then invalid.[3] In fact, the Neyman-Pearson theory of statistical tests (see Lehmann (1997)) establishes that asymptotic efficiency assessments cannot be made when confidence intervals are not of the same effective size. Our explicit formulas permit the construction of valid confidence intervals. They also show that second order bias correction is considerably more difficult for MCC and MCFD estimators: MCC and MCFD involve additional second order biases because they are approximations of MCMD-W and MCMD-P estimators. Finally, they provide the necessary tools for assessing the relative efficiencies of different Monte Carlo estimators of derivatives (see Duffie and Glynn (1995) for a discussion of efficient simulation designs for Monte Carlo estimators of conditional expectations).

Section 2 provides a brief introduction to Malliavin calculus. Section 3 studies the convergence of the various MC estimators of derivatives: MCMD, MCC and MCFD. The computational requirements differ across estimators. In general MCMD and MCCFD estimators, which converge at a faster rate, require the simulation of auxiliary processes such as Malliavin derivatives or processes with perturbed initial values. It follows that there is a trade-off between speed of convergence and execution time. Section 4 studies this trade-off by performing a numerical comparison of the three methods in the context of two examples. The first example is a linear dynamic portfolio choice problem; the second one addresses the issue of risk management for digital options when the underlying price follows a process with constant elasticity of variance. These numerical studies

[3] A discussion of this issue can be found in Detemple, Garcia and Rindisbacher (2004). They show that the coverage probability is positively related to the asymptotic variance when the second order bias is ignored.

illustrate that the faster convergence rate of MCMD estimators results in an efficiency gain in terms of CPU time and percentage root mean square error. Section 5 concludes the paper. Appendix A contains a result from Detemple, Garcia and Rindisbacher (2004) that is used to characterize the second order bias for the different methods under consideration. Proofs are collected in appendix B. Appendix C presents an integration by parts argument that can be employed to find MCMD estimators when transition probabilities are unknown and payoff functions are non-smooth.

2 A Primer on Malliavin Calculus

The Malliavin calculus is a calculus of variations for stochastic processes defined on a Wiener space (the smallest state space on which a Brownian motion can be defined). This calculus applies to Wiener functionals, i.e. random variables and stochastic processes that depend on the trajectories of a Brownian motion. It enables us to measure the effects of a small variation in the trajectory of the underlying Brownian motion on this functional. This section presents elementary operations associated with the Malliavin calculus. A detailed treatment can be found in Nualart (1995).

2.1 The Malliavin Derivative of a Smooth Brownian Functional

Let $(t_1, ..., t_n)$ be a partition of the interval $[0, T]$ and let F be a random variable of the form

$$F \equiv f\left(W_{t_1}, ..., W_{t_n}\right),$$

where f is an infinite times continuously differentiable, bounded function and W is a univariate Brownian motion process. The random variable F depends (smoothly) on the Brownian motion W sampled at a finite number of points in the interval $[0, T]$. It is called a *smooth Brownian functional.*

The Malliavin derivative of F is the change in F due to a change in the path of W. To formalize this notion consider a time t such that $t_1 < ... < t_{k-1} < t \leq t_k ... < t_n$ and suppose that W_s is perturbed to $W_s + \varepsilon$ for all $s \geq t$. The Malliavin derivative of F at time t, which is written as $\mathcal{D}_t F$, is defined as

$$\mathcal{D}_t F \equiv \left.\frac{\partial f\left(W_{t_1}, ..., W_{t_{k-1}}, W_{t_k} + \varepsilon, ..., W_{t_n} + \varepsilon\right)}{\partial \varepsilon}\right|_{\epsilon=0} = \sum_{i=k}^{n} f_i\left(W_{t_1}, ..., W_{t_k}, ..., W_{t_n}\right), \tag{6}$$

where f_i is the derivative with respect to the i^{th} argument of f.

A simple example will illustrate the concept. Suppose that the price of a stock follows a geometric Brownian motion with drift μ and volatility coefficient σ. At time T the stock price is $S_T = f(W_T)$ where the function $f(x)$ is $f(x) \equiv S_0 \exp\left((\mu - \frac{1}{2}\sigma^2)T + \sigma x\right)$. An application of the definition shows that the Malliavin derivative of S_T is

$$\mathcal{D}_t S_T = \left.\frac{\partial f(W_T + \varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} = \sigma S_0 \exp\left((\mu - \frac{1}{2}\sigma^2)T + \sigma W_T\right) = \sigma S_T.$$

In this example S_T depends only on the Brownian motion at T. The Malliavin derivative is then the same as the derivative with respect to W_T. This reflects the fact that a perturbation of the path of the Brownian motion from time t onward affects S_T only through the terminal value W_T.

2.2 The Malliavin Derivative of an Ito Integral and the Chain Rule

The notion of a Malliavin derivative can be extended to random variables that depend on the path of the Brownian motion over a continuous interval $[0,T]$. This extension uses the fact that a path-dependent functional can be approximated by a suitable sequence of smooth Brownian functionals. In short, the Malliavin derivative of the path-dependent functional is the limit of the Malliavin derivatives of the smooth Brownian functionals in the approximating sequence. The space of random variables for which Malliavin derivatives are defined is called $\mathbb{D}^{1,2}$. This space is the completion of the set of smooth Brownian functionals with respect to the norm $\|F\|_{1,2} \equiv \left(\mathbf{E}\left[F^2\right] + \mathbf{E}\left[\int_0^T \|\mathcal{D}_t F\|^2 dt\right]\right)^{1/2}$.

This extension enables us to define Malliavin derivatives of stochastic integrals in a natural manner. For instance, consider the stochastic integral $F(W) = \int_0^T h(t)dW_t$, where $h(t)$ is a function of time. By Ito's lemma we can write $F(W) = h(T)W_T - \int_0^T W_t dh(t)$. Taking a perturbation of W_s to $W_s + \varepsilon$ for all $s \in [t, \infty)$ shows that

$$F(W + \varepsilon 1_{[t,\infty)}) - F(W) = h(T)\varepsilon 1_{[t,\infty)}(T) - \int_0^T \varepsilon 1_{[t,\infty)}(s)dh(s)$$

where $1_{[t,\infty)}(s)$ is the indicator function (i.e. $1_{[t,\infty)}(s) = 1$ if $s \in [t,\infty)$; $= 0$ otherwise) and therefore

$$\mathcal{D}_t F = \lim_{\varepsilon \to 0} \frac{F(W + \varepsilon 1_{[t,\infty)}) - F(W)}{\varepsilon} = h(T)1_{[t,\infty)}(T) - \int_0^T 1_{[t,\infty)}(s)dh(s) = h(t).$$

That is, the Malliavin derivative of F at date t is the volatility $h(t)$ of the stochastic integral at t. In other words, the Malliavin derivative $\mathcal{D}_t F$ measures the sensitivity of the random variable F to the Brownian innovation at t.

Assume now that the integrand $h(t) = h(W,t)$ is a progressively measurable process. In this case, with the notation $W_s^\varepsilon = W_s + \varepsilon 1_{[t,\infty)}(s)$, we can write

$$\begin{aligned}
F(W + \varepsilon 1_{[t,\infty)}) - F(W) &= \int_0^T h(W^\varepsilon, s)dW_s^\varepsilon - \int_0^T h(W,s)dW_s \\
&= \int_0^T \left(h(W^\varepsilon, s) - h(W,s)\right) dW_s + \int_0^T h(W^\varepsilon, s)d\left(W_s^\varepsilon - W_s\right) \\
&= \int_0^T \left(h(W^\varepsilon, s) - h(W,s)\right) dW_s + \varepsilon \int_0^T h(W^\varepsilon, s)d1_{[t,\infty)}(s) \\
&= \int_0^T \left(h(W^\varepsilon, s) - h(W,s)\right) dW_s + \varepsilon h(W^\varepsilon, t)
\end{aligned}$$

where the last equality follows from the fact that $d1_{[t,\infty)}(s) = \delta_0(s-t)$ is the Dirac delta function at the point $s - t = 0$.[4] It follows that

$$\begin{aligned}
\mathcal{D}_t F &= \lim_{\varepsilon \to 0} \frac{F(W + \varepsilon 1_{[t,\infty)}) - F(W)}{\varepsilon} = \int_0^T \mathcal{D}_t h(W,s) dW_s + h(W,t) \\
&= \int_t^T \mathcal{D}_t h(W,s) dW_s + h(W,t)
\end{aligned}$$

where, by definition, $\mathcal{D}_t h(W,s) = \lim_{\varepsilon \to 0} \left(h(W^\varepsilon, s) - h(W,s) \right) / \varepsilon$. Similarly for a (random) Riemann integral $F(W) = \int_0^T h(W,s)ds$ we obtain $\mathcal{D}_t F = \int_0^T \mathcal{D}_t h(W,s)ds = \int_t^T \mathcal{D}_t h(W,s)ds$.

Typical applications in finance involve functions that depend on several path-dependent random variables. These cases can be handled using the chain rule of Malliavin calculus. Suppose that $F = (F_1, ..., F_n)$ is a vector of random variables in $\mathbb{D}^{1,2}$, and that ϕ is a differentiable function of F with bounded derivatives. Then,

$$\mathcal{D}_t \phi(F) = \sum_{i=1}^n \frac{\partial \phi}{\partial x_i}(F) \mathcal{D}_t F_i.$$

A standard application of this rule arises when a stock return satisfies an Ito process with progressively measurable coefficients (μ, σ), i.e. the price satisfies $dS_t/S_t = \mu_t dt + \sigma_t dW_t$. The terminal value is

$$S_T = S_0 \exp\left(\int_0^T \left(\mu_s - \frac{1}{2}\sigma_s^2 \right) ds + \int_0^T \sigma_s dW_s \right)$$

and the chain rule of Malliavin calculus gives

$$\mathcal{D}_t S_T = S_T \left(\int_t^T \left(\mathcal{D}_t \mu_s - \sigma_s \mathcal{D}_t \sigma_s \right) ds + \int_t^T \mathcal{D}_t \sigma_s dW_s + \sigma_t \right)$$

where we used $\mathcal{D}_t \sigma_s^2 = 2\sigma_s \mathcal{D}_t \sigma_s$ and the formulas above for the derivatives of an Ito integral and a random Riemann integral.

2.3 Malliavin Derivatives of Solutions of Stochastic Differential Equations

The rules of Malliavin calculus described above enable use to calculate the derivative of the solution of a stochastic differential equation. Suppose, for instance, that a state variable X solves the equation $dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$, subject to some initial condition $X_0 = x$, where W is a standard Brownian motion. Equivalently, we can write the equation in integral form as

$$X_t = x + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dW_s.$$

[4]The Dirac delta function $\delta_0(s-t)$ is such that $\delta_0(s-t) = 0$ if $s \neq t$ and $\int_0^T f(s)\delta_0(s-t)ds = f(t)$.

Applying the results from the previous sections it is easy to verify that $\mathcal{D}_t X_s$ satisfies

$$\mathcal{D}_t X_s = \int_t^s \partial\mu(X_v)\mathcal{D}_t X_v dv + \int_t^s \partial\sigma(X_v)\mathcal{D}_t X_v dW_v + \sigma(X_t)$$

(recall that $\mathcal{D}_t x = 0$). This linear equation can be solved to yield

$$\mathcal{D}_t X_s = \sigma(X_t)\mathcal{E}\left(\int_t^T \partial\mu(X_s)ds + \partial\sigma(X_s)dW_s\right)$$

for all $s \geq t$, where $\mathcal{E}$ denotes the stochastic exponential.[5]

Note that the Malliavin derivative of a stochastic differential equation is related to the derivative of the solution of the SDE with respect to the initial condition. The Brownian functional $X_s(x)$ which gives the position at time s of the state variable as a stochastic mapping of its initial position $X_t = x$, is called the *stochastic flow*. Ikeda and Watanabe (1981) and Kunita (1991) show that the stochastic flow solves the same SDE as the diffusion itself. More importantly they show that its gradient $\nabla_{t,x} X_s(x) \equiv \lim_{\tau\to 0} \frac{1}{\tau}\left(X_s(x+\tau) - X_s(x)\right)$ (i.e. the change in the flow induced by an infinitesimal perturbations of the initial state) solves an SDE with linear coefficients and initial condition given by the identity matrix. The process $\nabla_{t,x} X_s(x)$ is called the *tangent* process.

As the Malliavin derivative solves the same SDE but with a different initial condition, it follows immediately that $\mathcal{D}_t X_s = \sigma(X_t)\nabla_{t,x} X_s(x)$: the Malliavin derivative is equal to the tangent $\nabla_{t,x} X_s(x)$ multiplied by the volatility coefficient of the process.

2.4 The Clark-Ocone Formula

Our next result is a very useful formula known as the Clark-Ocone formula. This formula ties into the Martingale Representation Theorem which states that a martingale adapted to a Brownian filtration can be written as a stochastic integral with respect to Brownian motion. Malliavin calculus, in effect, gives an explicit expression for the integrand in this representation. In other words, it identifies the volatility coefficient of the martingale.

The Clark-Ocone formula states that any random variable $F \in \mathbb{D}^{1,2}$ can be decomposed as

$$F = \mathbf{E}[F] + \int_0^T \mathbf{E}_s\left[\mathcal{D}_s F\right] dW_s,$$

where $\mathbf{E}_s[\cdot]$ is the conditional expectation at time s. For the special case of a martingale $M_t = \mathbf{E}_t[F]$, closed by the random variable F, one obtains $M_t = M_0 + \int_0^t \mathbf{E}_s\left[\mathcal{D}_s F\right] dW_s$.

[5] Ordinary exponentials solve linear ODEs. Likewise, stochastic exponentials solve linear SDEs. More precisely, $\mathcal{E}(M)$ solves $d\mathcal{E}(M)_t = \mathcal{E}(M)_t dM_t$ subject to $\mathcal{E}(M)_0 = 1$. For univariate continuous local martingales M we have $\mathcal{E}(M)_T = \exp\left(M_T - \frac{1}{2}[M]_T\right)$, where $[M]$ is the quadratic variation of M (i.e. $[M]$ is the unique, predictable process such that $[M]_0 = 0$ and $M^2 - [M]$ is a continuous local martingale (Karatzas and Shreve (1991), Problem 5.17)).

2.5 Multivariate Wiener Processes

To conclude this presentation of Malliavin calculus we briefly indicate generalizations of definitions and results to the case of multivariate Brownian motions. Suppose that $W = (W_1, ..., W_d)'$ is a d-dimensional Brownian motion and that $F \equiv f(W_{t_1}, ..., W_{t_n})$ is a smooth Brownian functional, that depends on the vector Brownian motion W sampled at a finite number of time points.

In this multivariate case one can define a Malliavin derivative with respect to each component of the multidimensional Brownian motion. The Malliavin derivative is a d-dimensional row vector $\mathcal{D}_t F = (\mathcal{D}_{1t}F, ..., \mathcal{D}_{dt}F)$ where, for $j = 1, ..., d$, the derivative $\mathcal{D}_{jt}F$ is obtained by perturbing the j^{th} component of the Brownian motion. In other word $\mathcal{D}_{jt}F$ satisfies definition (6) where W_{js} is perturbed to $W_{js} + \varepsilon$ for $s \geq t$ and other components of the Brownian motion are left unchanged.

All the rules of Malliavin calculus described for a univariate Brownian motion also apply to the multidimensional case.

3 Simulation-based Estimators of Derivatives

Suppose that we wish to estimate the derivative of a function $f(t, x) = \mathbf{E}_{t,x}[h(X_T)]$ where h is continuously differentiable and X is a n-dimensional vector of state variables with dynamics

$$dX_t = A(X_t)dt + \sum_{j=1}^{d} B_j(X_t)dW_t^j; \quad X_0 = x.$$

The coefficients A and B are assumed to satisfy local Lipschitz and linear growth conditions (see Karatzas and Shreve (1991), Theorem 2.5, p. 287). This section presents three simulation methods that can be used to compute the derivative of interest.

3.1 Monte Carlo with Malliavin Derivative (MCMD)

Our first approach is based on Malliavin calculus. Suppose that we take the Malliavin derivative on both sides of the expression $f(t, x) = \mathbf{E}_{t,x}[h(X_T)]$ where $x = X_t$. The chain rule of Malliavin calculus and the commutativity of the Malliavin derivative and the conditional expectation operators (see Nualart (1995) proposition 1.2.4 page 32) gives

$$\partial_x f(t, X_t)\mathcal{D}_t X_t = \mathcal{D}_t f(t, X_t) = \mathcal{D}_t \mathbf{E}_{t,x}[h(X_T)] = \mathbf{E}_{t,x}[\mathcal{D}_t h(X_T)] = \mathbf{E}_{t,x}[\partial h(X_T)\mathcal{D}_t X_T], \qquad (7)$$

where $\mathcal{D}_t X_T$ is the terminal value of the solution of the linear stochastic differential equation

$$d\mathcal{D}_t X_v = \left(\partial A(X_v)dv + \sum_{j=1}^{d} \partial B_j(X_v)dW_v^j \right) \mathcal{D}_t X_v \quad \text{and} \quad \lim_{v \downarrow t} \mathcal{D}_t X_v = B(X_t). \qquad (8)$$

In general, the random variable $\mathcal{D}_t X_T$ will depend on the whole path of the underlying Brownian motion, making it difficult to apply convergence results tailored for expectations of functions of the terminal value of a diffusion. Two avenues can be pursued to bring us back to this more conventional form. Both of these identify a function g and a diffusion Y such that

$$\partial_x f(t,x) = \mathbf{E}_{t,x}[g(Y_T)]. \tag{9}$$

A Monte Carlo estimator can be readily constructed from this expression, by replacing the conditional probability measure with its empirical counterpart that places equal probability on independent replications of the terminal value of the Euler discretized diffusion Y_T^N,

$$[\partial_x f]^{M,N}(t,x) = \frac{1}{M}\sum_{i=1}^{M} g(Y_T^{i,N}). \tag{10}$$

This estimator of the derivative $\partial_x f(t,x)$ is of the same form as the Monte Carlo estimator of the function $f(t,x)$ and will share its convergence properties.

3.1.1 Malliavin Derivative Path Estimators (MCMD-P)

A formula such as (9) can be found either by conditioning on the terminal value of the state variables or by introducing an augmented diffusion that captures the joint behavior of X and its Malliavin derivative $\mathcal{D}_t X$. The first transformation works as follows. Using the law of iterated expectations in the last conditional expectation of (7) gives

$$\partial_x f(t,x)B(x) = \mathbf{E}_{t,x}[\partial h(X_T)\mathbf{E}_{t,x}[\mathcal{D}_t X_T | X_T]]. \tag{11}$$

The desired formula (9) follows if we select $g(Y_T) \equiv g(t,x,X_T)$ with the definition

$$g(t,x,z) \equiv \partial h(z)k(t,x,z)C(x) \tag{12}$$

where

$$k(t,x,z) \equiv \mathbf{E}_{t,x}[\mathcal{D}_t X_T | X_T = z] \text{ and } C(x) \equiv B(x)'(B(x)B(x)')^{-1}. \tag{13}$$

Clearly, a Monte Carlo estimator based on this transformation, is easy to implement only if the conditional expectation $k(t,x,z)$ has an explicit form. This is a strong condition and cases where this approach works are rare.

The second transformation applies even when an explicit form for k is unknown. It relies on an expansion of the state space that adds the $n \times d$-vector of Malliavin derivative $\mathcal{D}_t X_v$ to the set of state variables. The augmented system of state variables $Y' = [Y_1', Y_2']$, with $Y_1 \equiv X$ and $Y_2' \equiv [\mathcal{D}_{1t}X', ..., \mathcal{D}_{dt}X']$, satisfies the stochastic differential equation

$$dY_v = \begin{bmatrix} A \\ \partial A Y_2 \end{bmatrix}(Y_v)dv + \sum_{j=1}^{d} \begin{bmatrix} B_j \\ \partial B_j Y_2 \end{bmatrix}(Y_v)dW_v^j \tag{14}$$

subject to the initial condition $Y_t' = [X_t', B_1(Y_{1t})', ..., B_d(Y_{1t})']$. The derivative of the function can be written as in (9) with the definition

$$g(y) \equiv (\partial h(y_1))y_2 C(y_1). \tag{15}$$

A setting of practical interest for which the function k in (12) can be found in explicit form is the Black and Scholes model. For $dX_t = X_t\,(Adt + BdW_t)$ we obtain $\mathcal{D}_t X_T = BX_T$ and therefore $k(t,x,X_T) = BX_T$ with $g(t,x,X_T) = \partial h(X_T)(X_T/x)$. Here, the estimator based on (12) is identical to the path derivative estimator of Broadie and Glasserman (1996).[6]

As mentioned in section 2.3, Malliavin derivatives are proportional to tangent processes. They can therefore be used to formalize the notion of "path derivative" for arbitrary diffusion models. More precisely, note that the Malliavin derivative satisfies $\mathcal{D}_t X_v C(X_t) = \nabla_{t,x} X_v(X_t)$ where $\nabla_{t,x} X_v(X_t)$ is the derivative of the stochastic flow $X_v(X_t)$ with respect to its initial position at time t (see Section 2.3 for the univariate case). Given this relation and the natural interpretation of the derivative of the stochastic flow as a path derivative we call estimators based on (12) or (15) Monte Carlo Malliavin derivative path estimator (MCMD-P).

3.1.2 Malliavin Derivative Weight Estimators (MCMD-W)

The estimators presented so far require that h be at least continuously differentiable. Unfortunately, for some applications in finance, such as managing the risks associated with a position in digital options, this differentiability assumption is not satisfied. Moreover, expressions for derivatives are difficult to obtain for several types of financial contracts, such as mortgage-backed securities.

In these instances one can proceed in two ways. First, assume that h fails to be continuously differentiable but suppose that the transition density $p(t,x,s,y)$ of the diffusion exists and is known. Under these conditions the conditional expectation

$$f(t, X_t) = \int_{\mathbb{R}} h(z)p(t, X_t, T, z)dz.$$

[6] Broadie and Glasserman (BG) use this example to introduce the idea of path derivatives. As $X_T = X_t F$ where the random variable $F = \exp\left((A - \frac{1}{2}B^2)(T-t) + B(W_T - W_t)\right)$ is independent of X_t, the path derivative can be defined as an ordinary derivative of the linear function $X_t F$ with respect to X_t. The same property holds in the example with stochastic volatility that they provide. For cases where the relevant SDEs do not have explicit solutions BG define path derivatives for the Euler discretized processes (see Proposition 7 page 284). For this approximation scheme, a deterministic map links the initial and terminal values of the discretized processes for each sub-interval.

can be Malliavin-differentiated on both sides to obtain

$$\partial_x f(t, X_t)\mathcal{D}_t X_t = \int_{\mathbb{R}} h(z)\partial_x p(t, X_t, T, z)dz\mathcal{D}_t X_t = \mathbf{E}_{t,X_t}\left[h(X_T)\partial_x \log p(t, X_t, T, X_T)\right]\mathcal{D}_t X_t.$$

The resulting probabilistic representation of the derivative

$$\partial_x f(t, X_t) = \mathbf{E}_{t,X_t}\left[h(X_T)\partial_x \log p(t, X_t, T, X_T)\right]$$

is in the form (9) with the choice

$$g(t, x, T, z) \equiv h(z)\partial_x \log p(t, x, T, z). \tag{16}$$

Estimators based on this formula correspond to the likelihood ratio or score function estimators proposed by Broadie and Glasserman (1996) in the context of security pricing with known transition density.[7] As these estimates are again of the form (9) their convergence properties are identical to those of MCMD-P estimators (12), (15), based on Malliavin derivatives. However, applications of this method are limited because knowledge of the score function $\partial_x \log p(t, x, T, z)$ is required.

Next we derive an estimator that is of the same form for unknown transition densities. For this note that the probabilistic representation of the derivative is given as the expected value of the payoff function multiplied by a "weight"

$$H_{t,T}^{score} \equiv \partial_x \log p(t, x, T, X_T), \tag{17}$$

equal to the score. For geometric Brownian motion, Broadie and Glasserman (1996) obtain the "score-weight" by an integration-by-parts argument. Appendix C shows how to generalize this idea. An integration-by-parts argument for Malliavin calculus establishes,[8]

$$\partial_x f(t, x) = \mathbf{E}_{t,x}\left[\partial h(X_T)\nabla_{t,x} X_T(x)\right] = \mathbf{E}_{t,x}\left[H_{t,T}^{(\alpha)} h(X_T)\right] \tag{18}$$

where

$$H_{t,T}^{(\alpha)} \equiv \int_t^T (dW_s)'\nu_{t,s}^{(\alpha)} \tag{19}$$

$$\nu_{t,s}^{(\alpha)} \equiv C(X_s)\nabla_{t,x} X_s(x)\alpha_{t,s}, \tag{20}$$

for some progressively measurable process α such that $\int_t^T \alpha_{t,s} ds = I_n$, the identity matrix of dimension n, and C defined in (12). The right hand side of (18) is again of the form (9) with

$$g(y) \equiv y_2 h(y_1) \tag{21}$$

[7] See Rubinstein and Shapiro (1993) for a general discussion of this method.

[8] This result is the same as in Fournié et al. (1999). Our derivation of the formula, in Appendix C, relies on different arguments and does not use stochastic calculus for non-anticipative processes.

and expanded vector $Y_v' \equiv [X_v', H_{t,v}^{(\alpha)\prime}]$, for $v \in [t,T]$. Here, the desired formula was found by adding the Malliavin weight $H_{t,T}^{(\alpha)}$ to the set of state variables. Note also that the function g depends on the payoff function h, rather than its derivative. This feature is important for implementation. It implies that a Monte Carlo method, based on the expanded vector of state variables Y, can be combined with a discretization scheme to estimate the derivative $\partial_x f(t,x)$ even though the payoff function is non-differentiable and the transition density is unknown. As the payoff is multiplied by the weight $H_{t,T}^{(\alpha)}$, in (18), we call the resulting estimator a Monte Carlo Malliavin derivative weight estimator (MCMD-W).[9] The fact that the likelihood ratio/score estimator has a similar structure, suggests that MCMD-W is a generalization. We now show why this is indeed the case.

Fournié et al. (2001) establish that the Malliavin weight corresponding to the score provides the estimator with minimal mean-square error. As the score is a function of the terminal value of the diffusion the optimal Malliavin weight must be as well. Malliavin weights that are deterministic functions of the terminal value of the diffusion can be obtained from (21), using the law of iterated expectations: it suffices to condition on the terminal value within the conditional expectation to obtain a "weight" of this form. By definition this conditional expectation attains the minimal mean square error among all random variables that are functions of the terminal value. As the weight that attains the minimal mean square error is unique, this "weight" must be identical to the score. That is,

$$\mathbf{E}_{t,x}\left[H_{t,T}^{(\alpha)} \middle| X_T \right] = \partial_x \log p(t,x,T,X_T) \equiv H_{t,T}^{score}, \tag{22}$$

for any admissible choice of α.[10] It follows that MCMD-W estimators do indeed generalize the score or likelihood ratio method.[11]

Following the arguments in the proof in Appendix C, it is interesting to note that the Malliavin weight is in fact nothing else but an abstract score, i.e. a derivative of a log-likelihood-ratio,

$$H_{t,T}^{(\alpha)} = \left(\partial_\lambda \log \frac{d\mathbf{P}_{t,x}^{\lambda}}{d\mathbf{P}_{t,x}} \right)_{|\lambda=0} \tag{23}$$

for a particular density given by,

$$\frac{d\mathbf{P}_{t,x}^{\lambda}}{d\mathbf{P}_{t,x}} = \mathcal{E}\left(-\lambda \int_t^{\cdot} (\nu_{t,s}^{(\alpha)})' dW_s \right)_T, \tag{24}$$

[9]The implementation of MCMD-W requires a discretization scheme to calculate the Malliavin weight.

[10]This result can also be established using a probabilistic representation of the transition density, that is itself the derivative of a conditional expectation of an indicator function.

[11]We also conclude that the score estimator, an element of the class of MCMD-W estimators, is similar to the estimator (12)-(13), an element of the class of MCMD-P estimators. Both can be derived by conditioning expectations of the Malliavin derivative on the terminal value of the diffusion. Given that these conditional expectations are generally unavailable, both estimators are often infeasible in practice. In particular, the corresponding MCMD-W estimator can not be calculated in explicit form if the log-likelihood ratio is unknown.

where derivatives are taken with respect to a perturbation parameter λ, that controls shifts of the Brownian paths, in the direction of the suitably chosen process $\nu_{t,s}^{(\alpha)}$ defined in (20). The Malliavin weight is therefore a score weight corresponding to a particular log-likelihood obtained from (24). It is found using the same integration-by-parts idea as the traditional score estimator.

3.1.3 Malliavin Derivative Mixed Estimators (MCMD-M)

MCMD-P estimators involve the derivative of the payoff function h and therefore take into account the local sensitivity of the payoff function with respect to variations in the terminal value caused by infinitesimal perturbations of the initial value. In contrast, MCMD-W estimators use weights $H_{t,T}^{(\alpha)}$ that are the same for different payoff functions. These weights are therefore independent of the derivative of the payoff function h. This difference in sensitivity results in a higher variance for MCMD-W estimators. This is confirmed by the experiments for geometric Brownian motion carried out in Broadie and Glasserman (1996).[12]

Fournié et. al. (1999) resolve this difficulty by using a localization procedure which, for geometric Brownian motion, improves Malliavin weight estimators. The basic idea behind their approach is an additive decomposition of the payoff function in a once continuously differentiable function plus a differentiable function with discontinuous derivative but support restricted to compact intervals $(x' - \delta, x' + \delta)$ around points of discontinuity x' of its derivatives. If we use MCMD-P for the continuously differentiable part and MCMD-W for the remaining part containing the discontinuities of the derivative of the payoff function we obtain their localized estimators. Given the exact nature of this decomposition, for arbitrary choices of localization parameters δ, convergence properties of the estimators are not affected. We call their estimator a Monte Carlo Malliavin derivative mixture (MCMD-M) estimator.

3.1.4 The Asymptotic Error Behavior of Malliavin Derivative-based Estimators

All three MCMD estimators are based on probabilistic representations of the same form. Using Malliavin calculus, we can define a possibly augmented diffusion Y and a function g such that $\partial_x f(t,x)$ has the probabilistic representation (9). This formula is valid even when transition densities are unknown and/or when the payoff functions are not smooth. This representation has the same form as the one for the function itself, and thus, achieves the optimal rate of convergence for Monte Carlo estimators implied by the Central Limit theorem for i.i.d. random variables.

To describe the asymptotic behavior of MCMD estimators of derivatives, a result of Detemple, Garcia and Rindisbacher (2004) on asymptotic distributions of Monte Carlo estimators for Euler discretized diffusions, can be applied. The following theorem describes the weak limit of the

[12]Fournié et. al. (2001) provide theoretical results explaining the observation of Broadie and Glasserman (1996).

estimation error.[13]

Theorem 1: *Consider a function $g \in \mathcal{C}^3(\mathbb{R}^d)$ and suppose that $g(X_T) \in \mathbb{D}^{1,2}$. Also assume that the conditions of Theorem 5 in Appendix A hold, and*

$$\lim_{r\to\infty} \mathbf{E}_{t,x}\left[\mathbf{1}_{\{g(X_T)^2>r\}} g(X_T)^2\right] = 0. \tag{25}$$

Then, as $M \to \infty$,

$$\sqrt{M}\left(\partial_x f^{M,N_M}(t,x) - \partial_x f(t,x)\right) \Rightarrow \epsilon^{md}\frac{1}{2}K_{t,T}(x) + L_{t,T}(x)$$

where $N_M \to \infty$, as $M \to \infty$, $\epsilon^{md} = \lim_{M\to\infty} \sqrt{M}/N_M$ *and $L_{t,T}(x)$ is the terminal value of a Gaussian martingale with (deterministic) quadratic variation and conditional variance given by*

$$[L,L]_{t,T} = \int_t^T \mathbf{E}_{t,x}[N_s(N_s)']ds = \mathbf{VAR}_{t,x}[g(X_T)]$$
$$N_s = \mathbf{E}_{s,x}[\partial g(X_T)\mathcal{D}_s X_T].$$

The second order bias function $K_{t,T}(x)$ is defined in equation (40) in Appendix A for $t = 0$.

The theorem shows that the asymptotic law of the estimator has two parts. The Gaussian martingale L results from the Central Limit theorem, in the approximation of the expectation by an empirical mean over independent replications. The function $K_{t,T}(x)$ appears because using the Euler scheme amounts to sampling from random variables that are only convergent approximations of the true terminal point of the diffusion. It represents a second order discretization bias. This bias disappears if the end point of the diffusion can be simulated directly.

Condition (25) is a uniform integrability condition for centered second moments. A sufficient condition for (25) is that $g(X_T)$ be $\mathbf{L}^p$-bounded, for some $p > 2$. In particular, the existence of third order moments for $g(X_T)$ is sufficient.[14]

3.2 Monte Carlo Covariation (MCC)

The covariation representation

$$\partial_x f(t,x) = \lim_{1/\tau\to\infty} \mathbf{E}_{t,x}\left[g(X_T)\left(\frac{W_{t+\tau}-W_t}{\tau}\right)'\right]C(x)$$

[13] Let S be a metric space and $\mathcal{S}$ its Borel sets. A sequence of random variables X^N is said to converge weakly to a random variable X, denoted by $X^N \Rightarrow X$ whenever, with $\mathbf{P}_{X^N} \equiv \mathbf{P}\circ(X^N)^{-1}$ and $\mathbf{P}_X \equiv \mathbf{P}\circ X^{-1}$, we have $\int_S f(s)d\mathbf{P}_{X^N}(s) \to \int_S f(s)d\mathbf{P}_X(s)$ for all continuous and bounded functions f on S.

[14] The formula for the second order bias, K, requires that g be three times continuously differentiable. Results of Bally and Talay (1996) guarantee that the same convergence result holds even if the payoff function is not differentiable. A similar type of efficiency result for Monte Carlo estimators of functions expressed as conditional expectations of diffusions, can be found in Duffie and Glynn (1995), but without an explicit formula for the second order bias.

naturally suggests the estimator

$$\widetilde{\partial_x f(t,x)}^{M,N,\tau} = \left(\frac{1}{M}\sum_{i=1}^{M} g(X_T^{i,N})\left(\frac{W_{t+\tau}^i - W_t^i}{\tau}\right)'\right) C(x), \tag{26}$$

where C is defined in (12) and X_T^N is generated using a Euler approximation of the stochastic differential equation based on N discretization points. This simple approach was first proposed by Cvitanic, Goukasian and Zapatero (2002, 2003), who implement it in the context of asset allocation problems and risk management problems for derivatives. The method seems attractive from a computational point of view as it does not require the simulation of auxiliary processes such as Malliavin derivatives, Malliavin weights, or diffusions with perturbed initial values.

The asymptotic distribution of the error is as follows.

Theorem 2: *Consider a function $g \in \mathcal{C}^3(\mathbb{R}^d)$ and suppose that $g(X_T) \in \mathbb{D}^{1,2}$. Let $\Delta_\tau W_t \equiv W_{t+\tau} - W_t$ and $K_T(x)$ be defined as in Appendix A, equation (40).*

Define the events

$$F_{t,T}(N,\tau,r) = \left\{\left\| N\left(g(X_T^N) - g(X_T)\right)\frac{\Delta_\tau W_t}{\tau} - \frac{1}{2} I_{t,T}^{N,\tau}(x)\right\| > r\right\},$$

$$G_{t,T}(\tau,r) = \left\{\left\| g(X_T)\frac{\Delta_\tau W_t}{\tau} - \mathbf{E}_{t,x}\left[g(X_T)\frac{\Delta_\tau W_t}{\tau}\right]\right\| > r\right\},$$

where $I_{t,T}^{N,\tau}(x) = \mathbf{E}_{t,x}\left[N\left(g(X_T^N) - g(X_T)\right)\frac{\Delta_\tau W_t}{\tau}\right]$, and suppose that the conditions

$$\lim_{r\to\infty}\limsup_{1/\tau,N\to\infty}\mathbf{E}_{t,x}\left[\mathbf{1}_{F_{t,T}(N,r,\tau)}\left\| N\left(g(X_T^N) - g(X_T)\right)\frac{\Delta_\tau W_t}{\tau} - \frac{1}{2} I_{t,T}^{N,\tau}(x)\right\|\right] = 0, \tag{27}$$

$$\lim_{r\to\infty}\limsup_{1/\tau\to\infty}\mathbf{E}_{t,x}\left[\mathbf{1}_{G_{t,T}(\tau,r)}\left\| g(X_T)\frac{\Delta_\tau W_t}{\tau} - \mathbf{E}_{t,x}\left[g(X_T)\frac{\Delta_\tau W_t}{\tau}\right]\right\|^2\right] = 0 \tag{28}$$

hold. Then, as $M \to \infty$,

$$\begin{aligned} M^{1/3}\left(\widetilde{\partial_x f(t,x)}^{M,N_M,\tau_M} - \partial_x f(t,x)\right) \;\Rightarrow\; & \varepsilon_1^c \partial_v \mathbf{E}_{t,x}[\partial_x f(v,X_v)B(X_v)]_{|v=t} C(x) \\ & + \varepsilon_2^c \frac{1}{2}\partial K_{t,T}(x) + O_{t,T}(x)C(x) \end{aligned}$$

where $N_M, 1/\tau_M \to \infty$ when $M \to \infty$, $\varepsilon_1^c = \lim_{M\to\infty} M^{1/3}\tau_M$ and $\varepsilon_2^c = \lim_{M\to\infty} M^{1/3}/N_M$, C is defined in (12), and where $O_{t,T}$ is the terminal value of a Gaussian martingale with (deterministic) quadratic variation $[O,O]_{t,T}(x) = \mathbf{E}_{t,x}[g(X_T)^2] I_d$.

The uniform integrability condition (27) is necessary and sufficient for the convergence of the mean of the second order discretization bias to the expectation of the limiting error distribution. A sufficient condition for (27) is uniform $\mathbf{L}^p$-boundedness of $N(g(X_T^N) - g(X_T))\frac{\Delta_\tau W_t}{\tau}$ for some $p > 1$, jointly in τ and N. This sufficient condition is satisfied, in particular, if the payoff is bounded. The second uniform integrability condition (28) is sufficient for the Lindeberg condition for triangular arrays to hold. This standard condition is invoked to obtain the Gaussian limit process from a Central Limit theorem (see Kallenberg (1997), Theorem 4.12, p. 69). Sufficient conditions for (28) are uniform $\mathbf{L}^p$-boundedness of $g(X_T)\frac{\Delta_\tau W_t}{\tau}$ for some $p > 2$ in τ.

Theorem 2 shows that the best convergence rate for the estimator based on the covariation is slower than that for the estimator using Malliavin derivatives. The efficient MC scheme requires an eight-fold increase in the number of independent replications M along with a cut of the initial shift τ by half when the number discretization points N is doubled.

The second order bias, in this scheme, has two components. The first, $\frac{1}{2}\partial K_{t,T}(x)$, is due to the fact that sampling from the true distribution of the terminal point of the SDE is not feasible. The second, $\partial_v \mathbf{E}_{t,x}[\partial_x f(v, X_v)B(X_v)]_{|v=t}$, comes from the fact that one must deal with the derivative of the second order bias. This derivative appears because MCC effectively estimates the smoothed derivative $\frac{1}{\tau}\int_t^{t+\tau} \partial_x f(v, X_v)B(X_v)dvC(x)$, instead of $\partial_x f(t, x)$. MCC estimators are therefore only correct in the limit as $1/\tau \to \infty$. In practice, the selection of τ seems to be a non-trivial task and the resulting second order bias is a matter of concern.

In order to produce asymptotic confidence intervals which do not suffer from size distortion the second order bias must be calculated. This task is considerably more difficult for this method than for MCMD because the derivative of the second order bias must be computed. However, as the computational effort needed to calculate the MCC estimator is smaller, it is not uniformly dominated by the MCMD estimator.

The comparison of MCC with MCMD-W reveals that MCC effectively approximates the instantaneous Malliavin weight $H_{t,T}^{(\alpha^{mcc})}$, obtained for the choice $\alpha_{t,s}^{mcc} = \frac{1}{\tau}\mathbf{1}_{(t,t+\tau]}(s)$, by its Euler approximation with step size τ, $[\hat{H}_{t,T}^{(\alpha^{mcc})}]^\tau = \left(\frac{W_{t+\tau} - W_t}{\tau}\right)' C(X_t)$. The MCC method therefore uses the exact weight only in the limit when $1/\tau \to \infty$. For fixed τ, these approximate Malliavin weights introduce an additional second order bias and reduce the overall convergence rate.

3.3 Monte Carlo with Finite Difference (MCFD)

An estimate of (11) can also be produced by approximating the derivative inside the expectation by a finite difference, in the manner of numerical PDE methods. If the function is then evaluated by averaging over independent replications one obtains a Monte Carlo version of the well known finite difference method for PDEs. The motivation for this estimator is the limiting result

$$\partial_{x_j} f(t, x) = \lim_{1/\tau_j \to \infty} \frac{\mathbf{E}_{t,x}\left[g(X_T(x + \alpha_j \tau_j e_j))\right] - \mathbf{E}_{t,x}\left[g(X_T(x - (1 - \alpha_j)\tau_j e_j))\right]}{\tau_j} \tag{29}$$

for $\alpha_j \in [0,1]$, $j = 1,\ldots,d$, where $X_T(x)$ is the diffusion process started at $X_t = x$ and $e_j = [0,...,1,...,0]'$ is the j^{th} unit vector of dimension d. The choice $\alpha = 1$ corresponds to single forward differences, $\alpha = 0$ to single backward differences, and $\alpha = 1/2$ to central differences.

The relation above suggests the Monte Carlo finite difference (MCFD) estimator

$$\widehat{\partial_y f(t,x)}^{M,N,\tau_j,\alpha_j} = \left(\frac{\frac{1}{M}\sum_{i=1}^{M}\left[g(X_T^{i,N}(x+\alpha_j\tau_j e_j)) - g(X_T^{i,N}(x-(1-\alpha_j)\tau_j e_j))\right]}{\tau_j} \right)_{j=1,\ldots,d} \tag{30}$$

where $X_T^{i,N}$ is an approximation of X_T based on the Euler scheme using N discretization points. This perturbation approach was proposed by Glynn (1989) and Glasserman (1991) and its convergence properties were studied by L'Ecuyer and Perron (1994).[15] All these papers consider a setup where sampling from the true distribution is feasible. In our context the underlying variables satisfy nonlinear diffusions which, in general, prevents sampling from the true distribution. Instead, a discretization scheme is employed to approximate the state variables. The asymptotic distribution will then depend on two sources of error.

Our next theorem establishes the convergence properties for this procedure. It shows that the convergence rate depends on the choices of α, hence on the difference scheme selected, as well as on the perturbation parameter τ.

Theorem 3: *Consider a function $g \in \mathcal{C}^3(\mathbb{R}^d)$ and suppose that $g(X_T) \in \mathbb{D}^{1,2}$. Let $K_{t,T}(x)$ be defined as in Appendix A, equation (40). Define the events*

$$F_{t,T}^{j}(N,\tau_j,r) = \left\{ \left| N\nabla_{t,x_j}^{\tau_j,\alpha_j} g(X_T^N(x)) - \mathbf{E}_{t,x}\left[N\nabla_{t,x_j}^{\tau_j,\alpha_j} g(X_T^N(x))\right]\right| > r \right\}$$

$$G_{t,T}^{j}(\tau_j,r) = \left\{ \left| \nabla_{t,x_j}^{\tau_j,\alpha_j} g(X_T(x)) - \mathbf{E}_{t,x}\left[\nabla_{t,x_j}^{\tau_j,\alpha_j} g(X_T(x))\right]\right| > r \right\}$$

where

$$\nabla_{t,x_j}^{\tau_j,\alpha_j} g(X_T^N(x)) \equiv \frac{g(X_T^N(x+\alpha_j\tau_j e_j)) - g(X_T^N(x-(1-\alpha_j)\tau_j e_j))}{\tau_j}$$

and $\nabla_{t,x_j}^{\tau_j,\alpha_j} g(X_T(x))$ is defined in a similar manner, substituting X_T for X_T^N. Suppose that the conditions

$$\lim_{r\to\infty}\limsup_{1/\tau_j,N\to\infty} \mathbf{E}_{t,x}\left[\mathbf{1}_{F_{t,T}^{j}(N,\tau_j,r)}\left|N\nabla_{t,x_j}^{\tau_j,\alpha_j} g(X_T^N) - \mathbf{E}_{t,x}\left[N\nabla_{t,x_j}^{\tau_j,\alpha_j} g(X_T^N(x))\right]\right|\right] = 0 \tag{31}$$

$$\lim_{r\to\infty}\limsup_{1/\tau_j\to\infty} \mathbf{E}_{t,x}\left[\mathbf{1}_{G_{t,T}^{j}(\tau_j,r)}\left|\nabla_{t,x_j}^{\tau_j,\alpha_j} g(X_T(x)) - \mathbf{E}_{t,x}\left[\nabla_{t,x_j}^{\tau_j,\alpha_j} g(X_T(x))\right]\right|^2\right] = 0 \tag{32}$$

hold, for all $j = 1,\ldots,d$. Then, as $M \to \infty$,

[15] For a unified view of the perturbation method and the likelihood ratio/score method see L'Ecuyer (1990).

(i) if $\alpha_j = 1/2$ (MCCFD) we have

$$M^{1/2}\left(\widehat{\partial_{x_j} f(t,x)}^{M,N_M,\tau_{j,M},1/2} - \partial_{x_j} f(t,x)\right) \Rightarrow \varepsilon_{j1}^{fcd}\frac{1}{24}\partial_{x_j}^3 f(t,x) + \varepsilon_{j2}^{fcd}\frac{1}{2}\partial_{x_j} K_{t,T}(x) + Q_{t,T}^j(x)$$

where $N_M, 1/\tau_{j,M} \to \infty$ when $M \to \infty$, $\varepsilon_{j1}^{fcd} = \lim_{M\to\infty} M^{1/4}\tau_{j,M}$ and $\varepsilon_{j2}^{fcd} = \lim_{M\to\infty} M^{1/2}/N_M$,
(ii) if $\alpha_j \neq 1/2$ (MCBFD and MCFFD)

$$M^{1/2}\left(\widehat{\partial_{x_j} f(t,x)}^{M,N_M,\tau_{j,M},\alpha_j} - \partial_{x_j} f(t,x)\right) \Rightarrow \varepsilon_{j1}^{fd}\delta(\alpha_j)\partial_{x_j}^2 f(t,x) + \varepsilon_{j2}^{fd}\frac{1}{2}\partial_{x_j} K_{t,T}(x) + Q_{t,T}^j(x)$$

where $N_M, 1/\tau_{j,M} \to \infty$ when $M \to \infty$, with $\delta(\alpha_j) = (2\alpha_j - 1)/2$, $\varepsilon_{1j}^{fd} = \lim_{M\to\infty} M^{1/2}\tau_{j,M}$ and $\varepsilon_{j2}^{fd} = \lim_{M\to\infty} M^{1/2}/N_M$.

The random variable $Q_{t,T}^j(x)$ is the j^{th} element of the terminal value of a Gaussian martingale with quadratic variation $[Q,Q]_{t,T}(x) = \mathbf{E}_{t,x}\left[\int_t^T L(v,X_v)L(v,X_v)'dv\right]$, where $L(v,X_v)' = \mathbf{E}_{v,X_v}\left[\mathcal{D}_v\left(\partial g(X_T)\mathcal{D}_v X_T C(X_v)\right)\right]$ with C defined in (12).

For smooth payoff functions such that $g(X_T) \in \mathbb{D}^{1,2}$, the speed of convergence of MCFD estimators is the same as for the MCMD estimator. But in contrast to MCMD-P estimators, MCFD estimators have an additional second order bias caused by the finite difference approximations of both the derivative of the payoff function and the tangent process.

The MCFD estimator can be viewed as an estimator where the Malliavin derivatives are approximated by a finite difference. This rests on the fact that a Malliavin derivative in a diffusion context corresponds to the tangent process. The finite differences used for the MCFD estimator converge to the derivative of the stochastic flow with respect to the initial condition. This follows from the relation

$$\frac{g(X_T(x+\alpha_j\tau_j e_j)) - g(X_T(x-(1-\alpha_j)\tau_j e_j))}{\tau_j} = J_T^{1,\tau_j} J_T^{2,\tau_j}$$

where J_T^{1,τ_j} is a finite difference approximation of the derivative of the payoff function,

$$J_T^{1,\tau_j} = \frac{g(X_T(x+\alpha_j\tau_j e_j)) - g(X_T(x-(1-\alpha_j)\tau_j e_j))}{X_T(x+\alpha_j\tau_j e_j) - X_T(x-(1-\alpha_j)\tau_j e_j)}$$

and J_T^{2,τ_j} is a finite difference approximation of the derivative of the stochastic flow with respect to its initial condition

$$J_T^{2,\tau_j} = \frac{X_T(x+\alpha_j\tau_j e_j) - X_T(x-(1-\alpha_j)\tau_j e_j)}{\tau_j}.$$

As $J_T^{1,\tau_j} \Rightarrow \partial g(X_T)$ and $J_T^{2,\tau_j} \Rightarrow \mathcal{D}_t X_T C(x) e_j$ as $1/\tau_j \to \infty$ we see that, indeed, MCFD estimators are approximations of MCMD-P estimators. But if the payoff function is differentiable, the variance of the finite difference approximation is of order $\mathbf{O}(\tau_{j,M}^2)$, and therefore does not explode as $\tau_{j,M} \to 0$.[16] This explains why the convergence of MCFD and MCMD are the same.

But it is important to note that to obtain efficient estimators, the perturbation parameters $\tau_{j,M}$ have to be chosen with care. To cut the length of the asymptotic confidence interval by half we must quadruple the number of replications, double the number of discretization points and cut the initial perturbation by half for forward or backward differences and divide by $\sqrt{2}$ for central differences.

Using the true Malliavin derivatives does not require finding the optimal choice of the initial perturbation. It also has the advantage of eliminating the error induced by a finite difference approximation of the tangent process, one of the terms in the second order bias.

Our next result shows that for discontinuous functions $g(X_T) \notin \mathbb{D}^{1,2}$, MCMD-W estimators outperform MCFD estimators in terms of convergence speed. Indicator functions are a good example of discontinuous functions. They arise in risk management applications such as Delta-hedging for digital options or Delta-Gamma hedging for vanilla call options. Our next theorem summarizes asymptotic convergence properties covering those cases.

Theorem 4: *Consider a function g such that $g(x) = g^c(x) + \sum_{j=1}^{\infty} \gamma_j \mathbf{1}_{B_j}(x)$ where $B_i \cap B_j = \emptyset$ for $i \neq j$, with $g^c(x) \in \mathcal{C}^3(\mathbb{R}^d)$ and suppose that $g^c(X_T) \in \mathbb{D}^{1,2}$. Let $K_T(x)$ be defined as in Theorem 5 of Appendix A, equation (40). Suppose that the conditions (31) and (32) of Theorem 3 hold, for all $j = 1, \ldots, d$. Then, as $M \to \infty$,*
(i) if $\alpha_j = 1/2$ (MCCFD) we have

$$M^{2/5}\left(\widehat{\partial_{x_j} f(t,x)}^{M,N_M,\tau_{j,M},1/2} - \partial_{x_j} f(t,x)\right) \Rightarrow \varepsilon_{j1}^{fcd} \frac{1}{24} \partial_{x_j}^3 f(t,x) + \varepsilon_{j2}^{fcd} \frac{1}{2} \partial_{x_j} K_{t,T}(x) + Q_{t,T}^j(x)$$

where $N_M, 1/\tau_{j,M} \to \infty$ when $M \to \infty$, $\varepsilon_{j1}^{fcd} = \lim_{M\to\infty} M^{1/5} \tau_{j,M}$ and $\varepsilon_{j2}^{fcd} = \lim_{M\to\infty} M^{2/5}/N_M$,
(ii) if $\alpha_j \neq 1/2$ (MCBFD and MCFFD)

$$M^{1/3}\left(\widehat{\partial_{x_j} f(t,x)}^{M,N_M,\tau_{j,M},\alpha_j} - \partial_{x_j} f(t,x)\right) \Rightarrow \varepsilon_{j1}^{fd} \delta(\alpha_j) \partial_{x_j}^2 f(t,x) + \varepsilon_{j2}^{fd} \frac{1}{2} \partial_{x_j} K_{t,T}(x) + Q_{t,T}^j(x)$$

where $N_M, 1/\tau_{j,M} \to \infty$ when $M \to \infty$, with $\delta(\alpha_j) = (2\alpha_j - 1)/2$, $\varepsilon_{1j}^{fd} = \lim_{M\to\infty} M^{1/3} \tau_{j,M}$ and $\varepsilon_{j2}^{fd} = \lim_{M\to\infty} M^{1/3}/N_M$.

The random variable $Q_{t,T}^j(x)$ is the terminal value of a Gaussian martingale with quadratic

[16] The notation $\mathbf{O}(\cdot)$ denotes Landau's "at most of order". See footnote (20) for a precise definition.

variation

$$
\begin{aligned}
[Q^j, Q^j]_{t,T}(x) \;=\; & \sum_{k=1}^{\infty} \gamma_k^2 \left(2\alpha_j - 1\right) \partial_{x_j} \mathbf{P}_{t,x} \left(X_T(x) \in B_k\right) \\
& -2\alpha_j \sum_{k,l=1}^{\infty} \gamma_k \gamma_l \partial_{x_j} \mathbf{P}_{t,x} \left(\{X_T(x) \in B_k\} \cap \left\{X_T(x') \in B_l\right\}\right)_{|x'=x} \\
& +2(1-\alpha_j) \sum_{k,l=1}^{\infty} \gamma_k \gamma_l \partial_{x'_j} \mathbf{P}_{t,x} \left(\{X_T(x) \in B_k\} \cap \left\{X_T(x') \in B_l\right\}\right)_{|x'=x}, \quad (33)
\end{aligned}
$$

and such that $Q^j_{t,T}(x)$ *and* $Q^k_{t,T}(x)$ *are mutually independent for* $j \neq k$ *and all* $x \in \mathbb{R}^d$*.*

Theorem 4 shows that the best convergence rate for the estimators based on finite differences for discontinuous payoff functions is lower than that for the estimators using Malliavin weights. It is faster than for estimators based on the covariation, only for central differences. For central differences, the efficient Monte Carlo scheme mandates an increase in the number of independent replications M by a factor of $2^{5/2}$ when the number discretization points N is doubled. In addition, one must simultaneously cut the initial shift τ by a factor of $\sqrt{2}$. For discontinuous payoff functions, the efficient scheme for forward/backward differences is the same as the one based on the covariation.

As this scheme also produces a numerical approximate of an approximation of the derivative, the second order bias has two components. The first one, $\frac{1}{2}\partial K_{t,T}(x)$, is the same as the one for the covariation estimator. The second one depends on whether central or other differences are used. The faster convergence rate stems from the fact that central differences are second order accurate approximations of the function to be approximated.

Similarly, as shown in the previous section, MCC weights are approximations of MCMD weights. Therefore MCMD-W estimators converge faster than MCC estimators. The slower convergence rate of MCC and MCFD estimators, relative to MCMD-W and MCMD-P estimators, is also accompanied by additional second order biases in the asymptotic limit distribution.

MCMD, MCFD and MCC estimators differ in terms of the number of auxiliary processes that have to be simulated. MCMD-P estimators require the simulation of Malliavin derivatives, whereas for MCMD-W estimators one has to calculate a Malliavin weight. MCFD estimators involve the calculation of perturbed diffusions that approximate the Malliavin derivative, whereas MCC estimators do not require auxiliary processes, because they approximate an instantaneous Malliavin weight. Given these differences in computational requirements, the ordering of the methods in terms of convergence speed is not necessarily preserved when rankings are based on CPU time. The next section investigates the efficiency of MCMD, MCFD and MCC estimators.

4 Numerical Examples

This section compares the performances of the various methods reviewed in applications to optimal portfolio choice and to the hedging of digital options.

4.1 Portfolio Choice

In the dynamic portfolio choice problem of Merton (1971) stock returns P^i and state variables Y satisfy the joint diffusion

$$dP_t^i/P_t^i = r(t,Y_t)dt + \sigma_i(t,Y_t)'[\theta(t,Y_t)dt + dW_t], \qquad i = 1, ..., d$$

$$dY_t = \mu^Y(t,Y_t)dt + \sigma^Y(t,Y_t)dW_t$$

where r is the riskless short rate and θ the market price of risk (MPR). The classic solution of this problem writes the optimal portfolio policy (i.e. the fractions of wealth invested in stocks) as

$$\pi_t'\sigma(t,Y) = \frac{\partial_x V}{-x\partial_{xx}V}\theta(t,Y)' + \frac{\partial_{xy}V}{-x\partial_{xx}V}\sigma^Y(t,Y)$$

where $V(t,x,y) = \sup_\pi \mathbf{E}[u(X_T^\pi)|X_t = x, Y_t = y]$ is the value function. In this expression X_T^π stands for the terminal wealth resulting from a policy π (X_t is wealth at date t) and $u(\cdot)$ is the utility function. The value function solves the Hamilton-Jacobi-Bellman partial differential equation

$$\begin{aligned} 0 &= \partial_t V + \partial_y V \mu^Y + \frac{1}{2}trace\{\partial_{yy}V\sigma^Y(\sigma^Y)'\} + x\partial_x V r \\ &\quad -\frac{1}{2}x^2\partial_{xx}V\|\frac{\partial_x V}{-x\partial_{xx}V}\theta' + \frac{\partial_{xy}V}{-x\partial_{xx}V}\sigma^Y\|^2 \end{aligned}$$

subject to the boundary condition $V(T,x) = u(T,x)$. Merton (1971) emphasized the importance of the intertemporal hedging component of the portfolio (the second term in the portfolio formula), which is caused by fluctuations in investment opportunities.

For constant relative risk aversion (CRRA) $u(x) = x^{1-R}/(1-R)$ the hedging demand is the only part that is not in explicit form. Simple manipulations show that $V(t,x,y) = \frac{x^{1-R}}{1-R}f(t,y)^R$ where $f(t,y)$ solves the linear PDE

$$\mathcal{L}_t f - \rho\partial_y f\sigma^Y\theta + \left[\frac{1}{2}\rho(\rho-1)\|\theta\|^2 - \rho r\right]f = 0, \tag{34}$$

subject to the boundary condition $f(T,y) = 1$. In (34) $\mathcal{L}_t f = \partial_t f + \partial_y f\mu^Y + \frac{1}{2}trace\{\partial_{yy}f\sigma^Y(\sigma^Y)'\}$ is the infinitesimal generator of the diffusion process for the state variables and $\rho = 1 - 1/R$. Expressed in terms of f the optimal portfolio becomes

$$\pi_t'\sigma(t,Y) = \frac{1}{R}\theta(t,Y)' + \frac{\partial f}{f}\sigma^Y(t,Y).$$

With constant relative risk aversion the fractions of wealth in stocks are independent of wealth.

The Feynman-Kac formula (Karatzas and Shreve (1991), Theorem 7.6., p. 366) links the solution of (34) to the Monte Carlo methods described in the previous section. Let $k(t, Y_t) \equiv \mathbf{E}_{t,Y_t}[\xi_{t,T}^{\rho}]$, where $\xi_{t,T} \equiv \exp\left(-\int_t^T r_s ds - \int_t^T \theta_s' dW_s - \frac{1}{2}\int_t^T \theta_s'\theta_s ds\right)$ is the state price density implied by the market structure (S, Y, θ, r). Passing to the new measure $\tilde{\mathbf{P}}$, under which $\tilde{W}_t = W_t + \rho \int_0^t \theta(s, Y_s)ds$ is a Brownian motion, gives

$$k(t, Y_t) = \tilde{\mathbf{E}}_{t,Y_t}\left[\exp\left(-\rho\int_t^T r(s, Y_s)ds + \frac{1}{2}\rho(\rho-1)\int_t^T \|\theta(s, Y_s)\|^2 ds\right)\right]$$

where $\tilde{\mathbf{E}}$ is the expectation under $\tilde{\mathbf{P}}$ and $dY_t = \tilde{\mu}^Y(t, Y_t)dt + \sigma^Y(t, Y_t)d\tilde{W}_t$ with $\tilde{\mu}^Y = \mu^Y - \rho\sigma^Y\theta$. The Feynman-Kac formula then shows that

$$\tilde{\mathcal{L}}_t k + \left[\frac{1}{2}\rho(\rho-1)\|\theta\|^2 - \rho r\right] k = 0; \text{ and } k(T, Y) = 1 \text{ for all } Y \in \mathbb{R}^d.$$

Using $\tilde{\mathcal{L}}_t k = \mathcal{L}_t k + \partial_y k(-\rho\sigma^Y\theta)$ gives $\mathcal{L}_t k - \rho\partial_y k\sigma^Y\theta + \left[\frac{1}{2}\rho(\rho-1)\|\theta\|^2 - \rho r\right] k = 0$, subject to $k(T, Y_T) = 1$. In light of (34) we conclude that $f = k$. It follows that $f(t, y) = \mathbf{E}_{t,y}[\xi_{t,T}^{\rho}]$.

Let us now review the three Monte Carlo methods that can be used to estimate the hedging demand $(\partial f/f)\sigma^Y$. The first one is based on the limiting result

$$\pi_t'\sigma_t = \lim_{\tau\to 0}\frac{1}{\tau}\mathbf{E}_{t,Y_t}\left[\left(\frac{X_{t+\tau} - X_t}{X_t}\right)(W_{t+\tau} - W_t)\right] = \lim_{\tau\to 0}\frac{1}{\tau}\mathbf{E}_{t,Y_t}\left[\xi_{t,t+\tau}X_{t,t+\tau}\frac{(W_{t+\tau} - W_t)}{\xi_{t,t+\tau}}\right].$$

This limit follows from the fact that the optimal portfolio is, up to a scaling factor, equal to the covariation between optimal wealth and the Brownian motion ($d[X, W]_t/X_t = \pi_t'\sigma_t dt$). As optimal wealth equals $\xi_t X_t = \lambda^{-1/R}\xi_t^{\rho}\mathbf{E}_{t,Y_t}[\xi_{t,T}^{\rho}]$ where λ is the Lagrange multiplier for the static budget constraint (see Karatzas, Lehoczky and Shreve (1987)) we also have $\xi_{t+\tau}X_{t+\tau}/\xi_t X_t \equiv \xi_{t,t+\tau}X_{t,t+\tau} = \mathbf{E}_{t+\tau,Y_{t+\tau}}[\xi_{t,T}^{\rho}]/\mathbf{E}_{t,Y_t}[\xi_{t,T}^{\rho}]$. Combining these expressions shows that

$$\pi_t'\sigma_t = \lim_{\tau\to 0}\frac{1}{\tau}\frac{\mathbf{E}_{t,Y_t}\left[\xi_{t,T}^{\rho}\frac{(W_{t+\tau} - W_t)}{\xi_{t,t+\tau}}\right]}{\mathbf{E}_{t,Y_t}\left[\xi_{t,T}^{\rho}\right]}.$$

This formula is the basis for the approach proposed by Cvitanic, Goukasian and Zapatero (2002, 2003). They suggest computing the portfolio based on the formula on the right hand side with τ fixed. This approach is clearly just a special case of MCC methods presented in section 2.2.

The second Monte Carlo method under consideration is based on the expression obtained by taking Malliavin derivatives on both sides of $f(t, Y_t) = \mathbf{E}_{t,Y_t}[\xi_{t,T}^{\rho}]$. For the left hand side, this gives

$$\mathcal{D}_t f(t, Y_t) = \partial_y f(t, Y_t)\sigma^Y(t, Y_t).$$

For the right hand side, we obtain

$$\mathcal{D}_t \mathbf{E}_{t,Y_t}[\xi^{\rho}_{t,T}] = -\rho f(t, Y_t)(a(t, Y_t) + b(t, Y_t)),$$

where the functions a and b are

$$a(t, Y_t) \equiv \frac{\mathbf{E}_{t,Y_t}[\xi^{\rho}_{t,T} \int_t^T \mathcal{D}_t r_s ds]}{\mathbf{E}_{t,Y_t}[\xi^{\rho}_{t,T}]} \qquad b(t, Y_t) \equiv \frac{\mathbf{E}_{t,Y_t}[\xi^{\rho}_{t,T} \int_t^T (dW_s + \theta_s ds)' \mathcal{D}_t \theta_s ds]}{\mathbf{E}_{t,Y_t}[\xi^{\rho}_{t,T}]}$$

with $\mathcal{D}_t r_s = \partial_y r(s, Y_s) \mathcal{D}_t Y_s$ and $\mathcal{D}_t \theta_s = \partial_y \theta(s, Y_s) \mathcal{D}_t Y_s$. The optimal portfolio becomes

$$\pi'_t \sigma_t = \frac{1}{R} \theta'_t - \rho(a(t, Y_t) + b(t, Y_t)).$$

The MCMD-P estimator provides estimates for the functions a and b.

Finally, recall that the hedging demand is $\partial_y f(t, y)/f(t, y)$ with $f(t, y) = \mathbf{E}_{t,y}[\xi^{\rho}_{t,T}]$. As the state price density $\xi_{t,T}$ starts at $\xi_{t,t} = 1$ and depends only on the state variables Y, through the coefficients $\theta(t, y), r(t, y)$, we obtain

$$\pi'_t \sigma_t = \frac{1}{R} \theta'_t + \frac{\lim_{\tau \to 0} \frac{1}{\tau} \left(\mathbf{E}_{t,Y_t + \alpha\tau}[\xi^{\rho}_{t,T}] - \mathbf{E}_{t,Y_t - (1-\alpha)\tau}[\xi^{\rho}_{t,T}] \right)}{\mathbf{E}_{t,Y_t}[\xi^{\rho}_{t,T}]} \sigma^Y$$

where $\alpha \in \{1, 1/2, 0\}$. MCFD estimators are obtained by estimating the conditional expectations by the empirical mean over independent replications of the terminal value of the diffusion for the state price density, $\xi_{t,T}$. The estimate is computed using a Euler scheme starting at 1 and starting the driving diffusion Y at points Y_t, $Y_t + \alpha\tau$ and $Y_t + (1 - \alpha)\tau$. This yields MCBFD for $\alpha = 0$, MCCFD for $\alpha = \frac{1}{2}$ and MCFFD for $\alpha = 1$.

Table 1 summarizes the estimators for the different methods. For all methods the parameter N corresponds to the number of discretization points in time and the parameter M to the number of replications. For finite difference methods the parameter τ is the perturbation of the initial value used to calculate numerical derivatives of the function $f^{M,N}$; for the MCC estimator this parameter represents the time step for the Brownian increment needed to calculate the covariation.

[Insert Table 1 about here]

Table 2 describes the approximation schemes used to calculate the components of the estimators above. It identifies the discretized processes for the state variables and their Malliavin derivatives and the discretized hedges.

[Insert Table 2 about here]

The tables show that all methods, except the one based on Malliavin derivatives, involve three convergence parameters, M, N, τ. This last parameter appears because MCFD and MCC do not approximate the optimal portfolio policy, but just a convergent approximation of it. As a result they involve an additional second order bias that slows down their optimal convergence rate. MCFD and MCC estimators are computed by drawing random variables whose variance depends on the discretization parameter τ. To find the error distribution this variance is normalized by $\sqrt{\tau}$ in order to satisfy the Lindeberg condition and apply a Central Limit theorem for i.i.d random variables. The perturbation parameter τ must be controlled along with the number of Monte Carlo replications in order to keep the variance of the estimator finite.

The computational requirements of these various MC approaches are very different. Forward and backward MCFD require the simulation of one auxiliary process, namely the process with perturbed initial condition. For central differences each process must be perturbed twice. The method based on Malliavin derivatives (MCMD) requires, for each state variable, the simulation of a Malliavin derivative with respect to each Brownian motion. In contrast the covariation method (MCC) does not require auxiliary processes. From these observations and the previous results it is clear that no method dominates in all dimensions. Indeed, the higher convergence speed of MCMD comes at the cost of having to simulate additional auxiliary processes. The covariation method, that does not require auxiliary processes, will always dominate the others in terms of computation time. Its convergence rate, however, is slower.

Let us now illustrate the performance of the various methods for the linear model in Wachter (2002). In this model the short rate r is constant, whereas the MPR θ follows an OU-process

$$d\theta_t = A(\overline{\theta} - \theta_t)dt + \Sigma dW_t; \quad \theta_0 \text{ given}, \tag{35}$$

where $A, \overline{\theta}$ and Σ are positive constants.

As shown in Wachter (2002), if the determinant condition $\Sigma^{-2}A^2 + \rho(1 + 2\Sigma^{-1}A) \geq 0$ holds, where $\rho = 1 - 1/R$ and $R > 1$, the optimal portfolio weight is a linear function of state variables. If we define the constants $G \equiv -\Sigma^{-1}A + \sqrt{\Sigma^{-2}A^2 + \rho(1 + 2\Sigma^{-1}A)}$ and $\alpha = 2(A + \Sigma G)$, the optimal demand for the stock of an investor with CRRA preferences over terminal wealth is $\pi_t = \pi_{1t} + \pi_{2t}$ where $\pi_{1t} = X_t(1/R)(\sigma_t)^{-1}\theta_t$ i s the mean-variance demand and

$$\pi_{2t} = -\frac{\rho}{R}\left[\Phi_1(t) + \Phi_2(t)\theta_t\right]\Sigma\sigma^{-1},$$

with

$$\Phi_1(t) \equiv \frac{2\left(1 - \exp\left(-\frac{1}{2}\alpha(T-t)\right)\right)^2}{\alpha\left(\alpha + (\rho - G)\Sigma\left(1 - \exp\left(-\alpha(T-t)\right)\right)\right)}A\overline{\theta} \tag{36}$$

$$\Phi_2(t) \equiv \frac{1 - \exp\left(-\alpha(T-t)\right)}{\alpha + (\rho - G)\Sigma\left(1 - \exp\left(-\alpha(T-t)\right)\right)} \tag{37}$$

represents the intertemporal hedging demand.

Estimates for the parameters are: $A = 0.043875$, $\overline{\theta} = 0.1667$, $\Sigma = -0.0727$. The interest rate is fixed at $r = 0.06$ and the stock volatility at $\sigma = 0.3158$. The initial MPR is $\theta_t = \overline{\theta} = 0.1667$.

To compare the methods we proceed as follows. We consider a mutual fund with 100 different types of clients who can classified in terms of 10 different investment horizons, ranging from 1 to 10 years, and 10 different risk tolerance profiles, with relative risk aversions ranging from 1.5 to 6. For each of these configurations the optimal portfolio and the corresponding error are computed. Note that Monte Carlo methods are particularly well suited to perform this task. While PDE methods require the resolution of the relevant PDEs for each configuration of risk aversion separately, all the Monte Carlo methods use the common estimator $\xi_{t,T}^{N}$ for all risk tolerance profiles.

The experimental setup is as follows. Along the optimal convergence path we simulate six parameter combinations for (M, N, τ) and report the percentage mean square error based on the 100 portfolio weights.

[Insert Figure 1 about here]

We see that MCC does worst whereas MCMD-P, MCBFD and MCFFD do best. These performance rankings correspond to the ordering of the convergence speeds. Central finite differences (MCCFD) does not seem to perform as well as forward finite differences (MCFFD) and backward finite difference (MCBFD). Overall, these simulation results confirm the theoretical findings: MCMD-P and MCFD dominate MCC.

It is important to note that, as the true portfolio weight is affine, the MCFD estimators for the portfolio weight are not sensitive to the choice of the initial perturbation parameter τ. For general nonlinear models, the choice of the perturbation parameter for MCFD and MCC is a non-trivial exercise. If τ is too large, the approximation of the derivatives will be poor. Conversely, if τ is too small the variance of the estimator and therefore the RMSE will explode. It follows that initial perturbation parameters have to be selected using preliminary tests. To perform such tests without prior knowledge about the solution is particularly difficult for multivariate problems. For a random choice of these parameters MCMD-P will most likely strictly dominate MCFD estimators. This additional specification issue does not arise for MCMD-P estimators.

4.2 Risk Management for Digital Options

The previous section illustrates the superior performance of MCMD-P in the context of an asset allocation problem with smooth utility (i.e. payoff) function. We now consider a risk management problem with a non-smooth payoff involving digital options. Hedging digital options is a difficult task because the payoff function, $\mathbf{1}_{\{X_T > K\}}$, is discontinuous at the strike K. Its derivative, which intervenes in the Delta hedge, is the Dirac delta function with mass at the point of discontinuity, i.e. the derivative is null everywhere except at the point K, where it becomes infinite. Resolving this hedging problem is clearly of practical importance. Moreover, any method that can be used to implement hedges for digital options also has ramifications for standard options, as the Gamma of a plain vanilla European-style option exhibits a similar non-smoothness at the strike.[17]

To see that the estimation of the Delta of a digital option is a delicate task it suffices to consider its MCMD-P estimator. Inspection reveals that this estimator is not feasible when the joint transition density of the price process and its Malliavin derivative is unknown. As the probability of the event $\{X_T = K\}$ is null, averaging over independent replications of the Dirac point mass gives an estimate of the hedge identically equal to zero, for any finite number of replications.

In contrast to MCMD-P estimators, MCMD-W estimators do not depend on the local curvature of the payoff function. As a result they remain practically feasible, even if the support of the derivative of the payoff function is concentrated at a single point. As shown by our previous convergence results they are asymptotically optimal estimators of derivatives. We now illustrate their computational efficiency in a simulation experiment involving digital options.

Throughout this section we assume that the underlying asset price follows a process with constant elasticity of variance (CEV),

$$dX_t = X_t \left((r - q)dt + \sigma X_t^{\beta} dW_t \right), \quad X_0 \text{ given},$$

where r is the riskfree rate, q the continuously compounded dividend yield and W is a Brownian motion under the risk neutral measure. Cox (1975) and Emanuel and MacBeth (1982) provide an exact option pricing formula for this model. Both use the fact that the transition density of a CEV process is a non-central chi-square. Knowledge of this density function permits the derivation of an exact formula for the Delta hedge of any option with payoff contingent upon the terminal value of the underlying asset price.

Let us consider a digital option with maturity T, strike K and payoff function $h(T, x) = e^{-rT}\mathbf{1}_{\{x>K\}}$. The price is $f(t, x) = \mathbf{E}_{t,x}[h(T, X_T)]$ where $\mathbf{E}_{t,x}[\cdot]$ is the conditional expectation at date t under the risk neutral measure. MCMD-W, MCC and MCFD estimators for the Delta hedge of this digital option are easily derived from the general expressions for Monte Carlo estimators of derivatives provided in section 3. Formulas can be found in Table 3.

[Insert Table 3 about here].

[17] Broadie and Glasserman (1996) propose a method, based on path derivatives, to calculate the Gamma of a European option in the Black- Scholes setting. For geometric Brownian motion, the path derivative is $\nabla_{t,x}X_T = X_T/x$ where X_T has known transition density $p(t, x, T, \cdot)$. As the derivative of the indicator $\mathbf{1}_{\{X_T \leq K\}}$ is the Dirac delta at K the derivative of the digital becomes $\partial_x \mathbf{E}_{t,x}[\mathbf{1}_{\{X_T \leq K\}}] = \mathbf{E}_{t,x}[\delta_{\{X_T = K\}}\nabla_{t,x}X_T] = \frac{K}{x}p(t, x, T, K)$. This approach, to calculate the path derivative estimator of a non-smooth payoff function, does not work when the transition density is unknown or when the Malliavin derivative of the underlying price is not a deterministic function of this price. The same applies to the MCMD-P estimator of a non-smooth function.

Implementation of MCMD-W and MCMD-M estimators requires the calculation of Malliavin weights. Euler approximations of these weights are given in Table 4.

[Insert Table 4 about here].

In order to assess the relative efficiencies of the various hedging estimators for digital options a large scale experiment is performed as follows. In a first step the parameters $[X, T, r, q, \sigma, \beta]$ of the model are drawn from independent distributions. In a second step the Delta hedge of the digital option is computed, for each draw of parameters, using the various methods. Errors, relative to the true value of the hedge, and computation times are recorded. These calculations are performed using several values of the design parameters N, M and τ. In a last step measures of computation speed (average computation time) and accuracy (root mean square relative error) are computed over the sample of parameter draws. This computation is carried out for each estimator and for each set of values for the design parameters N, M and τ. The end result is a relation, for each estimation method, between computation speed and accuracy as a function of the design parameters.

The choice, in step two, of the design parameters N, M and τ is made in the following manner. The base values for these parameters are set at $N_1 = 10$, $M_1 = 1000$ and $\tau_1 = 0.001$. The increasing sequence $N_i = N_1 + 10 \times (i-1)$ where i increases from 1 to 6 is then considered for the first parameter. The other parameters, M and τ, are adjusted, along this sequence, so as to satisfy the optimal growth restrictions identified in Theorems 1-3.

The parameter distributions are selected to produce a wide sample of market conditions. The strike price is fixed at $K = 100$ and initial asset prices are drawn from a uniform distribution with support $[70, 130]$. Dividend yields are uniformly distributed over $[0, 0.1]$. The interest rate is zero with probability 0.2 and uniform $[0, 0.1]$ with probability 0.8. Similarly, time to maturity is drawn from a mixture of uniform distributions: with probability 0.75 the maturity date is uniform over $[0.1, 1]$ and with probability 0.25 it is uniform over $[1, 5]$. The volatility parameters of the CEV process are such that σ is uniformly distributed over $[0.1, 0.5]$ and β over $[-0.9, 0]$. Finally, the localization parameter δ for MCMD-M is set at $X_0/10$, where X_0 is the initial stock price.

A total of $2,500$ parameter configurations were drawn from the distributions described above. Out of this sample configurations for which the option prices were less than a penny or Delta hedges less than 0.001 were eliminated. This left an effective sample of 990 “good” parameter values.[18]

Figure 2 illustrates the results. First note that the finite difference estimators (MCFD) are all less efficient than estimators involving exact (MCMD-W and MCMD-M) or approximate (MCC)

[18] The elimination of over 50% of the draws is a restriction on the random sampling scheme that does not favor any of the methods considered. This rejection rate can be reduced by sampling so that the volatility of returns σX^{β} belongs to the interval $[0.1, 0.5]$ with uniform probability. As this restriction on joint draws of (β, σ) was not imposed in our design we frequently drew low volatilities in cases where the option was far out of the money.

weights. Among weight estimators, which do not involve derivatives of the payoff function, the order is the one predicted by the convergence rates in Theorems 1 and 3.

The results in Theorem 2 show that MCC estimators have a lower convergence rate than MCCFD estimators. In light of this, our result that MCC estimators perform significantly better than MCCFD estimators for hedging digital options, may seem counterintuitive. To understand this, it is important to remember that performance depends on both the convergence speed and the convergence constant. The finding that MCC estimators outperform MCFD estimators even though their convergence rate is lower, suggests that the convergence constant for MCFD estimators is considerably larger. Why is this the case? The convergence constant of MCFD estimators is small if both the finite difference approximation of the derivative of the payoff and the finite difference approximation of the Malliavin derivative are accurate. In contrast, MCC estimators do not involve approximations of derivatives of the payoff function. Their convergence constant is determined by the error resulting from the Euler approximation of the instantaneous Malliavin weight. In the portfolio application of the previous section, the payoff function is smooth and MCFD far outperforms MCC. The reversed efficiency results for the Delta-hedge of the digital option provides evidence that it is the approximation of the degenerate derivative of the discontinuous payoff function that is responsible for the under-performance of MCFD.

Further evidence on this point can be inferred from the dominance of MCMD-M over MCMD-W. As shown in section 3.1.3, MCMD-M estimators are additive mixtures of MCMD-W and MCMD-P estimators. Localization ensures that the Malliavin weight in the MCMD-M estimator is concentrated around a compact interval containing the points of discontinuities of the payoff. Outside this interval, MCMD-M estimators have the same structure as MCMD-P estimators. Hence, as MCFD estimators are approximate MCMD-P estimators (see section 3.3), they also approximate MCMD-M estimators outside the interval containing the points of discontinuities. Inside this interval, MCMD-M estimators share the structure of MCMD-W estimators, and are therefore approximated by MCC estimators (see section 3.2). If the interval of discontinuity did not matter, the dominance of MCMD-M over MCMD-W would suggest that MCFD ought to dominate MCC. The reverse ordering of MCFD and MCC in the experiment indicates that the poor performance of MCFD is due to an inaccurate finite difference approximation of the degenerate derivative of the payoff.

As the derivative of the digital payoff function is degenerate at the strike, an MCMD-P estimate of the option's Delta is null. The localization in MCMD-M can be viewed as an attempt to derive an estimator that is "close" in structure to MCMD-P, but does not rely on finite difference approximations of the degenerate derivative. In fact, the MCMD-P component of an MCMD-M estimator receives more weight if the length of the interval containing the points of discontinuity is small enough. From this perspective, MCMD-M estimators are "close" to MCMD-P estimators, and our finding that MCMD-M estimators outperform MCMD-W estimators for discontinuous functions, is consistent with the results of Broadie and Glasserman (1996) showing that MCMD-P estimators dominate MCMD-W estimators, in the variance sense, for continuous payoff functions.

[Insert Figure 2 about here]

MCC estimators are less efficient than MCMD estimators because, as shown in section 3, they are based on a Euler approximation of an instantaneous Malliavin weight. In contrast, MCMD-M and MCMD-W estimators use exact Malliavin weights and choices of the parameter α leading to smoothed time averages of instantaneous Malliavin weights. It is important to remark that MCMD-M and MCMD-W are more efficient than MCC estimators even though for a fixed number of replications and discretization points, they take almost twice as long to calculate. This further highlights the poor performance of approximate MCC-weights relative to exact MCMD-weights. Our results also suggest that choosing $s = T$ within the class of α's of the form $\mathbf{1}_{[t,s]}(\cdot)/(s-t)$, where $s \in [t, T]$, is in fact optimal. Resulting Malliavin-weights are smoothed time averages of instantaneous Malliavin-weights over the full remaining life of the option and therefore have a lower variance than instantaneous Malliavin-weights obtained with $s = t$.

5 Conclusion

This paper provides explicit formulas for asymptotic distributions associated with various Monte Carlo estimators of derivatives of functions of diffusions. Estimators based on Malliavin derivatives were shown to be the only ones that preserve the optimal convergence rate for Monte Carlo schemes. Alternative schemes, such as those based on the covariation with the underlying Wiener process or on finite differences, converge at slower rates. They also suffer from additional second order biases. Our explicit expressions for the second order biases can be used to assess the size distortion of confidence intervals that ignore the second order bias. Second order bias correction was found to be easier to implement for estimators based on Malliavin derivatives. Given that asymptotic confidence intervals are only valid when the second order bias is taken into account this advantage of MCMD is an additional argument in favor of its use for estimating derivatives in diffusion settings.

6 Appendix

6.1 Appendix A: Expected Approximation Errors

The proofs of our results, in Appendix B, rely on a theorem of Detemple, Garcia and Rindisbacher (2004) that characterizes the asymptotic expected approximation error. To state this result we need to introduce the following notation. Recall that ∂A, ∂B_j are $d \times d$ matrices of Jacobians. The tangent process, in this multivariate setup, solves the linear equation

$$d\nabla_{t,x}X_s = \left(\partial A(X_s)ds + \sum_{j=1}^{d}\partial B_j(X_s)dW_s^j\right)\nabla_{t,x}X_s \text{ with } \nabla_{t,x}X_t = I_n. \tag{38}$$

With this notation we have,

Theorem 5: *Let $g \in \mathcal{C}^3(\mathbb{R}^d)$ be such that the uniform integrability condition*

$$\lim_{r\to\infty}\limsup_{N}\mathbf{E}_t\left[\mathbf{1}_{\{\|N(g(X_T^N)-g(X_T))\|>r\}}N|g(X_T^N)-g(X_T)|\right] = 0 \tag{39}$$

holds ($\mathbf{P}$-a.s.). Then, as $N \to \infty$,

$$N\mathbf{E}_t\left[g(X_T^N)-g(X_T)\right] \to \frac{1}{2}K_{t,T}(x) \equiv \frac{1}{2}\mathbf{E}_t\left[\partial g(X_T)V_1(t,T)+V_2(t,T)\right] \tag{40}$$

where the random variables $V_1(t,T)$ and $V_2(t,T)$ are

$$\begin{aligned} V_1(t,T) \;=\; & -\nabla_{t,x}X_T\int_t^T(\nabla_{t,x}X_s)^{-1}\left(\partial A(X_s)dX_s + \sum_{j=1}^{d}[\partial B_jA - \sum_{i=1}^{d}(\partial B_j)(\partial B_j)B_i](X_s)dW_s^j\right) \\ & +\nabla_{t,x}X_T\int_t^T(\nabla_{t,x}X_s)^{-1}\sum_{j=1}^{d}\left([\partial B_j\partial B_jA](X_s) - \int_0^T\sum_{k,l=1}^{d}[\partial_k(\partial_lAB_{l,j})B_{k,j}](X_s)\right)ds \\ & +\nabla_{t,x}X_T\int_t^T(\nabla_{t,x}X_s)^{-1}\sum_{i,j=1}^{d}\left([\partial[\partial B_j\partial B_jB_i]B_i - \partial B_i\partial B_j\partial B_jB_i](X_s)\right)ds \end{aligned}$$

$$V_2(t,T) = -\int_t^T\sum_{i,j=1}^{d}\nu_{i,j}(s,T)ds.$$

where $\nu_{i,j}(s,T) \equiv [h^{i,j}(\nabla_{t,x}X_\cdot)^{-1}[(\partial B_j)B_i](X),W^i]_s$ with $h_t^{i,j} \equiv \mathbf{E}_t\left[\mathcal{D}_{jt}\left(\partial g(X_T)\nabla_{t,x}X_Te_i\right)\right]$. An explicit expression for $\nu_{i,j}(s,T)$ is in Detemple, Garcia and Rindisbacher (2004).

In order to derive the asymptotic expected approximation error, on the right hand side of (40), an expansion, based on the mean value theorem, of the error $g(X_T^N)-g(X_T)$ is used. This gives a linear SDE with four autonomous components. Three of these autonomous terms converge at the rate $1/N$. The last one converges at the rate $1/\sqrt{N}$, but has zero expectation. This difference in the convergence behavior of the autonomous terms in the error expansion explains the presence of two parts in the second order bias $K_{t,T}(x)$. The limits of the first three terms give rise to $V_1(t,T)$. The fourth term leads to $V_2(t,T)$.

The uniform integrability assumption in the theorem enables us to find the asymptotic expected approximation error by taking the expectation of the weak limit of the error expansion.

6.2 Appendix B: Proofs

In what follows we set $t = 0$, without loss of generality, and write $\mathbf{E}_0$ for $\mathbf{E}_{0,x}$, $\mathbf{P}_0$ for $\mathbf{P}_{0,x}$, ∇_x for $\nabla_{0,x}$, and $K_T(x)$ for $K_{0,T}(x)$.

Proof of Theorem 2: Recall that $f(0,x) \equiv \mathbf{E}_0[g(X_T)]$ and that the MCC estimator $\widetilde{\partial_x f(0,x)}^{M,N,\tau}$ is given in (26). The error of the MCC estimator, $\widetilde{\varepsilon}(0,x;M,N,\tau) \equiv \widetilde{\partial_x f(0,x)}^{M,N,\tau} - \partial_x f(0,x)$, can be expanded as follows

$$\begin{aligned}
\widetilde{\varepsilon}(0,x;M,N,\tau) &= \mathbf{E}_0\left[g(X_T)\frac{\Delta_\tau W_0}{\tau}\right]C(x) - \partial_x f(0,x) \\
&\quad + \frac{1}{M}\sum_{i=1}^{M}\left(g(X_T^{i,N}) - g(X_T^i)\right)\frac{\Delta_\tau W_0^i}{\tau}C(x) \\
&\quad + \frac{1}{M}\sum_{i=1}^{M}\left(g(X_T^i)\frac{\Delta_\tau W_0^i}{\tau} - \mathbf{E}_0\left[g(X_T)\frac{\Delta_\tau W_0}{\tau}\right]\right)C(x)
\end{aligned}$$

where $\Delta_\tau W_0 = W_\tau - W_0$ and C is the generalized inverse of B defined in (12). In this expansion, $X_T^{i,N}$ is an independent replication of the Euler discretization of the diffusion given in (4), X_T^i (resp. W_T^i) is an independent replication of the terminal value of the diffusion (resp. of the Brownian motion driving the SDE) and $\Delta_\tau W_0^i$ is an independent replication of the increment in the Brownian motion.

Our next three lemmas give the asymptotic errors for the three terms in this expansion.

Lemma 1: *Under the assumption of Theorem 2 we have*

$$\frac{1}{\tau}\left(\mathbf{E}_0\left[g(X_T)\frac{\Delta_\tau W_0}{\tau}\right]C(x) - \partial_x f(0,x)\right) \to \frac{1}{2}\left(\partial_v \mathbf{E}_0\left[\partial_x f(v,X_v)B(X_v)\right]\right)_{|v=0}C(x)$$

as $1/\tau \to \infty$.

Proof of Lemma 1: The Clark-Ocone formula $g(X_T) = \mathbf{E}_0\left[g(X_T)\right] + \int_0^T \mathbf{E}_v[\mathcal{D}_v g(X_T)]dW_v$ gives

$$\frac{1}{\tau}\left(\mathbf{E}_0\left[g(X_T)\frac{\Delta_\tau W_0}{\tau}\right]C(x) - \partial_x f(0,x)\right) = \frac{1}{\tau^2}\int_0^\tau \beta(v)dv$$

with $\beta(v) \equiv \mathbf{E}_0\left[\mathbf{E}_v[\mathcal{D}_v g(X_T)]C(x) - \partial_x f(0,x)\right]$. Given that $\beta(0) = 0$, a second order Taylor approximation of $\beta(v)$ yields $\lim_{1/\tau\to\infty}\frac{1}{\tau^2}\int_0^\tau \beta(v)dv = \frac{1}{2}\partial_v\beta(v)_{|v=0}$. But

$$\begin{aligned}
\partial_v\beta(v)_{|v=0} &= \left(\partial_v \mathbf{E}_0\left[\mathbf{E}_v[\mathcal{D}_v g(X_T)]C(x) - \partial_x f(0,x)\right]\right)_{|v=0} \\
&= \left(\partial_v \mathbf{E}_0\left[\partial_x f(v,X_v)B(X_v)\right]\right)_{|v=0}C(x)
\end{aligned}$$

where the last equality uses $\mathbf{E}_v[\mathcal{D}_v g(X_T)] = \partial_x f(v, X_v)B(X_v)$. Substituting the expression for $\partial_v \beta(v)_{|v=0}$ in the limit above establishes the result in the lemma. ■

Lemma 2: *Let $K_T(x)$ be given by equation (40) in Theorem 5. Under the assumption of Theorem 2 we have, as $M \to \infty$,*

$$N_M \frac{1}{M} \sum_{i=1}^{M} \left(g(X_T^{i,N_M}) - g(X_T^i) \right) \frac{\Delta_{\tau_M} W_0^i}{\tau_M} \to \frac{1}{2} \partial K_T(x) B(x)$$

in probability, when $N_M, 1/\tau_M \to \infty$ as $M \to \infty$.

Proof of Lemma 2: We proceed in two steps. In the first step we find a sequential limit. In the second step we establish that joint limits and sequential limits are the same. Define,

$$e_T^{M,N,\tau} \equiv \frac{1}{M} \sum_{i=1}^{M} N \left(g(X_T^{i,N_M}) - g(X_T^i) \right) \frac{\Delta_\tau W_0^i}{\tau} - \frac{1}{2} \partial K_T(x) B(x),$$

and note that $e_T^{M,N,\tau} = e_{1,T}^{M,N,\tau} + e_{2,T}^{N,\tau} + e_{3,T}^{N}$, with

$$e_{1,T}^{M,N,\tau} \equiv \frac{1}{M} \sum_{i=1}^{M} N \left(g(X_T^{i,N}) - g(X_T^i) \right) \frac{\Delta_\tau W_0^i}{\tau} - \mathbf{E}_0 \left[N \left(g(X_T^N) - g(X_T) \right) \frac{\Delta_\tau W_0}{\tau} \right]$$

$$e_{2,T}^{N,\tau} \equiv \mathbf{E}_0 \left[\frac{1}{\tau} \int_0^\tau [\mathcal{D}_v - \mathcal{D}_0] N \left(g(X_T^N) - g(X_T) \right) dv \right]$$

$$e_{3,T}^{N} \equiv \mathcal{D}_0 N \mathbf{E}_0 \left[g(X_T^N) - g(X_T) \right] - \frac{1}{2} \partial K_T(x) B(x).$$

The expression for $e_{2,T}^{N,\tau}$ is obtained by using the Clark-Ocone formula

$$N \left(g(X_T^N) - g(X_T) \right) = \mathbf{E}_0 \left[N \left(g(X_T^N) - g(X_T) \right) \right] + \int_0^T \mathbf{E}_v \left[\mathcal{D}_v N \left(g(X_T^N) - g(X_T) \right) \right] dW_v$$

to conclude

$$\mathbf{E}_0 \left[N \left(g(X_T^N) - g(X_T) \right) \frac{\Delta_\tau W_0}{\tau} \right] = \mathbf{E}_0 \left[\frac{1}{\tau} \int_0^\tau \mathbf{E}_v \left[\mathcal{D}_v N \left(g(X_T^N) - g(X_T) \right) \right] dv \right].$$

By the weak law of large numbers for i.i.d random sequences the first term vanishes in probability for fixed N and τ, i.e. $\mathbf{P}-\lim_{M\to\infty} e_{1,T}^{M,N,\tau} = 0$ for all N, τ. By the continuous differentiability of the Riemann integral $\lim_{\tau\to\infty} e_{2,T}^{N,\tau} = 0$ for fixed N. Finally, by Theorem 5 that shows $N\mathbf{E}_0[g(X_T^N) - g(X_T)] \to \frac{1}{2} K_T(x)$ and by the chain rule of Malliavin that gives $\mathcal{D}_0 K_T(x) = \partial K_T(x) B(x)$ we obtain $\lim_{N\to\infty} e_{3,T}^{N} = 0$. We conclude that $e_T^{M,N,\tau} \to 0$ in probability, if sequentially $M, 1/\tau, N \to \infty$.

Next, we show that this sequential limit corresponds to the joint limit (i.e. when $M, N, 1/\tau \to \infty$ jointly) and thus to the limit along any "diagonal". Invoking arguments in the proof of Theorem 4.2

in Billingsley (1968) (see also Lemma 6 in Phillips and Moon (1996)), it is necessary and sufficient to show that

$$\limsup_{M,N,1/\tau\to\infty} \mathbf{P}_0\left(\left\|e_{1,T}^{M,N,\tau}\right\| > \epsilon\right) = 0 \tag{41}$$

$$\limsup_{N,1/\tau\to\infty} \left\|e_{2,T}^{N,\tau}\right\| = 0. \tag{42}$$

Condition (41) is proved by applying the Markov-type inequality,[19]

$$\mathbf{P}_0\left(\frac{1}{M}\left\|\sum_{i=1}^{M} H_T^{i,N,\tau}\right\| > \epsilon\right) \leq \frac{\mathbf{E}_0\left[\|H_T^{N,\tau}\|\mathbf{1}_{\{\|H_T^{N,\tau}\|>M^\gamma\epsilon\}}\right]}{\epsilon M} + M^{\gamma-1}\mathbf{P}_0\left(\|H_T^{N,\tau}\| \leq \epsilon M^\gamma\right) \tag{43}$$

for $\epsilon > 0$ and $\gamma \in]0,1[$ to the random variable

$$H_T^{i,N,\tau} \equiv N\left(g(X_T^{i,N}) - g(X_T^i)\right)\frac{\Delta_\tau W_0^i}{\tau} - \mathbf{E}_0\left[N\left(g(X_T^N) - g(X_T)\right)\frac{\Delta_\tau W_0}{\tau}\right].$$

Let $M, N, 1/\tau \to \infty$. The uniform integrability condition (27) with $r = M^\gamma\epsilon$ establishes that the first term on the right hand side of (43) converges to zero for all $\epsilon > 0$. The second term also converges to zero as $M^{\gamma-1} \to 0$ when $M \to \infty$. We conclude that the left hand side converges to zero as $M, N, 1/\tau \to \infty$, thereby establishing (41).

For condition (42) note the inequality

$$\sup_{N\geq N_0, 1/\tau\geq 1/\tau_0} \|e_{2,T}^{N,\tau}\| \leq \bar{e}_{2,T}^{N_0,\tau_0} \equiv \sup_{N\geq N_0}\left\|\mathbf{E}_0\left[\sup_{v\in(0,\tau_0]} \mathcal{D}_v Z_T^N - \mathcal{D}_0 Z_T^N\right]\right\|,$$

where $\sup_{v\in(0,\tau_0]} \mathcal{D}_v Z_T^N - \mathcal{D}_0 Z_T^N$ is non-decreasing in τ_0 and null for $\tau_0 = 0$. Taking the infimum over $N_0 > 0, 1/\tau_0 > 0$ on both sides gives (42) as $\inf_{N_0>0,1/\tau_0>0} \bar{e}_{2,T}^{N_0,\tau_0} = 0$. ■

Lemma 3: *Under the assumption of Theorem 2 we have*

$$\sqrt{\tau_M}\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\left(g(X_T^i)\frac{\Delta_{\tau_M} W_0^i}{\tau_M} - \mathbf{E}_0\left[g(X_T^i)\frac{\Delta_{\tau_M} W_0}{\tau_M}\right]\right) \Rightarrow O_T(x) \tag{44}$$

where $1/\tau_M \to \infty$ as $M \to \infty$ and where $O_T(x) \sim N\left(0, \mathbf{E}_0[g(X_T)^2]\right)$.

[19] This inequality holds for a sequence of i.i.d. random vectors Z^i and any $\epsilon > 0$, as $\mathbf{E}\left[\|Z\|\mathbf{1}_{\|Z\|>\epsilon M^\gamma}\right] = \mathbf{E}\left[\|Z\|\right] - \mathbf{E}\left[\|Z\|\mathbf{1}_{\|Z\|\leq\epsilon M^\gamma}\right]$, by the triangle inequality and the Markov inequality (Kallenberg (1997), Lemma 3.1 page 40),

$$\begin{aligned}
\mathbf{E}\left[\|Z\|\mathbf{1}_{\{\|Z\|>\epsilon M^\gamma\}}\right] &\geq \mathbf{E}\left[\|Z\|\right] - \epsilon M^\gamma \mathbf{E}\left[\mathbf{1}_{\{\|Z\|\leq\epsilon M^\gamma\}}\right] \geq \mathbf{E}\left[\left\|\frac{1}{M}\sum_{i=1}^{M} Z^i\right\|\right] - \epsilon M^\gamma \mathbf{P}\left(\|Z\| \leq \epsilon M^\gamma\right) \\
&\geq \mathbf{P}\left(\left\|\frac{1}{M}\sum_{i=1}^{M} Z^i\right\| > \epsilon\right)\epsilon M - \epsilon M^\gamma \mathbf{P}\left(\|Z\| \leq \epsilon M^\gamma\right).
\end{aligned}$$

Proof of Lemma 3: First note that Ito's lemma implies

$$(\Delta_\tau W_t)(\Delta_\tau W_t)' = \int_t^{t+\tau} dW_s (W_s - W_t)' + \int_t^{t+\tau} (W_s - W_t)(dW_s)' + \tau I_d,$$

while the Clark-Ocone formula gives

$$g(X_T)^2 = \mathbf{E}_0[g(X_T)^2] + \sum_{j=1}^d \int_0^T \mathbf{E}_v[\mathcal{D}_{jv} g(X_T)^2] dW_v^j$$

$$g(X_T) = \mathbf{E}_0[g(X_T)] + \sum_{j=1}^d \int_0^T \mathbf{E}_v[\mathcal{D}_{jv} g(X_T)] dW_v^j.$$

It follows that

$$\begin{aligned}
m_{2T}^\tau &\equiv \mathbf{E}_0\left[g(X_T)^2 (\Delta_\tau W_0)(\Delta_\tau W_0)'\right] \\
&= \mathbf{E}_0\left[\int_0^\tau \mathbf{E}_v\left[\mathcal{D}_v g(X_T)^2\right] dW_v \left(\int_0^\tau (W_s - W_0)(dW_s)'\right)\right] \\
&\quad + \mathbf{E}_0\left[\int_0^\tau \mathbf{E}_v\left[\mathcal{D}_v g(X_T)^2\right] dW_v \left(\int_0^\tau dW_s (W_s - W_0)'\right)\right] + \tau \mathbf{E}_0\left[g(X_T)^2\right] I_d \\
&= \mathbf{E}_0\left[\int_0^\tau (W_v - W_0)\mathbf{E}_v\left[\mathcal{D}_v g(X_T)^2\right] dv\right] \\
&\quad + \mathbf{E}_0\left[\int_0^\tau \left(\mathbf{E}_v\left[\mathcal{D}_v g(X_T)^2\right]\right)' (W_v - W_0)' dv\right] + \tau \mathbf{E}_0\left[g(X_T)^2\right] I_d,
\end{aligned}$$

and that $m_{1T}^\tau \equiv \mathbf{E}_0\left[g(X_T)\Delta_\tau W_0\right] = \mathbf{E}_0\left[\int_0^\tau \mathbf{E}_v\left[\mathcal{D}_v g(X_T)\right] dv\right]$. From these expressions we obtain the conditional variance at $t = 0$,

$$\begin{aligned}
\mathbf{VAR}_0\left[g(X_T)\frac{\Delta_\tau W_0}{\sqrt{\tau}}\right] &= \frac{1}{\tau} m_{2T}^\tau - \tau\left(\frac{1}{\tau} m_{1T}^\tau\right)\left(\frac{1}{\tau} m_{1T}^\tau\right)' \\
&= \frac{1}{\tau}\mathbf{E}_0\left[\int_0^\tau (W_v - W_0)\mathbf{E}_v\left[\mathcal{D}_v g(X_T)^2\right] dv\right] \\
&\quad + \frac{1}{\tau}\mathbf{E}_0\left[\int_0^\tau \mathbf{E}_v\left(\left[\mathcal{D}_v g(X_T)^2\right]\right)' (W_v - W_0)' dv\right] + \mathbf{E}_0\left[g(X_T)^2\right] I_d \\
&\quad - \tau\left(\frac{1}{\tau}\mathbf{E}_0\left[\int_0^\tau \mathbf{E}_v\left[\mathcal{D}_v g(X_T)\right] dv\right]\right)\left(\frac{1}{\tau}\mathbf{E}_0\left[\int_0^\tau \mathbf{E}_v\left[\mathcal{D}_v g(X_T)\right] dv\right]\right)'.
\end{aligned}$$

Given that terms one, two and four converge to zero as $1/\tau \to \infty$ the limit

$$\lim_{1/\tau \to \infty} \mathbf{VAR}_0\left[g(X_T)\frac{\Delta_\tau W_0}{\sqrt{\tau}}\right] = \mathbf{E}_0\left[g(X_T)^2\right] I_d$$

holds. The uniform integrability condition (28), which is sufficient for the Lindeberg Central Limit theorem for independent random variables to hold, can now be invoked to conclude that the weak limit is Gaussian with variance $\mathbf{E}_0\left[g(X_T)^2\right] I_d$. ■

Proof of Theorem 2 (continued): Combining Lemmas 1, 2 and 3, shows that[20]

$$\widetilde{\partial_x f(0,x)}^{M,N_M,\tau_M} - \partial_x f(0,x) = \mathbf{O_P}\left(\frac{1}{N_M} + \frac{1}{\sqrt{M\tau_M}} + \left(\frac{1}{\tau_M}\right)^{-1}\right).$$

With the selection $1/\tau_M = M^\delta/\varepsilon_1^c$ and $N_M = M^\gamma/\varepsilon_2^c$ for some $\varepsilon_1^c, \varepsilon_2^c, \delta, \gamma > 0$, it is seen that the efficient scheme is obtained for $(\gamma, \delta) = \inf\left(\arg\max_{\gamma,\delta} \min\{\delta, (1-\delta)/2, \gamma\}\right) = (1/3, 1/3)$. This proves the theorem. ■

Proof of Theorem 3: The error $\widehat{\varepsilon}_j\left(0, x; M, N, \tau_j, \alpha_j\right) \equiv \widehat{\partial_{x_j} f(0,x)}^{M,N,\tau_j} - \partial_{x_j} f(0,x)$ of the MCFD estimators $\widehat{\partial_{x_j} f(0,x)}^{M,N,\tau_j}$ given in (30), can be expanded for arbitrary $\alpha_j \in \mathbb{R}$, as:

$$\widehat{\varepsilon}_j\left(0, x; M, N, \tau_j, \alpha_j\right) = R_{1j}(0, x; \alpha_j, \tau_j) + R_{2j}(0, x; M, N, \tau_j, \alpha_j) + R_{3j}(0, x; M, \tau_j, \alpha_j)$$

with

$$R_{1j}\left(0, x; \tau_j, \alpha_j\right) \equiv \frac{\mathbf{E}_0\left[g\left(X_T(x + \alpha_j \tau_j e_j)\right) - g\left(X_T(x - (1-\alpha_j)\tau_j e_j)\right)\right]}{\tau_j} - \partial_{x_j} f(0,x) \tag{45}$$

$$\begin{aligned} R_{2j}\left(0, x; M, N, \tau_j, \alpha_j\right) \equiv\ & \frac{1}{M}\sum_{i=1}^{M}\left(\frac{g\left(X_T^{i,N}(x + \alpha_j \tau_j e_j)\right) - g\left(X_T^{i,N}(x - (1-\alpha_j)\tau_j e_j)\right)}{\tau_j}\right) \\ & -\frac{1}{M}\sum_{i=1}^{M}\left(\frac{g\left(X_T^{i}(x + \alpha_j \tau_j e_j)\right) - g\left(X_T^{i}(x - (1-\alpha_j)\tau_j e_j)\right)}{\tau_j}\right) \end{aligned} \tag{46}$$

$$\begin{aligned} R_{3j}\left(0, x; M, \tau_j, \alpha_j\right) \equiv\ & \frac{1}{M}\sum_{i=1}^{M}\left(\frac{g\left(X_T^{i}(x + \alpha_j \tau_j e_j)\right) - g\left(X_T^{i}(x - (1-\alpha_j)\tau_j e_j)\right)}{\tau_j}\right) \\ & -\mathbf{E}_0\left[\frac{g\left(X_T(x + \alpha_j \tau_j e_j)\right) - g\left(X_T(x - (1-\alpha_j)\tau_j e_j)\right)}{\tau_j}\right]. \end{aligned} \tag{47}$$

In this expansion, $e_j = [0, ..., 1, ..., 0]'$ is the j^{th} unit vector of dimension d, $X_T^{i,N}(x)$ is an independent replication of the Euler discretization of the SDE based on N points started at $X_0^i = x$, and $X_T^i(x)$ is an independent replication of the continuous diffusion with initial value $X_0^i = x$.

Our next three lemmas provide the asymptotic errors for R_{1j}, R_{2j} and R_{3j}.

[20] The symbol $\mathbf{O_P}(x)$ stands for "*at most of order x in probability*". A sequence of random variables Z^N is $\mathbf{O_P}(N^k)$, if for every $\epsilon > 0$, there exists a real number r such that $\mathbf{P}(N^{-k}|Z^N| > r) < \epsilon$ for all N. If $N^{-k}|Z^N| \le r$ for all N pointwise, the sequence Z^N is said to be *at most of order* N^k. In this case we write $\mathbf{O}(N^k)$. In contrast, a sequence Z^N is *of smaller order than* N^k, denoted by $\mathbf{o}(N^k)$, if $N^{-k}|Z^N| \to 0$ as $N \to \infty$.

Lemma 4: *Let $R_{1j}(0,x;\tau_j,\alpha_j)$ be defined by (45) and suppose that $f \in \mathcal{C}^3(\mathbb{R}^d)$. Then*

$$\begin{cases} \lim_{1/\tau_j\to\infty} \frac{1}{\tau_j^2} R_{1j}(0,x;\tau_j,\alpha_j) = \frac{1}{24}\partial^3_{x_j} f(0,x) & \text{if } \alpha_j = 1/2 \\ \\ \lim_{1/\tau_j\to\infty} \frac{1}{\tau_j} R_{1j}(0,x;\tau_j,\alpha_j) = \frac{2\alpha_j - 1}{2}\partial^2_{x_j} f(0,x) & \text{if } \alpha_j \neq 1/2. \end{cases}$$

Proof of Lemma 4: Given the differentiability assumption on f we can use Taylor series expansions for $f(0, x + \alpha_j\tau_j e_j)$ and $f(0, x - (1-\alpha_j)\tau_j e_j)$ around $f(0,x)$ to write (see footnote 20 for a definition of $\mathbf{o}(\cdot)$)

$$R_{1j}(0,x;\tau_j,\alpha_j) = \frac{1}{2}\left(\alpha_j^2 - (1-\alpha_j)^2\right)\partial^2_{x_j} f(0,x)\tau_j + \frac{1}{6}\left(\alpha_j^3 + (1-\alpha_j)^3\right)\partial^3_{x_j} f(0,x)\tau_j^2 + \mathbf{o}\left(\left(\frac{1}{\tau_j}\right)^{-2}\right).$$

If $\alpha_j = 1/2$ the first term vanishes and $R_{1j}(0,x;\tau_j,\alpha_j)$ is of order $(1/\tau_j)^{-2}$. If $\alpha_j \neq 1/2$ the second term in the expansion is asymptotically negligible and $R_{1j}(0,x;\tau_j,\alpha_j)$ is of order $(1/\tau_j)^{-1}$. The limits announced are obtained from this Taylor expansion. ■

Lemma 5: *Let $K_T(x)$ be defined by (40), $R_{2j}(0,x;M,N,\tau_j,\alpha_j)$ by (46) and select $N_M, \tau_{j,M}$ such that $N_M, 1/\tau_{j,M} \to \infty$ when $M \to \infty$. Suppose that the assumptions of Theorem 3 hold. Then, for all $\alpha_j \in [0,1]$ we have $N_M R_{2j}(0,x;M,N_M,\tau_{j,M},\alpha_j) \to \frac{1}{2}\partial K_T(x)$ in probability as $M \to \infty$.*

Proof of Lemma 5: We proceed in two steps as in the proof of Lemma 2. Under the assumptions of Theorem 3 we can apply the Clark-Ocone formula to find $R_{2j}(0,x;M,N,\tau_j) = R_{21j}(0,x;M,N,\tau_j) - R_{22j}(0,x;M,N,\tau_j)$ with

$$\begin{aligned} R_{21j}(0,x;M,N,\tau_j,\alpha_j) &\equiv \frac{1}{M}\sum_{i=1}^{M} \frac{g\left(X_T^{i,N}(x+\alpha_j\tau_j e_j)\right) - g\left(X_T^i(x+\alpha_j\tau_j e_j)\right)}{\tau_j} \\ R_{22j}(0,x;M,N,\tau_j,\tau_j) &\equiv \frac{1}{M}\sum_{i=1}^{M} \frac{g\left(X_T^{i,N}(x-(1-\alpha_j)\tau_j e_j)\right) - g\left(X_T^i(x-(1-\alpha_j)\tau_j e_j)\right)}{\tau_j}. \end{aligned}$$

The law of large numbers and the arguments in the proof of Theorem 5 in Detemple, Garcia and Rindisbacher (2004) imply

$$\begin{aligned} NR_{21j}(0,x;M,N,\tau_j,\alpha_j) &\to \frac{1}{2}\frac{K_T(x+\alpha_j\tau_j e_j)}{\tau_j}, \\ NR_{22j}(0,x;M,N,\tau_j,\alpha_j) &\to \frac{1}{2}\frac{K_T(x-(1-\alpha_j)\tau_j e_j)}{\tau_j} \end{aligned}$$

in probability, as $M, N \to \infty$ sequentially. If next $1/\tau_j \to \infty$, we obtain $NR_{2j}(0,x;M,N,\tau_j,\alpha_j) \to \frac{1}{2}\partial_j K_T(x)$. This establishes that $NR_{2j}(0,x;M,N,\tau_j,\alpha_j) \to \frac{1}{2}\partial_j K_T(x)$ in probability as $M, N, 1/\tau_j \to \infty$ sequentially.

The remainder of the proof parallels the proof of Lemma 2. To show that the same limit holds if $M, N, 1/\tau^j \to \infty$ jointly, and therefore along any "diagonal", it is sufficient to show that

$$\limsup_{M,N,1/\tau_j\to\infty} \mathbf{P}_0\left(\left|\frac{1}{M}\sum_{i=1}^{M} N\nabla_{x_j}^{\tau_j,\alpha_j} g\left(X_T^{i,N}(x)\right) - \mathbf{E}_0\left[N\nabla_{x_j}^{\tau_j,\alpha_j} g\left(X_T^N(x)\right)\right]\right| > \epsilon\right) = 0 \tag{48}$$

for all $\epsilon > 0$ and

$$\limsup_{N,1/\tau_j\to\infty} \left|\mathbf{E}_0\left[N\nabla_{x_j}^{\tau_j,\alpha_j} g\left(X_T^N(x)\right)\right] - \partial_{x_j} N\mathbf{E}_0\left[g(X_T^N(x)) - g(X_T(x))\right]\right| = 0. \tag{49}$$

As in the proof of Lemma 2, (48) follows from the Markov-type inequality (43) and the uniform integrability condition (31). Similarly (49) is satisfied, because

$$\partial_{x_j} N\mathbf{E}_0\left[g(X_T^N(x)) - g(X_T(x))\right] = N\mathbf{E}_0\left[\nabla_{x_j}(g(X_T^N(x)) - g(X_T(x)))\right]$$

and under the assumptions of Theorem 5,

$$\limsup_{1/\tau_j,N\to\infty} \mathbf{E}_0\left[\nabla_{x_j}^{\tau_j,\alpha_j} N(g(X_T^N(x)) - g(X_T(x)))\right] = \limsup_{N\to\infty} \mathbf{E}_0\left[\nabla_{x_j} N(g(X_T^N(x)) - g(X_T(x)))\right].$$

We conclude that the joint limit corresponds to the sequential limit. ■

Lemma 6: *Let $R_{3j}(0, x; M, \tau_j)$ be defined by (47) and select $\tau_{j,M}$ such that $1/\tau_{j,M} \to \infty$ when $M \to \infty$. Suppose that the assumptions of Theorem 3 hold. For $j = 1, \ldots, d$ and for all $\alpha_j \in [0, 1]$ we have the limit, $\left(\sqrt{M} R_{3j}(0, x; M, \tau_{j,M})\right)_{j=1,\ldots,d} \Rightarrow Q_T$ as $M \to \infty$, where $Q_T \sim N\left(0, \mathbf{E}_0[\int_0^T L_v L_v' dv]\right)$ with $L_v' = \mathbf{E}_v\left[\mathcal{D}_v\left(\partial g(X_T)\mathcal{D}_v X_T C(X_v)\right)\right]$.*

Proof of Lemma 6: Let us first find the asymptotic variance of

$$H_T^{\tau_j} \equiv \frac{g\left(X_T(x + \alpha_j\tau_j e_j)\right) - g\left(X_T(x - (1 - \alpha_j)\tau_j e_j)\right)}{\tau_j}.$$

As $g(X_T(x)) \in \mathbb{D}^{1,2}$, we can apply the Clark-Ocone formula to obtain

$$H_T^{\tau_j} - \mathbf{E}_0[H_T^{\tau_j}] = \int_0^T \mathbf{E}_v[\mathcal{D}_v H_T^{\tau_j}] dW_v \equiv \int_0^T (L_v^{\tau_j})' dW_v$$

and therefore, $\mathbf{VAR}_0[H_T^{\tau_j}] = \mathbf{E}_0\left[[H^{\tau_j}, H^{\tau_j}]_T\right] = \int_0^T \mathbf{E}_0[(L_v^{\tau_j})' L_v^{\tau_j}] dv$.

Define $f(v, X_v(x)) \equiv \mathbf{E}_{v,X_v(x)}\left[g(X_T(X_v(x)))\right]$. Commutativity of the Malliavin derivative and the conditional expectation gives

$$\begin{aligned}
(L_v^{\tau_j})' &= \mathcal{D}_v \mathbf{E}_v[H_T^{\tau_j}] = \mathcal{D}_v\left(\frac{f(v, X_v(x + \alpha_j e_j\tau_j)) - f(v, X_v(x - (1 - \alpha_j)e_j\tau_j))}{\tau_j}\right) \\
&= \frac{1}{\tau_j}\partial_x f(v, X_v(x + \alpha_j e_j\tau_j)) B(X_v(x + \alpha_j e_j\tau_j)) \\
&\quad -\frac{1}{\tau_j}\partial_x f(v, X_v(x - (1 - \alpha_j)e_j\tau_j)) B(X_v(x - (1 - \alpha_j)e_j\tau_j)).
\end{aligned}$$

Using $X(x-(1-\alpha_j)e_j\tau_j) - X(x+\alpha_j e_j \tau_j) \to 0$ $\mathbf{P}_0$-a.s., as $\tau_j \to 0$ we obtain

$$(L_v^{\tau_j})' - \left(\frac{\partial_x f(v, X_v(x+\alpha_j e_j \tau_j)) - \partial_x f(v, X_v(x-(1-\alpha_j)e_j\tau_j))}{\tau_j}\right) B(X_v) \to 0$$

$\mathbf{P}_0$-a.s., as $\tau_j \to 0$, or, equivalently, $(L_v^{\tau_j})' - \partial^2_{x,x_j} f(v, X_v) B(X_v) \to 0$ $\mathbf{P}_0$-a.s., as $\tau_j \to 0$. From $\partial^2_{x,x_j} f(v, X_v) B(X_v) = \mathcal{D}_v \partial_{x_j} f(v, X_v)$, we conclude $(L_v^{\tau_j})' \to \mathcal{D}_v \partial_{x_j} f(v, X_v)$, $\mathbf{P}_0$-a.s. as $\tau_j \to 0$, and hence $[H^{\tau_j}, H^{\tau_j}]_T \to \int_0^T L'_{jv} L_{jv} dv$ $\mathbf{P}_0$-a.s., with $L'_{jv} = \mathcal{D}_v \partial_{x_j} f(v, X_v) = \mathbf{E}_v \left[\mathcal{D}_v \left(\partial g(X_T) \nabla_{x_j} X_T\right)\right]$. By assumption (32), $L_v^{\tau_j}$ is uniformly integrable so that almost sure convergence implies convergence of the mean. We conclude

$$\lim_{\tau_j \to 0} \mathbf{VAR}_0 \left[H_T^{\tau_j}\right] = \mathbf{E}_0 \left[\int_0^T L'_{jv} L_{jv} ds\right].$$

This analysis holds for all $j = 1, ..., d$. With $L_v = [L'_{jv}]_{j=1,\ldots,d}$ we obtain the asymptotic variance $\left(\mathbf{VAR}_0 \left[H_T^{\tau_j}\right]\right)_{j=1,\ldots,d} = \mathbf{E}_0 \left[\int_0^T L_v L'_v dv\right]$.

The asymptotic distribution follows from the Lindeberg Central Limit theorem for independent random variables. The uniform integrability condition (32) ensures that the Lindeberg condition for application of this theorem is satisfied. ■

Proof of Theorem 3 (continued): Given the results of Lemmas 4, 5 and 6, we see that

$$\widehat{\partial_x f(0,x)}^{M,N_M,\tau_{j,M}} - \partial_x f(0,x) = \begin{cases} \mathbf{O_P}\left(\frac{1}{N_M} + \frac{1}{\sqrt{M}} + \left(\frac{1}{\tau_{j,M}}\right)^{-2}\right) & \text{if } \alpha_j = 1/2 \\ \mathbf{O_P}\left(\frac{1}{N_M} + \frac{1}{\sqrt{M}} + \left(\frac{1}{\tau_{j,M}}\right)^{-1}\right) & \text{if } \alpha_j \neq 1/2. \end{cases}$$

Choose $1/\tau_{j,M} = M^{\delta_j}/\varepsilon^{\kappa}_{j1}$ and $N_M = M^{\gamma}/\varepsilon^{\kappa}_2$ for some $\varepsilon^{\kappa}_{j1}, \varepsilon^{\kappa}_2, \delta_j, \gamma > 0$ and $\kappa \in \{fcd, fd\}$. If $\alpha_j = 1/2$ the efficient scheme is obtained for $(\gamma, \delta_j) = \inf\left(\arg\max_{\gamma,\delta_j} \min\{2\delta_j, 1/2, \gamma\}\right) = (1/2, 1/4)$. For $\alpha_j \neq 1/2$, $(\gamma, \delta_j) = \inf\left(\arg\max_{\gamma,\delta_j} \min\{\delta_j, 1/2, \gamma\}\right) = (1/2, 1/2)$ provides the efficient scheme. This proves the theorem. ■

Proof of Theorem 4: The proof of Theorem 4 parallels the proof of Theorem 3. For the discontinuous functions under consideration, $\sum_{j=1}^{\infty} \gamma_j \mathbf{1}_{B_j}(x) \in \mathbb{D}^{1,2}$, Lemma 6 is replaced by:

Lemma 7: *Let $R_{3j}(0, x; M, \tau_j)$ be defined by (47) and select $\tau_{j,M}$ such that $1/\tau_{j,M} \to \infty$ when $M \to \infty$. Suppose that the assumptions of Theorem 4 hold. For $j = 1, \ldots, d$ and for all $\alpha_j \in [0, 1]$*

we have the limit, $\sqrt{M\tau_{j,M}}R_{3j}\left(0,x;M,\tau_{j,M}\right)\Rightarrow Q_T^j$ *as* $M\to\infty$*, where* $Q_T^j\sim N\left(0,V_0(x)\right)$ *with*

$$
\begin{aligned}
V_0(x) &= \sum_{k=1}^{\infty}\gamma_k^2\left(2\alpha_j-1\right)\partial_{x_j}\mathbf{P}_0\left(\{X_T(x)\in B_k\}\right)\\
&\quad -2\alpha_j\sum_{k,l=1}^{\infty}\gamma_k\gamma_l\partial_{x_j}\mathbf{P}_0\left(\{X_T(x)\in B_k\}\cap\left\{X_T(x')\in B_l\right\}\right)_{|x'=x}\\
&\quad +2(1-\alpha_j)\sum_{k,l=1}^{\infty}\gamma_k\gamma_l\partial_{x_j'}\mathbf{P}_0\left(\{X_T(x)\in B_k\}\cap\left\{X_T(x')\in B_l\right\}\right)_{|x'=x},
\end{aligned}
$$

and where Q_T^j *and* Q_T^k *are mutually independent.*

Proof of Lemma 7: We first calculate the asymptotic variance of $g(x)=g^c(x)+g^d(x)$ where $g^d(x)=\sum_{j=1}^{\infty}\gamma_j\mathbf{1}_{B_j}(X_T)$. Let $H_T^{\tau_j}\equiv H_T^{c,\tau_j}+H_T^{d,\tau_j}$ with

$$
\begin{aligned}
H_T^{c,\tau_j} &\equiv \frac{g^c\left(X_T(x+\alpha_j\tau_je_j)\right)-g^c\left(X_T(x-(1-\alpha_j)\tau_je_j)\right)}{\sqrt{\tau_j}}\\
H_T^{d,\tau_j} &\equiv \frac{g^d\left(X_T(x+\alpha_j\tau_je_j)\right)-g^d\left(X_T(x-(1-\alpha_j)\tau_je_j)\right)}{\sqrt{\tau_j}}.
\end{aligned}
$$

Using $g^c(x)g^d(x)=0$, we obtain $\mathbf{VAR}_0\left[H_T^{\tau_j}\right]=\mathbf{VAR}_0\left[H_T^{c,\tau_j}\right]+\mathbf{VAR}_0\left[H_T^{d,\tau_j}\right]$.

From the assumption $g^c(X_T)\in\mathbb{D}^{1,2}$ and Lemma 6, we get $\mathbf{VAR}_0\left[H_T^{c,\tau_j}\right]=\mathbf{O}((1/\tau_j)^{-1/2})$ when $1/\tau_j\to\infty$, where $\mathbf{O}(\cdot)$ is defined in footnote 20

Next, we show that $\lim_{\tau_i,\tau_j\to 0}\left(\mathbf{COV}_0[H_T^{d,\tau_i},H_T^{d,\tau_j}]-\mathbf{E}_0\left[H_T^{d,\tau_i}H_T^{d,\tau_j}\right]\right)=0$ for all i,j. As

$$
\mathbf{COV}_0[H_T^{d,\tau_i},H_T^{d,\tau_j}]=\mathbf{E}_0\left[H_T^{d,\tau_i}H_T^{d,\tau_j}\right]-\mathbf{E}_0\left[H_T^{d,\tau_i}\right]\mathbf{E}_0\left[H_T^{d,\tau_j}\right],
$$

and $\frac{1}{\sqrt{\tau_j}}\mathbf{E}_0[H_T^{c,\tau_j}]\to\partial_{x_j}f(0,x)$ we have $\mathbf{E}_0[H_T^{d,\tau_i}]\mathbf{E}_0[H_T^{d,\tau_j}]=\mathbf{O}\left(((1/\tau_i)(1/\tau_j))^{-1/2}\right)$, and therefore $\mathbf{E}_0[H_T^{d,\tau_i}]\mathbf{E}_0[H_T^{d,\tau_j}]\to 0$, $\mathbf{P}_0-$a.s., as $1/\tau_i,1/\tau_j\to\infty$.

Hence,

$$
\left(\mathbf{VAR}_0[H_T^{d,\tau_j}]-\mathbf{E}_0\left[\left(H_T^{d,\tau_j}\right)^2\right]\right)\to 0,\mathbf{P}_0-a.s.,\ \text{as}\ \frac{1}{\tau_j}\to\infty \tag{50}
$$

$$
\left(\mathbf{COV}_0[H_T^{d,\tau_i},H_T^{d,\tau_j}]-\mathbf{E}_0\left[H_T^{d,\tau_i}H_T^{d,\tau_j}\right]\right)\to 0,\mathbf{P}_0-a.s.,\ \text{as}\ \frac{1}{\tau_i},\frac{1}{\tau_j}\to\infty. \tag{51}
$$

To derive the asymptotic limits of the variance and the covariance we therefore need the limits of $\mathbf{E}\left[\left(H_T^{d,\tau_j}\right)^2\right]$ and $\mathbf{E}\left[H_T^{d,\tau_i}H_T^{d,\tau_j}\right]$.

To find the limits of $\mathbf{E}_0\left[\left(H_T^{d,\tau_j}\right)^2\right]$ and $\mathbf{E}_0\left[H_T^{d,\tau_i}H_T^{d,\tau_j}\right]$, note that

$$\begin{aligned} H_T^{d,\tau_i}H_T^{d,\tau_j} &= \frac{1}{\sqrt{\tau_i\tau_j}}g^d\left(X_T(x+\alpha_i\tau_i e_i)\right)g^d\left(X_T(x+\alpha_j\tau_j e_j)\right) & (52)\\ &+\frac{1}{\sqrt{\tau_i\tau_j}}g^d\left(X_T(x-(1-\alpha_i)\tau_i e_i)\right)g^d\left(X_T(x-(1-\alpha_j)\tau_j e_j)\right)\\ &-\frac{1}{\sqrt{\tau_i\tau_j}}g^d\left(X_T(x+\alpha_i\tau_i e_i)\right)g^d\left(X_T(x-(1-\alpha_j)\tau_j e_j)\right)\\ &-\frac{1}{\sqrt{\tau_i\tau_j}}g^d\left(X_T(x-(1-\alpha_i)\tau_i e_i)\right)g^d\left(X_T(x+\alpha_j)\tau_j e_j)\right) & (53)\end{aligned}$$

where $g^d\left(X_T(x)\right)g^d\left(X_T(x')\right)\equiv\sum_{k,l=1}^{\infty}\gamma_k\gamma_l\mathbf{1}_{B_k}(X_T(x))\mathbf{1}_{B_l}(X_T(x'))$. With the definitions $p_k(x)\equiv\mathbf{P}_0(X_T(x)\in B_k)$ and $q_{kl}(x,x')\equiv\mathbf{P}_0\left(\{X_T(x)\in B_k\}\cap\{X_T(x')\in B_l\}\right)$ and using $p_k(x)=q_{kk}(x,x)$ and $q_{kl}(x,x)=0$ for $k\neq l$, we obtain

$$\begin{aligned} \mathbf{E}_0\left[\left(H_T^{d,\tau_j}\right)^2\right] &= \sum_{k=1}^{\infty}\gamma_k^2\left(\frac{p_k(x+\alpha_j e_j\tau_j)-p_k(x)}{\tau_j}+\frac{p_k(x-(1-\alpha_j)e_j\tau_j)-p_k(x)}{\tau_j}\right)\\ &-2\sum_{k=1}^{\infty}\gamma_k^2\left(\frac{q_{kk}(x+\alpha_j e_j\tau_j,x-(1-\alpha_j)e_j\tau_j)-q_{kk}(x,x)}{\tau_j}\right)\\ &-2\sum_{\substack{k,l=1\\k\neq l}}^{\infty}\gamma_k\gamma_l\left(\frac{q_{kl}(x+\alpha_j e_j\tau_j,x-(1-\alpha_j)e_j\tau_j)-q_{kl}(x,x)}{\tau_j}\right)\end{aligned}$$

and therefore with $V_0^{\tau_j}(x)\equiv\mathbf{VAR}_0[H_T^{d,\tau_j}]$,

$$V_0^{\tau_j}(x)\to\sum_{k=1}^{\infty}\gamma_k^2\left(2\alpha_j-1\right)\partial_{x_j}p_k(x)-2\sum_{k,l=1}^{\infty}\gamma_k\gamma_l\left(\alpha_j\partial_{x_j}q_{kl}(x,x')_{|x'=x}-(1-\alpha_j)\partial_{x'_j}q_{kl}(x,x')_{|x'=x}\right),$$

$\mathbf{P}_0$-a.s., as $1/\tau_j\to\infty$.

Next, we show that $\mathbf{E}_0[H_T^{d,\tau_i}H_T^{d,\tau_j}]\to 0$ $\mathbf{P}_0$-a.s. when $i\neq j$ and $1/\tau_i,1/\tau_j\to\infty$. Using (52) we obtain

$$\mathbf{E}_0[H_T^{d,\tau_i}H_T^{d,\tau_j}]=\sum_{k,l=1}^{\infty}\gamma_k\gamma_l h_{kl}(x,\alpha,\tau_i,\tau_j)$$

where

$$\begin{aligned} h_{kl}(x,\alpha,\tau_i,\alpha_j,\tau_j) &\equiv \frac{q_{kl}(x+\alpha_i e_i\tau_i,x+\alpha_j e_j\tau_j)-q_{kl}(x,x)}{\sqrt{\tau_i\tau_j}}\\ &+\frac{q_{kl}(x-(1-\alpha_i)e_i\tau_i,x-(1-\alpha_j)e_j\tau_j)-q_{kl}(x,x)}{\sqrt{\tau_i\tau_j}}\\ &-\frac{q_{kl}(x+\alpha_i e_i\tau_i,x-(1-\alpha_j)e_j\tau_j)-q_{kl}(x,x)}{\sqrt{\tau_i\tau_j}}\\ &-\frac{q_{kl}(x-(1-\alpha_i)e_i\tau_i,x+\alpha_j e_j\tau_j)-q_{kl}(x,x)}{\sqrt{\tau_i\tau_j}}.\end{aligned}$$

It follows that

$$\begin{aligned} h_{kl}(x,\alpha,\tau_i,\alpha_j,\tau_j) &= \alpha_i \partial_{x_i} q_{kl}(x,x)\sqrt{\frac{\tau_i}{\tau_j}} + \alpha_j \partial_{x'_j} q_{kl}(x,x')_{|x'=x}\sqrt{\frac{\tau_j}{\tau_i}} \\ &\quad -(1-\alpha_i)\partial_{x_i} q_{kl}(x,x)\sqrt{\frac{\tau_i}{\tau_j}} - (1-\alpha_j)\partial_{x'_j} q_{kl}(x,x')_{|x'=x}\sqrt{\frac{\tau_j}{\tau_i}} \\ &\quad -\alpha_i \partial_{x_i} q_{kl}(x,x)\sqrt{\frac{\tau_i}{\tau_j}} + (1-\alpha_j)\partial_{x'_j} q_{kl}(x,x')_{|x'=x}\sqrt{\frac{\tau_j}{\tau_i}} \\ &\quad +(1-\alpha_i)\partial_{x_i} q_{kl}(x,x)\sqrt{\frac{\tau_i}{\tau_j}} - \alpha_j \partial_{x'_j} q_{kl}(x,x')_{|x'=x}\sqrt{\frac{\tau_j}{\tau_i}} + \mathbf{o}(1) \\ &= \mathbf{o}(1). \end{aligned}$$

Hence $h_{kl}(x,\alpha,\tau_i,\tau_j) \to 0, \mathbf{P}_0$−a.s. as $1/\tau_i, 1/\tau_j \to \infty$. This shows that H_T^{d,τ_i} and H_T^{d,τ_j} are asymptotically uncorrelated, for $i \neq j$.

The asymptotic distribution follows from the Lindeberg Central Limit theorem for independent random variables: the uniform integrability condition (32) ensures that Lindeberg's condition is satisfied. ■

Proof of Theorem 4 (continued): The results of Lemmas 4, 5 and 7 give

$$\widehat{\partial_x f(0,x)}^{M,N_M,\tau_{j,M}} - \partial_x f(0,x) = \begin{cases} \mathbf{O_P}\left(\frac{1}{N_M} + \frac{1}{\sqrt{M\tau_{j,M}}} + \left(\frac{1}{\tau_{j,M}}\right)^{-2}\right) & \text{if } \alpha_j = 1/2 \\ \mathbf{O_P}\left(\frac{1}{N_M} + \frac{1}{\sqrt{M\tau_{j,M}}} + \left(\frac{1}{\tau_{j,M}}\right)^{-1}\right) & \text{if } \alpha_j \neq 1/2. \end{cases}$$

Choose $1/\tau_{j,M} = M^{\delta_j}/\varepsilon_{j1}^{\kappa}$ and $N_M = M^{\gamma}/\varepsilon_2^{\kappa}$ for some $\varepsilon_{j1}^{\kappa}, \varepsilon_2^{\kappa}, \delta_j, \gamma > 0$ and $\kappa \in \{fcd, fd\}$. For $\alpha_j = 1/2$, the efficient scheme is attained for $(\gamma,\delta_j) = \inf\left(\arg\max_{\delta_j,\gamma} \min\{2\delta_j, (1-\delta_j)/2, \gamma\}\right) = (2/5, 1/5)$. For $\alpha_j \neq 1/2$, $(\gamma,\delta_j) = \inf\left(\arg\max_{\gamma,\delta_j} \min\{\delta_j, (1-\delta_j)/2, \gamma\}\right) = (1/3, 1/3)$ provides the efficient scheme. This proves the theorem. ■

6.3 Appendix C: Abstract Integration by Parts

This Appendix shows how to use an integration by parts argument to handle a non-differentiable payoff function when the transition density is unknown.[21] Consider a diffusion with volatility coefficient $B \in \mathcal{C}^1(\mathbb{R}^n \times \mathbb{R}^d)$ such that $rank(B(x)) = n$ for all $x \in \mathbb{R}^n$ and define the adapted shift on the Wiener space, $\Omega = \mathcal{C}^0([0,T];\mathbb{R}^d)$,

$$\theta^{\lambda}(\omega)_v = \omega_v + \int_t^v \nu_{t,s}(\omega)\lambda ds.$$

[21] The presentation is based on Bismut's approach to Malliavin calculus (see Bichteler, Gravereaux and Jacod (1987) for a comparison of this approach with Malliavin's original approach).

This perturbation of the state space depends on a parameter $\lambda \in \mathbb{R}^n$ and a progressively measurable $\mathbb{R}^n \times \mathbb{R}^d$-valued matrix process $\nu_{t,s}$. To obtain our integration by parts formula we pass to a new measure $\mathbf{P}^{\lambda}_{t,x}$ under which $W_v + \int_t^v \nu_{t,s}\lambda ds$, $v \geq t$, is a standard Brownian motion process. The existence of this probability measure requires, by Girsanov's theorem,

$$\mathbf{E}_{t,x}\left[\mathcal{E}\left(-\int_t^v (\nu_{t,s}\lambda)' dW_s\right)\right] = 1, \qquad \text{for all } v \geq t.$$

Under this condition we can define $\frac{d\mathbf{P}^{\lambda}_{t,x}}{d\mathbf{P}_{t,x}} \equiv \mathcal{E}\left(-\int_t^T (\nu_{t,s}\lambda)' dW_s\right)$.

Taking derivatives with respect to the parameter λ on both sides of the equality

$$\mathbf{E}_{t,x}\left[\frac{d\mathbf{P}^{\lambda}_{t,x}}{d\mathbf{P}_{t,x}} h(X_T(\theta^{\lambda}))\right] = \mathbf{E}_{t,x}\left[h(X_T)\right] \tag{54}$$

and evaluating the resulting expressions at $\lambda = 0$, gives, for $h \in \mathcal{C}^1(\mathbb{R}^d)$

$$\mathbf{E}_{t,x}\left[\left(\partial_{\lambda}\frac{d\mathbf{P}^{\lambda}_{t,x}}{d\mathbf{P}_{t,x}}\right) h(X_T(\theta^{\lambda})) + \frac{d\mathbf{P}^{\lambda}_{t,x}}{d\mathbf{P}_{t,x}} \partial h(X_T(\theta^{\lambda}))\partial_{\lambda} X_T(\theta^{\lambda})\right]_{|\lambda=0} = 0.$$

Using $\partial_{\lambda}\frac{d\mathbf{P}^{\lambda}_{t,x}}{d\mathbf{P}_{t,x}} = \frac{d\mathbf{P}^{\lambda}_{t,x}}{d\mathbf{P}_{t,x}}\partial_{\lambda}\left(\int_t^T (-\nu_{t,s}\lambda)' dW_s - \frac{1}{2}\int_t^T \|\nu_{t,s}\lambda\|^2 ds\right)$ and $\frac{d\mathbf{P}^{\lambda}_{t,x}}{d\mathbf{P}_{t,x}}_{|\lambda=0} = 1$ gives

$$\frac{d\mathbf{P}^{\lambda}_{t,x}}{d\mathbf{P}_{t,x}}_{|\lambda=0} = -\int_t^T dW_s' \nu_{t,s}.$$

Substituting in (54) yields the abstract integration by parts formula,

$$\mathbf{E}_{t,x}\left[\left(\int_t^T (dW_s)'\nu_{t,s}\right) h(X_T(\theta^{\lambda}))_{|\lambda=0}\right] = \mathbf{E}_{t,x}\left[\partial h(X_T)\left(\partial_{\lambda} X_T(\theta^{\lambda})\right)_{|\lambda=0}\right]. \tag{55}$$

To identify $\left(\partial_{\lambda} X_T(\theta^{\lambda})\right)_{|\lambda=0}$ differentiate the integral representation

$$X_T(\theta^{\lambda}) = x + \int_0^T A\left(X_s(\theta^{\lambda})\right) ds + \sum_{j=1}^d \int_0^T B_j\left(X_s(\theta^{\lambda})\right)\left(dW_s^j + \int_t^T \nu_{t,s}\lambda ds\right)$$

with respect to λ, to obtain

$$\left(\partial_{\lambda} X_T(\theta^{\lambda})\right)_{|\lambda=0} = \int_0^T \left(\partial A(X_s)ds + \sum_{j=1}^d \partial B_j(X_s) dW_s^j\right)\left(\partial_{\lambda} X_s(\theta^{\lambda})\right)_{|\lambda=0} + \int_0^T B(X_s)\nu_{t,s} ds.$$

The solution of this linear SDE is

$$\left(\partial_{\lambda} X_T(\theta^{\lambda})\right)_{|\lambda=0} = \nabla_{t,x} X_T(X_t) \int_t^T \left(\nabla_{t,x} X_s(X_t)\right)^{-1} B(X_s)\nu_{t,s} ds$$

where $\nabla_{t,x} X_T(X_t)$ is the solution of (38).[22] The process $\nabla_{t,x} X_T(X_t)$ is the tangent process.

For C defined in (12), selecting the process $\nu_{t,s}^{(\alpha)} \equiv C(X_s)\nabla_{t,x}X_s(X_t)\alpha_{t,s}$ for some progressively measurable process $\alpha_{t,\cdot}$ such that $\int_t^T \alpha_{t,s}ds = I_n$, and substituting in (55) gives (18), that is

$$\mathbf{E}_{t,x}\left[\left(\int_t^T (dW_s)' C(X_s)\nabla_{t,x}X_s(X_t)\alpha_{t,s}\right) h(X_T)\right] = \mathbf{E}_{t,x}\left[\partial h(X_T)\nabla_{t,x}X_T(X_t)\right].$$

To establish the same result when $h \in \mathbf{L}^2$ is not continuously differentiable we use the fact that any square integrable function can be approximated by a sequence of infinitely differentiable functions with compact support. This enables us to proceed by first using the approximating sequence and then passing to the limit. See Fournié et. al. (1999) for detailed arguments.

References

Bally, V., D. Talay. 1996. The law of the Euler scheme for stochastic differential equations (I) : convergence rate of the distribution function. *Probab. Theory Related Fields* 104(1) 43-60.

Bichteler, K., J. B. Gravereaux, J. Jacod. 1987. *Malliavin Calculus for Processes with Jumps.* Gordon Breach Publishing Group.

Broadie, M., P. Glasserman. 1996. Estimating Security Price Derivatives using Simulation. *Management Sci.* 42(2) 269-285.

Cox, J. 1975. Notes on Option Pricing I: Constant Elasticity of Variance Diffusions. Unpublished note, Stanford University, Graduate School of Business.

Cvitanic, J., L. Goukasian, F. Zapatero. 2002. Hedging with Monte Carlo Simulation. E. Kontoghiorghes, B. Rustem and S. Siokos eds. *Computational Methods in Decision-Making, Economics and Finance.* Kluwer Academic Publishers, 339-353.

Cvitanic, J., L. Goukasian, F. Zapatero. 2003. Monte Carlo Computation of Optimal Portfolios in Complete Markets. *J. Econ. Dynam. Control* 27(6) 971-986.

Detemple, J. B., R. Garcia, M. Rindisbacher. 2003. A Monte Carlo Method for Optimal Portfolios. *J. Finance* 58(1) 401-446.

Detemple, J. B., R. Garcia, M. Rindisbacher. 2004. Asymptotic Properties of Monte Carlo Estimators of Diffusion Processes. Working Paper, CIRANO, forthcoming *J. Econometrics.*

[22] Note that $\left(\partial_\lambda X_T(\theta^\lambda)\right)_{|\lambda=0}$ exists whenever the stochastic flow is differentiable with respect to the initial condition. A sufficient condition for this is $X_T \in \mathbb{D}^{1,2}$, which is satisfied when the diffusion is in the domain of the Malliavin derivative operator. See Bichteler, Gravereaux, Jacod (1986) for more on the differentiability of perturbed diffusions.

Duffie, D., P. Glynn. 1995. Efficient Monte Carlo Simulation of Security Prices.] *Ann. Appl. Probab.* 5(4) 1995: 897-905.

Emanuel, D. C., J. D. MacBeth. 1982. Further Results on the Constant Elasticity of Variance Call Option Pricing Model. *J. Fin. Quant. Anal.* 17(3) 533-554.

Fournié, E., J-M. Lasry, J. Lebuchoux, P-L. Lions, N. Touzi. 1999. Applications of Malliavin Calculus to Monte Carlo Methods in Finance. *Finance Stochastics* 3(4) 391-412.

Fournié, E., J-M. Lasry, J. Lebuchoux, P-L. Lions. 2001. Applications of Malliavin Calculus to Monte Carlo Methods in Finance II. *Finance Stochastics* 5(2) 201-236.

Glasserman, P. 1991. *Gradient Estimation via Perturbation Analysis.* Kluwer Academic Publishers, Norwell, MA.

Glynn, P. W. 1989. Optimization of Stochastic Systems via Simulation. *Proceedings of the 1989 Winter Simulation Conference.* Society for Computer Simulation, San Diego, 90-105.

Ikeda, N., S. Watanabe. 1981. *Stochastic Differential Equations and Diffusion Processes.* North Holland.

Kallenberg, O. 1997. *Foundations of Modern Probability Theory.* Springer-Verlag, New York.

Karatzas, I., S. Shreve. 1991. *Brownian Motion and Stochastic Calculus.* Second Edition, Springer-Verlag, New York.

Kloeden, P., E. Platen. 1997. *Numerical solution of stochastic differential equations.* Springer-Verlag, New York.

Kunita, H. 1991. *Stochastic Flows and Stochastic Differential Equations.* Cambridge University Press.

L'Ecuyer, P. 1990. A Unified View of IPA, SF, and LR Gradient Estimation Techniques. *Management Sci.* 36(11) 1364-1383.

L'Ecuyer, P., G. Perron. 1994. On the Convergence Rate of IPA and FDC Derivative Estimators. *Oper. Res.* 42(4) 643-656.

Lehmann, E. L. 1997. *Testing Statistical Hypothesis.* Second Edition, Springer-Verlag, New York.

Merton, R. C. 1971. Optimum Consumption and Portfolio Rules in a Continuous Time Model. *J. Econ. Theory* 3(4) 273-413.

Phillips, P. C. B., H. R. Moon. 1999. Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica* 67(5) 1057-1111.

Rubinstein, R. Y., A. Shapiro. 1993. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method.* Wiley, Chichester and New York.

Wachter, J. 2002. Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets. *J. Finan. Quant. Anal.* 37(1) 63-91.

Table 1: Monte Carlo Estimators for Hedging Demand with CRRA Preferences

Monte Carlo Finite Difference Method (MCFD)
$\left(\pi_t^{hedge}\right)^{M,N,\tau}(t,y) = \sigma(t,y)^{-1}\left[\sigma^Y(t,y)'\left[\frac{f^{M,N}(t,y+\alpha\tau_j e_j) - f^{M,N}(t,y-(1-\alpha)\tau_j e_j)}{f^{M,N}(t,y)}\right]_{j=1,\ldots,d}\right]$
$f^{M,N}(t,z) = \frac{1}{M}\sum_{i=1}^{M}(\xi_{t,T}^{i,N})^{\rho}$ with $Y_t = z$
(MCFFD: $\alpha = 1$), (MCCFD: $\alpha = \frac{1}{2}$) and (MCBFD: $\alpha = 0$)
Monte Carlo Covariation (MCC)
$\left(\pi_t^{hedge}\right)^{M,N,\tau}(t,y) = \sigma(t,y)^{-1}\left[\frac{1}{\tau}\left[\frac{\sum_{i=1}^{M}(\xi_{t,T}^{y,i,N})^{\rho}\frac{W_{j,t+\tau}^{i} - W_{j,t}^{i}}{\xi_{t,t+\tau}^{y,i,N}}}{\sum_{i=1}^{M}(\xi_{t,T}^{y,i,N})^{\rho}}\right]_{j=1,\ldots,d} - \frac{1}{R}\theta(t,y)\right]$
Monte Carlo Malliavin Derivatives (MCMD)
$\left(\pi_t^{hedge}\right)^{M,N,\tau}(t,y) = \sigma(t,y)^{-1}\left[(-\rho)(a^{N,M}(t,y) + b^{N,M}(t,y))\right]$

Table 2: Euler Scheme for Components of Portfolio Estimator

State Variables ($h = \frac{T-t}{N}$)
$\Delta Y_{t+kN}^{N}(y) = \mu^Y(t+kh, Y_{t+kh}^{N}(y))h + \sigma^Y(t+kh, Y_{t+kh}^{N}(y))\Delta W_{t+kh}$
$k = 1,\ldots,N$ and $Y_t^N(y) = y$
$\Delta\xi_{t,t+kh}^{N}(y) = -\xi_{t,t+kh}[r(t+kh, Y_{t+kh}^{N}(y))h + \theta(t+kh, Y_{t+kh}^{N}(y))'\Delta W_{t+kh}]$
$k = 1,\ldots,N$ and $\xi_{t,t}^{N}(y) = 1$
Hedging Terms ($h = \frac{T-t}{N}$)
$[a^{M,N} + b^{M,N}](t,y) = \frac{\sum_{i=1}^{M}(H_{t,T}^{a,i,N} + H_{t,T}^{b,i,N})(\xi_{t,T}^{i,N}(y))^{\rho}}{\sum_{i=1}^{M}(\xi_{t,T}^{i,N}(y))^{\rho}}$
$\Delta H_{t,t+kh}^{a,N} = ([\mathcal{D}_t Y_{t+kh}(y)]^N)'(\partial r(t+kh, Y_{t+kh}^{N}(y)))'h$
$k = 1,\ldots,N$ and $H_{t,t}^{a,N}(y) = 0$
$\Delta H_{t,t+kh}^{b,N} = ([\mathcal{D}_t Y_{t+kh}(y)]^N)'(\partial\theta(t+kh, Y_{t+kh}^{N}(y)))'[\theta(t+kh, Y_{t+kh}^{N}(y))h + \Delta W_{t+kh}]$
$k = 1,\ldots,N$ and $H_{t,t}^{b,N}(y) = 0$
Malliavin Derivatives ($h = \frac{T-t}{N}$)
$\Delta[\mathcal{D}_t Y_{t+kh}(y)]^N = [\partial_y\mu^Y(t+kh, Y_{t+kh}^{N}(y))h + \sum_{j=1}^{d}\partial_y\sigma_j^Y(t+kh, Y_{t+kh}^{N}(y))\Delta W_{t+kh}^{j}]\mathcal{D}_t Y_{t+kh}^{N}(y)]^N$
$k = 1,\ldots,N$ and $[\mathcal{D}_t Y_t(y)]^N = \sigma^Y(t,y)$

Table 3: Delta Hedge Estimators of Digital Options in the CEV Model ($h(T,x) = e^{-rT}\mathbf{1}_{\{x>K\}}$)

Monte Carlo Finite Difference Method (MCFD)
$\widetilde{[\partial_x f]}^{M,N,\tau}(t,x) = \frac{1}{M}\sum_{i=1}^{M} \frac{h(T,X_T^{i,N}(x+\alpha\tau)) - h(T,X_T^{i,N}(x-(1-\alpha)\tau))}{\tau}$ (MCFFD: $\alpha = 1$), (MCCFD: $\alpha = \frac{1}{2}$) and (MCBFD: $\alpha = 0$)
Monte Carlo Covariation (MCC)
$\widehat{[\partial_x f]}^{M,N,\tau}(t,x) = \sum_{i=1}^{M} h(T,X_T^{i,N}(x)) \left(\frac{W_{t+\tau}^i - W_t^i}{\tau}\right) \left(\sigma x^{1+\beta}\right)^{-1}$
Monte Carlo Malliavin Weights (MCMD-W)
$[\partial_x f]^{M,N}(t,x) = \frac{1}{M}\sum_{i=1}^{M} h(T,X_T^{i,N}(x))[H_{t,T}^{\left(\mathbf{1}_{[t,T]}(\cdot)/(T-t)\right)}]^{i,N}$
Monte Carlo Malliavin Mixture (MCMD-M)
$\widehat{[\partial_x f]}^{M,N}(t,x) = \frac{1}{M}\sum_{i=1}^{M} h_\delta(T,X_T^{i,N}(x) \left[H_{t,T}^{\left(\mathbf{1}_{[t,T]}(\cdot)/(T-t)\right)}(x)\right]^{i,N} + \frac{1}{M}\sum_{i=1}^{M} k_\delta(T,X_T^{i,N}) \left[\mathcal{D}_t X_T(x)\right]^{i,N}$ $h_\delta(T,x) = h(T,x) - e^{-rT}\frac{(x-(K-\delta))^+\mathbf{1}_{\{x\leq x+\delta\}}}{2\delta} + \mathbf{1}_{\{x\geq K+\delta\}}$ and $k_\delta(x) = \frac{e^{-rT}}{\sigma x^{1+\beta}}\left(\frac{\mathbf{1}_{\{x\geq K-\delta\}}\mathbf{1}_{\{x\leq K+\delta\}}}{2\delta}\right)$

Table 4: Components of Delta Hedge Estimators of Digital Options in the CEV Model

State Variables ($h = \frac{T-t}{N}$)
$\Delta X_{t+kh}^N(x) = X_{t+kh}^N(x)((r-q)h + \sigma(X_{t+kh}^N)^\beta \Delta W_{t+kh})$ $k = 1,\ldots,N$ and $X_t^N(x) = x$
Malliavin Derivatives ($h = \frac{T-t}{N}$)
$\Delta[\mathcal{D}_t X_{t+kh}(x)]^N = [\mathcal{D}_t X_{t+kh}(x)]^N \left((r-q)h + \sigma X_{t+kh}^N(x)^{\beta-1}\left(X_{t+kh}^N(x)^{\beta-1} + b\right)\Delta W_{t+kh}\right)$ $k = 1,\ldots,N$ and $[\mathcal{D}_t X_t(x)]^N = \sigma x^{1+\beta}$
Malliavin Weights ($h = \frac{T-t}{N}$)
$\left[H_{t,T}^{\left(\mathbf{1}_{[t,T]}(\cdot)/T\right)}\right]^N = \frac{1}{T}\sum_{k=1}^{N}\left(\left(\sigma X_{t+kh}^N\right)^{1+\beta}\right)^{-2}[\mathcal{D}_t X_{t+kh}(x)]^N \Delta W_{t+kh}$

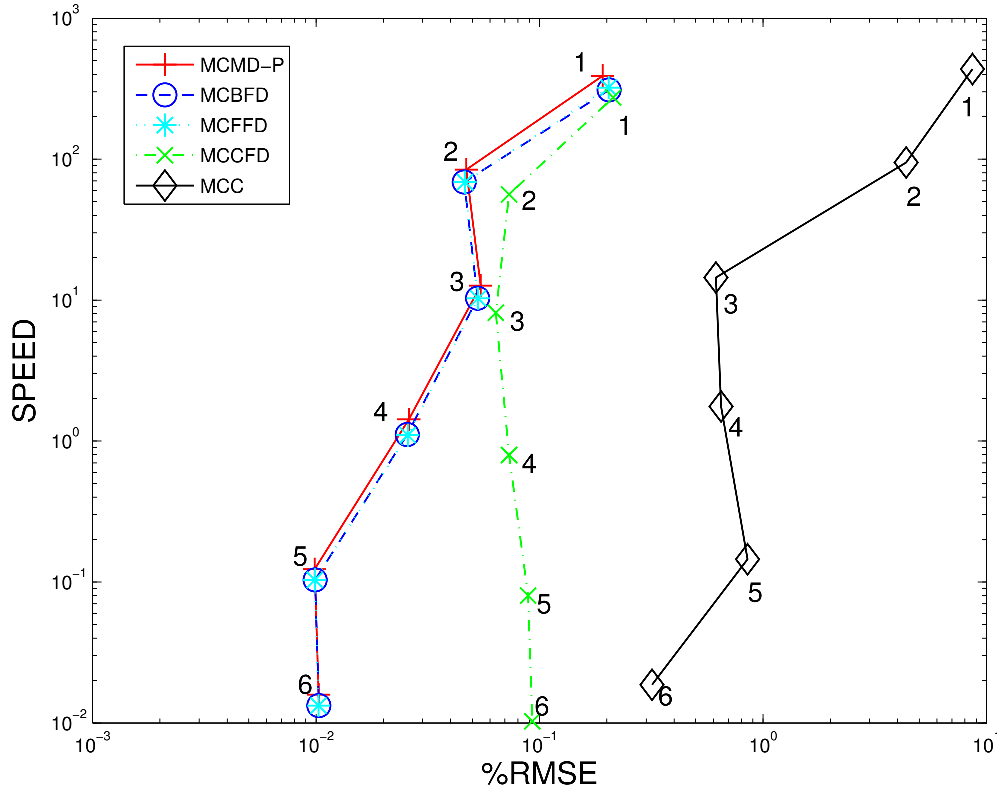

Figure 1: Efficiency Plot: % RMSE and Speed for 100 Portfolios (Risk aversion $R = 1.5, \ldots, 6$ and Horizon $T = 1, \ldots, 10$.) Speed is measured as the inverse of CPU time. The following simulation setup is used. For $i = 1, ..., 6$, we use $+^i$ for Malliavin derivatives (MCMD-P) with $N = 5 \times 2^{i-1} \times T$ and $M = 200 \times 2^{2(i-1)}$; $\times^i$ for Central finite differences (MCCFD) with $N = 5 \times 2^{i-1} \times T$, $M = 200 \times 2^{2(i-1)}$ and $\tau = 2^{(1-i)/2} \times 0.01$; $*^i$ for Forward finite differences (MCFFD) with $N = 5 \times 2^{i-1} \times T$ and $M = 200 \times 2^{2(i-1)}$ and $\tau = 2^{1-i} \times 0.01$; $\bigcirc^i$ for Backward finite differences (MCBFD) with $N = 5 \times 2^{i-1} \times T$ and $M = 200 \times 2^{2(i-1)}$ and $\tau = 2^{1-i} \times 0.01$; $\Diamond^i$ for Cross variation (MCC) with $N = 5 \times 2^{i-1} \times T$, $M = 200 \times 2^{3(i-1)}$ and $\tau = 2^{(i-1)} \times 0.2$. % RMSEs are calculated relative to "true" values obtained from the Malliavin derivative estimator with Doss transformation (see Detemple, Garcia and Rindisbacher (2003)) with $N = 1,000$ and $M = 3,000,000$.

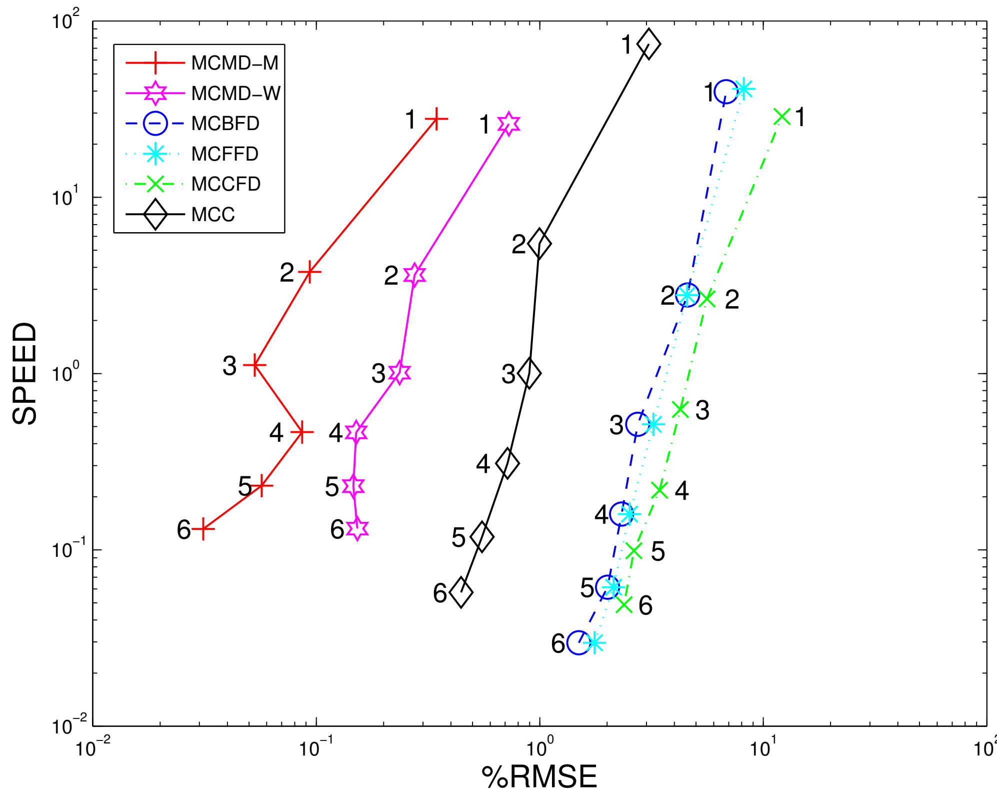

Figure 2: Efficiency Plot: % RMSE and Speed for 990 parameterizations of digital options when the underlying asset follows a CEV process. Speed is measured as the inverse of CPU-time. The following simulation setup is used. For $i = 1, ..., 6$, we use $+^i$ and $\bigstar^i$ for Malliavin derivatives (MCMD-M and MCMD-W) with $N = 10 \times i$ and $M = 1000 \times i^2$; $\times^i$ for Central finite differences (MCCFD) with $N = 10 \times i$, $M = [1000 \times i^{5/2}]$ and $\tau = 0.001 \times i^{-2}$; $*^i$ for Forward finite differences (MCFFD) with $N = 10 \times i$ and $M = 1000 \times i^3$ and $\tau = 0.001 \times i^{-3}$; $\bigcirc^i$ for Backward finite differences (MCBFD) with $N = 10 \times i$ and $M = 1000 \times i^3$ and $\tau = 0.001 \times i^{-3}$; $\diamondsuit^i$ for Cross variation (MCC) with $N = 10 \times i$, $M = 1000 \times i^3$ and $\tau = 0.001 \times i^{-3}$.