

Estimating Greeks in simulating Lévy-driven models

Paul Glasserman

Columbia Business School, 403 Uris Hall, New York, NY 10027, USA;
email: pg20@columbia.edu

Zongjian Liu

Department of Industrial Engineering and Operations Research,
Columbia University, 313A Mudd Building, New York,
NY 10027, USA; email: zl2115@columbia.edu

We develop methods for estimating price sensitivities by simulation for Lévy-driven models. The methods combine pathwise derivatives and likelihood ratio method estimators with alternative approaches to approximating and simulating Lévy processes. We develop estimators based on exact sampling of increments, time-change representations of Lévy processes, saddlepoint approximations to the score functions of the increments, compound Poisson approximations and compound Poisson approximations with Brownian approximations to small jumps. We discuss the relative merits of these various alternatives, both in theory and in practice, and we illustrate their use through examples.

1 INTRODUCTION

Jump processes have been widely proposed and analyzed as models of asset prices, exchange rates, interest rates, commodities and other financial variables. Compared with pure-diffusion processes, models with jumps are often found to provide a better fit to option prices, to return distributions (producing distributions with higher kurtosis) and to time-series properties. Lévy processes and, more generally, models driven by Lévy processes provide the natural framework for building continuous-time models with jumps (see Cont and Tankov (2004) and Schoutens (2003) for extensive overviews of Lévy-driven models in finance).

Option prices in Lévy models can sometimes be calculated through numerical transform inversion, and sometimes through the numerical solution of partial integro-

The authors thank the referees for detailed comments and suggestions that helped improve the paper.

differential equations; but some combinations of payouts and models require Monte Carlo simulation. In applying Monte Carlo simulation, generating paths of the underlying Lévy process can present a challenge, particularly for infinite activity processes, which jump infinitely many times in a finite time interval. Various exact and approximate methods have been proposed for simulating specific cases and broad categories of Lévy processes.

This paper focuses on the problem of estimating price sensitivities, that is, Greeks, in simulating Lévy-driven models. Price sensitivities are fundamental to hedging option positions against changes in underlying assets. In a model with jumps, exact replication of an option through delta hedging is typically impossible in theory as well as in practice; nevertheless, Greeks remain important. Various criteria have been proposed to gauge hedge effectiveness and define optimal hedging in this context; the optimal hedge is often a combination of the traditional delta and a separate jump term (see Chapter 10 of Cont and Tankov (2004)). Sensitivities to model parameters provide a mechanism for hedging model risk and changes in parameter values unrelated to changes in underlying prices; these types of sensitivities are as relevant to jump models as they are to diffusion models. Price sensitivities are also potentially useful in model calibration, which involves minimizing over model parameters to bring model prices as close as possible to market prices.

There are two broad approaches to estimating price sensitivities by Monte Carlo: methods that differentiate paths and methods that differentiate probabilities. In the first category, the pathwise method estimates derivatives by calculating the derivative of an option payout with respect to the underlying asset, and multiplying this by the derivative of the underlying path with respect to the parameter of interest. In the second category, the likelihood ratio method (LRM) works by differentiating the density of the underlying model. Depending on the application, each method has potential advantages (see Chapter 7 of Glasserman (2004) for general background on these methods).

In this paper we develop pathwise and LRM estimators for Lévy-driven models. Although these general approaches to estimating Greeks by Monte Carlo are well established, the Lévy context introduces challenges and solutions that merit special consideration. Any method for approximate simulation of Lévy processes gives rise to methods for approximate simulation of Greeks. However, derivatives from an approximation are not well defined until we specify how the approximation changes with the parameters of interest. This flexibility can have a significant impact on the quality of the sensitivity estimates. We will also see that the choice of approximation method is often more important for the quality of sensitivity estimates than for the prices themselves.

We summarize the methods considered in this paper as follows.

- Section 3 considers cases in which exact (unbiased) derivative estimates are available. This includes delta estimates using the pathwise method, and LRM estimators when the density of the Lévy increments is known. Many Lévy models of interest can be represented as time-changed Brownian motions in which the time change is given by an increasing Lévy process; we show how this representation leads to pathwise and LRM estimators.
- Section 4 derives LRM estimators based on saddlepoint approximations. Often, the densities of Lévy increments needed for LRM are known only through their characteristic functions; saddlepoint methods provide a way of approximating the necessary densities and their derivatives from the characteristic function.
- Section 5 derives pathwise and LRM estimators from compound Poisson approximations to infinite activity Lévy processes. A compound Poisson approximation truncates the arrival of jumps to a finite rate and is thus easy to simulate; a second-order Brownian approximation to the truncated small jumps can reduce the truncation error (Asmussen and Rosinski (2001); Cont and Tankov (2004); Signahl (2003)). In its most straightforward form, the pathwise method is inapplicable with these approximations because of discontinuities in the Poisson process, but we show how to circumvent this difficulty to obtain consistent estimators. We will see that different truncation rules (and different correction terms) lead to alternative estimators. For both pathwise and LRM estimators, second-order corrections to the truncation error dramatically improve the accuracy (reduce the bias) of sensitivity estimates; indeed, in our numerical examples, these corrections are even more important for the sensitivity estimates than for the price estimates.

To provide the necessary background for these techniques, Section 2 reviews some general properties of Lévy processes and some specific examples. Section 6 summarizes our observations on the methods we study.

2 BACKGROUND ON LÉVY PROCESSES AND DERIVATIVE ESTIMATION

2.1 Lévy processes

A stochastic process X , $X(0) = 0$, is a Lévy process if it has stationary, independent increments and is continuous in probability; the last condition means that the process has no fixed jump times. Loosely speaking, a Lévy process is the sum of a deterministic drift, a Brownian motion and jump terms. This is made precise through the Lévy–Itô decomposition (see, for example, Sato (1999)).

If the total arrival rate of jumps in X is finite, X can be decomposed as:

$$X(t) = bt + \sqrt{A}W(t) + \sum_{k=1}^{N(t)} Y_k \quad (1)$$

where A and b are constants, Y_k are independent and identically distributed (iid) jump sizes, N is a Poisson process, W is a standard Brownian motion, and N , W and $\{Y_k\}$ are independent of each other. In the generalization to an infinite arrival rate of jumps, the jumps can no longer be described through a compound Poisson process and must instead be described through a Lévy measure.

We will suppose that the Lévy measure has a density q on \mathbb{R} . The integral of q over some set $C \subseteq \mathbb{R}$ may be interpreted as the arrival rate of jumps of size in C . If the integral of q over all of \mathbb{R} is some finite value m , then the jump component of X is compound Poisson with arrival rate m , and the jump sizes have probability density $q(\cdot)/m$. But our focus is on infinite activity processes for which the total mass of q is infinite. The Lévy density is required to satisfy:

$$\int (y^2 \wedge 1)q(y) dy < \infty$$

If:

$$\int |y|q(y) dy < \infty$$

then X has finite variation; the case of infinite variation is called the compensated case. Let:

$$Q(x) = \int_{y \leq x} q(y) dy \quad \text{and} \quad \bar{Q}(x) = \int_{y \geq x} q(y) dy$$

We assume throughout that $q > 0$ on all \mathbb{R} , and this condition makes Q strictly increasing.

A Lévy process X is characterized by the transform:

$$\begin{aligned} \Phi_t(s) &= E[\exp(sX(t))] \\ &= \exp\left(t\left(bs + \frac{1}{2}As^2 + \int_{\mathbb{R}} (e^{sy} - 1 - sy\mathbf{1}_{|y| \leq 1})q(y) dy\right)\right) \end{aligned} \quad (2)$$

with b and A interpreted as in (1) and q the Lévy density. (The integrand in the exponent is truncated to cover the compensated case; truncation is unnecessary if X has finite variation.) For some processes, $\Phi_t(s)$ is infinite unless s is purely imaginary, but we will assume throughout that the region of convergence (see Widder (1941)) of $\Phi_t(s)$ contains an interval $(-s_-, s_+)$, where $s_- > 0$ and $s_+ > 0$. We let $\Phi(s) = \Phi_1(s)$ and then $\Phi_t(s) = (\Phi(s))^t$. We say X has the Lévy triplet (b, A, q) . We focus on the case $A = 0$ in which X has no Brownian component.

2.2 Examples

We now turn to some examples. A standard construction of an asset price process S sets:

$$S_t = S_0 \exp(at + X(t)) \quad (3)$$

in the same way that geometric Brownian motion is constructed from ordinary Brownian motion. In (3), the constant a is chosen to make the discounted expectation of S_t equal to S_0 , ie, $E[e^{-rt} S_t] = S_0$, where r is the risk-free interest rate.

EXAMPLE 2.1 (Variance gamma model) The variance gamma (VG) model (see Madan *et al* (1998)) uses (3) with X the VG process, which can be defined by setting:

$$X(t) = X(t; \rho, \nu, \theta) = \theta \Gamma(t; 1, \nu) + \rho W(\Gamma(t; 1, \nu)) \quad (4)$$

where W is a standard Brownian motion and $\Gamma(t; 1, \nu)$ is a gamma process with mean rate 1 and variance rate ν . The gamma process is an increasing Lévy process whose increments are gamma distributed; in particular, $\Gamma(t; 1, \nu)$ has a gamma distribution with shape parameter t/ν and scale parameter ν . The VG process has finite variation.

EXAMPLE 2.2 (Normal inverse Gaussian model) The normal inverse Gaussian (NIG) model (Barndorff-Nielsen (1998)) uses (3) with X the NIG process, which can be defined by setting:

$$X(t) = X(t; \alpha, \beta, \mu, \delta) = \mu t + \beta I_G(t; \delta, \gamma) + W(I_G(t; \delta, \gamma)) \quad (5)$$

where $I_G(t; \delta, \gamma)$ is an inverse Gaussian process, an increasing Lévy process whose increments have inverse Gaussian distributions with parameters δ and γ and $\gamma = \sqrt{\alpha^2 - \beta^2}$. The NIG process belongs to the compensated case having infinite variation.

EXAMPLE 2.3 (CGMY model) This is similar to the previous two examples, but with X now taken to be the CGMY process, defined in Carr *et al* (2002). The CGMY process has finite variation for $0 < Y < 1$ and infinite variation for $1 \leq Y < 2$.

EXAMPLE 2.4 (Ornstein–Uhlenbeck processes) The specification in (3) is a standard way to construct asset price processes from Lévy processes, but it is by no means the only construction. A Lévy-driven Ornstein–Uhlenbeck process Y is defined by setting:

$$dY(t) = -\lambda Y(t) dt + dX(t) \quad (6)$$

with X a Lévy process. The resulting process Y can then be used to model volatility, for example.

In simulating Y , we may apply an Euler approximation to discretize the evolution Y with a step size $\Delta t = T/m$. This gives the approximation:

$$Y_T = \sum_{k=1}^m (1 - \lambda \Delta t)^{k-1} X_{\Delta t, k} \quad (7)$$

where $X_{\Delta t, k}$ are iid and have the same distribution as $X(\Delta t)$.

2.3 Derivative estimation

To discuss the problem of estimating price sensitivities, we first need to formulate the pricing problem. We let V denote the (discounted) payout of some derivative security. The payout is a function of $(S_{t_1}, \dots, S_{t_m})$, the level of an underlying asset S at fixed dates $t_1 < \dots < t_m$, and we sometimes write $V = V(S) = V(S_{t_1}, \dots, S_{t_m})$, depending on context. We let $\Delta_1 = t_1$ and $\Delta_i = t_i - t_{i-1}$, $i = 2, \dots, m$, and we suppose that S is a function of the independent increments $(X(t_1), X(t_2) - X(t_1), \dots, X(t_m) - X(t_{m-1}))$; each $X(t_i) - X(t_{i-1})$ has the same distribution as $X(\Delta_i)$. In a slight abuse of notation, we will write (X_1, \dots, X_m) for a vector of independent random variables in which X_i has the distribution of $X(\Delta_i)$. In estimating $E[V(S)]$, we simulate (X_1, \dots, X_m) , map these to S , and then evaluate $V(S)$.

We use θ to denote a generic parameter with respect to which we want to estimate sensitivities. The parameter θ could be S_0 in (3) or any of the parameters appearing in the examples above. More generally, θ may be a parameter of q and b , in which case we write q_θ and b_θ to stress this dependence. A dot over a variable, as in \dot{b}_θ and $\dot{q}_\theta(x)$, indicates a derivative with respect to the parameter; a prime indicates a derivative with respect to another variable, such as the argument of a density or a payout function. We sometimes subscript the expectation operator by θ , as in writing $E_\theta[V(S)]$ to stress dependence on θ . Our objective is to estimate:

$$\frac{d}{d\theta} E_\theta[V(S)]$$

As noted in the Introduction, pathwise estimates of derivatives are based on differentiating the dependence on a parameter along each path, and LRM estimators are based on differentiating probability densities. Thus, to derive pathwise estimators, we need to give $S = S_\theta$ a functional dependence on the parameter; the pathwise estimator is then:

$$\sum_{i=1}^m \frac{\partial V}{\partial S_\theta(t_i)} \frac{\partial S_\theta(t_i)}{\partial \theta}$$

This requires some smoothness in the payout (Lipschitz continuity, for example) and differentiability of the underlying asset with respect to the parameter. Recall that S

is constructed from X and θ may be a parameter of X , in which case we would differentiate S with respect to X and X with respect to θ . Thus, the key issue for the applicability of the pathwise method in the Lévy context is the interpretation and calculation of expressions of the form $dX_\theta(t)/d\theta$.

The likelihood ratio method proceeds by putting the dependence on θ in the density of X (or S), rather than in the path itself. We let $f_{t,\theta}$ denote the density of $X(t)$; Theorem 27.4 of Sato (1999) ensures that a density exists if q has infinite mass. The joint density of the vector $X = (X_1, \dots, X_m)$ of independent increments is then:

$$f_\theta(X) = f_{\Delta_1,\theta}(X_1) \cdots f_{\Delta_m,\theta}(X_m) \equiv f_{1,\theta}(X_1) \cdots f_{m,\theta}(X_m)$$

and we may write the expected payout as:

$$E_\theta[V(S)] = \int V(s) f_\theta(x) dx$$

where $x = (x_1, \dots, x_m)$ and $s = (s_1, \dots, s_m)$. This moves all the θ -dependence of X into the density. After differentiating the density, the LRM estimator eventually takes the form:

$$V(S) \mathcal{S}_\theta(X)$$

where $\mathcal{S}_\theta(\cdot)$ is called the score function. The score function and the derivation of the LRM estimator are reviewed in the next section. The key issue for the applicability of LRM in the Lévy context is the calculation of the score function.

3 EXACT METHODS

The main focus of this paper is on the derivation and analysis of approximation techniques for intractable derivative estimation problems in the Lévy context. Before proceeding to more difficult cases, we briefly discuss settings in which no approximations are needed.

3.1 Pathwise derivative for delta

Delta is the derivative of the price with respect to S_0 . In the case of (3), the pathwise derivative:

$$\frac{\partial}{\partial S_0} V(S_T) = V'(S_T) \exp(aT + X(T)) = V'(S_T) \frac{S_T}{S_0}$$

generalizes a familiar expression for delta estimates based on geometric Brownian motion.

For the multidimensional case, writing $S = (S_1, \dots, S_m)$, we have a natural extension:

$$\frac{\partial}{\partial S_0} V(S) = \sum_{i=1}^m \frac{\partial V(S)}{\partial S_i} \frac{S_i}{S_0} = (S_0)^{-1} \sum_{i=1}^m \frac{\partial V(S)}{\partial S_i} S_i$$

3.2 Parametric derivatives of random variables

We briefly review the construction of Suri (1983) leading to an expression for $dX_\theta/d\theta$ for a random variable X whose distribution depends on a parameter θ .

Suppose that the dependence on θ is given through a density f_θ and distribution function F_θ . If, say, f_θ is positive almost everywhere, we may set:

$$X_\theta = F_\theta^{-1}(U), \quad U \sim \text{Unif}[0, 1]$$

to give $X = X_\theta$ a functional dependence on θ while preserving the correct distribution F_θ at each θ . Then $F_\theta(X_\theta) = F_\theta(F_\theta^{-1}(U)) = U$. Taking the derivative with respect to θ , we have:

$$\frac{\partial}{\partial \theta} F_\theta(X_\theta) + f_\theta(X_\theta) \frac{\partial}{\partial \theta} X_\theta = 0$$

and therefore:

$$\frac{\partial}{\partial \theta} X_\theta = -(f_\theta(X_\theta))^{-1} \frac{\partial}{\partial \theta} F_\theta(X_\theta) \quad (8)$$

This solves the key step in computing the pathwise estimator if we can evaluate f_θ and \dot{F}_θ .

3.3 The likelihood ratio method with known densities

To derive the LRM estimator, we begin with a one-dimensional example in which the underlying asset is given by (3) and the payout V is a function only of S_T for some fixed time T :

$$\begin{aligned} E_\theta[V(S)] &= \int_{\mathbb{R}} V(S_0 \exp(a_\theta T + x)) f_\theta(x) dx \\ &= \int_{\mathbb{R}} V(S_0 \exp(x)) f_\theta(x - a_\theta T) dx \end{aligned} \quad (9)$$

We have written the constant a in (3) as a_θ because this coefficient will indeed typically change with the distribution of X . The rightmost expression in (9) moves all dependence on θ out of the payout V and into the density, as required for LRM.

Differentiating with respect to θ , bringing the derivative inside the integral, and writing $f'(x)$ for the derivative with respect to x , we get:

$$\begin{aligned}
 \frac{d}{d\theta} E_\theta[V(S)] &= \int_{\mathbb{R}} V(S_0 \exp(x)) \frac{\partial}{\partial \theta} f_\theta(x - a_\theta T) dx \\
 &= \int_{\mathbb{R}} V(S_0 \exp(x)) (\dot{f}_\theta(x - a_\theta T) - \dot{a}_\theta T f'_\theta(x - a_\theta T)) dx \\
 &= \int_{\mathbb{R}} V(S_0 \exp(a_\theta T + x)) \frac{\dot{f}_\theta(x) - \dot{a}_\theta T f'_\theta(x)}{f_\theta(x)} f_\theta(x) dx \\
 &= E_\theta \left[V(S) \frac{\dot{f}_\theta(X) - \dot{a}_\theta T f'_\theta(X)}{f_\theta(X)} \right] \tag{10}
 \end{aligned}$$

Conditions ensuring the validity of this derivation are given in Proposition 7.3.5 of Asmussen and Glynn (2007), for example. When the interchange of derivative and integral holds, the expression inside the expectation on the right in (10) provides an unbiased estimator of the sensitivity on the left. We will write this estimator as $V(S)\mathbf{S}_\theta(X)$, using the score function:

$$\mathbf{S}_\theta(x) = \frac{\dot{f}_\theta(x) - \dot{a}_\theta T f'_\theta(x)}{f_\theta(x)} \tag{11}$$

In the multidimensional case, we have:

$$S_{t_{i+1}} = S_{t_i} \exp(a_\theta \Delta_{i+1} + X_{i+1})$$

and the X_i are independent. The LRM estimator of the derivative of $E_\theta[V(S)]$ is then:

$$V(S)\mathbf{S}_\theta(X), \quad \mathbf{S}_\theta(x) = \sum_{i=1}^m \frac{\dot{f}_{i,\theta}(x_i) - \dot{a}_\theta \Delta_i f'_{i,\theta}(x_i)}{f_{i,\theta}(x_i)}$$

An especially important case is delta, for which $\theta = S_0$; however, S_0 is not a parameter of the Lévy process X . In this case, by changing the variable x_1 to $x_1 - \log S_0$, we get:

$$\begin{aligned}
 E_\theta[V(S)] &= \int V(\exp(a_\theta t_1 + x_1), s_2, \dots, s_m) \\
 &\quad \times f_{1,\theta}(x - \log S_0) f_{2,\theta}(x_2) \cdots f_{m,\theta}(x_m) dx_1 dx_2 \cdots dx_m
 \end{aligned}$$

So the score function is $\mathbf{S}_\theta(x) = -f'_{1,\theta}(x_1)/S_0$.

These calculations produce the LRM estimators of interest in substantial generality, at least in principle. For Lévy processes, the relevant densities and derivatives are often unknown or given by cumbersome expressions. In some cases, they are known

only through their characteristic functions. In Glasserman and Liu (2008), we use the numerical transform inversion to calculate $f_{t,\theta}(x)$, $\dot{f}_{t,\theta}(x)$ and $f'_{t,\theta}(x)$ numerically from their transforms, and we analyze the convergence order. Kienitz (2008) approximates the derivative of the density with a finite-difference approximation.

3.4 Using a time-change representation

Several important classes of Lévy processes can be represented as subordinated Brownian motions, ie, as a Brownian motion evaluated under a random time change in which the time change is an increasing Lévy process independent of the Brownian motion. See Carr and Wu (2004) for applications of this idea in option pricing. The VG, NIG and CGMY processes all admit such a representation.

Suppose, then, that we can write the Lévy process X as:

$$X(t) = \mu Y(t) + \sigma W(Y(t)) \quad (12)$$

where W is a standard Brownian motion and Y is an increasing Lévy process (a subordinator) independent of W . Such a representation is potentially useful in simulating X , particularly if it is easier to generate the increments of Y than to generate the increments of X directly. We will investigate the implications of such a representation for derivative estimation when μ , σ and $Y(t)$ may depend on a parameter θ . To stress the dependence on the parameter, we subscript the variables by θ and write, eg, $X_\theta(t)$ for $X(t)$.

3.4.1 Pathwise derivative estimation

If the parameter θ affects μ or σ but not Y , then the paths $\{X_\theta(t), 0 \leq t \leq T\}$ are differentiable in θ and we have:

$$\frac{\partial}{\partial \theta} X_\theta(t) = \dot{\mu}_\theta Y(t) + \dot{\sigma}_\theta W(Y(t))$$

If Y does depend on θ , then an attempt to differentiate X_θ pathwise in (12) would seem to require differentiating the Brownian motion W with respect to its time parameter, which is clearly impossible. However, we may write the increments of X_θ in the form:

$$X_\theta(t_{i+1}) - X_\theta(t_i) = \mu_\theta [Y_\theta(t_{i+1}) - Y_\theta(t_i)] + \sigma_\theta \sqrt{[Y_\theta(t_{i+1}) - Y_\theta(t_i)]} Z_{i+1} \quad (13)$$

where Z_1, Z_2, \dots are independent, standard normal random variables. It is now evident that we can differentiate both sides of this equation with respect to θ , provided we can differentiate Y_θ with respect to θ . Thus, somewhat paradoxically, the time-change representation allows us to differentiate X_θ pathwise at any discrete set of dates but breaks down in the continuous-time limit.

We illustrate these ideas with examples.

EXAMPLE 3.1 (VG process) In the time-change representation (4), the subordinator is a gamma process $Y(t) = \Gamma(t; 1, \nu)$. The parameter θ has a specific meaning in (4), which leads to:

$$\frac{\partial}{\partial \theta} X(t; \rho, \nu, \theta) = \Gamma(t; 1, \nu)$$

For the parameter ρ , we get:

$$\frac{\partial}{\partial \rho} X(t; \rho, \nu, \theta) = W(\Gamma(t; 1, \nu))$$

The parameter ν is more interesting because it affects the gamma process. A gamma random variable $\Gamma(\alpha, \beta)$ has the same distribution as $\beta\Gamma(\alpha, 1)$; so, for any fixed t , $\Gamma(t/\nu, \nu)$ has the same distribution as $\nu\Gamma(t/\nu, 1)$. If we make this substitution in our representation of $X(t; \rho, \nu, \theta)$, then differentiating (4) with respect to ν yields:

$$\begin{aligned} \frac{\partial}{\partial \nu} X(t; \rho, \nu, \theta) &= \theta \frac{\partial}{\partial \nu} (\nu\Gamma(t/\nu, 1)) + \rho Z \frac{\partial}{\partial \nu} \sqrt{\nu\Gamma(t/\nu, 1)} \\ &= \theta\Gamma(t/\nu, 1) + \rho Z \frac{\sqrt{\Gamma(t/\nu, 1)}}{2\sqrt{\nu}} \\ &\quad \times \left(\theta\nu + \rho Z \frac{\sqrt{\nu}}{2\sqrt{\Gamma(t/\nu, 1)}} \right) \frac{\partial \Gamma(t/\nu, 1)}{\partial \nu} \end{aligned}$$

with Z a standard normal random variable. In this expression, the derivative $\partial\Gamma(t/\nu, 1)/\partial\nu$ of a gamma random variable is obtained from (8) and is equal to:

$$\frac{\partial \Gamma(t/\nu, 1)}{\partial \nu} = \frac{t}{\nu^2} \frac{F_{\Gamma(t/\nu-1,1)}(\Gamma(t/\nu, 1)) - \psi(t/\nu) F_{\Gamma(t/\nu,1)}(\Gamma(t/\nu, 1))}{f_{\Gamma(t/\nu,1)}(\Gamma(t/\nu, 1))}$$

where $F_{\Gamma(\alpha,\beta)}$ and $f_{\Gamma(\alpha,\beta)}$ are the cumulative distribution function (CDF) and probability density function (PDF) of a gamma distribution with parameters α and β , and $\psi(x)$ is the digamma function, $\psi(x) = \Gamma'(x)/\Gamma(x)$.

The VG process admits an alternative representation as the difference of two independent gamma processes:

$$X(t; \rho, \nu, \theta) = \beta_+ \Gamma_+ \left(\frac{t}{\nu}, 1 \right) - \beta_- \Gamma_- \left(\frac{t}{\nu}, 1 \right)$$

with β_{\pm} functions of ρ, ν and θ (see Madan *et al* (1998) for details). This representation leads to alternative pathwise derivatives.

EXAMPLE 3.2 (NIG process) In the time-change representation (5) of the NIG process, the subordinator is an inverse Gaussian process, $Y(t) = I_G(t; \delta, \gamma)$. The parameter μ in (5) is straightforward. For parameters α, β and δ , we need to employ (13) to

get the derivatives of $X(t; \alpha, \beta, \mu, \delta)$ and employ (8) to get the derivative of $I_G(t; \delta, \gamma)$. The density of $I_G(t; \delta, \gamma)$ is (see Barndorff-Nielsen (1998)):

$$(2\pi)^{-1/2} t \delta e^{t\delta\gamma} x^{-3/2} \exp\left(-\frac{1}{2}(t^2\delta^2x^{-1} + \gamma^2x)\right)$$

In the case of parameter δ , for example, the derivative of the CDF of $I_G(t; \delta, \gamma)$ with respect to δ is:

$$\begin{aligned} \frac{\partial F_{I_G(t; \delta, \gamma)}(x)}{\partial \delta} &= 2t\gamma e^{2t\delta\gamma} F_{N(0,1)}\left(-\gamma\sqrt{x} - \frac{t\delta}{\sqrt{x}}\right) - \frac{t}{\sqrt{x}} f_{N(0,1)}\left(\gamma\sqrt{x} - \frac{t\delta}{\sqrt{x}}\right) \\ &\quad - \frac{t}{\sqrt{x}} e^{2t\delta\gamma} f_{N(0,1)}\left(-\gamma\sqrt{x} - \frac{t\delta}{\sqrt{x}}\right) \end{aligned}$$

where $F_{N(0,1)}$ and $f_{N(0,1)}$ are the standard normal CDF and PDF.

Madan and Yor (2005) show how to express the CGMY process as a time-changed Brownian motion. However, in this case the law of the subordinator Y is known only through its characteristic function: a case considered in Glasserman and Liu (2008). The Ornstein–Uhlenbeck example (6) does not fit the random time-change framework; however, we may directly differentiate both sides of (7) to get:

$$\frac{\partial Y_T}{\partial \theta} = \sum_{k=1}^m (1 - \lambda \Delta t)^{k-1} \frac{\partial X_{\Delta t, k}}{\partial \theta}$$

The case $\theta = \lambda$ can also be handled explicitly.

3.4.2 LRM

A time-change representation is potentially very useful for LRM estimation, particularly when the density of $Y(t)$ is known but that of $X(t)$ is not. Expression (13) represents the increments of X as a mixture of normal random variables, and this in turn leads to a “mixed” score function using Y as well as X .

To illustrate, we begin with the case in which the payout V is a function only of S_T . Write $f_{Y(t)}$ for the density of $Y(t)$. Then we have:

$$E_\theta[V(S_T)] = \int_{-\infty}^{\infty} \int_0^{\infty} V(S_0 \exp(a_\theta T + x)) f_{N(\mu y, \sigma \sqrt{y})}(x) f_{Y(T)}(y) dy dx$$

where $f_{N(\mu y, \sigma \sqrt{y})}(x)$ is the density function of $N(\mu y, \sigma \sqrt{y})$. The joint density of $X(T)$ and $Y(T)$ is $f_\theta(x, y) = f_{N(\mu y, \sigma \sqrt{y})}(x) f_{Y(T)}(y)$, and the density of $X(T)$ is:

$$f_\theta(x) = \int_0^{\infty} f_{N(\mu y, \sigma \sqrt{y})}(x) f_{Y(T)}(y) dy \quad (14)$$

To calculate the score function, we first write:

$$E_\theta[V(S_T)] = \int_{-\infty}^{\infty} \int_0^{\infty} V(S_0 \exp(x)) f_{N(\mu y, \sigma \sqrt{y})}(x - a_\theta T) f_{Y(T)}(y) dy dx$$

Now we differentiate with respect to θ to get:

$$\begin{aligned} \frac{d}{d\theta} E_\theta[V(S_T)] &= \frac{d}{d\theta} \int_{-\infty}^{\infty} \int_0^{\infty} V(S_0 \exp(x)) f_{N(\mu y, \sigma \sqrt{y})}(x - a_\theta T) f_{Y(T)}(y) dy dx \\ &= \int_{-\infty}^{\infty} \int_0^{\infty} V(S_0 \exp(x)) (f_{Y(T)}(y)) (\dot{f}_{N(\mu y, \sigma \sqrt{y})}(x - a_\theta T) \\ &\quad - \dot{a}_\theta T f'_{N(\mu y, \sigma \sqrt{y})}(x - a_\theta T) + f_{N(\mu y, \sigma \sqrt{y})}(x - a_\theta T) \dot{f}_{Y(T)}(y)) dy dx \\ &= \int_{-\infty}^{\infty} \int_0^{\infty} V(S_0 \exp(a_\theta T + x)) (f_{Y(T)}(y)) (\dot{f}_{N(\mu y, \sigma \sqrt{y})}(x) \\ &\quad - \dot{a}_\theta T f'_{N(\mu y, \sigma \sqrt{y})}(x) + f_{N(\mu y, \sigma \sqrt{y})}(x) \dot{f}_{Y(T)}(y)) dy dx \end{aligned}$$

The score function therefore is:

$$\begin{aligned} \mathcal{S}_\theta(x, y) &= \frac{\dot{f}_{N(\mu y, \sigma \sqrt{y})}(x) - \dot{a}_\theta T f'_{N(\mu y, \sigma \sqrt{y})}(x)}{f_{N(\mu y, \sigma \sqrt{y})}(x)} + \frac{\dot{f}_{Y(T)}(y)}{f_{Y(T)}(y)} \\ &\equiv \mathcal{S}_\theta(x | y) + \mathcal{S}_\theta(y) \end{aligned}$$

The multidimensional case is similar. By applying (14), and following the same reasoning as in Section 3.3, we have for vectors of increments $x = (x_1, \dots, x_m)$ and $y = (y_1, \dots, y_m)$:

$$\begin{aligned} \mathcal{S}_\theta(x, y) &= \sum_{i=1}^m \frac{\dot{f}_{N(\mu y_i, \sigma \sqrt{y_i})}(x_i) - \dot{a}_\theta \Delta_i f'_{N(\mu y_i, \sigma \sqrt{y_i})}(x_i)}{f_{N(\mu y_i, \sigma \sqrt{y_i})}(x_i)} + \frac{\dot{f}_{Y(\Delta_i)}(y_i)}{f_{Y(\Delta_i)}(y_i)} \\ &\equiv \sum_{i=1}^m (\mathcal{S}_\theta(x_i | y_i) + \mathcal{S}_\theta(y_i)) \end{aligned}$$

Although the time-change representation significantly expands the scope of models for which exact LRM is feasible (ie, models for which we know the density of Y but not of X), this generality comes at the price of increasing variance. The following proposition indicates that using the score function of X results in lower variance.

PROPOSITION 3.3 *With the notation above:*

$$\text{var}(V(X)\mathcal{S}_\theta(X, Y)) \geq \text{var}(V(X)\mathcal{S}_\theta(X))$$

PROOF Note that:

$$f'_\theta(x) = \int_0^\infty f'_{N(\mu y, \sigma \sqrt{y})}(x) f_{Y(t)}(y) dy$$

and:

$$\dot{f}_\theta(x) = \int_0^\infty (\dot{f}_{N(\mu y, \sigma \sqrt{y})}(x) f_{Y(t)}(y) + f_{N(\mu y, \sigma \sqrt{y})}(x) \dot{f}_{Y(t)}(y)) dy$$

so, denoting $Y(T)$ by Y , we have:

$$\begin{aligned} E[\mathbf{S}_\theta(X, Y) | X = x] &= \int_0^\infty \mathbf{S}_\theta(x, y) \frac{f_\theta(x, y)}{f_\theta(x)} dy \\ &= \frac{\dot{f}_\theta(x) - \dot{a}_\theta T f'_\theta(x)}{f_\theta(x)} \\ &= \mathbf{S}_\theta(x) \end{aligned}$$

by (11). Thus:

$$\begin{aligned} \text{var}[V(X)\mathbf{S}_\theta(X, Y)] &= \text{var}[E[V(X)\mathbf{S}_\theta(X, Y) | X]] + E[\text{var}[V(X)\mathbf{S}_\theta(X, Y) | X]] \\ &= \text{var}[V(X)\mathbf{S}_\theta(X)] + E[\text{var}[V(X)\mathbf{S}_\theta(X, Y) | X]] \\ &\geq \text{var}[V(X)\mathbf{S}_\theta(X)] \end{aligned}$$

A similar argument applies in the multidimensional case. \square

We now illustrate these points with examples. To apply LRM directly (without the time-change representation), we use numerical transform inversion, as in Glasserman and Liu (2008). We compare this with pathwise and LRM estimates using the time-change representation.

We use the VG and NIG models as test cases, pricing European call options with:

$$V(x) = \exp(-rT) \max\{S_0 \exp(aT + x) - K, 0\}$$

where K is the strike price, T is the maturity and r is the risk-free discount rate.

For parameters in the VG model, we use $S_0 = 100$, $K = 100$, $r = 0.05$, $T = 1$, $\rho = 0.2$, $\nu = 1$ and $\theta = -0.15$. For the NIG model, we take values from K ellezi and Webber (2004): $S_0 = 100$, $K = 100$, $r = 0.1$, $T = 1$, $\alpha = 28.42141$, $\beta = -15.08623$, $\theta = 0.31694$ and $\mu = 0.05851$. We also test a discretely monitored Asian call option using the NIG model. With monthly averaging over a one-year horizon, we get 12 steps in the average. We used this example with transform inversion in Glasserman and Liu (2008), which allows comparison with the time-change representation.

TABLE 1 Standard deviations for sensitivity estimation methods and computing times (in seconds) in parentheses.

	Number of simulations	Numerical transform inversion LRM	Time-changed LRM	Time-changed pathwise method
VG-EU	1×10^5	0.338 (2.75)	0.363 (1.26)	0.118 (1.12)
NIG-EU	1×10^6	0.066 (5.03)	0.186 (1.39)	0.016 (9.67)
NIG-AS	1×10^6	0.073 (56.61)	0.172 (18.09)	0.010 (116.21)

VG-EU = VG European call option example, NIG-EU = NIG European call option example, NIG-AS = NIG Asian call option example. The second column gives the number of replications used.

For the VG model, we consider sensitivities to ρ and run 500,000 simulation trials for all three methods. For the NIG model, we consider sensitivities to δ and run 1 million simulation trials for all three methods and for both the European call option example and the Asian call option example. In Table 1, we report the standard errors (and computing times) for the three methods and three examples. In all cases, the time-change pathwise method provides the lowest standard deviations, followed by numerical transform inversion LRM. As expected from Proposition 3.3, the time-changed LRM has higher variance than direct LRM, but the time-change score function is easier to calculate because it involves simpler probability densities. Our results are calculated using MATLAB running on a laptop computer with a 1.6 GHz Pentium M processor and 1 Gb of RAM. We have not attempted to optimize computing times in our MATLAB implementations, and relative speeds could change if the methods are implemented in a lower-level programming language.

4 SADDLEPOINT APPROXIMATION TO THE SCORE FUNCTIONS

4.1 The method

In this section we propose and test a method for approximating the score function. The method is relevant to models for which the underlying density is unknown or difficult to work with but for which the Laplace transform is available. In this setting, numerical transform inversion could be used, as in Glasserman and Liu (2008); saddlepoint approximation provides a somewhat simpler, if less accurate, alternative.

Saddlepoint approximations are typically used to approximate probability densities or distribution functions (see Jensen (1995) for an overview). We will use them to approximate score functions. Suppose that a random variable X has density f_θ and cumulant generating function (CGF):

$$K_\theta(s) = \log E_\theta[\exp(sX)]$$

assumed finite in a neighborhood of the origin. A saddlepoint expansion of $f_\theta(x)$ takes the form:

$$f_\theta(x) = \frac{1}{\sqrt{2\pi K_\theta''(\hat{s})}} \exp\{K_\theta(\hat{s}) - \hat{s}x\} \{1 + (\frac{1}{8}\lambda_{4,\theta}(\hat{s}) - \frac{5}{24}\lambda_{3,\theta}^2(\hat{s})) + \dots\} \quad (15)$$

(Jensen (1995, p. 27)), where \hat{s} is called the saddlepoint and satisfies $K_\theta'(\hat{s}) = x$. The CGF is strictly convex (unless X is a constant), so there is at most one root to this equation; the value of \hat{s} depends on θ . The higher-order terms are defined by:

$$\lambda_{n,\theta}(s) = \frac{K_\theta^{(n)}(s)}{[K_\theta''(s)]^{n/2}}$$

To approximate the score, we will differentiate an approximation to f_θ . To stress the dependence of \hat{s} on x and θ , we write it as $\hat{s} = \hat{s}_\theta(x)$. The derivative of $\hat{s}_\theta(x)$ with respect to θ is:

$$\frac{\partial \hat{s}_\theta(x)}{\partial \theta} = -(K_\theta''(\hat{s}))^{-1} \dot{K}_\theta'(\hat{s}) \quad (16)$$

(see Appendix A for details). In approximating f_θ based on (15), we can use:

$$\frac{1}{\sqrt{2\pi K_\theta''(\hat{s})}} \exp\{K_\theta(\hat{s}) - \hat{s}x\} \quad (17)$$

or:

$$\frac{1}{\sqrt{2\pi K_\theta''(\hat{s})}} \exp\{K_\theta(\hat{s}) - \hat{s}x\} \{1 + (\frac{1}{8}\lambda_{4,\theta}(\hat{s}) - \frac{5}{24}\lambda_{3,\theta}^2(\hat{s}))\} \quad (18)$$

or even include more terms. These are all saddlepoint approximations to the densities.

By taking the logarithm of both sides of (15) and differentiating with respect to θ , we get the score function $\mathbf{S}_\theta(x)$ (see Appendix A for details):

$$\mathbf{S}_\theta(x) = \frac{1}{2[K_\theta''(\hat{s})]^2} (K_\theta'''(\hat{s})\dot{K}_\theta'(\hat{s}) - K_\theta''(\hat{s})\dot{K}_\theta''(\hat{s})) + \dot{K}_\theta(\hat{s}) + T_\theta \quad (19)$$

where T_θ is the derivative of $\log\{1 + (\frac{1}{8}\lambda_{4,\theta}(\hat{s}) - \frac{5}{24}\lambda_{3,\theta}^2(\hat{s})) + \dots\}$ with respect to θ . If we drop T_θ , we obtain a first-order saddlepoint approximation to the score function:

$$\mathbf{S}_\theta(x) \approx \hat{\mathbf{S}}_\theta(x) = \frac{1}{2[K_\theta''(\hat{s})]^2} (K_\theta'''(\hat{s})\dot{K}_\theta'(\hat{s}) - K_\theta''(\hat{s})\dot{K}_\theta''(\hat{s})) + \dot{K}_\theta(\hat{s}) \quad (20)$$

The same approximation to the score results from first approximating $f_\theta(x)$ by (17) and then taking the logarithmic derivative with respect to θ . But, by starting from the saddlepoint expansion, we have a more precise derivation and more flexibility to

get a higher-order approximation by including more terms from T_θ . For example, corresponding to (18), we can approximate T_θ by:

$$T_\theta \approx \frac{\frac{1}{8}\dot{\lambda}_{4,\theta}(\hat{s}) - \frac{5}{12}\lambda_{3,\theta}(\hat{s})\dot{\lambda}_{3,\theta}(\hat{s})}{1 + (\frac{1}{8}\lambda_{4,\theta}(\hat{s}) - \frac{5}{24}\lambda_{3,\theta}^2(\hat{s}))} \cdot \frac{\partial \hat{s}_\theta(x)}{\partial \theta} \quad (21)$$

to obtain a second-order approximation. Higher-order approximations are obtained by approximating T_θ with more terms.

A basic property of any score function is that it has mean zero. However, this property need not be shared by the saddlepoint approximation $\hat{\mathbf{S}}_\theta(X)$. We would therefore like to center the approximation by replacing it with:

$$\hat{\mathbf{S}}_\theta(x) - E[\hat{\mathbf{S}}_\theta(X)] \quad (22)$$

In practice, we are unlikely to know the expected value of the approximate score, but we can estimate it using the sample mean, and subtract the sample mean to center the approximation.

4.2 Numerical examples

We now illustrate the saddlepoint approximation to the score function with examples.

EXAMPLE 4.1 (Normal distribution) The CGF of $N(\mu, \sigma^2)$ is $K(s) = \mu s + \frac{1}{2}\sigma^2 s^2$. For any x , the saddlepoint is $\hat{s} = (x - \mu)/\sigma$. In this case, the first-order saddlepoint approximation is exact for both parameters μ and σ , as is easily verified by direct calculation.

EXAMPLE 4.2 (Gamma distribution) The CGF of the gamma distribution with shape parameter α and scale parameter β is $K(s) = \alpha \log \beta - \alpha \log(\beta - s)$. For the parameter β , the first-order saddlepoint approximation gives the exact score function $\mathbf{S}_\beta(x) = \alpha/\beta - x$. For the parameter α , the first-order saddlepoint approximation is $\hat{\mathbf{S}}_\alpha(x) = \log(\beta x) - \log \alpha + 1/(2\alpha)$, while the true score function is $\mathbf{S}_\alpha(x) = \log(\beta x) - \Gamma'(\alpha)$. The difference is:

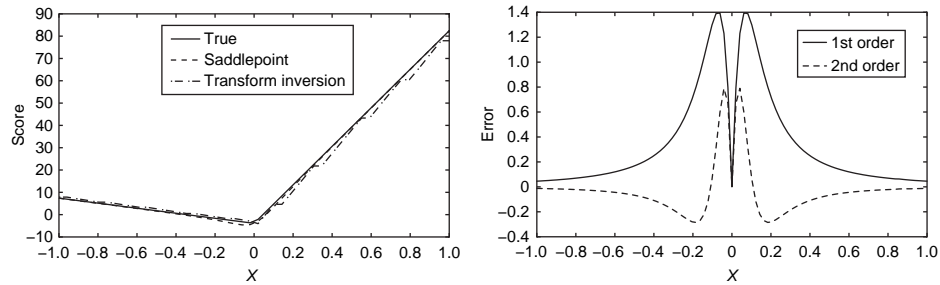
$$\hat{\mathbf{S}}_\alpha(x) - \mathbf{S}_\alpha(x) = \Gamma'(\alpha) - \log \alpha + \frac{1}{2\alpha} = O(\alpha^{-2})$$

and is thus small for large α .

EXAMPLE 4.3 (VG process) The CGF of a VG process with the parameter set (ρ, ν, θ) at time T is $K_T(s) = -T \log(1 - \theta \nu s + \rho^2 \nu \frac{1}{2} s^2)/\nu$. The saddlepoint is thus given by the root of a quadratic equation and is easy to compute. The VG density is available as an infinite integral (see formula 6 in Madan *et al* (1998)):

$$f_{\text{vg}}(x) = \int_0^\infty \frac{1}{\rho \sqrt{2\pi y}} \exp\left(-\frac{(x - \theta y)^2}{2\rho^2 y}\right) \frac{y^{T/\nu-1} \exp(-y/\nu)}{\nu^{T/\nu} \Gamma(T/\nu)} dy$$

FIGURE 1 Comparison of the true score function, the first-order saddlepoint approximation and the approximation through numerical transform inversion for parameter ρ in the VG model.



The right panel shows the errors in the first-order and second-order saddlepoint approximations to the score function.

TABLE 2 Estimates of the expected value of the approximate score function for each parameter using first-order and second-order saddlepoint approximations.

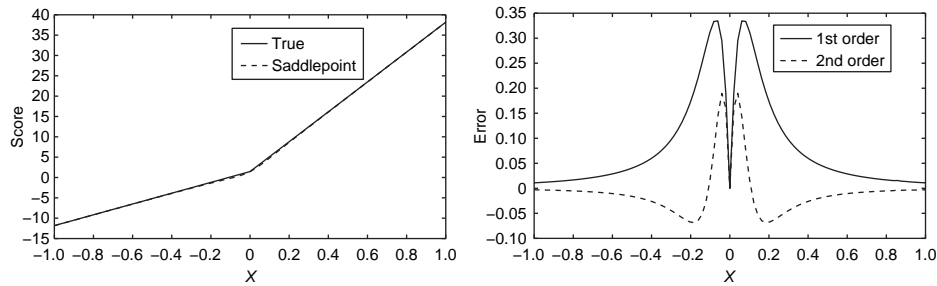
	First order	Second order
ρ	-0.799 (0.005)	-0.143 (0.005)
θ	-0.192 (0.004)	-0.034 (0.004)
ν	-0.039 (0.0002)	-0.001 (0.0003)

The numbers in parentheses are standard errors.

The derivative of the density with respect to any of the parameters ρ , ν or θ is also an infinite integral. This makes the saddlepoint approximation to the score function potentially attractive.

In the left panel of Figure 1, we consider the parameter ρ and compare the true score function, the first-order saddlepoint approximation and the approximation from numerical transform inversion (using the method in Glasserman and Liu (2008)). The parameter values are the same as in Section 3.4.2. The true score function is evaluated by truncating the infinite integrals in the density and its derivative to $[0, 20]$ and applying the “quadl” function in MATLAB with a tolerance 1×10^{-8} . The transform inversion calculations use the method in Glasserman and Liu (2008) with parameters as given there in the third row of Table 2. In the right panel of Figure 1, we graph the errors in the first-order and second-order saddlepoint approximations. In this example, including more terms in (A.6) to approximate T_θ produces a more accurate approximation.

FIGURE 2 Comparison of the true score function and the first-order saddlepoint approximation for parameter θ in the VG model.



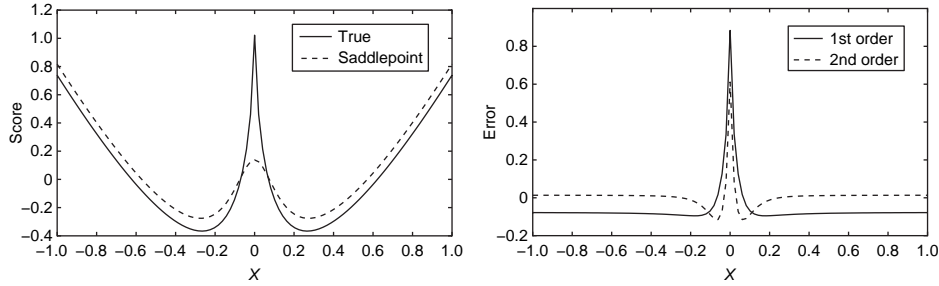
The right panel shows the errors in the first-order and second-order saddlepoint approximations to the score function.

In Figure 2 and Figure 3 on the next page, we consider the parameters θ and ν . The true score functions are obtained following the same steps as used for ρ . Overall, the approximations are excellent for θ but less accurate for ν ; in both cases, keeping more terms improves accuracy.

Next, we test the accuracy of the saddlepoint approximations in sensitivity estimation problems, applying LRM with an approximate score function. We use the same parameter values as in the example of Section 3.4.2. We estimate the expected value of the approximate score function; recall that the true score function has mean zero. The expected score is the derivative of the constant 1 with respect to the parameter. The results are reported in Table 2 on the facing page, based on 1 million replications for each case. The results show a clear bias away from zero.

To correct for this bias, we can center each approximate score function by subtracting its sample mean across all replications and then use the centered approximated score in the LRM estimator (see the discussion surrounding (22)). We test this idea with the European option example specified in Section 3.4.2. The results are summarized in Table 3 on the next page. The table estimates the bias in LRM estimates using saddlepoint approximations and centered saddlepoint approximations, relative to the reference values. The reference values are obtained through a finite-difference approximation to numerically calculated option prices. The simulation results in the table are based on 1 million replications. In all cases, the second-order approximation performs better than the first-order approximation. But the results also indicate the importance of centering the approximate score; indeed, centering appears to be even more important to improving accuracy than including additional terms in the approximation.

FIGURE 3 Comparison of the true score function and the first-order saddlepoint approximation for parameter ν in the VG model.



The right panel shows the errors in the first-order and second-order saddlepoint approximations to the score function.

TABLE 3 Estimated bias in LRM estimates in the VG model using first-order and second-order saddlepoint approximations to the score function.

	Reference values	Error in sensitivity		Centered	
		1st order	2nd order	1st order	2nd order
ρ	23.04	-10.95 (0.26)	-3.44 (0.26)	-1.95 (0.27)	-1.83 (0.27)
θ	-17.33	-2.48 (0.13)	-0.67 (0.13)	-0.32 (0.14)	-0.29 (0.14)
ν	0.55	-1.24 (0.003)	-0.08 (0.005)	-0.80 (0.01)	-0.07 (0.01)

The estimates in the last two columns center the approximate score by its sample mean. The numbers in parentheses are standard errors.

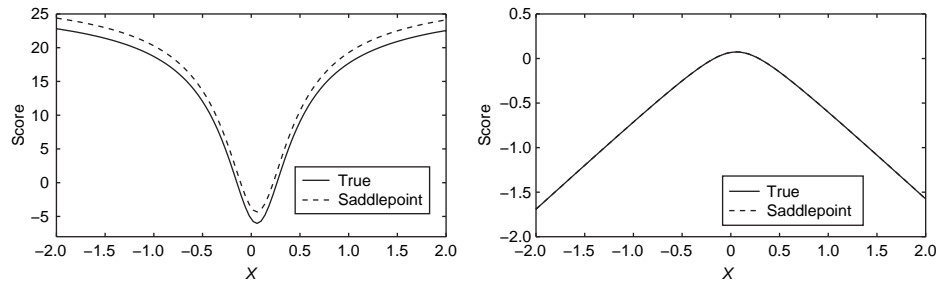
EXAMPLE 4.4 (NIG process) The CGF of an NIG process with parameters $(\alpha, \beta, \mu, \delta)$ at time 1 is:

$$K_{\alpha, \beta, \mu, \delta}(s) = \delta(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + s)^2}) + \mu s$$

For the parameter β , the first-order saddlepoint approximation (20) gives the exact score function; for parameters α, μ and δ , the accuracy of the first-order approximation varies. Explicit expressions are given in Appendix B. Figure 4 on the facing page compares the true score function and the first-order saddlepoint approximation for parameters δ and α . In the case of α , the approximation is virtually exact.

As in the VG example, we estimate the expected approximate score and use the approximate score to estimate option price sensitivities, with and without centering. In this example, we use only the first-order approximation. We use the parameters and European option specified in the example of Section 3.4.2. The option price is insensitive to μ , so we consider only α and δ .

FIGURE 4 Comparison of the true score function and the first-order saddlepoint approximation for parameters δ (left panel) and α (right panel) in the NIG model.



For parameter α , the approximation is virtually exact, so the two curves cannot be distinguished in the figure.

TABLE 4 Estimated mean of the approximate score, reference values for option price sensitivities to each parameter, and biases in LRM estimators for these sensitivities using the approximated score and the centered approximate score in the NIG model.

	Approximate score	Reference value	Sensitivity	Centered
δ	1.658 (0.005)	5.81	19.03 (0.04)	0.19 (0.05)
α	0.001 (0.00007)	-0.15	0.01 (0.0006)	0.01 (0.0006)

Results based on 1 million replications are reported in Table 4. The table shows the estimated mean of the approximate score, a reference value for each sensitivity, the estimated bias in LRM estimates using the approximate score, and the estimated bias when the approximate score is centered. As in the VG example, centering has a big impact.

5 COMPOUND POISSON AND NORMAL APPROXIMATIONS

In this section we take a fundamentally different approach to approximating Lévy processes and developing sensitivity estimates associated with these approximations. Section 3 worked with exact representations of the Lévy process or the distribution of its increments, and Section 4 used approximations to the density of the Lévy increments, and in Glasserman and Liu (2008) we started from the Laplace transform of the density. Now we approximate the dynamics of the Lévy process directly. By truncating small jumps, we arrive at a compound Poisson approximation for the jump component. Building on the work of Asmussen and Rosinski (2001) and others, we

then refine the compound Poisson approximation by making a normal (or, more generally, second-order) approximation to the truncated jump. We then develop derivative estimates associated with these approximations.

An approximation to the Lévy process does not automatically determine the derivative of an approximation option price, much less an associated estimator. The derivative is not defined until we specify how the approximation changes with the parameter of interest. Indeed, we have considerable flexibility in how we specify this parametric dependence, and this flexibility is key to the development of pathwise estimators, in particular.

In the rest of this section, we first detail the approximations, and then derive two types of pathwise estimators: a combined pathwise–LRM estimator and a pure LRM estimator. Along the way, we discuss the relative advantages of the various methods.

5.1 Truncating small jumps

Our first step is to approximate the Lévy process X by a process in which all jumps of a size smaller than ε have been removed. The arrival of jumps of magnitude greater than or equal to ε defines a Poisson process N with rate:

$$\lambda = \int_{|y| \geq \varepsilon} q_\theta(y) \, dy$$

These jumps are iid with density $q_\theta(y)/\lambda$ for $|y| \geq \varepsilon$ and 0 otherwise. The approximation takes the form:

$$X_1(t) = c_\theta t + \sum_{k=1}^{N(t)} Y_k \quad (23)$$

where Y_1, Y_2, \dots are the iid jumps. In order to keep the overall mean of the process unchanged and have $E[X_1(t)] = E[X(t)]$ for all t , we need to choose c_θ so that:

$$c_\theta + \int_{|y| \geq \varepsilon} y q_\theta(y) \, dy = b_\theta + \int_{|y| > 1} y q_\theta(y) \, dy$$

the expression on the left is the drift in (23), and the expression on the right is what we get for the original process through (2). Thus:

$$c_\theta = b_\theta - \int_{\varepsilon \leq |y| \leq 1} y q_\theta(y) \, dy$$

A different truncation level in (2) would produce a different value of b_θ but would leave c_θ unchanged.

Set $X_\varepsilon(t) = X(t) - X_1(t)$, which at $t = 1$ has CGF:

$$K_{\varepsilon, \theta}(s) = \int_{-\varepsilon}^{\varepsilon} (e^{sy} - 1 - sy) q_\theta(y) \, dy$$

and moment generating function (MGF) $\Phi_{\varepsilon,\theta}(s) = \exp(K_{\varepsilon,\theta}(s))$. By construction, the mean of this process is $E[X_\varepsilon(t)] = K'_{\varepsilon,\theta}(0) = 0$. The variance rate is:

$$\sigma_\theta^2(\varepsilon) \equiv \text{var}(X_\varepsilon(1)) = K''_{\varepsilon,\theta}(0) = \int_{-\varepsilon}^{\varepsilon} y^2 q_\theta(y) dy$$

This measures how much variance we have removed from X by removing the small jumps. To compensate, we might consider an approximation of the form:

$$X_2(t) = X_1(t) + \sigma_\theta(\varepsilon)W(t) \quad (24)$$

with W being a standard Brownian motion. Here, the small jumps of the original process have been approximated by a Brownian motion with the same variance. Asmussen and Rosinski (2001), Chapter 12 of Cont and Tankov (2004) and Signahl (2003) give conditions and criteria under which X_2 is indeed a better approximation than X_1 . An alternative to (23) (pointed out by Cont) would be to choose the mean to preserve the martingale property of the asset price (3). This adjustment is more complicated, so we focus on the simpler case of matching the mean of the Lévy process.

Our goal is to use these approximations to the Lévy processes to approximate the sensitivities. We use pathwise and LRM estimators and combinations of the two.

5.2 Pathwise method: truncated jumps

5.2.1 The method

We now turn to the derivation of sensitivity estimates. We will subscript variables by θ where we want to stress the dependence on the parameter.

In attempting to apply the pathwise method to $X_{1,\theta}$ or $X_{2,\theta}$, we immediately confront two obstacles. First, the Poisson random variable $N(t)$ is inevitably discontinuous in λ , so if λ varies with θ then changes in the parameter will introduce discontinuities in $X_{1,\theta}$ and $X_{2,\theta}$. Second, the jump-size random variables $Y_{k,\theta}$ are potentially discontinuous in θ , as a result of the jump truncation. Suppose, in particular, that we set:

$$Y_\theta = \frac{Q_{\theta,\varepsilon}^{-1}(U)}{\lambda_\theta}$$

with $Q_{\theta,\varepsilon}/\lambda_\theta$ the CDF of jumps of magnitude greater than ε . The density of Y_θ is zero on $(-\varepsilon, \varepsilon)$, so the CDF is flat on this interval and its inverse has a discontinuity. If the point of discontinuity of the inverse varies with θ , it will introduce a discontinuity in Y_θ . The discontinuity in Y_θ occurs when a change in θ causes the jump size to move from $(-\infty, -\varepsilon]$ to $[\varepsilon, \infty)$, or vice versa. There is no way to move smoothly from one of these intervals to the other. (For examples of other contexts of sensitivity

estimation with discontinuities, see, for example, Fu and Hu (1997) and Hong and Liu (2007).)

To address both issues, we separate positive and negative jumps and write the original compound Poisson process in (23) as the difference of two compound Poisson processes with positive jumps. In other words, we rewrite $X_1(t)$ as:

$$X_1(t) = c_\theta t + \sum_{k=1}^{N^+(t)} Y_k^+ - \sum_{k=1}^{N^-(t)} Y_k^- \quad (25)$$

where $N^+(t)$ and $N^-(t)$ are independent with arrival rates:

$$\lambda^+ = \int_{y \geq \varepsilon} q_\theta(y) dy \quad \text{and} \quad \lambda^- = \int_{y \geq \varepsilon} q_\theta(-y) dy$$

The positive jumps Y_k^+ have density $q_\theta(y)/\lambda^+$, $y \geq \varepsilon$, the negative jumps Y_k^- have density $q_\theta(-y)/\lambda^-$, $y \geq \varepsilon$. With this separation, we can have each $Y_{k,\theta}^+$ and $Y_{k,\theta}^-$ vary smoothly with θ .

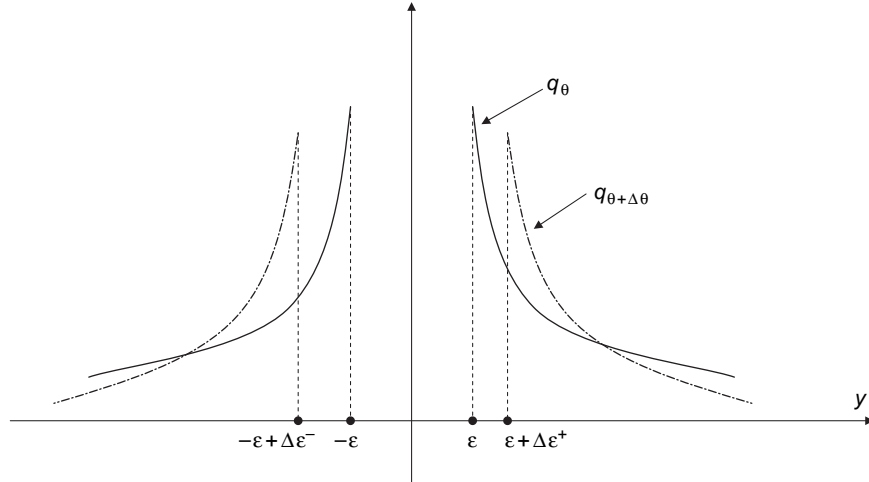
Even if the sizes of the jumps change smoothly, we still have to address potential discontinuities in the number of jumps in a fixed interval. To prevent this possibility, we would like to keep λ_θ^+ and λ_θ^- constant as θ varies. For this, we can use the flexibility we have in specifying how small jumps are truncated. In particular, we can let ε vary with θ , and we can use different truncation points ε_θ^\pm for positive and negative jumps as we move θ from its initial value. By requiring that $d\lambda_\theta^+/d\theta = 0$ and $d\lambda_\theta^-/d\theta = 0$, we obtain the derivatives of ε_θ^+ and ε_θ^- with respect to θ at ε :

$$\left. \frac{d\varepsilon_\theta^+}{d\theta} \right|_{\varepsilon_\theta^+ = \varepsilon} = \frac{\int_\varepsilon^\infty \dot{q}_\theta(y) dy}{q_\theta(\varepsilon)} \quad \text{and} \quad \left. \frac{d\varepsilon_\theta^-}{d\theta} \right|_{\varepsilon_\theta^- = \varepsilon} = \frac{\int_\varepsilon^\infty \dot{q}_\theta(-y) dy}{q_\theta(-\varepsilon)} \quad (26)$$

This construction is illustrated in Figure 5 on the facing page. As we vary θ , we move the truncation thresholds so that the area under the Lévy density on each side remains fixed. This ensures that the number of positive jumps and the number of negative jumps remain fixed, so we do not introduce a discontinuity in N_θ^+ or N_θ^- , nor do we have a jump flip from one side to the other. We have, in effect, defined a direction along which the compound Poisson approximation changes smoothly in θ .

We can define a pathwise estimator by differentiating along this direction. The resulting derivative of $X_{1,\theta}(t)$ is:

$$\frac{\partial}{\partial \theta} X_{1,\theta}(t) = \dot{c}_\theta t + \sum_{k=1}^{N^+(t)} \frac{\partial}{\partial \theta} Y_{k,\theta}^+ - \sum_{k=1}^{N^-(t)} \frac{\partial}{\partial \theta} Y_{k,\theta}^- \quad (27)$$

FIGURE 5 Variation of ε^\pm with θ , keeping λ_θ^+ and λ_θ^- independent of θ .


where:

$$\begin{aligned} \dot{c}_\theta &= \dot{b}_\theta - \int_{\varepsilon \leq |y| \leq 1} y \dot{q}_\theta(y) dy + \varepsilon q_\theta(\varepsilon) \left. \frac{d\varepsilon_\theta^+}{d\theta} \right|_{\varepsilon_\theta^+ = \varepsilon} - \varepsilon q_\theta(-\varepsilon) \left. \frac{d\varepsilon_\theta^-}{d\theta} \right|_{\varepsilon_\theta^- = \varepsilon} \\ &= \dot{b}_\theta - \int_{\varepsilon \leq |y| \leq 1} y \dot{q}_\theta(y) dy + \varepsilon \left(\int_\varepsilon^\infty \dot{q}_\theta(y) dy - \int_\varepsilon^\infty \dot{q}_\theta(-y) dy \right) \end{aligned}$$

The derivatives of $Y_{k,\theta}^\pm$ are calculated using (8).

If we start from $X_{2,\theta}(t)$, we need the derivative of $\sigma_\theta^2(\varepsilon)$, given as follows:

$$\begin{aligned} \dot{\sigma}_\theta^2(\varepsilon) &= \varepsilon^2 q_\theta(\varepsilon) \left. \frac{d\varepsilon_\theta^+}{d\theta} \right|_{\varepsilon_\theta^+ = \varepsilon} + \int_0^\varepsilon y^2 \dot{q}_\theta(y) dy \\ &\quad + \varepsilon^2 q_\theta(-\varepsilon) \left. \frac{d\varepsilon_\theta^-}{d\theta} \right|_{\varepsilon_\theta^- = \varepsilon} + \int_{-\varepsilon}^0 y^2 \dot{q}_\theta(y) dy \\ &= \int_{-\varepsilon}^\varepsilon y^2 \dot{q}_\theta(y) dy + \varepsilon^2 \int_{|y| \geq \varepsilon} \dot{q}_\theta(y) dy \end{aligned} \quad (28)$$

The derivative of $X_{2,\theta}(t)$ with respect to θ is then:

$$\frac{\partial}{\partial \theta} X_{2,\theta}(t) = \dot{\sigma}_\theta(\varepsilon) W(t) + \frac{\partial}{\partial \theta} X_{1,\theta}(t) \quad (29)$$

An alternative approach to approximating and simulating Lévy processes uses a series representation, as in Rosinski (2001). For finite variation Lévy processes, the

compound Poisson approximation (25) is equivalent to a particular truncation of the series representation of $X(t)$. As shown in Liu (2008), the pathwise derivative obtained by differentiating the truncated series is the same as (27). So we do not separately pursue the idea of derivative estimation through the series representation.

5.2.2 Convergence rates of the bias

In this section we discuss our general approach to the convergence of the estimators (27) and (29) as the truncation level ε becomes small. A detailed analysis is contained in the technical appendix at the end of the paper. Asmussen and Rosinski (2001), Cont and Tankov (2004) and Signahl (2003) analyze the convergence of the approximations X_1 and X_2 themselves. Our analysis builds on Signahl (2003), in particular.

For simplicity, we just consider $t = 1$ and omit the argument t from the notation, writing, eg, $X_\theta(t)$ as X_θ . If we could apply the pathwise method directly to X_θ , we would be estimating:

$$\frac{d}{d\theta} E[V(X_{\varepsilon,\theta} + X_{1,\theta})] = E \left[V'(X_{\varepsilon,\theta} + X_{1,\theta}) \left(\frac{\partial}{\partial \theta} X_{\varepsilon,\theta} + \frac{\partial}{\partial \theta} X_{1,\theta} \right) \right]$$

If $X_{1,\theta}$ is used as the approximation, then we estimate:

$$\frac{d}{d\theta} E[V(X_{1,\theta})] = E \left[V'(X_{1,\theta}) \frac{\partial}{\partial \theta} X_{1,\theta} \right]$$

The bias therefore is:

$$\frac{d}{d\theta} (E[V(X_{\varepsilon,\theta} + X_{1,\theta})] - E[V(X_{1,\theta})]) \quad (30)$$

Signahl (2003) analyzes differences in expectations of the type in (30), without derivatives, both for X_1 and X_2 . He applies a Taylor approximation to V and thus reduces the question of convergence of the expectation of V to the convergence of the moments of the residual $X_{\varepsilon,\theta}$. Under appropriate conditions, he finds that the error using X_1 is $O(\varepsilon^2)$ whereas the error using X_2 is $O(\varepsilon^3)$. In our setting, we will weaken the conditions on V (Signahl (2003) and Cont and Tankov (2004) assume that V has bounded derivatives) and we will need to include new terms arising from pathwise differentiation, in addition to the moments of the residual.

We let:

$$V_{1,\theta}(x) = E[V(x + X_{1,\theta})]$$

so (30) equals:

$$\frac{d}{d\theta} E[V_{1,\theta}(X_{\varepsilon,\theta}) - V_{1,\theta}(0)]$$

The analysis proceeds by differentiating in θ and then applying a Taylor approximation to the function $V_{1,\theta}$ to get:

$$\begin{aligned} \frac{d}{d\theta}(E[V_{1,\theta}(X_{\varepsilon,\theta}) - V_{1,\theta}(0)]) &= E\left[\dot{V}_{1,\theta}(X_{\varepsilon,\theta}) + V'_{1,\theta}(X_{\varepsilon,\theta})\frac{\partial}{\partial\theta}X_{\varepsilon,\theta} - \dot{V}_{1,\theta}(0)\right] \\ &= \frac{1}{2}E[\dot{V}''_{1,\theta}(\xi_1(X_{\varepsilon,\theta}))X_{\varepsilon,\theta}^2] \\ &\quad + \frac{1}{2}E\left[V''_{1,\theta}(\xi_2(X_{\varepsilon,\theta}))X_{\varepsilon,\theta}\frac{\partial}{\partial\theta}X_{\varepsilon,\theta}\right] \end{aligned} \quad (31)$$

where $\xi_i(X_{\varepsilon,\theta})$, $i = 1, 2$, lie between 0 and $X_{\varepsilon,\theta}$. Once these steps are justified, the key is then to determine the order of the two terms in (31).

This analysis requires several conditions, starting with the following conditions on V .

ASSUMPTION 5.1 *We assume that $V(x)$ is four times differentiable. For $x > 0$, $0 \leq |V^{(n)}(x)| \leq C_v e^{v+x}$, and for $x < 0$, $0 \leq |V^{(n)}(x)| \leq C_v e^{v-|x|}$ for some constants $C_v > 0$, $v_+ \in (0, s_+)$ and $v_- \in (0, s_-)$, $n = 0, \dots, 4$.*

The differentiability assumed here is not directly applicable to option payouts. We can always find a series of differentiable functions approximating a less smooth payout, and we can replace a payout with its conditional expectation shortly before maturity. However, our analysis does not address these additional sources of approximation error. In practice, the key distinction is generally between payouts that are Lipschitz continuous in the underlying asset price (such as standard calls and puts) and payouts with discontinuities. The pathwise method is generally inapplicable with discontinuities. The condition at $n = 0$ ensures that $E[V(X_\theta)]$ exists and is finite.

Our analysis also imposes conditions on the Lévy density q_θ and the MGF. As these are quite involved even to state, we omit them and refer the reader to the technical appendix or Liu (2008). The conditions are verified in our applications to the VG and NIG models.

We stress that all our results on convergence hold in the big- O sense and are thus upper bounds. The exact convergence rate in each case is potentially faster than indicated by our results.

5.2.3 Examples

We now illustrate the ideas of the previous two subsections with examples.

EXAMPLE 5.2 (VG model) For the VG model with parameters (ρ, v, θ) , the Lévy measure is:

$$q_{\text{vg}}(y) = \begin{cases} C \exp(-My)y^{-1}, & y > 0 \\ C \exp(Gy)|y|^{-1}, & y < 0 \end{cases}$$

where $C = 1/\nu$, and M and G are functions of (θ, ρ, ν) (see Madan *et al* (1998) for details). The VG process has finite variation and the first component of the Lévy triplet is $b = \int_{-1}^1 y q_{\text{vg}}(y) dy$. The region of convergence is:

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

We consider the sensitivity to ρ and write M_ρ and G_ρ for emphasis. Applying (8), we obtain that the derivatives of Y_ρ^+ and Y_ρ^- with respect to ρ are equal to:

$$\frac{\partial}{\partial\rho} Y_\rho^+ = -\dot{M}_\rho \frac{Y_\rho^+}{M_\rho} \quad \text{and} \quad \frac{\partial}{\partial\rho} Y_\rho^- = \dot{G}_\rho \frac{Y_\rho^-}{G_\rho}$$

The derivative of $q_{\text{vg},\rho}$ with respect to ρ is:

$$\dot{q}_{\text{vg},\rho}(y) = \begin{cases} -\dot{M}_\rho \exp(-M_\rho y) / M_\rho, & y > 0 \\ \dot{G}_\rho \exp(G_\rho y) / G_\rho, & y < 0 \end{cases}$$

PROPOSITION 5.3 *Suppose that Assumption 5.1 holds and $T \geq 3\nu/4$. Then, in the VG model, the bias in the pathwise derivative estimate for ρ is $O(\varepsilon^2)$ using $X_1(T)$, and $O(\varepsilon^3)$ using $X_2(T)$.*

We test these methods with the European call option example defined in Section 3.4.2. In Table 5 on the facing page, we run 5 million simulation trials and report the estimated biases using the compound Poisson (CP) approximation X_1 and the corrected (CP-N) approximation X_2 for prices and sensitivities (and the corresponding standard errors for each case). The CP results in the last column illustrate the convergence rate of the bias; the next three estimates (not included in the table) at $\varepsilon = 1/32$, $1/64$ and $1/128$ are -0.77 , -0.22 and -0.05 , each with an estimated standard error of 0.03 . The errors roughly decrease like $O(\varepsilon^2)$ when ε becomes small. This is in line with our analysis, though the smoothness required on the payout function in our analysis does not hold in this example.

Table 5 on the facing page also compares the CP approximation X_1 and the CP-N approximation X_2 . The normal approximation dramatically improves accuracy, and it has an even bigger impact on the sensitivity estimates than on the price estimates. The bias becomes so small that it is difficult to estimate accurately for small ε , so we only take ε as small as $1/16$.

EXAMPLE 5.4 (NIG model) The Lévy measure of the NIG process is:

$$q_{\text{nig}}(y) = \pi^{-1} \delta \alpha |y|^{-1} K_1(\alpha|y|) e^{\beta y}$$

where $K_1(\cdot)$ is the modified Bessel function of the third kind with index 1. The NIG process is a compensated Lévy process, and the first component of the Lévy triplet is

TABLE 5 Comparison of estimated biases for prices and sensitivities to ρ using the VG model.

ε	Approximation	Price	ρ -sensitivity
1/2	CP	-6.55 (0.01)	-32.75 (0.01)
1/2	CP-N	0.26 (0.01)	4.85 (0.03)
1/4	CP	-2.67 (0.01)	-18.27 (0.02)
1/4	CP-N	-0.09 (0.01)	1.84 (0.03)
1/8	CP	-0.91 (0.01)	-6.24 (0.03)
1/8	CP-N	-0.04 (0.01)	0.11 (0.03)
1/16	CP	-0.31 (0.01)	-2.52 (0.03)
1/16	CP-N	-0.01 (0.01)	0.02 (0.03)

“CP” indicates the compound Poisson approximation X_1 and CP-N includes the normal correction, X_2 . The reference values for the price and sensitivity are 11.27 and 23.04.

(see Barndorff-Nielsen (1997)):

$$b = 2\pi^{-1}\delta\alpha \int_0^1 \sinh(\beta y) K_1(\alpha y) dy$$

The region of convergence includes $(-\alpha - \beta, \alpha - \beta)$. We consider sensitivities to the parameter δ . It is easy to see that $\dot{q}_{\text{nig},\delta}(y) = q_{\text{nig},\delta}(y)/\delta$. We also have:

$$\sigma_\delta^2(\varepsilon) = 2\pi^{-1}\delta\alpha \int_0^\varepsilon y \cosh(\beta y) K_1(\alpha y) dy$$

Interestingly, the convergence rates in this example are slower than in the previous one.

PROPOSITION 5.5 *Suppose that Assumption 5.1 holds. Then, in the NIG model, the bias in the pathwise derivative estimate for δ is $O(\varepsilon)$ using $X_1(T)$, and $O(\varepsilon^2)$ using $X_2(T)$.*

We test the approximations with the European call option example in Section 3.4.2. To sample from $q_{\text{nig}}(y)/\lambda$, we use the acceptance–rejection method. By the properties of K_1 , we can use the densities of two shifted exponential distributions as the bounding densities: for Y^+ , the bounding density is $(\alpha - \beta) \exp(-(\alpha - \beta)(y - \varepsilon))$; for Y^- , the bounding density is $(\alpha + \beta) \exp(-(\alpha + \beta)(y - \varepsilon))$.

Simulating the compound Poisson approximation for the NIG model is more computationally demanding than for the VG model. In the NIG model, the jump rate λ increases very quickly as ε approaches 0. Also, with:

$$\frac{\partial}{\partial \delta} Y_\delta^+ = (\delta q_{\text{nig}}(Y_\delta^+))^{-1} \int_{Y_\delta^+}^\infty q_{\text{nig}}(y) dy$$

TABLE 6 Comparison of simulation errors for prices and sensitivities to δ using the NIG model.

ε	Approximation	Standard option		Asian option	
		Price	δ -sensitivity	Price	δ -sensitivity
1/2	CP	-2.72 (0.01)	-8.61 (0.01)	-1.72 (0.01)	-5.18 (0.01)
1/2	CP-N	-0.11 (0.01)	0.34 (0.02)	-0.07 (0.02)	0.28 (0.03)
1/4	CP	-2.50 (0.01)	-8.39 (0.01)	-1.60 (0.01)	-5.07 (0.01)
1/4	CP-N	-0.11 (0.01)	0.27 (0.02)	-0.07 (0.02)	0.24 (0.03)
1/8	CP	-1.76 (0.01)	-6.47 (0.01)	-1.11 (0.01)	-4.19 (0.02)
1/8	CP-N	-0.08 (0.01)	0.07 (0.02)	-0.03 (0.02)	0.08 (0.03)
1/16	CP	-1.22 (0.01)	-5.24 (0.01)	-0.74 (0.01)	-3.12 (0.01)
1/16	CP-N	-0.01 (0.01)	0.01 (0.02)	-0.01 (0.02)	-0.07 (0.03)

CP indicates the compound Poisson approximation X_1 and CP-N includes the normal correction, X_2 . The reference values for the price and sensitivity are 11.36 and 5.81 for the standard option, and 6.34 and 3.71 for the Asian option.

and:

$$\frac{\partial}{\partial \delta} Y_{\delta}^{-} = (\delta q_{\text{nig}}(-Y_{\delta}^{-}))^{-1} \int_{Y_{\delta}^{-}}^{\infty} q_{\text{nig}}(-y) dy$$

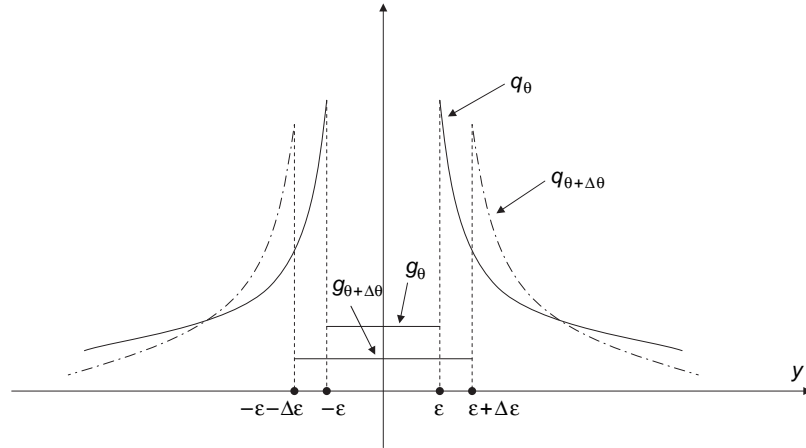
the computation of the derivatives involves the numerical integral of the modified Bessel function. The integral could be tabulated to accelerate calculation.

The third and fourth columns of Table 6 report results based on 1 million replications. As in the VG example, the CP-N approximation substantially outperforms the CP approximation, and the improvement is greater for the sensitivity than for the price. However, the convergence orders in ε for the two methods are difficult to discern from the simulation results, presumably because the ε values are not yet small enough for the asymptotic regime.

EXAMPLE 5.6 (Path-dependent option) We consider an Asian option example using the NIG model. The option is based on averaging monthly values over a one-year horizon, and the strike is equal to S_0 . Results based on 1 million replications are reported in the last two columns of Table 6. It shows the same general pattern of results as the European option.

5.3 Pathwise method: truncated density

In the previous section, we derived pathwise estimators for compound Poisson approximations by separating positive and negative jumps and adjusting the truncation levels to keep the arrival rates of both constant. In this section, we develop an alternative approximation method in which, rather than truncate all small jumps, we

FIGURE 6 Changing ε with θ while keeping $\tilde{\lambda}$ fixed.


Here, ε and $-\varepsilon$ move in opposite directions with the same magnitude $\Delta\varepsilon = \varepsilon_{\theta+\Delta\theta} - \varepsilon_{\theta}$.

instead truncate their density. Here, too, we will develop first-order and second-order approximations.

5.3.1 The method

We introduce a jump-size random variable \tilde{Y}_{θ} , which has the same density as Y_{θ} on $(-\infty, -\varepsilon] \cup [\varepsilon, \infty)$, but with a positive density on $(-\varepsilon, \varepsilon)$. In the most general version of the method, we choose the density of \tilde{Y}_{θ} on $(-\varepsilon, \varepsilon)$ to be proportional to a positive function g_{θ} . This will allow \tilde{Y}_{θ} to change continuously with θ and eliminates the need to separate positive and negative jumps.

We let:

$$\tilde{\lambda} = \int_{-\varepsilon}^{\varepsilon} g_{\theta}(y) dy + \lambda$$

denote the total arrival rate of jumps. We now require $d\tilde{\lambda}/d\theta = 0$ to make the pathwise method applicable to the new compound Poisson process. To achieve this, we let ε depend on θ and denote it by ε_{θ} . Setting $d\tilde{\lambda}/d\theta = 0$, we find that we must have:

$$\dot{\varepsilon}_{\theta} = \frac{\int_{|y| \geq \varepsilon_{\theta}} \dot{q}_{\theta}(y) dy + \int_{|y| \leq \varepsilon_{\theta}} \dot{g}_{\theta}(y) dy}{q_{\theta}(-\varepsilon_{\theta}) + q_{\theta}(\varepsilon_{\theta}) - g_{\theta}(\varepsilon_{\theta}) - g_{\theta}(-\varepsilon_{\theta})} \quad (32)$$

in contrast with (26). This mechanism for varying ε with θ is illustrated in Figure 6.

Let:

$$\tilde{X}_{1,\theta}(t) = \tilde{c}_\theta t + \sum_{k=1}^{\tilde{N}(t)} \tilde{Y}_{k,\theta}$$

where $\tilde{N}(t)$ is a Poisson process with arrival rate $\tilde{\lambda}_\theta$, the $\tilde{Y}_{k,\theta}$ are iid with the same distribution as \tilde{Y}_θ and:

$$\tilde{c}_\theta = c_\theta - \int_{-\varepsilon_\theta}^{\varepsilon_\theta} y g_\theta(y) dy$$

Then $\tilde{X}_{1,\theta}(t)$ provides an alternative compound Poisson approximation to X_θ .

The derivative of $\tilde{X}_{1,\theta}(t)$ is:

$$\frac{\partial}{\partial \theta} \tilde{X}_{1,\theta}(t) = \frac{d\tilde{c}_\theta}{d\theta} t + \sum_{k=1}^{\tilde{N}(t)} \frac{\partial}{\partial \theta} \tilde{Y}_{k,\theta}$$

where:

$$\begin{aligned} \frac{d\tilde{c}_\theta}{d\theta} = & \dot{\mu}_\theta - \int_{\varepsilon_\theta \leq |y| \leq 1} y \dot{q}_\theta(y) dy - \int_{-\varepsilon_\theta}^{\varepsilon_\theta} y \dot{g}_\theta(y) dy \\ & + \varepsilon_\theta \dot{\varepsilon}_\theta (q_\theta(\varepsilon_\theta) - q_\theta(-\varepsilon_\theta) - g_\theta(\varepsilon_\theta) + g_\theta(-\varepsilon_\theta)) \end{aligned}$$

With the normal correction, we get:

$$\tilde{X}_{2,\theta}(t) = \tilde{\sigma}_\theta(\varepsilon_\theta) W(t) + \tilde{X}_{1,\theta}(t)$$

where $\tilde{\sigma}_\theta^2(\varepsilon_\theta) = \sigma_\theta^2(\varepsilon_\theta) - \sigma_{g,\theta}^2(\varepsilon_\theta)$ and:

$$\sigma_{g,\theta}^2(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} y^2 g_\theta(y) dy$$

So:

$$\frac{\partial}{\partial \theta} \tilde{X}_{2,\theta}(t) = \frac{d\tilde{\sigma}_\theta(\varepsilon_\theta)}{d\theta} W(t) + \frac{\partial}{\partial \theta} \tilde{X}_{1,\theta}(t)$$

with:

$$\begin{aligned} \frac{d\tilde{\sigma}_\theta^2(\varepsilon_\theta)}{d\theta} = & \frac{d\varepsilon_\theta}{d\theta} \varepsilon_\theta^2 (q_\theta(\varepsilon_\theta) + q_\theta(-\varepsilon_\theta) - g_\theta(\varepsilon_\theta) - g_\theta(-\varepsilon_\theta)) \\ & + \int_{-\varepsilon_\theta}^{\varepsilon_\theta} y^2 (\dot{q}_\theta(y) - \dot{g}_\theta(y)) dy \\ = & \varepsilon_\theta^2 \left(\int_{|y| \geq \varepsilon_\theta} \dot{q}_\theta(y) dy + \int_{-\varepsilon_\theta}^{\varepsilon_\theta} \dot{g}_\theta(y) dy \right) \\ & + \int_{-\varepsilon_\theta}^{\varepsilon_\theta} y^2 (\dot{q}_\theta(y) - \dot{g}_\theta(y)) dy \end{aligned} \quad (33)$$

A simple choice of $g_\theta(y)$ would be to set it equal to some constant G , which may or may not depend on θ but should satisfy the constraint that $\tilde{\sigma}_\theta^2(\varepsilon_\theta) \geq 0$. If G does not depend on θ , many formulas simplify. For example, $\dot{\varepsilon}_\theta$ reduces to:

$$\dot{\varepsilon}_\theta = \frac{\int_{|y| \geq \varepsilon_\theta} \dot{q}_\theta(y) dy}{q_\theta(-\varepsilon_\theta) + q_\theta(\varepsilon_\theta) - 2G}$$

Also, on $(-\varepsilon_\theta, \varepsilon_\theta)$, $(\partial/\partial\theta)\tilde{Y}_\theta$ becomes:

$$\frac{\partial}{\partial\theta} \tilde{Y}_\theta = \frac{(G - q_\theta(\varepsilon_\theta))\dot{\varepsilon}_\theta + \partial/\partial\theta \bar{Q}_\theta(\varepsilon_\theta)}{G}$$

On the other hand, to achieve a higher order of convergence, it is better to choose G depending on θ so that $\tilde{\sigma}_\theta^2(\varepsilon_\theta) \equiv 0$ for all θ , which yields:

$$G_\theta = \frac{3}{2}\varepsilon_\theta^{-3}\sigma_\theta^2(\varepsilon_\theta)$$

This, however, leads to more complicated expressions and additional computation.

5.3.2 Examples

EXAMPLE 5.7 (VG model) We use the same VG example as before and let $g_\rho(y) \equiv \frac{1}{2}$ to make $\tilde{\sigma}_\rho^2(\varepsilon_\rho) \geq 0$ for the ε_ρ values we try. The derivative of \tilde{Y}_ρ with respect to ρ is:

$$\frac{\partial}{\partial\rho} \tilde{Y}_\rho = \begin{cases} \dot{G}_\rho \tilde{Y}_\rho / G_\rho, & \tilde{Y}_\rho \leq -\varepsilon_\rho \\ -\dot{M}_\rho \tilde{Y}_\rho / M_\rho, & \tilde{Y}_\rho \geq \varepsilon_\rho \\ (1 - 2q_{\text{vg}}(\varepsilon_\rho))\dot{\varepsilon}_\rho - 2C \dot{M}_\rho \cdot e^{-\varepsilon_\rho M_\rho} / M_\rho, & |\tilde{Y}_\rho| < \varepsilon_\rho \end{cases} \quad (34)$$

PROPOSITION 5.8 *Suppose that Assumption 5.1 holds and $T \geq 3\nu/4$. Then, in the VG model, the bias in the pathwise derivative estimate for ρ is $O(\varepsilon_\rho^2)$ using $\tilde{X}_1(T)$, and $O(\varepsilon_\rho^3)$ using $\tilde{X}_2(T)$.*

As we repeatedly halve ε , reducing it from $\frac{1}{2}$ to $\frac{1}{128}$, our estimates of the bias using the compound Poisson approximation \tilde{X}_1 , based on 5 million replications, are $-28.69, -26.21, -5.84, -2.52, -0.71, -0.18$ and -0.01 , with standard errors of approximately 0.04 to 0.05. Thus, the error decreases roughly like $O(\varepsilon^2)$, as in our discussion of Table 5 on page 31.

As in the previous section, adding the normal correction dramatically improves accuracy. We present numerical results for this comparison after introducing alternative estimators in the next two sections.

EXAMPLE 5.9 (NIG model) We use the same NIG example as before. The bounding density is now a truncated double exponential distribution:

$$q_{\text{de}}(y) = \begin{cases} C_{\text{de}} \exp(-(\alpha - \beta)y), & y \geq \varepsilon_\delta \\ C_{\text{de}} \exp(-(\alpha + \beta)|y|), & y \leq -\varepsilon_\delta \end{cases} \quad (35)$$

where the constant C_{de} normalizes $q_{\text{de}}(y)$.

We let $g_\delta(y) \equiv 0.2$ to make $\tilde{\sigma}_\delta^2(\varepsilon_\delta) \geq 0$ for the ε_δ values we try. The following result is established along the same lines as the previous examples.

PROPOSITION 5.10 *Suppose that Assumption 5.1 holds. Then, in the NIG model, the bias in the pathwise derivative estimate for δ is $O(\varepsilon_\delta)$ using $\tilde{X}_1(T)$, and $O(\varepsilon_\delta^2)$ using $\tilde{X}_2(T)$.*

5.4 Pathwise and likelihood ratio methods combined

In Sections 5.2 and 5.3, we let ε depend on θ to make λ_θ independent of θ and thus avoid introducing discontinuities. In this section, we consider an alternative approach that keeps ε fixed (and thus lets $\lambda = \lambda_\theta$ vary) but uses LRM for the dependence of λ on θ and uses pathwise differentiation for the jump sizes (and the normal correction). L'Ecuyer (1990) combines these methods in other settings.

If ε is independent of θ , we have:

$$\dot{\lambda}_\theta = \int_{|y| \geq \varepsilon} \dot{q}_\theta(y) dy, \quad \dot{\lambda}_\theta^+ = \int_{y \geq \varepsilon} \dot{q}_\theta(y) dy \quad \text{and} \quad \dot{\lambda}_\theta^- = \int_{y \geq \varepsilon} \dot{q}_\theta(-y) dy$$

Also, we have:

$$\dot{\sigma}_\theta^2(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} y^2 \dot{q}_\theta(y) dy \quad (36)$$

which is different from the derivative $\sigma_\theta^2(\varepsilon)$ in Sections 5.2 and 5.3.

To derive an estimator of $dE_\theta[V(X_{1,\theta}(t))]/d\theta$, we use the representation in (25). With the Poisson probabilities:

$$p_n^\pm(\theta) = \exp(-\lambda_\theta^\pm t) \frac{(\lambda_\theta^\pm t)^n}{n!}$$

we have:

$$\begin{aligned} & \frac{d}{d\theta} E_\theta[V(X_{1,\theta}(t))] \\ &= \frac{d}{d\theta} \sum_{m,n=0}^{\infty} p_m^+(\theta) p_n^-(\theta) E \left[V \left(c_\theta t + \sum_{k=1}^m Y_{k,\theta}^+ - \sum_{j=1}^n Y_{j,\theta}^- \right) \right] \\ &= \sum_{m,n=0}^{\infty} \left(E[V(X_{1,\theta}(t))] \left(\frac{dp_m^+(\theta)}{d\theta} p_n^-(\theta) + p_m^+(\theta) \frac{dp_n^-(\theta)}{d\theta} \right) \right. \\ & \quad \left. + p_m^+(\theta) p_n^-(\theta) \frac{d}{d\theta} E \left[V \left(c_\theta t + \sum_{k=1}^m Y_{k,\theta}^+ - \sum_{j=1}^n Y_{j,\theta}^- \right) \right] \right) \quad (37) \end{aligned}$$

Simple algebra yields:

$$\frac{d}{d\theta} p_n^\pm(\theta) = p_n^\pm(\theta) \dot{\lambda}_\theta^\pm \left(\frac{n}{\lambda_\theta^\pm} - t \right)$$

so the first term in (37) equals:

$$\sum_{m,n=0}^{\infty} \left(E[V(X_{1,\theta}(t))] p_m^+(\theta) p_n^-(\theta) \left(\dot{\lambda}_\theta^+ \left(\frac{m}{\lambda_\theta^+} - t \right) + \dot{\lambda}_\theta^- \left(\frac{n}{\lambda_\theta^-} - t \right) \right) \right)$$

Differentiating the second term in (37), we get:

$$\sum_{m,n=0}^{\infty} \left(p_m^+(\theta) p_n^-(\theta) E \left[V'(X_{1,\theta}(t)) \left(\dot{c}_\theta t + \sum_{k=1}^m \frac{\partial \bar{P}_\theta^+(Y_{k,\theta}^+)/\partial \theta}{p^+(Y_{k,\theta}^+)} + \sum_{j=1}^n \frac{\partial \bar{P}_\theta^-(Y_{j,\theta}^-)/\partial \theta}{p(Y_{j,\theta}^-)} \right) \right] \right)$$

So (37) is equal to:

$$E_\theta \left[\left(\dot{\lambda}_\theta^+ \left(\frac{N^+(t)}{\lambda_\theta^+} - t \right) + \dot{\lambda}_\theta^- \left(\frac{N^-(t)}{\lambda_\theta^-} - t \right) \right) V(X_{1,\theta}(t)) + V'(X_{1,\theta}(t)) \left(\dot{c}_\theta t + \sum_{k=1}^m \frac{\partial \bar{P}_\theta^+(Y_{k,\theta}^+)/\partial \theta}{p^+(Y_{k,\theta}^+)} + \sum_{j=1}^n \frac{\partial \bar{P}_\theta^-(Y_{j,\theta}^-)/\partial \theta}{p(Y_{j,\theta}^-)} \right) \right] \quad (38)$$

where:

$$\begin{aligned} p_\theta^\pm(y) &= \frac{q_\theta(\pm y)}{\lambda_\theta^\pm} \\ \bar{P}_\theta^\pm(y) &= \frac{\bar{Q}_\theta^\pm(y)}{\lambda_\theta^\pm} \\ &= \int_y^\infty \frac{q_\theta(\pm y) dy}{\lambda_\theta^\pm} \\ \dot{c}_\theta &= \dot{b}_\theta - \int_{\varepsilon \leq |y| \leq 1} \dot{q}_\theta(y) dy \end{aligned}$$

The expression inside the expectation in (38) provides an estimator of the derivative.

A similar expression applies if we include the normal correction for the variance.

Note that:

$$\frac{\partial \bar{P}_\theta^+(y)/\partial \theta}{p_\theta^+(y)} = \frac{\partial \bar{Q}_\theta^+(y)/\partial \theta}{q_\theta(y)} - \frac{\dot{\lambda}_\theta^+}{\lambda_\theta^+} \frac{\bar{Q}_\theta^+(y)}{q_\theta(y)}$$

Since, in general, $\dot{\lambda}_\theta^+ \neq 0$, we have:

$$(p^+(y))^{-1} \frac{\partial}{\partial \theta} \bar{P}^+(y) \neq (q(y))^{-1} \frac{\partial}{\partial \theta} \bar{Q}^+(y)$$

and the corresponding conclusion for negative jumps.

The convergence rate analysis for the bias in Section 5.2 goes through essentially as before. Because ε no longer depends on θ , the MGF of $\hat{f}_{\varepsilon,\theta}(x)$ is simply:

$$\Phi_{\varepsilon,\theta}(s) \int_{-\varepsilon}^{\varepsilon} (e^{sy} - 1 - sy)\dot{q}_{\theta}(y) dy$$

and the definition of $D_{k,\theta}(\varepsilon)$ is changed to:

$$D_{k,\theta}(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} |y^k \dot{q}_{\theta}(y)| dy$$

With these changes, Theorem C 3 holds as before.

We postpone discussion of numerical results until the next section and merely point out some implementation aspects of our running examples. In the VG model, the pathwise-LRM estimator takes longer to compute than the estimators of Sections 5.2 and 5.3 because $(p_{\rho}^{\pm}(y))^{-1}(\partial/\partial\rho)\bar{P}_{\rho}^{\pm}(y)$ involves the computation of the infinite integral in $\bar{Q}_{\rho}^{\pm}(y)$. In the NIG model, the pathwise-LRM estimator is much faster to evaluate because $p_{\delta}^{\pm}(y)^{-1}(\partial/\partial\delta)\bar{P}_{\delta}^{\pm}(y) \equiv 0$, so for NIG the estimator in (38) is simply:

$$\left(\dot{\lambda}_{\delta}^{+} \left(\frac{N^{+}(t)}{\lambda_{\delta}^{+}} - t \right) + \dot{\lambda}_{\delta}^{-} \left(\frac{N^{-}(t)}{\lambda_{\delta}^{-}} - t \right) \right) V(X_{1,\delta}) + \dot{c}_{\delta} t V'(X_{1,\delta}(t))$$

which does not involve computation of the integral of the modified Bessel function.

5.5 Likelihood ratio method estimator

In this section, we derive pure LRM estimators for the CP and CP-N approximations. The score function for a compound Poisson process is:

$$\mathcal{S}_{1,\theta} \sum_{k=1}^{N(t)} Y_k = \dot{\lambda}_{\theta} \left(\frac{N(t)}{\lambda_{\theta}} - t \right) + \sum_{k=1}^{N(t)} \mathcal{S}_{\theta}(Y_k) \quad (39)$$

where $\mathcal{S}_{\theta}(Y_k)$ is the score function of Y_k . If we use X_1 as the approximation, then the sensitivity estimator is:

$$V'(X_1)\dot{c}_{\theta}t + V(X_1)\mathcal{S}_{1,\theta} \sum_{k=1}^{N(t)} Y_k$$

Because the parameter θ appears in the drift c_{θ} and not just in the density of the compound sum, the estimator involves differentiating V . This imposes a limitation on the applicability of the estimator, requiring, for example, that V be Lipschitz continuous.

TABLE 7 Estimated errors in derivative estimates in the VG model with parameter ρ .

ε	Approx.	PW1	PW2	PW-LRM	LRM
1/2	CP	-32.75 (0.01)	-28.69 (0.02)	-18.76 (0.04)	-18.76 (0.04)
1/2	CP-N	4.85 (0.03)	-12.55 (0.64)	2.11 (0.05)	2.14 (0.08)
1/4	CP	-18.27 (0.02)	-26.21 (0.05)	-4.79 (0.07)	-4.67 (0.08)
1/4	CP-N	1.84 (0.03)	0.45 (0.05)	1.26 (0.08)	1.44 (0.10)
1/8	CP	-6.24 (0.03)	-5.84 (0.04)	-1.71 (0.09)	-1.85 (0.11)
1/8	CP-N	0.11 (0.03)	-0.15 (0.04)	0.27 (0.09)	0.15 (0.12)
1/16	CP	-2.52 (0.03)	-2.52 (0.04)	-0.22 (0.09)	-0.21 (0.12)
1/16	CP-N	0.02 (0.03)	-0.02 (0.04)	0.01 (0.09)	0.04 (0.12)

The last four columns compare the estimators in Sections 5.2–5.5. The reference value for the sensitivity is 23.04.

But if we use the approximation X_2 , then we can make c_θ the drift of the Brownian correction and thus move the dependence on θ out of V and into the normal distribution of the correction. Because $X_2(t)$ is the independent sum of $c_\theta t + \sigma_\theta(\varepsilon)W(t)$ and $\sum_{k=1}^{N(t)} Y_k$, the score function is now the sum of the score functions of these two pieces:

$$\mathcal{S}_{2,\theta}\left(x + \sum_{k=1}^{N(t)} y_k\right) = \mathcal{S}_{1,\theta} \sum_{k=1}^{N(t)} y_k - \frac{c_\theta t - x}{\sigma_\theta^2(\varepsilon)t} \dot{c}_\theta t + \left(\frac{(c_\theta t - x)^2}{\sigma_\theta^3(\varepsilon)t} - \frac{1}{\sigma_\theta(\varepsilon)}\right) \dot{\sigma}_\theta(\varepsilon) \quad (40)$$

where $\dot{\sigma}_\theta^2(\varepsilon)$ is given by (36). So, the LRM estimator is $V(X_2)\mathcal{S}_{2,\theta}(X_2)$ and it does not require the Lipschitz condition on V . Here, the Brownian term serves not only to correct the level of the variance, but also to “spread out” the dependence on θ in the drift.

The convergence rates follow along the same lines as in the previous sections. Here, $\sigma_\theta^2(\varepsilon)$ and $D_{2,\theta}(\varepsilon)$ are the same as those in Section 5.4, and the convergence rates for both approximations are of the same order as those in Section 5.4.

In the VG model, the score function of Y_k with parameter ρ is:

$$\mathcal{S}_\rho(y) = \mathbf{1}_{y < 0} \cdot y \dot{G}_\rho - \mathbf{1}_{y > 0} \cdot y \dot{M}_\rho - \frac{\dot{\lambda}_\rho}{\lambda_\rho}$$

In the NIG model, the score function of Y_k with parameter δ is simply 0. The full dependence on the parameter is embedded in the arrival rate (and the drift). As a consequence, the pure LRM estimator coincides with the combined pathwise–LRM estimator in this case.

TABLE 8 Estimated errors in derivative estimates in the NIG model with parameter δ .

ε	Approx.	PW1	PW2	PW + LRM/LRM
1/2	CP	-8.61 (0.01)	-8.63 (0.01)	-8.59 (0.01)
1/2	CP-N	0.34 (0.02)	-3.66 (0.07)	0.36 (0.02)
1/4	CP	-8.39 (0.01)	-8.58 (0.01)	-7.86 (0.01)
1/4	CP-N	0.27 (0.02)	0.45 (0.02)	0.32 (0.02)
1/8	CP	-6.47 (0.01)	-8.19 (0.01)	-5.82 (0.02)
1/8	CP-N	0.07 (0.02)	0.03 (0.02)	0.14 (0.02)
1/16	CP	-5.24 (0.01)	-7.41 (0.01)	-3.86 (0.02)
1/16	CP-N	0.01 (0.02)	-0.06 (0.02)	0.03 (0.03)

The last three columns compare the estimators in Sections 5.2–5.5; the combined pathwise–LRM estimator and pure LRM estimator coincide in this case. The reference value for the sensitivity is 5.81.

5.6 Numerical results

We now compare the estimators of Sections 5.2–5.5 on our running examples: European options in the VG and NIG models and an Asian option in the NIG model. In each case, we compare estimators derived from compound Poisson approximations with (CP-N) and without (CP) the normal correction term. In reporting results, we refer to the pathwise estimators of Sections 5.2 and 5.3 as PW1 and PW2; we refer to the combined method of Section 5.4 as PW + LRM; and we refer to the pure LRM estimator of Section 5.5 as LRM.

Table 7 on the preceding page reports estimated errors in the VG model, based on 5 million replications. For ease of comparison, we have repeated the values for PW1 from Table 5 on page 31. Table 8 and Table 10 on the facing page report estimated errors (for European and Asian options, respectively) in the NIG model, based on 1 million replications. For ease of comparison, we have repeated the values for PW1 from Table 6 on page 32. Computing times in seconds are reported in Table 9 on the facing page. The increase in computing time as ε decreases varies widely because of differences in the dependence on ε of the truncated intensity λ . We have not attempted to optimize our MATLAB implementations, and relative speeds could change if the methods are implemented in a lower-level programming language.

The most salient and consistent feature of the numerical results is the dramatic improvement achieved through inclusion of the normal correction term. This effect overwhelms any other difference among the estimators. At small values of ε , all the CP-N estimators have similar errors. In the VG model, the estimators that use LRM for the jumps generally have lower bias (particularly among the CP-N estimators with larger ε values and among all the CP estimators), but the pathwise estimators have lower variance. The differences are less consistent in the NIG examples.

TABLE 9 Computing times in seconds.

ε	VG model				NIG model		
	PW1	PW2	PW+ LRM	LRM	PW1	PW2	PW+ LRM/LRM
1/2	49	46	110	38	25	44	16
1/4	51	47	299	40	56	60	18
1/8	59	51	878	43	796	625	40
1/16	75	56	1,649	50	5,326	5,310	190

The increase in computing time as ε decreases varies widely because of differences in the dependence on ε of the truncated intensity λ .

TABLE 10 Estimated errors in derivative estimates for an Asian option in the NIG model with parameter δ .

ε	Approx.	PW1	PW2	PW + LRM/LRM
1/2	CP	-5.18 (0.01)	-5.11 (0.01)	-5.17 (0.01)
1/2	CP-N	0.28 (0.03)	-1.32 (0.13)	0.24 (0.03)
1/4	CP	-5.07 (0.01)	-4.85 (0.01)	-4.78 (0.01)
1/4	CP-N	0.24 (0.03)	0.27 (0.03)	0.22 (0.03)
1/8	CP	-4.19 (0.02)	-4.18 (0.01)	-3.35 (0.01)
1/8	CP-N	0.08 (0.03)	0.05 (0.03)	0.11 (0.03)
1/16	CP	-3.12 (0.01)	-2.98 (0.02)	-2.23 (0.04)
1/16	CP-N	-0.07 (0.03)	-0.01 (0.04)	0.06 (0.06)

The last three columns compare the estimators in Sections 5.2–5.5; the combined pathwise–LRM estimator and pure LRM estimator coincide in this case. The reference value for the sensitivity is 3.71.

In comparing errors, it should be stressed that the bias in each case results not from the choice of estimator (pathwise or LRM) but rather from the choice of approximation and, especially, how the approximation changes with θ . Indeed, each of the estimators is unbiased for the particular “directional” derivative specified in its construction. Once we specify how ε , λ , Y_k and so on vary with θ , the bias is the difference between the derivative of the expectation of the corresponding CP or CP-N approximation and the derivative of the expectation with respect to the true Lévy process. In every case, the bias vanishes as ε becomes small (the various directional derivatives eventually all coincide) but the errors could potentially be very different at larger values of ε .

In contrast to the bias, the variance does depend on the estimation methodology and not just the form of the approximation. In most applications, LRM estimators have larger variance than pathwise estimators when both are applicable, so it is noteworthy that the standard errors in our comparisons are generally close across methods.

The variability of LRM estimators results from the score function, whose variability increases with the number of terms. In the Lévy process setting, we would expect to see higher variance in LRM estimators over longer time horizons.

Given the overall consistency in performance of the CP-N-based estimators, the main factors in choosing among them should be computational costs and ease of implementation. These factors are very much model dependent, as they involve sampling from truncated Lévy densities, integrating Lévy densities and differentiating various model-specific quantities with respect to the parameter of interest.

6 SUMMARY

We have developed a variety of methods for estimating price sensitivities through simulation of Lévy-driven models. The methods combine pathwise and likelihood ratio method estimation techniques with alternative approaches to approximating and simulating Lévy processes. Each method, both for simulation and for derivative estimation, has potential advantages, depending on the particular model and the information available. We summarize our conclusions as follows:

- Given an exact method for simulating the increments of the driving Lévy process, direct application of the pathwise method will often yield the best sensitivity estimates, at least for Lipschitz payouts. A representation of a Lévy process as a time-changed Brownian motion can be useful in implementing pathwise estimators, though this approach breaks down in the continuous-time limit if the parameter of interest affects the time change.
- When the probability density of the Lévy increments is known, LRM estimation is directly applicable using the score function derived from the density. A time-change representation is often useful in deriving an expression for the density as a mixture of normals. The resulting score function is often more tractable but the “mixed” LRM estimator has higher variance than one based directly on the score function for the increments.
- When the density is known only through its characteristic function or Laplace transform, an LRM estimator can be implemented through numerical transform inversion, as in Glasserman and Liu (2008). Saddlepoint approximations offer an alternative to numerical transform inversion and provide nearly closed-form expressions for the score function. They can be fast and easy to evaluate, but their accuracy is unpredictable. Centering the approximate score by its sample mean substantially improves the accuracy of the resulting LRM estimator.
- A compound Poisson process can approximate the dynamics of an infinite activity Lévy process, rather than just the distribution of its discrete increments. We

have considered approximations that truncate all small jumps or truncate the density of small jumps. Adding a deterministic drift centers the process at its original mean, and adding a Brownian correction matches the original variance. Alternative approximations and combinations of pathwise derivatives and LRM lead to alternative estimators. The Brownian correction can theoretically improve convergence by an order of magnitude (from $O(\varepsilon^2)$ to $O(\varepsilon^3)$ in the VG model, from $O(\varepsilon)$ to $O(\varepsilon^2)$ in the NIG model) and has a dramatic effect in numerical tests.

- Of the compound-Poisson-based methods, the simplest to implement is often a combination that uses LRM for the dependence of the Poisson rate on the parameter and pathwise derivatives for the jump sizes and the mean and variance of the Brownian correction.

APPENDIX A: SADDLEPOINT APPROXIMATION DERIVATION

Taking logarithms on both sides of (15), we get:

$$\begin{aligned} \log f_\theta(x) = & -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(K_\theta''(\hat{s})) + K_\theta(\hat{s}) - \hat{s}x \\ & + \log\{1 + (\frac{1}{8}\lambda_{4,\theta}(\hat{s}) - \frac{5}{24}\lambda_{3,\theta}^2(\hat{s})) + \dots\} \end{aligned} \quad (\text{A.1})$$

The derivative of $K_\theta^{(n)}(\hat{s}_\theta(x))$ with respect to θ is:

$$\frac{\partial K_\theta^{(n)}(\hat{s}_\theta(x))}{\partial \theta} = \dot{K}_\theta^{(n)}(\hat{s}) + K_\theta^{(n+1)}(\hat{s}) \frac{\partial \hat{s}_\theta(x)}{\partial \theta} \quad (\text{A.2})$$

Now apply this formula to $K_\theta'(\hat{s}_\theta(x))$. Since $K_\theta'(\hat{s}) = x$, differentiating both sides with respect to θ yields:

$$\dot{K}_\theta'(\hat{s}) + K_\theta''(\hat{s}) \frac{\partial \hat{s}_\theta(x)}{\partial \theta} = 0 \quad (\text{A.3})$$

and thus:

$$\frac{\partial \hat{s}_\theta(x)}{\partial \theta} = -(K_\theta''(\hat{s}))^{-1} \dot{K}_\theta'(\hat{s}) \quad (\text{A.4})$$

Applying (A.2) to $K_\theta(\hat{s}_\theta(x))$ and $K_\theta''(\hat{s}_\theta(x))$, we obtain:

$$\frac{\partial K_\theta(\hat{s}_\theta(x))}{\partial \theta} = \dot{K}_\theta(\hat{s}) + K_\theta'(\hat{s}) \frac{\partial \hat{s}_\theta(x)}{\partial \theta}$$

and:

$$\frac{\partial K_\theta''(\hat{s}_\theta(x))}{\partial \theta} = \dot{K}_\theta''(\hat{s}) + K_\theta'''(\hat{s}) \frac{\partial \hat{s}_\theta(x)}{\partial \theta}$$

Substituting these expressions into (A.1) and applying (A.4), we get:

$$\frac{\partial \log f_\theta(x)}{\partial \theta} = \frac{1}{2[K_\theta''(\hat{s})]^2} (K_\theta'''(\hat{s})\dot{K}_\theta'(\hat{s}) - K_\theta''(\hat{s})\dot{K}_\theta''(\hat{s})) + \dot{K}_\theta(\hat{s}) + T_\theta \quad (\text{A.5})$$

where T_θ is the derivative of $\log\{1 + (\frac{1}{8}\lambda_{4,\theta}(\hat{s}) - \frac{5}{24}\lambda_{3,\theta}^2(\hat{s})) + \dots\}$ with respect to θ :

$$T_\theta = \frac{\frac{1}{8}\dot{\lambda}_{4,\theta}(\hat{s}) - \frac{5}{12}\lambda_{3,\theta}(\hat{s})\dot{\lambda}_{3,\theta}(\hat{s}) + \dots}{1 + (\frac{1}{8}\lambda_{4,\theta}(\hat{s}) - \frac{5}{24}\lambda_{3,\theta}^2(\hat{s})) + \dots} \cdot \frac{\partial \hat{s}_\theta(x)}{\partial \theta} \quad (\text{A.6})$$

with $\dot{\lambda}_{n,\theta}(\hat{s})$ given by:

$$\begin{aligned} \dot{\lambda}_{n,\theta}(\hat{s}) = & \frac{1}{[K_\theta''(\hat{s})]^{n/2+1}} (K_\theta''(\hat{s})\dot{K}_\theta^{(n)}(\hat{s}) - K_\theta^{(n+1)}(\hat{s})\dot{K}_\theta'(\hat{s})) \\ & - \frac{nK_\theta^{(n)}(\hat{s})}{2[K_\theta''(\hat{s})]^{n/2+2}} (K_\theta''(\hat{s})\dot{K}_\theta''(\hat{s}) - K_\theta'''(\hat{s})\dot{K}_\theta'(\hat{s})) \end{aligned}$$

APPENDIX B: SADDLEPOINT APPROXIMATION TO NIG PROCESS

Solving $K'_{\alpha,\beta,\mu,\delta}(s) = x$ yields:

$$\hat{s} = \frac{\alpha(x - \mu)}{\sqrt{\delta^2 + (x - \mu)^2}} - \beta$$

For the parameter β , the saddlepoint approximation to the score function is:

$$\hat{S}_\beta(x) = x - \mu - \delta\beta(\alpha^2 - \beta^2)^{-1/2}$$

which can be verified to be the true score function. For the other parameters, the approximate and exact score functions are as follows:

$$\hat{S}_\mu(x) = \frac{3(x - \mu)}{2(\delta^2 + (x - \mu)^2)} + \frac{\alpha(x - \mu)}{\sqrt{\delta^2 + (x - \mu)^2}} - \beta$$

$$S_\mu(x) = \frac{(x - \mu)}{\delta^2 + (x - \mu)^2} - \frac{\alpha(x - \mu)}{\sqrt{\delta^2 + (x - \mu)^2}} \frac{K_1'}{K_1} \left(\delta\alpha q \left(\frac{x - \mu}{\delta} \right) \right) - \beta$$

$$\hat{S}_\alpha(x) = \frac{1}{2\alpha} + \delta\alpha(\alpha^2 - \beta^2)^{-1/2} - \sqrt{\delta^2 + (x - \mu)^2}$$

$$S_\alpha(x) = \frac{1}{\alpha} + \delta\alpha(\alpha^2 - \beta^2)^{-1/2} + \sqrt{\delta^2 + (x - \mu)^2} \frac{K_1'}{K_1} \left(\delta\alpha q \left(\frac{x - \mu}{\delta} \right) \right)$$

$$\hat{S}_\delta(x) = \sqrt{\alpha^2 - \beta^2} + \frac{3(x - \mu)^2}{2\delta(\delta^2 + (x - \mu)^2)} - \frac{\delta\alpha}{\sqrt{\delta^2 + (x - \mu)^2}}$$

$$S_\delta(x) = \sqrt{\alpha^2 - \beta^2} + \frac{(x - \mu)^2}{\delta(\delta^2 + (x - \mu)^2)} + \frac{\delta\alpha}{\sqrt{\delta^2 + (x - \mu)^2}} \frac{K_1'}{K_1} \left(\delta\alpha q \left(\frac{x - \mu}{\delta} \right) \right)$$

APPENDIX C: TECHNICAL APPENDIX

In this appendix, we discuss the conditions we use to analyze the convergence of various estimators as the truncation level ε becomes small. Asmussen and Rosinski (2001), Cont and Tankov (2004) and Signahl (2003) analyze the convergence of the approximations X_1 and X_2 themselves. Our analysis builds on Signahl (2003), in particular. We give conditions that allow us to identify the order of convergence of the estimators and discuss these conditions. This appendix discusses the conditions; proofs of the convergence results are given in Liu (2008).

C.1 Convergence rates of the bias: truncated jumps

We begin by considering the estimators (27) and (29) as the truncation level ε becomes small.

C.1.1 A general result

For simplicity, we just consider $t = 1$ and omit the argument t from the notation, writing, for example, $X_\theta(t)$ as X_θ . If we could apply the pathwise method directly to X_θ , we would be estimating:

$$\frac{d}{d\theta} E[V(X_{\varepsilon,\theta} + X_{1,\theta})] = E \left[V'(X_{\varepsilon,\theta} + X_{1,\theta}) \left(\frac{\partial}{\partial \theta} X_{\varepsilon,\theta} + \frac{\partial}{\partial \theta} X_{1,\theta} \right) \right]$$

If $X_{1,\theta}$ is used as the approximation, then we estimate:

$$\frac{d}{d\theta} E[V(X_{1,\theta})] = E \left[V'(X_{1,\theta}) \frac{\partial}{\partial \theta} X_{1,\theta} \right]$$

The bias therefore is:

$$\frac{d}{d\theta} (E[V(X_{\varepsilon,\theta} + X_{1,\theta})] - E[V(X_{1,\theta})]) \quad (C.1)$$

Signahl (2003) analyzes differences in expectations of the type in (C.1), without derivatives, both for X_1 and X_2 . He applies a Taylor approximation to V and thus reduces the question of convergence of the expectation of V to the convergence of the moments of the residual $X_{\varepsilon,\theta}$. Under appropriate conditions, he finds that the error using X_1 is $O(\varepsilon^2)$ whereas the error using X_2 is $O(\varepsilon^3)$. In our setting, we will weaken the conditions on V (Signahl (2003) and Cont and Tankov (2004) assume V has bounded derivatives) and we will need to include new terms arising from pathwise differentiation, in addition to the moments of the residual.

We let:

$$M_{k,\theta}(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} |y|^k q_\theta(y) dy \quad \text{for } k = 1, 2$$

and:

$$M_{3,\theta}(\varepsilon) = \max \left\{ \int_{-\varepsilon}^{\varepsilon} |y|^3 q_{\theta}(y) \, dy, \sigma_{\theta}^4(\varepsilon) \right\}$$

Then $M_{2,\theta}(\varepsilon) = \sigma_{\theta}^2(\varepsilon)$, and $M_{1,\theta}(\varepsilon)$ exists only in the finite variation case. Also note that in the finite variation case, as shown in Signahl (2003), Jensen's inequality gives:

$$\begin{aligned} \sigma_{\theta}^4(\varepsilon) &= \left(\int_{-\varepsilon}^{\varepsilon} |y|^2 q_{\theta}(y) \, dy \right)^2 \\ &= \left(\int_{-\varepsilon}^{\varepsilon} M_{1,\theta}(\varepsilon) |y| \cdot \frac{|y| q_{\theta}(y)}{M_{1,\theta}(\varepsilon)} \, dy \right)^2 \\ &\leq \int_{-\varepsilon}^{\varepsilon} M_{1,\theta}^2(\varepsilon) |y|^2 \cdot \frac{|y| q_{\theta}(y)}{M_{1,\theta}(\varepsilon)} \, dy \\ &= M_{1,\theta}(\varepsilon) \int_{-\varepsilon}^{\varepsilon} |y|^3 q_{\theta}(y) \, dy \end{aligned}$$

So, for sufficiently small ε , in the finite variation case:

$$M_{3,\theta}(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} |y|^3 q_{\theta}(y) \, dy$$

We also let:

$$V_{1,\theta}(x) = E[V(x + X_{1,\theta})]$$

so (C.1) equals:

$$\frac{d}{d\theta} E[V_{1,\theta}(X_{\varepsilon,\theta}) - V_{1,\theta}(0)]$$

The analysis proceeds by differentiating in θ and then applying a Taylor approximation to the function $V_{1,\theta}$ to get:

$$\begin{aligned} \frac{d}{d\theta} (E[V_{1,\theta}(X_{\varepsilon,\theta}) - V_{1,\theta}(0)]) &= E \left[\dot{V}_{1,\theta}(X_{\varepsilon,\theta}) + V'_{1,\theta}(X_{\varepsilon,\theta}) \frac{\partial}{\partial \theta} X_{\varepsilon,\theta} - \dot{V}_{1,\theta}(0) \right] \\ &= \frac{1}{2} E[\dot{V}''_{1,\theta}(\xi_1(X_{\varepsilon,\theta})) X_{\varepsilon,\theta}^2] \\ &\quad + \frac{1}{2} E \left[V''_{1,\theta}(\xi_2(X_{\varepsilon,\theta})) X_{\varepsilon,\theta} \frac{\partial}{\partial \theta} X_{\varepsilon,\theta} \right] \quad (\text{C.2}) \end{aligned}$$

where $\xi_i(X_{\varepsilon,\theta})$, $i = 1, 2$, lie between 0 and $X_{\varepsilon,\theta}$. Once these steps are justified, the key is then to determine the order of the two terms in (C.2).

This analysis requires several conditions, starting with the following conditions on V .

ASSUMPTION C 1 *We assume that $V(x)$ is four times differentiable. For $x > 0$, $0 \leq |V^{(n)}(x)| \leq C_v e^{v_+ x}$, and for $x < 0$, $0 \leq |V^{(n)}(x)| \leq C_v e^{v_- |x|}$ for some constants $C_v > 0$, $v_+ \in (0, s_+)$ and $v_- \in (0, s_-)$, $n = 0, \dots, 4$.*

The differentiability assumed here is not directly applicable to option payouts. We can always find a series of differentiable functions approximating a less smooth payout, and we can replace a payout with its conditional expectation shortly before maturity. However, our analysis does not address these additional sources of approximation error. In practice, the key distinction is generally between payouts that are Lipschitz continuous in the underlying asset price (such as standard calls and puts) and payouts with discontinuities. The pathwise method is generally inapplicable with discontinuities. The condition at $n = 0$ ensures that $E[V(X_\theta)]$ exists and is finite.

Our second assumption imposes conditions on the Lévy density q_θ and the MGF $\Phi_{\varepsilon,\theta}$ extended to the complex plane. The first item formally states conditions that we introduced earlier.

ASSUMPTION C 2

- 1) $q_\theta(y) > 0$ for all $y \in \mathbb{R}$; $q_\theta(y)$ is differentiable in θ ; and:

$$\int_{\mathbb{R}} q_\theta(y) dy = \infty$$

- 2) For any θ_0 there is an integrable function H_q for which, in the finite variation case, $(e^{sy}\mathbf{1}_{|y|\geq 1} + y\mathbf{1}_{|y|<1})|\dot{q}_\theta(y)| \leq H_q(y)$, and in the compensated case, $(e^{sy}\mathbf{1}_{|y|\geq 1} + y^2\mathbf{1}_{|y|<1})|\dot{q}_\theta(y)| \leq H_q(y)$ for all θ in a neighborhood of θ_0 , and for all $s \in (-s_-, s_+)$.

- 3) For any $s \in (-s_-, s_+)$, $|\Phi_\theta(s + iu)u|^2$ as a function of u is integrable.

Condition 1) of Assumption C 2 ensures that X_ε has a density, denoted by $f_{\varepsilon,\theta}(x)$, due to Theorem 27.4 of Sato (1999); we denote its CDF by $F_{\varepsilon,\theta}(x)$. Condition 2) ensures, for any $\varepsilon > 0$, the interchangeability of integration and differentiation in the following equation:

$$\frac{d}{d\theta} \int_{\mathbb{R}} e^{sy} q_\theta(y) \mathbf{1}_{|y|\geq \varepsilon} dy = \int_{\mathbb{R}} e^{sy} \dot{q}_\theta(y) \mathbf{1}_{|y|\geq \varepsilon} dy \quad \text{for } s \in (-s_-, s_+)$$

Using the bound:

$$\left| e^{s+iu} - \sum_{k=0}^n \frac{(s+iu)^k}{k!} \right| \leq \min \left\{ (e^{|s|} + 1) \frac{|s+iu|^n}{n!}, e^{|s|} \frac{|s+iu|^{n+1}}{(n+1)!} \right\} \quad (\text{C.3})$$

for $s, u \in \mathbb{R}$, condition 2) also implies that:

$$\frac{d}{d\theta} \int_{-\varepsilon}^{\varepsilon} (e^{sy} - 1 - sy) q_\theta(y) dy = \int_{-\varepsilon}^{\varepsilon} (e^{sy} - 1 - sy) \dot{q}_\theta(y) dy$$

Condition 3) of Assumption C 2 depends on the decay rate of $|\Phi_\theta(s + iu)|$ along a vertical line in the complex plane. If $|\Phi_\theta(s + iu)|$ has an exponential decay, then

condition 3) is automatically satisfied. Orey (1968) gives a condition on q_θ implying that $|\Phi_\theta(s + iu)|$ decays exponentially in u for any fixed s . If $|\Phi_\theta(s + iu)|$ has a power decay in u , Theorem 28.4 of Sato (1999) provides a way to estimate the power decay rate based on the behavior of q_θ near 0.

Returning to (C.2), we define:

$$D_{k,\theta}(\varepsilon) = \varepsilon^k \int_{|y| \geq \varepsilon} |\dot{q}_\theta(y)| dy + \int_{-\varepsilon}^{\varepsilon} |y|^k \dot{q}_\theta(y) dy$$

It is easy to see that $D_{3,\theta}(\varepsilon)$ is always of higher order than $D_{2,\theta}(\varepsilon)$.

Liu (2008) shows, with a weaker Assumption C2 and modified definition of $D_{k,\theta}$, that the first term in (C.2) is $O(M_{2,\theta}(\varepsilon))$ and that the second term in (C.2) is $O(D_{2,\theta}(\varepsilon))$. We will show in the next subsection that with our more restrictive Assumption C2 and the definition of $D_{k,\theta}$, this conclusion still holds. This means that the overall convergence order of the sensitivity estimate is bounded by whichever of $M_{2,\theta}(\varepsilon)$ and $D_{2,\theta}(\varepsilon)$ has the lower convergence order.

A similar analysis applies to $X_{2,\theta}$ and the error:

$$\frac{d}{d\theta} (E[V(X_{\varepsilon,\theta} + X_{1,\theta})] - E[V(\sigma_\theta(\varepsilon)W + X_{1,\theta})]) \quad (\text{C.4})$$

In this case, the convergence order is determined by whichever of $M_{3,\theta}(\varepsilon)$ and $D_{3,\theta}(\varepsilon)$ has the lower order. This demonstrates that including the normal correction term does indeed result in higher-order convergence, because $M_{3,\theta}(\varepsilon)$ and $D_{3,\theta}(\varepsilon)$ have higher order than $M_{2,\theta}(\varepsilon)$ and $D_{2,\theta}(\varepsilon)$, respectively. We summarize this discussion in the following theorem.

THEOREM C3 *Under Assumptions C1 and C2, the order of the bias in (C.1) is the order of the maximum of $M_{2,\theta}(\varepsilon)$ and $D_{2,\theta}(\varepsilon)$; the order of the bias in (C.4) is the order of the maximum of $M_{3,\theta}(\varepsilon)$ and $D_{3,\theta}(\varepsilon)$.*

We stress that all our results on convergence hold in the big- O sense and are thus upper bounds. The exact convergence rate in each case is potentially faster than indicated by our results.

C.1.2 Proof of Theorem C3

First, let us show that the MGF of $\dot{F}_{\varepsilon,\theta}(x)$ is $\dot{\Phi}_{\varepsilon,\theta}(s)/s$, $s \in (-s_-, s_+)$. Let $\dot{\Phi}_{\varepsilon,\theta}(s) = \Phi_{\varepsilon,\theta}(s)R_{\varepsilon,\theta}(s)$. Then:

$$\begin{aligned} R_{\varepsilon,\theta}(s) &= \int_{-\varepsilon}^{\varepsilon} (e^{sy} - 1 - sy)\dot{q}_\theta(y) dy + (e^{s\varepsilon} - 1 - s\varepsilon) \int_{\varepsilon}^{\infty} \dot{q}_\theta(y) dy \\ &\quad + (e^{-s\varepsilon} - 1 + s\varepsilon) \int_{\varepsilon}^{\infty} \dot{q}_\theta(-y) dy \end{aligned}$$

Applying (C.3), we can bound $|R_{\varepsilon,\theta}(s + iu)|$ as follows:

$$\begin{aligned} |R_{\varepsilon,\theta}(s + iu)| &\leq \frac{1}{2}e^{|\varepsilon|}|s + iu|^2 \int_{-\varepsilon}^{\varepsilon} |y^2 \dot{q}_{\theta}(y)| dy \\ &\quad + \frac{1}{2}e^{|\varepsilon|}|s + iu|^2 \varepsilon^2 \int_{|y| \geq \varepsilon} |\dot{q}_{\theta}(y)| dy \\ &= \frac{1}{2}e^{|\varepsilon|}|s + iu|^2 D_{2,\theta}(\varepsilon) \end{aligned} \quad (\text{C.5})$$

Then:

$$\left| \frac{\dot{\Phi}_{\varepsilon,\theta}(s + iu)}{s + iu} \right| \leq \frac{1}{2}e^{|\varepsilon|} D_{2,\theta}(\varepsilon) |\Phi_{\varepsilon,\theta}(s + iu)| |s + iu| \quad (\text{C.6})$$

where the right-hand side is an integrable function of u by condition 3) of Assumption C 2. So $|\dot{\Phi}_{\varepsilon,\theta}(s + iu)/(s + iu)|$ is an integrable function of u for any $s \in (-s_-, s_+)$.

Since $\Phi_{\varepsilon,\theta}(s)/s, s \in (-s_-, 0)$ is the MGF of $F_{\varepsilon,\theta}(x)$, and $-\Phi_{\varepsilon,\theta}(s)/s, s \in (0, s_+)$ is the MGF of $\bar{F}_{\varepsilon,\theta}(x) = 1 - F_{\varepsilon,\theta}(x)$, we have the inverse formula for $F_{\varepsilon,\theta}(x)$ (similar formula for $\bar{F}_{\varepsilon,\theta}(x)$):

$$F_{\varepsilon,\theta}(x) = \frac{e^{-sx}}{2\pi i} \int_{-\infty}^{\infty} e^{-iux} \frac{\Phi_{\varepsilon,\theta}(s + iu)}{s + iu} du$$

This, combined with (C.6), shows that $\dot{\Phi}_{\varepsilon,\theta}(s)/s$ is the MGF of $\dot{F}_{\varepsilon,\theta}(x)$.

Now we return to the proof of Theorem C 3. The first term of (C.2) is $O(M_{2,\theta}(\varepsilon))$, as shown in Liu (2008). We need to show the second term is $O(D_{2,\theta}(\varepsilon))$ using the new Assumption C 2 and definition of $D_{k,\theta}$.

Applying (8) and noting that $\dot{\Phi}_{\varepsilon,\theta}(s)/s$ is the MGF of $\dot{F}_{\varepsilon,\theta}(x)$, we have:

$$\begin{aligned} &\left| E \left[V''_{1,\theta}(\xi_2(X_\varepsilon)) X_\varepsilon \frac{\partial}{\partial \theta} X_\varepsilon \right] \right| \\ &= \left| \int_{\mathbb{R}} V''_{1,\theta}(\xi_2(x)) x \dot{F}_{\varepsilon,\theta}(x) dx \right| \\ &\leq C_v \int_0^\infty e^{v+x} |x \dot{F}_{\varepsilon,\theta}(x)| dx + C_v \int_{-\infty}^0 e^{-v-x} |x \dot{F}_{\varepsilon,\theta}(x)| dx \\ &\leq C_v \left(\int_0^\infty x^2 e^{-2\eta x} dx \right)^{1/2} \left(\int_{\mathbb{R}} (e^{(v+\eta)x} \dot{F}_{\varepsilon,\theta}(x))^2 dx \right)^{1/2} \\ &\quad + C_v \left(\int_0^\infty x^2 e^{2\eta x} dx \right)^{1/2} \left(\int_{\mathbb{R}} (e^{(v-\eta)x} \dot{F}_{\varepsilon,\theta}(x))^2 dx \right)^{1/2} \end{aligned}$$

for small $\eta > 0$ by the Schwartz inequality, and:

$$\begin{aligned} & \left| E \left[V''_{1,\theta}(\xi_2(X_\varepsilon)) X_\varepsilon \frac{\partial}{\partial \theta} X_\varepsilon \right] \right| \\ &= \frac{C_v}{\sqrt{2\eta}} \left(\frac{1}{2\pi} \int_{\mathbb{R}} \left| \frac{\dot{\Phi}_{\varepsilon,\theta}(v_+ + \eta + iu)}{v_+ + \eta + iu} \right|^2 du \right)^{1/2} \\ & \quad + \frac{C_v}{\sqrt{2\eta}} \left(\frac{1}{2\pi} \int_{\mathbb{R}} \left| \frac{\dot{\Phi}_{\varepsilon,\theta}(v_- + \eta + iu)}{v_- + \eta + iu} \right|^2 du \right)^{1/2} \end{aligned} \quad (\text{C.7})$$

The last equation is due to the Planchrel equality. By (C.5) and by condition 3) of Assumption C 2, (C.7) is bounded by $D_{2,\theta}(\varepsilon)$ times some constant. Therefore, the second term of (C.2) is $O(D_{2,\theta}(\varepsilon))$, which completes the proof.

C.1.3 Convergence rate of the bias in the VG and NIG models

As a consequence of Theorem C 3, we have the following proposition.

PROPOSITION C 4 *Suppose that Assumption C 1 holds and $T \geq 3\nu/4$. Then, in the VG model, the bias in the pathwise derivative estimate for ρ is $O(\varepsilon^2)$ using $X_1(T)$, and $O(\varepsilon^3)$ using $X_2(T)$.*

PROOF First, we verify that the VG process satisfies Assumption C 2. Condition 1) is easily verified. Since there are constants $c_1 > 0$ and $c_2 > 0$ such that $\dot{q}_{\text{vg},\rho}(y) \leq c_1 \exp(-c_2|y|)$, condition 2) is satisfied. From the form of $q_{\text{vg},\rho}$, we see that $M_{1,\rho}(\varepsilon) = O(\varepsilon)$ and $M_{2,\rho}(\varepsilon) = O(\varepsilon^2)$. Also, we have $|\Phi_{\text{vg},\rho}(s + iu)| = O(|u|^{-2t/\nu})$. So if $4t/\nu > 3$, then $|\Phi_{\text{vg},\rho}(s + iu)u|^2$ is integrable and condition 3) is satisfied.

With condition 3) of Assumption C 2 satisfied, we have $D_{2,\theta}(\varepsilon) = O(\varepsilon^2)$. So, by Theorem C 3, the error in the pathwise estimate using $X_1(T)$ is $O(\varepsilon^2)$. Similar calculations give $M_{3,\theta}(\varepsilon) = O(\varepsilon^4)$ and $D_{3,\theta}(\varepsilon) = O(\varepsilon^3)$, so the error using $X_2(T)$ is $O(\varepsilon^3)$. \square

In the NIG model, we consider sensitivities to the parameter δ . Interestingly, the convergence rates in this example are slower than in the previous one.

PROPOSITION C 5 *Suppose that Assumption C 1 holds. Then, in the NIG model, the bias in the pathwise derivative estimate for δ is $O(\varepsilon)$ using $X_1(T)$, and $O(\varepsilon^2)$ using $X_2(T)$.*

PROOF We need two properties of the function K_1 . Applying a bound for modified Bessel functions (Laforgia (1991)), it follows that:

$$\frac{K_1(x)}{K_1(y)} > e^{y-x}, \quad 0 < x < y$$

So $K_1(\alpha|y|) < \exp(\alpha\varepsilon - \alpha|y|)K_1(\alpha\varepsilon)$ for $|y| \geq \varepsilon$. The second property we need is $K_1(x) \sim 1/x$ as $x \rightarrow 0$, by which we have $\sigma_\delta^2(\varepsilon) = O(\varepsilon)$.

Now we check Assumption C2. Condition 1) is easily verified. By the properties of K_1 just noted, there exist $c_1 > 0$ and $c_2 > 0$, such that:

$$(e^{sy} \mathbf{1}_{|y| \geq 1} + y^2 \mathbf{1}_{|y| < 1}) |\dot{q}_{\text{nig}}(y)| \leq c_1 \exp(-c_2|y|)$$

Since $|\Phi_{\text{nig},\delta}(s + iu)|$ has an exponential decay along the vertical line, condition 3) is satisfied. Direct calculation shows that $D_{2,\delta}(\varepsilon)$ is $O(\varepsilon)$, and $M_{3,\delta}(\varepsilon)$ and $D_{3,\delta}(\varepsilon)$ are both $O(\varepsilon^2)$. \square

C.2 Pathwise method: truncated density

C.2.1 The method

We introduce a jump-size random variable \tilde{Y}_θ which has the same density as Y_θ on $(-\infty, -\varepsilon] \cup [\varepsilon, \infty)$, but with a positive density on $(-\varepsilon, \varepsilon)$. We choose the density of \tilde{Y}_θ on $(-\varepsilon, \varepsilon)$ to be proportional to a positive function g_θ . This will allow \tilde{Y}_θ to change continuously with θ and eliminates the need to separate positive and negative jumps. We impose the following assumption on $g_\theta(y)$.

ASSUMPTION C6 *The function $g_\theta(y)$ is integrable on $(-\varepsilon, \varepsilon)$ and differentiable in θ . Moreover, $\dot{g}_\theta(y)$ is bounded by an integrable function $H_g(y)$.*

We let:

$$\tilde{\lambda} = \int_{-\varepsilon}^{\varepsilon} g_\theta(y) dy + \lambda$$

denote the total arrival rate of jumps. We now require $d\tilde{\lambda}/d\theta = 0$ to make the pathwise method applicable to the new compound Poisson process. To achieve this, we let ε depend on θ and denote it by ε_θ . Setting $d\tilde{\lambda}/d\theta = 0$, we find that we must have:

$$\dot{\varepsilon}_\theta = \frac{\int_{|y| \geq \varepsilon_\theta} \dot{q}_\theta(y) dy + \int_{|y| \leq \varepsilon_\theta} \dot{g}_\theta(y) dy}{q_\theta(-\varepsilon_\theta) + q_\theta(\varepsilon_\theta) - g_\theta(\varepsilon_\theta) - g_\theta(-\varepsilon_\theta)} \quad (\text{C.8})$$

in contrast with (26). This mechanism for varying ε with θ is illustrated in Figure 6 on page 33.

The density of \tilde{Y}_θ is:

$$f_{\tilde{Y}_\theta}(y) = \begin{cases} \frac{q_\theta(y)}{\tilde{\lambda}}, & |y| \geq \varepsilon_\theta \\ \frac{g_\theta(y)}{\tilde{\lambda}}, & |y| < \varepsilon_\theta \end{cases}$$

The derivative of \tilde{Y}_θ with respect to θ is then given by:

$$\frac{\partial}{\partial \theta} \tilde{Y}_\theta = \begin{cases} \frac{-\dot{Q}_\theta(\tilde{Y}_\theta)}{q_\theta(\tilde{Y}_\theta)}, & \tilde{Y}_\theta \leq -\varepsilon_\theta \\ \frac{\dot{Q}_\theta(\tilde{Y}_\theta)}{q_\theta(\tilde{Y}_\theta)}, & \tilde{Y}_\theta \geq \varepsilon_\theta \\ \frac{\int_{\tilde{Y}_\theta}^{\varepsilon_\theta} \dot{g}_\theta(y) dy + (g(\varepsilon_\theta) - q_\theta(\varepsilon_\theta))\dot{\varepsilon}_\theta + \dot{Q}_\theta(\varepsilon_\theta)}{g_\theta(\tilde{Y}_\theta)}, & |\tilde{Y}_\theta| < \varepsilon_\theta \end{cases} \quad (\text{C.9})$$

Let:

$$\tilde{X}_{1,\theta}(t) = \tilde{c}_\theta t + \sum_{k=1}^{\tilde{N}(t)} \tilde{Y}_{k,\theta}$$

where $\tilde{N}(t)$ is a Poisson process with arrival rate $\tilde{\lambda}_\theta$, the $\tilde{Y}_{k,\theta}$ are iid with the same distribution as \tilde{Y}_θ and:

$$\tilde{c}_\theta = c_\theta - \int_{-\varepsilon_\theta}^{\varepsilon_\theta} y g_\theta(y) dy$$

Then $\tilde{X}_{1,\theta}(t)$ provides an alternative compound Poisson approximation to X_θ .

The derivative of $\tilde{X}_{1,\theta}(t)$ is:

$$\frac{\partial}{\partial \theta} \tilde{X}_{1,\theta}(t) = \frac{d\tilde{c}_\theta}{d\theta} t + \sum_{k=1}^{\tilde{N}(t)} \frac{\partial}{\partial \theta} \tilde{Y}_{k,\theta}$$

where:

$$\begin{aligned} \frac{d\tilde{c}_\theta}{d\theta} = & \dot{\mu}_\theta - \int_{\varepsilon_\theta \leq |y| \leq 1} y \dot{q}_\theta(y) dy - \int_{-\varepsilon_\theta}^{\varepsilon_\theta} y \dot{g}_\theta(y) dy \\ & + \varepsilon_\theta \dot{\varepsilon}_\theta (q_\theta(\varepsilon_\theta) - q_\theta(-\varepsilon_\theta) - g_\theta(\varepsilon_\theta) + g_\theta(-\varepsilon_\theta)) \end{aligned}$$

With the normal correction, we get:

$$\tilde{X}_{2,\theta}(t) = \tilde{\sigma}_\theta(\varepsilon_\theta) W(t) + \tilde{X}_{1,\theta}(t)$$

where $\tilde{\sigma}_\theta^2(\varepsilon_\theta) = \sigma_\theta^2(\varepsilon_\theta) - \sigma_{g,\theta}^2(\varepsilon_\theta)$, and:

$$\sigma_{g,\theta}^2(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} y^2 g_\theta(y) dy$$

So:

$$\frac{\partial}{\partial \theta} \tilde{X}_{2,\theta}(t) = \frac{d\tilde{\sigma}_\theta(\varepsilon_\theta)}{d\theta} W(t) + \frac{\partial}{\partial \theta} \tilde{X}_{1,\theta}(t)$$

with:

$$\begin{aligned}
 \frac{d\tilde{\sigma}_\theta^2(\varepsilon_\theta)}{d\theta} &= \frac{d\varepsilon_\theta}{d\theta} \varepsilon_\theta^2 (q_\theta(\varepsilon_\theta) + q_\theta(-\varepsilon_\theta) - g_\theta(\varepsilon_\theta) - g_\theta(-\varepsilon_\theta)) \\
 &\quad + \int_{-\varepsilon_\theta}^{\varepsilon_\theta} y^2 (\dot{q}_\theta(y) - \dot{g}_\theta(y)) dy \\
 &= \varepsilon_\theta^2 \left(\int_{|y| \geq \varepsilon_\theta} \dot{q}_\theta(y) dy + \int_{-\varepsilon_\theta}^{\varepsilon_\theta} \dot{g}_\theta(y) dy \right) \\
 &\quad + \int_{-\varepsilon_\theta}^{\varepsilon_\theta} y^2 (\dot{q}_\theta(y) - \dot{g}_\theta(y)) dy \tag{C.10}
 \end{aligned}$$

C.2.2 Convergence rate of the bias

The analysis goes through as before, with some changes in definition. Corresponding to $M_{k,\theta}(\varepsilon)$ and $D_{k,\theta}(\varepsilon)$, we define, for $k = 1, 2, 3, 4$:

$$\begin{aligned}
 \tilde{M}_{k,\theta}(\varepsilon) &= M_{k,\theta}(\varepsilon) + \int_{-\varepsilon}^{\varepsilon} |y|^k g_\theta(y) dy \\
 \tilde{D}_{k,\theta}(\varepsilon) &= D_{k,\theta}(\varepsilon) + \varepsilon^k \int_{-\varepsilon}^{\varepsilon} |\dot{g}_\theta(y)| dy
 \end{aligned}$$

With these definitions, we state the following result, proved in Liu (2008).

THEOREM C 7 *Under Assumptions C1–C6, the order of the bias of the pathwise estimate using $\tilde{X}_{1,\theta}$ is the order of the maximum of $\tilde{M}_{2,\theta}(\varepsilon)$ and $\tilde{D}_{2,\theta}(\varepsilon)$; the order of the bias using $\tilde{X}_{2,\theta}$ is the order of the maximum of $\tilde{M}_{3,\theta}(\varepsilon)$ and $\tilde{D}_{3,\theta}(\varepsilon)$.*

A simple choice of $g_\theta(y)$ would be to set it equal to some constant G , which may or may not depend on θ but should satisfy the constraint that $\tilde{\sigma}_\theta^2(\varepsilon_\theta) \geq 0$. If G does not depend on θ , a lot of formulas introduced above simplify. For example, $\dot{\varepsilon}_\theta$ reduces to:

$$\dot{\varepsilon}_\theta = \frac{\int_{|y| \geq \varepsilon_\theta} \dot{q}_\theta(y) dy}{q_\theta(-\varepsilon_\theta) + q_\theta(\varepsilon_\theta) - 2G}$$

Also on $(-\varepsilon_\theta, \varepsilon_\theta)$, $(\partial/\partial\theta)\tilde{Y}_\theta$ becomes:

$$\frac{\partial}{\partial\theta} \tilde{Y}_\theta = \frac{(G - q_\theta(\varepsilon_\theta))\dot{\varepsilon}_\theta + \frac{\partial}{\partial\theta} \bar{Q}_\theta(\varepsilon_\theta)}{G}$$

On the other hand, to achieve a higher order of convergence, it is better to choose G depending on θ so that $\tilde{\sigma}_\theta^2(\varepsilon_\theta) \equiv 0$ for all θ , which yields:

$$G_\theta = \frac{3}{2} \varepsilon_\theta^{-3} \sigma_\theta^2(\varepsilon_\theta)$$

Then:

$$\dot{\varepsilon}_\theta = \frac{\varepsilon_\theta^2 \int_{|y| \geq \varepsilon_\theta} \dot{q}_\theta(y) dy + 3 \int_{-\varepsilon_\theta}^{\varepsilon_\theta} y^2 \dot{q}_\theta(y) dy}{-2\varepsilon_\theta^2 (q_\theta(-\varepsilon_\theta) + q_\theta(\varepsilon_\theta)) - 3\varepsilon_\theta^{-3} \sigma_\theta^2(\varepsilon_\theta)}$$

$$\dot{G}_\theta = -\frac{3}{4} \varepsilon_\theta^{-3} \left(\varepsilon_\theta^2 \int_{|y| \geq \varepsilon_\theta} \dot{q}_\theta(y) dy + \int_{-\varepsilon_\theta}^{\varepsilon_\theta} y^2 \dot{q}_\theta(y) dy \right)$$

and:

$$\frac{\partial}{\partial \theta} \tilde{Y}_\theta = \frac{(\varepsilon_\theta - \tilde{Y}_\theta) \dot{G}_\theta + (G_\theta - q_\theta(\varepsilon_\theta)) \dot{\varepsilon}_\theta + \partial / \partial \theta \bar{Q}_\theta(\varepsilon_\theta)}{G_\theta}$$

for $\tilde{Y}_\theta \in (-\varepsilon_\theta, \varepsilon_\theta)$. This, however, involves more computation.

C.2.3 Convergence rates of the bias in the VG and NIG models

PROPOSITION C 8 *Suppose that Assumption C 1 holds and $T \geq 3\nu/4$. Then, in the VG model, the bias in the pathwise derivative estimate for ρ is $O(\varepsilon_\rho^2)$ using $\tilde{X}_1(T)$ and $O(\varepsilon_\rho^3)$ using $\tilde{X}_2(T)$.*

PROOF By simple algebra:

$$\tilde{\sigma}_\rho^2(\varepsilon_\rho) = \sigma_\rho^2(\varepsilon_\rho) - \varepsilon_\rho^3/3 = O(\varepsilon_\rho^2), \quad \tilde{D}_{2,\theta}(\varepsilon_\rho) = D_{2,\rho}(\varepsilon_\rho) = O(\varepsilon_\rho^2)$$

and:

$$\tilde{M}_{3,\rho}(\varepsilon_\rho) = O(\varepsilon_\rho^3), \quad \tilde{D}_{3,\theta}(\varepsilon_\rho) = D_{3,\rho}(\varepsilon_\rho) = O(\varepsilon_\rho^3)$$

□

The following result is established along the same lines as the previous examples.

PROPOSITION C 9 *Suppose that Assumption C 1 holds. Then, in the NIG model, the bias in the pathwise derivative estimate for δ is $O(\varepsilon_\delta)$ using $\tilde{X}_1(T)$, and $O(\varepsilon_\delta^2)$ using $\tilde{X}_2(T)$.*

C.3 Convergence rate of the bias in the pathwise–LRM method

The convergence rate analysis for the bias in Section 5.2 goes through essentially as before. Because ε no longer depends on θ , the MGF of $f_{\varepsilon,\theta}(x)$ is simply:

$$\Phi_{\varepsilon,\theta}(s) = \int_{-\varepsilon}^{\varepsilon} (e^{sy} - 1 - sy) \dot{q}_\theta(y) dy$$

and the definition of $D_{k,\theta}(\varepsilon)$ is changed to:

$$D_{k,\theta}(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} |y^k \dot{q}_\theta(y)| dy$$

With these changes, Theorem C 3 holds as before.

REFERENCES

- Asmussen, S., and Glynn, P. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer.
- Asmussen, S., and Rosinski, J. (2001). Approximations of small jumps of Lévy processes with a view towards simulation. *Journal of Applied Probability* **38**, 482–493.
- Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modeling. *Scandinavian Journal of Statistics* **24**, 1–13.
- Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance and Stochastics* **2**, 41–68.
- Carr, P., Geman, H., Madan, D. B., and Yor, M. (2002). The fine structure of asset returns: an empirical investigation. *Journal of Business* **75**, 305–332.
- Carr, P., and Wu, L. (2004). Time-changed Lévy processes and option pricing. *Journal of Financial Economics* **17**, 113–141.
- Cont, R., and Tankov, P. (2004). *Financial Modelling with Jump Processes*. Chapman and Hall/CRC, Boca Raton, FL.
- Fu, M.C., and Hu, J.-Q. (1997). *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*. Kluwer Academic Publishers, Norwell, MA.
- Glasserman, P. (2004) *Monte Carlo Methods in Financial Engineering*. Springer, New York.
- Glasserman, P., and Liu, Z. (2008). Sensitivity estimates from characteristic functions. *Operations Research*, to appear.
- Hong, L. J., and Liu, G. (2007). Pathwise estimation of probability sensitivities through terminating or steady-state simulations. Working Paper, Hong Kong University of Science and Technology.
- Jensen, J. L. (1995). *Saddlepoint Approximations*. Oxford University Press.
- Kéllezi, E., and Webber, N. (2004). Valuing Bermudan options when asset returns are Lévy processes. *Quantitative Finance* **4**, 87–100.
- Kienitz, J. (2008). A note on Monte Carlo Greeks for jump diffusions and other Lévy models. Working Paper, Deutsche Postbank AG, Bonn, Germany.
- Laforgia, A. (1991). Bounds for modified Bessel functions. *Journal of Computational and Applied Mathematics* **34**, 263–267.
- L'Ecuyer, P. (1990). A unified view of the IPA, SF, and LR gradient estimation techniques. *Management Science* **36**, 1364–1383.
- Liu, Z. (2008). Sensitivity estimates for Lévy-driven models in finance. PhD Thesis, Department of Industrial Engineering and Operations Research, Columbia University.
- Madan, D., Carr, P., and Chang, E. (1998). The variance gamma process and option pricing. *European Finance Review* **2**, 79–105.
- Madan, D., and Yor, M. (2005). CGMY and Meixner subordinators are absolutely continuous with respect to one-sided stable subordinators. Working Paper, University of Maryland.
- Orey, S. (1968). On continuity properties of infinitely divisible distribution functions. *Annals of Mathematical Statistics* **39**, 936–937.
- Rosinski, J. (2001). Series representations of Lévy processes from the perspective of point processes. In *Lévy Processes: Theory and Applications*, Barndorff-Nielsen, O. E., Mikosch, T., and Resnick, S. I. (eds), pp. 401–415. Birkhäuser.
- Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.

- Schoutens, W. (2003). *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley.
- Signahl, M. (2003). On error rates in normal approximations and simulation schemes for Lévy processes. *Stochastic Models* **19**, 287–298.
- Suri, R. (1983). Implementation of sensitivity calculations on a Monte Carlo experiment. *Journal of Optimization Theory and Applications* **40**, 625–630.
- Widder, D. V. (1941). *The Laplace Transform*. Princeton University Press.