Sensitivity Estimates from Characteristic Functions

Paul Glasserman
Columbia Business School, Columbia University, New York, New York 10027, pg20@columbia.edu

Zongjian Liu
IEOR Department, Columbia University, New York, New York 10027, zl2115@columbia.edu

The likelihood ratio method (LRM) is a technique for estimating derivatives of expectations through simulation. LRM estimators are constructed from the derivatives of probability densities of inputs to a simulation. We investigate the application of the likelihood ratio method for sensitivity estimation when the relevant densities for the underlying model are known only through their characteristic functions or Laplace transforms. This problem arises in financial applications, where sensitivities are used for managing risk and where a substantial class of models have transition densities known only through their transforms. We quantify various sources of errors arising when numerical transform inversion is used to sample through the characteristic function and to evaluate the density and its derivative, as required in LRM. This analysis provides guidance for setting parameters in the method to accelerate convergence.

Subject classifications: simulation.
Area of review: Simulation.
History: Received January 2008; revisions received July 2008, July 2009, November 2009; accepted December 2009.

1. Introduction

Simulation-based derivative estimates are useful in sensitivity analysis and as inputs to optimization. In financial applications, sensitivity estimates are important in measuring and managing risk. Financial intermediaries hedge the risks they incur in buying and selling options by trading in the underlying assets; the hedging strategy is determined by the sensitivity of the option price to the underlying assets and other parameters.

The simulation literature provides several approaches to parameter sensitivity estimation; see Asmussen and Glynn (2007, Chapter 7) for an overview and references. The likelihood ratio method (LRM) constructs derivative estimators from the derivatives of probability densities associated with the simulation model. In a discrete-event simulation, the relevant densities would usually be those of interarrival or processing times; in financial applications, the relevant densities are typically those associated with the transition laws of the underlying assets. LRM is particularly attractive when there are discontinuities in the payoff or performance associated with simulated paths or discontinuities in the dynamics of the underlying model.

This paper develops and analyzes the use of LRM when the relevant densities are known only through their characteristic functions or Laplace transforms. Our primary motivation for investigating this problem lies in financial applications, where many important classes of models are characterized through Laplace or Fourier transforms. This is often the case for models driven by Lévy processes (see, e.g., Cont and Tankov 2004, Schoutens 2003) and also the affine class of jump-diffusion models studied in Duffie et al. (2000), which includes many of the most widely used models of asset prices. In other application domains, including telecommunications (e.g., Mikosch et al. 2002), stable laws arise in models with self-similarity; with few exceptions, the densities of stable laws are known only through their characteristic functions. Lévy-driven models also arise in queueing (Kella and Whitt 1992), insurance risk (Klüppelberg et al. 2004), and other areas of applied probability.

We analyze LRM estimators calculated through numerical transform inversion using, in particular, the method of Abate and Whitt (1992). The random weight used in the LRM estimator—the score function—involves both a probability density and its derivative; we numerically invert the transform of each. We analyze the various sources of error that arise in estimating expectations and their sensitivities using this approach. In particular, we focus on the simulation impact of three sources of numerical error: the grid spacing used in building an approximation to a distribution, the discretization error in the transform inversion integral, and the truncation error in the inversion integral. Building on the error analysis of Abate and Whitt (1992), we arrive at the following conclusions for the bias resulting from the combination of transform inversion and simulation in the problems we consider:

- the bias from approximating the distribution is quadratic in the grid spacing;
- the bias from the discretization error is exponentially small in the integration step size; and
- the magnitude of the bias from truncation depends on the decay of the modulus of the complex Laplace transform of the density or its derivative.

1611
This analysis provides guidance on the optimal allocation of computational effort across the various parts of the algorithm in order to balance the three sources of error. We will see that the decay rate of the transform can vary widely across models. Moreover, the relevant rate can be different for an expectation and its sensitivity. Thus, our main conclusion is that to understand the computational cost of this method, the key step is to examine the decay of the transform of the density and its derivative. We provide both theoretical and experimental support for our conclusions.

The implementation of simulation-based LRM estimators involves the underlying probability densities—and thus, in our setting, the use of transform inversion—in two ways: we need to know how to sample from the relevant densities in order to simulate the model, and we need to evaluate the density and its derivative in order to calculate the score function that defines the LRM estimators. There are general methods for sampling from transforms (see Devroye 1981) and specific methods for specific distributions that do not require numerical inversion—see, for example, the methods in §1.7 of Samorodnitsky and Taqqu (1994) for stable distributions. However, these methods do not address the problem of computing the score function. In order to keep our investigation generic, we analyze a method in which numerical transform inversion is used both for sampling and for evaluation of the score. We precompute the cumulative distribution function on a grid and use this approximate distribution function in generating samples. The spacing in this grid is the first of the three sources of error listed above.

The problem of numerical inversion of Fourier and Laplace transforms has been studied extensively. Relative to this literature, our contribution lies in analyzing how the errors in transform inversion affect simulation results when transforms are used for sampling and as an ingredient in sensitivity estimation. There are many inversion techniques to choose from and many choices to make in combining these techniques with simulation—for example, whether to invert the transform for a density or the transform for its distribution function. We will see that the application to simulation motivates certain choices. Given our focus on applying LRM, we have also imposed an important consistency condition between the sampling method and the method used to approximate the score function—namely, that the expectation of the approximate score function be zero with respect to the approximating distribution, just as the exact score function has zero mean with respect to the exact distribution. This constraint drives some of the details of our approach. Indeed, this constraint would not ordinarily hold if one combined an arbitrary sampling method with a score function approximated through transform inversion.

The rest of this paper is organized as follows. Section 2 provides more detailed background and motivation for our investigation, and it introduces examples to which we return in our numerical tests. Section 3 lays out the steps in our method, including numerical transform inversion, approximate sampling, and sensitivity estimation. Section 5 presents our error analysis for both expectations and their sensitivities. Section 6 provides a summary algorithm and illustrates our results through numerical examples. Longer proofs are deferred to an electronic companion, which is available as part of the online version at http://or.journal.informs.org/. Some of the results in this paper appear in preliminary form in Glasserman and Liu (2010).

2. Background and Motivation

To fix ideas, let the random variable $X$ have a density $g_\theta$ that depends on a parameter $\theta$, and consider the expectation of $V(X)$, for some nonnegative function $V$. For example, $X$ may represent the price of an underlying asset (or its logarithm) at some future date, and $V(X)$ the discounted payoff of an option on the asset. To stress the dependence of the expected value of $V(X)$ on the parameter $\theta$, we subscript the expectation by $\theta$ and write

$$E_\theta[V(X)] = \int V(x)g_\theta(x)\,dx.$$  

In using LRM to estimate the sensitivity with respect to $\theta$, the key LRM identity is

$$\frac{d}{d\theta}E_\theta[V(X)] = E_\theta\left[V(X)\frac{\hat{g}_\theta(x)}{g_\theta(x)}\right], \tag{1}$$

with $\hat{g}_\theta$ denoting the derivative of $g_\theta$ with respect to the parameter $\theta$. This follows from bringing the derivative inside the integral and then multiplying and dividing by $g_\theta$. When this identity holds (and it does under mild regularity conditions—see Asmussen and Glynn 2007, Proposition 7.3.5), the expression inside the expectation on the right in (1) provides an unbiased estimator of the sensitivity on the left. We will write this estimator as $V(X)S_\theta(X)$ with

$$S_\theta(x) = \frac{\hat{g}_\theta(x)}{g_\theta(x)}$$

the score function. For a function $V(X_1,\ldots,X_m)$ of independent random variables $X_1,\ldots,X_m$, each with density $g_\theta$, the LRM estimator of the derivative of $E_\theta[V(X_1,\ldots,X_m)]$ is

$$V(X_1,\ldots,X_m)S_\theta(X_1,\ldots,X_m),$$

$$S_\theta(x_1,\ldots,x_m) = \sum_{i=1}^m \frac{\hat{g}_\theta(x_i)}{g_\theta(x_i)}. \tag{2}$$

The LRM estimator is easy to use when the density $g_\theta$ and its derivative are readily available. Moreover, it has the important practical feature of allowing the estimation of sensitivities for many different functions $V$ using a single score function $S_\theta$, and it does not require differentiability of $V$. The drawback of LRM is that it often suffers from large variance. As explained in the introduction, here we
investigate the application of LRM when the density is not explicitly available but is known through its characteristic function or through its Laplace transform.

Financial applications provide an important source of motivation for this investigation. In the most extensively studied models of asset prices with jumps, the price of an asset at time $T$ is given by $A_T = A_0 \exp(aT + X_T)$, where $X_T$ is the time-$T$ value of a Lévy process with $X_0 = 0$, and $A_0$ and $a$ are constants. A Lévy process has stationary independent increments, so its increments have infinitely divisible distributions; such distributions are often specified through their characteristic functions, via the Lévy-Khinchine formula (as in, e.g., Sato 1999, p. 37). For example, in the variance gamma (VG) model of Madan et al. (1998) with parameters $\rho$, $\nu$, and $\theta$, the Laplace transform of $X_T$ is given by

$$L_{VG}(t) = E[e^{-tX_T}] = \left(\frac{1}{1 + t\nu - \rho^2\nu^2/2}\right)^{T/\nu}$$

(3)

for all complex $t$ in a neighborhood of the origin.

As a simple example of a discounted payoff function, we could take

$$e^{-rT} \max\{0, A_T - K\} = e^{-rT} \max\{0, A_0e^{\theta T + X_T} \} = V(X_T).$$

(4)

Of particular interest are sensitivities with respect to $A_0$ (the option’s delta) and $\rho$, which is a type of volatility parameter. As discussed in Madan et al. (1998), $\theta$ determines the mean and the sign of the skewness of the distribution, $\nu$ determines its kurtosis. Through a change of variables, we can move $A_0$ and $a$ into the density and thus into the Laplace transform.

In the normal inverse Gaussian (NIG) model introduced by Barndorff-Nielsen (1998), we have with model parameters $\alpha$, $\beta$, $\mu$, and $\theta$,

$$L_{NIG}(t) = E[e^{-tX_T}] = \exp(\theta T [\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta - t)^2}] - \mu Tr)$$

(5)

for all complex $t$ in a neighborhood of the origin. We will examine sensitivities with respect to $\theta$ and $A_0$. We also note an important contrast between the VG and NIG models that is relevant to our analysis: along a vertical line in the complex plane, the modulus of $L_{NIG}$ decays exponentially, whereas that of $L_{VG}$ has a power decay.

The marginal distributions of Lévy processes are sufficiently flexible to allow fits to many sources of data, including asset returns. However, because Lévy processes necessarily have independent increments, they cannot produce serial dependence or clustering as models of log asset prices. Barndorff-Nielsen and Shephard (2001) introduce serial dependence in a process $Y$ driven by a Lévy process $X$ through a stochastic differential equation of the form

$$dY_t = -\lambda Y_t dt + dX_t,$$

(6)

$\lambda$ a constant. When $X$ is Brownian motion, this defines an ordinary Ornstein-Uhlenbeck (OU) process; using other Lévy processes for $X$ provides greater flexibility in shaping the marginals of $Y$. In this setting, even the characteristic function of $Y_T$ may be unknown. In practice, one would ordinarily simulate a discrete-time approximation to (6). By viewing $Y$ as a function of the increments of $X$, we may then estimate sensitivities of $E[V(Y_T)]$ using the score function for these increments. Thus, the methods we develop using the characteristic functions for VG and NIG models extend to OU-type processes driven by these Lévy processes.

In another extension of these models, Carr et al. (2003) apply random time changes to Lévy processes; here as well, tractability is limited to transforms of the transition laws. The same is true for the broad class of affine jump-diffusion models characterized in Duffie et al. (2000). Therefore, our method is relevant to these models as well.

It is sometimes possible to find specific methods or approximations to estimate sensitivities in specific models. The VG and NIG processes, for example, can both be represented as time-changed Brownian motions, so a score function can be constructed from the normal distribution of the Brownian motion and the mixing distribution of the random time change. Broadie and Kaya (2006) develop an exact method for certain affine processes. However, the objective of this paper is to develop and analyze a general-purpose method that requires little more than a characteristic function as input. A generic method is useful in the absence of special features of a model; it also serves as a benchmark against which to compare approximations and other methods. In subsequent work (Glasserman and Liu 2010), we have tested approximations and alternative methods using the method developed here as a benchmark.

3. Overview of the Method

In this section, we present the key steps in the method we analyze, explaining how we sample from the characteristic function and how we evaluate the score function. We first detail the one-dimensional case, then discuss the multidimensional generalization.

Let $G_\theta$ denote the distribution function associated with the density $g_\theta$. If we knew $G_\theta$, we could sample from this distribution by setting $X = G_\theta^{-1}(U)$, with $U$ uniformly distributed on $[0, 1]$. Similarly, if we knew $g_\theta$ and $\dot{g}_\theta$, we could evaluate the score function. Thus, we need methods to accomplish the following tasks, starting from the characteristic function or Laplace transform of $g_\theta$: We need to generate samples from $G_\theta$, and we need to approximate $g_\theta$ and $\dot{g}_\theta$ in order to evaluate the score function.

With these objectives in mind, we can outline our procedure as follows:

- On a finite grid of $x$ values, we precompute values of $g_\theta$, $\dot{g}_\theta$, and $\ddot{g}_\theta$. 

—Using the Abate-Whitt (1992) algorithm, each transform inversion is approximated using a finite weighted sum of transform values, given in (13).

—Each inversion is subject to two types of error: a discretization error (defined in (14)) and a truncation error (defined in (15)).

- At each simulation step, we
  —generate $\tilde{X} = \tilde{g}_0^{-1}(U), U \sim \text{Unif}[0,1]$, from our approximation $\tilde{G}_g$, as in (9); and
  —evaluate the approximate score function $\tilde{S}_g = \frac{\delta}{\delta \theta} \tilde{g}_0$ at $\tilde{X}$, using (8) and (10).

- At the end of each path, we evaluate the function $V$ and the sum of the approximate score values, as in (2).

The details of this procedure are specified in this section and the next and then summarized in an algorithm in §6.1. The rest of the paper is devoted to analyzing the details of this procedure.

The two-sided Laplace transform of a function $f$ is given by

$$L_f(t) = \int_{-\infty}^{\infty} e^{-tx} f(x) \, dx,$$

where $t$ is a complex variable. This transform is two-sided because the lower limit of integration is $-\infty$ rather than zero. For background, see Widder (1941, Ch. VI).

For the transform $L_{\delta g_0}$ of $g_0$, we suppose that the region of convergence in the complex plane includes a real interval $(\sigma_0, \sigma_0)$, where $\sigma_0 > 0$ and $\sigma_0 > 0$. (Of the examples in §2, this excludes stable distributions, although it includes tempered stable distributions.) By Widder (1941 Theorem 5b, p. 242), we have $L_{\delta g_0}(t) = L_{g_0}(t)/t$ for Re($t$) $\in (0, \sigma_0)$, and we have $L_{\delta g_0}(t) = -(\delta g_0)(t)/t$ for Re($t$) $\in (\sigma_0, 0)$, where $\delta g_0 = 1 - G_0$. The transform of $\delta g_0$, the derivative of the transform of $g_0$ under, for example, the condition in the following lemma, follows from Proposition 7.3.5 in Asmussen and Glynn (2007).

**Lemma 1.** If there is an integrable function $H$ for which $e^{-\alpha x} |\delta g_0(x)| \leq H(x)$ for all real $x$ for all $\theta$ in a neighborhood of $\theta_0$, then

$$L_{\delta g_0}(t) = \int_{-\infty}^{\infty} e^{-tx} \frac{\partial}{\partial \theta} g_0(x) \, dx = \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} e^{-tx} g_0(x) \, dx = \frac{\partial}{\partial \theta} L_{g_0}(t)$$

at $\theta = \theta_0$ where Re($t$) $\in (\sigma_0, \sigma_0)$.

Thus, from the transform of $g_0$, we get the transforms of $G_g$, $\delta g_0$, and $\tilde{g}_0$. By restricting to a subinterval if necessary, we assume that the region of convergence of $L_{\delta g_0}$ includes $(\sigma_0, \sigma_0)$.

Using numerical transform inversion (discussed in detail in §4), we can approximate the value of $G_0(x)$ at any $x$. Our overall approach will be to build a piecewise linear approximation $\tilde{G}_g$ to the function $G_g$, extract from that a piecewise-constant approximation $\tilde{g}_0$ to the density $g_0$, and then differentiate to get an approximation $\tilde{\delta g}_0$ to the differentiated density $\delta g_0$. At each step, we seek to ensure that $\tilde{G}_g$ and $\tilde{g}_0$ have the properties of a probability distribution and density, respectively, and that the approximate score function $\tilde{S}_g / \tilde{\delta g}_0$ will have expected value zero when evaluated at samples drawn from $G_g$.

To accomplish these objectives, we construct (in §4) an equally spaced grid $\{x_j, j \in J\}$ on the $x$-axis, where $J$ is a finite index set and $x_{j-1} - x_j = \delta$ for $j \in J$. At each grid point $x_j$, we use numerical transform inversion to calculate $G_j = G_g(x_j)$. Between grid points, we use linear interpolation: for any $x \in (x_{j-1}, x_j)$, we set

$$\tilde{G}_g(x) = \frac{x-x_{j-1}}{\delta} G_j + \frac{x_j-x}{\delta} G_{j-1}. \quad (7)$$

To ensure that $\tilde{G}_g$ is increasing, we require $G_j \geq G_{j-1}$, for all $j \in J$. This is not automatically guaranteed because...
of numerical error in transform inversion; we enforce this property when we discuss inversion in §4. Also, we choose grid points to make \( G_{\text{min}} \approx 0 \) and \( G_{\text{max}} \approx 1 \), where \( G_{\text{min}} \equiv G_{\text{min},j}, G_{\text{max}} \equiv G_{\text{max},j} \) with \( j_{\text{min}} = \min \{ j \in J \} \), \( x_{\text{min}} = x_{j_{\text{min}}} \), and \( j_{\text{max}} \) and \( x_{\text{max}} \) defined accordingly. For \( x < x_{\text{min}} \), we set \( G_{\theta}(x) = 0 \); for \( x > x_{\text{max}} \), we set \( G_{\theta}(x) = 1 \).

We use a uniform grid for simplicity of implementation and error analysis. It may be possible to achieve better results with a grid tailored to a specific \( G_{\theta} \) (or \( V \)), but this would require information about the distribution (or payoff) that might not be readily available. Our results are likely to carry over to settings in which the largest and smallest grid spacing remain within a constant multiple of each other as the grid is refined. One could also consider approximating the tail of the distribution by, e.g., an exponential beyond \( x_{\text{min}} \) and \( x_{\text{max}} \).

We denote by \( \hat{X} \) a random variable with distribution \( G_{\theta}(x) \). The density of \( \hat{X} \) is denoted by \( \hat{g}_{\theta}(x) \) and is given by a piecewise constant function,

\[
\hat{g}_{\theta}(x) = \begin{cases} 
(G_j - G_{j-1})/\delta, & \text{if } x \in [x_{j-1}, x_j), \ j \in J; \\
0, & \text{if } x < x_{\text{min}} \text{ or } x > x_{\text{max}}.
\end{cases} \tag{8}
\]

To sample from \( G_{\theta} \), we generate \( U \) uniformly over \( (G_{\text{min}}, G_{\text{max}}) \) and set

\[
\hat{X} = U \hat{X} + x_{j-1} G_j - x_j G_{j-1}, \tag{9}
\]

with \( j \) the index for which \( G_{j-1} \leq U < G_j \).

By sampling from \( G_{\theta}(x) \), we can estimate \( E_d[V(\hat{X})] \), with \( E_d \) indicating that \( \hat{X} \sim G_{\theta} \). In order to estimate the sensitivity \( E_s[V(\hat{X}) \hat{S}_{\theta}(\hat{X})] \), where \( \hat{S}_{\theta}(x) = \hat{g}_{\theta}(x)/\hat{g}_{\theta}(x) \) and \( \hat{g}_{\theta}(x) = \partial \hat{g}_{\theta}(x)/\partial \theta \), we compute \( \hat{g}_{\theta}(x) \) as follows:

\[
\hat{g}_{\theta}(x) = \begin{cases} 
(G_j - G_{j-1})/\delta, & \text{if } x \in [x_{j-1}, x_j), \ j \in J; \\
0, & \text{if } x < x_{\text{min}} \text{ or } x > x_{\text{max}}.
\end{cases} \tag{10}
\]

where \( G_j \approx \hat{G}_{\theta}(x_j) \) is calculated through numerical inversion of the transform of \( \hat{G}_{\theta} = dG_{\theta}/d\theta \). So, as we compute each \( G_j \) to construct the approximation \( \hat{G}_{\theta} \), we also compute \( \hat{G}_j \) to be able to evaluate \( \hat{g}_{\theta}(x) \). We fix the values \( \hat{G}_{\text{min}} = 0 \) and \( \hat{G}_{\text{max}} = 0 \); this ensures that \( \hat{g}_{\theta} \) integrates to zero over \( [x_{\text{min}}, x_{\text{max}}] \), and thus that the expected score is zero in the sense that

\[
\int_{-\infty}^{\infty} \hat{S}_{\theta}(x) \hat{g}_{\theta}(x) \, dx = \int_{x_{\text{min}}}^{x_{\text{max}}} \hat{g}_{\theta}(x) \, dx = 0.
\]

Once these values are computed and stored, sampling is easy and fast (it is \( O(\log |J|) \) using binary search), so the key question is the quality of the approximation; i.e., the difference between \( E_d[V(\hat{X})] \) and \( E_d[V(X)] \), and the difference between \( E_s[V(\hat{X}) \hat{S}_{\theta}(\hat{X})] \) and \( E_s[V(X)\hat{S}_{\theta}(X)] \). These differences have several sources, including numerical errors in transform inversion and discretization errors in the approximation \( \hat{G}_{\theta} \).

In the multidimensional case, \( X \) is a vector \((X_1, \ldots, X_m)\) of independent random variables. For example, in a \( L \)-driven model, these could be the increments of an underlying \( L \)-process. We sample each \( X_i, i = 1, \ldots, m \), from its distribution using (9). Because the \( X_i \) are independent, their joint density is the product of their marginal densities, and the score function \( S_{\theta} \) for the vector is the sum of the score functions of the individual components.

### 4. The Fourier-Series Method for Laplace Inversion

Abate and Whitt (1992) derived and analyzed a Fourier-series inversion formula for the one-sided Laplace transform, and we follow their approach and build on their analysis. Cai et al. (2007) extend the Abate-Whitt algorithm to two-sided Laplace transforms, which yields the following expression for a continuous function \( f \) in terms of its transform \( L_f \), for any \( \sigma \) in the interval of convergence \((\sigma, \sigma)\):

\[
f(x) = \frac{e^{\sigma x}}{\pi} \int_0^\infty \left( \Re[L_f(\sigma + i\omega)] \cos(\omega x) - \Im[L_f(\sigma + i\omega)] \sin(\omega x) \right) \, d\omega. \tag{11}
\]

We abbreviate this formula as \( f(x) = I_{\sigma, x}(L_f) \). We can (and will) allow the choice of \( \sigma \) to depend on \( x \).

Employing the trapezoidal rule to numerically evaluate the infinite integral in (11) with a step size \( h \) gives the approximation

\[
I_{\sigma, x}^h(L_f) = \frac{h e^{\sigma x}}{2\pi} L_f(\sigma) + \frac{h e^{\sigma x}}{\pi} \sum_{k=1}^{N} \left( \Re[L_f(\sigma + ikh)] \cos(khx) - \Im[L_f(\sigma + ikh)] \sin(khx) \right). \tag{12}
\]

See Abate and Whitt (1992, pp. 18–19) for a discussion of the advantages of the trapezoidal rule over higher-order quadrature rules in this setting. As in Abate and Whitt (1992), we truncate the infinite sum in (12); keeping only the first \( N \) terms of the series in (12) yields

\[
I_{\sigma, x}^{N,h}(L_f) = \frac{h e^{\sigma x}}{2\pi} L_f(\sigma) + \frac{h e^{\sigma x}}{\pi} \sum_{k=1}^{N} \left( \Re[L_f(\sigma + ikh)] \cos(khx) - \Im[L_f(\sigma + ikh)] \sin(khx) \right). \tag{13}
\]

We call \( T_p = Nh \) the truncation point.

By applying this method to \( L_{\theta}(\theta) \), we obtain the approximations \( I_{\sigma, x}^N(L_{\theta}) \) and \( I_{\sigma, x}^{N,h}(L_{\theta}) \). The discretization error at \( x \) resulting from step size \( h \) is

\[
ed_h(x) = I_{\sigma, x}(L_{\theta}) - g_{\theta}(x); \tag{14}
\]
we show in Appendix B that \( \epsilon'_\sigma(x) \geq 0 \). The truncation error is

\[
e'_\sigma(x) = I_{\sigma,x}^{N,L}(L_0) - I_{\sigma,x}^{N,L}(L_{-\infty}). \tag{15}
\]

Thus, \( I_{\sigma,x}^{N,L}(L_0) \geq g_0(x) + e'_\sigma(x) \). We similarly define the discretization error \( e^{\sigma}_{\sigma_0} = I_{\sigma_0,x}^{N,L}(L_{-\infty}) - I_{\sigma_0,x}^{N,L}(L_0) \) and the truncation error \( e'_\sigma(x) = I_{\sigma,x}^{N,L}(L_{-\infty}) - I_{\sigma,x}^{N,L}(L_0) \) in approximating \( \hat{g}_0(x) \). Note that the truncation error could be negative, so \( I_{\sigma,x}^{N,L}(L_0) \), as well as \( I_{\sigma,x}^{N,L}(L_{-\infty}) \), could be negative too.

We will apply the Fourier-series method in a way that ensures monotonicity of \( G_j, j \in J \), and ensures that \( G_{l_{\text{min}}} \) approaches 0 and \( G_{l_{\text{max}}} \) approaches \(-\infty\) and \(+\infty\), respectively. First, we make the following observation about the behavior of the inversion method at extreme values of \( x \):

**Proposition 1.** For any \( \sigma \in (0, \sigma_0) \),

\[
I_{\sigma,x}^{N,L}(L_{\sigma_0}) \to 0 \quad \text{as} \quad x \to -\infty;
\]

for any \( \sigma \in (\sigma_0, 0) \),

\[
I_{\sigma,x}^{N,L}(L_{\sigma_0}) \to 0 \quad \text{as} \quad x \to +\infty.
\]

**Proof.** From the formulas for \( I_{\sigma,x}^{N,L}(L_{\sigma_0}) \) and \( I_{\sigma,x}^{N,L}(L_{\sigma_0}) \), we see that \( I_{\sigma,x}^{N,L}(L_{\sigma_0}) = O(e^{\sigma_\infty}) \) as \( x \to -\infty \) and \( I_{\sigma,x}^{N,L}(L_{\sigma_0}) = O(e^{\sigma_\infty}) \) as \( x \to +\infty \), which yields the conclusion. \( \square \)

From this result, we see that in order for the \( G_j \) to approach 0 and 1 at extreme values of \( x_{\text{min}} \) and \( x_{\text{max}} \), we can pick \( \sigma_\in \in (0, \sigma) \) and \( \sigma_{-} \in (\sigma, 0) \), and let

\[
G_j = \begin{cases} I_{\sigma,x}^{N,L}(L_{\sigma_0}), & \text{if } x_{\text{min}} \leq x_j \leq 0; \\ 1 - I_{\sigma,x}^{N,L}(L_{\sigma_0}), & \text{if } x_j > 0. \end{cases} \tag{16}
\]

The transforms and their approximations may diverge if \( \sigma_{-} \) is chosen equal to \( \sigma_0 \) or \( \sigma_+ \) is chosen equal to \( \sigma \) (recall that the interval \((\sigma_0, \sigma)\) determines the region of convergence of \( L_{\sigma_0} \)). To ensure stability, the values of \( \sigma_{-} \) and \( \sigma_+ \) should not be chosen right next to the boundaries; with this qualification, we have found that the results are not very sensitive to the choices of \( \sigma_{-} \) and \( \sigma_+ \). The question of choosing the vertical line in the complex plane for the inversion integral is investigated in Cai et al. (2007).

For the monotonicity of the \( G_j \), we will use the following property of the Fourier-series method.

**Proposition 2.** Let \( f \) be a density and \( F \) its distribution function. Suppose the interval \((\sigma_0, \sigma)\) is within the region of convergence of \( L_{\sigma_0} \) and \( L_{\sigma_0} \), where \( \sigma_1 < 0 \) and \( \sigma_2 > 0 \). Then for any \( \sigma \in (0, \sigma_2) \),

\[
\frac{d}{dx} I_{\sigma,x}^{N,L}(L_{\sigma}) = I_{\sigma,x}^{N,L}(F_{\sigma,x}). \tag{17}
\]

Similarly, if \( \tilde{F} = 1 - F \), then for any \( \sigma \in (\sigma_1, 0) \),

\[
\frac{d}{dx} I_{\sigma,x}^{N,L}(L_{\sigma}) = -I_{\sigma,x}^{N,L}(F_{\sigma,x}). \tag{18}
\]

Because \( I_{\sigma,x}^{N,L}(L_{\sigma}) = g_0(x) + e'_\sigma(x) \) and \( e'_\sigma(x) \geq 0 \), we may conclude that \( I_{\sigma,x}^{N,L}(L_{\sigma}) \) is nonnegative for all sufficiently large \( N \) at any point at which \( g_0(x) \) is strictly positive. From Proposition 2, we see that nonnegativity of \( I_{\sigma,x}^{N,L}(L_{\sigma}) \) implies monotonicity of \( I_{\sigma,x}^{N,L}(L_{\sigma_0}) \) and \( I_{\sigma_0,x}^{N,L}(L_{\sigma_0}) \). In practice, we do not know how large \( N \) needs to be, so we apply the following rule: if it happens that \( G_{j_{\sigma_0}} < G_{j_{\sigma_0}-1} \) for some \( j_{\sigma_0} \), we simply let \( G_{j_{\sigma_0}} = G_{j_{\sigma_0}-1} \) to make \( G_j, j \in J \) a monotonically increasing sequence. (Alternatively, one might increase \( N \) adaptively, but we have not explored this possibility.) The steps we use to construct the sequence \( G_j \) are as follows:

1. Let \( x_0 = E_0[X] = -L_{\sigma_0}(0) \) and compute \( G_0 \) using (16).
2. Let \( x_j = x_0 + j\delta \) and \( x_{-j} = x_0 - j\delta \). Compute \( G_{\sigma_{-j}} \) by (16). At any \( x_j \) close to zero (say \( |x_j| < \epsilon \) for some tolerance \( \epsilon \) much smaller than \( \delta \)), we use a very large truncation point to ensure accuracy as we switch from one case in (16) to the other. After getting \( G_j, G_{-j} \), we enforce monotonicity by adjusting their values according to the following rule: If \( G_j < G_{j-1} \), then set \( G_j = G_{j-1} \); if \( G_{-j} > G_{-j-1} \), then set \( G_{-j} = G_{-j-1} \).
3. We continue for \( j = 1, 2, \ldots \) until we find \( j_{\text{max}} > 0 \) and \( j_{\text{min}} < 0 \) such that \( G_{j_{\text{max}}} \approx 1 \) and \( G_{j_{\text{min}}} \approx 0 \); we set \( J = \{ j_{\text{min}} + 1, \ldots, j_{\text{max}} - 1 \} \) and use \( x_j, j \in J \) as our grid.

In the next section, we discuss the errors in estimating expectations \( E_0[V(X)] \) (which we refer to generically as prices) and sensitivities \( E_0[V(X)S_j(X)] \). It will be important to keep in mind that our calculation (16) uses \( \sigma_{-} \in (\sigma_0, 0) \) in computing values at \( x > 0 \), and uses \( \sigma_{+} \in (0, \sigma) \) for all \( x < 0 \).

## 5. Error Analysis

### 5.1. Error Analysis for Prices

In this section, we analyze the error in estimating a price, i.e., the difference between \( E_0[\hat{V}(\hat{X})] \) and \( E_0[V(X)] \). For simplicity, we let

\[
I_{\sigma,x}^{N,L}(L_{\sigma_0}) = \begin{cases} I_{\sigma,x}^{N,L}(L_{\sigma_0}), & \text{if } x_{\text{min}} \leq 0, \\ 0, & \text{if } x > 0, \end{cases} \tag{19}
\]

and let \( e_j(x) = 1[x > 0]e_{\sigma_0}^j(x) + 1[x \leq 0]e_{\sigma_0}^j(x) \) and \( e_j(x) = 1[x > 0]e_{\sigma_0}^j(x) + 1[x \leq 0]e_{\sigma_0}^j(x) \), where \( 1\{\cdot\} \) is the indicator function.

We can decompose the error using

\[
|E_0[\hat{V}(\hat{X})] - E_0[V(X)]| = \left| \int_{x_{\text{min}}}^{x_{\text{max}}} V(x)\tilde{g}_0(x) \, dx - \int_{-\infty}^{\infty} V(x)g_0(x) \, dx \right| \leq \left| \int_{x_{\text{min}}}^{x_{\text{max}}} V(x)\tilde{g}_0(x) \, dx - \int_{-\infty}^{\infty} V(x)I_{\sigma,x}^{N,L}(L_{\sigma_0}) \, dx \right| \tag{20}
\]

\[
+ \left| \int_{-\infty}^{\infty} V(x)(I_{\sigma,x}^{N,L}(L_{\sigma_0}) - g_0(x)) \, dx \right|. \tag{21}
\]
We will analyze (21) first and then turn to (20). If \( V \geq 0 \), then \( e_d \geq 0 \) implies that
\[
\left| \int_{-\infty}^{\infty} V(x) (\mathcal{L}^N_{\mathcal{L}_\theta} - g_{\theta})(x) \right| = \left| \int_{-\infty}^{\infty} V(x) e_d(x) + e_r(x) \right| \leq \int_{-\infty}^{\infty} V(x) e_d(x) dx + \int_{-\infty}^{\infty} V(x) e_r(x) dx.
\]
(22)

To bound the error, we impose the following condition on \( L_{\mathcal{L}_\theta} \):

**Assumption 1.** For any \( \sigma \) in \((\sigma_r, \sigma_\alpha)\), as \( \omega \to \infty \),
\[
|\text{Re}[L_{\mathcal{L}_\theta}(\sigma + i\omega)]| = O(\lambda^{-\alpha} \text{Re}[L_{\mathcal{L}_\theta}(\sigma + i\omega)])
\]
and
\[
|\text{Im}[L_{\mathcal{L}_\theta}(\sigma + i\omega)]| = O(\lambda^{-\alpha} \text{Im}[L_{\mathcal{L}_\theta}(\sigma + i\omega)])
\]
uniformly in \( \lambda \geq 1 \), for some \( \alpha_k > 0 \) and \( \alpha_l > 0 \).

This assumption is broadly applicable. For example, it holds if \( \text{Re}[L_{\mathcal{L}_\theta}(\sigma + i\omega)] \) and \( \text{Im}[L_{\mathcal{L}_\theta}(\sigma + i\omega)] \) are regularly varying functions of \( \omega \) with negative indices, or if \( -\log(\text{Re}[L_{\mathcal{L}_\theta}(\sigma + i\omega)]) \) and \( -\log(\text{Im}[L_{\mathcal{L}_\theta}(\sigma + i\omega)]) \) are regularly varying functions with positive indices. Regular variation makes precise the notion of a power decay (see, e.g., Bingham et al. 1987 for background), so Assumption 1 is consistent with both a power decay and an exponential decay of the real and imaginary parts of the Laplace transform in the indicated directions. In particular, this assumption holds for the variance gamma and normal inverse Gaussian models we use in our numerical tests.

We impose the following condition on the payoff function \( V \):

**Assumption 2.** For \( x > 0 \), \( 0 \leq V(x) \leq C_x e^{x^\gamma} \), and for \( x < 0 \), \( 0 \leq V(x) \leq C_x e^{x^\gamma} \), for some constants \( C_x > 0 \), \( \gamma \in (0, -\sigma_i) \), and \( \gamma \in (-\sigma_r, 0) \). Also, we assume \( V \) to be differentiable at all but finitely many points \( x \).

This assumption is more than sufficient to ensure that \( E_\theta[V(X)] \) exists, and it is satisfied by standard option payoffs, such as the call option payoff in (4). This assumption together with Lemma 1 also ensures the validity of the LRM identity in (1) through Proposition 7.3.5 of Asmussen and Glynn (2007).

For fixed \( \sigma_+ \in (\sigma_r, 0) \) and \( \sigma_- \in (0, \sigma_r) \), let
\[
M_\lambda(T_p) = \left| L_{\mathcal{L}_\theta}(\sigma_+ + iT_p) \right| \quad \text{and} \quad M(T_p) = \max\{M_+(T_p), M_-(T_p)\}.
\]

We can now state the following result, which is proved in Appendix B.

**Proposition 3.** Under Assumptions 1 and 2, we can find \( \sigma_- \in (\sigma_r, 0) \) and \( \sigma_+ \in (0, \sigma_r) \) such that
\[
\int_{-\infty}^{\infty} V(x) e_d(x) dx = O(e^{-C/h}),
\]
for some constant \( C > 0 \), and
\[
\int_{-\infty}^{\infty} V(x) e_r(x) dx = O(M(T_p)).
\]

Through (22), this determines the order of (21). We next decompose (20) as
\[
\int_{-\infty}^{\infty} V(x) \hat{g}_\theta(x) dx - \int_{-\infty}^{\infty} V(x) I_{\mathcal{L}_\theta}^N(h(L_{\mathcal{L}_\theta})) dx \leq \int_{-\infty}^{\infty} V(x) I_{\mathcal{L}_\theta}^N(h(L_{\mathcal{L}_\theta})) dx + \int_{-\infty}^{\infty} V(x) \hat{g}_\theta(x) dx - I_{\mathcal{L}_\theta}^N(h(L_{\mathcal{L}_\theta})) dx.
\]

For the last term, we have the following result, proved in Appendix C:

**Lemma 2.** If \( V \) is bounded on the interval \([x_{j-1}, x_j]\), then
\[
\int_{x_{j-1}}^{x_j} V(x) \hat{g}_\theta(x) dx - I_{\mathcal{L}_\theta}^N(h(L_{\mathcal{L}_\theta})) dx = O(\delta^3).
\]

If, furthermore, \( V \) is differentiable, then
\[
\int_{x_{j-1}}^{x_j} V(x) \hat{g}_\theta(x) dx - I_{\mathcal{L}_\theta}^N(h(L_{\mathcal{L}_\theta})) dx = O(\delta^3).
\]

Through this lemma, we arrive at the following result:

**Proposition 4.** If \( V \) is bounded on \([x_{\min}, x_{\max}]\) and differentiable at all but finitely many points,
\[
\int_{x_{\min}}^{x_{\max}} V(x) \hat{g}_\theta(x) dx - I_{\mathcal{L}_\theta}^N(h(L_{\mathcal{L}_\theta})) dx = O(\delta^3).
\]

For the two tail errors over \((-\infty, x_{\min})\) and \((x_{\max}, \infty)\), we have the following proposition, proved in Appendix D:

**Proposition 5.** If \( G_{\min} = \epsilon \), then
\[
\int_{-\infty}^{x_{\min}} V(x) I_{\mathcal{L}_\theta}^N(h(L_{\mathcal{L}_\theta})) dx = O(e^{(\sigma_- + v_-)/\sigma_-})
\]
and if \( G_{\max} = 1 - \epsilon \), then
\[
\int_{x_{\max}}^{\infty} V(x) I_{\mathcal{L}_\theta}^N(h(L_{\mathcal{L}_\theta})) dx = O(e^{(\sigma_- + v_-)/\sigma_-}).
\]

Because both \((\sigma_- + v_-)/\sigma_-\) and \((\sigma_- + v_-)/\sigma_-\) are strictly between 0 and 1, this result indicates that
\[
\int_{-\infty}^{x_{\min}} V(x) I_{\mathcal{L}_\theta}^N(h(L_{\mathcal{L}_\theta})) dx \quad \text{and} \quad \int_{x_{\max}}^{\infty} V(x) I_{\mathcal{L}_\theta}^N(h(L_{\mathcal{L}_\theta})) dx
\]
are negligible compared to
\[
\int_{x_{\min}}^{x_{\max}} V(x) (\hat{g}_\theta(x) - I_{\mathcal{L}_\theta}^N(h(L_{\mathcal{L}_\theta}))) dx.
\]

With this specification, we can combine Propositions 3 and 4 to quantify the pricing error:

**Theorem 1.** Under Assumptions 1 and 2,
\[
|E_\theta[V(\hat{X})] - E_\theta[V(X)]| = O(\delta^3 + e^{-C/h} + M(T_p)).
\]
5.2. Error Analysis for Sensitivities

We now extend our analysis to the error in estimating the sensitivities, bounding \( |E_g[V(\hat{X})\hat{S}_\theta(\hat{X})] - E_g[V(X)S_\theta(X)]| \). Much as in the previous section, we define

\[
N^N, h(L_{\bar{\theta}}) = \begin{cases} 
N^N, h(L_{\bar{\theta}}), & \text{if } x \leq 0 \\
N^N, h(L_{\bar{\theta}}), & \text{if } x > 0,
\end{cases}
\]

and we let \( \hat{e}_g(x) = I_1[x > 0]E_{\sigma_g}(x) + I_1[x \leq 0]E_{\sigma_g}^1(x) \) and \( \hat{e}_s(x) = I_1[x > 0]E_{\sigma_s}(x) + I_1[x \leq 0]E_{\sigma_s}^1(x) \).

We bound the error in the sensitivity estimate as

\[
|E_g[V(\hat{X})\hat{S}_\theta(\hat{X})] - E_g[V(X)S_\theta(X)]| \leq \int_{-\infty}^{\infty} V(x)\hat{\theta}_\theta(x) dx - \int_{-\infty}^{\infty} V(x)\hat{\theta}_\theta(x) dx \leq \int_{-\infty}^{\infty} V(x)\hat{\theta}_\theta(x) dx - \int_{-\infty}^{\infty} V(x)N^N, h(L_{\bar{\theta}}) dx + \int_{-\infty}^{\infty} V(x)(N^N, h(L_{\bar{\theta}}) - \hat{\theta}_\theta(x)) dx.
\]

This bound is very similar to the one we used for the error in the price estimate, but now with derivatives of \( g_\theta \). Much as in Assumption 1, we impose

ASSUMPTION 3. For any \( \sigma \) in \((\sigma_-, \sigma_+)\), as \( \omega \to \infty \),

\[
|\text{Re}[L_{\bar{\theta}}(\sigma + i\lambda\omega)]| = O(\lambda^{-\hat{C}}|\text{Re}[L_{\bar{\theta}}(\sigma + i\omega)]|)
\]

and

\[
|\text{Im}[L_{\bar{\theta}}(\sigma + i\lambda\omega)]| = O(\lambda^{-\hat{C}}|\text{Im}[L_{\bar{\theta}}(\sigma + i\omega)]|)
\]

uniformly in \( \lambda \geq 1 \), for some \( \hat{\alpha}_\delta > 0 \) and \( \hat{\alpha}_s > 0 \).

For fixed \( \sigma_- \) and \( \sigma_+ \), let

\[
\hat{M}_\hat{\sigma}(T_p) = |L_{\bar{\theta}}(\sigma_\pm + iT_p)| \quad \text{and} \quad \hat{M}(T_p) = \max\{\hat{M}_\hat{\sigma}(T_p), \hat{M}_\hat{\sigma}(T_p)\}.
\]

With these assumptions and definitions, the previous analysis carries over with appropriate modification (the function \( g_\theta \) now replaced by \( \hat{\theta}_\theta \)), leading to the following result:

THEOREM 2. Under Assumptions 2 and 3, using the same \( \sigma_- \) and \( \sigma_+ \), as in Proposition 3,

\[
\int_{-\infty}^{\infty} V(x)|\hat{\theta}_\theta(x)| dx = O(e^{-\hat{C}/h}),
\]

for some positive constant \( \hat{C} \),

\[
\left|\int_{-\infty}^{\infty} V(x)\hat{\theta}_s(x) dx\right| = O(\hat{M}(T_p))
\]

and

\[
|E_g[V(\hat{X})\hat{S}_\theta(\hat{X})] - E_g[V(X)S_\theta(X)]| = O(\delta^2 + e^{-\hat{C}/h} + \hat{M}(T_p)).
\]

5.3. The Multidimensional Case

Similar conclusions can be reached for the multidimensional case on the error analysis. First, for the payoff function \( V(x_1, \ldots, x_n) \), we assume

ASSUMPTION 4. \( V(x_1, \ldots, x_m) \leq V_1(x_1) \cdots V_m(x_m) \) for some functions \( V_1, \ldots, V_m \) satisfying Assumption 2.

Each \( V_i \) in this condition is bounded by the exponential of a linear function. When a Lévy process is used to model a log asset price, as in (4), Assumption 4 allows us to bound a path-dependent payoff using linear functions of the underlying asset, and this should cover most cases of practical interest. Under this assumption, the error analysis in the multidimensional case is very similar to that in the one-dimensional case. The details are in the electronic companion to this paper.

6. Algorithm and Numerical Examples

In this section, we encapsulate our analysis into an overall algorithm and then illustrate the theoretical results with numerical examples. We investigate the convergence rate of our method as we vary the grid spacing \( \delta \), the step size \( h \), and the truncation parameter \( T_p \), using our error estimates to guide the relative magnitudes of the changes in the parameters.

6.1. Algorithm

In the previous sections, we have focused on the bias in estimating prices and sensitivities. As a measure of overall simulation error, we use mean square error (MSE), which is the sum of the squared bias and the estimator variance. If we use \( N_s \) simulation trials, then the MSE for the price estimate is

\[
\text{MSE}_{\text{price}} = (\delta^2 + e^{-C/h} + \hat{M}(T_p))^2 + \frac{\text{Var}_{\text{price}}}{N_s},
\]

and for the sensitivity, the MSE is

\[
\text{MSE}_{\text{sen}} = (\delta^2 + e^{-C/h} + \hat{M}(T_p))^2 + \frac{\text{Var}_{\text{sen}}}{N_s},
\]

where \( \text{Var}_{\text{price}} \) and \( \text{Var}_{\text{sen}} \) denote the variance per replication of the price estimate and sensitivity estimate, respectively. To drive each MSE to 0, we need to decrease \( \delta \) and \( h \) and increase \( T_p \) and \( N_s \). For efficiency, we should do so at rates consistent with their impact on the MSE.

Algorithm

INITIALIZE \( \delta, h, T_p, \) and \( N_s \)
REPEAT until stopping condition is satisfied:
1. BUILD grid \( \{x_i\} \) and compute \( \hat{G}_\theta, \hat{S}_\theta, \) and \( \hat{\theta}_\theta \) at these grid points;
2. FOR replications \( n = 1, \ldots, N_s \)
   (a) GENERATE \( \hat{X}_i \);
   (b) EVALUATE \( V(\hat{X}_i) \) and \( \hat{S}_\theta(\hat{X}_i) \);
3. RETURN sample means of $V(\hat{X})$ and $V(\hat{X})\tilde{S}_n(\hat{X})$

4. UPDATE parameters: $\delta \rightarrow \delta^i$, $h \rightarrow h^i$, $T_p \rightarrow T_p^i$,

   $N_s \rightarrow N_s^i$

The grid construction is detailed in Steps 1–3 in §4; the function evaluations and sampling mechanism are specified in (7)–(10). The theoretical analysis in §5 allows us to specify the parameter update rule in the last step to maximize the rate of decrease of the mean square error. For example, from (23) and (24), we see that if we increase the number of replications by a factor of $c$, (so $N_s \rightarrow cN_s$), then we should refine the grid size by a factor of $c^{1/4}$ (so $\delta \rightarrow \delta/c^{1/4}$). However, the relative allocation of effort between these parameters and the truncation point $T_p$ depends on the underlying model through $M(T_p)$ and $M(T_p^i)$; we illustrate this point in our examples, where we make the update rule explicit.

The iterative refinement of the computational parameters requires a stopping condition. Ideally, one would like to stop when a target level of accuracy and precision is reached. The variance per replication is easily estimated using the sample variance. The bias is more difficult to estimate because the exact value of the price or sensitivity is unknown; a simple and standard approach is to use the difference in estimates from one iteration to the next as an estimate of the remaining bias. The bias and variance estimates can then be combined to estimate the MSE, and the algorithm stopped once the MSE falls below a desired tolerance.

Our asymptotic results do not provide much guidance on initializing the computational parameters in the first step of the algorithm. Similar issues arise across many applications of simulation. In discretizing a diffusion, for example, one seldom if ever has solid grounds for choosing the initial size of the time step. Even in the absence of bias, choosing an initial number of replications requires some problem-specific information about the magnitude of the variance. Although theoretical support for initial parameter values can be difficult to find, practical considerations often dictate at least rough values for the number of replications, the number of grid points, and integration parameters $h$ and $T_p$, as in our examples below.

We use the VG and NIG models as test cases. These two models exhibit qualitatively different dependence on $T_p$; this will be reflected in the rates we use to vary the error parameters and in the overall convergence rate of the MSE. The VG and NIG models have been studied extensively in previous work and offer some tractability that makes them useful as test cases.

6.2. VG Model

We begin by considering the VG model with the call option payoff in (4). From a VG process $X$, we define an asset price by setting $A_T = A_0 \exp(aT + X_T)$, for some initial price $A_0$ and parameter $a$. Formula (22) in Madan et al. (1998) gives

$$a = r + \frac{1}{\nu} \log(1 - \theta\nu - \rho^2\nu/2),$$

with $\rho$, $\nu$, and $\theta$ parameters of the model. Madan et al. (1998) provide an option pricing formula for the VG model, which is useful in evaluating the error in our simulation estimates.

The Laplace transform of $X_T$ appears in (3). Its region of convergence is the vertical strip in the complex plane that intersects the real axis on the interval

$$\left(\frac{\theta\nu - \sqrt{\theta^2\nu^2 + 2\rho^2\nu^2}, \theta\nu + \sqrt{\theta^2\nu^2 + 2\rho^2\nu^2}}{\rho^2\nu}\right).$$

For any $\sigma$ in this interval, $[L_{VG}(\sigma + i\omega)]$ has a power decay (as $\omega \to \infty$) with rate $2T/\nu$. Thus,

$$M(T_p) = O(T_p^{-2T/\nu}),$$

and the MSE for the price estimate is therefore

$$\text{MSE}_{\text{price, VG}} = (O(\delta^2 + e^{-C/h} + T_p^{-2T/\nu}))^2 + \frac{\text{Var}_{\text{price}}}{N_s}.$$  

To reduce the MSE, we would like, ideally, to increase $T_p$ and $N_s$ and decrease $\delta$ and $h$ in a way that equates the magnitude of the error reduction in each source of error. Otherwise, we would be devoting too much computational effort to some parts of the algorithm and insufficient effort to others. This yields the following update rule for the algorithm in §6.1:

UPDATE: $T_p \rightarrow \sqrt{10}T_p$, $\delta \rightarrow 10^{-7/2}\delta$,

$N_s \rightarrow 10^{2T/\nu}N_s$, $h^{-1} \rightarrow h^{-1} + T \log 10/C\nu$.

(Our choice of $C$ is specified in the proof of Proposition 3.) With these changes, the RMSE (the square root of the MSE) for the price estimate should decrease by a factor of $10^{7/\nu}$.

The number 10 could be replaced with any other constant in the update rule and in this conclusion. All of these statements should be understood in the big-$O$ sense provided by our results.

Although our focus is on the impact of $T_p$ and the other parameters on the convergence of the RMSE, the tail behavior identified in (26) also provides information about the density of $X_T$: the decay rate of the transform characterizes the smoothness of the density. In particular, from (26) and Feller (1971, Lemma 4, p. 517), we find that if $2T/\nu \leq 1$, then the density of $X_T$ does not have an integrable derivative and may fail to be differentiable. In this case, LRM would not be applicable, even if we had a convenient expression for the density. We therefore require $2T/\nu > 1$.

In our examples, we use the following values for the VG process and the call option payoff: $A_0 = 100$, $r = 5%$. 

Table 1. European call prices and derivatives for VG model.

<table>
<thead>
<tr>
<th>ν</th>
<th>Call price</th>
<th>A₀-derivative</th>
<th>ρ-derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.2669</td>
<td>0.7282</td>
<td>23.0434</td>
</tr>
<tr>
<td>0.5</td>
<td>10.9292</td>
<td>0.6927</td>
<td>28.5971</td>
</tr>
</tbody>
</table>

\[ T = 1, \; \rho_0 = 0.2, \; \theta = -0.15, \text{ and we set } K = 100 \text{ or } 125. \text{ We compare results at } \nu = 1 \text{ and } \nu = 0.5. \text{ Using the formula in Madan et al. (1998) for the prices of European call options, we get the values in Table 1, against which we compare the simulation estimates.}

To test our sensitivity estimates, we calculate sensitivities with respect to the model parameter \( p \) and the initial price \( A_0 \) of the underlying asset. By applying finite difference approximations to the formula for option prices, we get the derivative values in Table 1.

To apply LRM, we need to move the dependence on \( A_0 \) and \( \rho \) into the density; recall from (25) that \( a \) is a function of \( \rho \). We therefore work with the random variable \( \log A_0 + a T + X_T \), whose Laplace transform is \( A_0^a \exp(-a T) L_{\text{VG}}(t) \). For the parameter \( A_0 \), the Laplace transform of the partial derivative is \(-iL_{\text{VG}}(t)/A_0; \text{ for the parameter } \rho, \text{ the Laplace transform of the derivative is } \partial(A_0^a \exp(-a T) L_{\text{VG}}(t))/\partial \rho. \text{ In both cases, we get } M(T_p) = O(T_p^{-2}) \), so the sensitivity MSE is

\[
\text{MSE}_{\text{sen, VG}} = \left( \delta^2 + \frac{C}{h} + T_p^{-2} \right)^2 + \frac{\text{Var}_{\text{sen}}}{N_s}. \tag{28}
\]

Recall that we have required \( 2T/\nu > 1 \) for the differentiability of the density, and this ensures that the exponent of \( T_p \) is negative.

The impact of the truncation point \( T_p \) in the sensitivity MSE (28) differs from that in the price MSE (27) and results in a slower overall rate of convergence. For example, with \( \nu = 1 \), we get \( 2T/\nu = 2 \), so the optimal RMSE for the price is \( O(T_p^{-2}) \), whereas for the sensitivity it is \( O(T_p^{-1}) \). Thus, to decrease the price RMSE by a factor of 10, we increase the truncation point by a factor of \( \sqrt{10} \), but to decrease the sensitivity RMSE by a factor of 10 we increase the truncation point by a factor of 10:

**UPDATE:** \( T_p \to 10T_p, \; \delta \to 10^{-1/2} \delta, \; N_s \to 10N_s, \; h^{-1} \to h^{-1} + \log 10/C. \)

A corresponding rule applies for \( \nu = 0.5 \).

Table 2 shows numerical results for price estimates with \( \nu = 1 \). From each row to the next, we multiply \( T_p \) by \( \sqrt{10} \) and change the other parameters at the corresponding rates. The initial values are set to make the modulus of \( \nu \) reasonably small (some number between 0.1 and 0.5) and \( J \) have several points. In the “Error” column, we report the absolute difference between the simulation mean and the formula price. The column labeled “SE” shows the standard error of each estimate. To estimate the standard errors precisely, we simulate all rows for 5E7 replications. We then multiply the standard error obtained by 100 for the first row, 10 for the second row and 1 for the third row. In general agreement with our analysis, the order of magnitude of the price error decreases by roughly a factor of 10 from each row to the next. In order to get reliable estimates for our comparison, we use a larger number of replications than would be optimal under our analysis. In practice, we would try to set the value of \( N_s \) to make the standard error approximately equal to the bias.

Our results are calculated using MATLAB running on a laptop computer with a 1.6 GHz Pentium M processor and 1 G of RAM. Even at our largest values of \( |J| \), building the grid takes less than a tenth of a second. Once the grid is built, we can run 100,000 replications in under a second. Because we tabulate the transform inversion results, the only computational parameter that affects the computing time per replication is the number of grid points \(|J|\). As a point of comparison, with \(|J| = 87 \) (as in the third row of Table 2), sampling from the grid takes approximately 3.7 times as long as sampling from the normal distribution using MATLAB’s built-in implementation of the inverse normal distribution. The sampling time could potentially be accelerated using a cut-point method or alias table.

Table 2 also shows numerical results for the sensitivities with \( \nu = 1 \). The sensitivity errors decrease by approximately \( \sqrt{10} \) from one row to the next, in line with our analysis, although the convergence of the derivative with respect to \( \rho \) is a bit faster than expected.

Table 3 shows results for \( \nu = 0.5 \). In this case, the modulus of the Laplace transform decays more quickly, so we increase \( T_p \) by a factor of \( \sqrt{10} \) from one row to the next. This should decrease the price error by a factor of 10 and the sensitivity error by a factor of \( 10^{3/4} \approx 5.6 \) in each case.

Table 2. Results for prices and derivatives on \( A_0 \) and \( \rho \) in VG model, with \( \nu = 1 \).

| \( T_p \) | \( \delta \) | \( N_s \) | \( |J| \) | Price | \( A_0 \)-derivative | \( \rho \)-derivative |
|--------|--------|--------|--------|--------|-----------------|-----------------|
|        |        |        |       | Mean    | Error | SE   | Mean    | Error | SE   |
| 10     | 0.5    | 5E3    | 12    | 13.073  | 1.806 | 0.2999      | 0.559         | 0.169 | 0.0323 |
| 31.6   | 0.16   | 5E5    | 37    | 11.588  | 0.321 | 0.0201      | 0.772         | 0.043 | 0.0025 |
| 100    | 0.05   | 5E7    | 87    | 11.300  | 0.032 | 0.0018      | 0.743         | 0.014 | 0.0002 |

Note. The top panel is for \( K = 100 \), and the bottom panel is for \( K = 125 \).
The results in the tables are roughly in line with these predictions.

### 6.3. NIG Model

For an NIG process $X$ at time $T$ with parameters $(\alpha, \beta, \theta, \mu)$, the Laplace transform is given in (5). The region of convergence is the vertical strip in the complex plane that intersects the real line on the interval $[\beta - \alpha, \beta + \alpha]$. Whereas the Laplace transform of the VG process exhibits a power decay, that of the NIG process exhibits an exponential decay; more precisely, $| L_{\text{NIG}}(\sigma + i0) | \sim \exp(-\theta T \omega)$ as $\omega \to \infty$, for any fixed $\sigma$ in the convergence interval.

To define an asset price $A_T$ from $X_T$ as in (4), the appropriate value of $a$ is

$$a = r - \mu - \theta \left( \sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + 1)^2} \right).$$

(29)

For the parameters, we use the following values from Këllezi and Webber (2004):

$$A_0 = 100 \quad K = 100 \quad r = 0.1 \quad T = 1 \quad \alpha = 28.42141 \quad \beta = -15.08623 \quad \theta = 0.31694 \quad \mu = 0.05851.$$

We also consider the case $K = 125$. As in Këllezi and Webber (2004), numerical integration produces a call price of 11.3599 for these parameters. Numerical differentiation yields 0.8124 as the derivative with respect to $A_0$ and 5.8165 as the derivative with respect to $\theta$.

By following the same steps as in the VG example, we find that the MSEs for the price and the two sensitivities for the NIG process all have the same order of convergence, given by

$$\text{MSE}_{[\text{price, sen}, \text{NIG}]} = \left( O(\delta^2 + e^{-C/h} + e^{-\theta T_p}) \right)^2 + \frac{\text{Var}_{[\text{price, sen}, \text{NIG}]} }{N_t}.$$  

(30)

Notice that the dependence on $T_p$ now shows an exponential decay rather than a power decay. So, to target a factor of 10 reduction in the RMSE, we use the following rule:

**UPDATE:**

$T_p \rightarrow T_p + \log 10/\theta C$, $\delta \rightarrow 10^{-1/2} \delta$, $N_t \rightarrow 10^{2N_t}$, $h^{-1} \rightarrow h^{-1} + \log 10/C$.

The resulting increment in $T_p$ is around 7.265 in our example.

### 6.4. A Path-Dependent Option

We now turn to a multidimensional example and consider the pricing of a discretely monitored Asian option in the

---

**Table 3.** Results for prices and derivatives on $A_0$ and $\rho$ in VG model, with $\nu = 0.5$.

| $T_p$ | $\delta$ | $N_t$ | $|J|$ | Price Mean | Error | SE | $A_0$-derivative Mean | Error | SE | $\rho$-derivative Mean | Error | SE |
|-------|----------|-------|------|-----------|--------|---|----------------------|--------|---|----------------------|--------|---|
| 10    | 0.5      | 5E3   | 8    | 13.179    | 2.250  | 0.286 | 0.529                | 0.163  | 0.0296 | 37.472                | 8.875  | 3.638 |
| 17.8  | 0.16     | 5E5   | 23   | 11.307    | 0.378  | 0.0209 | 0.752                | 0.060  | 0.0024 | 26.670                | 1.927  | 0.358 |
| 31.6  | 0.05     | 5E7   | 67   | 10.964    | 0.035  | 0.0020 | 0.704                | 0.011  | 0.0002 | 28.461                | 0.136  | 0.036 |
| 10    | 0.5      | 5E3   | 8    | 4.066     | 2.129  | 0.1945 | 0.344                | 0.133  | 0.0216 | 35.835                | 10.288 | 2.7194 |
| 17.8  | 0.16     | 5E5   | 23   | 2.399     | 0.462  | 0.0105 | 0.254                | 0.043  | 0.0012 | 24.129                | −1.418 | 0.2510 |
| 31.6  | 0.05     | 5E7   | 67   | 1.982     | 0.045  | 0.0009 | 0.217                | 0.007  | 0.0001 | 25.453                | −0.094 | 0.0243 |

**Note.** The top panel is for $K = 100$, and the bottom panel is for $K = 125$.

---

**Table 4.** Results for prices and derivatives on $A_0$ and $\theta$ in NIG model.

| $T_p$ | $\delta$ | $N_t$ | $|J|$ | Price Mean | Error | SE | $A_0$-derivative Mean | Error | SE | $\theta$-derivative Mean | Error | SE |
|-------|----------|-------|------|-----------|--------|---|----------------------|--------|---|----------------------|--------|---|
| 20    | 0.25     | 1E4   | 9    | 13.082    | 1.722  | 0.1315 | 0.7641                | 0.0483 | 0.0179 | 2.075                | 3.741  | 0.7400 |
| 27.3  | 0.079    | 1E6   | 22   | 11.501    | 0.141  | 0.0107 | 0.8048                | 0.0076 | 0.0019 | 5.463                | 0.353  | 0.0665 |
| 34.5  | 0.025    | 1E8   | 66   | 11.373    | 0.014  | 0.0010 | 0.8116                | 0.0008 | 0.0002 | 5.784                | 0.032  | 0.0068 |
| 20    | 0.25     | 1E4   | 9    | 2.234     | 1.208  | 0.0542 | 0.2148                | 0.0298 | 0.0104 | 4.356                | −1.231 | 0.4404 |
| 27.3  | 0.079    | 1E6   | 22   | 1.148     | 0.123  | 0.0036 | 0.1931                | 0.0081 | 0.0008 | 5.345                | −0.242 | 0.0318 |
| 34.5  | 0.025    | 1E8   | 66   | 1.038     | 0.013  | 0.0003 | 0.1858                | 0.0008 | 0.0001 | 5.557                | −0.030 | 0.0032 |

**Note.** The top panel is for $K = 100$, and the bottom panel is for $K = 125$. 

---
NIG model. The payoff of the Asian option is a function of the average level of the underlying asset over \( m \) equally spaced averaging dates and is given by

\[
\max \left( \frac{1}{m} \sum_{k=1}^{m} A_k, 0 \right),
\]

where

\[
A_k = A_0 \exp \left( a \Delta t + \sum_{l=1}^{k} X_{\Delta t, l} \right) = A_{(k-1) \Delta t} \exp (a \Delta t + X_{\Delta t, k})
\]

and the \( X_{\Delta t, k}, k = 1, \ldots, m \), are independent increments of the NIG process. To formulate the problem in the setting of §5.3, we set

\[
V(x_1, x_2, \ldots, x_m) = \max \left( \frac{1}{m} \sum_{k=1}^{m} \exp \left( a \Delta t + \sum_{l=1}^{k} x_l \right) - K, 0 \right).
\]

It is easy to see that this function satisfies Assumption 4.

As in Equation (30), the MSE is

\[
\text{MSE}_{\text{[price, sen].NIG}}^{\text{Mult}} = \frac{\left( O(\delta^2) + e^{-C/\delta} + e^{-\theta \Delta t} \right)^2 + \frac{\text{Var}_{\text{[price, sen].NIG}}^{\text{Mult}}}{N_j}}{N_j}. \tag{31}
\]

In order to measure the error in our method, we need benchmark values of the prices and sensitivities against which to compare. We estimate accurate values using 108 simulation replications and exact sampling from the NIG distribution using its representation as a continuous mixture of normals. We use finite difference approximations with common random numbers to estimate sensitivities. This would be prohibitively time consuming in practice but is necessary to get accurate benchmark values for comparison. We use \( m = 12 \) steps in the average, which corresponds to monthly price fixings with a maturity of \( T = 1 \) year. We set \( K = 100 \). We estimate a price of 6.335, a sensitivity to \( A_0 \) of 0.7525, and a sensitivity to \( \theta \) of 3.71. We believe these estimates are accurate to the number of decimal places reported.

We set the initial values \( \delta = 0.1 \) and \( T_p = 50 \) to make the modulus of the Laplace transform reasonably small. The initial value of \( h \) is obtained by setting \( e^{-C/\delta} = \delta^2 \). From each row to the next, we use the following rule:

**UPDATE:** \( T_p \rightarrow T_p + \log 10/\theta \Delta t \), \( \delta \rightarrow 10^{-1/2} \delta \),

\[
N_j \rightarrow 10^2 N_j, \quad h^{-1} \rightarrow h^{-1} + \log 10/C.
\]

The results are shown in Table 5 and are in general agreement with our theoretical analysis. The derivative with respect to \( \theta \) in the first row has a very large error; we attribute this to the small value of \( |J| \) in the first row.

### 6.5. Integrated OU Process

The Asian option gives rise to a multidimensional problem because the payoff of the option is path dependent. We now consider an example that is multidimensional because of the dynamics of the underlying asset, rather than the form of the payoff. As in Barndorff-Nielsen and Shephard (2001), we introduce a process \( Y \) with the dynamics in (6), in which \( X \) is a Lévy process. In our example, we take \( X \) to be an NIG process and interpret \( Y \) as the logarithm of the price of an underlying asset on which we price a standard European call option. This reduces to a problem of estimating the expectation of a function of \( Y_t \) and its sensitivities.

We use a Euler scheme for the evolution of \( Y_t \), and thus replace (6) with the recursion

\[
Y_{t+\Delta t} = (1 - \lambda \Delta t) Y_t + X_{t+\Delta t} - X_t.
\]

If we set \( \Delta t = T/m \), then

\[
Y_T = \sum_{k=1}^{m} (1 - \lambda \Delta t)^{k-1} X_{\Delta t, k}, \tag{32}
\]

where \( X_{\Delta t, k}, k = 1, \ldots, m \), are i.i.d. increments of \( X \) over consecutive intervals of length \( \Delta t \). Because (32) represents \( Y_T \) as a linear combination of independent NIG increments, it would allow us to write the Laplace transform of the Euler approximation to \( Y_T \) in terms of the NIG Laplace transform; however, for the purpose of testing our results, we will treat this as a multidimensional problem with a payoff that depends on \( X_{\Delta t, k}, k = 1, \ldots, m \).

We use the NIG parameters from §6.3 with \( K = 100 \), and we take \( \lambda = 2 \) and \( m = 12 \). By forcing \( E[e^{\theta T} Y_T] = 1 \), we get

\[
a = r - \frac{\theta \Delta t}{T} \sum_{k=1}^{m} \left( \sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + (1 - \lambda \Delta t)^{k-1})^2} \right). \tag{33}
\]

The payoff function is \( V(x_1, x_2, \ldots, x_m) = \max \{A_0 \exp (aT + y_T) - K, 0\} \). where \( y_T = \sum_{k=1}^{m} (1 - \lambda \Delta t)^{k-1} x_k \). This \( V \) satisfies Assumption 4, which leads us to

\[
\text{MSE}_{\text{[price, sen].NIG}}^{\text{OU}} = \frac{\left( O(\delta^2) + e^{-C/\delta} + e^{-\theta \Delta t} \right)^2 + \frac{\text{Var}_{\text{[price, sen].NIG}}^{\text{OU}}}{N_j}}{N_j}. \tag{33}
\]

As in §6.4, we estimate benchmark values through extensive simulation with 108 replications. This gives a price of 9.896, a sensitivity to \( A_0 \) of 0.9261, and a sensitivity to \( \theta \) of 1.61. Table 6 illustrates the convergence of the price and derivative estimates computed through transform inversion, and these are in general agreement with our theoretical analysis.
7. Summary
We have analyzed and tested a method for estimating sensitivities by simulation using the likelihood ratio method when the underlying density is known only through its characteristic function or Laplace transform. The method uses numerical transform inversion and incurs several types of error; we have presented results on the convergence rates of these errors and illustrated these results through numerical examples. The main determinant of the overall convergence rate is the truncation point used in the transform inversion. The convergence rate determined by the truncation point can vary qualitatively between models and, within a given model, can be different for expectations and their sensitivities. In our examples, the variance gamma model exhibits polynomial convergence for both expectations and sensitivities, with a slower rate for the sensitivities; the normal inverse Gaussian model exhibits exponential convergence for both expectations and sensitivities.

8. Electronic Companion
An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Acknowledgments
The authors thank the associate editor and referees for many comments and suggestions that helped improve the clarity of this paper. This research is supported in part by NSF grant DMS-0410234.

References

Table 6. European call prices and derivatives in OU model.

| $T_p$ | $\delta$ | $N_j$ | $|J|$ | Price Mean | Price Error | Price SE | $A_\theta$-derivative Mean | $A_\theta$-derivative Error | $A_\theta$-derivative SE | $\theta$-derivative Mean | $\theta$-derivative Error | $\theta$-derivative SE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 0.1 | 1E4 | 10 | 13.155 | 3.259 | 0.0517 | 1.052 | 0.126 | 0.0341 | –3.31 | 4.91 | 0.9648 |
| 137 | 0.032 | 1E6 | 31 | 9.958 | 0.062 | 0.0063 | 0.905 | 0.021 | 0.0044 | 1.37 | 0.24 | 0.1392 |
| 224 | 0.01 | 1E8 | 95 | 9.920 | 0.006 | 0.0006 | 0.924 | 0.002 | 0.0005 | 1.59 | 0.02 | 0.0151 |