Journal of Computational Finance (1–42) Volume 11/Number 3, Spring 2008

Simple and efficient simulation of the Heston stochastic volatility model

Leif Andersen
Banc of America Securities, 9 West 57th Street, New York, NY 10019, USA;
email: leif.andersen@bofasecurities.com

Stochastic volatility models are increasingly important in practical derivatives pricing applications, yet relatively little work has been undertaken in the development of practical Monte Carlo simulation methods for this class of models. This paper considers several new algorithms for time-discretization and Monte Carlo simulation of Heston-type stochastic volatility models. The algorithms are based on a careful analysis of the properties of affine stochastic volatility diffusions, and are straightforward and quick to implement and execute. Tests on realistic model parameterizations reveal that the computational efficiency and robustness of the simulation schemes proposed in the paper compare favorably with existing methods.

1 INTRODUCTION

Square-root diffusions take a central role in several important models in finance, including the CIR interest rate model (Cox *et al* (1985)), the Heston stochastic volatility model (Heston (1993)), and the general affine model in (Duffie *et al* (2000)). Of particular interest to us here is the Heston model, where a recent reformulation of the original Fourier integrals in Heston (1993) (see Lewis (2001) and Lipton (2002), and also Carr and Madan (1999) and Lee (2004)) has made computations of European option prices numerically stable and efficient, allowing for quick model calibration to market prices. Partly as a result of this development, there has been much interest recently in embedding the Heston diffusion dynamics (or variations thereof) in derivatives pricing models, as a means to capture volatility smiles and skews in quoted markets for options. For interest rate applications, see, for example, Andersen and Andreasen (2002), Andersen and Brotherton-Ratcliffe (2005) and Piterbarg (2005); for foreign exchange see Andreasen (2006); for equity options, see the aforementioned Lewis (2001) and Lipton (2002).

Many practical applications of models with Heston dynamics involve the pricing and hedging of path-dependent securities, which, in turn, nearly always require the introduction of Monte Carlo methods. Despite the fact that the Heston model is nearly 15 years old, there has been remarkably little research into

The author is indebted to Jesper Andreasen, Victor Piterbarg and Vladimir Piterbarg for insights and assistance. A special thank you goes to Mike Staunton whose VBA implementation uncovered several typos in earlier drafts.

efficient discretization of the continuous-time Heston dynamics for purposes of Monte Carlo simulation. Recently, however, a few papers have emerged. Kahl and Jackel (2005) propose application of an implicit Milstein scheme for the square-root diffusion of the variance process, coupled with a particular discretization for the asset process; the scheme is designed to work particularly well for cases where there is significant negative correlation between the asset process and the variance process. Broadie and Kaya (2006) have developed a completely bias[1]-free scheme based on acceptance–rejection sampling of the variance process coupled with certain Fourier inversion computations. While elegant, the Broadie and Kaya (2006) scheme has a number of practical drawbacks, including complexity and lack of computational speed. At the cost of introducing some (small) amount of bias, Smith (2007) has been able to improve on the latter, by introducing certain approximations to the characteristic function of time-integrated variance, enabling efficient caching techniques. A much simpler approach was considered by Lord *et al* (2006) who introduce an Euler scheme equipped with certain rules to deal with the fact that the variance process can become negative in a direct Euler discretization. The authors conclude that the computational efficiency of the scheme exceeds that of the more complicated schemes in Kahl and Jackel (2005) and Broadie and Kaya (2006). While this conclusion overall appears sound, the resulting discretization scheme is largely heuristic and uses essentially none of the known analytical results for the Heston model. For many cases of practical interest, the resulting discretization bias in the Euler scheme is, as we shall see later, quite high at a practical number of time steps.

The contribution of this paper is two-fold. First, we propose a number of schemes designed to extract selected elements of the full-blown Broadie–Kaya scheme, yet retain the speed, simplicity and ease of implementation that characterize the Euler scheme. Second, we provide numerical tests of the resulting schemes on realistic – and challenging – market data. In our opinion, most of the tests done in previous literature are far too easy to pass, typically involving at-the-money options with short maturities and overly benign model parameters (low correlations, low volatility-of-variance and high variance mean reversion). In practical applications on interest rate and foreign exchange (FX) markets, implied model parameters are often quite extreme and option maturities can easily stretch to 15–20 years or more.

The paper is organized as follows: in Section 2 we outline the basic Heston dynamics and summarize a few relevant facts about the processes involved, and we also briefly introduce existing discretization schemes from the literature, for later comparative experiments; in Section 3 we consider two different schemes to discretize the variance process, paying attention to computational issues throughout; Section 4 discusses how to combine the variance process discretization with

[1] We remind the reader that approximating the expectation of a (function of a) continuous-time diffusion process by Monte Carlo methods involves two types of errors: (1) the statistical error ("noise") common to all Monte Carlo applications; and (2) a bias caused by the specific way the continuous-time diffusion process is approximated by a discrete-time process. See Section 5 for further details.

a discretization for the asset process, with an emphasis on techniques for the case where the variance process and the asset process have correlation far from zero; numerical comparisons can be found in Section 5, and a brief discussion of extensions is in Section 6; finally, Section 7 concludes the paper.

2 HESTON MODEL BASICS

2.1 SDE and basic properties

The Heston model is defined by the coupled two-dimensional stochastic differential equation (SDE):

$$\mathrm{d}X(t)/X(t) = \sqrt{V(t)}\,\mathrm{d}W_X(t) \tag{1}$$

$$\mathrm{d}V(t) = \kappa(\theta - V(t))\,\mathrm{d}t + \varepsilon\sqrt{V(t)}\,\mathrm{d}W_V(t) \tag{2}$$

where $\kappa, \theta, \varepsilon$ are strictly positive constants, and where W_X and W_V are scalar Brownian motions in some probability measure; we assume that $\mathrm{d}W_X(t) \cdot \mathrm{d}W_Y(t) = \rho\,\mathrm{d}t$, where the correlation ρ is some constant in $[-1, 1]$. $X(t)$ represents an asset price process (eg, a stock, an FX rate and so forth) and is assumed to be a martingale in the chosen probability measure; adding a drift to X is trivial and is omitted for notational simplicity. $V(t)$ represents the instantaneous variance of relative changes to $X(t)$, in the sense that the quadratic variation of $\mathrm{d}X(t)/X(t)$ over $[t, t + \mathrm{d}t]$ is $V(t)\,\mathrm{d}t$. $V(t)$ is modeled as a mean-reverting square-root diffusion, with dynamics similar to the celebrated interest rate model of Cox *et al* (1985).

Several analytical results exist for the behavior of the process (1)–(2), see, for example, Andersen and Piterbarg (2005), Dufresne (2001) and Cox *et al* (1985). We list a few of these results here.

PROPOSITION 1 *Let $F_{\chi^2}(y; \nu, \lambda)$ be the cumulative distribution function for the non-central chi-squared distribution with ν degrees of freedom and non-centrality parameter λ:*

$$F_{\chi^2}(z; \nu, \lambda) = \mathrm{e}^{-\lambda/2} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!\,2^{\nu/2+j}\,\Gamma(\nu/2 + j)} \int_0^z x^{\nu/2+j-1}\,\mathrm{e}^{-x/2}\,\mathrm{d}x$$

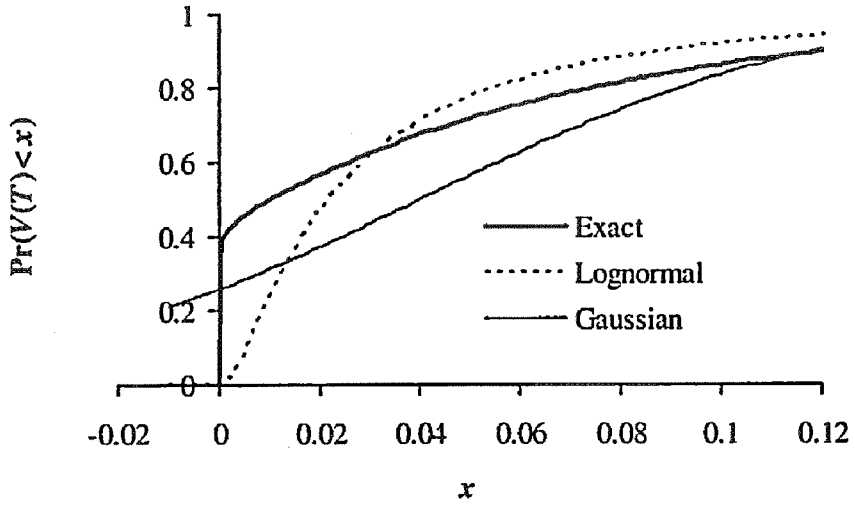
For the process (2) define:

$$d = 4\kappa\theta/\varepsilon^2; \quad n(t, T) = \frac{4\kappa\,\mathrm{e}^{-\kappa(T-t)}}{\varepsilon^2(1 - \mathrm{e}^{-\kappa(T-t)})}, \quad T > t$$

Let $T > t$. Conditional on $V(t)$, $V(T)$ is distributed as $\mathrm{e}^{-\kappa(T-t)}/n(t, T)$ times a non-central chi-squared distribution with d degrees of freedom and non-centrality parameter $V(t)n(t, T)$. That is:

$$\Pr(V(T) < x | V(t)) = F_{\chi^2}\left(\frac{x \cdot n(t, T)}{\mathrm{e}^{-\kappa(T-t)}}; d, V(t) \cdot n(t, T)\right)$$

FIGURE 1 Cumulative distribution of V.



Note: The figure shows the cumulative distribution function for $V(T)$ given $V(0)$, with $T = 0.1$. Model parameters were $V(0) = \theta = 4\%$, $\kappa = 50\%$ and $\varepsilon = 100\%$. The lognormal and Gaussian distributions in the graph were parameterized by matching mean and variances to the exact distribution of $V(T)$.

From the known properties of the non-central chi-squared distribution, the following corollary easily follows.

COROLLARY 1 *Let $T > t$. Conditional on $V(t)$, $V(T)$ has the following first two moments:*

$$\mathrm{E}(V(T)|V(t)) = \theta + (V(t) - \theta)\,\mathrm{e}^{-\kappa(T-t)}$$

$$\mathrm{Var}(V(T)|V(t)) = \frac{V(t)\varepsilon^2\,\mathrm{e}^{-\kappa(T-t)}}{\kappa}(1 - \mathrm{e}^{-\kappa(T-t)}) + \frac{\theta\varepsilon^2}{2\kappa}(1 - \mathrm{e}^{-\kappa(T-t)})^2$$

We note that the variance of $V(T)$ grows with increasing ε (volatility of variance) and decreasing κ (mean reversion speed). For reference, Appendix A lists the exact moments of $\ln X$, as well as the covariance between $\ln X$ and V.

PROPOSITION 2 *Assume that $V(0) > 0$. If $2\kappa\theta \geq \varepsilon^2$ then the process for V can never reach zero. If $2\kappa\theta < \varepsilon^2$, the origin is accessible and strongly reflecting.*

In typical applications, $2\kappa\theta$ is often significantly below ε^2, so the likelihood of hitting zero is often quite significant. Indeed, the process for V often has a strong affinity for the area around the origin, as is evident from the distribution graph in Figure 1. For comparison, we have superimposed Gaussian and lognormal distributions matched to the first two moments of V; evidently neither of these distributions are particularly good proxies for the true distribution of V.

Conditional on $X(t)$, the characteristic function for $\ln X(T)$ is known in closed form; see, for example, Heston (1993). As a consequence, Fourier-based

expressions for the price of European call and put options can be worked out. The following formulation, in the style of Lipton (2002), is convenient for computations.

PROPOSITION 3 *Consider a call option paying* $(X(T) - K)^+$ *at time* T. *The time zero (undiscounted) price of the call option is:*

$$\mathrm{E}((X(T) - K)^+) = X(0) - \frac{K}{2\pi}\int_{-\infty}^{\infty} \frac{\exp((1/2 - \mathrm{i}k)\ln(X(0)/K) + h_1 - (k^2 + 1/4)h_2)V(0)}{k^2 + 1/4}\,\mathrm{d}k$$

where i *is the complex unit, and*

$$h_1 = -\frac{\kappa\theta}{\varepsilon^2}\left(\partial_+ T + 2\ln\left(\frac{\partial_- + \partial_+ \mathrm{e}^{-\xi T}}{2\xi}\right)\right)$$

$$h_2 = -\frac{1 - \mathrm{e}^{-\xi T}}{\partial_- + \partial_+ \mathrm{e}^{-\xi T}}, \quad \hat{\kappa} = \kappa - \rho\varepsilon/2$$

$$\partial_\pm = \xi \mp (\mathrm{i}k\rho\varepsilon + \hat{\kappa}), \quad \xi = \sqrt{k^2\varepsilon^2(1 - \rho^2) + 2\mathrm{i}k\varepsilon\rho\hat{\kappa} + \hat{\kappa}^2 + \varepsilon^2/4}$$

For numerical work, it is generally useful to recognize that the process for $X(t)$ is often relatively close to geometric Brownian motion, making it sensible to work with logarithms of $X(t)$, rather than $X(t)$ itself. An application of Ito's lemma shows that (1)–(2) is equivalent to:

$$\mathrm{d}\ln X(t) = -\tfrac{1}{2}V(t)\,\mathrm{d}t + \sqrt{V(t)}\,\mathrm{d}W_X(t) \tag{3}$$

$$\mathrm{d}V(t) = \kappa(\theta - V(t))\,\mathrm{d}t + \varepsilon\sqrt{V(t)}\,\mathrm{d}W_V(t) \tag{4}$$

All of the numerical works shall center on this formulation of the model dynamics.

2.2 Path simulation

Given some arbitrary set of discrete times $\mathcal{T} = \{t_i\}_{i=1}^N$, consider now the problem of generating random paths of the pair $(X(t), V(t))$ for all $t \in \mathcal{T}$. This would be required, for instance, in the pricing of path-dependent securities with payout functions that depend on observations of $X(t)$ at a given finite set of dates. To devise such a scheme, it suffices to contemplate the more fundamental question of how, for an arbitrary increment Δ, to generate a random sample of $(X(t + \Delta), V(t + \Delta))$ given $(X(t), V(t))$; repeated application of the resulting one-period scheme (with Δ generally different at each date in $\mathcal{T}$) will produce a full path $(X(t), V(t))_{t\in\mathcal{T}}$. Below, we outline a few previously proposed techniques for updating X and V from time t to time $t + \Delta$.

2.3 Euler scheme

Using $\hat{X}$ and $\hat{V}$ to denote discrete-time approximations to X and V, respectively, a basic Euler scheme for (3)–(4) would take the form:

$$\ln \hat{X}(t+\Delta) = \ln \hat{X}(t) - \tfrac{1}{2}\hat{V}(t)\Delta + \sqrt{\hat{V}(t)}\, Z_X \sqrt{\Delta}$$

$$\hat{V}(t+\Delta) = V(t) + \kappa(\theta - \hat{V}(t))\Delta + \varepsilon\sqrt{\hat{V}(t)}\, Z_V \sqrt{\Delta} \tag{5}$$

where Z_X and Z_V are standardized Gaussian variables with correlation ρ. Note that generation of Z_X and Z_V on a computer can be done by setting:

$$Z_V = \Phi^{-1}(U_1)$$

$$Z_X = \rho Z_V + \sqrt{1-\rho^2}\,\Phi^{-1}(U_2)$$

where U_1 and U_2 are independent uniform samples, and where Φ^{-1} is the inverse cumulative Gaussian distribution function. Computation of Φ^{-1} can be accomplished, for instance, by the algorithm in Moro (1995), at relatively small computational cost.

One immediate problem with the scheme above is that the discrete process for V can become negative with non-zero probability, which in turn would make computation of $\sqrt{\hat{V}}$ impossible and cause the time stepping scheme to fail. To get around this problem, several remedies have been proposed in the literature; see Lord *et al* (2006) for a review of various "fixes". The scheme that appears to produce the smallest discretization bias can be written in the form:

$$\ln \hat{X}(t+\Delta) = \ln \hat{X}(t) - \tfrac{1}{2}\hat{V}(t)^{+}\Delta + \sqrt{\hat{V}(t)^{+}}\, Z_X \sqrt{\Delta} \tag{6}$$

$$\hat{V}(t+\Delta) = \hat{V}(t) + \kappa(\theta - \hat{V}(t)^{+})\Delta + \varepsilon\sqrt{\hat{V}(t)^{+}}\, Z_V \sqrt{\Delta} \tag{7}$$

where we use the notation $x^{+} = \max(x, 0)$. In Lord *et al* (2006) this scheme is denoted as being "full truncation"; its main characteristic is that the process for V is allowed to go below zero, at which point the process for V becomes deterministic with an upward drift of $\kappa\theta$.

2.4 Kahl–Jackel scheme

Kahl and Jackel (2005) suggest discretizing the V-process using an implicit Milstein scheme, coupled with their "IJK" discretization for the stock process.

Specifically, they propose the scheme:

$$\ln \hat{X}(t+\Delta) = \ln \hat{X}(t) - \frac{\Delta}{4}(\hat{V}(t+\Delta) + \hat{V}(t)) + \rho\sqrt{\hat{V}(t)}\, Z_V\sqrt{\Delta} + \frac{1}{2}\left(\sqrt{\hat{V}(t+\Delta)} + \sqrt{\hat{V}(t)}\right)(Z_X\sqrt{\Delta} - \rho Z_V\sqrt{\Delta}) + \frac{1}{4}\varepsilon\rho\Delta(Z_V^2 - 1) \tag{8}$$

$$\hat{V}(t+\Delta) = \frac{\hat{V}(t) + \kappa\theta\Delta + \varepsilon\sqrt{\hat{V}(t)}\, Z_V\sqrt{\Delta} + \frac{1}{4}\varepsilon^2\Delta(Z_V^2 - 1)}{1 + \kappa\Delta} \tag{9}$$

It is easy to verify that this discretization scheme will result in positive paths for the V process if $4\kappa\theta > \varepsilon^2$. As argued earlier in connection with Proposition 2, this restriction is rarely satisfied in practice, and one typically finds that the sampling scheme for V produces negative values with substantial probability. Unfortunately, Kahl and Jackel (2005) do not provide a solution for this problem, but it seems reasonable to use a truncation scheme similar to that behind (6)–(7). That is, whenever $\hat{V}$ drops below zero, we use (7), and simultaneously make sure to use $\hat{V}(t+\Delta)^+$ and $\hat{V}(t)^+$, rather than $\hat{V}(t+\Delta)$ and $\hat{V}(t)$, in (8).

2.5 Broadie–Kaya scheme

In Broadie and Kaya (2006), $V(t+\Delta)$ is sampled directly from the known distribution in Proposition 1. As direct inversion of the distribution function for $V(t+\Delta)$ is numerically expensive, an acceptance–rejection technique is used instead. Loosely, the scheme involves sampling from a Poisson distribution followed by an acceptance–rejection sample from a central chi-squared distribution with its degree-of-freedom parameter determined by the outcome of the Poisson draw. See Broadie and Kaya (2006) or Glasserman (2003) for details. With V being drawn from its exact probability distribution, the resulting sampling scheme for the V process is completely bias-free.

To obtain a bias-free scheme for sampling the asset price process, first integrate the SDE for $V(t)$, to yield:

$$V(t+\Delta) = V(t) + \int_t^{t+\Delta} \kappa(\theta - V(u))\,\mathrm{d}u + \varepsilon\int_t^{t+\Delta} \sqrt{V(u)}\,\mathrm{d}W_V(u)$$

or, equivalently:

$$\int_t^{t+\Delta} \sqrt{V(u)}\,\mathrm{d}W_V(u) = \varepsilon^{-1}\left(V(t+\Delta) - V(t) - \kappa\theta\Delta + \kappa\int_t^{t+\Delta} V(u)\,\mathrm{d}u\right) \tag{10}$$

A Cholesky decomposition shows that:

$$\mathrm{d}\ln X(t) = -\tfrac{1}{2}V(t)\,\mathrm{d}t + \rho\sqrt{V(t)}\,\mathrm{d}W_V(t) + \sqrt{1-\rho^2}\sqrt{V(t)}\,\mathrm{d}W(t)$$

where W is a Brownian motion independent of W_V. In integral form:

$$\ln X(t+\Delta) = \ln X(t) + \frac{\rho}{\varepsilon}(V(t+\Delta) - V(t) - \kappa\theta\Delta) + \left(\frac{\kappa\rho}{\varepsilon} - \frac{1}{2}\right)\int_t^{t+\Delta} V(u)\,\mathrm{d}u + \sqrt{1-\rho^2}\int_t^{t+\Delta}\sqrt{V(u)}\,\mathrm{d}W(u) \tag{11}$$

where we have used (10). Conditional on $V(t+\Delta)$ and $\int_t^{t+\Delta} V(u)\,\mathrm{d}u$, it is clear that the distribution of $\ln X(t+\Delta)$ is Gaussian with easily computable moments. After first sampling $V(t+\Delta)$ from the non-central chi-squared distribution (as described above), one then performs the following steps:

(1) conditional on $V(t+\Delta)$ (and $V(t)$) draw a sample $\int_t^{t+\Delta} V(u)\,\mathrm{d}u$;
(2) conditional on $V(t+\Delta)$ and $\int_t^{t+\Delta} V(u)\,\mathrm{d}u$, use (11) to draw a sample of $\ln X(t+\Delta)$ from a Gaussian distribution.

While execution of the second step is straightforward, the first one is not, as the necessary conditional distribution of $\int_t^{t+\Delta} V(u)\,\mathrm{d}u$ is not known in closed form. Broadie and Kaya (2006) are, however, able to derive the characteristic function, which they can numerically Fourier invert to generate the conditional cumulative distribution function for $\int_t^{t+\Delta} V(u)\,\mathrm{d}u$. Numerical inversion of this distribution function over a uniform random variable finally allows for generation of a sample of $\int_t^{t+\Delta} V(u)\,\mathrm{d}u$. The total algorithm requires great care in numerical discretization to prevent introduction of noticeable biases and is further complicated by the fact that the characteristic function for $\int_t^{t+\Delta} V(u)\,\mathrm{d}u$ contains two modified Bessel functions[2] (each of which represents an infinite series).

The Broadie–Kaya algorithm can be extended to delta and gamma computations (see Broadie and Kaya (2004)) and is bias-free by construction. However, in its pure form its complexity and lack of speed often limits practical use to benchmarking of theoretical values against which more practical schemes can be measured; see Lord *et al* (2006) for numerical cost–benefit comparison against the Euler scheme (6)–(7).[3] Also, the use of an acceptance–rejection scheme in the simulation of $V(t+\Delta)$ is inconvenient in many trading and risk management applications, as perturbation of model parameters may alter the total number of pseudo-random uniform numbers needed to generate a path of X and V. Techniques based on caches or separated random number streams can

[2] While the technique of Smith (2007) still requires the computation of Bessel functions, his slight modification of the characteristic function allows for caching of calculation results, improving overall speed.

[3] Broadie and Kaya (2006) also compare their method to an Euler scheme, but one with a sub-optimal way of handling negative values of V.

be introduced to circumvent this,[4] at the potential cost of introducing additional complexity into an already challenging simulation algorithm.

2.6 Other discretization schemes

For the special case of zero correlation, Andersen and Brotherton-Ratcliffe (2005) use an Euler scheme for $\ln X$ and suggest approximating the process for V as a lognormal variable, with moments fitted to the true moments given in Corollary 1. Unlike a standard Euler scheme in V, this scheme insures that the V process stays strictly positive. Still, we know from Figure 1 that the distribution for V is not particularly close to lognormal, and we typically find that the computational performance of the scheme in Andersen and Brotherton-Ratcliffe (2005) is comparable to that of the Euler scheme (6)–(7).

The textbook Glasserman (2003) briefly considers applications of a standard Milstein scheme (see, for example, Kloeden and Platen (1999)) on the Heston model; the results demonstrate a somewhat erratic convergence behavior for European call option pricing. The test case considered in Glasserman (2003) has quite benign parameters, as ε is only 30%, about three times lower than values typically used in practice; if one increases ε to 100%, it can be verified that the Milstein scheme essentially breaks apart. This is not surprising, given that the drift term (ie, the term that multiplies Δ) contains a factor $V(t)^{-1/2}$, which leads to poor numerical performance for the cases where there is a high likelihood of the V-process reaching zero. As also pointed out in Glasserman (2003), applications of the Milstein scheme lack theoretical support as the SDE for V fails to satisfy certain smoothness conditions. We cannot recommend application of the standard Milstein scheme as a way to discretize the V-process and shall not discuss it further here. We do, however, consider the *implicit* Milstein scheme suggested in Kahl and Jackel (2005) in the numerical tests in Section 5.

3 PROPOSED DISCRETIZATION SCHEMES FOR V

Before commencing on the description of the new V-discretization schemes we shall test in this paper, let us briefly consider a few qualitative properties of the true distribution for V (see Proposition 1). First, it is known (see Johnson *et al* (1995, p. 450)) that the non-central chi-squared distribution approaches a Gaussian distribution as the non-centrality parameter approaches ∞. From Proposition 1, we know that $V(t+\Delta)$ is proportional to a non-central chi-squared distribution with non-centrality parameter $V(t) \cdot n(t, t+\Delta)$, where n is independent of $V(t)$.

[4] As noted by Mark Broadie, one way to avoid acceptance–rejection techniques would be to combine numerical root-search with known algorithms (see, for example, Ding (1992)) to create a large three-dimensional cache of the inverse of $F_{\chi'^2}(z; \nu, \lambda)$ for all conceivable values of ν and λ. Interpolation into such a table could then generate samples of V by the standard inverse transform method. As the parameter space is potentially very large, such "brute-force" caching would have its own challenges (eg, the dimensioning of the cache and design of inter- and extrapolation rules).

In other words, for sufficiently large[5] $V(t)$ a good proxy for $V(t+\Delta)$ would be a Gaussian variable with the first two moments fitted to match those given in Corollary 1.

For small $V(t)$, on the other hand, the non-centrality parameter approaches zero, and the distribution of $V(t+\Delta)$ becomes proportional to that of an ordinary (central) chi-squared distribution with $4\kappa\theta/\varepsilon^2$ degrees of freedom. We recall that the density of a central chi-squared distribution with ν degrees of freedom is:

$$f_{\chi^2}(x;\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \mathrm{e}^{-x/2} x^{\nu/2-1} \tag{12}$$

For many cases of practical relevance, $4\kappa\theta/\varepsilon^2 \ll 2$, so the presence of the term $x^{\nu/2-1}$ in (12) implies that, for small $V(t)$, the density of $V(t+\Delta)$ will be very large around zero; see Figure 1 for visual confirmation. It should be clear that approximation of $V(t+\Delta)$ with a Gaussian variable is typically *not* accurate when $V(t)$ is close to zero.

3.1 The truncated Gaussian scheme

In this scheme the idea is to sample from a moment-matched Gaussian density where all probability mass below zero is inserted into a delta-function at the origin. For large values of $V(t)$ (where the likelihood of reaching zero is low) this scheme will automatically reproduce the asymptotic behavior of $V(t+\Delta)$ described earlier. For small $V(t)$, the resulting scheme will approximate the chi-squared density in (12) by a mass in zero combined with an upper density tail proportional to $\mathrm{e}^{-x^2/2}$. Given the near-singular behavior of (12) around the origin, this does not seem like an unreasonable approximation, as shown in Figure 2 (compare with Figure 1).

In summary, the truncated Gaussian (TG) scheme writes:

$$\hat{V}(t+\Delta) = (\mu + \sigma \cdot Z_V)^+ \tag{13}$$

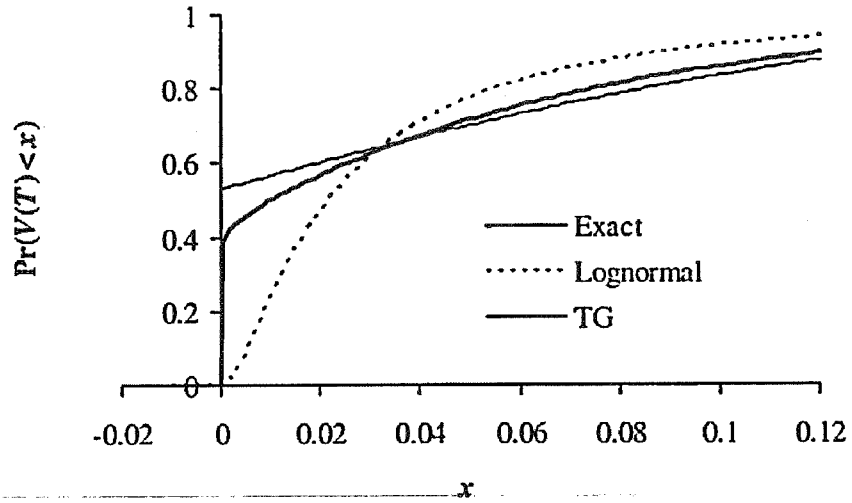
where Z_V is a standard Gaussian random variable, and μ and σ are constants that will depend on the time step Δ and $\hat{V}(t)$, as well as the parameters in the SDE for V.

3.1.1 Computing μ and σ

To set μ and σ, we will proceed to match both $\mathrm{E}(\hat{V}(t+\Delta))$ and $\mathrm{E}(\hat{V}(t+\Delta)^2)$ to the exact values of $\mathrm{E}(V(t+\Delta)|V(t)=\hat{V}(t))$ and $\mathrm{E}(V(t+\Delta)^2|V(t)=\hat{V}(t))$ as computed from Corollary 1. The result is listed in the proposition below.

[5] Note that $n(t, t+\Delta)$ goes to infinity for $\Delta \downarrow 0$, so what constitutes a large enough value of $V(t)$ for the distribution of $V(t+\Delta)$ to be well approximated by a Gaussian depends on the size of the time step, of course.

FIGURE 2 Cumulative distribution of V.



Note: The figure shows the cumulative distribution function for $V(T)$ given $V(0)$, with $T = 0.1$. Model parameters were as in Figure 1. The TG approximating distribution was computed by moment matching a truncated Gaussian distribution, as described in Section 3.1.1.

PROPOSITION 4 *Let* $\phi(x) = (2\pi)^{-1/2}\,\mathrm{e}^{-x^2/2}$ *be the standard Gaussian density, and define a function* $r : \mathbb{R} \to \mathbb{R}$ *by the relation:*

$$r(x)\phi(r(x)) + \Phi(r(x))(1 + r(x)^2) = (1 + x)(\phi(r(x)) + r(x)\Phi(r(x)))^2$$

Also set $m \equiv \mathrm{E}(V(t+\Delta)|V(t) = \hat{V}(t))$, $s^2 \equiv \mathrm{Var}(V(t+\Delta)|V(t) = \hat{V}(t))$ *and* $\psi \equiv s^2/m^2 > 0$. *If* $\hat{V}(t+\Delta)$ *is generated by the TG scheme (13), with parameter settings:*

$$\mu = \frac{m}{\phi(r(\psi))r(\psi)^{-1} + \Phi(r(\psi))}, \quad \sigma = \frac{m}{\phi(r(\psi)) + r(\psi)\Phi(r(\psi))} \tag{14}$$

then $\mathrm{E}(\hat{V}(t+\Delta)) = m$ *and* $\mathrm{Var}(\hat{V}(t+\Delta)) = s^2$.

PROOF For (13), an easy computation shows that:

$$\mathrm{E}(\hat{V}(t+\Delta)) = \int_{-\mu/\sigma}^{\infty} (\mu + \sigma x)\phi(x)\,\mathrm{d}x = \mu\Phi(\mu/\sigma) + \sigma\phi(\mu/\sigma) \tag{15}$$

$$\mathrm{E}(\hat{V}(t+\Delta)^2) = \int_{-\mu/\sigma}^{\infty} (\mu + \sigma x)^2\phi(x)\,\mathrm{d}x = \mathrm{E}(\hat{V}(t+\Delta))\mu + \sigma^2\Phi(\mu/\sigma) \tag{16}$$

Due to the non-linear form of Equations (15)–(16), the moment-matching exercise cannot be carried out analytically, so we will have to rely on numerical methods. For reasons of computation efficiency, however, a naive brute-force approach that employs a two-dimensional root-search routine at each time step in the scheme is obviously out of the question. Instead, we proceed by defining the ratio $r = \mu/\sigma$.

Matching the mean (15) to m results in:

$$\mu = \frac{m}{r^{-1}\phi(r) + \Phi(r)}, \quad \sigma = r^{-1}\mu = \frac{m}{\phi(r) + r\Phi(r)}$$

Inserting this expression into (16), along with $\sigma = r\mu$, we get, after a few rearrangements:

$$\mathrm{E}(\hat{V}(t+\Delta)^2) = m^2\left(\frac{1}{r^{-1}\phi(r) + \Phi(r)} + \frac{\Phi(r)}{(\phi(r) + r\Phi(r))^2}\right)$$

Matching (16) to $s^2 + m^2$ then yields:

$$s^2 + m^2 = m^2\left(\frac{1}{r^{-1}\phi(r) + \Phi(r)} + \frac{\Phi(r)}{(\phi(r) + r\Phi(r))^2}\right)$$

With $\psi = s^2/m^2 > 0$, this equation can be rearranged to:

$$r\phi(r) + \Phi(r)(1 + r^2) = (1 + \psi)(\phi(r) + r\Phi(r))^2$$

Clearly, then, r is only a function of ψ, ie, $r = r(\psi)$. □

Recovery of the function r must be done by numerical root-search, but the function is generic and can be mapped out once and for all, completely independent of any model or simulation setup. In practice, we would carry out this mapping on a discrete, equidistant grid for ψ (to allow for easy look-up), on a bounded domain. To determine the limits of this domain, we notice, from Corollary 1, that

$$m = \theta + (\hat{V}(t) - \theta)\,\mathrm{e}^{-\kappa\Delta} \tag{17}$$

$$s^2 = \frac{\hat{V}(t)\varepsilon^2\,\mathrm{e}^{-\kappa\Delta}}{\kappa}(1 - \mathrm{e}^{-\kappa\Delta}) + \frac{\theta\varepsilon^2}{2\kappa}(1 - \mathrm{e}^{-\kappa\Delta})^2 \tag{18}$$

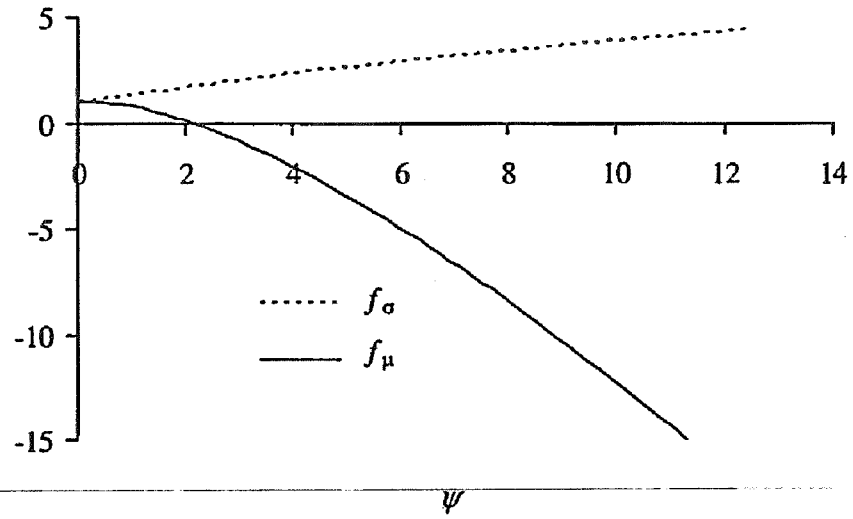
such that:

$$\psi = \frac{s^2}{m^2} = \frac{(\hat{V}(t)\varepsilon^2\,\mathrm{e}^{-\kappa\Delta}/\kappa)(1 - \mathrm{e}^{-\kappa\Delta}) + (\theta\varepsilon^2/2\kappa)(1 - \mathrm{e}^{-\kappa\Delta})^2}{(\theta + (\hat{V}(t) - \theta)\,\mathrm{e}^{-\kappa\Delta})^2} \tag{19}$$

Differentiating this expression with respect to $\hat{V}(t)$ shows that $\partial\psi/\partial\hat{V}(t) < 0$ for all $\hat{V}(t) \geq 0$, such that the largest possible value for ψ is obtained for $\hat{V}(t) = 0$ and the smallest possible value for $\hat{V}(t) = \infty$. Inserting these values for $\hat{V}(t)$ into (19) shows that $\psi \in (0, \varepsilon^2/(2\kappa\theta)]$.

In practice, there is no need to map $r(\psi)$ all the way down to $\psi = 0$; if the probability of $\hat{V}(t+\Delta)$ reaching zero, is negligible, we can skip the moment-fitting step entirely and simply set $\mu = m$ and $\sigma = s$. If we introduce a confidence multiplier α (a number around 4 or 5), we can decide to skip the fitting step when $m/s = \psi^{-1/2} > \alpha$. In practice, the relevant domain for ψ on which we,

FIGURE 3 Functions f_μ and f_σ.



as a minimum, need to map the function $r(\psi)$ is thus:

$$\psi \in [1/\alpha^2, \varepsilon^2/(2\kappa\theta)] \tag{20}$$

As a final computational trick, note that, once we have established the function r, we can write, from (14):

$$\mu = f_\mu(\psi) \cdot m, \quad f_\mu(\psi) = \frac{r(\psi)}{\phi(r(\psi)) + r(\psi)\Phi(r(\psi))} \tag{21}$$

$$\sigma = f_\sigma(\psi) \cdot s, \quad f_\sigma(\psi) = \frac{\psi^{-1/2}}{\phi(r(\psi)) + r(\psi)\Phi(r(\psi))} \tag{22}$$

The two functions $f_\mu(\psi)$ and $f_\sigma(\psi)$ are ultimately what we should cache on a computer once and for all, on an equidistant grid for ψ large enough to span the domain (20). Figure 3 shows the functions $f_\mu(\psi)$ and $f_\sigma(\psi)$. Intuitively, shifting the left tail mass of a Gaussian into a delta-function at zero will, all things equal, raise the mean and lower the variance relative to the original Gaussian distribution. To counter these effects, for large values of ψ (which correspond to small values of V) f_μ becomes significantly negative and f_σ becomes substantially larger than one. For instance, for the model parameters used in Figure 2, when $\hat{V}(t) = 0$ we get $f_\mu = -49.4$ and $f_\sigma = 6.65$. Naive truncation schemes (such as certain Euler schemes) that assume $f_\mu \approx f_\sigma \approx 1$ not surprisingly have large biases.

3.1.2 Summary of the TG algorithm

Assume now that we have proceeded to map out $f_\mu(\psi)$ and $f_\sigma(\psi)$ on a domain for ψ at least as large as (20). The detailed algorithm for the TG simulation step from $\hat{V}(t)$ to $\hat{V}(t + \Delta)$ is then as follows:

(1) given $\hat{V}(t)$, compute[6] m and s^2 from Equations (17) and (18);
(2) compute $\psi = s^2/m^2$ and look up $f_\mu(\psi)$ and $f_\sigma(\psi)$ from the cache;
(3) compute μ and σ according to Equations (21) and (22);
(4) draw a uniform random number U_V;
(5) compute $Z_V = \Phi^{-1}(U_V)$, eg, using the algorithm in Moro (1995); and
(6) use (13), ie, set $\hat{V}(t+\Delta) = (\mu + \sigma Z_V)^+$.

If implemented intelligently, apart from the computation of Z_V, a step in the TG scheme should only involve a handful of simple algebraic operations (+, −, /, and *) and should have speed comparable to the Euler step (6)–(7).

3.2 The quadratic-exponential scheme

The TG scheme models the upper tail of the density for $\hat{V}(t+\Delta)$ as being proportional to $e^{-x^2/2}$. For low values of $\hat{V}(t)$, however, this density decay is too fast, as is obvious from (12). We now introduce a scheme that is designed to address this issue; as an added benefit, the resulting scheme will not require the same amount of precaching as was necessary for the TG scheme.

We derive our new scheme in steps. The first step is based on an observation that a non-central chi-squared with moderate or high non-centrality parameter can be well represented by a power function applied to a Gaussian variable. See Patnaik (1949), Pearson (1959) and Piterbarg (2003), as well as the survey in Kahl and Jackel (2005). While there is evidence that a cubic transformation of a Gaussian variable is preferable, such a scheme could not preserve non-negative values for the V process and we abandon it in favor of a quadratic representation, along the lines of Patnaik (1949). Specifically, for sufficiently large[7] values of $\hat{V}(t)$, we write:

$$\hat{V}(t+\Delta) = a(b + Z_V)^2 \tag{23}$$

where Z_V is a standard Gaussian random variable, and a and b are certain constants, to be determined by moment matching. These constants will depend on the time step Δ and $\hat{V}(t)$, as well as the parameters in the SDE for V.

The scheme (23) does not work well for low values of $\hat{V}(t)$ – in fact the moment-matching exercise mail fail entirely – so we supplement it with a scheme to be used for low values of $\hat{V}(t)$. For this, we take inspiration from the asymptotic density in (12) and use an approximated density for $\hat{V}(t+\Delta)$ of the form:

$$\Pr(\hat{V}(t+\Delta) \in [x, x+\mathrm{d}x]) \approx (p\delta(0) + \beta(1-p)\,e^{-\beta x})\,\mathrm{d}x, \quad x \geq 0 \tag{24}$$

[6]Notice that these computations involve an exponential $\exp(-\kappa\Delta)$. Needless to say, this exponential, which only depends on the size of the time step, should be computed outside the Monte Carlo loop. For extra efficiency, we could go further and consider writing $\psi = (\hat{V}(t)k_1 + k_2)^{-2}(\hat{V}(t)k_3 + k_4)$, where the four Δ-dependent constants $k_1, \ldots, k_4$ can be computed before the Monte Carlo simulation starts. To the extent that the time grid is non-equidistant, and/or the parameters in the SDE for V are functions of time, we will need to cache such data for each time step. The computational overhead to do this is trivial, of course.

[7]Recall that the non-centrality parameter in the exact distribution for $V(t+\Delta)$ is proportional to $V(t)$. We shall shortly make it precise what we mean with "sufficiently large".

where δ is a Dirac delta function, and p and β are non-negative constants to be determined. As in the TG scheme, we have a probability mass at the origin, but now the strength of this mass (p) is explicitly specified, rather than implied from other parameters. The mass at the origin is supplemented with an exponential tail, qualitatively similar to that of the density (12). It can be verified that if $p \in [0, 1]$ and $\beta \geq 0$, then (24) constitutes a valid density function. Figure 4 in Section 3.2.3 below demonstrates the quality of the approximations (23) and (24); generally the quadratic-exponential (QE) approximations are more accurate than the TG approximations.

Sampling according to (24) is straightforward and efficient. To see this, first we integrate (24) to generate a cumulative distribution function:

$$\Psi(x) = \Pr(\hat{V}(t+\Delta) \leq x) = p + (1-p)(1 - \mathrm{e}^{-\beta x}), \quad x \geq 0$$

We notice that the inverse of Ψ is readily computable:

$$\Psi^{-1}(u) = \Psi^{-1}(u; p, \beta) = \begin{cases} 0, & 0 \leq u \leq p \\ \beta^{-1} \ln\left(\dfrac{1-p}{1-u}\right), & p < u \leq 1 \end{cases} \tag{25}$$

By the standard inverse distribution function method, we thus get the simple sampling scheme:

$$\hat{V}(t+\Delta) = \Psi^{-1}(U_V; p, \beta) \tag{26}$$

where U_V is a draw from a uniform distribution. Note that this scheme is extremely fast to execute.

Equations (23) and (26) together define our QE discretization scheme. What remains is the determination of the constants a, b, p and β, as well as a rule for when to switch from (23) to (26).

3.2.1 Computing a and b

Our strategy is again to determine a and b by moment-matching techniques.

PROPOSITION 5 *Let m and s be as defined in Proposition 4 (Equations (17) and (18)), and set $\psi = s^2/m^2$. Provided that $\psi \leq 2$, set:*

$$b^2 = 2\psi^{-1} - 1 + \sqrt{2\psi^{-1}}\sqrt{2\psi^{-1} - 1} \geq 0 \tag{27}$$

and:

$$a = \frac{m}{1+b^2} \tag{28}$$

Let $\hat{V}(t+\Delta)$ be as defined in (23); then:

$$\mathrm{E}(\hat{V}(t+\Delta)) = m \quad \textit{and} \quad \mathrm{Var}(\hat{V}(t+\Delta)) = s^2.$$

PROOF We first recognize that (23) describes $\hat{V}(t+\Delta)$ as being distributed as a times a non-central chi-squared distribution with one degree of freedom and non-centrality parameter b^2 (see, for example, Johnson *et al* (1995)). From known results, it follows that:

$$\mathrm{E}(\hat{V}(t+\Delta)) = a(1+b^2), \quad \mathrm{Var}(\hat{V}(t+\Delta)) = 2a^2(1+2b^2)$$

Equating these moments to the exact values m and s^2 gives the equation system:

$$a(1+b^2) = m, \quad 2a^2(1+2b^2) = s^2$$

Set $x = b^2$ and $\psi = s^2/m^2$. Elimination of a yields:

$$x^2 + 2x(1-2\psi^{-1}) + 1 - 2\psi^{-1} = 0$$

Evaluation of the discriminant of this second-order equation shows that a solution is only possible if $\psi \leq 2$. Under this constraint, the solution for $x = b^2$ is (27). □

We emphasize that the values of a and b in Proposition 5 only apply for the case where $\psi \leq 2$. For higher values of ψ (corresponding to low values of $\hat{V}(t)$), the scheme will fail.

3.2.2 Computing p and β

PROPOSITION 6 *Let m, s and ψ be as defined in Proposition 4. Assume that $\psi \geq 1$ and set:*

$$p = \frac{\psi - 1}{\psi + 1} \in [0, 1) \tag{29}$$

and:

$$\beta = \frac{1-p}{m} = \frac{2}{m(\psi+1)} > 0 \tag{30}$$

Let $\hat{V}(t+\Delta)$ be as defined in (26); then:

$$\mathrm{E}(\hat{V}(t+\Delta)) = m \quad \textit{and} \quad \mathrm{Var}(\hat{V}(t+\Delta)) = s^2$$

PROOF By direct integration of the density (24), it is easy to show that:

$$\mathrm{E}(\hat{V}(t+\Delta)) = \frac{1-p}{\beta}, \quad \mathrm{Var}(\hat{V}(t+\Delta)) = \frac{1-p^2}{\beta^2}$$

Enforcing the moment-matching conditions results in the equation system:

$$\frac{1-p}{\beta} = m, \quad \frac{1-p^2}{\beta^2} = s^2$$

Elimination of β yields:

$$(1+\psi)p^2 - 2\psi p + \psi - 1 = 0$$

where $\psi = s^2/m^2$. This system will always have exactly one solution for p less than 1, namely that in (29). Equation (30) then immediately follows. We stress that for the solution (29)–(30) to make sense, we need p to be non-negative. That is, we must demand that $\psi \geq 1$. □

3.2.3 Switching rule

With $\psi = \mathrm{Var}(V(t+\Delta)|V(t) = \hat{V}(t)) \times \mathrm{E}(V(t+\Delta)|V(t) = \hat{V}(t))^{-2}$, we have shown that the quadratic sampling scheme (23) can only be moment matched for $\psi \leq 2$. On the other hand, the exponential scheme (26) can only be moment matched for $\psi \geq 1$. Fortunately, these domains of applicability overlap, such that at least one of the two schemes can always be used. A natural procedure is to introduce some critical level $\psi_c \in [1, 2]$, and use (23) if $\psi \leq \psi_c$ and (26) otherwise. The exact choice for ψ_c appears to have relatively small effects on the quality of the overall simulation scheme; we use $\psi_c = 1.5$ in our numerical tests. We note from (19) that, for any fixed value of $\hat{V}(t)$, $\lim_{\Delta \downarrow 0} \psi = 0$, so as the time step is reduced the need to use (26) becomes increasingly remote. For practical-sized values of the time steps, however, the switching likelihood is often very substantial.

At this point, it may be worth considering whether one could dispense of the switching rule by, say, relaxing the requirements that both the first and second moments of the V-process be matched exactly. Piterbarg (2003), for instance, uses a quadratic scheme similar to (23) but only fits the first moment when $\psi > 2$ (an event that in many practically relevant model settings will have significant probability). There are, however, no speed benefits to such a scheme and, as one would intuitively expect, numerical tests generally show a marked deterioration in numerical performance relative to our full switching scheme.

To illustrate the quality of the QE approximation to the true distribution of V, we consider two different cases, one on each side of the trigger condition for ψ. See Figure 4 for the results.

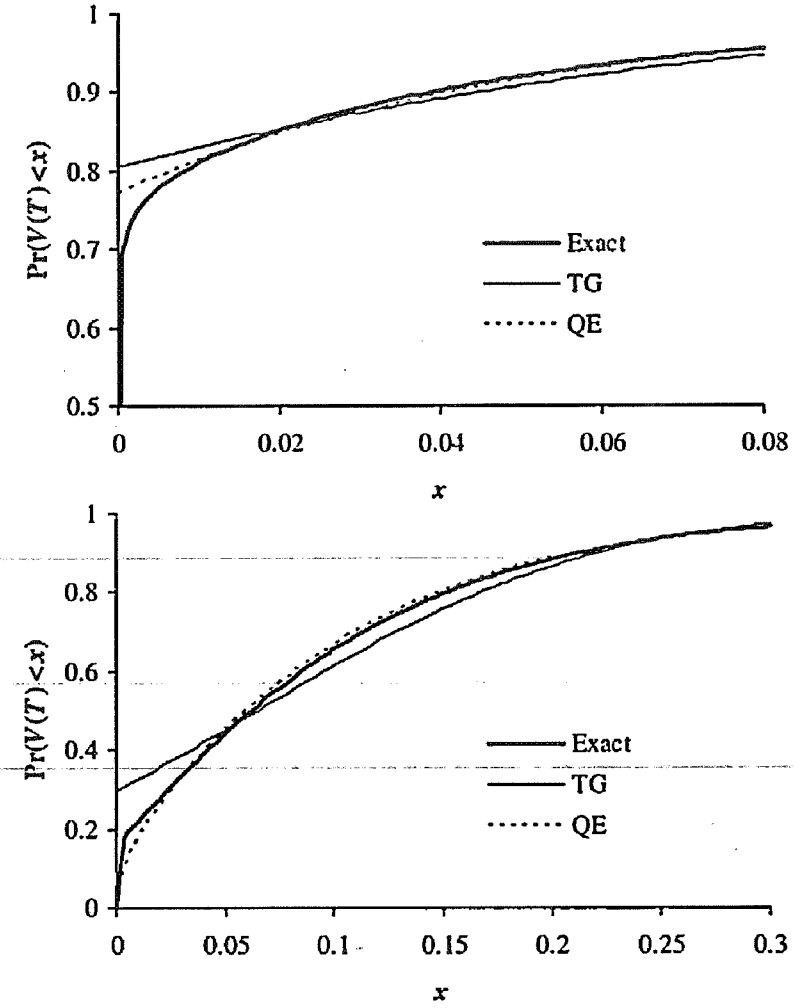
3.2.4 Summary of QE algorithm

Assume that some arbitrary level $\psi_c \in [1, 2]$ has been selected. The detailed algorithm for the QE simulation step from $\hat{V}(t)$ to $\hat{V}(t+\Delta)$ is then:

(1) given $\hat{V}(t)$, compute m and s^2 from Equations (17) and (18);
(2) compute $\psi = s^2/m^2$;
(3) draw a uniform random number U_V;
(4) **If** $\psi \leq \psi_c$:
   (a) compute a and b from Equations (28) and (27),
   (b) compute $Z_V = \Phi^{-1}(U_V)$, eg, using the algorithm in Moro (1995),
   (c) use (23), ie, set $\hat{V}(t+\Delta) = a(b + Z_V)^2$; and
(5) **Otherwise**, if $\psi > \psi_c$:
   (a) compute β and p according to Equations (30) and (29),
   (b) use (26), ie, set $\hat{V}(t+\Delta) = \Psi^{-1}(U_V; p, \beta)$, where Ψ^{-1} is given in (25).

As before, exponentials used in computation of m and s^2 should be precached; see Footnote 6.

FIGURE 4 Cumulative distribution of V.



Note: The figures show the cumulative distribution function for $V(T)$ given $V(0)$, with $T = 0.1$. Model parameters are as in Figure 1, but $V(0)$ has been lowered to $V(0) = 1\%$ in the upper figure and raised to $V(0) = 9\%$ in the lower figure.

3.2.5 Extensions

Schemes TG and QE both capture the near-singular behavior of V around the origin by inserting a Dirac mass at $V = 0$. The real V density, however, does not have such a mass and one wonders whether a more careful characterization of the behavior at $V = 0$ may be possible. Inspection of the limiting chi-squared density (12) shows that, for example, one could consider replacing in (24) the Dirac mass in zero with a term of the form constant $\cdot\, x^q$, for some constant q between -1 and 0. This idea indeed leads to a tractable sampling scheme, of potential use when a very accurate approximation for small V is required; see Appendix B for details. For most practical applications, scheme QE as listed above is accurate enough, so we do not pursue further extensions in the main paper.

4 PROPOSED DISCRETIZATION SCHEMES FOR X

We start our discussion about the discretization of the X process by considering a scheme that does *not* work well. The rationale for the failure of this scheme, however, is quite illuminating and will guide us to a better scheme, proposed in Section 4.2.

4.1 How *not* to discretize the X-process

For concreteness, assume first that we have chosen to use the TG scheme in Section 3.1 as our method of choice for the generation of random paths for the variance process V. That is, advancement of V on the time interval $[t, t+\Delta]$ takes the form:

$$\hat{V}(t+\Delta) = (\mu + \sigma Z_V)^+$$

where μ and σ are certain moment-matched constants and Z_V is a Gaussian random variable. Suppose that we combine this scheme with an Euler scheme in $\ln X$ (as in (7), but with no need to truncate V at zero):

$$\ln \hat{X}(t+\Delta) = \ln \hat{X}(t) - \tfrac{1}{2}\hat{V}(t)\Delta + \sqrt{\hat{V}(t)}\, Z_X \sqrt{\Delta} \tag{31}$$

where Z_X is another Gaussian random variable. It is quite tempting to set the correlation between Z_X and Z_V equal to ρ – that is, the correlation between the driving Brownian motions in the SDE (3)–(4) – but is this, in fact, reasonable? To analyze this, we first notice that the strongly non-linear relationship between $\hat{V}(t+\Delta)$ and Z_V will imply that the effective correlation between $\ln \hat{X}(t+\Delta)$ and $\hat{V}(t+\Delta)$ will be closer to zero than ρ for the cases where $\Pr(\mu + \sigma Z_V < 0)$ is significant, as it would be if $\hat{V}(t)$ was close to zero. In reality, however, it can be verified from the results in Appendix A that the true correlation between $\ln X(t+\Delta)$ and $V(t+\Delta)$ (conditioned on $V(t)$ and $X(t)$) will always be close to ρ, even for large values of Δ and when $V(t)$ is close to the origin.

If one were to nevertheless ignore the problem of "leaking correlation" and insist on using (31), at practical levels of Δ one would experience a strong tendency for the Monte Carlo simulation to generate too feeble effective correlation and, consequently, paths of X with poor distribution tails. In call option pricing terms, this would manifest itself in an overall poor ability to price options with strikes away from at-the-money.

4.2 Discretization scheme for X

In light of the problems highlighted above, we abandon naive Euler discretization for $\ln X$ and instead turn our focus on the exact representation (11):

$$\ln X(t+\Delta) = \ln X(t) + \frac{\rho}{\varepsilon}(V(t+\Delta) - V(t) - \kappa\theta\Delta) + \left(\frac{\kappa\rho}{\varepsilon} - \frac{1}{2}\right)\int_t^{t+\Delta} V(u)\,\mathrm{d}u + \sqrt{1-\rho^2}\int_t^{t+\Delta}\sqrt{V(u)}\,\mathrm{d}W(u)$$

In the expression for $\ln X(t+\Delta)$ the term $(\rho/\varepsilon)V(t+\Delta)$ is the key driver of the correlation between $X(t+\Delta)$ and $V(t+\Delta)$; any discretization scheme for $\ln X$ should attempt to keep this term.

To use (11) in the discretization of $\ln X$, we need to consider how to handle the time integral of V. Rather than using Fourier methods, here we simply write:

$$\int_t^{t+\Delta} V(u)\,\mathrm{d}u \approx \Delta[\gamma_1 V(t) + \gamma_2 V(t+\Delta)] \tag{32}$$

for certain constants γ_1 and γ_2. There are multiple ways for setting γ_1 and γ_2, the simplest being the Euler-like setting $\gamma_1 = 1$, $\gamma_2 = 0$. A central discretization, on the other hand, would set $\gamma_1 = \gamma_2 = \frac{1}{2}$. A more sophisticated (but slower) approach could be based on moment matching; the interested reader can find the exact moments for $\int_t^{t+\Delta} V(u)\,\mathrm{d}u$ in Dufresne (2001, p. 16).

As W is independent of V, conditional on $V(t)$ and $\int_t^{t+\Delta} V(u)\,\mathrm{d}u$, the Ito integral:

$$\int_t^{t+\Delta} \sqrt{V(u)}\,\mathrm{d}W(u)$$

is Gaussian with mean zero and variance $\int_t^{t+\Delta} V(u)\,\mathrm{d}u$. With our approximation (32), this leads us to propose the following natural discretization scheme:

$$\begin{aligned}
\ln \hat{X}(t+\Delta) &= \ln \hat{X}(t) + \frac{\rho}{\varepsilon}(\hat{V}(t+\Delta) - \hat{V}(t) - \kappa\theta\Delta) \\
&\quad + \Delta\left(\frac{\kappa\rho}{\varepsilon} - \frac{1}{2}\right)(\gamma_1 \hat{V}(t) + \gamma_2 \hat{V}(t+\Delta)) \\
&\quad + \sqrt{\Delta}\sqrt{1-\rho^2}\sqrt{\gamma_1 \hat{V}(t) + \gamma_2 \hat{V}(t+\Delta)} \cdot Z \\
&= \ln \hat{X}(t) + K_0 + K_1 \hat{V}(t) + K_2 \hat{V}(t+\Delta) \\
&\quad + \sqrt{K_3 \hat{V}(t) + K_4 \hat{V}(t+\Delta)} \cdot Z
\end{aligned} \tag{33}$$

where Z is a standard Gaussian random variable, independent of $\hat{V}$, and $K_0, \ldots, K_4$ are given by:

$$K_0 = -\frac{\rho\kappa\theta}{\varepsilon}\Delta, \quad K_1 = \gamma_1\Delta\left(\frac{\kappa\rho}{\varepsilon} - \frac{1}{2}\right) - \frac{\rho}{\varepsilon}$$

$$K_2 = \gamma_2\Delta\left(\frac{\kappa\rho}{\varepsilon} - \frac{1}{2}\right) + \frac{\rho}{\varepsilon}, \quad K_3 = \gamma_1\Delta(1-\rho^2), \quad K_4 = \gamma_2\Delta(1-\rho^2)$$

Note that K_i, $i = 0, \ldots, 4$, depend on the time step as well as on the constants γ_1 and γ_2.

For given values of γ_1 and γ_2, the scheme constitutes our proposed discretization scheme for $\ln X$. It is to be combined with a simulation scheme for V, in the following fashion:

(1) given $\hat{V}(t)$, generate $\hat{V}(t+\Delta)$ using one of the time stepping schemes in Section 3;
(2) draw a uniform random number U, independent of all random numbers used for $\hat{V}(t+\Delta)$;
(3) set $Z = \Phi^{-1}(U)$, eg, using the algorithm in Moro (1995); and
(4) given $\ln \hat{X}(t)\hat{V}(t)$, and the value for $\hat{V}(t+\Delta)$ computed in Step (1), compute $\ln \hat{X}(t+\Delta)$ from (33).

We note that Step (1) can, if desired, be replaced by an acceptance–rejection algorithm (or perhaps the cache methodology mentioned in Footnote 4).[8]

4.3 Martingale correction: regularity

The scheme (33) is equivalent to:

$$\hat{X}(t+\Delta) = \hat{X}(t)\exp(K_0 + K_1\hat{V}(t)) \times \exp(K_2\hat{V}(t+\Delta) + \sqrt{K_3\hat{V}(t) + K_4\hat{V}(t+\Delta)}\cdot Z) \tag{34}$$

As discussed in Andersen and Piterbarg (2005), the continuous-time process for X may not have finite higher moments, but will always be a martingale. That is:

$$\mathrm{E}(X(t+\Delta)|X(t)) = X(t) < \infty$$

In contrast, (33) will not satisfy an equivalent discrete-time martingale condition,[9] ie:

$$\mathrm{E}(\hat{X}(t+\Delta)|\hat{X}(t)) \neq \hat{X}(t)$$

The practical relevance of this is often minor, as the net drift away from the martingale is typically very small and controllable by reduction of the time step. Also, the ability to hit the mean of the distribution for X does not necessarily translate itself into better prices for options. Nevertheless, in the spirit of Glasserman and Zhao (1999), let us examine whether it is possible to modify our sampling scheme for X to strictly enforce $\mathrm{E}(\hat{X}(t+\Delta)|\hat{X}(t)) = \hat{X}(t)$. As part of this, we will also examine whether there might be parameter settings where the X process blows up, in the sense that $\mathrm{E}(\hat{X}(t+\Delta)|\hat{X}(t)) = \infty$.

PROPOSITION 7 *Let K_i, $i = 1, \ldots, 4$, be as defined in Equation (33). Define:*

$$M = \mathrm{E}(\mathrm{e}^{A\hat{V}(t+\Delta)}|\hat{V}(t)) > 0, \quad A = K_2 + \frac{1}{2}K_4 = \frac{\rho}{\varepsilon}(1+\kappa\gamma_2\Delta) - \frac{1}{2}\gamma_2\Delta\rho^2$$

[8] Attempts at numerical comparison of such a scheme with the QE scheme turned out somewhat inconclusive, as a very large number of Monte Carlo simulations was required to detect any noticeable differences in bias between the two schemes. If V is sampled exactly from a chi-squared distribution, analytical martingale correction (see Section 4.3) of the process for X is still possible. The required expressions are available from the author upon request.

[9] The Euler scheme (6)–(7) satisfies $\mathrm{E}(\hat{X}(t+\Delta)|\hat{X}(t)) = \hat{X}(t)$, whereas the Kahl–Jackel scheme (8)–(9) does not.

If $M < \infty$, then $\mathrm{E}(\hat{X}(t+\Delta)|\hat{X}(t)) < \infty$. Assuming that M is finite, set:

$$K_0^* = -\ln M - (K_1 + \tfrac{1}{2}K_3)\hat{V}(t) \tag{35}$$

and:

$$\ln \hat{X}(t+\Delta) = \ln \hat{X}(t) + K_0^* + K_1\hat{V}(t) + K_2\hat{V}(t+\Delta) + \sqrt{K_3\hat{V}(t) + K_4\hat{V}(t+\Delta)} \cdot Z \tag{36}$$

where Z is a standard Gaussian random variable. In this case,

$$\mathrm{E}(\hat{X}(t+\Delta)|\hat{X}(t)) = \hat{X}(t)$$

PROOF By iterated conditional expectations, from (36) we note that (suppressing the implicit conditioning on $\hat{V}(t)$):

$$\begin{aligned}
&\mathrm{E}(\hat{X}(t+\Delta)|\hat{X}(t)) \\
&\quad = \mathrm{E}(\mathrm{E}(\hat{X}(t+\Delta)|\hat{X}(t), \hat{V}(t+\Delta))) \\
&\quad = \hat{X}(t)\,\mathrm{e}^{K_0^* + K_1\hat{V}(t)}\mathrm{E}\left(\mathrm{e}^{K_2\hat{V}(t+\Delta)}\mathrm{E}\left(\mathrm{e}^{\sqrt{K_3\hat{V}(t)+K_4\hat{V}(t+\Delta)}\cdot Z}|\hat{X}(t), \hat{V}(t+\Delta)\right)\right) \\
&\quad = \hat{X}(t)\,\mathrm{e}^{K_0^* + (K_1+\frac{1}{2}K_3)\hat{V}(t)}\mathrm{E}(\mathrm{e}^{A\hat{V}(t+\Delta)})
\end{aligned} \tag{37}$$

where the third step uses a known result for lognormal distributions and where we have defined $A = K_2 + \frac{1}{2}K_4$. For $\mathrm{E}(\hat{X}(t+\Delta)|\hat{X}(t))$ to equal $\hat{X}(t)$, we evidently require:

$$\mathrm{e}^{K_0^* + (K_1+\frac{1}{2}K_3)\hat{V}(t)}M = 1$$

which is (35). □

To summarize, the martingale corrected scheme in Proposition 7 involves substituting K_0^* for K_0 in the basic scheme (33). As stated in the proposition, for this to be possible – and indeed for (33) to be meaningful in the first place – we require that $M = \mathrm{E}(\exp(A\hat{V}(t+\Delta)|\hat{V}(t)))$ be finite. Assuming that $\gamma_2 \geq 0$ (which would always be the case in practice) and $\hat{V}(t+\Delta) \geq 0$ (which is always the case for the schemes in Section 3), it can be verified that $A \leq 0$ for $\rho \leq 0$, which in turn shows that:

$$\rho \leq 0 \Rightarrow M < \infty \tag{38}$$

This is an obvious consequence of the fact that $\mathrm{e}^{A\hat{V}(t+\Delta)}$ here is bounded to the interval [0, 1].

In practical applications, we virtually always have $\rho \leq 0$, in which case (33) is safe to use. Positive correlations may occur, of course, in which case we will need to examine the discretization scheme for V in more detail. We proceed to do so below.

4.3.1 The TG scheme

PROPOSITION 8 *Let $\hat{V}(t+\Delta) = (\mu + \sigma Z_V)^+$ (the TG scheme). Then, for any value of A, we have:*

$$\mathrm{E}(\mathrm{e}^{A\hat{V}(t+\Delta)}|\hat{V}(t)) = \mathrm{e}^{A\mu+\frac{1}{2}A^2\sigma^2}\Phi(d_+) + \Phi(-d_-) \tag{39}$$

with:

$$d_+ = \frac{\mu}{\sigma} + A\sigma, \quad d_- = \frac{\mu}{\sigma}$$

PROOF Here $\hat{V}(t+\Delta) = (\mu + \sigma Z_V)^+$ where μ and σ depend on $\hat{V}(t)$. For $A \geq 0$, we have:

$$\begin{aligned}\mathrm{E}(\mathrm{e}^{A\hat{V}(t+\Delta)}|\hat{V}(t)) &= \mathrm{E}(\max(\exp(A\mu + A\sigma Z_V), 1)) \\ &= 1 + \mathrm{E}(\max(\exp(A\mu + A\sigma Z_V) - 1, 0)), \quad A \geq 0\end{aligned}$$

For $A < 0$, on the other hand, the same arguments lead to:

$$\mathrm{E}(\mathrm{e}^{A\hat{V}(t+\Delta)}|\hat{V}(t)) = 1 - \mathrm{E}(\max(1 - \exp(A\mu + A\sigma Z_V), 0))$$

Irrespective of the sign for A, the variable $\exp(A\mu + A\sigma Z_V)$ is lognormal (as Z_V is Gaussian), so standard results from Black–Scholes analysis can be used to compute the above option-like expectation. The final result, which is always finite irrespective of the value of A, is (39). □

While the result for $\mathrm{E}(\mathrm{e}^{A\hat{V}(t+\Delta)}|\hat{V}(t))$ is available in closed form, it is rather complicated and not particularly efficient to compute inside a discretization loop, as required in martingale correction by (35). Caching techniques can help, of course, but they become cumbersome for general (non-equidistant) time grids.

4.3.2 The QE scheme

PROPOSITION 9 *Let the QE scheme be given as in Section 3.2, and characterized by constants a, b, β and p computed from Propositions 5 and 6. Let $\psi = s^2/m^2$, with m and s given in (17) and (18). Also, let $\psi_c \in [1, 2]$ be given. If $\psi \leq \psi_c$, then:*

$$\mathrm{E}(\mathrm{e}^{A\hat{V}(t+\Delta)}|\hat{V}(t)) = \frac{\exp(Ab^2a/(1-2Aa))}{\sqrt{1-2Aa}} \tag{40}$$

where A must satisfy:

$$A < \frac{1}{2a} \tag{41}$$

If, on the other hand, $\psi > \psi_c$, then:

$$\mathrm{E}(\mathrm{e}^{A\hat{V}(t+\Delta)}|\hat{V}(t)) = p + \frac{\beta(1-p)}{\beta - A} \tag{42}$$

provided that:

$$A < \beta \tag{43}$$

PROOF For $\psi \le \psi_c$, we recall that the QE scheme sets $\hat{V}(t+\Delta) = a(b+Z_V)^2$, the distribution of which is a times a non-central chi-squared distribution with one degree of freedom and non-centrality parameter b^2 (see Section 3.2.1). From known results for the non-central chi-squared distribution (specifically, its moment-generating function), we get the result (40). For this expectation to exist, we must demand that $aA < 1/2$.

For $\psi > \psi_c$, we have:

$$\mathrm{E}(\mathrm{e}^{A\hat{V}(t+\Delta)}|\hat{V}(t)) = p + \beta(1-p)\int_0^\infty \mathrm{e}^{[A-\beta]u}\,\mathrm{d}u = p + \frac{\beta(1-p)}{\beta - A}$$

provided that $A < \beta$ (otherwise the expectation does not exist). □

We emphasize that for the QE scheme the expectation $\mathrm{E}(\mathrm{e}^{A\hat{V}(t+\Delta)}|\hat{V}(t))$ does not exist for all values of A. Of the two regularity conditions (41) and (43), the first is, in practice, the most restrictive.[10] To get a feeling for how restrictive (41) is, we recall that:

$$A = \frac{\rho}{\varepsilon}(1+\kappa\gamma_2\Delta) - \frac{1}{2}\gamma_2\Delta\rho^2$$

We also notice that the definition of a in (28) guarantees that $a \ge 4\kappa^{-1}\varepsilon^2(1-\mathrm{e}^{-\kappa\Delta})$ always, so (41) will certainly be satisfied if:

$$\frac{\rho}{\varepsilon}(1+\kappa\gamma_2\Delta) - \frac{1}{2}\gamma_2\Delta\rho^2 < \frac{2\kappa}{\varepsilon^2(1-\mathrm{e}^{-\kappa\Delta})}$$

If $\rho > 0$, this imposes a limit on the size of the time step, roughly $\rho\varepsilon\Delta < 2$. As ε is normally around 50–150%, the resulting restriction on the time step is not likely to be a practical problem, even for quite large positive values of ρ.

Application of the result of Proposition 7 to enforce the martingale condition in the discretization of X here is straightforward and convenient. Specifically, in (36) we simply set:

$$K_0^* = \begin{cases} -\dfrac{Ab^2a}{1-2Aa} + \dfrac{1}{2}\ln(1-2Aa) - \left(K_1 + \dfrac{1}{2}K_3\right)\hat{V}(t), & \psi \le \psi_c \\ -\ln\left(p + \dfrac{\beta(1-p)}{\beta - A}\right) - \left(K_1 + \dfrac{1}{2}K_3\right)\hat{V}(t), & \psi > \psi_c \end{cases}$$

4.4 Convergence considerations

A formal analysis of the convergence properties for the schemes proposed in this paper is difficult and complicated by the fact that the X process may not have any high-order moments. As such, the usual examination of (weak) convergence of expectations of polynomials of X is not always meaningful. While we could, in principle, undertake an examination of the convergence of expectations on

[10] Recall that $\beta = 2/(m(\psi+1)) \approx m^{-1}$, where m is typically a small number for $\psi > \psi_c$.

selected slow-growing payouts of X (eg, call options), the technicalities of such an analysis are considerable and we skip it. (See Lord *et al* (2006) for examples of this type of analysis.) Instead, we focus on a simpler concept, namely that of *weak consistency*. As shown in Kloeden and Platen (1999, p. 328) there is a strong link between weak consistency and weak convergence.

PROPOSITION 10 *Assume that $\gamma_1 + \gamma_2$ in (33) approaches 1 for $\Delta \to 0$. The TG and QE schemes are then both weakly consistent. That is, conditional on $\hat{X}(t)$ and $\hat{V}(t)$, we have for both schemes:*

$$\begin{aligned} \lim_{\Delta\to 0} \mathrm{E}\left(\frac{\ln \hat{X}(t+\Delta) - \ln \hat{X}(t)}{\Delta}\right) &= -\frac{1}{2}\hat{V}(t) \\ \lim_{\Delta\to 0} \mathrm{Var}\left(\frac{\ln \hat{X}(t+\Delta) - \ln \hat{X}(t)}{\sqrt{\Delta}}\right) &= \hat{V}(t) \end{aligned} \tag{44}$$

$$\begin{aligned} \lim_{\Delta\to 0} \mathrm{E}\left(\frac{\hat{V}(t+\Delta) - \hat{V}(t)}{\Delta}\right) &= \kappa(\theta - \hat{V}(t)) \\ \lim_{\Delta\to 0} \mathrm{Var}\left(\frac{\hat{V}(t+\Delta) - \hat{V}(t)}{\sqrt{\Delta}}\right) &= \varepsilon^2 \hat{V}(t) \end{aligned} \tag{45}$$

$$\lim_{\Delta\to 0} \mathrm{Cov}\left(\frac{\hat{V}(t+\Delta) - \hat{V}(t)}{\sqrt{\Delta}}, \frac{\ln \hat{X}(t+\Delta) - \ln \hat{X}(t)}{\sqrt{\Delta}}\right) = \rho\varepsilon\hat{V}(t) \tag{46}$$

PROOF Both conditions in (45) are clearly satisfied, as the TG and QE schemes are based on exact matches of the first two conditional moments of $\hat{V}(t+\Delta)$. From (33) we also have (suppressing conditioning on $\hat{X}(t)$ and $\hat{V}(t)$):

$$\begin{aligned} &\frac{\mathrm{E}(\ln \hat{X}(t+\Delta)) - \ln \hat{X}(t)}{\Delta} \\ &\quad = \frac{(\rho/\varepsilon)(\mathrm{E}(\hat{V}(t+\Delta)) - \hat{V}(t) - \kappa\theta\Delta)}{\Delta} \\ &\qquad + \left(\frac{\kappa\rho}{\varepsilon} - \frac{1}{2}\right)(\gamma_1 \hat{V}(t) + \gamma_2 \mathrm{E}(\hat{V}(t+\Delta))) \\ &\quad \to \frac{\rho}{\varepsilon}(\kappa(\theta - \hat{V}(t)) - \kappa\theta) + \left(\frac{\kappa\rho}{\varepsilon} - \frac{1}{2}\right)\hat{V}(t) = -\frac{1}{2}\hat{V}(t) \end{aligned}$$

The second part of (44) is proved the same way. Equation (46) follows from the observation that the form of (33) implies that:

$$\begin{aligned} \mathrm{Cov}(\hat{V}(t+\Delta), \ln \hat{X}(t+\Delta)) &= \mathrm{Cov}\left(\hat{V}(t+\Delta), \frac{\rho}{\varepsilon}\hat{V}(t+\Delta)\right) \\ &= \frac{\rho}{\varepsilon}\mathrm{Var}(\hat{V}(t+\Delta)) \end{aligned}$$

□

5 NUMERICAL TESTS

To test our discretization schemes, we turn to the pricing of European call options in the Heston model. This constitutes a standard test case, as prices can be computed with great precision from the analytical result in Proposition 3. We consider a call option C maturing at time T with strike K; let the exact option price at time 0 be $C(0)$. Using a discretization scheme that approximates $X(T)$ with $\hat{X}(T)$, we can establish an approximation $\hat{C}(0)$ to the option price by computing the expectation:

$$\hat{C}(0) = \mathrm{E}((\hat{X}(T) - K)^+)$$

Due to the errors introduced by the discretization of time, $\hat{C}(0)$ is generally not equal to $C(0)$. We define the bias e of a discretization scheme as:

$$e = C(0) - \hat{C}(0) \tag{47}$$

Clearly, e is a function of the time step Δ used in the discretization scheme; we are interested in establishing the function $e(\Delta)$ for the schemes outlined in previous chapters.

In (47), $C(0)$ can be computed by the technique in Proposition 3. To estimate $\hat{C}(0)$, we use Monte Carlo methods. Specifically, for a given discretization scheme for $\hat{X}$, we draw N independent samples of $\hat{X}^{(1)}(T)$, $\hat{X}^{(2)}(T)$, ..., $\hat{X}^{(N)}(T)$ using an equidistant time-grid with fixed step Δ; $\hat{C}(0)$ is then estimated in standard Monte Carlo fashion as:

$$\hat{C}(0) \approx \frac{1}{N} \sum_{i=1}^{N} (\hat{X}^{(i)}(T) - K)^+$$

The right-hand side of this equation is a random variable with mean $\hat{C}(0)$ and a standard deviation ("Monte Carlo error") of order $O(N^{-1/2})$. Using a sufficiently high number N of samples, we can keep the standard deviation low and obtain a high-accuracy estimate for $\hat{C}(0)$.

Having outlined the basic procedure to measure bias, let us set up some specific test cases. As discussed in Section 1, in our tests we wish to use parameters and option characteristics that are challenging and practically relevant. For this, we consider three quite different settings, listed in Table 1.

Loosely, the data of Case I are representative of the market for long-dated FX options, such as those that are embedded in the popular power-reverse dual contract. Case II could be considered representative for a long-dated interest rate option, and Case III has model parameters that may be encountered in equity options markets. Case III is similar to test cases prevalent in the existing literature; we expect it to be the most straightforward to handle numerically. For all test cases, we examine option prices at three levels of the strike: $K = 70$, $K = 100$ and $K = 140$.

In our numerical results, we use the following discretization schemes: the Euler scheme (6)–(7); the Kahl–Jackel scheme (8)–(9), denoted "IM–IJK";

TABLE 1 Test cases for numerical experiments.

	Case I	Case II	Case III
ε	1	0.9	1
κ	0.5	0.3	1
ρ	−0.9	−0.5	−0.3
T	10	15	5
$V(0), \theta$	4%	4%	9%

Note: In all cases $V(0) = \theta$ and $X(0) = 100$.

the TG scheme of Section 3.1; and the QE scheme of Section 3.2. For the latter two schemes, we use (33) to discretize $\ln X$, using central discretization ($\gamma_1 = \gamma_2 = 1/2$). In (33), we work both with and without martingale corrections (see (36)); we use "TG-M" and "QE-M" to label the martingale-corrected versions of schemes TG and QE, respectively. To keep the sample standard deviation low, all tests were run using a high[11] number of paths, $N = 10^6$. The random number generator used for all numerical experiments was[12] ran2 from Press *et al* (1992).

5.1 Results for Case I

For Case I, Table 2 lists Monte Carlo estimates of the bias $e(\Delta)$, for values of Δ ranging from 1/32 year to 1 year.

From the table, we see the following.

- The TG and, in particular, the QE schemes both have biases that are substantially lower than that of the Euler scheme: for at-the-money or out-of-the-money options, the bias of the TG/QE schemes at a value of $\Delta = 1$ or $\Delta = 1/2$ is roughly comparable to that of the Euler scheme at $\Delta = 1/32$.
- The QE scheme outperforms the TG scheme and has a bias that converges very rapidly as the time step is reduced: for all strikes in the table, a simulation step of $\Delta = 1/2$ or $\Delta = 1/4$ is sufficient to render the bias for the QE scheme statistically insignificant, even at 10^6 paths.
- The TG and QE schemes are robust with respect to option moneyness, with particularly strong performance for out-of-the money options.
- Adding a martingale correction to the TG and QE schemes generally lowers the bias further relative to the basic schemes, particularly (and not surprisingly) for the in-the-money options with $K = 70$.

[11] Ideally, we would have liked to have used an even higher number of sample paths, as the biases of the new schemes are quite low (as we shall see). Practical computing limitations, however, makes it difficult to increase the number of paths: for 32 steps per year, pricing a 15-year option in the Heston model requires drawing 960 random numbers for each path, so at 10^6 paths we already need about a billion random numbers (and associated manipulation of these numbers to increment X and V) to compute a single option price.

[12] This generator has a period of 2×10^{18}, which ensures that there is no chance of period exhaustion in our trials, despite the very large number of random numbers needed.

TABLE 2 Estimated bias (e) in test Case I.

Δ	Euler	IM+IJK	TG	QE	TG-M	QE-M
			$K = 100$			
1	−6.394 (0.029)	−57.648 (0.107)	−1.290 (0.013)	−1.022 (0.013)	−0.338 (0.012)	−0.233 (0.013)
1/2	−3.685 (0.021)	−32.977 (0.070)	−0.606 (0.013)	−0.311 (0.013)	−0.262 (0.013)	−0.133 (0.013)
1/4	−2.048 (0.017)	−18.427 (0.046)	−0.321 (0.013)	−0.049 (0.013)	−0.165 (0.013)	−0.002* (0.013)
1/8	−1.051 (0.015)	−10.121 (0.033)	−0.231 (0.013)	−0.002* (0.013)	−0.138 (0.013)	0.006* (0.013)
1/16	−0.516 (0.014)	−5.327 (0.024)	−0.158 (0.013)	0.004* (0.013)	−0.089 (0.013)	0.005* (0.013)
1/32	−0.243 (0.014)	−2.717 (0.020)	−0.126 (0.013)	−0.009* (0.013)	−0.062 (0.013)	−0.009* (0.013)
			$K = 140$			
1	−4.273 (0.019)	−51.611 (0.094)	0.091 (0.002)	0.077 (0.002)	0.108 (0.002)	0.086 (0.002)
1/2	−1.913 (0.010)	−28.153 (0.057)	0.027 (0.002)	0.023 (0.002)	0.043 (0.002)	0.025 (0.003)
1/4	−0.756 (0.006)	−14.785 (0.033)	0.011 (0.003)	0.004* (0.003)	0.023 (0.002)	0.004* (0.003)
1/8	−0.269 (0.004)	−7.527 (0.020)	0.007* (0.003)	−0.002* (0.003)	0.016 (0.002)	−0.002* (0.003)
1/16	−0.105 (0.003)	−3.608 (0.012)	0.007* (0.003)	0.000* (0.003)	0.013 (0.002)	0.000* (0.003)
1/32	−0.045 (0.003)	−1.636 (0.007)	0.001* (0.003)	0.000* (0.003)	0.005* (0.003)	0.000* (0.003)
			$K = 70$			
1	−3.955 (0.038)	−52.570 (0.116)	−1.203 (0.023)	−0.853 (0.023)	−0.231 (0.022)	−0.114 (0.022)
1/2	−2.180 (0.030)	−28.333 (0.078)	−0.593 (0.023)	−0.172 (0.023)	−0.181 (0.022)	0.012* (0.023)
1/4	−1.222 (0.026)	−14.535 (0.055)	−0.398 (0.022)	0.003* (0.023)	−0.171 (0.022)	0.025* (0.022)
1/8	−0.603 (0.024)	−7.242 (0.041)	−0.306 (0.022)	0.006* (0.023)	−0.147 (0.022)	0.008* (0.022)
1/16	−0.268 (0.023)	−3.530 (0.033)	−0.202 (0.022)	0.004* (0.022)	−0.069* (0.023)	0.003* (0.022)
1/32	−0.109 (0.023)	−1.786 (0.028)	−0.172 (0.022)	−0.020* (0.022)	−0.042* (0.023)	−0.021* (0.022)

Note: Numbers in parentheses are sample standard deviations. Starred results are those that are not statistically significant at the level of three sample standard deviations.

- The Euler scheme becomes increasingly competitive relative to the TG scheme when the strike is lowered. This is a consequence of the fact that the Euler scheme by construction is bias-free for the case $K = 0$, whereas the TG scheme is not (but TG-M is).
- The IM-IJK scheme does poorly, with biases that are substantially larger than those of the Euler scheme.

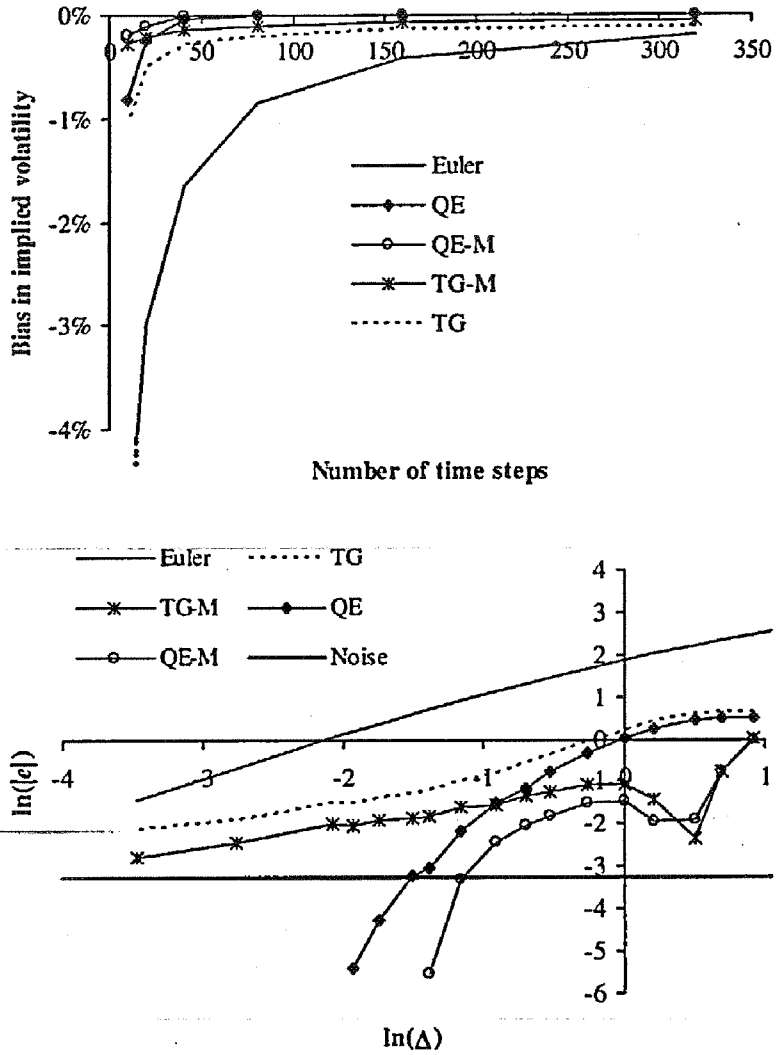
The poor performance of the IM-IJK scheme is surprising, given that this scheme is supposed to be particularly efficient for the setup that we consider in Case I, namely strong negative correlation ρ. To investigate whether the poor performance was caused by either the Milstein scheme (for the V process) or the IJK scheme (for X), we ran a series of tests where we combined an Euler scheme for V with the IJK scheme for X; the results were similar to those for IM-IJK in Table 2, suggesting that the IJK scheme for X is the main reason for the large biases.

To visualize some of the results in Table 2, consider the case $K = 100$, say, and let us convert biases in the table into errors in implied Black–Scholes volatility. The upper panel of Figure 5 shows some of the results; the superior performance of our new discretization techniques relative to the Euler scheme are obvious. Graphs for other values of K are similar.

One might at this point consider the problem of establishing empirical convergence order[13] for the various schemes covered in Table 2. An immediate problem here is the fact that our new schemes (QE and QE-M in particular) have biases that are so low that the bulk of the numbers in Table 2 are not statistically significant, and to make them so would require an impractical amount of computing effort. Instead, additional runs were undertaken with coarser time steps than those in the table, in the hope that a bias pattern might emerge from these runs. The results are listed in the lower panel of Figure 5. The Euler scheme has a convergence order of 1 (as expected), whereas the QE and QE-M schemes converge at a rate that is substantially higher than linear, but – at least for the values of Δ in the graph – no fixed convergence rate can be established. The convergence rate for the TG and TG-M schemes is lower than for the other two schemes and appears to be around 0.5. As a consequence, when the step size is reduced further than in the figure, the Euler scheme will (in all likelihood) eventually produce less biased results than the TG and TG-M schemes. This, however, will often be of limited practical relevance, as the precision of the TG and TG-M schemes is often adequate for applications long before the convergence "cross-over" point is reached. In particular, when running a practical number of paths ($\ll 10^6$), Monte Carlo noise for a practical number of paths will often overwhelm the bias of the TG and TG-M schemes, even when only a handful of time steps per year are used. A more penetrating analysis of the various trade-offs between bias and Monte Carlo noise could be performed along the lines of Duffie and Glynn (1995), but

[13] Recall that a discretization scheme has order n if the absolute value of the bias e decays as constant $\cdot \Delta^n$.

FIGURE 5 Convergence of bias.



Note: For test Case I and $K = 100$, the figures above show the convergence of the estimated call option price bias (e) as the time step Δ is reduced. The upper figure converts the bias into an error in implied Black–Scholes volatility and has the total number of time steps ($= T/\Delta$) on the x-axis. The lower figure graphs the logarithm of the absolute value of the bias against the logarithm of the time step. The "Noise" graph indicates the approximate level of the logarithmic bias below which it becomes statistically insignificant at the level of three sample standard deviations.

we skip it here as the QE scheme is so obviously the winner for the tests above. For notes on computation times, see Section 5.3.

5.2 Results for Case II and Case III

Tables 3 and 4 list estimated call option price biases for test cases II and III. Results for Case II are qualitatively and quantitatively very similar to those of Case I, with the TG and QE schemes outperforming the Euler scheme, which in turn outperforms the IM-IJK scheme. Of the schemes proposed in this paper, the TG scheme is again slightly worse than the QE scheme, which here performs very

strongly with all biases being statistically insignificant with just two steps per year. Adding martingale correction to the TG and QE schemes appears to yield less benefits than for test Case I above, although some improvements can be seen for $K = 70$.

For the less challenging Case III, our new schemes still perform significantly better than both the Euler and IM-IJK schemes, but now the IM-IJK scheme produces somewhat reasonable results, although the convergence of the bias is rather erratic. Thus, while there apparently are parameter combinations for which the IM-IJK scheme can be used, the scheme is not robust. In particular, as the variance of the V-process is increased – through a decrease of κ and/or an increase in ε – the IM-IJK scheme performs increasingly poorly.

5.3 Computational times

When comparing the numerical efficiency of discretization schemes, one needs to consider both the bias of the individual schemes, as well as the time it takes to compute a sample path. A scheme that computes very fast but has a large bias may in fact be preferable to a slower scheme with a low bias, to the extent that one can use a substantially smaller time step in the former scheme than in the latter for a fixed computational budget. For reference, Table 5 lists computing times for all schemes used in Sections 5.1 and 5.2, measured relative to the computing time of the Euler scheme.[14] The numbers were averages for all runs in Tables 1–4. As the QE scheme here is only marginally slower than the Euler scheme, the strong results of the QE scheme in Sections 5.1 and 5.2 makes it clearly preferable to the Euler scheme and it should be the method of choice. Martingale correction of the QE scheme (that is, the QE-M scheme) takes only a little extra time, and can be expected to yield some modest benefit for in-the-money options. Note that percentage timing differences between the various schemes would, of course, shrink in applications involving payouts more costly to compute than that of a simple European call option.

We note that the martingale-corrected TG scheme here is slower than the other schemes by a factor larger than two, which is a consequence of the fact that the martingale correction for the TG scheme is rather involved (and also, in part, a consequence of the fact that we did not bother to attempt to cache or otherwise optimize the algorithm). In light of the often modest gains associated with martingale correction, in most cases it should not be activated for the TG scheme.

6 EXTENSIONS

Before we conclude the paper, let us consider a few possible extensions.

[14]The computer used was a standard laptop PC with a Pentium CPU running at 1.6 GHz.

TABLE 3 Estimated bias (e) in test Case II.

Δ	Euler	IM+IJK	TG	QE	TG-M	QE-M
			$K = 100$			
1	−7.039 (0.073)	−22.496 (0.073)	0.516 (0.046)	0.459 (0.041)	0.694 (0.045)	0.528 (0.041)
1/2	−4.184 (0.064)	−12.231 (0.069)	0.249 (0.061)	0.108* (0.044)	0.357 (0.059)	0.118* (0.045)
1/4	−2.187 (0.050)	−6.590 (0.053)	0.102* (0.058)	0.019* (0.047)	0.178 (0.056)	0.019* (0.047)
1/8	−1.104 (0.044)	−3.178 (0.044)	0.056* (0.045)	0.019* (0.046)	0.100* (0.045)	0.019 (0.046)
1/16	−0.664 (0.073)	−1.674 (0.089)	0.029* (0.062)	−0.051* (0.050)	0.061* (0.062)	−0.042* (0.050)
1/32	−0.277 (0.042)	−0.711 (0.042)	0.072* (0.042)	0.026* (0.041)	0.074* (0.042)	0.026* (0.041)
			$K = 140$			
1	−6.067 (0.067)	−17.579 (0.062)	0.452 (0.040)	0.362 (0.035)	0.486 (0.040)	0.324 (0.035)
1/2	−3.351 (0.058)	−8.621 (0.061)	0.196 (0.057)	0.021* (0.039)	0.248 (0.055)	0.006* (0.039)
1/4	−1.611 (0.043)	−4.093 (0.045)	0.078* (0.054)	−0.001* (0.041)	0.123* (0.052)	−0.006* (0.041)
1/8	−0.749 (0.038)	−1.706 (0.036)	0.063* (0.039)	0.009* (0.041)	0.093* (0.039)	0.007* (0.041)
1/16	−0.481 (0.070)	−0.853 (0.086)	0.045* (0.058)	−0.053* (0.044)	0.047* (0.057)	−0.054* (0.044)
1/32	−0.172 (0.036)	−0.297 (0.035)	0.082* (0.035)	0.010* (0.034)	0.056* (0.036)	0.010* (0.034)
			$K = 70$			
1	−4.565 (0.078)	−19.482 (0.080)	−0.337 (0.050)	−0.161 (0.046)	−0.114* (0.050)	−0.070* (0.046)
1/2	−2.698 (0.069)	−9.866 (0.075)	−0.172* (0.064)	−0.090* (0.049)	−0.037* (0.063)	−0.076* (0.050)
1/4	−1.326 (0.055)	−4.999 (0.059)	−0.151* (0.062)	−0.016* (0.052)	−0.049* (0.060)	−0.015* (0.052)
1/8	−0.632 (0.050)	−2.246 (0.050)	−0.115* (0.050)	0.021* (0.051)	−0.060* (0.050)	0.021* (0.051)
1/16	−0.413 (0.077)	−1.229 (0.092)	−0.147* (0.066)	−0.053* (0.054)	−0.065* (0.065)	−0.054* (0.054)
1/32	−0.150 (0.048)	−0.517 (0.047)	−0.032* (0.047)	0.033* (0.047)	0.049* (0.047)	0.033* (0.047)

Note: Numbers in parentheses are sample standard deviations. Starred results are those that are not statistically significant at the level of three sample standard deviations.

TABLE 4 Estimated bias (e) in test Case III.

Δ	Euler	IM+IJK	TG	QE	TG-M	QE-M
			$K = 100$			
1	−4.365 (0.074)	0.258 (0.049)	0.483 (0.054)	0.372 (0.052)	0.634 (0.055)	0.492 (0.053)
1/2	−2.277 (0.062)	0.940 (0.050)	0.235 (0.053)	0.123* (0.054)	0.291 (0.053)	0.144* (0.054)
1/4	−1.119 (0.057)	0.955 (0.054)	0.078* (0.055)	−0.084* (0.057)	0.104* (0.055)	−0.077 * (0.057)
1/8	−0.439 (0.054)	0.862 (0.053)	0.116* (0.054)	0.024* (0.053)	0.121* (0.054)	0.025* (0.053)
1/16	−0.235 (0.053)	0.534 (0.052)	0.079* (0.053)	−0.078* (0.054)	0.087* (0.053)	−0.077* (0.054)
1/32	−0.060* (0.055)	0.308 (0.054)	0.024* (0.055)	0.023* (0.055)	0.028* (0.055)	0.023* (0.055)
			$K = 140$			
1	−4.495 (0.066)	0.924 (0.040)	0.728 (0.046)	0.557 (0.044)	0.707 (0.047)	0.529 (0.045)
1/2	−2.264 (0.054)	1.104 (0.042)	0.332 (0.045)	0.164 (0.046)	0.334 (0.045)	0.132* (0.046)
1/4	−1.092 (0.048)	0.967 (0.046)	0.151 (0.047)	−0.071* (0.049)	0.159 (0.047)	−0.074* (0.049)
1/8	−0.432 (0.045)	0.808 (0.045)	0.169 (0.046)	0.020* (0.044)	0.176 (0.046)	0.019* (0.044)
1/16	−0.209 (0.045)	0.507 (0.044)	0.094* (0.045)	−0.067* (0.046)	0.098* (0.045)	−0.067* (0.046)
1/32	−0.076* (0.047)	0.269 (0.046)	0.030* (0.046)	0.005* (0.047)*	0.032 (0.046)	0.005* (0.047)
			$K = 70$			
1	−2.957 (0.080)	0.219 (0.056)	−0.328 (0.060)	−0.188 (0.058)	−0.113* (0.061)	−0.010* (0.059)
1/2	−1.522 (0.068)	0.832 (0.057)	−0.136* (0.060)	−0.100* (0.060)	−0.058* (0.060)	−0.052* (0.061)
1/4	−0.737 (0.063)	0.714 (0.060)	−0.134* (0.061)	−0.124* (0.063)	−0.098* (0.061)	−0.113* (0.063)
1/8	−0.245 (0.061)	0.636 (0.059)	0.020* (0.060)	0.028* (0.059)	0.039* (0.060)	0.031* (0.059)
1/16	−0.166* (0.060)	0.358 (0.059)	−0.006* (0.060)	−0.100* (0.061)	0.006* (0.060)	−0.099* (0.061)
1/32	−0.011* (0.061)	0.208 (0.060)	−0.022* (0.061)	0.044* (0.062)	−0.015* (0.061)	0.044* (0.062)

Note: Numbers in parentheses are sample standard deviations. Starred results are those that are not statistically significant at the level of three sample standard deviations.

TABLE 5 Average computation times, relative to the Euler scheme.

IM-IJK	TG	QE	TG-M	QE-M
1.23	1.28	1.21	2.92	1.38

6.1 Displacement

In interest rate applications, it is often technically convenient to assume that $\rho = 0$ in the Heston model. As this generally does not produce option prices that calibrate well to the market, a separate "local volatility" mechanism is introduced into the model to mimic the effect of negative correlation between X and V. A standard model (see, for example, Andersen and Andreasen (2002) or Piterbarg (2005)) replaces (2) in the Heston SDE with:

$$\mathrm{d}X(t) = (hX(t) + k)\sqrt{V(t)}\,\mathrm{d}W_X(t)$$

where h and k are positive constants. Let $X^*(t) = hX(t) + k$ and $V^*(t) = h^2V(t)$. An application of Ito's lemma then shows that:

$$\mathrm{d}X^*(t) = X^*(t)\sqrt{V^*(t)}\,\mathrm{d}W_X(t),$$
$$\mathrm{d}V^*(t) = \kappa(h^2\theta - V^*(t))\,\mathrm{d}t + h\varepsilon\sqrt{V^*(t)}\,\mathrm{d}W_V(t)$$

This vector SDE can be discretized with the methods in this paper; the resulting path for $X^*(t)$ can be translated into paths of $X(t)$ by the relation $X(t) = (X^*(t) - k)/h$.

It is equally easy to introduce a displacement in the process for V, which allows us to work with processes of the form:

$$\mathrm{d}V(t) = \kappa(\theta - V(t))\,\mathrm{d}t + \varepsilon\sqrt{h + V(t)}\,\mathrm{d}W_V(t)$$

where h is some constant. We leave the details to the reader.

6.2 Time-dependent parameters

In some applications, certain parameters of the Heston SDE are functions of time. One such application can be found in Andersen and Andreasen (2002), where the process for $X(t)$ is written as:

$$\mathrm{d}X(t)/X(t) = \lambda(t)\sqrt{V(t)}\,\mathrm{d}W_X(t) \tag{48}$$

where λ is a bounded deterministic function of time. In Andersen and Andreasen (2002), the process for $V(t)$ has constant parameters and can easily be discretized by the schemes in Section 3. To handle (48), we could assume that λ can be approximated as being piecewise flat on $[t, t + \Delta]$ with value $\bar{\lambda}$; for instance, we could set $\bar{\lambda}$ to $(\lambda(t) + \lambda(t + \Delta))/2$. This leads to a trivial modification of the

sampling scheme for $\ln X$:

$$\ln \hat{X}(t+\Delta) = \ln \hat{X}(t) + \frac{\rho}{\varepsilon}\bar{\lambda}(\hat{V}(t+\Delta) - \hat{V}(t) - \kappa\theta\Delta) + \Delta\bar{\lambda}\left(\frac{\kappa\rho}{\varepsilon} - \frac{\bar{\lambda}}{2}\right)(\gamma_1\hat{V}(t) + \gamma_2\hat{V}(t+\Delta)) + \bar{\lambda}\sqrt{1-\rho^2}\sqrt{\gamma_1\hat{V}(t) + \gamma_2\hat{V}(t+\Delta)}\cdot Z$$

where the notation is the same as in (33). For the more general case where the parameters of the process for $V(t)$ may also depend on time (as in Piterbarg (2005)), we proceed in the same fashion and approximate all parameters as being piecewise flat on the discretization grid; this, in turn, allows for application of all schemes in this paper. See (Glasserman 2003, p. 130) for similar ideas.

6.3 Jumps in stock price and variance

Both Broadie and Kaya (2006) and Lord *et al* (2006) consider a model where a Poisson jump term is added to the basic Heston dynamics of the X-process. Specifically, we write:

$$\mathrm{d}\ln X = -\tfrac{1}{2}V(t)\,\mathrm{d}t - \eta\bar{\mu}\,\mathrm{d}t + \sqrt{V(t)}\,\mathrm{d}W_X(t) + J(t)\,\mathrm{d}N(t)$$

where $N(t)$ is a Poisson process with intensity η and $J(t)$ is Gaussian. Both N and J are assumed to be independent of the Brownian motions for V and X. Adding the term $J(t)\,\mathrm{d}N(t)$ to the process of $\ln X$ will induce jumps: if $N(t)$ increments at time t, the X process jumps to $X(t)\,\mathrm{e}^{J(t)}$. Note that we have added a martingale compensation drift $\eta\bar{\mu}$ to keep X a martingale; $\bar{\mu}$ is given by $\bar{\mu} = \mathrm{E}(\mathrm{e}^J) - 1$.

Simulation in the model is trivial due to the independence assumption. Specifically, we can write:

$$\ln X(t) = \ln X^*(t) + Z(t), \quad Z(t) = \int_0^T J(t)\,\mathrm{d}N(t)$$

where $\ln X^*(t)$ is governed by a standard jump-free Heston model. The simulation techniques developed in this paper can be used to generate paths for $\ln X^*(t)$ (incorporation of the drift $\eta\bar{\mu}$ is trivial) and paths of $Z(t)$ can be generated by overlaying samples from a Poisson distribution with Gaussian jumps.

In some applications, jumps may also be added to the variance process V. The proper way to generate paths in this case is to first draw all the jump times of the V process and then use one of the discretization schemes for V (see Section 3) *between* these jump dates. We trust that the user can intuitively grasp how this is done; the paper by Broadie and Kaya (2006) contains further details.

7 CONCLUSION

In this paper, we have considered two new discretization schemes – denoted TG and QE – to be used in Monte Carlo simulation of Heston (and Heston-like)

models. The schemes also have applications for simulation of affine models in more generality. Our proposed discretization schemes are based on careful analysis of the true, and often rather problematic, behavior of affine square-root processes, combined with a mechanism to generate the correct amount of co-dependence between the variance process and the asset process. The schemes introduced in this paper are simple to implement and generally yield very substantial efficiency improvements over existing methods. Of the schemes considered, the QE scheme should be the default choice, due to its simplicity and strong performance; martingale correction (the QE-M scheme) is optional. The TG scheme has considerable intuitive appeal, but has sublinear convergence and generally performs somewhat worse than QE at practically relevant time steps. In the TG scheme, the variance process is simulated by applying a guaranteed monotonic transformation to a Gaussian random variable; this may make this scheme more natural to use than the QE scheme in multi-asset applications that involve several correlated variance processes. Examination of such multidimensional applications is left for future research.

Computational performance tests of the proposed schemes were carried out using realistic and challenging model parameters and payout characteristics. While all our new schemes were robust under changes to model parameters and option moneyness, some schemes in the existing literature did not do as well as expected. The "fixed" Euler scheme of Lord *et al* (2006) has acceptable behavior but generally requires substantially more time steps than any of our new schemes before biases are reduced to acceptable levels. Still, for benign parameters (eg, where $2\kappa\theta > \varepsilon^2$) we find that the scheme in Lord *et al* (2006) has good performance and would likely be a perfectly adequate choice. The scheme in Kahl and Jackel (2005) was not robust in our tests and returned very high biases in some cases. Even with benign model parameters, the scheme did not consistently outperform Euler stepping.

While the schemes in this paper are already significant improvements over existing methods, we do not doubt that additional performance can be teased out of the fundamental ideas of the paper. Experiments with better approximations to time integrals of the variance process – perhaps using bridge-type processes or moment matching (as suggested in Section 4.2) – may be one avenue to pursue in future research. Suitable applications of the results in Appendix B when the V process is close to zero might also reduce bias even further, as might, say, more complicated switching rules in the QE/QE-M schemes. For such high-precision results to have much practical relevance, however, methods must be introduced to reduce Monte Carlo noise below the levels we encountered in this paper. Construction of such variance reduction methods is yet another topic that may be pursued in future research.

APPENDIX A MOMENTS OF V AND $\ln X$

PROPOSITION A1 *For some $T > 0$, consider the joint characteristic function:*

$$\varphi(u, v) = \mathrm{E}(\mathrm{e}^{iuV(T)+ivx(T)}), \quad x(T) = \ln X(T)/X(0)$$

where X and V are characterized by the vector SDE (3)–(4). Define:

$$d(v) = \sqrt{(\mathrm{i}v\rho\varepsilon - \kappa)^2 + v^2\varepsilon^2 + \varepsilon^2 \mathrm{i}v}$$

$$Q(u, v) = \frac{\alpha_+(v) - \mathrm{i}u}{\alpha_-(v) - \mathrm{i}u}, \quad \alpha_\pm(v) = \frac{\kappa - \mathrm{i}v\rho\varepsilon \pm d(v)}{\varepsilon^2}$$

Then:

$$\varphi(u, v) = \mathrm{e}^{C(T;u,v)+D(T;u,v)V(0)} \tag{A.1}$$

with:

$$D(T; u, v) = \alpha_+(v)\frac{1 - Q(u, v)\,\mathrm{e}^{d(v)T}\alpha_-(v)/\alpha_+(v)}{1 - Q(u, v)\,\mathrm{e}^{d(v)T}}$$

$$C(\tau; u, v) = \kappa\theta\left[\alpha_+(v)\tau + \frac{\alpha_-(v) - \alpha_+(v)}{d(v)}\ln\left(\frac{Q(u, v)\,\mathrm{e}^{d(v)\tau} - 1}{Q(u, v) - 1}\right)\right]$$

PROOF Let:

$$q(t, V, x; u, v) = \mathrm{E}(\mathrm{e}^{\mathrm{i}uV(T)+\mathrm{i}vx(T)} | V(t) = V, x(t) = x)$$

such that $\varphi(u, v) = q(0, V(0), 0; u, v)$. From standard results for diffusion processes, q must satisfy a PDE:

$$\frac{\partial q}{\partial t} - \frac{1}{2}V\frac{\partial q}{\partial x} + \kappa(\theta - V)\frac{\partial q}{\partial V} + \rho\varepsilon V\frac{\partial^2 q}{\partial V \partial x} + \frac{1}{2}V\frac{\partial^2 q}{\partial x^2} + \frac{1}{2}\varepsilon^2 V\frac{\partial^2 q}{\partial V^2} = 0$$

subject to the terminal boundary condition $q(T, V, x; u, v) = \mathrm{e}^{\mathrm{i}uV+\mathrm{i}vx}$. The affine form of our equations suggest that:

$$q(t, V, x; u, v) = \mathrm{e}^{C(T-t;u,v)+D(T-t;u,v)V+\mathrm{i}vx}$$

Insertion of this expression into the PDE above yields a Riccati system of ordinary ODEs for C and D, which can be solved by separation of variables. The result (A.1) then follows. □

Equipped with the characteristic function $\varphi(u, v)$ as computed in (A.1), we can (assisted by a symbolic calculus computer package) establish various moments of $\ln X(T)$ and $V(T)$ by differentiation. First, let us define a few auxiliary variables:

$$\Omega_1 = \mathrm{e}^{-2\kappa T}\varepsilon^2 + 4\,\mathrm{e}^{-\kappa T}((1 + \kappa T)\varepsilon^2 - 2\rho\kappa\varepsilon(2 + \kappa T) + 2\kappa^2) + (2\kappa T - 5)\varepsilon^2 - 8\rho\kappa\varepsilon(\kappa T - 2) + 8\kappa^2(\kappa T - 1)$$

$$\Omega_2 = -\mathrm{e}^{-2\kappa T}\varepsilon^2 + 2\,\mathrm{e}^{-\kappa T}(-\kappa T\varepsilon^2 + 2\rho\varepsilon\kappa(1 + \kappa T) - 2\kappa^2) + \varepsilon^2 - 4\kappa\rho\varepsilon + 4\kappa^2$$

$$\Omega_3 = \mathrm{e}^{-2\kappa T} + 2\kappa\,\mathrm{e}^{-\kappa T}\left(T - \frac{2\rho}{\varepsilon}(1 + \kappa T)\right) + \frac{4\kappa\rho - \varepsilon}{\varepsilon}$$

$$\Omega_4 = \mathrm{e}^{-\kappa T}\left(1 - \kappa T + \frac{2\rho\kappa^2 T}{\varepsilon}\right) - \mathrm{e}^{-2\kappa T}$$

With these definitions, we have the results in Table A.1.

TABLE A.1 Exact and first-order moments.

Moment	Value	$o(T)$ limit
$\mathrm{E}(\ln X(T))$	$\ln X(0) + \frac{1}{2\kappa}(\theta - V(0))(1 - \mathrm{e}^{-\kappa T}) - \frac{1}{2}\theta T$	$\ln X(0) - \frac{1}{2}V(0)T$
$\mathrm{Var}(\ln X(T))$	$\frac{\theta}{8\kappa^3}\Omega_1 + \frac{V(0)}{4\kappa^3}\Omega_2$	$V(0)T$
$\mathrm{E}(V(T))$	$\theta + (V(0) - \theta)\,\mathrm{e}^{-\kappa T}$	$V(0) + \kappa(\theta - V(0))T$
$\mathrm{Var}(V(T))$	$\frac{V(0)\varepsilon^2}{\kappa}(\mathrm{e}^{-\kappa T} - \mathrm{e}^{-2\kappa T}) + \frac{\theta\varepsilon^2}{2\kappa}(1 - \mathrm{e}^{-\kappa T})^2$	$V(0)\varepsilon^2 T$
$\mathrm{Cov}(\ln X(T), V(T))$	$\frac{\theta\varepsilon^2}{4\kappa^2}\Omega_3 + \frac{V(0)\varepsilon^2}{2\kappa^2}\Omega_4$	$\rho V(0)\varepsilon T$

APPENDIX B REFINEMENT OF THE QUADRATIC-EXPONENTIAL SCHEME FOR SMALL V

First, let us consider a pure central chi-squared distribution with ν degrees of freedom; the relevant density is given in (12). We assume that $\nu \leq 2$ and wish to approximate the true density by an expression of the form:

$$h(x) = \begin{cases} C_1 x^q, & 0 \leq x < x_c \\ C_2\,\mathrm{e}^{-\beta x}, & x \geq x_c \end{cases} \tag{B.1}$$

The following result establishes the constants C_1, C_2, β and x_c by moment matching.

PROPOSITION B1 *Let Q be a random variable with chi-squared density (12), where $0 < \nu \leq 2$. Let Y_ν be a variable with density h in (B.1), and define $q = \nu/2 - 1 \in (-1, 0)$. Set $x_c = -q$ and:*

$$k_2 = \frac{2q^2}{(q+3)(q+2)^2}$$

$$k_1 = q^2\left(\frac{q+1}{q+3} - 2\frac{(q+1)(3q+4)}{(q+2)^2}\right) - 4(q+1)(q+2)$$

$$k_0 = 2\left(\frac{(q+1)(3q+4)}{q+2}\right)^2$$

$$y = \frac{-k_1 - \sqrt{k_1^2 - 4k_2k_0}}{2k_2}$$

For Y_v and Q to have identical first and second moments, we must have:

$$C_1 = (1-y)(q+1)\cdot(-q)^{-(q+1)} \tag{B.2}$$

$$\beta = \left(\left(2q+2-\frac{C_1}{q+2}(-q)^{q+2}\right)y^{-1}+q\right)^{-1} \tag{B.3}$$

$$C_2 = y\beta\,\mathrm{e}^{-\beta q} \tag{B.4}$$

At these parameter values, $h(x)$ is a proper density.

PROOF By direct computation, we notice that the cumulative distribution corresponding to density h is:

$$H(x) = \Pr(Y_v \le x) = \begin{cases} \dfrac{C_1}{q+1}x^{q+1}, & 0 \le x \le x_c \\ \dfrac{C_1}{q+1}x_c^{q+1} + \dfrac{C_2}{\beta}(\mathrm{e}^{-\beta x_c} - \mathrm{e}^{-\beta x}), & x > x_c \end{cases} \tag{B.5}$$

In particular:

$$H(\infty) = \frac{C_1}{q+1}x_c^{q+1} + \frac{C_2}{\beta}\,\mathrm{e}^{-\beta x_c}$$

Straightforward integration shows that:

$$\mathrm{E}(Y_v) = \frac{C_1}{q+2}x_c^{q+2} + \frac{C_2}{\beta}\,\mathrm{e}^{-\beta x_c}(x_c+\beta^{-1})$$

$$\mathrm{E}(Y_v^2) = \frac{C_1}{q+3}x_c^{q+3} + \frac{C_2}{\beta}\,\mathrm{e}^{-\beta x_c}(x_c^2 + 2x_c\beta^{-1} + 2\beta^{-2})$$

To establish a reasonable value for x_c, let us first note that, for $v = 2$ and $v \to 0$, the form of $h(x)$ becomes exactly identical to that of a central chi-squared density, provided that we set $x_c = 0$ and $x_c \to 1$, respectively. Assuming linearity, a pragmatic general choice for x_c is then to set:

$$x_c = 1 - v/2 = -q$$

To find the remaining constants (C_1, C_2, β), we match the first two moments against those of the true chi-squared distribution. As a chi-squared distribution with v degrees of freedom has mean v and variance $2v$, we get, after inclusion of the condition $H(\infty) = 1$, the following system of equations:

$$\frac{C_1}{q+1}x_c^{q+1} + y = 1 \tag{B.6}$$

$$\frac{C_1}{q+2}x_c^{q+2} + (x_c+\beta^{-1})y = v \tag{B.7}$$

$$\frac{C_1}{q+3}x_c^{q+3} + (x_c^2 + 2x_c\beta^{-1} + 2\beta^{-2})y = v^2 + 2v \tag{B.8}$$

Here, we have defined $y = y(C_2, \beta) = (C_2/\beta)\, e^{-\beta x_c}$.

To solve (B.6)–(B.8), we eliminate C_1 and β, to yield a single equation in y. Specifically, (B.6) allows us to write C_1 as a function of y and Equation (B.7) then allows us to also write β as a function of y. Insertion of the resulting expressions for C_1 and β in (B.8) then yields, after several trivial rearrangements and use of $x_c = -q$:

$$k_2 y^2 + k_1 y + k_0 = 0 \tag{B.9}$$

where the constants k_0, k_1 and k_2 were defined above. Solution of (B.9) yields the result for y listed in the proposition; it can be verified that the solution always exists for the range of q covered in the proposition. Inserting y in (B.6)–(B.8) (with $x_c = -q$) yields the results of (B.2)–(B.4). □

Having established a workable approximation for a chi-squared distribution with low degrees of freedom, let us consider how we can use this in an approximation for a *non-central* chi-squared distribution.

PROPOSITION B2 *Let the random variable R be distributed according to a non-central chi-squared distribution with d degrees of freedom and non-centrality parameter λ. Set $c = (d + 2\lambda)/(d + \lambda)$, and assume that $(d + \lambda)/c \le 2$. The distribution of R can be approximated by:*

$$\Pr(R \le x) = \Pr(Y_\nu \le x/c) \tag{B.10}$$

where $\nu \approx (d + \lambda)/2c$ and the distribution of Y_ν is given by the density h in Proposition B1. In particular, the first two moments of Y_ν and R coincide.

PROOF Following the *ansatz* in Patnaik (1949), one can approximate R as being distributed as c times a central chi-squared distribution with ν degrees of freedom. Appropriate values for c and ν can be found by moment matching to be:

$$c = (d + 2\lambda)/(d + \lambda), \quad \nu = (d + \lambda)/c$$

Assuming that $\nu \le 2$, the result in Proposition B1 can then be used to establish the result (B.10). □

With the result in Proposition B2, we immediately get the following result.

PROPOSITION B3 *Consider the Heston variance process (2). For some positive time step Δ, define $k = e^{-\kappa\Delta} n(t, t + \Delta)^{-1}$, where $n(t, t + \Delta)$ is defined in Proposition 1. Set $d = 4\kappa\theta/\varepsilon^2$, $\lambda = \hat{V}(t) n(t, t + \Delta)$, $c = (d + 2\lambda)/(d + \lambda)$ and $q = (d + \lambda)/(2c) - 1$. Assuming that $q \in (-1, 0)$, then, as an approximation:*

$$\Pr(V(t + \Delta) \le x | V(t)) \approx H\left(\frac{x}{kc}\right) \tag{B.11}$$

where $H(x)$ is given in (B.5) with $x_c = -q$, and C_1, C_2 and β are computed as prescribed in Proposition B1. In particular, with $V(t + \Delta)$ distributed according to (B.11), $\mathrm{E}(V(t + \Delta)|V(t)) = m$ and $\mathrm{Var}(V(t + \Delta)|V(t)) = s^2$, where m and s are the true moments given in (17) and (18).

PROOF We recall from Proposition 1 that, conditional on $V(t)$, $V(t+\Delta)$ is k times a non-central chi-squared distributed with d degrees of freedom and non-centrality parameter λ. The result of the proposition then follows directly from Proposition B2 above. □

The result above hinges on the condition $q \in (-1, 0)$ or, equivalently, $0 < (d+\lambda)/(2c) < 1$. Only the upper bound of 1 is relevant here; it translates into $(d+\lambda)^2/(d+2\lambda) < 2$. Insertion of the definitions of d and λ followed by a few rearrangements reduce this to the requirement $\psi > 1$, where $\psi = s^2/m^2$. This restriction coincides (not surprisingly) with that of the scheme (26), allowing us to substitute in the QE scheme (26) with sampling from the distribution (B.11).

Sampling from (B.11) requires inversion of the cumulative distribution function H. The form of H, however, allows this to be done in closed form. Specifically, we have:

$$H^{-1}(u) = \begin{cases} \left(\dfrac{(q+1)u}{C_1}\right)^{1/(q+1)}, & 0 \le u \le u_c \\ -\beta^{-1}\ln\left(e^{\beta q} - \dfrac{u-u_c}{C_2}\right), & u_c < u \le 1 \end{cases}$$

$$u_c \equiv \frac{C_1}{q+1}(-q)^{q+1}$$

REFERENCES

Andersen, L., and Andreasen, J. (2002). Volatile volatilities. *Risk* **15**, 163–168.

Andersen, L., and Brotherton-Ratcliffe, R. (2005). Extended LIBOR market models with stochastic volatility. *The Journal of Computational Finance* **9**(1), 1–40.

Andersen, L., and Piterbarg, V. (2005). Moment explosions in stochastic volatility models. *Finance and Stochastics* **11**(1), 29–50.

Andreasen, J. (2006). Long-dated FX hybrids with stochastic volatility. Working Paper, Bank of America.

Broadie, M., and Kaya, Ö. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research* **54**(2), 217–231.

Broadie, M., and Kaya, Ö. (2004). Exact simulation of option greeks under stochastic volatility and jump diffusion models. *Proceedings of the 2004 Winter Simulation Conference*, Ingalls, R. G., *et al* (eds).

Carr, P., and Madan, D. (1999). Option pricing and the fast Fourier transform. *The Journal of Computational Finance* **2**(4), 61–73.

Cox, J., Ingersoll, J., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica* **53**(2), 385–407.

Ding, C. G. (1992). Algorithm AS275: computing the non-central chi-square distribution function. *Applied Statistics* **41**, 478–482.

Duffie, D., and Glynn, P. (1995). Efficient Monte Carlo simulation of security prices. *Annals of Applied Probability* **5**, 897–905.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica* **68**, 1343–1376.

Dufresne, D. (2001). The integrated square-root process. Working Paper, University of Montreal.

Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer, New York.

Glasserman, P. and Zhao, X. (1999). Arbitrage-free discretization of log-normal forward LIBOR and swap rate models. *Finance and Stochastics* **4**, 35–68.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**(2), 327–343.

Johnson, N., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2. Wiley Interscience, New York.

Kahl, C., and Jackel, P. (2005). Fast strong approximation Monte-Carlo schemes for stochastic volatility models. Working Paper, ABN AMRO and University of Wuppertal.

Lee, R. (2004). Option pricing by transform methods: extensions, unification, and error control. *The Journal of Computational Finance* **7**(3), 51–86.

Lewis, A. (2001). *Option Valuation Under Stochastic Volatility*. Finance Press, Newport Beach.

Lipton, A. (2002). The vol smile problem. *Risk* **15**(2), 61–65.

Lord, R., Koekkoek, R., and van Dijk, D. (2006). A comparison of biased simulation schemes for stochastic volatility models. Working Paper, Tinbergen Institute.

Kloeden, P., and Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations*, 3rd edn. Springer, New York.

Moro, B. (1995). The full Monte. *Risk* **8**(2), 57–58.

Patnaik, P. (1949). The non-central χ^2- and F-distributions and their applications. *Biometrika* **36**, 202–232.

Pearson, E. (1959). Note on an approximation to the distribution of non-central χ^2. *Biometrika* **46**, 364.

Piterbarg, V. (2003). Discretizing processes used in stochastic volatility models. Working Paper, Bank of America.

Piterbarg, V. (2005). Stochastic volatility model with time-dependent skew. *Applied Mathematical Finance* **12**, 147–185.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical Recipes in C*. Cambridge University Press, New York.

Smith, R. (2007). An almost exact simulation model for the Heston model. *The Journal of Computational Finance* **11**, 1.