Non-Affine Option Pricing

KYRIAKOS CHOURDAKIS

KYRIAKOS CHOURDAKIS is a lecturer in financial economics at Queen Mary, University of London in the U.K.
k.chourdakis@qmul.ac.uk

This article presents a procedure to approximate European option prices for models with stochastic volatilities and jumps, focusing on the non-affine case. Unlike simulation-based methods and methods that solve the relevant partial differential equations numerically, the approach exploits the probabilistic structure of the specification. A continuous-time Markov chain is constructed that mimics the underlying volatility processes. Option prices can consequently be derived for the approximating process in semi-closed form, that is, up to an integration error.

It is shown that as the state space of this chain becomes denser, the discrete state process converges to its continuous state counterpart, allowing a good approximation of the derivative contracts in question. The approach is generic and can be applied to a very wide class of stochastic volatility models, possibly augmented with jumps and a correlation structure. An experiment using three common models shows that an easy-to-implement approximation based on a relatively coarse grid can lead to very accurate option prices.

Models with stochastic volatilities and jumps are widely used by both academicians and practitioners to capture the salient features of various financial time series. Heston [1993] was the first to recognize that knowledge of the characteristic function of the log-return density can lead to efficient computation of option prices. Bakshi and Madan [2000] generalize this approach, proving the relationship between characteristic functions and option prices. Bakshi, Cao, and Chen [1997] provide examples of such characteristic functions for the jump-diffusion square root stochastic volatility model, while Duffie, Pan, and Singleton [2000] treat a general class of processes that they call *affine jump diffusions* (AJD).

The AJD class comprises processes whose components (drift, covariance matrix, and jump intensities) have an affine dependence on the state vector. Duffie, Pan, and Singleton [2000] show that the characteristic function of such processes will have an exponential affine form, with the coefficients solving a system of complex-valued ordinary differential equations. For specific models, such as the one in Bakshi, Cao, and Chen [1997], this system admits closed-form solutions; in such cases the option prices can be represented in semiclosed form by numerical inversion of the characteristic function.

Despite its richness, the class of AJDs might not be able to capture all features of the data. Affine square root volatility processes, such as the one in Heston [1993], do not allow rapid volatility increases such as are observed in the data; non-affine specifications, such as the log-variance models, are better able to accommodate such features.[1]

In addition, models with strong non-linearities in the drift or the volatility of the volatility constitute departures from the AJD class.[2] Other promising non-affine specifications include ones where the spot/volatility correlation and the jump distribution explicitly depend on the state process too.

Closed-form solutions for options in a non-affine environment are yet to be developed. Thus, one is limited to numerical approximations, normally constructed either by numerically solving the relevant partial differential equations or by simulating the underlying asset.

We present a generic methodology that allows good approximation of the characteristic function of the log-return density for the non-affine cases. This is achieved by building a continuous-time Markov chain with a progressively denser state space, which can approximate virtually any state process quite well. The correlation structure and the jump densities are also allowed to depend on the same state process.

The approximation could be viewed in the same light as the trinomial tree building procedure of Hull and White [1994]; in both cases, the instantaneous drift and volatility are replicated.[3] The approach can be seen as a generalization of the two-regime model of Naik [1993]; option prices for multistate regime-switching models follow naturally in semiclosed form.

The approximation of diffusions using Markov chains is not a novel approach. It has with roots in the seminal work of Kushner, summarized in Kushner [1990] and Dupuis and Kushner [2001]. For the diffusion to be approximated, a continuous-time Markov chain is constructed for a given state space, with probabilities that preserve the instantaneous drift and volatility structure.

For any given state, only the neighboring states need to be (instantaneously) reached for this construct, resembling a trinomial tree in continuous time.[4] If a jump diffusion is approximated, more states need to be reached.[5]

For ease of exposition, we consider one volatility factor. Multiple factors are a natural generalization, discussed in Kushner [1990], while approximations of jump diffusions are covered in Kushner and DiMasi [1978].

In the case of the option pricing problem, we show in Chourdakis [2002] that the characteristic function for hidden Markov chain models with jumps is available in closed form. Therefore, one can easily compute option prices for the approximating model in semiclosed form by inverting the characteristic function (following, for example, Bakshi and Madan [2000]).

As the Markov chain converges to the underlying process, the option prices based on this specification will converge to the true ones. A number of practical considerations are also suggested, providing a number of techniques to use if computation time is expensive.

Three experiments are run in order to assess the usefulness of this approach. Two affine specifications and one non-affine specification are used for this investigation. The square root stochastic volatility process without and with jumps is used as a yardstick, as the option prices can also be retrieved in semiclosed form as in Bakshi, Cao, and Chen [1997].

It is found that even a relatively coarse 20-point grid can offer very accurate prices in the absence of jumps. A tighter 80-point grid is necessary when jumps are allowed. The log-variance process serves as the non-affine example, where the approximated prices easily match their simulated counterparts.

I. APPROXIMATION PROCEDURE

Suppose that the data-generating process, under risk-neutrality, is summarized as

$$\mathrm{d}\log Y_t = \mu(x_{t-})\mathrm{d}t + \sigma(x_{t-})\mathrm{d}W_t + \mathrm{d}q(x_{t-})J(x_{t-}) \tag{1}$$

$$\mathrm{d}x_t = \alpha(x_{t-})\mathrm{d}t + \beta(x_{t-})\mathrm{d}Z_t \tag{2}$$

where x indicates the state process, which is assumed to be stationary; $\sigma(x)$ transforms the state process into the volatility of the diffusion; $q(x)$ is a Poisson process with intensity $\lambda(x)$; and $J(x)$ is the jump size. Finally, the covariance $\mathrm{d}\langle W, Z\rangle = \rho(x)\,\mathrm{d}\,t$ is allowed to be state-dependent too.

Concept of Local Consistency

The process x is approximated by a Markov chain x^h that lies in a discrete subset $\mathcal{R}^h \subset \mathbb{R}$ where h denotes the spacing between discrete points.[6] The instantaneous transition probabilities are collected in the rate matrix of the Markov chain $\mathbf{Q}^h$. The transitions over an infinitesimal interval δ are given by

$$\mathsf{P}\left\{x^h(t+\delta) = e_j | x^h(t) = e_i\right\} = 1_{(j=i)} + q^h_{j,i} + o(\delta)$$

where $\mathbf{1}_{(j=i)}$ is the indicator function ($\mathbf{1}_{(j=i)} = \mathbf{1}$ if $(j=i)$, and $\mathbf{1}_{(j=i)} = 0$ otherwise). The instantaneous transition probabilities are defined in such a way as to maintain the instantaneous drift and the volatility of x, obeying the local consistency concept as defined in Kushner [1990].

Suppose that the grid $\mathcal{R}^h$ tends to cover the domain of x as $h \downdownarrows 0$. The associated filtration is generated as $\mathcal{F}^h(t) = \sigma\ \{x^h(u),\ 0 \leqslant u \leqslant t\}$, and E_t^h denotes the expectation taken, conditional on this filtration. The Markov chain $x^h(t)$ that lies in the grid $\mathcal{R}^h$ will approximate the diffusion with infinitesimal drift $)\alpha(x)$ and variance $\beta(x)$ if the functions α and β are such that weak uniqueness holds for the stochastic differential Equation (2) and for some $\varrho > 0$ and $\delta > 0$, with $\Delta x^h(t)\ x^h(t+\delta) - x^h(t)$:

$$
\begin{aligned}
\mathsf{E}_t^h\left\{\Delta x^h(t)\right\} &= \alpha\left(x^h(t)\right)\delta + O(h^\varrho \delta) \\
\mathsf{E}_t^h\left\{\Delta x^h(t) - \mathsf{E}_t^h\left\{\Delta x^h(t)\right\}\right\}^2 &= \beta^2\left(x^h(t)\right)\delta + O(h^\varrho \delta) \\
\left|\Delta x^h(t)\right| &= O(h)
\end{aligned}
$$

One such approximation scheme can be found in Piccioni [1987]. The rate matrix $\mathbf{Q}^h$ is of the Jacobi form, with elements given by:

$$
\begin{aligned}
q_{i-1,i}^h &= \frac{1}{2h^2}\beta^2(x_i^h) + \frac{1}{h}\alpha^-(x_i^h) \\
q_{i,i}^h &= -\frac{1}{h^2}\beta^2(x_i^h) - \frac{1}{h}|\alpha(x_i^h)| \\
q_{i+1,i}^h &= \frac{1}{2h^2}\beta^2(x_i^h) + \frac{1}{h}\alpha^+(x_i^h)
\end{aligned} \tag{3}
$$

for all interior states, while the first and last [n-th] state are made reflective.[7] The superscripts $\pm$ denote the positive/negative part of a function, and the grid $\mathcal{R}^h = \{x_1^h, x_2^h, \ldots, x_n^h\}$. More accurate results for coarser grids can be constructed by setting the "corrected" matrix elements:

$$
\begin{aligned}
q_{i-1,i}^h &= \frac{1}{2h^2}\beta^2(x_i^h) - \frac{1}{2h}\alpha(x_i^h) \\
q_{i,i}^h &= -\frac{1}{h^2}\beta^2(x_i^h) \\
q_{i+1,i}^h &= \frac{1}{2h^2}\beta^2(x_i^h) + \frac{1}{2h}\alpha(x_i^h)
\end{aligned} \tag{4}
$$

Of course for the transition probabilities to be well defined, the grid cannot be too coarse; an upper value $h < \frac{\beta^2(x_i^h)}{|\alpha(x_i^h)|}$ has to be set. If this condition is violated, the uncorrected set of rate matrix elements can be used.

We consider two well-established specifications: the square root volatility process of Heston [1993], and the log-variance process of Melino and Turnbull [1990]. The former belongs to the affine class and is used to assess the quality of the approximation, while the latter is non-affine and illustrates the usefulness of the approach in those settings. The jump-diffusion square root volatility process of Bakshi, Cao, and Chen [1997] is considered to examine how the approximation procedure deals with the discontinuous case.

Square Root Process

The square root stochastic volatility model takes the form

$$
\begin{aligned}
\mathrm{d}\log Y_t &= r\mathrm{d}t - \frac{1}{2}x_{t-} + \sqrt{x_{t-}}\mathrm{d}W_t \\
\mathrm{d}x_t &= \vartheta(\bar{v} - x_{t-})\mathrm{d}t + \varphi\sqrt{x_{t-}}\mathrm{d}Z_t
\end{aligned}
$$

where the state variable x denotes the latent variance. After applying Itô's lemma to the state process in order to take logarithms, we parameterize the process as:

$$
\begin{aligned}
\mathrm{d}\log Y_t &= r\mathrm{d}t - \frac{1}{2}\exp x_{t-} + \exp\frac{x_{t-}}{2}\mathrm{d}W_t \\
\mathrm{d}x_t &= \vartheta\left(\left(\bar{v} - \frac{\varphi^2}{2\vartheta}\right)\exp(-x_{t-}) - 1\right)\mathrm{d}t + \varphi\exp\frac{-x_{t-}}{2}\mathrm{d}Z_t
\end{aligned}
$$

where x now denotes the logarithm of the variance.

The various functionals take the form

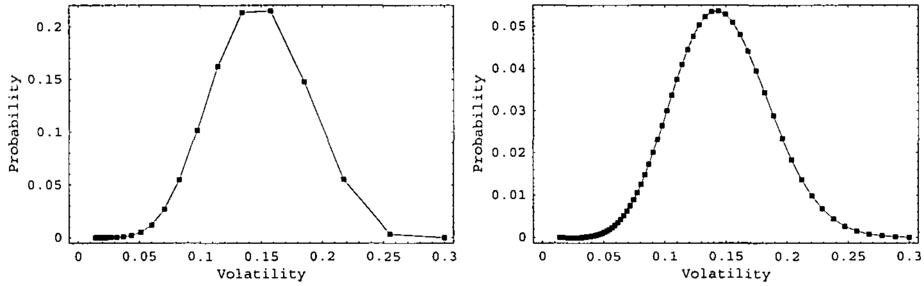
$$
\begin{aligned}
\sigma &: x \to \exp\frac{x}{2} \\
\alpha &: x \to \vartheta\left(\left(\bar{v} - \frac{\varphi^2}{2\vartheta}\right)\exp(-x) - 1\right) \\
\beta &: x \to \varphi\exp\frac{-x}{2}
\end{aligned}
$$

Assuming that $\vartheta > 0$, which imposes stationarity, and that $\bar{v} > \frac{\varphi^2}{2\vartheta}$, which is necessary for the variance to remain positive, one can apply Equations (4) to compute the non-zero elements of the rate matrix $\mathbf{Q}^h$. This matrix will have typical "corrected" elements:

$$
\begin{aligned}
q_{i-1,i}^h &= \frac{\varphi^2}{2h^2}\exp(-x_i^h) - \frac{\vartheta}{h}\left(\left(\bar{v} - \frac{\varphi^2}{2\vartheta}\right)\exp(-x_i^h) - 1\right) \\
q_{i,i}^h &= -\frac{\varphi^2}{h^2}\exp(-x_i^h) \\
q_{i+1,i}^h &= \frac{\varphi^2}{2h^2}\exp(-x_i^h) + \frac{\vartheta}{h}\left(\left(\bar{v} - \frac{\varphi^2}{2\vartheta}\right)\exp(-x_i^h) - 1\right)
\end{aligned}
$$

Exhibit 1

Examples of Grid Construction and Conditional Distribution of Volatility for Heston Model



A. 20-Point Density

B. 80-Point Density

Initial variance set at $v_0 = 0.02$.

Log-Variance Process

The non-affine example is the log-variance stochastic volatility model of Melino and Turnbull [1990], which has been estimated in Andersen, Benzoni, and Lund [1998], Chourdakis [2002], and Chernov et al. [2003]:

$$\mathrm{d}\log Y_t = r\mathrm{d}t - \frac{1}{2}\exp x_{t-} + \exp\frac{x_{t-}}{2}\mathrm{d}W_t$$

$$\mathrm{d}x_t = \vartheta(\log\bar{v} - x_{t-})\mathrm{d}t + \varphi\mathrm{d}Z_t$$

where the process x is the logarithm of the variance.

In this case, the various functionals take the form:

$$\sigma : x \to \exp\frac{x}{2}$$

$$\alpha : x \to \vartheta\left(\log\bar{v} - x\right)$$

$$\beta : x \to \varphi$$

Assuming that $\vartheta > 0$, a condition that ensures stationarity, the "corrected" rate matrix will have the typical elements:

$$q^h_{i-1,i} = \frac{\varphi^2}{2h^2} - \frac{\vartheta}{h}\left(\log\bar{v} - x^h_i\right)$$

$$q^h_{i,i} = -\frac{\varphi^2}{h^2}$$

$$q^h_{i+1,i} = \frac{\varphi^2}{2h^2} + \frac{\vartheta}{h}\left(\log\bar{v} - x^h_i\right)$$

In both examples, the instantaneous probabilities are intuitive. As the distance from the long-run mean increases, so does the probability of moving toward this long-run mean. As the discretization interval $h\downarrow 0$, the elements $q^h_{i-1,i}$, $q^h_{i-1,i} \to +\infty$, while $q^h_{i,i} \to -\infty$, implying that the trinomial structure converges to a binomial one with zero probability of remaining at the current state. In addition, it is trivial to verify that the local consistency conditions are satisfied.

Volatility Forecasts

The theory of continuous-time Markov chains can be used to forecast future volatilities. Given the initial volatility $\sigma(0)$ and its corresponding value on the grid $x(0) = e_i$, the conditional distribution of the future volatility can be easily expressed by the matrix exponential:

$$\left\{\mathsf{P}\left\{x^h(t) = e_j | x^h(0) = e_i\right\}\right\}_{j=1}^{n} = \iota^{\mathsf{T}} \cdot \exp\{t\mathbf{Q}^h\}$$

These conditional densities can provide useful information for construction of the approximating chain. In particular, the probabilities can be used to adjust the width of the grid in such a way as to ensure that as many as possible realizations of the latent process remain feasible.

Exhibit 1 presents the conditional densities for the Heston model. (The calibrating parameters of the process are the same as in Exhibit 4, Panel B.) The logarithm of the variance is modeled, and the support of the distribution is truncated to the interval $\mathcal{R} = \log(0.01 \times \bar{v}\}, 4.50 \times \bar{v}) \subset \mathbb{R}$. One can observe that this choice ensures that only variances with negligible probabilities are omitted.

Densities for two grid sets are presented. In Exhibit 1, Panel A, the interval $\mathcal{R}$ is divided using 20 equidistant points. In Panel B, the number of points is increased to 80.

Correlation Structure

State-dependent correlations can also be constructed, by rewriting an approximation of Equation (1), given that $x_{t-} = x_i^h$, as:

$$\mathrm{d}\log Y_t = \left\{\mu(x_i^h) - \rho(x_i^h)\frac{\sigma(x_i^h)}{\beta(x_i^h)}\alpha(x_i^h)\right\}\mathrm{d}t +$$

$$\sqrt{1-\rho^2(x_i^h)}\sigma(x_i^h)\mathrm{d}W_t^* +$$

$$\rho(x_i^h)\frac{\sigma(x_i^h)}{\beta(x_i^h)}\Delta x_t^h + \mathrm{d}q(x_i^h)J(x_i^h) \tag{5}$$

where the state changes Δx_t^h are given by

$$\Delta x_t^h = \begin{cases} +h & \text{with prob. } q_{i+1,i}^h \mathrm{d}t + o(\mathrm{d}t) \\ -h & \text{with prob. } q_{i-1,i}^h \mathrm{d}t + o(\mathrm{d}t) \\ 0 & \text{with prob. } 1-(q_{i+1,i}^h + q_{i-1,i}^h)\mathrm{d}t + o(\mathrm{d}t) \end{cases} \tag{6}$$

The process Δx_t^h is therefore usually zero, except when the Markov chain changes state. This is of course in line with the intuition behind the spot-volatility correlation, which dictates that *changes* in the volatility should be accompanied by *changes* in the conditional mean. This approach also follows Naik [1993] where fixed size jumps are considered to add correlation in a regime-switching framework.

As an example, if a constant correlation ρ is permitted, the log-variance process admits the approximation

$$\mathrm{d}\log Y_t = \left\{r - \frac{\exp x_i^h}{2} - \frac{\rho}{\varphi}\vartheta\left(\log\bar{v} - x_i^h\right)\exp\frac{x_i^h}{2}\right\}\mathrm{d}t +$$

$$\sqrt{1-\rho^2}\exp\frac{x_i^h}{2}\mathrm{d}W_t^* + \exp\frac{x_i^h}{2}\frac{\rho}{\varphi}\mathrm{d}x_t^h$$

II. OPTION PRICING PROCEDURE

We next explain the pricing of options and the various considerations involved.

Characteristic Function

We show in a theorem how to compute the characteristic function of a state-dependent process in closed form. This will serve as the approximation of the characteristic function of the initial process in Equations (1) and (2). The proof of the theorem can be found in Chourdakis [2002].

Theorem (characteristic function): Consider a process of the form (5) with the underlying Markov chain having rate matrix $\mathbf{Q}^h$ *with elements given in (6). Then, the characteristic function of* $\log\frac{Y(t)}{Y(0)}$*, given the initial state* $x(0) = x_i^h$*, will satisfy*

$$\phi_i(u) = \iota^\mathsf{T}\cdot\exp\{t\mathbf{B}(u)\}\cdot\mathbf{e}_i$$

where ι *is an* $(n\times 1)$ *vector of ones, and* $\mathbf{e}_i$ *is the i-th* $(n\times 1)$ *unit vector. The matrix function* $\mathbf{B}(u)$ *has elements of the form:*

$$\beta_{j,i}(u) = \begin{cases} q_{i,i}^h + \Psi_i^h(u) & \textit{when } j = i \\ q_{j,i}^h\Psi_{i\pm}^h(u) & \textit{when } j = i\pm 1 \\ 0 & \textit{otherwise} \end{cases}$$

The functions Ψ *are defined as*

$$\Psi_i^h(u) = \mathtt{i}\mu(x_i^h)u - \frac{1}{2}\left(1-\rho^2(x_i^h)\right)\sigma^2(x_i^h)u^2 +$$

$$\lambda(x_i^h)\left(\phi_J(u;x_i^h) - 1\right)$$

$$\Psi_{i\pm}^h(u) = \exp\left\{\pm\mathtt{i}\rho(x_i^h)\frac{\sigma(x_i^h)}{\beta(x_i^h)}hu\right\}$$

Finally, ϕ_J *is the characteristic function of the jump size.*

Option Pricing

The approximate characteristic function underlies the computation of option prices. Appendix A shows formally that under certain conditions option prices based on the approximating Markov chain converge to their true counterparts. The option price of the approximating model can be found by inverting the characteristic function of the log-return on maturity.

Bakshi and Madan [2000] show that under general assumptions, a call option with strike K and time to maturity t will be priced as:

$$C = S(0)\Pi_1 - K\exp\{-rt\}\Pi_2$$

where

$$\Pi_n = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left\{\frac{\phi_{[n]}(u)\exp\{-\mathrm{i}\theta\log\frac{K}{Y(0)}\}}{\mathrm{i}\theta}\right\}du$$

for $n = 1, 2$, and quantities defined as follows:

$$\phi_{[2]}(u) = \phi(u)$$
$$\phi_{[1]}(u) = \frac{\phi(u-\mathrm{i})}{\phi(-\mathrm{i})}$$

The function ϕ is the characteristic function of $\log\{Y(t)/Y(0)\}$. The result of the theorem leads to the approximation $C^h \approx C$ of the price, where the approximate characteristic function ϕ_i^h is used instead of ϕ. The subscript i indicates that the initial variance is equal to $\sigma^2(e_i)$, which is equal to the true initial variance.

Computational Considerations

The numerical inversion of the characteristic function can be avoided if one uses the FFT method of Carr and Madan [1999]. The FFT procedure will reduce the number of computations from $O(N^2)$ to $O(N\log_2 N)$, significantly shortening computation time. The FFT procedure requires N number of characteristic function calls, demanding N matrix exponential computations.

This approach, although theoretically appealing, might stretch computing resources to unacceptable levels. To avoid this, one can use knowledge of the characteristic function to compute the moments, and use them to compute the option prices. The moments are given as derivatives of ϕ_i^h, and the k-th uncentered moment is given by

$$M_k = \frac{1}{\mathrm{i}^k}\iota^{\mathsf{T}} \cdot \tilde{\mathbf{B}}_k^* \cdot \mathbf{e}_i$$

Say that the first m moments are used for the density approximation. Appendix B shows how the whole set of matrices $\{\tilde{\mathbf{B}}_k^*\}_{k=1}^m$ can be simultaneously computed. The procedure is very efficient, requiring the numerical computation of only one matrix exponential. The matrix exponentials of sparse matrices such as the ones involved here can be computed very rapidly using the ExpoKit procedures of Sidje [1998].

There are a number of ways to approximate a density based on the moments. Jarrow and Rudd [1982] and Jondeau and Rockinger [1998] give such methodologies using Hermite polynomials or Gram-Charlier expansions around the normal distribution. Gallant and Nychka [1987] construct a so-called semi-non-parametric density, widely used for simulation-based estimation procedures.

Although they are simple to implement, these methodologies generate densities that are typically multimodal and in some cases negative. This feature is not generally encountered in financial time series, which typically exhibit high skewness and kurtosis, but are typically unimodal. Moreover, if departures from normality are substantial, the inadequacies of the series approach are revealed in the tails of the distribution (as noted in Kendall and Stuart [1977]), which can make them unacceptable for option pricing purposes.

Here the Pearson IV density is used (see, for example, Kendall and Stuart [1977]), which can accommodate higher moments without sacrificing its unimodality and non-negativity. The Pearson IV density is based only on the first four moments, making its implementation straightforward.

Using this approach, one numerical integration is needed, and the call price is given by:

$$C = \exp\{-rT\}\int_{\log\frac{K}{Y(0)}}^{\infty} (Y(0)\exp y - K) f_P(y)dy$$

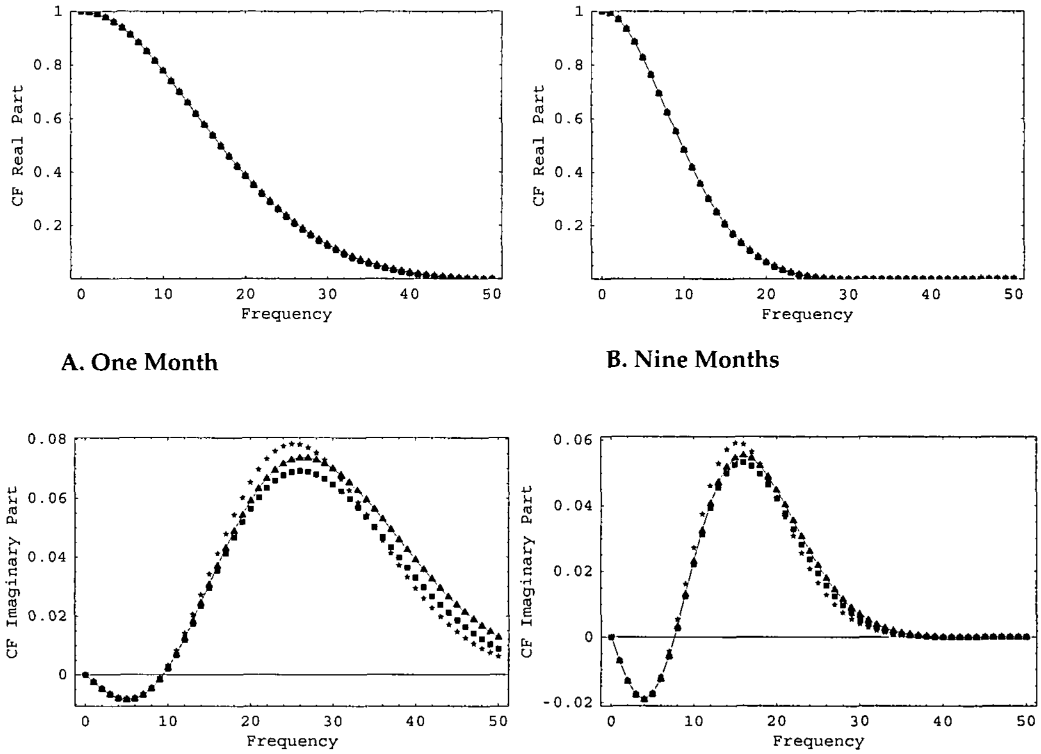
where f_P is the Pearson IV approximating density. Appendix C gives the form of the Pearson IV density, as well as the necessary condition. If the Pearson conditions are violated, an Edgeworth expansion is used instead.[8]

When a parametric density is fit to the theoretical moments, a second approximation is introduced. To assess the quality of both approximation steps, Exhibit 2 presents the real and imaginary parts for the approximate characteristic functions of the Heston model, whose characteristic function is available in closed form. Therefore, useful comparisons can be made.

Together with the exact characteristic function, three approximations are given. The first approximation uses the Markov chain method, the second approximation the Pearson IV density, and the third approximation an Edgeworth expansion.

One can clearly see the high quality of the Markov chain approximation. Both the real and imaginary parts of the characteristic function are matched. Note that there are apparent discrepancies between the true and the approximate densities, however. This suggests that a pro-

EXHIBIT 2
Approximate Characteristic Functions for Different Maturities



A. One Month
B. Nine Months
C. One Month
D. Nine Months

Solid line: true characteristic function for the Heston model. Triangles: approximate characteristic function derived from the Markov chain approximation method. Boxes: characteristic function after the Pearson IV density is fitted to the moments. Stars: characteristic function after the Edgeworth expansion density is fitted to the moments.

cedure that exploits the characteristic function directly should price options more accurately, but when we investigate pricing performance, we find that the density approximation gives very accurate option prices.

III. ASSESSMENT OF THE APPROXIMATIONS

We conduct an experiment to assess the pricing performance of the approximation procedure. Three different specifications are used in order to examine the robustness of the pricing performance with respect to different models, different parameter values, and the presence of jumps. The base parameters for these experiments are reported in Exhibit 3, following parameters estimated in the literature.[9]

Square Root Case

The square root stochastic volatility model (SV-H) of Heston [1993] is the cornerstone of this experiment.

EXHIBIT 3
Parameter Set Used in Calibration Exercise

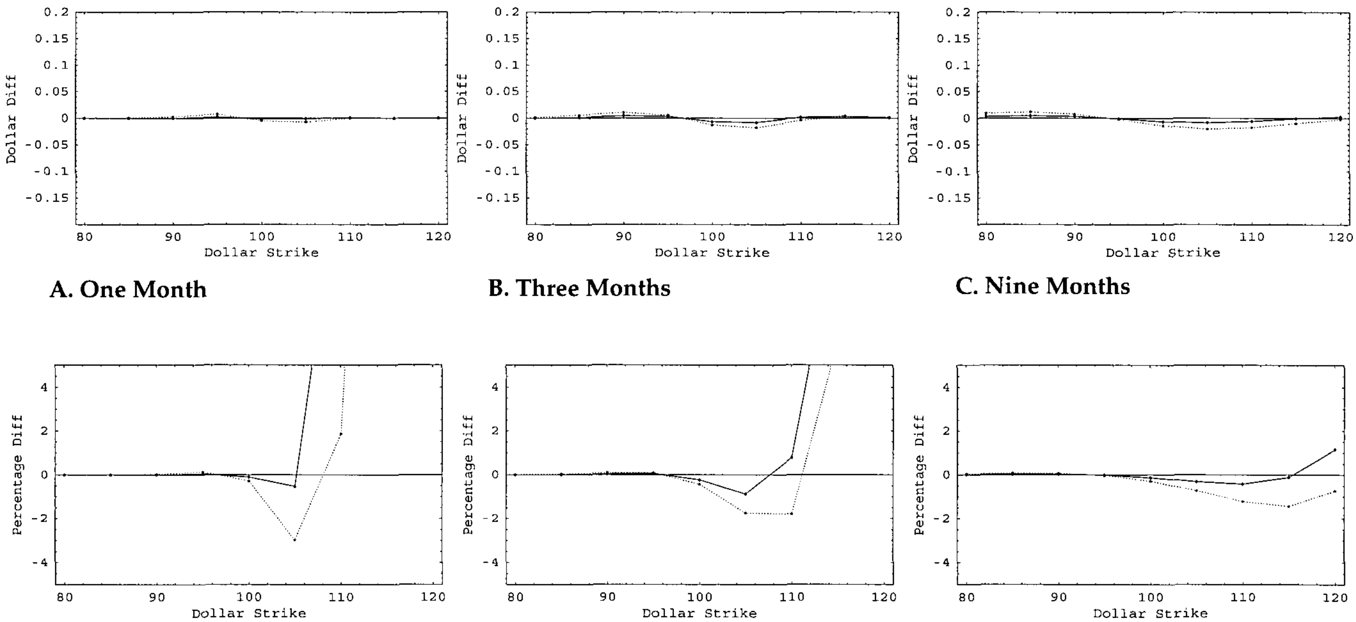
	μ	ϑ	$\bar{v}$	φ	ρ	λ_J	μ_J	σ_J
SV-H	0.0	7.0	0.02	0.5	−0.5			
SV-LV	0.0	5.0	ln 0.02	0.5	−0.5			
SV-H-J	0.0	7.0	0.02	0.5	−0.5	0.5	0.0	0.05

SV-H: Square root SV model. SV-LV: Log-variance SV model. SV-H-J: SV-H model augmented with jumps.

This specification belongs in the affine class, and its characteristic function is known explicitly, rendering option prices available in closed form. It can be used as a yardstick to explore the efficiency of the Markov chain approximation algorithm across different moneyness and maturity intervals.

Exhibit 4 presents the pricing errors of the approximation for different maturities.[10] The good performance of the Markov chain approximation is apparent. Errors are smaller than \$0.02 for options on an asset with spot

EXHIBIT 4
Maturity Effects



A. One Month B. Three Months C. Nine Months

D. One Month E. Three Months F. Nine Months

Dollar difference $C_{Heston} - C_{MCApprox}$ (Panels A-C), and relative difference $100 \times [(C_{Heston} - C_{MCApprox})/C_{Heston}]$ (Panels D-E) for different maturities. Solid line: approximation based on 80 points; dotted line: approximation based on 20 points. The parameter vector is $\{S_0, v_0, r, \vartheta, \overline{v}, \varphi, \rho\} = \{\$100, 0.02, 0, 4, 0.02, 0.3, -0.5\}$.

price \$100.00. The approximations show a very good fit even for relatively short maturities, and relative errors are reduced as the maturity lengthens.

Exhibits 5, 6, 7, and 8 examine the robustness of the approximation with respect to the initial volatility (v_0); the strength of mean-reversion parameter (ϑ); the volatility of volatility (φ); and the stock-volatility correlation (ρ). The time horizon in this experiment is set at three months.

The results are as expected. Pricing errors increase slightly as the volatility process becomes more erratic, namely, when mean reversion weakens, or when there is greater initial volatility, volatility of volatility, or correlation.

Nevertheless the pricing performance is very robust, and pricing errors remain below \$0.05. The at-the-money contracts are typically overpriced by 1% to 4%. Overall, the performance of the 20-point grid is close to the 80-point one, indicating that a coarse grid can be very effective in capturing the features of the underlying asset price process.

Jump-Diffusion Stochastic Volatility Case

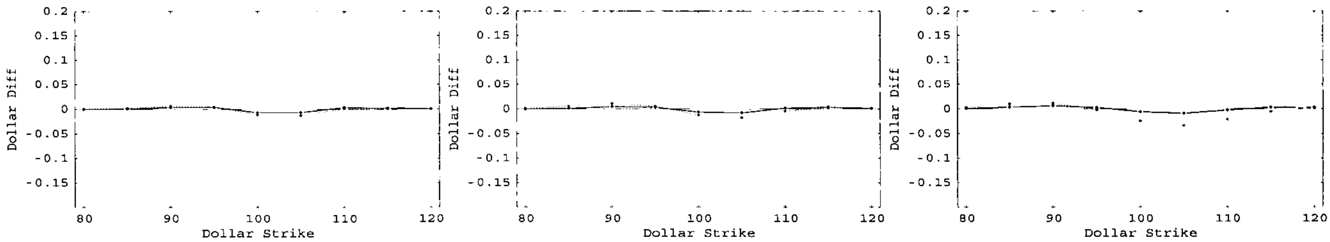
The presence of jumps is a well-documented feature of asset returns, examined in detail in Merton [1976], Bakshi, Cao, and Chen [1997], and Bates [1998, 2000]. An SV-H model augmented with simple Poisson/lognormal jumps that are independent of the state processes (SV-H-J) is in the affine class, with the characteristic function given in Bakshi, Cao, and Chen [1997], among others. Such a specification is investigated here, in order to examine whether or not the presence of such jumps affects the accuracy and efficiency of the algorithm.

Exhibit 9 presents the pricing performance of the approximation for different maturities, one, three, and nine months. The pricing errors are generally below \$0.10 for the 80-point grid. When a coarser grid is employed the errors increase rapidly, indicating that unlike the pure diffusion case a tight grid is significant when jumps are present.

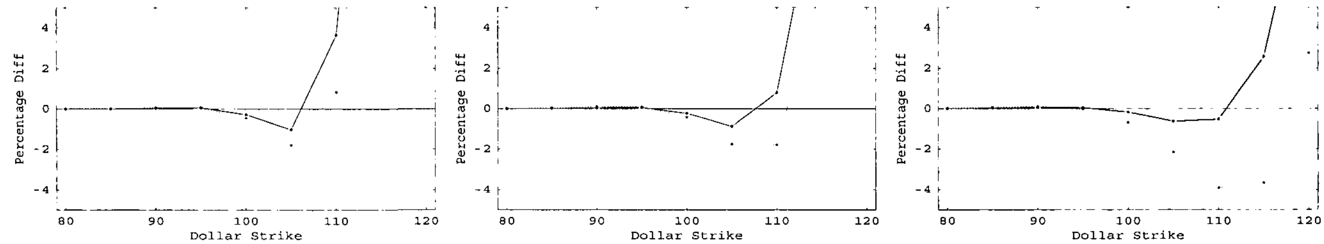
Exhibit 10 shows how the pricing errors change according to jump parameters. Extreme values for the expected jump size (μ_J), the jump volatility (σ_J), and the jump intensity (λ_J) are used to investigate their effects.

One can see the sensitivity of the approximation, mainly to changes in the distributional parameters of the jump size. Even in the extreme cases, however, the tight 80-point grid is fairly robust, and the pricing errors are less

EXHIBIT 5
Initial Volatility Effects



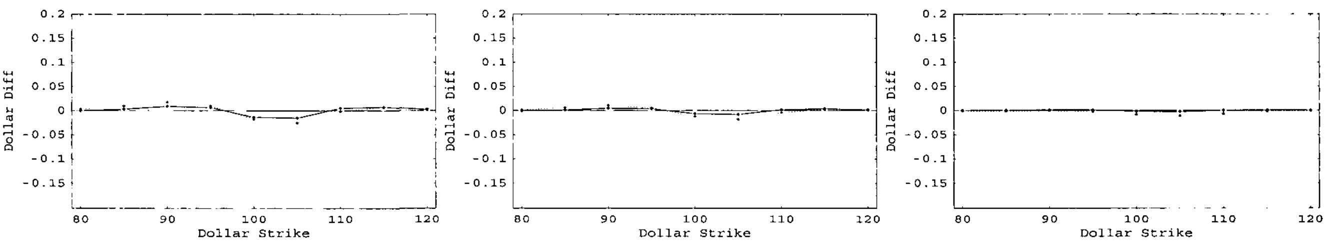
A. $v_0 = 0.01$ B. $v_0 = 0.02$ C. $v_0 = 0.04$



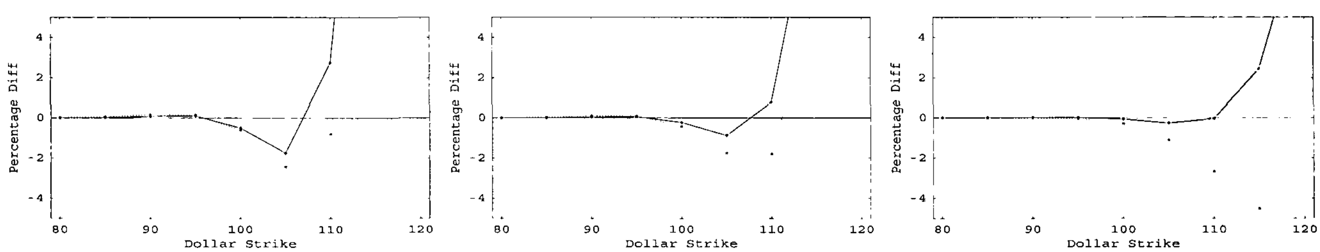
D. $v_0 = 0.01$ E. $v_0 = 0.02$ F. $v_0 = 0.04$

Dollar difference $C_{Heston} - C_{MCApprox}$ (Panels A-C), and relative difference $100 \times [(C_{Heston} - C_{MCApprox})/C_{Heston}]$ (Panels D-E) for different initial volatility values. Solid line: approximation based on 80 points; dotted line: approximation based on 20 points. Remaining parameter vector set at $\{S_0, r, \vartheta, \overline{v}, \phi, \rho\} = \{\$100, 0, 4, 0.02, 0.3, -0.5\}$. Maturity set at three months.

EXHIBIT 6
Mean Reversion Strength Effects



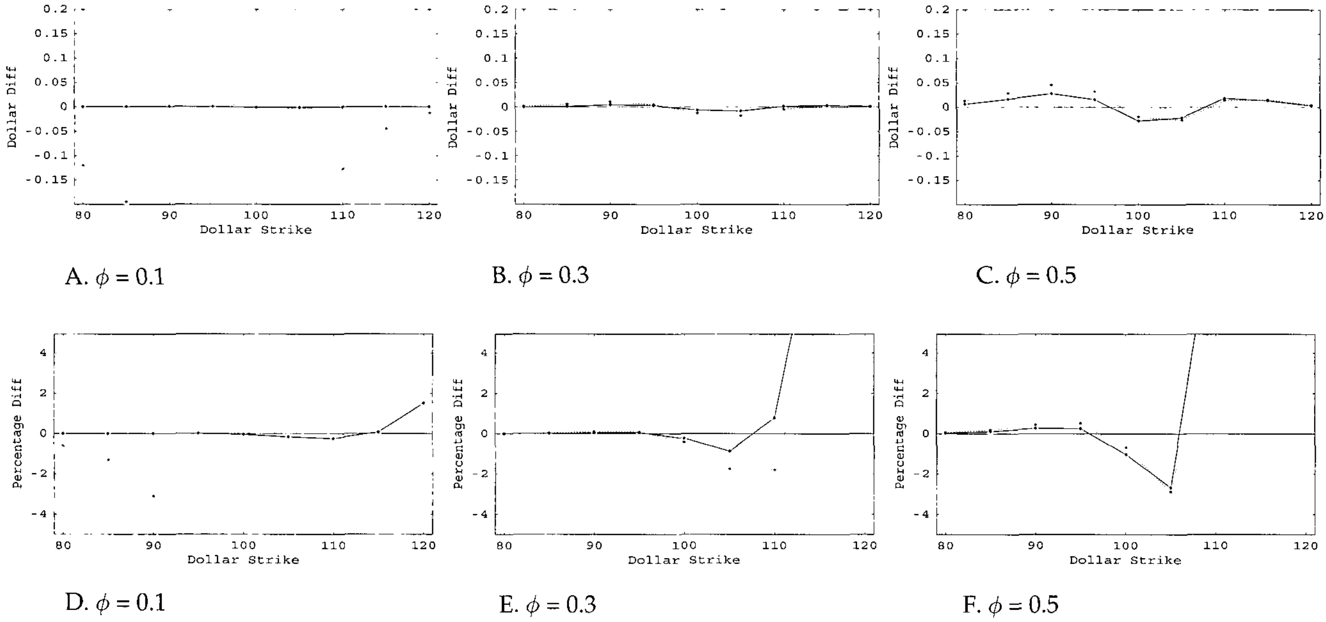
A. $\theta = 4$ B. $\theta = 7$ C. $\theta = 14$



D. $\theta = 4$ E. $\theta = 7$ F. $\theta = 14$

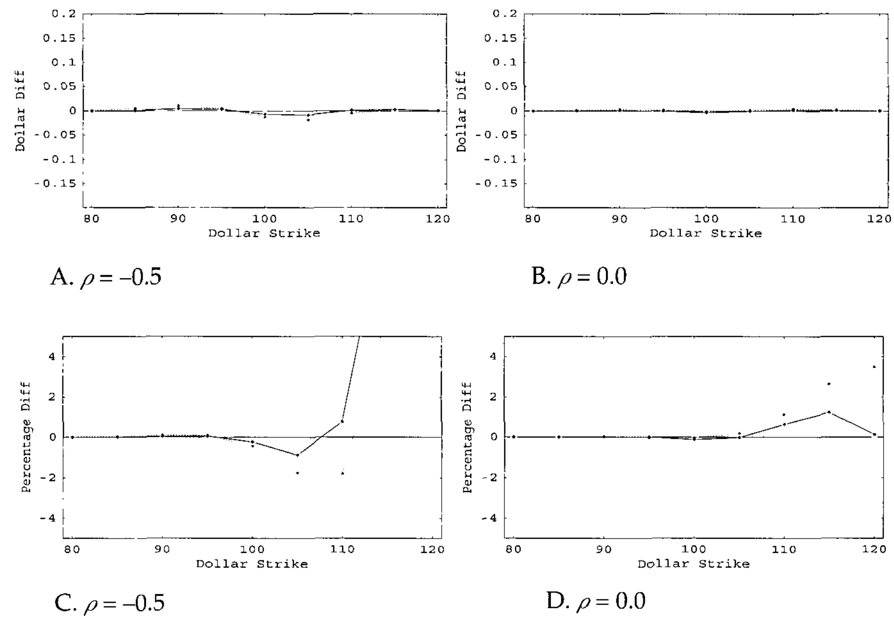
Dollar difference $C_{Heston} - C_{MCApprox}$ (Panels A-C), and relative difference $100 \times [(C_{Heston} - C_{MCApprox})/C_{Heston}]$ (Panels D-E) for different values of the mean reversion strength parameter θ. Solid line: approximation based on 100 points; dotted line: approximation based on 50 points. Remaining parameter vector set at $\{S_0, v_0, r, \overline{v}, \varphi, \rho\} = \{\$100, 0.02, 0, 0.02, 0.3, -0.5\}$. Maturity set at three months.

EXHIBIT 7
Volatility of Volatility Effects



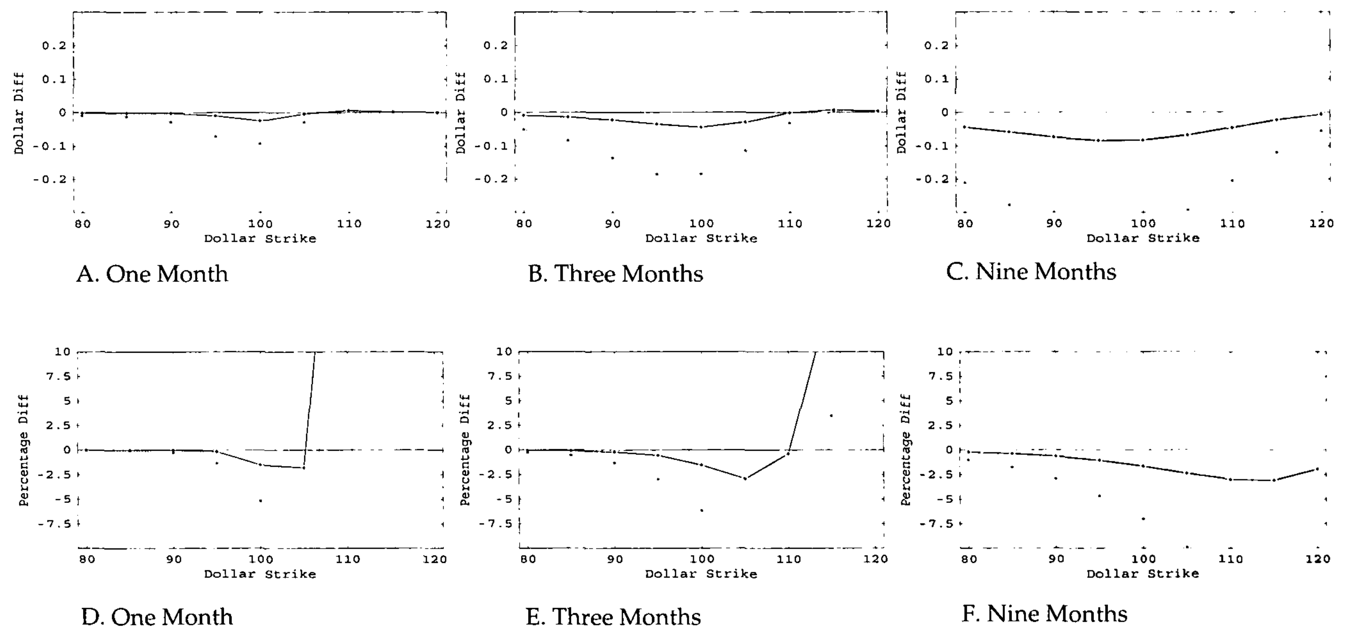
Dollar difference $C_{Heston} - C_{MCApprox}$ (Panels A-C), and relative difference $100 \times [(C_{Heston} - C_{MCApprox})/C_{Heston}]$ (Panels D-E) for different values of the volatility-of-volatility parameter ϕ. Solid line: approximation based on 80 points; dotted line: approximation based on 20 points. Remaining parameter vector set at $\{S_0, v_0, r, \vartheta, \overline{v}, \rho\} = \{\$100, 0.02, 0, 7, 0.02, -0.5\}$. Maturity set at three months.

EXHIBIT 8
Correlation Effects



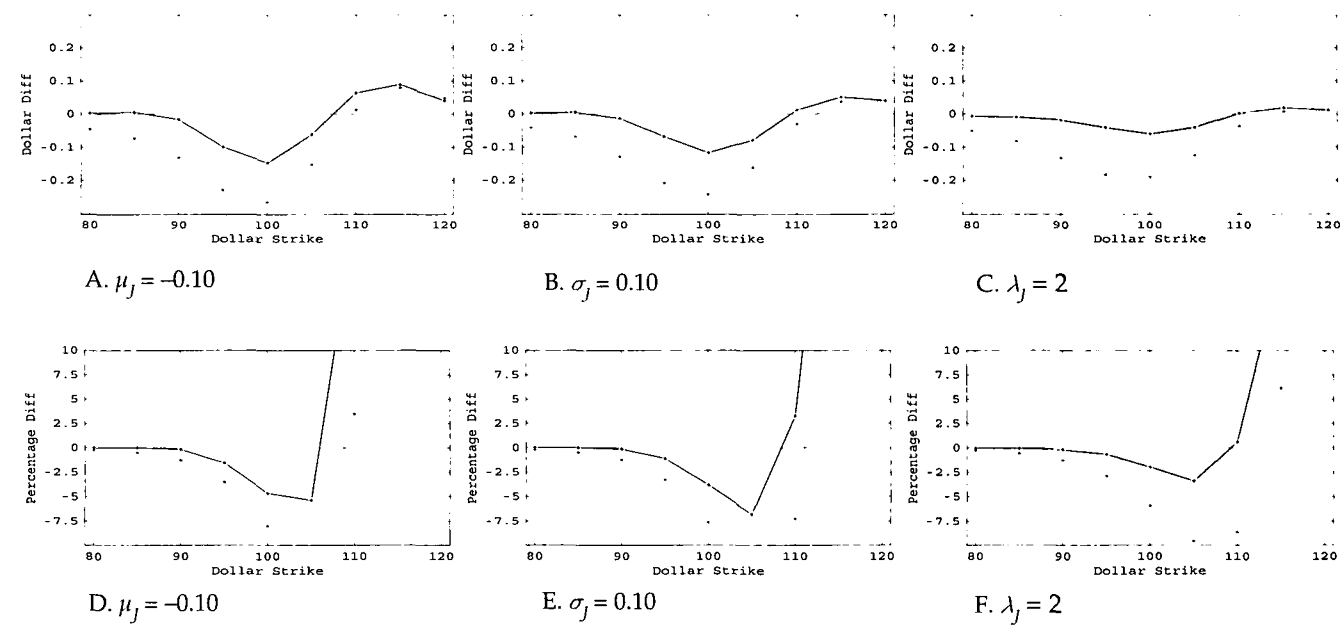
Dollar difference $C_{Heston} - C_{MCApprox}$ (Panels A-B), and relative difference $100 \times [(C_{Heston} - C_{MCApprox})/C_{Heston}]$ (Panels C-D) for different values of the correlation parameter ρ. Solid line: approximation based on 80 points; dotted line: approximation based on 20 points. Remaining parameter vector set at $\{S_0, v_0, r, \vartheta, \overline{v}, \varphi\} = \{\$100, 0.02, 0, 7, 0.02, 0.3\}$. Maturity set at three months.

EXHIBIT 9
Jump Effects for Different Maturities



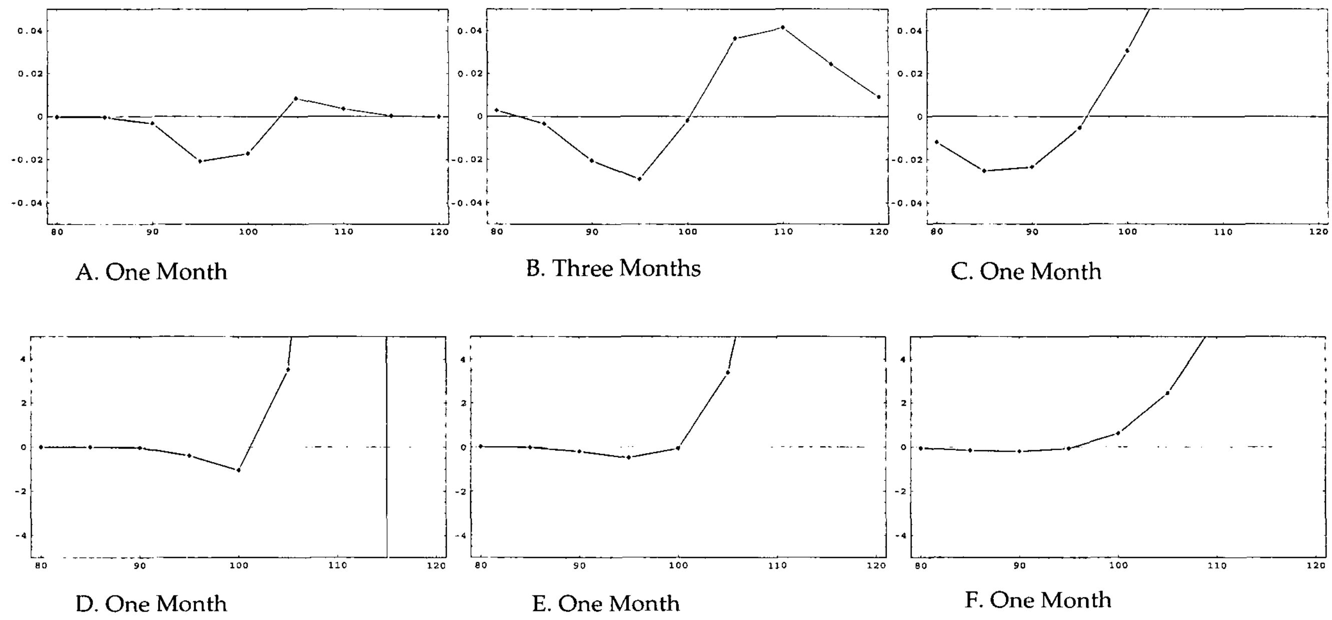
Dollar difference $C_{BCC} - C_{MCApprox}$ (Panels A-C), and relative difference $100 \times [(C_{BCC} - C_{MCApprox})/C_{BCC}]$ (Panels D-E) when jumps are included. Solid line: approximation based on 80 points; dotted line: approximation based on 20 points. Remaining parameter vector set at $\{S_0, v_0, r, \vartheta, \overline{v}, \varphi, \rho, \lambda_J, \mu_J, \sigma_J\}$ = {\$100, 0.02, 0, 7, 0.02, 0.3, −0.5, 0.5, 0, 0.05}.

EXHIBIT 10
Jump Parameter Effects



Dollar difference $C_{BCC} - C_{MCApprox}$ (Panels A-C), and relative difference $100 \times [(C_{BCC} - C_{MCApprox})/C_{BCC}]$ (Panels D-E) when the jump intensity (mean, volatility) is equal to 2 (−0.10, 0.10). Solid line: approximation based on 80 points; dotted line: approximation based on 20 points. Remaining parameter vector set at $\{S_0, v_0, r, \vartheta, \overline{v}, \varphi, \rho, \lambda_J, \mu_J, \sigma_J\}$ = {\$100, 0.02, 0, 7, 0.02, 0.3, −0.5, 0.5, 0, 0.05}. Maturity set at three months.

EXHIBIT 11
Price Differences for Non-Affine Log-Variance Model



Solid line: approximation based on 80 points. Parameter vector set at $\{S_0, v_0, r, \vartheta, \bar{v}, \rho,\} = \{\$100, 0.02, 0, 5, 0.02, 0.5, -0.5\}$.

than \$0.15 for a three-month call with underlying price of \$100.00. In relative terms, the worst mispricing occurs near the money, where the call is overpriced by as much as 5%.

Non-Affine Case

Pricing errors for the non-affine log-variance model (SV-LV) of Melino and Turnbull [1990] are presented in Exhibit 11 to illustrate the usefulness of the approach in computing option prices for such specifications. The log-variance stochastic volatility model (SV-LV) is used as the non-affine example, and the true option prices are computed using simulation.[11]

Overall, the approximation displays excellent performance; the pricing errors are under \$0.04. The relative mispricing of near-the-money contracts varies between zero and 2%.

IV. CONCLUSIONS

Our approximation methodology can be used to price options for a very general class of models with stochastic volatilities and jumps. The procedure can be very efficient in the case of non-affine specifications, where semiclosed-form solutions do not exist. The approximation is based on replication of the underlying diffusions using a progressively denser continuous-time Markov chain.

Although the approach is generic, allowing multifactor volatility structures and jumps that depend on the underlying processes, we illustrate it using three one-factor widely used models to assess the performance of the approximation procedure. When compared to their true (closed-form or simulated) results, option prices appear to be approximated well. The density approximation and the derived call prices are robust in terms of the model parameters and the nuisance parameters that dictate truncation of the processes' support. It is demonstrated furthermore that even a relatively coarse grid can result in quite accurate prices.

APPENDIX A

Convergence of Option Prices

This appendix shows that the option prices C^h computed using the Markov chain approximation converge to their true counterparts C.

Formally, the price of a European exercise derivative is given by $C^h = \mathsf{E}_0^h g(S_T^h)$, where g is the terminal payoffs, S^h is the underlying asset price at the maturity T, based on the Markov chain approximation procedure, and E_0^h is understood as the (risk-neutral) expectation at time zero, under the probability measure dictated by the Markov chain approximating process. $C = \mathsf{E}_0 g(S_T)$, where S_T and E_0 are the asset price and the expectation based on the diffusion in Equations (1) and (2).

As pointed out in Jacod and Shiryaev [1987, p. 516] and Kushner [1990], the weak convergence $x^h \Rightarrow x$, and subsequently $S^h \Rightarrow S$ will imply convergence in the distribution of $S_T^h \xrightarrow{\mathscr{D}} S_T$. Under the continuous mapping theorem one can conclude that $g(S_T^h) \xrightarrow{\mathscr{D}} g(S_T)$ for any continuous payoff function g, implying the convergence of the corresponding derivative prices $C^h = \mathsf{E}_0 g(S_T^h) \to \mathsf{E}_0 g(S_T) = C$ (see, for example, Billingsley [1968]).

The condition for this approximation and weak convergence to be valid are very general, namely, that the instantaneous drift and volatility are locally Lipschitz continuous functions. Then local consistency can be achieved. Local consistency ensures the convergence of option prices. For an example where the local consistency is not satisfied and the convergence breakes down, see Hubalek and Schachermayer [1998]. Weak convergence would ensure the convergence of all derivatives that have continuous payoffs.

APPENDIX B

Derivatives of the Matrix Exponential and Computation of $\tilde{\mathbf{B}}_k^*$

The procedure for differentiating the matrix exponential can be found in Mathias [1997]. Specifically, in order to calculate the first m moments, one has to take the first m derivatives, which can be obtained simultaneously using a procedure as follows:

1. Create a sequence of matrices

$$\mathbf{B}_k^* = \tfrac{1}{k!}\mathbf{B}^{(k)}(0) \text{ for } k \in \{0, \ldots, m\}$$

where $\mathbf{B}^{(k)}(0)$ has as elements the k-th derivatives of the elements of $\mathbf{B}(u)$ evaluated at $u = 0$. The elements of these matrices will have the form: For $k = 0$:

$$\mathbf{B}_0^*(u) = \mathrm{Q}^h$$

For $k = 1$, the matrix $\mathbf{B}_1^*$ will have elements:

$$\begin{cases} \mathrm{i}\mu(x_i^h) + \lambda(x_i^h)\phi_J'(0; x_i^h) & \text{when } j = i \\ q_{j,i}^h \left\{ \pm \mathrm{i}\rho(x_i^h)\frac{\sigma(x_i^h)}{\beta(x_i^h)} h \right\} & \text{when } j = i \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

For $k = 2$, the matrix $\mathbf{B}_2^*$ will have elements:

$$\frac{1}{2} \times \begin{cases} -\left(1 - \rho^2(x_i^h)\right)\sigma^2(x_i^h) + \lambda(x_i^h)\phi_J''(0; x_i^h) & \text{when } j = i \\ q_{j,i}^h \left\{ \pm \mathrm{i}\rho(x_i^h)\frac{\sigma(x_i^h)}{\beta(x_i^h)} h \right\}^2 & \text{when } j = i \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

For $k \geq 3$, the matrix $\mathbf{B}_k^*$ will have elements:

$$\frac{1}{k} \times \begin{cases} \lambda(x_i^h)\phi_J^{(k)}(0; x_i^h) & \text{when } j = i \\ q_{j,i}^h \left\{ \pm \mathrm{i}\rho(x_i^h)\frac{\sigma(x_i^h)}{\beta(x_i^h)} h \right\}^k & \text{when } j = i \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

2. Construct the block upper triangular block Toeplitz [BUTBT] matrix, using as blocks the elements of the matrix sequence above:

$$\tilde{\mathbf{B}} = \begin{pmatrix} \mathbf{B}_0^* & \mathbf{B}_1^* & \cdots & \mathbf{B}_m^* \\ \mathbf{0} & \mathbf{B}_0^* & \cdots & \mathbf{B}_{m-1}^* \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_0^* \end{pmatrix}$$

3. The matrix exponential of the matrix $\exp \tilde{\mathbf{B}}$ will have as block elements the $(N \times N)$ matrices of derivatives:

$$\left.\tfrac{1}{k!}\tfrac{\mathrm{d}^k}{\mathrm{d}\theta^k} \exp \mathbf{B}(u)\right|_{u=0}$$

In other words, the k-th derivative of the matrix exponential function $\mathbf{B}(u)$ evaluated at $u = 0$ will be given by $k!$ times the $(1, k)$ block [with dimensions $(N \times N)$] of the matrix $\exp \tilde{\mathbf{B}}$, denoted $\tilde{\mathbf{B}}_k^*$.

APPENDIX C

Pearson IV Density

Say that the first four *uncentered* moments are m_1 through m_4. The corresponding *centered* moments are given by:

$$\mu_2 = m_2 - m_1^2$$
$$\mu_3 = m_3 - 3m_2m_1 + 2m_1^3$$
$$\mu_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4$$

Measures of the skewness and kurtosis are given by:

$$\beta_1 = \left(\frac{\mu_3}{\mu_2^{3/2}}\right)^2$$
$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

The Pearson IV density exists if the condition:

$$4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6) > \beta_1(\beta_2 + 3)^2$$

is satisfied. Exhibit C illustrates the solution of the inequalities, in terms of the skewness and kurtosis coefficients. Then, one can define:

$$r = \frac{6(\beta_2 - \beta_1 - 1)}{2\beta_2 - 3\beta_1 - 6}$$
$$\nu = -\text{sign}(m_3)\frac{r(r-2)\sqrt{\beta_1}}{\sqrt{16(r-1) - \beta_1(r-2)^2}}$$
$$\alpha = \sqrt{\frac{\mu_2}{16}(16(r-1) - \beta_1(r-2)^2)}$$
$$m = \frac{r+2}{2}$$

Define the function f^* and its moments as

$$f^*(u) = \frac{\left(1 + \frac{u^2}{\alpha^2}\right)^{-m} \exp\left\{-\nu \tan^{-1}\frac{u}{\alpha}\right\}}{\int_{-\infty}^{+\infty}\left(1 + \frac{z^2}{\alpha^2}\right)^{-m} \exp\left\{-\nu \tan^{-1}\frac{z}{\alpha}\right\} \mathrm{d}z}$$

$$m_1^* = \int_{-\infty}^{+\infty} z f^*(z)\mathrm{d}z$$

$$m_2^* = \int_{-\infty}^{+\infty} (z - m_1^*)^2 f^*(z)\mathrm{d}z$$

The numerical integrals are very smooth, and are computed very efficiently. The Pearson IV density is constructed as

$$f_P(u) = \sqrt{\frac{m_2^*}{\mu_2}} f^*\left(\sqrt{\frac{m_2^*}{\mu_2}}(u - m_1^*) - m_1\right)$$

ENDNOTES

[1]For example, Chernov et al. [2003] add a jump component to the affine Heston [1993] specification, in order to compare its empirical performance with the non-affine log-variance specification. As discussed in Ghysels, Harvey, and Renault [1996], such an Ornstein-Uhlenbeck diffusion for the logarithm of the variance can be seen as the continuous-time limit of the popular AR(1)-EGARCH (1, 1).

[2]Aït-Sahalia [1999] suggests that the short interest rate exhibits a variable degree of mean reversion. He finds that the mean reversion is much stronger, the farther it is from the long-run mean. Similar arguments could well be made for the volatility process.

[3]There are two main differences between our approach of this paper and Hull and White [1994]: 1) the approximation here is set in continuous time, and 2) the underlying volatility rather than the asset price is approximated using the Markov chain.

[4]The trinomial tree has to be thought of in an *instantaneous* sense. In any positive time interval all states are reachable.

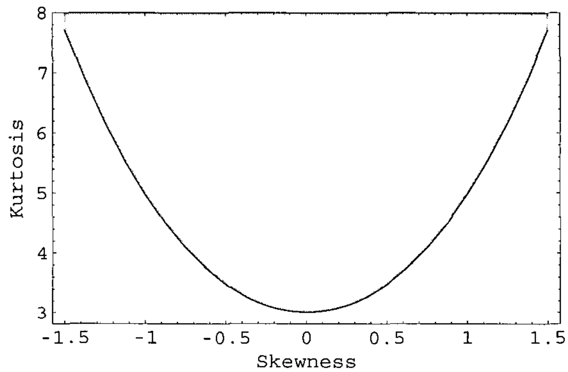
It is interesting to note that procedures such as trinomial trees must be used for the numerical solution of various PDEs encountered in finance; see, for instance, Hull [2003]. In these procedures *both* the time and the state space are discretized; an example is given in Benzoni [1999]. Here a different purely probabilistic approach is taken, where only the state space needs to be discretized. From another standpoint, Aït-Sahalia [2002] uses a similar Markov chain to investigate whether a discrete sample belongs to a diffusion.

[5]The process that is approximated is the volatility process. Jump diffusion here indicates a stochastic volatility with jumps *in the volatility*, and not the conditional mean. Normal jumps in the conditional mean are included in the first place.

[6]The points need not be equidistant. An equal distance is used here for ease of exposition, and since without loss of generality non-linearities are introduced by the function $\sigma(x)$.

[7]One could employ other schemes, such as maintain the instantaneous mean and volatility matching at the endpoints of the interval too. This would require transition from the endpoints to states that are not adjacent, rendering the rate matrix non-Jacobi and putting an extra burden on computation. Asymptotically all such schemes are equivalent, and indeed experimentation shows that even in small samples the differences are

EXHIBIT C
Admissible Skewness/Kurtosis Area for Pearson IV Density



negligible. The border states can alternatively be made absorbent. In experiments we conducted, the differences are also negligible.

[8]The necessary condition requires that the kurtosis exhibited be high enough for moderate values of skewness. In the Monte Carlo experiment, this is always satisfied, and the Pearson approximation is always used. If the condition fails, implying that the distribution is close to normal, density series expansions are justified instead.

[9]The literature on estimating and filtering stochastic volatility models is vast. Andersen, Benzoni, and Lund [1998], Chernov and Ghysels [2000], Chourdakis [2002], Pan [2002], and Eraker, Johannes, and Polson [2003], among others, serve as a guide in the choice of parameter values.

[10]The volatility support is truncated into the interval (0.01 $\times \bar{v}$ 4.5 $\times \bar{v}$), which excludes realizations with probabilities of order 10^{-6} or lower. Experiments with larger grids reveal that the results are very robust with respect to the width of the support, given that it is wide enough to include the bulk of the probability mass. Two grids are considered: a relatively coarse 20-point grid, and a denser 80-point grid. Option prices are based on the Pearson IV approximating density where applicable, and on an Edgeworth expansion when not.

[11]The simulations were performed using the Euler discretization scheme of the relevant diffusions. A time step of approximately 0.001 years is typically used, and 100,000 paths are typically simulated. An antithetic variable procedure is employed to reduce the variance.

REFERENCES

Aït-Sahalia, Y. "Telling from Discrete Data whether the Underlying Continuous-Time Model is a Diffusion." *Journal of Finance*, 57 (2002), pp. 2075–2112.

——. "Transition Densities for Interest Rate and Other Nonlinear Diffusions." *Journal of Finance*, 54 (1999), pp. 1361–1395.

Andersen, T.G., L. Benzoni, and J. Lund. "Estimating Jump Diffusions for Equity Returns." Working paper, Kellogg Graduate School of Management, Northwestern University, 1998.

Bakshi, G., C. Cao, and Z. Chen. "Empirical Performance of Alternative Option Pricing Models." *Journal of Finance*, 5 (1997), pp. 2003–2049.

Bakshi, G., and D. Madan. "Spanning and Derivative-Security Valuation." *Journal of Financial Economics*, 55 (2000), pp. 205–238.

Bates, D.S. "Post-'87 Crash Fears in S&P 500 Futures Options." *Journal of Econometrics*, 94 (2000), pp. 181–238.

——. "Pricing Options under Jump Diffusion Processes." Working Paper 37/88, The Wharton School, University of Pennsylvania, 1998.

Benzoni, L. "Pricing Options Under Stochastic Volatility: An Econometric Analysis." Working paper, Kellogg Graduate School of Management, Northwestern University, 1999.

Billingsley, P. *Convergence of Probability Measures*. New York: Wiley, 1968.

Carr, P., and D. Madan. "Option Valuation Using the Fast Fourier Transform." *Journal of Computational Finance*, 2 (1999), pp. 61–73.

Chernov, M., A.R. Gallant, E. Ghysels, and G. Tauchen. "Alternative Models for Stock Price Dynamics." *Journal of Econometrics*, 116 (2003), pp. 225–257.

Chernov, M., and E. Ghysels. "A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Option Valuation." *Journal of Financial Economics*, 56 (2000), pp. 407–458.

Chourdakis, K. "Continuous Time Regime Switching Models and Applications in Estimating Processes with Stochastic Volatility and Jumps." Working Paper 464, Queen Mary, University of London, 2002.

Duffie, D., J. Pan, and K. Singleton. "Transform Analysis and Asset Pricing for Affine Jump–Diffusions." *Econometrica*, 68 (2000), pp. 1343–1376.

Dupuis, P.G., and H.J. Kushner. *Numerical Methods for Stochastic Control Problems in Continuous Time*. Vol. 24 of *Applications of Mathematics*. New York: Springer-Verlag, 2001.

Eraker, B., M. Johannes, and N. Polson. "The Impact of Jumps in Returns and Volatility." *Journal of Finance*, 53 (2003), pp. 1269–1300.

Galland, R.A., and D.W. Nychka. "Seminonparametric Maximum Likelihood Estimation." *Econometrica*, 55 (1987), pp. 363–390.

Ghysels, E., A. Harvey, and E. Renault. "Stochastic Volatility." In G. Maddala and C. Rao, eds., *Handbook of Statistics*, 14, *Statistical Methods in Finance*. Amsterdam: North Holland, 1996.

Heston, S.L. "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options." *Review of Financial Studies*, 6 (1993), pp. 327–344.

Hubalek, F., and W. Schachermayer. "When Does Convergence of Asset Price Processes Imply Convergence of Option Prices?" *Mathematical Finance*, 8(4) (1998), pp. 215-233.

Hull, J.C. *Options, Futures and Other Derivatives*, 5th ed. Englewood Cliffs, NJ: Prentice-Hall, 2003.

Hull, J.C., and A. White. "Numerical Procedures for Implementing Term Structure Models I: Single Factor Models." *The Journal of Derivatives*, 2 (1994), pp. 7–16.

Jacod, J., and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. New York: Springer-Verlag, 1987.

Jarrow, R., and A. Rudd. "Approximate Option Valuation for Arbitrary Stochastic Processes." *Journal of Financial Economics*, 10 (1982), pp. 347–349.

Jondeau, E., and M. Rockinger. "Estimating Gram-Charlier Expansions Under Positivity Constraints." Working Paper 653/1998, Cahier de Recherche du Groupe, HEC, 1998.

Kendall, M., and A. Stuart. *The Advanced Theory of Statistics*, vol. I, 4th ed. London: Charles Griffin and Co, 1977.

Kushner, H.J. "Numerical Methods for Stochastic Control Problems in Continuous Time." *SIAM Journal of Control and Optimization*, 28 (5) (1990), pp. 999–1048.

Kushner, H.J., and G. DiMasi. "Approximations of Functionals and Optimal Control Problems on Jump Diffusion Processes." *Journal of Mathematical Analysis and Applications*, 63 (1978), pp. 772–800.

Mathias, R. "A Chain Rule for Matrix Functions and Applications." *SIAM Journal on Matrix Analysis and Applications*, 17 (2) (1997), pp. 610–620.

Melino, A., and M. Turnbull. "Pricing Foreign Currency Options with Stochastic Volatility." *Journal of Econometrics*, 45, (1990), pp. 239–265.

Merton, R. "Option Pricing When the Underlying Stock Returns are Discontinuous." *Journal of Financial Economics*, 4 (1976), pp. 125–144.

Naik, V. "Option Valuation and Hedging Strategies with Jumps in the Volatility of Asset Returns." *Journal of Finance*, 48 (1993), pp. 1969–1984.

Pan, J. "The Jump Risk-Premia Implicit in Options: Evidence from an Integrated Time Series Study." *Journal of Financial Economics*, 63 (2002), pp. 3–50.

Piccioni, M. "Convergence of Implicit Discretization Schemes for Linear Differential Equations with Application to Filtering." In G.D. Prato and L. Tubaro, eds., *Stochastic PDEs and Applications, Lecture Notes in Mathematics 1236*. New York: Springer-Verlag, 1987.

Sidje, R.B. "EXPOKIT: A Software Package for Computing Matrix Exponentials." *ACM Trans. Math. Softw.*, 24 (1) (1998), pp. 130–156.