Asymptotic Properties of Monte Carlo Estimators of Derivatives
Author(s): Jérôme Detemple, René Garcia and Marcel Rindisbacher
Source: *Management Science*, Vol. 51, No. 11 (Nov., 2005), pp. 1657-1675
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/20110454
Accessed: 29-03-2017 16:11 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Management Science*

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

MANAGEMENT SCIENCE
Vol. 51, No. 11, November 2005, pp. 1657–1675
ISSN 0025-1909 | EISSN 1526-5501 | 05 | 5111 | 1657

DOI 10.1287/mnsc.1050.0398
© 2005 INFORMS

Asymptotic Properties of Monte Carlo Estimators of Derivatives

Jérôme Detemple
School of Management, Boston University, 595 Commonwealth Avenue, Boston, Massachusetts 02215, and CIRANO, 2020 University Street, 25th Floor, Montréal, Québec, H3A 2A5 Canada, detemple@bu.edu

René Garcia
Economics Department, Université de Montréal, Montréal, Québec, H3A 2A5 Canada; CIREQ, Montréal, Canada; and CIRANO, 2020 University Street, 25th Floor, Montréal, Québec, H3A 2A5 Canada, rene.garcia@umontreal.ca

Marcel Rindisbacher
J.L. Rotman School of Management, University of Toronto, Toronto, Ontario, Canada, and CIRANO, 2020 University Street, 25th Floor, Montréal, Québec, H3A 2A5 Canada, rindisbm@rotman.utoronto.ca

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, 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.

Key words: simulation; derivative estimation; Malliavin path; Malliavin weight; covariation; finite difference; likelihood ratio; weak convergence

History: Accepted by Ravi Jagannathan, finance; received April 24, 2002. This paper was with the authors 1 year and 8 months for 1 revision.

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

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

where

$$\mathscr{L}_t f(t,x) = \partial_t f(t,x) + \partial_x f(t,x) A + \tfrac{1}{2}\operatorname{trace}(\partial_{xx} f(t,x) BB')$$

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 §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, $i = 1, \ldots, M$ of

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

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

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 MC 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 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

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

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

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

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, $M^{-1/2}$.

[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 ($W = \text{const.} \times N^2$), 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 ($W = \text{const.} \times M$), this convergence rate is also $W^{-1/2}$.

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 (see Broadie and Glasserman 1996) is based on the partial derivative of the deterministic function that 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 that associates the terminal value to the initial value, that is, 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 §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 (available online at http://mansci.pubs.informs.org/ecompanion.html), 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 emphasize 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.

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

Cvitanic et al. (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 §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 nondiffusion 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 §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 is the discretization error due to the finite number of discretization points, N, in the approximation (4) of (2). The second is the MC 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 nonzero 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 MC estimators of derivatives (see Duffie and Glynn 1995 for a discussion of efficient simulation designs for MC 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 that 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 addresses the issue of risk management for digital options when the underlying price follows a process with constant elasticity of variance. These numerical studies 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 et al. (2004) that is used

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

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

to characterize the second-order bias for the different methods under consideration. Proofs are 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 nonsmooth. All appendices can be found in an online supplement at http://mansci.pubs.informs.org/ecompanion.html.

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

$$F \equiv f(W_{t_1}, \ldots, W_{t_n}),$$

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 $t_1 < \cdots < t_{k-1} < t \le t_k \cdots < t_n$ and suppose that W_s is perturbed to $W_s + \varepsilon$ for all $s \ge t$. The Malliavin derivative of F at time t, which is written as $\mathcal{D}_t F$, is defined as

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

where f_i is the derivative with respect to the ith 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 $f(x) \equiv S_0 \exp((\mu - (1/2)\sigma^2)T + \sigma x)$. An application of the definition shows that the Malliavin derivative of S_T is

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

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 $\|F\|_{1,2} \equiv (\mathbf{E}[F^2] + \mathbf{E}[\int_0^T \|\mathcal{D}_t F\|^2 \, dt])^{1/2}$.

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

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

the notation $W_s^\varepsilon = W_s + \varepsilon 1_{[t,\infty)}(s)$, we can write

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

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

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

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

$$\mathscr{D}_t F = \int_0^T \mathscr{D}_t h(W,s)\, ds = \int_t^T \mathscr{D}_t h(W,s)\, ds.$$

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 $F = (F_1, \ldots, F_n)$ is a vector of random variables in $\mathbb{D}^{1,2}$, and that ϕ is a differentiable function of F with bounded derivatives. Then,

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

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

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

and the chain rule of Malliavin calculus gives

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

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

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

2.3. Malliavin Derivatives of Solutions of Stochastic Differential Equations

The rules of Malliavin calculus described above enable us 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

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

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

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

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

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

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

Note that the Malliavin derivative of an SDE 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 $\nabla_{t,x} X_s(x) \equiv \lim_{\tau\to 0}(1/\tau)(X_s(x+\tau) - X_s(x))$ (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 $\mathscr{D}_t X_s = \sigma(X_t)\nabla_{t,x} X_s(x)$: The Malliavin derivative is equal to the tangent $\nabla_{t,x} X_s(x)$ multiplied by the volatility coefficient of the process.

[5] Ordinary exponentials solve linear ODEs. Likewise, stochastic exponentials solve linear SDEs. More precisely, $\mathscr{E}(M)$ solves $d\mathscr{E}(M)_t = \mathscr{E}(M)_t dM_t$ subject to $\mathscr{E}(M)_0 = 1$. For univariate continuous local martingales M, we have $\mathscr{E}(M)_T = \exp(M_T - (1/2)[M]_T)$, 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)).

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

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

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

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 $M_t = M_0 + \int_0^t \mathbf{E}_s[\mathcal{D}_s F]\, dW_s$.

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 $W = (W_1, \ldots, W_d)'$ is a d-dimensional Brownian motion and that $F \equiv f(W_{t_1}, \ldots, W_{t_n})$ 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 $\mathcal{D}_t F = (\mathcal{D}_{1t} F, \ldots, \mathcal{D}_{dt} F)$, where, for $j = 1, \ldots, d$, the derivative $\mathcal{D}_{jt} F$ is obtained by perturbing the jth component of the Brownian motion. In other words, $\mathcal{D}_{jt} F$ 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 an 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, p. 32) gives

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

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

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

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

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

An MC 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,

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

This estimator of the derivative $\partial_x f(t, x)$ is of the same form as the MC 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

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

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

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

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

where

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

Clearly, an MC 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 $Y_2' \equiv [\mathcal{D}_{1t}X', \ldots, \mathcal{D}_{dt}X']$, satisfies the SDE

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

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

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

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 $dX_t = X_t(A dt + B dW_t)$, we obtain $\mathcal{D}_t X_T = B X_T$, 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 §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 §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).

[6] Broadie and Glasserman use this example to introduce the idea of path derivatives. As $X_T = X_t F$, where the random variable $F = \exp((A - (1/2)B^2)(T-t) + B(W_T - W_t))$ 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, Broadie and Glasserman define path derivatives for the Euler discretized processes (see Proposition 7, p. 284). For this approximation scheme, a deterministic map links the initial and terminal values of the discretized processes for each subinterval.

3.1.2. Malliavin Derivative Weight Estimators (MCMD-W). The estimators presented so far require that h be at least continuous. 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 continuous but suppose that the transition density $p(t,x,s,y)$ of the diffusion exists and is known. Under these conditions the conditional expectation

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

can be Malliavin-differentiated on both sides to obtain

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

The resulting probabilistic representation of the derivative

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

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) and (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"

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

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

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

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

establishes[8]

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

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

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

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 an MC 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 nondifferentiable 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. Because 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. Because the weight that attains the minimal mean square error is unique, this "weight" must be identical to the score. That is,

$$\mathbf{E}_{t,x}\big[H_{t,T}^{(\alpha)} \mid X_T\big] = \partial_x \log p(t,x,T,X_T) \equiv H_{t,T}^{\text{score}}, \tag{22}$$

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,

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

for a particular density given by

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

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

[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 nonanticipative processes.

[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 cannot be calculated in explicit form if the log-likelihood ratio is unknown.

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

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

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 MC 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 et al. (2004) on asymptotic distributions of MC estimators for Euler discretized diffusions can be applied. The following theorem describes the weak limit of the estimation error.[13]

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

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

Then, as $M\to\infty$,

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

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*

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

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 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

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

naturally suggests the estimator

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

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 et al. (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* $g\in\mathscr{C}^3(\mathbb{R}^d)$ *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 Equation* (40) *in Appendix* A.

Define the events

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

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

[13] Let S be a metric space and $\mathscr{S}$ 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 MC 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.

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

where $I^{N,\tau}_{t,T}(x) = \mathbf{E}_{t,x}[N(g(X^N_T) - g(X_T))(\Delta_\tau W_t/\tau)]$, *and suppose that the conditions*

$$\lim_{r\to\infty}\limsup_{1/\tau,\,N\to\infty}\mathbf{E}_{t,x}\bigg[\mathbf{1}_{F_{t,T}(N,r,\tau)} \cdot \bigg\|N\big(g(X^N_T)-g(X_T)\big)\frac{\Delta_\tau W_t}{\tau}-\frac{1}{2}I^{N,\tau}_{t,T}(x)\bigg\|\bigg]=0, \quad (27)$$

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

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

$$M^{1/3}\big(\widehat{\partial_x f(t,x)}^{M,N_M,\tau_M} - \partial_x f(t,x)\big) \Rightarrow \varepsilon^c_1\big[\partial_v \mathbf{E}_{t,x}[\partial_x f(v,X_v)B(X_v)]\big]_{v=t}C(x) + \varepsilon^c_2\tfrac{1}{2}\partial K_{t,T}(x) + O_{t,T}(x)C(x),$$

where $N_M, 1/\tau_M\to\infty$ *when* $M\to\infty$,

$$\varepsilon^c_1 = \lim_{M\to\infty} M^{1/3}\tau_M \quad \textit{and} \quad \varepsilon^c_2 = \lim_{M\to\infty} M^{1/3}/N_M,$$

C is defined as 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 $N(g(X^N_T) - g(X_T))(\Delta_\tau W_t/\tau)$ 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 the Central Limit Theorem (see Kallenberg 1997, Theorem 4.12, p. 69). Sufficient conditions for (28) are uniform $\mathbf{L}^p$-boundedness of $g(X_T)(\Delta_\tau W_t/\tau)$ 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, $(1/2)\partial K_{t,T}(x)$, 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 $(1/\tau)\cdot\int_t^{t+\tau}\partial_x f(v,X_v)B(X_v)\,dv C(x)$, 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 nontrivial task and the resulting second-order bias is a matter of concern.

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 $H^{(\alpha^{\mathrm{MCC}})}_{t,T}$, obtained for the choice $\alpha^{\mathrm{MCC}}_{t,s} = (1/\tau)\mathbf{1}_{(t,t+\tau]}(s)$, by its Euler approximation with step size τ,

$$[\widehat{H}^{(\alpha^{\mathrm{MCC}})}_{t,T}]^\tau = \bigg(\frac{W_{t+\tau}-W_t}{\tau}\bigg)' C(X_t).$$

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 an MC version of the well-known finite difference method for PDEs. The motivation for this estimator is the limiting result

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

for $\alpha_j\in[0,1]$, $j=1,\dots,d$, where $X_T(x)$ is the diffusion process started at $X_t = x$ and $e_j = [0,\dots,1,\dots,0]'$ is the jth 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 MC finite difference (MCFD) estimator

$$\widehat{\partial_y f(t,x)}^{M,N,\tau_j,\alpha_j} = \bigg(\frac{\frac{1}{M}\sum_{i=1}^M\big[g(X^{i,N}_T(x+\alpha_j\tau_j e_j)) - g(X^{i,N}_T(x-(1-\alpha_j)\tau_j e_j))\big]}{\tau_j}\bigg)_{j=1,\dots,d}, \quad (30)$$

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

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

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

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

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

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

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

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

(i) *If $\alpha_j = 1/2$ (MCCFD), we have*

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

where N_M, $1/\tau_{j,M} \to \infty$ when $M \to \infty$,

$$\varepsilon_{j1}^{fcd} = \lim_{M\to\infty} M^{1/4}\tau_{j,M}, \quad \text{and} \quad \varepsilon_{j2}^{fcd} = \lim_{M\to\infty} M^{1/2}/N_M.$$

(ii) *If $\alpha_j \neq 1/2$ (MCBFD and MCFFD), we have*

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

where N_M, $1/\tau_{j,M} \to \infty$ when $M \to \infty$, with

$$\delta(\alpha_j) = (2\alpha_j - 1)/2,$$

$$\varepsilon_{1j}^{fd} = \lim_{M\to\infty} M^{1/2}\tau_{j,M}, \quad \text{and} \quad \varepsilon_{j2}^{fd} = \lim_{M\to\infty} M^{1/2}/N_M.$$

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

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

Because $J_T^{1,\tau_j} \Rightarrow \partial g(X_T)$ and $J_T^{2,\tau_j} \Rightarrow \mathscr{D}_t X_T C(x) e_j$ 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

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

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

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 $B_i \cap B_j = \varnothing$ for $i \neq j$, with $g^c(x) \in \mathscr{C}^3(\mathbb{R}^d)$ and suppose that $g^c(X_T) \in \mathbb{D}^{1,2}$. Let $K_T(x)$ be defined as in Theorem 5 in Equation* (40) *in Appendix* A. *Suppose that 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*

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

where $N_M, 1/\tau_{j,M} \to \infty$ when $M \to \infty$,

$$\varepsilon_{j1}^{fcd} = \lim_{M\to\infty} M^{1/5} \tau_{j,M}, \quad \text{and} \quad \varepsilon_{j2}^{fcd} = \lim_{M\to\infty} M^{2/5}/N_M.$$

(ii) *If $\alpha_j \neq 1/2$ (MCBFD and MCFFD), we have*

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

where $N_M, 1/\tau_{j,M} \to \infty$ when $M \to \infty$, with

$$\delta(\alpha_j) = (2\alpha_j - 1)/2,$$
$$\varepsilon_{1j}^{fd} = \lim\nolimits_{M\to\infty} M^{1/3} \tau_{j,M}, \quad \text{and} \quad \varepsilon_{j2}^{fd} = \lim\nolimits_{M\to\infty} M^{1/3}/N_M.$$

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

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

and such that $Q_{t,T}^j(x)$ and $Q_{t,T}^k(x)$ 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 MC 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.

Because this scheme also produces a numerical approximate of an approximation of the derivative, the second-order bias has two components. The first, $(1/2)\partial K_{t,T}(x)$, is the same as the one for the covariation estimator. The second 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

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

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

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

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

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

where r is the riskless short rate and θ is 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 $V(t, x, y) = \sup_\pi \mathbf{E}[u(X_T^\pi) \mid X_t = x, Y_t = y]$ 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 PDE

$$0 = \partial_t V + \partial_y V\mu^Y + \tfrac{1}{2}\text{trace}\{\partial_{yy}V\sigma^Y(\sigma^Y)'\} + x\partial_x Vr - \tfrac{1}{2}x^2\partial_{xx}V\left\| \frac{\partial_x V}{-x\partial_{xx}V}\theta' + \frac{\partial_{xy}V}{-x\partial_{xx}V}\sigma^Y \right\|^2,$$

subject to the boundary condition $V(T, x) = u(T, x)$. Merton (1971) emphasizes 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

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

subject to the boundary condition $f(T, y) = 1$. In (34),

$$\mathcal{L}_t f = \partial_t f + \partial_y f\mu^Y + \tfrac{1}{2}\text{trace}\{\partial_{yy}f\sigma^Y(\sigma^Y)'\}$$

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 MC methods described in the previous section. Let $k(t, Y_t) \equiv \mathbf{E}_{t, Y_t}[\xi_{t,T}^\rho]$, where

$$\xi_{t,T} \equiv \exp\left(-\int_t^T r_s\,ds - \int_t^T \theta_s'\,dW_s - \tfrac{1}{2}\int_t^T \theta_s'\theta_s\,ds\right)$$

is the state price density implied by the market structure (S, Y, θ, r). Passing to the new measure $\widetilde{\mathbf{P}}$, under which $\widetilde{W}_t = W_t + \rho\int_0^t \theta(s, Y_s)\,ds$ is a Brownian motion, gives

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

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

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

Using $\widetilde{\mathcal{L}}_t k = \mathcal{L}_t k + \partial_y k(-\rho\sigma^Y\theta)$ gives

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

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 ($d[X, W]_t/X_t = \pi_t'\sigma_t dt$). Because

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

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 et al. 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

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

This formula is the basis for the approach proposed by Cvitanic et al. (2002, 2003). They suggest computing the portfolio based on the formula on the right-hand side of the above equation with τ fixed. This approach is clearly just a special case of MCC methods presented in §2.2.

The second MC 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

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

where functions a and b are

$$a(t, Y_t) \equiv \frac{\mathbf{E}_{t, Y_t}[\xi_{t,T}^\rho \int_t^T \mathcal{D}_t r_s \, ds]}{\mathbf{E}_{t, Y_t}[\xi_{t,T}^\rho]},$$

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

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 functions a and b.

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

$$\pi_t' \sigma_t = \frac{1}{R} \theta_t' + \frac{\lim_{\tau \to 0} \frac{1}{\tau} (\mathbf{E}_{t, Y_t + \alpha\tau}[\xi_{t,T}^\rho] - \mathbf{E}_{t, Y_t - (1-\alpha)\tau}[\xi_{t,T}^\rho])}{\mathbf{E}_{t, Y_t}[\xi_{t,T}^\rho]} \sigma^Y,$$

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 $\alpha = 1/2$, and MCFFD for $\alpha = 1$.

Table 1 Monte Carlo Estimators for Hedging Demand with CRRA Preferences

Monte Carlo finite difference method (MCFD)

$$(\pi_t^{\text{hedge}})^{M,N,\tau}(t, y) = \sigma(t, y)^{-1} \left[\sigma^Y(t, y)' \left[\frac{f^{M,N}(t, y + \alpha\tau_j e_j) - f^{M,N}(t, y - (1-\alpha)\tau_j e_j)}{f^{M,N}(t, y)} \right]_{j=1,\dots,d} \right]$$

$$f^{M,N}(t, z) = \frac{1}{M} \sum_{i=1}^M (\xi_{t,T}^{i,N})^\rho \quad \text{with } Y_t = z$$

(MCFFD: $\alpha = 1$), (MCCFD: $\alpha = 1/2$), and (MCBFD: $\alpha = 0$)

Monte Carlo covariation (MCC)

$$(\pi_t^{\text{hedge}})^{M,N,\tau}(t, y) = \sigma(t, y)^{-1} \cdot \left[\frac{1}{\tau} \left[\frac{\sum_{i=1}^M (\xi_{t,T}^{y,i,N})^\rho (W_{j,t+\tau}^i - W_{j,t}^i)/\xi_{t,t+\tau}^{y,i,N}}{\sum_{i=1}^M (\xi_{t,T}^{y,i,N})^\rho} \right]_{j=1,\dots,d} - \frac{1}{R} \theta(t, y) \right]$$

Monte Carlo Malliavin derivatives (MCMD)

$$(\pi_t^{\text{hedge}})^{M,N,\tau}(t, y) = \sigma(t, y)^{-1} [(-\rho)(a^{N,M}(t, y) + b^{N,M}(t, y))]$$

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.

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.

The tables show that all methods, except the one based on Malliavin derivatives, involve three convergence parameters, M, N, and 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 $\sqrt{\tau}$ to satisfy the Lindeberg condition and apply the Central Limit Theorem for i.i.d. random variables. The perturbation parameter τ must be controlled along with the number of MC replications 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

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

Table 2 Euler Scheme for Components of Portfolio Estimator

State variables $(h = (T-t)/N)$

$$\Delta Y^N_{t+kN}(y) = \mu^Y(t+kh, Y^N_{t+kh}(y))h + \sigma^Y(t+kh, Y^N_{t+kh}(y))\Delta W_{t+kh}$$

$$k = 1, \ldots, N \text{ and } Y^N_t(y) = y$$

$$\Delta \xi^N_{t,t+kh}(y) = -\xi_{t,t+kh}\left[r\left(t+kh, Y^N_{t+kh}(y)\right)h + \theta\left(t+kh, Y^N_{t+kh}(y)\right)'\Delta W_{t+kh}\right]$$

$$k = 1, \ldots, N \text{ and } \xi^N_{t,t}(y) = 1$$

Hedging terms $(h = (T-t)/N)$

$$[a^{M,N} + b^{M,N}](t,y) = \frac{\sum_{i=1}^M \left(H^{a,i,N}_{t,T} + H^{b,i,N}_{t,T}\right)\left(\xi^{i,N}_{t,T}(y)\right)^\rho}{\sum_{i=1}^M \left(\xi^{i,N}_{t,T}(y)\right)^\rho}$$

$$\Delta H^{a,N}_{t,t+kh} = ([\mathcal{D}_t Y_{t+kh}(y)]^N)'\left(\partial r\left(t+kh, Y^N_{t+kh}(y)\right)\right)'h$$

$$k = 1, \ldots, N \text{ and } H^{a,N}_{t,t}(y) = 0$$

$$\Delta H^{b,N}_{t,t+kh} = ([\mathcal{D}_t Y_{t+kh}(y)]^N)'\left(\partial \theta\left(t+kh, Y^N_{t+kh}(y)\right)\right)' \cdot \left[\theta\left(t+kh, Y^N_{t+kh}(y)\right)h + \Delta W_{t+kh}\right]$$

$$k = 1, \ldots, N \text{ and } H^{b,N}_{t,t}(y) = 0$$

Malliavin derivatives $(h = (T-t)/N)$

$$\Delta[\mathcal{D}_t Y_{t+kh}(y)]^N = \left[\partial_y \mu^Y(t+kh, Y^N_{t+kh}(y))h + \sum_{j=1}^d \partial_y \sigma^Y_j(t+kh, Y^N_{t+kh}(y))\Delta W^j_{t+kh}\right][\mathcal{D}_t Y_{t+kh}(y)]^N$$

$$k = 1, \ldots, N \text{ and } [\mathcal{D}_t Y_t(y)]^N = \sigma^Y(t,y)$$

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

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

where A, $\bar{\theta}$, 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 $\pi_{1t} = X_t(1/R)(\sigma_t)^{-1}\theta_t$ is the mean-variance demand and

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

with

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

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

represents the intertemporal hedging demand.

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

To compare the methods we proceed as follows. We consider a mutual fund with 100 different types of clients who can be 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 MC 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 MC methods use the common estimator $\xi^N_{t,T}$ for all risk tolerance profiles.

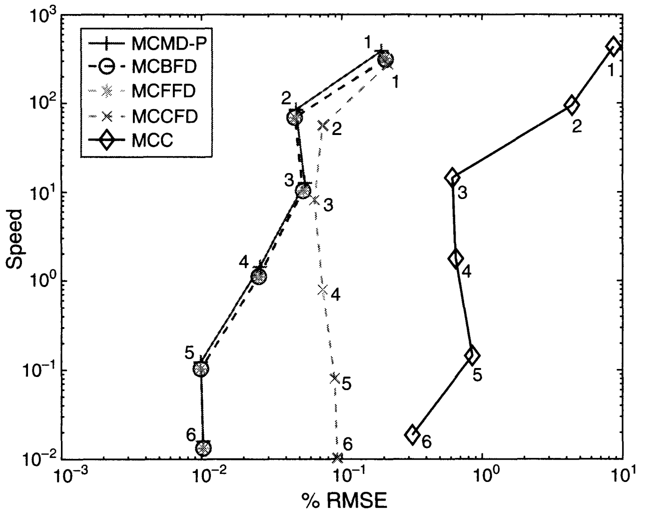
The experimental setup is as follows. Along the optimal convergence path we simulate six parameter combinations for (M, N, τ) and report the percentage mean square error based on the 100 portfolio weights. Figure 1 displays the results.

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 difference (MCCFD) does not seem to perform as well as forward finite difference (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, because 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 nontrivial 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.

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

Figure 1 Efficiency Plot: % RMSE and Speed for 100 Portfolios



Notes. 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 $i = 1, \ldots, 6$, 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 $N = 5 \times 2^{i-1} \times T$, $M = 200 \times 2^{2(i-1)}$, and $\tau = 2^{1-i} \times 0.01$; $\bigcirc^i$ for backward finite differences (MCBFD) with $N = 5 \times 2^{i-1} \times T$, $M = 200 \times 2^{2(i-1)}$, 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 et al. 2003) with $N = 1{,}000$ and $M = 3{,}000{,}000$.

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 a smooth utility (i.e., payoff) function. We now consider a risk management problem with a nonsmooth 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, that is, the derivative is null everywhere except at 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, because the Gamma of a plain vanilla European-style option exhibits a similar nonsmoothness 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. Because 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),

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

where r is the risk-free rate, q is 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 noncentral 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 MC estimators of derivatives provided in §3. Formulas can be found in Table 3. Implementation of MCMD-W and MCMD-M estimators requires the calculation of Malliavin weights. Euler approximations of these weights are given in Table 4.

To assess the relative efficiencies of the various hedging estimators for digital options, a large-scale experiment is performed as follows. In the first step, the parameters $[X, T, r, q, \sigma, \beta]$ of the model are

[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 $p(t, x, T, \cdot)$. Because the derivative of the indicator $\mathbf{1}_{\{X_T \le K\}}$ is the Dirac delta at K, the derivative of the digital becomes $\partial_x \mathbf{E}_{t,x}[\mathbf{1}_{\{X_T \le K\}}] = \mathbf{E}_{t,x}[\delta_{\{X_T = K\}} \nabla_{t,x} X_T] = (K/x)p(t, x, T, K)$. This approach, to calculate the path derivative estimator of a nonsmooth 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 nonsmooth function.

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

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: $\alpha = 1/2$), and (MCBFD: $\alpha = 0$)

Monte Carlo covariation (MCC)

$$\widehat{[\partial_x f]}^{M,N,\tau}(t,x) = \sum_{i=1}^{M} h(T, X_T^{i,N}(x))\left(\frac{W_{t+\tau}^i - W_t^i}{\tau}\right)(\sigma x^{1+\beta})^{-1}$$

Monte Carlo Malliavin weights (MCMD-W)

$$[\partial_x f]^{M,N}(t,x) = \frac{1}{M}\sum_{i=1}^{M} h(T, X_T^{i,N}(x))\left[H_{t,T}^{(\mathbf{1}_{[t,T]}(\cdot)/(T-t))}\right]^{i,N}$$

Monte Carlo Malliavin mixture (MCMD-M)

$$\widehat{[\partial_x f]}^{M,N}(t,x) = \frac{1}{M}\sum_{i=1}^{M} h_\delta(T, X_T^{i,N}(x))\left[H_{t,T}^{(\mathbf{1}_{[t,T]}(\cdot)/(T-t))}(x)\right]^{i,N} + \frac{1}{M}\sum_{i=1}^{M} k_\delta(T, X_T^{i,N})[\mathcal{D}_t X_T(x)]^{i,N}$$

$$h_\delta(T,x) = h(T,x) - e^{-rT}\frac{(x-(K-\delta))^+\mathbf{1}_{\{x\le x+\delta\}}}{2\delta} + \mathbf{1}_{\{x\ge K+\delta\}} \quad \text{and}$$

$$k_\delta(x) = \frac{e^{-rT}}{\sigma x^{1+\beta}}\left(\frac{\mathbf{1}_{\{x\ge K-\delta\}}\mathbf{1}_{\{x\le K+\delta\}}}{2\delta}\right)$$

drawn from independent distributions. In the 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 the 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 $N_1 = 10$, $M_1 = 1{,}000$, 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 to satisfy the optimal growth restrictions identified in Theorems 1–3.

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

State variables ($h = (T-t)/N$)

$$\Delta X_{t+kh}^N(x) = X_{t+kh}^N(x)\left((r-q)h + \sigma\left(X_{t+kh}^N\right)^\beta \Delta W_{t+kh}\right)$$

$$k = 1,\ldots,N \text{ and } X_t^N(x) = x$$

Malliavin derivatives ($h = (T-t)/N$)

$$\Delta[\mathcal{D}_t X_{t+kh}(x)]^N = [\mathcal{D}_t X_{t+kh}(x)]^N\left((r-q)h + \sigma X_{t+kh}^N(x)^{\beta-1} \cdot \left(X_{t+kh}^N(x)^{\beta-1} + b\right)\Delta W_{t+kh}\right)$$

$$k = 1,\ldots,N \text{ and } [\mathcal{D}_t X_t(x)]^N = \sigma x^{1+\beta}$$

Malliavin weights ($h = (T-t)/N$)

$$\left[H_{t,T}^{(\mathbf{1}_{[t,T]}(\cdot)/T)}\right]^N = \frac{1}{T}\sum_{k=1}^{N}\left(\left(\sigma X_{t+kh}^N\right)^{1+\beta}\right)^{-2}[\mathcal{D}_t X_{t+kh}(x)]^N \Delta W_{t+kh}$$

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]

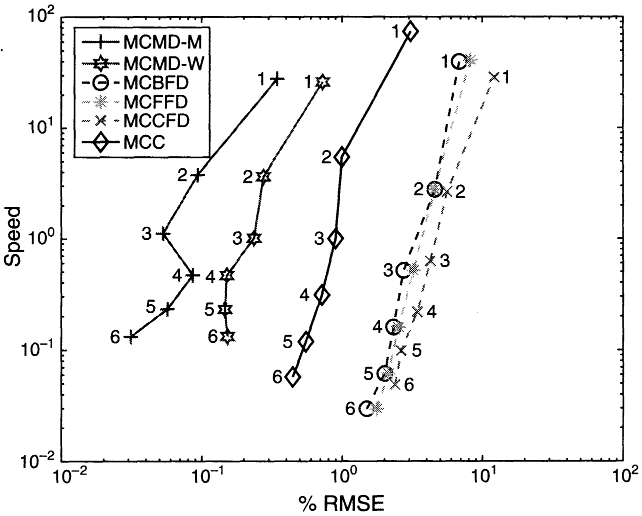
Figure 2 illustrates the results. First, note that the finite difference estimators (MCFD) are all less efficient than estimators involving exact (MCMD-W and MCMD-M) or approximate (MCC) weights. Among weight estimators, which do not involve derivatives of the payoff function, the order is the one predicted 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

[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. Because 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.

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

Figure 2 Efficiency Plot: % RMSE and Speed for 990 Parameterizations of Digital Options When the Underlying Asset Follows a CEV Process



Notes. Speed is measured as the inverse of CPU time. The following simulation setup is used: For $i = 1, \ldots, 6$, we use $+^i$ and $\star^i$ for Malliavin derivatives (MCMD-M and MCMD-W) with $N = 10 \times i$ and $M = 1{,}000 \times i^2$; $\times^i$ for central finite differences (MCCFD) with $N = 10 \times i$, $M = [1{,}000 \times i^{5/2}]$, and $\tau = 0.001 \times i^{-2}$; $*^i$ for forward finite differences (MCFFD) with $N = 10 \times i$, $M = 1{,}000 \times i^3$, and $\tau = 0.001 \times i^{-3}$; $\bigcirc^i$ for backward finite differences (MCBFD) with $N = 10 \times i$, $M = 1{,}000 \times i^3$, and $\tau = 0.001 \times i^{-3}$; $\diamond^i$ for cross variation (MCC) with $N = 10 \times i$, $M = 1{,}000 \times i^3$, and $\tau = 0.001 \times i^{-3}$.

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 provide 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 §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 §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 §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.

MCC estimators are less efficient than MCMD estimators because, as shown in §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 αs 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 MC 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 MC schemes. Alternative schemes, such as those

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms

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.

An online supplement to this paper is available on the *Management Science* website at http://mansci.pubs.informs.org/ecompanion.html.

Acknowledgments

Financial support from MITACS is gratefully acknowledged. The second author also thanks Hydro-Québec and the Bank of Canada for financial support.

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 and Breach, New York.

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, Graduate School of Business, Stanford University, Stanford, CA.

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

Cvitanic, J., L. Goukasian, F. Zapatero. 2003. Monte Carlo computation of optimal portfolios in complete markets. *J. Econom. Dynamics 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. *J. Econometrics*. Forthcoming.

Duffie, D., P. Glynn. 1995. Efficient Monte Carlo simulation of security prices. *Ann. Appl. Probab.* **5**(4) 897–905.

Emanuel, D. C., J. D. MacBeth. 1982. Further results on the constant elasticity of variance call option pricing model. *J. Financial Quant. Anal.* **17**(3) 533–554.

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.

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.

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

Glynn, P. W. 1989. Optimization of stochastic systems via simulation. *Proc. 1989 Winter Simulation Conf.* Society for Computer Simulation, San Diego, CA, 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*, 2nd ed. Springer-Verlag, New York.

Karatzas, I., J. P. Lemoczky, S. E. Shreve. 1987. Optimal portfolio and consumption decisions for a "small investor" on a finite horizon. *SIAM J. Control Optim.* **25**(5) 1557–1586.

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*, 2nd ed. Springer-Verlag, New York.

Merton, R. C. 1971. Optimum consumption and portfolio rules in a continuous time model. *J. Econom. Theory* **3**(4) 273–413.

Nualart, D. 1995. *The Malliavin Calculus and Related Topics*. Springer Verlag, New York.

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, UK, and New York.

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

This content downloaded from 132.204.3.57 on Wed, 29 Mar 2017 16:11:10 UTC
All use subject to http://about.jstor.org/terms