Estimation of continuous-time models with an application to equity volatility dynamics

Gurdip Bakshi, Nengjiu Ju, Hui Ou-Yang*

First Version: February 21, 2004
Current Version: January 24, 2005

*Bakshi and Ju are both at Department of Finance, Smith School of Business, University of Maryland, College Park, MD 20742. Ou-Yang is at Fuqua School of Business, Duke University, Durham, NC 27708. Bakshi can be reached at Tel: (301) 405-2261, email: gbakshi@rhsmith.umd.edu, Website: www.rhsmith.umd.edu/faculty/gbakshi/; Ju at Tel: (301) 405-2934, email: nju@rhsmith.umd.edu, and Website: www.rhsmith.umd.edu/faculty/nju/; Ou-Yang can be reached at Tel: (919) 660-3790, email: huiou@duke.edu, and Website: www.fuqua.duke.edu/faculty/alpha/huiou.htm. The authors gratefully acknowledge useful discussions with Yacine Aït-Sahalia, Doron Avramov, Charles Cao, Peter Carr, Ren-Raw Chen, Bob Dittmar, Steve Heston, Gautam Kaul, Mark Loewenstein, Dilip Madan, Matt Pritsker, Tyler Shumway, Guojun Wu, and Liuren Wu. Seminar participants at Michigan Business School, Rutgers University, and 2004 European Finance Association meetings (Maastricht) provided constructive comments. The feedback of an anonymous referee and Bill Schwert has substantially improved this paper. Computer codes used to implement the continuous-time models are available from any of the authors.

Estimation of continuous-time models with an application to equity volatility dynamics

Abstract

The treatment of this article renders closed-form density approximation feasible for univariate continuous-time models. Implementation methodology depends directly on the parametric-form of the drift and the diffusion of the primitive process and not its transformation to a unit-variance process. Offering methodological convenience, the approximation method relies on numerically evaluating one-dimensional integrals and circumvents existing dependence on intractable multidimensional integrals. Density-based inferences can now be drawn for a broader set of models of equity volatility. Our empirical results provide insights on crucial outstanding issues related to the ranking-orderings of continuous-time stochastic volatility models, the absence/presence of non-linearities in the drift function, and the desirability of pursuing more flexible diffusion function specifications.

JEL classification: G10; G11; G12; G13; C15; C32; C52.
Keywords: Continuous-time models, Maximum-likelihood estimation, Density approximation; Equity volatility; Volatility dynamics

Looking back over the history of financial economics there is little doubt that continuous-time stochastic processes have played a prominent role in the development of theoretically plausible and empirically testable models of portfolio selection, fixed income, equity derivatives, and expected stock returns. Whether the end-goal is martingale pricing or maximum-likelihood estimation, theory invariably requires the knowledge of the transition density of the economic forcing variable which is generally unamenable to closed-form characterization. In this sense the lack of analyticity of the density function has hampered empirical testing and the validation of alternative hypotheses about continuous-time models. To remedy this deficiency, Aït-Sahalia (1999, 2002) has proposed a method to approximate the transition density in a one-dimensional diffusion setting. Given its potential utility to the researcher in applied and theoretical work, the thrust of this article is to expand on the analytical density approach of Aït-Sahalia (1999, 2002) and our treatment renders the original method feasible for a substantially larger class of one-dimensional models. Armed with this modification we empirically implement the density approximation method to study the plausibility of general models of equity volatility. Density-based inferences allow us to disentangle issues connected with (i) the ranking-ordering of continuous-time volatility models, (ii) the presence of non-linearities in the drift function, and (iii) the desirability of adopting more flexible diffusion specifications.

The motivation for our analysis to expand on Aït-Sahalia (1999, 2002) derives from two considerations. First, in the context of one-dimensional diffusions $dX_t = \mu[X_t]\,dt + \sigma[X_t]\,dW_t$, for economic-variable X_t, extant density approximations hinge crucially on transforming X_t to a unit-variance process via $\int^X \frac{du}{\sigma[u]}$ and then on inverting $\int^X \frac{du}{\sigma[u]}$. This requirement has proved analytically challenging for some interesting economic models (see Bakshi and Ju (2002) for examples). Second, in the enhanced-method of Aït-Sahalia (1999, 2002), the recursively defined coefficients that fulfill the forward/backward equation have a multidimensional integral dependence and are seldom tractable outside of the constant elasticity of variance diffusion class. The framework of this paper overcomes both hurdles associated with implementing Aït-Sahalia (1999, 2002). Broadening the appeal of the methodology we show that the density approximation can be derived without reducing the primitive process to a unit-variance process and without analytically integrating and inverting $\int^X \frac{du}{\sigma[u]}$. The contribution of our approach also lies in determining the recursively defined expansion coefficients that exhibit at most a single integral dependence and consequently affords tractability. An advantage of this new approach is that it causes the density approximation to be virtually analytical for continuous-time models with nonlinear drift and diffusion functions of the general type analyzed in Aït-Sahalia (1996).

Market index volatility is one of the most fundamental variables determined in financial markets and is a particularly relevant input into option pricing, risk management systems, and volatility-based contingent claims (i.e., variance futures, and options on volatility). Despite the flurry of recent modeling efforts (see Andersen, Bollerslev, Diebold, and Ebens (2001), Duffie, Pan, and Singleton (2000), Heston (1993), and Jones (2003)) there is yet no consensus on the dynamic evolution of volatility in continuous-time. Exploiting the closed-form density approximation, our empirical analysis of daily market volatility provides evidence for a volatility process that has substantial nonlinear mean-reverting drift underpinnings.

Supporting a strand of drift specifications taking the parametric form $\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1}$, the finding of statistically significant $\alpha_2 < 0$ and $\alpha_3 > 0$ indicates drift-reversals at both high and low ends of the volatility spectrum. Volatility processes omitting a role for nonlinear diffusion coefficient $\sigma[X] \equiv \sqrt{\beta_1 X_t + \beta_2 X_t^{\beta_3}}$ with $\beta_3 > 2$ are structurally flawed and destined for unsatisfactory empirical performance. Thus, the inconsistency of affine volatility models (i.e., Heston (1993)) can be attributed to misspecified drift and diffusion coefficients. Overall, a market variance specification with nonlinear drift and diffusion function delivers the most desirable goodness-of-fit statistics and this new result has wide-ranging implications for pricing and trading of risks associated with equity and volatility derivatives.

The rest of the paper is divided into four parts. Section 1 discusses the enhanced density approximation method of Aït-Sahalia (1999, 2002) and develops results aimed at simplifying the multidimensional structure of the expansion coefficients up to fourth-order. Our characterizations are derived entirely in terms of the drift and diffusion of the underlying primitive process. Section 2 presents the density approximation for the encompassing model of Aït-Sahalia (1996). The next section is devoted to empirically evaluating continuous-time models of equity volatility. Section 4 summarizes our contributions and provides concluding statements.

1 Maximum-likelihood estimation of continuous-time models

Consider a one-dimensional diffusion process for a state variable X_t:

$$dX_t = \mu[X_t;\theta]\,dt + \sigma[X_t;\theta]\,dW_t, \tag{1}$$

where $\mu[X_t;\theta]$ and $\sigma[X_t;\theta]$ are respectively the coefficients of drift and diffusion, and θ represents the parameter vector in an bounded set $\Theta \subset R^K$. The maximum likelihood estimation of θ using discretely observed data requires the underlying transition density.

To facilitate empirical testing using density methods, Aït-Sahalia (1999, 2002) has developed two analytical density approximations. Of particular interest are the *enhanced formulae* in Aït-Sahalia (1999, 2002) which correspond to the limit where the order of the Hermite polynominals converges to infinity and is derived by forcing the coefficients to fulfill the Fokker-Plank-Kolmogorov partial differential equation. The contribution of this section is to propose a modification to the enhanced-method and shows that the resulting density approximation applies to a broader class of $\mu[X_t;\theta]$ and $\sigma[X_t;\theta]$.

1.1 Enhanced-method in Aït-Sahalia (1999, 2002)

Aït-Sahalia (1999, 2002) constructs a unit-variance process Y_t defined by:

$$Y_t \equiv \gamma[X;\theta] = \int^{X_t} \frac{du}{\sigma[u;\theta]}. \tag{2}$$

Letting $\gamma^{-1}[y;\theta]$ be the inverse function of $\gamma[X;\theta]$, the drift of dY_t is:

$$\mu_Y[y;\theta] = \frac{\mu\left[\gamma^{-1}[y;\theta];\theta\right]}{\sigma\left[\gamma^{-1}[y;\theta];\theta\right]} - \frac{1}{2}\frac{\partial \sigma}{\partial x}\left[\gamma^{-1}[y;\theta];\theta\right]. \tag{3}$$

Denoting $\phi[z] \equiv e^{-z^2/2}/\sqrt{2\pi}$ and Δ as a discrete time interval, Aït-Sahalia (1999, 2002) shows that the density of $Y_t = y$ can be approximated up to the K-th term by

$$p_Y^{(K)}[\Delta, y|y_0;\theta] = \Delta^{-1/2}\,\phi\left[\frac{y-y_0}{\Delta^{1/2}}\right]\exp\left(\int_{y_0}^{y}\mu_Y[w;\theta]\,dw\right)\sum_{k=0}^{K} c_k[y|y_0;\theta]\,\frac{\Delta^k}{k!}, \tag{4}$$

where $c_0[y|y_0,\theta] \equiv 1$ and, for $j \geq 1$, the recursive coefficients, $c_j[y|y_0;\theta]$, can be derived by solving:

$$c_j[y|y_0;\theta] = j\,(y-y_0)^{-j}\int_{y_0}^{y}(w-y_0)^{j-1}\left(\lambda[w]\,c_{j-1}[w] + \frac{1}{2}\,\frac{\partial^2 c_{j-1}[w|y_0;\theta]}{\partial w^2}\right)dw, \tag{5}$$

by defining

$$\lambda[y;\theta] \equiv -\frac{1}{2}\left(\mu_Y^2[y;\theta] + \frac{\partial \mu_Y[y;\theta]}{\partial y}\right). \tag{6}$$

The transition density of X_t is then obtained through the Jacobian formula as:

$$p_X^{(K)}[\Delta, x|x_0;\theta] \quad = \quad (\sigma[x;\theta])^{-1}\,p_Y^{(K)}[\Delta, \gamma[x]\,|\,\gamma[x_0];\theta]. \tag{7}$$

There are several aspects of this methodology that need fine-tuning. For the expansion in (4) to converge, the X_t process (1) must first be transformed to be sufficiently Gaussian. Based on the theoretical models adopted by Aït-Sahalia (1999) it is clear that when Y_t and $\gamma^{-1}[y;\theta]$ are in analytical closed-form, then $\mu_Y[y;\theta]$ and $\lambda[y;\theta]$ are completely analytical and the reduction to a unit-variance diffusion is feasible. Both Brandt and Santa-Clara (2002, p163) and Durham (2003, p465) have argued that - for a broad class of continuous-time models - the reduction step from X_t to Y_t is restrictive and can curb the appeal of the density approximation in empirical applications.

Recognizing this disparity between theory and implementation, Bakshi and Ju (2002) relax the requirement that both Y_t and $\gamma^{-1}[y;\theta]$ be known analytically, and explain how the Hermite expansions in the basic-approach of Aït-Sahalia (2002) can be reformulated so that only the numerical value of Y_t is needed. While the Bakshi and Ju (2002) refinement is appealing because it makes the density approximation possible for a wide class of $\sigma[X]$, there are reasons to believe that the method in (4) is superior if it can be transformed to make it apply to an equally wide class. The accuracy tests in Bakshi and Ju (2002) indicate that the approximation based on (4) is accurate with expansion coefficients as few as three, and is substantially more reliable (see also our comparisons to come in Table 1). The Hermite approach relies on expanding the density of Y_t around a standard normal while the expansion in (4) forces the density function to satisfy the Kolmogorov forward and backward equations to the order Δ^K, resulting in greater accuracy.

The second point relates to the determination of the recursively defined $c_j[y|y_0;\theta]$. From the form of (5) it may be observed that $c_1[y|y_0;\theta]$ and $c_2[y|y_0;\theta]$ can be derived by solving one-dimensional integrals and two-dimensional integrals respectively, and higher-dimensional integrals are involved in implementing the density approximation with $c_3[y|y_0;\theta]$ and beyond. Aït-Sahalia (1999) has solved the density function for models satisfying $dX_t = (\alpha_0+\alpha_1 X_t+\alpha_2\, X_t^2+\alpha_3\, X_t^{-1})\, dt+ \sigma\, X_t^{\rho}\, dW_t$ where each $c_j[y|y_0;\theta]$ is fully analytical. However, the expressions for leading terms $c_j[y|y_0;\theta]$ are still unknown under a general class of $\mu[X_t;\theta]$ and $\sigma[X_t;\theta]$. Particularly when $dX_t = (\alpha_0 + \alpha_1\, X_t + \alpha_2\, X_t^2 + \alpha_3\, X_t^{-1})\, dt + \sqrt{\beta_0 + \beta_1\, X_t + \beta_2\, X_t^{\beta_3}} dW_t$ the multidimensional integrals embedded in $c_j[y|y_0;\theta]$ remain unsolved, thereby eluding closed-form characterizations for $\{c_1[y|y_0;\theta],\, c_2[y|y_0;\theta],\, c_3[y|y_0;\theta],\, \cdots,\, c_K[y|y_0;\theta]\}$. The lack of analyticity of $c_j[y|y_0;\theta]$ is problematic: It has impaired the density-based inference of models with general $\mu[X;\theta]$ and $\sigma[X;\theta]$.

Citing this reason, Durham (2003), Durham and Gallant (2002), and Brandt and Santa-Clara (2002) have argued in favor of simulation-based methods. However, simulation methods can be cumbersome and computationally expensive. Our objective is to characterize transition densities for a broad parametric class of diffusion processes and exploit them for empirical testing and model selection.

1.2 Central elements of the modification

The proposed approximation method, outlined below, directly exploits the form of $\mu[X_t;\theta]$ and $\sigma[X_t;\theta]$ and bypasses closed-form reliance on $\mu_Y[y;\theta]$ and $\int^{X_t} \frac{du}{\sigma[u;\theta]}$ in practical applications. Our analytical contributions also ensure that each $c_j[y|y_0;\theta]$ contains at most one-dimensional integrals rather than a set of complex multidimensional integrals.

1.2.1 Circumventing the reliance of the approximation on $\mu_Y[y;\theta]$

Models for $\mu[X;\theta]$ and $\sigma[X;\theta]$ contained in Aït-Sahalia (1999) stress that when Y_t and $\gamma^{-1}[Y;\theta]$ are analytical so are $\mu_Y[Y]$ and $\lambda[Y]$. In this family of continuous-time models, equations (4), (5), and (7) constitute a conceptually simple and accurate approximation method. However, when Y_t and/or $\gamma^{-1}[Y;\theta]$ do not admit closed-form representation, it is expedient to reexpress all required density approximation components in terms of $\mu[X;\theta]$ and $\sigma[X;\theta]$ of the original X_t process.

Proposition 1 *Define the function*

$$f[X] \equiv \frac{\mu[X]}{\sigma[X]} - \frac{\sigma'[X]}{2} \tag{8}$$

by analogy with (3) with $\sigma'[X] \equiv \frac{\partial\sigma[X]}{\partial X}$. The following density approximation components can be obtained in terms of $\mu[X]$ and $\sigma[X]$ of the original process X_t (suppressing θ dependence):

$$\lambda[y] = -\frac{1}{2}\left(f^2[x] + f'[x]\,\sigma[x]\right), \tag{9}$$

$$\lambda'[y] \equiv \frac{\partial\lambda[y;\theta]}{\partial y} = -\frac{1}{2}\,\sigma[x]\,\left(f^2[x] + f'[x]\,\sigma[x]\right)' \tag{10}$$

$$\lambda''[y] \equiv \frac{\partial^2\lambda[y;\theta]}{\partial y^2} = -\frac{1}{2}\,\sigma[x]\,\left(\left(f^2[x] + f'[x]\sigma[x]\right)'\,\sigma[x]\right)' \tag{11}$$

$$y - y_0 = \int_{x_0}^{x} \frac{du}{\sigma[u]}, \tag{12}$$

$$\int_{y_0}^{y} \mu_Y[w]\,dw = \int_{x_0}^{x} f[u]\frac{du}{\sigma[u]}, \tag{13}$$

$$\int_{y_0}^{y} \lambda[w]\,dw = -\frac{1}{2}\int_{x_0}^{x}\left(f^2[u] + f'[u]\sigma[u]\right)\frac{du}{\sigma[u]}, \tag{14}$$

$$\int_{y_0}^{y} \lambda^2[w]\,dw = \frac{1}{4}\int_{x_0}^{x}\left(f^2[u] + f'[u]\sigma[u]\right)^2\frac{du}{\sigma[u]}. \tag{15}$$

Proof: Make the change of variable $x = \gamma^{-1}[y]$. From the definition of $\gamma[X]$ given in (2), we have $dy = dx/\sigma[x]$. From the definition of $\lambda[y]$ in (6) and the new variable x we have (9). The chain rule of differentiation implies $\lambda'[y] = \frac{\partial\lambda[y]}{\partial x}\frac{\partial x}{\partial y}$ in (10). Similarly, we obtain (11). Noting that $\mu_Y[y] = \mu[\gamma^{-1}[y]]/\sigma[\gamma^{-1}[y]] - \frac{\partial\sigma}{\partial x}(\gamma^{-1}[y])/2 = \mu[x]/\sigma[x] - \sigma'[x]/2$ and using $dy = dx/\sigma[x]$, we have (13). Using (9) and $dy = dx/\sigma[x]$, we obtain (14) and (15). □

1.2.2 Reduction of $c_j[y|y_0;\theta]$ to a set of one-dimensional integrals

Proposition 2 *For the recursively defined coefficients $c_j[y|y_0;\theta]$ in (5) and $\lambda[y;\theta]$ defined in (6), we may derive the higher-order $c_j[y|y_0;\theta]$ analytically with only one-dimensional integral dependence (suppressing the y_0 and θ arguments):*

$$
\begin{aligned}
c_1[y] &= \frac{1}{y-y_0}\int_{y_0}^{y}\lambda[w]\,dw, && (16)\\
c_2[y] &= c_1^2[y] + \frac{1}{(y-y_0)^2}(\lambda[y]+\lambda[y_0]-2\,c_1[y]), && (17)\\
c_3[y] &= c_1^3[y] + \frac{3}{(y-y_0)^2}\,(c_1[y](\lambda[y]+\lambda[y_0]) - 3c_2[y]) + \\
&\quad \frac{3}{(y-y_0)^3}\left(\frac{\lambda'[y]-\lambda'[y_0]}{2} + \int_{y_0}^{y}\lambda^2[w]\,dw\right), && (18)
\end{aligned}
$$

and,

$$
\begin{aligned}
c_4[y] &= c_1^4[y] + \frac{3}{(y-y_0)^2}(2\,\lambda[y]\,c_2[y] - 8\,c_3[y] + 2\,\lambda[y_0]\,c_1^2[y]) + \\
&\quad \frac{12\,c_1[y]}{(y-y_0)^3}\left(\frac{\lambda'[y]-\lambda'[y_0]}{2} + \int_{y_0}^{y}\lambda^2[w]\,dw\right) + \\
&\quad \frac{3}{(y-y_0)^4}(3\,\lambda^2[y] + 5\,\lambda^2[y_0] + 4\,\lambda[y]\,c_1[y] - 12\,c_2[y] + \lambda''[y] + \lambda''[y_0]). && (19)
\end{aligned}
$$

Proof: See Appendix A. □

To obtain coefficient $c_1[y]$ involves computing two simple integrals: $y - y_0 = \int_{x_0}^{x}\frac{du}{\sigma[u]}$ and $\int_{y_0}^{y}\lambda[w]\,dw$. Once $c_1[y]$ is obtained, $c_2[y]$ follows immediately and $c_3[y]$ merely requires $\int_{y_0}^{y}\lambda^2[w]\,dw$. Again, with $c_3[y]$ known, $c_4[y]$ can be easily recovered and involves no further integrals. Viewed from this perspective of solving one-dimensional integrals, the density approximation with $K = 4$ constitutes an efficient method. This proposition achieves the crucial task of reducing the recursively defined multidimensional integrals in $c_j[y_0;\theta]$ to those involving only the one-dimensional integrals outlined in Proposition 1.

With the relevant components in the density approximation expressed directly in terms of the original state variable X_t, its drift $\mu[X]$ and diffusion function $\sigma[X]$, the method under consideration can be applied to any selected scaler diffusion processes. The density approximation becomes

$$
p_X^{(K)}[\Delta, x|x_0;\theta] \approx \frac{\Delta^{-1/2}}{\sigma[X;\theta]}\,\phi\left[\frac{1}{\Delta^{1/2}}\int_{x_0}^{x}\frac{du}{\sigma[u]}\right]\exp\left(\int_{x_0}^{x} f[u]\frac{du}{\sigma[u]}\right)\sum_{k=0}^{4} c_k[\gamma[x]\,|\gamma[x_0];\theta]\,\frac{\Delta^k}{k!}, \qquad (20)
$$

where $\{c_1[y|y_0;\theta],\ldots,c_4[y|y_0;\theta]\}$ are presented in (16)-(19). Although some integrals still remain to be determined in our formulation in (13)-(15), but they are solely required for their numerical values. When $x-x_0$ is small, these integrals can be approximated by the Taylor series in Proposition 3 which renders the formulation completely analytical. Compared with the recursively defined multidimensional integrals the simplified $c_j[y]$'s are easier to evaluate and this connection will be highlighted in the context of the general model of Aït-Sahalia (1996). To guarantee that the density remains positive, the approximation for log-density is used: $\log\left(p_X^{(K)}[\Delta,x|x_0;\theta]\right)\approx -\log\left(2\pi\sigma^2[X;\theta]\Delta\right)/2-\left(\int_{x_0}^{x}\frac{du}{\sigma[u]}\right)^2/(2\Delta)+\int_{x_0}^{x}f[u]\frac{du}{\sigma[u]}+\sum_{k=0}^{4}\mathcal{C}_k[\gamma[x]\,|\gamma[x_0];\theta]\,\frac{\Delta^k}{k!}$, where $\mathcal{C}_1\equiv c_1$, $\mathcal{C}_2\equiv c_2-c_1^2$, $\mathcal{C}_3\equiv c_3-3c_2\,c_1+2c_1^3$, and $\mathcal{C}_4\equiv c_4-4c_3c_1-3c_2^2+12c_2c_1^2-6c_1^4$.

Aït-Sahalia (2003) has extended the density approximation method in Aït-Sahalia (1999, 2002) to higher-dimensional diffusion processes. It must be emphasized that the problem of determining $c_j[y_0|y_0;\theta]$ is even harder in the multivariate setting and we have been unable to work through the multivariate counterparts of Proposition 1 and Proposition 2. See Aït-Sahalia and Kimmel (2003) for closed-form likelihood expansions under affine multifactor models of the term structure.

Although not done here to preserve focus, the method can be adapted to approximate risk-neutral densities for valuing contingent claims when the characteristic function of the state-price density is unavailable (Bakshi and Madan (2000) and Duffie, Pan, and Singleton (2000)). Suppose the log stock price is $X_t\equiv\log(\mathcal{P}_t)$, then Ito's lemma implies $dX_t=(r-\frac{1}{2}\sigma[e^X;\theta])\,dt+\sigma[e^X;\theta]\,dW_t$. The risk-neutral density approximation $q_X^{(K)}[X]$ leads to European option prices with strike price $\mathcal{K}$ and maturity T as $e^{-rT}\int_{\log(\mathcal{K})}^{+\infty}(e^{X_{t+T}}-\mathcal{K})\,q_X^{(K)}[X_{t+T}]\,dX_{t+T}$.

1.2.3 Density approximation when $x-x_0$ and/or $y-y_0$ is small

For some applications (say, the spot interest rates), the difference between the two adjacent observations (x_0 and x) can be small. In such applications, Proposition 3 derives the corresponding closed-form approximation of the relevant integrals and renders the method totally analytical.

Proposition 3 *Let $E\equiv x-x_0$, $D\equiv y-y_0$, and define the successive partial derivative entities:*

$$\nu_i = \partial^i(1/\sigma[X_0])/\partial X_0^i, \tag{21}$$

$$\varphi_i = \partial^i(f[X_0]/\sigma[X_0])/\partial X_0^i \tag{22}$$

$$\lambda_i = \partial^i\lambda[y_0]/\partial y_0^i. \tag{23}$$

When $x-x_0$, equivalently $y-y_0$, is small, we have the following Taylor expansions of $\int_{x_0}^{x}\frac{du}{\sigma[u]}$,

$\int_{x_0}^{x} f[u] \frac{du}{\sigma[u]}$:

$$D = y - y_0 = \int_{x_0}^{x} \frac{du}{\sigma[u]} = \nu_0 E + \frac{\nu_1 E^2}{2} + \frac{\nu_2 E^3}{6} + \frac{\nu_3 E^4}{24} + \frac{\nu_4 E^5}{120} + \frac{\nu_5 E^6}{720}, \quad (24)$$

$$\int_{y_0}^{y} \mu_Y[w]\, dw = \int_{x_0}^{x} f[u] \frac{du}{\sigma[u]} = \varphi_0 E + \frac{\varphi_1 E^2}{2} + \frac{\varphi_2 E^3}{6} + \frac{\varphi_3 E^4}{24} + \frac{\varphi_4 E^5}{120} + \frac{\varphi_5 E^6}{720}, \quad (25)$$

and $c_1[y|y_0;\theta]$, $c_2[y|y_0;\theta]$, $c_3[y|y_0;\theta]$, *and* $c_4[y|y_0;\theta]$ *are presented in (62)-(65) of Appendix B.*

Owing to the results in (24)-(25) and (62)-(67), the approximation for the transition density (20) is now completely analytical. The partial derivatives, ν_i, φ_i and λ_i, are compact and can be conveniently programmed for the estimation of a wide class of continuous-time models. The tractability of $c_j[y_0|y_0;\theta]$ and the methodological dependence on $\mu[X;\theta]$ and $\sigma[X;\theta]$ are at the core of the analytical density approximation.

Upon further reflection, the coefficients in our Taylor series in Proposition 3 are related to the corresponding coefficients in Aït-Sahalia (2003) when the irreducible method is applied to the univariate case. Realize, however, in the irreducible multivariate modeling case, the expansion coefficients do not afford the tractability of integral representations and must be approximated by Taylor series in $x - x_0$ irrespective of whether $x - x_0$ is small or not. On the other hand, the integral equation (5) and the reduced coefficients in Proposition 2 hold for univariate diffusions and can be evaluated through efficient integration routines. In fact, Proposition 3 may be construed as stating that when $x - x_0$ is small, the appropriate integrals can be determined by appealing to Taylor series approximations making our method entirely analytical. In the irreducible multivariate setting, one must resort to Taylor approximations.

2 Density approximation for a parametric model class

Spurred by our characterizations in Proposition 1 through Proposition 3, this section applies the density approximation (20) to the following eight-parameter encompassing class of one-dimensional processes due to Aït-Sahalia (1996):

$$\mu[X;\theta] = \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1}, \quad (26)$$

$$\sigma[X;\theta] = \sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}, \quad (27)$$

subject to technical conditions (24a) through (24d) in Aït-Sahalia (1996). Here,

$$\theta \equiv (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2, \beta_3)\,. \quad (28)$$

Following the convention in Durham (2003), this general process is labeled as **GEN4** and, under appropriate restrictions, subsumes several theoretically appealing models:

GEN4: $dX_t = \left(\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1}\right) dt + \sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}\, dW_t.$

GEN2: $dX_t = (\alpha_0 + \alpha_1 X_t)\, dt + \sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}\, dW_t$

GEN1: $dX_t = \alpha_0\, dt + \sqrt{\beta_0 + \beta_1 X_t + \beta_2 x_t^{\beta_3}}\, dW_t$

CEV4: $dX_t = \left(\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1}\right) dt + \beta_2 X_t^{\beta_3}\, dW_t$

CEV2: $dX_t = (\alpha_0 + \alpha_1 X_t)\, dt + \beta_2 X_t^{\beta_3}\, dW_t$

CEV1: $dX_t = \alpha_0\, dt + \beta_0 X_t^{\beta_1}\, dW_t$

AFF: $dX_t = (\alpha_0 + \alpha_1 X_t)\, dt + \sqrt{\beta_0 + \beta_1 X_t}\, dW_t$

As argued earlier, to apply the enhanced-method of Aït-Sahalia (1999, 2002) to the process in (26)-(27) requires $Y_t \equiv \int^{X_t} \frac{du}{\sigma[u;\theta]} = \int \frac{du}{\sqrt{\beta_0 + \beta_1 u + \beta_2 u^{\beta_3}}}$, which has no known closed-form analytical representation. Thus, leaving aside the additional issue of multiple numerical integration determination of higher-order $c_j[y|y_0;\theta]$, none of the models in the GEN class are amenable to a density characterization under the approach of Aït-Sahalia (1999, 2002).

Returning to our methodology, we determine the components of $p_X^{(K)}[\Delta, x|x_0;\theta]$ in (20) by defining $V_0 \equiv \beta_0 + \beta_1 X_0 + \beta_2 X_0^{\beta_3}$. Clearly, $\nu_0^2 V_0 = 1$ where $\nu_0 \equiv 1/\sqrt{\beta_0 + \beta_1 X + \beta_2 X^{\beta_3}}$. Straightforward successive differentiation of $\nu_0^2 V_0 = 1$ with respect to X_0 yields the first five derivatives of ν_0 with respect to X_0:

$$\nu_1 = -\nu_0 V_1/(2V_0), \tag{29}$$

$$\nu_2 = -(3\nu_1 V_1 + \nu_0 V_2)/(2V_0), \tag{30}$$

$$\nu_3 = -(5\nu_2 V_1 + 4\nu_1 V_2 + \nu_0 V_3)/(2V_0), \tag{31}$$

$$\nu_4 = -(7\nu_3 V_1 + 9\nu_2 V_2 + 5\nu_1 V_3 + \nu_0 V_4)/(2V_0), \tag{32}$$

$$\nu_5 = -(9\nu_4 V_1 + 16\nu_3 V_2 + 14\nu_2 V_3 + 6\nu_1 V_4 + \nu_0 V_5)/(2V_0), \tag{33}$$

where $V_1 = \beta_1 + \beta_2\beta_3\, X_0^{\beta_3 - 1}$, $V_2 = \beta_2\beta_3(\beta_3 - 1)\, X_0^{\beta_3-2}$, $V_3 = (\beta_3 - 2)V_2/X_0$, $V_4 = (\beta_3 - 3)V_3/X_0$, and $V_5 = (\beta_3 - 4)V_4/X_0$ are the partial derivatives of V_0 with respect to X_0.

Proceeding, as in Proposition 3, we obtain $\varphi_i = \partial^i (f[X_0]/\sigma[X_0])/\partial X_0^i$. Recognize that $\varphi_0 = f[X_0]/\sigma[X_0] = \mu[X_0]/V_0 - V_1/(4V_0) = U_0/V_0$, where $U_0 = \mu[X_0] - V_1/4 = \alpha_0 - \beta_1/4 + \alpha_1 X_0 +$

$\alpha_2\, X_0^2 + \alpha_3/X_0 - \beta_2\beta_3\, X_0^{\beta_3 - 1}/4$. Thus, $\varphi_0 V_0 = U_0$. Successive differentiation of this equation produces the first five derivatives of φ_0 with respect to the state variable X_0,

$$\varphi_1 = (U_1 - \varphi_0)V_1/V_0, \tag{34}$$

$$\varphi_2 = (U_2 - 2\varphi_1 V_1 - \varphi_0 V_2)/V_0, \tag{35}$$

$$\varphi_3 = (U_3 - 3\varphi_2 V_1 - 3\varphi_1 V_2 - \varphi_0 V_3)/V_0, \tag{36}$$

$$\varphi_4 = (U_4 - 4\varphi_3 V_1 - 6\varphi_2 V_2 - 4\varphi_1 V_3 - \varphi_0 V_4)/V_0, \tag{37}$$

$$\varphi_5 = (U_5 - 5\varphi_4 V_1 - 10\varphi_3 V_2 - 10\varphi_2 V_3 - 5\varphi_1 V_4 - \varphi_0 V_5)/V_0, \tag{38}$$

where $U_1 = \alpha_1 + 2\alpha_2 X_0 - \alpha_3/X_0^2 - V_2/4$, $U_2 = 2\alpha_2 + 2\alpha_3/X_0^3 - V_3/4$, $U_3 = -6\alpha_3/X_0^4 - V_4/4$, $U_4 = 24\alpha_3/X_0^5 - V_5/4$, and $U_5 = -120\alpha_3/X_0^6 - V_6/4$ are the partial derivatives of U_0 with respect to X_0 and $V_6 = (\beta_3 - 5)V_5/X_0$. For $\mu[X_0]$ and $\sigma[X_0]$ governed via (26)-(27), the recursive nature of φ_i determines (25) of Proposition 3.

Finally, we characterize each $\lambda_i = \partial^i \lambda[y_0]/\partial y_0^i$ in (66)-(67) as a function of X_0. Based on the calculations in Appendix C, each required λ_i is:

$$\lambda_0 = H_0, \tag{39}$$

$$\lambda_1 = H_1 S_0, \tag{40}$$

$$\lambda_2 = H_2 S_0^2 + \lambda_1 S_1, \tag{41}$$

$$\lambda_3 = H_3 S_0^3 + 3\lambda_2 S_1 + \lambda_1(S_0 S_2 - 2S_1^2), \tag{42}$$

$$\lambda_4 = H_4 S_0^4 + 6\lambda_3 S_1 + \lambda_2(4S_0 S_2 - 11S_1^2) + \lambda_1(S_0^2 S_3 - 6S_0 S_1 S_2 + 6S_1^3), \tag{43}$$

$$\begin{aligned}\lambda_5 = {} & H_5 S_0^5 + 10\lambda_4 S_1 + \lambda_3(10S_0 S_2 - 35S_1^2) + \lambda_2(5S_0^2 S_3 - 40S_0 S_1 S_2 + 50S_1^3) + \\ & \lambda_1(S_0^3 S_4 - 8S_0^2 S_1 S_3 + 36S_0 S_1^2 S_2 - 6S_0^2 S_2^2 - 24S_1^4),\end{aligned} \tag{44}$$

$$\begin{aligned}\lambda_6 = {} & H_6 S_0^6 + 15\lambda_5 S_1 + \lambda_4(20S_0 S_2 - 85S_1^2) + \lambda_3(15S_0^2 S_3 - 150S_0 S_1 S_2 + 225S_1^3) + \\ & \lambda_2(6s_0^3 S_4 - 63S_0^2 S_1 S_3 + 346S_0 S_1^2 S_2 - 46S_0^2 S_2^2 - 274S_1^4) + \lambda_1(S_0^4 S_5 - \\ & 10S_0^3 S_1 S_4 + 6S_0^2 S_1^2 S_3 - 20S_0^3 S_2 S_3 - 240S_0 S_1^3 S_2 + 90S_0^2 S_1 S_2^2 + 120S_1^5),\end{aligned} \tag{45}$$

where S_0 through S_7 are shown in (68)-(75), and H_0 through H_6 are shown in (84)-(90). The analyticity of λ_i can now be used to build $c_1[y|y_0;\theta]$ through $c_4[y|y_0;\theta]$ in (62)-(65) and φ_i in (34)-(38) are employed for constructing $\int_{y_0}^{y} \mu_Y[w]\, dw$ expansion in (25). Given the choice of $\mu[X;\theta]$ and $\sigma[X;\theta]$, we obtain the density approximation for GEN4 through (20). Density functions for other continuous-time models are restricted special cases.

Before proceeding to empirical investigation, it is instructive to determine the accuracy of the density approximation under the proposed scheme for determining $c_j[y|y_0;\theta]$ in (20). To illustrate this aspect we pick two candidate continuous-time models by setting $\beta_0 \equiv 0$ in model AFF (i.e., the square-root model) and $\alpha_0 \equiv 0$ in model CEV2. Each candidate model has an exact density which allows comparison to the approximate density. Guided by Aït-Sahalia (1999) and Bakshi and Ju (2002) we compare the maximum absolute error of the approximate density relative to its exact density counterpart. Table 1 judges the worst possible approximation error by presenting MAXE $\equiv \max\left(|p^{\text{exact}}[x|x_0] - p^{\text{approx}}[x|x_0]|\right)$ and the maximum exact conditional density as $\max(p^{\text{exact}}[x|x_0]$. The key finding is that our density approximation is accurate for both $\Delta = 1/12$ and $\Delta = 1$ regardless of the underlying stochastic process. Inspection of the results also reveals little value-added by including $c_5[y|y_0;\theta]$ in $\sum_{k=0}^{K} c_k[\gamma[x]\,|\gamma[x_0];\theta]\,\frac{\Delta^k}{k!}$: the approximation with $K = 5$ marginally improves over $K = 4$ in some cases. The reason appears to be that the maximum absolute errors with $K = 4$ - which are in the order of 10^{-14} - have probably reached machine precision. However, as would be expected, embedding each additional term in the approximation tends to make the method progressively more accurate. To summarize the proposed density approximation method not only achieves the desired degree of accuracy, but applies to a class of economically relevant continuous-time models.

3 Evaluating continuous-time volatility models

Starting with Heston (1993), there is a tradition to model equity volatility as a continuous-time stochastic process with mean-reverting drift and square-root volatility. Despite the insight this model has enabled, there is, however, a growing consensus that the square-root specification is grossly misspecified (see Andersen, Benzoni and Lund (2002), Bakshi, Cao, and Chen (2000), Bates (2000), Elerian, Chib, and Shephard (2001), Eraker, Johannes, and Polson (2003), and Pan (2002), among others). With the exception of Jones (2003), who has provided evidence in favor of CEV models of the volatility process, the lack of closed-form density approximations has impeded progress on the testing of volatility processes beyond square-root. As such, several questions still remain unresolved with respect to the documented rejection of square-root volatility models: (i) Does the drift of the volatility process admit departures from linearity? (ii) Are volatility models with general $\sigma[X]$ more properly specified from empirical standpoints? (iii) What is the empirical potential of variance processes in the Aït-Sahalia (1996) class (26)-(27)? Issues connected with volatility modeling have bearing on the search for better performing option pricing models, parametric compensation for volatility risk, and the timing of volatility risks.

Before we can address aforementioned questions we need a suitable proxy for market volatility, which is intrinsically unobservable. Among the various possible choices at the daily frequency, the empirical literature has appealed to (i) GARCH volatility constructed from daily returns, (ii) intraday squared return measures (Andersen, Bollerslev, and Diebold (2003)), (iii) short-term near-money Black-Scholes implied volatility (Pan (2002)), and (iv) S&P 100 index option volatility, VIX (Jones (2003)). For reasons outlined in Jones (2003), we adopt the forward-looking VIX volatility extracted from index option prices in our empirical work. Thus, it must be understood at the outset that we are drawing conclusions about the desirability of stochastic volatility models based on the estimated dynamics of the VIX index.

Given to us by Chris Jones, the VIX is sampled over the period of July 1, 1988 to January 10, 2000 (2907 observations) and expressed in decimals. To aid comparisons with the empirical literature we let $X_t \equiv \mathrm{VIX}_t^2$ for $\{t = i\,\Delta \mid i = 0, \cdots, n\}$.

The estimate of θ in (28) is based on the log-likelihood function:

$$\max_{\theta} \ \mathcal{L}[\theta] \equiv \sum_{i=1}^{n} \log \left\{ p_X[\Delta, X_{i\Delta} | X_{(i-1)\Delta}; \theta] \right\} \tag{46}$$

where the transition density function, $p_X[\Delta, X_{i\Delta}|X_{(i-1)\Delta};\theta]$, is approximated via (20) with $\{c_1[y_0|y_0;\theta], \cdots, c_4[y_0|y_0;\theta]\}$ constructed separately for GEN4 and its nested variants. The efficiency and optimality of the maximum-likelihood estimator is discussed, among others, in Aït-Sahalia (1999, 2002).[1]

Table 2 displays maximum-likelihood model parameters, the standard errors (in parenthesis), the goodness-of-fit maximized log-likelihood values, and the rank-ordering of stochastic volatility models based on the Akaike Information Criterion (AIC). For each candidate model the estimate of β_0 in the diffusion function $\sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}$ was close to zero with no impact on log-likelihood. For this reason we throughout impose $\beta_0 \equiv 0$. In this case, the AFF specification collapses to the square-root model $dX_t = (\alpha_0 - \alpha_1 X_t)\, dt + \sqrt{\beta_1 X_t} dW_t$.

The failure of the square-root model is evident even in the presence of a high β_1 value of 0.1827 (i.e., volatility of volatility parameter, $\sqrt{\beta_1}$, is 0.4274). One drawback of this specification is that it is insufficiently flexible in fitting higher-order volatility moments: With sample VIX skewness and kurtosis of 2.39, and 9.67 respectively, this process can internalize large movements in the

[1] Demonstrating the computational superiority of the density approximation, the MLE computer code converged rapidly (in less than 1 minute on a 1 GHZ laptop computer) regardless of the drift and diffusion combination. Having a closed-form density approximation can therefore accelerate the speed of estimation several hundred fold relative to simulation-based approaches. We refer the reader to the discussion in Durham and Gallant (2002) and Li, Pearson, and Poteshman (2004) on the efficacy of alternative approaches and Brandt and Santa-Clara (2002) on the difficulty in estimating models using simulation-based methods. With finer Euler discretization and reasonably lengthy MCMC sampler draws, our computation time also measures favorably relative to the Bayesian methodology of Jones (2003).

underlying process only at the expense of an implausible level of β_1. Negative and statistically significant $\alpha_1 = -8.0369$ indicates speedy mean-reversion in the variance process and a model long-run volatility level of $\sqrt{-\alpha_0/\alpha_1} = 19.77\%$.

Keeping β_3 free in the CEV class rectifies modeling deficiencies of the AFF specification (that forces $\beta_3 \equiv 1/2$) as seen through a large jump in the log-likelihood. Transitioning from AFF to CEV2 increases the log-likelihood from 11400.28 to 12090.40, thereby rejecting AFF with a significant log-likelihood ratio statistic. Specifically, we compute the log-likelihood ratio statistic as minus twice the difference between the log-likelihood values of the restricted and the unrestricted models:

$$\mathcal{L}^* \equiv -2 \times (\mathcal{L}[\theta_R] - \mathcal{L}[\theta_U]) \tag{47}$$

which is distributed $\chi^2 [\dim[\theta_U] - \dim[\theta_R]]$. That the data favors the CEV class of variance processes over the AFF is further validated through high $-n/2$ AIC. Misspecification of $\sigma[X]$ is primarily responsible for the rejection of the affine volatility models.

Analyzing the MLE results across CEV1, CEV2 and CEV4 provides a number of fundamental insights about the dynamics of market variance. First, the highly significant α_2 and α_3 suggest the presence of non-linearities in the drift of the variance process. Consider CEV4: the estimate of α_2 is -166.72 (standard error of 60.53) and the estimate of α_3 is 0.0031 (standard error of 0.0015). Compared to the linear drift model CEV2 - which has a log-likelihood of 12090.40 - adding two additional non-linear drift parameters raises the log-likelihood to 12094.24. The resulting log-likelihood ratio statistic $\mathcal{L}^*$ is 7.68, which is bigger than $\chi^2[2]$ critical value of 6.0 (at 95% confidence level) and the linear drift model is rejected in favor of a non-linear drift in the variance process.

Omitting a role for $\alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3/X_t$ also worsens the performance of the stochastic volatility models versus CEV4. The realized value of the test statistic $\mathcal{L}^*$ is 9.18 which can be compared to the $\chi^2[3]$ critical value of 7.82. However, the same cannot be argued about linear drift versus constant drift accounting for the constant elasticity of variance structure $\sigma[X_t] = \beta_2 X_t^{\beta_3}$. The log-likelihood is virtually insensitive to the addition of $\alpha_1 X_t$ to the CEV1 specification: the small increase of $\mathcal{L}$ from 12089.65 to 12090.40 is insufficiently large to make $\mathcal{L}^* = 1.5$ statistically significant with 1 degree of freedom.

The exponent parameter β_3 is statistically significant in CEV1, CEV2, and CEV4 and ranges between 1.2732 to 1.2781 and several fold relative to the reported standard errors. The magnitude of β_3 is comparable but slightly higher than that reported by Jones (2003). An overarching conclusion is that $\beta_3 > 1$ is needed to match the time-series properties of the VIX index with the

CEV models (the one-sided t-test rejects the null hypothesis of $\beta_3 \leq 1$).

Recall the GEN class of variance processes shares the same drift and volatility structure as models in the CEV class except that the GEN class embeds an additional linear-term in the $\sigma[X]$ specification. The implementation of GEN1, GEN2, and GEN4 brings out several incremental facts about the behavior of the volatility processes. Comparing the log-likelihood and model estimates across CEV1 and GEN1, CEV2 and GEN2, and CEV4 and GEN4 establishes that the addition of $\beta_1 X_t$ as in $\sigma[X_t] = \sqrt{\beta_1 X_t + \beta_2 X_t^{\beta_3}}$ provides additional flexibility in fitting VIX dynamics. In each comparison, the $\mathcal{L}^*$ statistic ranges between 6.86 and 9.66 which is highly significant given $\chi^2[1] = 3.84$. Regardless of the functional form of the drift specification, this result can be interpreted to mean that the shape of $\sigma[X] = \sqrt{\beta_1 X_t + \beta_2 X_t^{\beta_3}}$ is statistically more attractive than $\sigma[X] = \beta_2 X_t^{\beta_3}$ in reconciling the path of the VIX index. The estimate of β_1 lies between 0.0141 and 0.0168 with a standard error of 0.0050 or 0.0051, making β_1 significant across all models in the GEN class.

Since the estimated value of β_0 is zero in the GEN class, the volatility function of market variance approaches zero as the variance itself approaches zero. However, the significant positive value of β_1 indicates that the volatility of the index variance process approaches zero at a lower rate than that in the CEV specifications. At higher variance levels, the volatility function in the GEN specifications increases faster than that in the CEV specifications. So, the GEN specifications accommodate greater volatility at both low and high variance levels. For the GEN specifications, the linear term $\beta_1 X$ becomes more pronounced than the nonlinear term $\beta_2 X^{\beta_3}$ for $X < (\beta_1/\beta_2)^{\frac{1}{\beta_3 - 1}}$. In the region $X < 0.0136$ in GEN4, the linear term $\beta_1 X$ is more heavily weighted than the non-linear component $\beta_2 X^{\beta_3}$ and vice-versa. Therefore, β_1 determines the behavior of the volatility function at low variance levels.

Results from GEN1, GEN2, and GEN4 strengthen our earlier conclusions from CEV models that support the existence of non-linearities in the drift function. As seen, the market variance data prefers GEN4 over both GEN1 and GEN2, while GEN1 and GEN2 are indistinguishable based on the log-likelihoods and log-likelihood ratio statistics. More precisely, the $\mathcal{L}^*$ is 11.72 between GEN1 and GEN4 and 10.48 between GEN2 and GEN4. The estimated α_2 of -166.72 (-209.35) for the CEV4 (GEN4) model is intuitive and guarantees a negative drift function in periods of pronounced market volatility. In the reverse when the market variance is low, $\mu[X]$ should contain a positive drift which dictates the positivity of α_3 and assures that zero is unattainable. The coexistence of statistically significant $\alpha_2 < 0$ and $\alpha_3 > 0$ suggests that mean-reversion and non-linearity of the drift function are robust phenomena in index volatility markets.

Elasticity parameters obtained from CEV and GEN models exhibit a mutually coherent mapping: the β_3 of GEN is slightly more than twice the CEV counterpart. The following interpretation holds for estimated β_3 of GEN4. With $\beta_3 = 2.8822$, the variance function $\sigma^2[X]$ is convex in variance, that is, $\frac{\partial \sigma^2[X;\theta]}{\partial X} = \beta_1 + \beta_2 \beta_3 X^{\beta_3 - 1} > 0$ and $\frac{\partial^2 \sigma^2[X;\theta]}{\partial X^2} = \beta_2 \beta_3 (\beta_3 - 1) X^{\beta_3 - 2} > 0$. The one-sided t-test overwhelmingly rejects the null hypothesis that $\beta_3 \leq 2$. This property has the implication that a well-performing variance process should have its variance function increasing at a rate faster than the level of market variance. Continuous-time volatility models violating such a property are likely headed for inconsistencies in the empirical dimension.

Because the most general specification, GEN4, exemplifies nonlinearity in the drift and diffusion function of the type described in (26)-(27), it rank-orders first with -$n/2$ AIC of 12092.07, followed next by GEN1 and GEN2 with -$n/2$ AIC of 12089.21 and 12088.83. Our density-based estimation implies that the linear AFF model displays unsatisfactory goodness-of-fit statistics and has dynamics most inconsistent with the observed movements in the VIX index. While GEN4 admits a more complex representation, the nonlinearity parameters are vital to generating a more realistic time-evolution of market volatility.

Before closing this section, note that the correct specification of interest-rate drift and diffusion is at the root of a contentious literature. How do patterns of non-linearity compare between the market variance and the interest-rate? To appreciate the similarities/differences, we explore two angles. First, the shapes of $\mu[.]$ and $\sigma[.]$ are plotted in Figure 1 for the GEN4 model. In this exercise, we focus on the estimate of θ in our Table 2 and the corresponding daily interest-rate estimates from Durham (2003): $\alpha_0 = -3.3157$, $\alpha_1 = 0.7328$, $\alpha_2 = -0.0503$, $\alpha_3 = 6.2555$, $\beta_0 = 0.3667$, $\beta_1 = -0.0275$, $\beta_2 = 0.003$, and $\beta_3 = 3.3732$. The visual impression from subplots A and B suggests that the mean-reversion in the interest-rate is relatively weaker in the tails of the distribution. This point can be confirmed by computing $\frac{1}{|\mu[r]|}\frac{\partial \mu[r]}{\partial r} = \{-155, 25\}$ at $r = \{0.01, 0.15\}$, versus $\frac{1}{|\mu[X]|}\frac{\partial \mu[X]}{\partial X} = \{-285, 45\}$ at $X = \{0.08^2, 0.316^2\}$. Furthermore, the close correspondence between the shapes in subplot C and subplot D indicates that the exponent parameter, β_3, is instrumental to the volatility specification of both interest-rate and market variance. Second, we compare model selection results for volatility and interest-rate dynamics. While Table 3 in Durham (2003) reveals that models with constant drift realistically captures interest-rate movements, our Table 2 shows that incorporating non-linearities in $\mu[X]$ offers goodness-of-fit improvements.

4 Concluding statements and summary

Density approximation is an issue whenever the researcher needs the state-price density for pricing purposes or for constructing the likelihood function in maximum-likelihood estimation. Building on Aït-Sahalia (1999, 2002) we have provided a method to approximate the transition density, and studied the empirical implications of estimating a much wider class of equity volatility models.

Our theoretical contribution in Proposition 1 and Proposition 2 shows how to express the recursively defined expansion coefficients in terms of one-dimensional integrals. To enhance the appeal of the methodology, the density approximation is derived in terms of the drift and diffusion function of the original state variable. This is done without incurring burdensome integration steps to reduce the state variable dynamics to a unit-variance process. Proposition 3 provides a technical treatment of the case when the necessary integrals can be Taylor series approximated and results in a solution for the expansion coefficients and the transition density. We illustrate the power of the methodology by deriving the density approximation for the general continuous-time model presented in Aït-Sahalia (1996).

Novel to the literature on equity volatility, our empirical results substantiate variance dynamics with nonlinear mean-reversion. Strong statistical evidence exists to support the presence of a nonlinear diffusion coefficient structure: the variance of variance function is composed of a term linear in variance plus a power function term in variance with an exponent that demands a value greater than two. The combined continuous-time variance model with the said properties furnishes reasonable goodness-of-fit statistics and produces superior performance metrics relative to its nested variants.

Other than the most obvious outlets for assessing the usefulness of the density approximation, our methods can be adapted to learn about the parametric nature of phenomena of significance to economists (a partial list includes electricity prices, intraday power volatility, exchange rate volatility distributions, and credit spreads) and for undertaking risk measurements under nonlinear drift and diffusion forcing processes. The field of density approximation is now amenable to analytical characterizations under a large class of non-standard continuous-time models.

Appendix A: Proof of Proposition 2

To outline a proof of the proposition, consider the determination of $c_2[y]$:

$$\begin{aligned} c_2[y] &= \frac{2}{(y-y_0)^2}\int_{y_0}^{y}(w-y_0)\left(\lambda[w]\,c_1[w]+\frac{1}{2}\frac{\partial^2 c_1[w]}{\partial w^2}\right)dw \\ &= \frac{2}{(y-y_0)^2}\int_{y_0}^{y}\lambda[w]\left(\int_{y_0}^{w}\lambda[z]\,dz\right)dw+\frac{1}{(y-y_0)^2}\int_{y_0}^{y}(w-y_0)\frac{\partial^2 c_1[w]}{\partial w^2}\,dw. \end{aligned} \tag{48}$$

The first part of (48) is equivalent to $c_1^2[y]$ as proved below:

$$\begin{aligned} \frac{2}{(y-y_0)^2}\int_{y_0}^{y}\lambda[w]\left(\int_{y_0}^{w}\lambda[z]\,dz\right)dw &= \frac{2}{(y-y_0)^2}\int_{y_0}^{y}\left(\int_{y_0}^{w}\lambda[z]\,dz\right)d\left(\int_{y_0}^{w}\lambda[z]\,dz\right) \\ &= \frac{1}{(y-y_0)^2}\left(\int_{y_0}^{w}\lambda[z]\,dz\right)^2\Bigg|_{y_0}^{y} \\ &= c_1^2[y]. \end{aligned} \tag{49}$$

Continuing, the second component of $c_2[y]$ reduces, by a repeated application of integration by parts, to:

$$\int_{y_0}^{y}(w-y_0)\frac{\partial^2 c_1[w]}{\partial w^2}dw = (w-y_0)\frac{\partial c_1[w]}{\partial w}\Bigg|_{y_0}^{y} - c_1[w]|_{y_0}^{y}. \tag{50}$$

Care must be exercised in evaluating the functions at the lower limit y_0. For example, $(w - y_0)\partial c_1[w]/\partial w$ does not evaluate to zero at $w = y_0$. To this end we note

$$\frac{\partial c_1[w]}{\partial w} = \frac{\lambda[w]-c_1[w]}{w-y_0}. \tag{51}$$

Consequently it then follows that

$$\int_{y_0}^{y}(w-y_0)\frac{\partial^2 c_1[w]}{\partial w^2}\,dw = \lambda[y]-2c_1[y]-\lambda[y_0]+2c_1[y_0] = \lambda[y]-2c_1[y]+\lambda[y_0], \tag{52}$$

where l'Hôpital's rule implies $c_1[y_0] = \lambda[y_0]$. Therefore,

$$c_2[y] = c_1^2[y] + \frac{1}{(y-y_0)^2}(\lambda[y]+\lambda[y_0]-2c_1[y]). \tag{53}$$

Now we consider $c_3[y]$, which is recursively defined as

$$c_3[y] = \frac{3}{(y-y_0)^3}\int_{y_0}^{y}(w-y_0)^2\left(\lambda[w]\,c_2[w]+\frac{1}{2}\frac{\partial^2 c_2[w]}{\partial w^2}\right)dw. \tag{54}$$

For clarity consider each part of (54) separately. Using (53) and skipping intermediate steps, $\frac{3}{(y-y_0)^3}\int_{y_0}^{y}(w-y_0)^2\lambda[w]c_2[w]dw$ simplifies to:

$$\begin{aligned} &\frac{3}{(y-y_0)^3}\int_{y_0}^{y}(w-y_0)^2\lambda[w]\left(c_1^2[w]+\frac{\lambda[w]+\lambda[y_0]-2c_1[w]}{(w-y_0)^2}\right)dw \\ =\;& c_1^3[y]+\frac{3}{(y-y_0)^3}\int_{y_0}^{y}\lambda[w](\lambda[w]+\lambda[y_0]-2c_1[w])\,dw. \end{aligned} \tag{55}$$

The second part of (54) has an analytical representation using integration by parts:

$$\begin{aligned} &\frac{3}{(y-y_0)^3}\int_{y_0}^{y}(w-y_0)^2\left(\frac{1}{2}\frac{\partial^2 c_2[w]}{\partial w^2}\right)dw = \frac{3(w-y_0)^2}{2(y-y_0)^3}\frac{\partial c_2[w]}{\partial w}\bigg|_{y_0}^{y} - \\ &\frac{3(w-y_0)c_2[w]}{(y-y_0)^3}\bigg|_{y_0}^{y}+\frac{3}{(y-y_0)^3}\int_{y_0}^{y}c_2[w]dw. \end{aligned} \tag{56}$$

It is straightforward to show that

$$\frac{\partial c_2[w]}{\partial w} = -\frac{2c_2[w]}{w-y_0}+\frac{2}{w-y_0}\left(\lambda[w]c_1[w]+\frac{1}{2}\frac{\partial^2 c_1[w]}{\partial w^2}\right), \tag{57}$$

and it holds that

$$\frac{\partial^2 c_1[w]}{\partial w^2} = \frac{\lambda'[w]}{w-y_0}-\frac{2(\lambda[w]-c_1[w])}{(w-y_0)^2}. \tag{58}$$

By a basic application of l'Hôpital's rule at the lower limit y_0

$$\begin{aligned} &\frac{3}{(y-y_0)^3}\int_{y_0}^{y}(w-y_0)^2\left(\frac{1}{2}\frac{\partial^2 c_2[w]}{\partial w^2}\right)dw = \frac{3}{(y-y_0)^2}(\lambda[y]\,c_1[y]-2c_2[y]) - \\ &\frac{3(\lambda[y]-c_1[y])}{(y-y_0)^4}+\frac{3\lambda'[y]}{2(y-y_0)^3}+\frac{3}{(y-y_0)^3}\int_{y_0}^{y}c_2[w]dw. \end{aligned} \tag{59}$$

Combining (55) and (59) we finally have

$$\begin{aligned} c_3[y] \;=\;& c_1^3[y]+\frac{3}{(y-y_0)^2}(c_1[y](\lambda[y]+\lambda[y_0])-2c_2[y])+\frac{3}{(y-y_0)^3}\left(\int_{y_0}^{y}\lambda^2[w]\,dw+\right. \\ &\left.\int_{y_0}^{y}c_2[w]\,dw-\int_{y_0}^{y}2\lambda[w]c_1[w]dw+\frac{\lambda'[y]}{2}\right)-\frac{3(\lambda[y]-c_1[y]}{(y-y_0)^4}. \end{aligned} \tag{60}$$

Now

$$
\begin{aligned}
&\int_{y_0}^{y} c_2[w]dw = \int_{y_0}^{y} \frac{2}{(w-y_0)^2} \int_{y_0}^{w} (z-y_0)\left(\lambda[z]c_1[z] + \frac{1}{2}\frac{\partial^2 c_1[z]}{\partial z^2}\right) dz\, dw = \\
&\int_{y_0}^{y}\int_{y_0}^{y} \frac{2}{(w-y_0)^2}(z-y_0)\left(\lambda[z]c_1[z] + \frac{1}{2}\frac{\partial^2 c_1[z]}{\partial z^2}\right) 1(z<w)\, dz\, dw = \\
&2\int_{y_0}^{y} (z-y_0)\left(\lambda[z]c_1[z] + \frac{1}{2}\frac{\partial^2 c_1[z]}{\partial z^2}\right)\left(\frac{1}{z-y_0} - \frac{1}{y-y_0}\right) dz = \\
&2\int_{y_0}^{y} \left(\lambda[z]c_1[z] + \frac{1}{2}\frac{\partial^2 c_1[z]}{\partial z^2}\right) dz - (y-y_0)\, c_2[y] = \\
&2\int_{y_0}^{y} \lambda[z]\, c_1[z]dz + \frac{\partial c_1[z]}{\partial z}\Big|_{y_0}^{y} - (y-y_0)\, c_2[y] = \\
&2\int_{y_0}^{y} \lambda(z)\, c_1[z]dz + \frac{\lambda[y]-c_1[y]}{y-y_0} - \frac{1}{2}\lambda'[y_0] - (y-y_0)\, c_2[y]. \qquad (61)
\end{aligned}
$$

Plugging (61) into (60) and rearranging

$$
\begin{aligned}
c_3[y] \quad = \quad & c_1^3[y] + \frac{3}{(y-y_0)^2}(c_1[y](\lambda[y]+\lambda[y_0]) - 3\, c_2[y]) + \\
& \frac{3}{(y-y_0)^3}\left(\frac{\lambda'[y]-\lambda'[y_0]}{2} + \int_{y_0}^{y} \lambda^2[w]\, dw\right).
\end{aligned}
$$

The proof of $c_4[y|y_0;\theta]$ is rather unwieldy and omitted to save on space (available from the authors).

Appendix B: Expressions for $c_j[y|y_0;\theta]$ in Proposition 3

$$
\begin{aligned}
c_1[y|y_0;\theta] \quad &= \quad \lambda_0 + \frac{\lambda_1 D}{2} + \frac{\lambda_2 D^2}{6} + \frac{\lambda_3 D^3}{24} + \frac{\lambda_4 D^4}{120} + \frac{\lambda_5 D^5}{720} + \frac{\lambda_6 D^6}{5040}, \qquad (62) \\
c_2[y|y_0;\theta] \quad &= \quad c_1^2[y_0] + \frac{\lambda_2}{6} + \frac{\lambda_3 D}{12} + \frac{\lambda_4 D^2}{40} + \frac{\lambda_5 D^3}{180} + \frac{\lambda_6 D^4}{1008}, \qquad (63) \\
c_3[y|y_0;\theta] \quad &= \quad c_1^3[y_0] + \left(\frac{\lambda_1^2}{4} + \frac{\lambda_0\lambda_2}{2} + \frac{\lambda_4}{40}\right) + \left(\frac{\lambda_1\lambda_2}{2} + \frac{\lambda_0\lambda_3}{4} + \frac{\lambda_5}{80}\right) D + \\
&\left(\frac{3\lambda_2^2}{20} + \frac{\lambda_1\lambda_3}{5} + \frac{3\lambda_0\lambda_4}{40} + \frac{\lambda_6}{280}\right) D^2 + \left(\frac{5\lambda_2\lambda_3}{48} + \frac{13\lambda_1\lambda_4}{240} + \frac{\lambda_0\lambda_5}{60}\right) D^3 + \\
&\left(\frac{23\lambda_3^2}{1344} + \frac{11\lambda_2\lambda_4}{420} + \frac{19\lambda_1\lambda_5}{1680} + \frac{\lambda_0\lambda_6}{336}\right) D^4 + \left(\frac{\lambda_3\lambda_4}{120} + \frac{\lambda_2\lambda_5}{192} + \frac{13\lambda_1\lambda_5}{6720}\right) D^5 \\
&+ \left(\frac{43\,\lambda_4^2}{4320} + \frac{7\lambda_3\lambda_5}{4320} + \frac{13\lambda_2\lambda_6}{15120}\right) D^6, \qquad (64) \\
c_4[y|y_0;\theta] \quad &= \quad c_1^4[y_0] + \left(\lambda_0\lambda_1^2 + \lambda_0^2\lambda_2 + \frac{3\lambda_2^2}{20} + \frac{\lambda_1\lambda_3}{5} + \frac{\lambda_0\lambda_4}{10} + \frac{\lambda_6}{280}\right) +
\end{aligned}
$$

$$\left(\frac{\lambda_3^2}{2}+2\lambda_0\lambda_1\lambda_2+\frac{\lambda_0^2\lambda_3}{2}+\frac{\lambda_2\lambda_3}{4}+\frac{3\lambda_1\lambda_4}{20}+\frac{\lambda_0\lambda_5}{20}\right)D+$$
$$\left(\frac{11\lambda_1^2\lambda_2}{12}+\frac{3\lambda_0\lambda_2^2}{5}+\frac{4\lambda_0\lambda_1\lambda_3}{5}+\frac{23\lambda_3^2}{336}+\frac{19\lambda_2\lambda_4}{168}+\frac{3\lambda_1\lambda_5}{56}+\frac{\lambda_0\lambda_6}{70}\right)D^2+$$
$$\left(\frac{7\lambda_1\lambda_2^2}{15}+\frac{19\lambda_1^2\lambda_3}{60}+\frac{5\lambda_0\lambda_2\lambda_3}{12}+\frac{13\lambda_0\lambda_1\lambda_4}{60}+\frac{\lambda_3\lambda_4}{20}+\frac{\lambda_0^5\lambda_5}{30}+\frac{\lambda_2\lambda_5}{30}+\right.$$
$$\left.\frac{11\lambda_1\lambda_6}{840}\right)D^3+\left(\frac{13\lambda_2^3}{180}+\frac{3\lambda_1\lambda_2\lambda_3}{10}+\frac{5\lambda_0\lambda_2\lambda_3}{12}+\frac{13\lambda_0\lambda_1\lambda_4}{60}+\frac{23\lambda_0\lambda_3^2}{336}+\right.$$
$$\frac{19\lambda_2^2\lambda_4}{420}+\frac{11\lambda_0\lambda_2\lambda_4}{105}+\frac{11\lambda_1\lambda_6}{840}+\frac{23\lambda_4^2}{2880}+\frac{\lambda_0^2\lambda_6}{168}+\frac{19\lambda_0\lambda_1\lambda_5}{420}+\frac{19\lambda_3\lambda_5}{1440}+$$
$$\left.\frac{37\lambda_2\lambda_6}{5040}\right)D^4+\left(\frac{\lambda_2^2\lambda_3}{15}+\frac{157\lambda_1\lambda_3^2}{3360}+\frac{181\lambda_1\lambda_2\lambda_4}{2520}+\frac{\lambda_0\lambda_3\lambda_4}{30}+\frac{79\lambda_1^2\lambda_5}{5040}+\right.$$
$$\left.\frac{\lambda_0\lambda_2\lambda_5}{48}+\frac{7\lambda_4\lambda_5}{1800}+\frac{13\lambda_0\lambda_1\lambda_6}{1680}+\frac{137\lambda_3\lambda_6}{50400}\right)D^5+\left(\frac{9\lambda_2\lambda_3^2}{448}+\frac{391\lambda_2^2\lambda_4}{25200}+\right.$$
$$\frac{79\lambda_1\lambda_3\lambda_4}{3600}+\frac{43\lambda_0\lambda_4^2}{10800}+\frac{139\lambda_1\lambda_2\lambda_5}{10800}+\frac{7\lambda_0\lambda_3\lambda_5}{1080}+\frac{71\lambda_5^2}{158400}+$$
$$\left.\frac{13\lambda_0\lambda_2\lambda_6}{3780}+\frac{281\lambda_4\lambda_6}{369600}\right)D^6. \qquad (65)$$

In terms of $\mu[X_0]$ and $\sigma[X_0]$, each required λ_i can be recursively derived as:

$$\lambda_0 = -(f^2[X_0]+f'[X_0]\,\sigma[X_0])/2, \qquad (66)$$
$$\lambda_i = \lambda'_{i-1}\,\sigma[X_0], \qquad i=1,2,3,4,5,6, \qquad (67)$$

which completes the characterization of the density approximation.

Appendix C: Expressions for λ_j in the GEN4 Model

To this end, let

$$S_0=\sigma[X_0]=\sqrt{\beta_0+\beta_1X_0+\beta_2X_0^{\beta_3}}. \qquad (68)$$

Then $S_0^2=\beta_0+\beta_1X_0+\beta_2X_0^{\beta_3}=V_0$. Differentiation leads to the following derivatives of S_0 with respect to X_0:

$$S_1 = V_1/(2S_0), \qquad (69)$$
$$S_2 = (V_2-2S_1^2)/(2S_0), \qquad (70)$$
$$S_3 = (V_3-6S_1S_2)/(2S_0), \qquad (71)$$

$$S_4 = (V_4 - 8S_1S_3 - 6S_2^2)/(2S_0), \quad (72)$$
$$S_5 = (V_5 - 10S_1S_4 - 20S_2S_3)/(2S_0), \quad (73)$$
$$S_6 = (V_6 - 12S_1S_5 - 30S_2S_4 - 20S_3^2)/(2S_0), \quad (74)$$
$$S_7 = (V_7 - 14S_1S_6 - 42S_2S_5 - 70S_3S_4)/(2S_0), \quad (75)$$

where $V_7 = (\beta_3 - 6)V_6/X_0$.

Let $F_0 = f[X] = \mu[X_0]/\sigma[X_0] - \sigma'[X_0]/2 = U_0/S_0$. So $S_0F_0 = U_0$. Obeying the above successive differentiation rules we arrive at:

$$F_0 = U_0/S_0, \quad (76)$$
$$F_1 = (U_1 - F_0S_1)/S_0, \quad (77)$$
$$F_2 = (U_2 - 2F_1S_1 - F_0S_2)/S_0, \quad (78)$$
$$F_3 = (U_3 - 3F_2S_1 - 3F_1S_2 - F_0S_3)/S_0, \quad (79)$$
$$F_4 = (U_4 - 4F_3S_1 - 6F_2S_2 - 4F_1S_3 - F_0S_4)/S_0, \quad (80)$$
$$F_5 = (U_5 - 5F_4S_1 - 10F_3S_2 - 10F_2S_3 - 5F_1S_4 - F_0S_5)/S_0, \quad (81)$$
$$F_6 = (U_6 - 6F_5S_1 - 15F_4S_2 - 20F_3S_3 - 15F_2S_4 - 6F_1S_5 - F_0S_6)/S_0, \quad (82)$$
$$F_7 = (U_7 - 7F_6S_1 - 21F_5S_2 - 35F_4S_3 - 35F_3S_4 - 21F_2S_5 - 7F_1S_6 - F_0S_7)/S_0, \quad (83)$$

where $U_6 = 720\alpha_3/X_0^7 - V_7/4$, $U_7 = -5040\alpha_3/X_0^8 - V_8/4$ and $V_8 = (\beta_3 - 7)V_7/X_0$.

We can now conveniently write

$$H_0 = -(f^2[X_0] + f'[X_0]\sigma[X_0])/2 = -(F_0^2 + F_1S_0)/2. \quad (84)$$

The first six derivatives of H_0 with respect to X_0 are:

$$H_1 = -(2F_0F_1 + F_2S_0 + F_1S_1)/2, \quad (85)$$
$$H_2 = -(2F_1^2 + 2F_0F_2 + 2F_2S_1 + F_3S_0 + F_1S_2)/2, \quad (86)$$
$$H_3 = -(6F_1F_2 + 2F_0F_3 + F_4S_0 + 3F_3S_1 + 3F_2S_2 + F_1S_3)/2, \quad (87)$$
$$H_4 = -(6F_2^2 + 8F_1F_3 + 2F_0F_4 + F_5S_0 + 4F_4S_1 + 6F_3S_2 + 4F_2S_3 + F_1S_4)/2, \quad (88)$$
$$H_5 = -(20F_2F_3 + 10F_1F_4 + 2F_0F_5 + F_6S0 + 5F_5S_1 + 10F_4S_2 + 10F_3S_3 + 5F_2S4 + F_1S_5)/2, \quad (89)$$
$$H_6 = -(20F_3^2 + 30F_2F_4 + 12F_1F_5 + 2F_0F_6 + F_7S0 + 6F_6S_1 + 15F_5S_2 + 20F_4S_3 + 15F_3S_4 + 6F_2S_5 + F_1S_6)/2. \quad (90)$$

References

Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9, 385-426.

Aït-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance* 54, 1361-1395.

Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusion: a closed-form approximation approach. *Econometrica* 70, 223-262.

Aït-Sahalia, Y., 2003. Closed-form likelihood expansions for multivariate diffusions. *Unpublished working paper.* Princeton University.

Aït-Sahalia, Y., and Kimmel, R., 2003. Estimating affine multifactor term structure models using closed-form likelihood expansions. *Unpublished working paper.* Princeton University.

Andersen, T., L. Benzoni, and Lund, J., 2002. An empirical investigation of continuous-time equity return models. *Journal of Finance* 57, 1239-1284.

Andersen, T., T. Bollerslev, F. Diebold, and Ebens, H., 2001. Modeling and forecasting realized volatility. *Econometrica* 71, 579-625.

Andersen, T., T. Bollerslev, and Diebold, F., 2003. Some like it smooth, and some like it rough: untangling continuous and jump components in measuring, modeling, and forecasting asset return volatility. *Unpublished working paper*, University of Pennsylvania # 03-025.

Bakshi, G., C. Cao, and Chen, Z., 2000. Pricing and hedging long-term options. *Journal of Econometrics* 94, 277-318.

Bakshi, G., and Ju, N., 2002. A refinement to Aït-Sahalia's maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Journal of Business* (forthcoming).

Bakshi, G., and Madan, D., 2000. Spanning and derivative security valuation. *Journal of Financial Economics* 55, No. 2, 205-238.

Bates, D., 2000. Post-87 crash fears in S&P 500 futures options. *Journal of Econometrics* 94, 181-238.

Brandt, M. and Santa-Clara, P., 2002, Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics* 63, 161-210.

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

Durham, G., 2003. Likelihood-based specification analysis of continuous-time models of short-term interest rate. *Journal of Financial Economics* 70, 463-487.

Durham, G. and Gallant, R., 2002. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business and Economic Statistics* 20, 297-316.

Elerian, D., S. Chib, and Shephard, N., 2001. Likelihood inference for discretely observed non-linear diffusions. *Econometrica* 69, 959-993.

Eraker, B., M. Johannes, and Polson, N., 2003. The impact of jumps in volatility and returns. *Journal of Finance* 58, No. 3, 1269-1300.

Gallant, R., and Tauchen, G., 1997. Estimation of continuous-time models for stock returns and interest rates. *Macroeconomics Dynamics* 1, 135-168.

Heston, S., 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.

Li, M., Pearson, N., and Poteshman, A., 2004. Conditional estimation of diffusion processes. *Journal of Financial Economics* 74 (1), 31-66.

Jones, C., 2003. The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics* 116, 181-224.

Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics* 63, No. 1, 3-50.

TABLE 1: Maximum Absolute Errors of the Density Approximation

Maximum absolute errors of the approximations are based on $K = 1, 2, 3, 4, 5$ and the Euler approximation. The approximate density is based on (20) with $c_j[\gamma[x]|\gamma[x_0];\theta]$ presented in Proposition 2. Entries corresponding to $\max(p|x_0)$ are the maximum conditional density. Computations involving Panels A and B use $\alpha_0 = 0.145 \times 0.0732$, $\alpha_1 = -0.145$, $\beta_0 = 0$, and $\beta_1 = 0.06521^2$. For the calculations performed in Panels C and D the initial stock price is fixed at $x_0 = \$50$, the initial volatility level is $\beta_2\, x_0^{\beta_3 - 1} = 0.3$, and α_1 and β_2 are allowed to vary. Tabulating the results by changing the elasticity of volatility coefficient β_3 provided the same conclusions and, therefore, omitted.

Panel A: $dX_t = (\alpha_0 + \alpha_1 X_t)\, dt + \sqrt{\beta_1 X_t}\, dW_t$, $t = 1/12$

x_0	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18
max(p|x_0) (10^1)	15.0	10.7	8.71	7.55	6.75	6.17	5.71	5.34	5.04
$K = 1$ (10^{-3})	6.46	1.32	0.67	0.36	0.18	0.32	0.15	1.13	2.70
$K = 2$ (10^{-6})	143	10.3	3.90	0.98	0.82	2.71	3.86	2.54	3.96
$K = 3$ (10^{-8})	89.7	4.11	1.34	0.26	0.31	0.15	1.36	2.83	3.26
$K = 4$ (10^{-10})	90.9	5.35	0.36	0.14	0.26	0.33	0.21	0.75	2.06
$K = 5$ (10^{-12})	153	8.27	6.57	5.95	8.55	9.82	14.9	5.94	5.98
Euler	7.17	3.73	2.55	1.93	1.59	1.35	1.17	1.04	0.94

Panel B: $dX_t = (\alpha_0 + \alpha_1 X_t) dt + \sqrt{\beta_1 X_t} dW_t$, $t = 1$

x_0	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18
max(p|x_0) (10^1)	4.39	3.25	2.69	2.35	2.11	1.94	1.80	1.68	1.59
$K = 1$ (10^{-2})	28.7	7.19	2.61	1.78	0.81	1.35	1.19	4.83	11.1
$K = 2$ (10^{-3})	79.8	4.96	2.23	0.71	0.50	1.35	2.00	1.83	2.45
$K = 3$ (10^{-4})	51.1	4.45	0.78	0.29	0.17	0.20	0.83	1.71	2.25
$K = 4$ (10^{-6})	117	41.7	5.1	0.67	1.44	2.27	1.81	5.38	14.4
$K = 5$ (10^{-7})	154	9.85	4.23	0.61	0.26	0.70	1.70	2.00	3.43
Euler	8.21	4.70	3.42	2.76	2.35	2.07	1.87	1.72	1.60

Panel C: $dX_t = \alpha_1 X_t + \beta_2 X_t^{\beta_3}\, dW_t$, $t = 1/12$

α_1	0.04	0.04	0.04	0.06	0.06	0.06	0.08	0.08	0.08
β_2	0.50	0.70	0.90	0.50	0.70	0.90	0.50	0.70	0.90
max(p|x_0) (10^{-2})	9.20	9.22	9.24	9.19	9.20	9.22	9.18	9.19	9.21
$K = 1$ (10^{-8})	1.48	0.55	0.11	15.05	6.26	1.12	47.50	22.69	6.24
$K = 2$ (10^{-11})	4.26	0.17	0.09	3.41	18.33	0.04	19.28	6.56	1.49
$K = 3$ (10^{-13})	1.56	0.32	2.83	2.64	0.36	2.31	7.10	1.24	2.02
$K = 4$ (10^{-14})	2.23	3.22	28.28	1.50	2.12	2.31	1.37	3.43	20.42
$K = 5$ (10^{-14})	2.28	3.22	28.28	1.54	2.12	2.31	1.34	3.42	20.42
Euler (10^{-3})	2.77	3.93	5.11	2.73	3.88	5.05	2.69	3.83	5.00

Panel D: $dX_t = \alpha_1 X_t + \beta_2 X_t^{\beta_3}\, dW_t$, $t = 1$

α_1	0.04	0.04	0.04	0.06	0.06	0.06	0.08	0.08	0.08
β_2	0.50	0.70	0.90	0.50	0.70	0.90	0.50	0.70	0.90
max(p|x_0) (10^{-2})	2.63	2.68	2.75	2.59	2.63	2.70	2.55	2.58	2.64
$K = 1$ (10^{-7})	8.19	2.51	0.49	65.17	26.24	4.66	206.58	95.23	25.89
$K = 2$ (10^{-9})	28.51	0.85	0.32	19.05	1.51	0.16	125.72	37.89	7.80
$K = 3$ (10^{-10})	13.20	0.05	0.00	17.29	1.46	0.03	43.33	6.92	0.11
$K = 4$ (10^{-12})	63.43	0.24	0.00	55.15	6.52	0.01	49.19	8.88	0.19
$K = 5$ (10^{-14})	456.61	0.57	0.39	0.04	1.16	0.52	325.76	2.23	0.37
Euler (10^{-3})	2.82	4.14	5.53	2.68	3.94	5.30	2.54	3.76	5.07

TABLE 2: Estimation Results for Market Index Variance

The encompassing model, **GEN4**, is $dX_t = \left(\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1}\right) dt + \sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}\, dW_t$. Nesting CEV1 and CEV2, the **CEV4** model is $dX_t = \left(\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1}\right) dt + \beta_2 X_t^{\beta_3}\, dW_t$. The **AFF** model is $dX_t = (\alpha_0 + \alpha_1 X_t)\, dt + \sqrt{\beta_0 + \beta_1 X_t}\, dW_t$. Throughout $\beta_0 \equiv 0$ as the estimate of β_0 was zero and has no impact on log-likelihood. The market index volatility is proxied by the daily VIX index. The daily data is sampled over the period of July 1, 1988 to January 10, 2000 (2907 observations). We take $X_t = \mathrm{VIX}_t^2$ and the VIX series is scaled by 100 to convert it into a decimal. The approximate density is analytical and displayed in (20) with $c_j[\gamma[x]|\gamma[x_0];\theta]$ presented in Proposition 2. Reported volatility model parameters and standard errors (in parenthesis) are based on maximizing the log-likelihood $\mathcal{L}[\theta] \equiv \sum_{i=1}^{n} \log\left\{p_X[\Delta, X_{i\Delta}|X_{(i-1)\Delta};\theta]\right\}$. The Akaike Information Criterion (AIC) is computed as $-2/n\,(\mathcal{L}[\theta] - \dim[\theta])$. Thus, a more properly specified model has higher $-\frac{n}{2}$AIC. Likelihood ratio test statistic for comparing nested models is $\mathcal{L}^* \equiv -2 \times (\mathcal{L}[\theta_R] - \mathcal{L}[\theta_U]) \sim \chi^2[df]$, where df is the number of exclusion restrictions and the 95% criterion values are

df	1	2	3	4
$\chi^2[df]$	3.84	6.0	7.82	9.50

. $\dim[\theta]$ is reported in curly brackets below $\mathcal{L}$.

Model	$\mathcal{L}$	$-\frac{n}{2}$ AIC	α_0	α_1	α_2	α_3	β_1	β_2	β_3
AFF	11400.28	11397.28	0.3141	-8.0369			0.1827		
	{3}		(0.0397)	(1.1977)			(0.0000)		
CEV1	12089.65	12086.65	0.0664					4.7825	1.2781
	{3}		(0.0114)					(0.1365)	(0.0214)
CEV2	12090.40	12086.40	0.0941	-1.4607				4.7046	1.2732
	{4}		(0.0255)	(1.1939)				(0.1365)	(0.0217)
CEV4	12094.24	12088.24	-0.3400	15.2476	-166.7249	0.0031		4.7645	1.2766
	{6}		(0.1846)	(6.3934)	(60.5296)	(0.0015)		(0.1381)	(0.0218)
GEN1	12093.21	12089.21	0.0654				0.0142	47.0436	2.8302
	{4}		(0.0117)				(0.0050)	(10.4937)	(0.1183)
GEN2	12093.83	12088.83	0.0920	-1.3540			0.0141	44.9773	2.8161
	{5}		(0.0267)	(1.2159)			(0.0051)	(10.0673)	(0.1187)
GEN4	12099.07	12092.07	-0.5537	21.3224	-209.3476	0.0051	0.0168	53.9726	2.8822
	{7}		(0.2162)	(7.3014)	(68.6506)	(0.0018)	(0.0050)	(12.4970)	(0.1226)

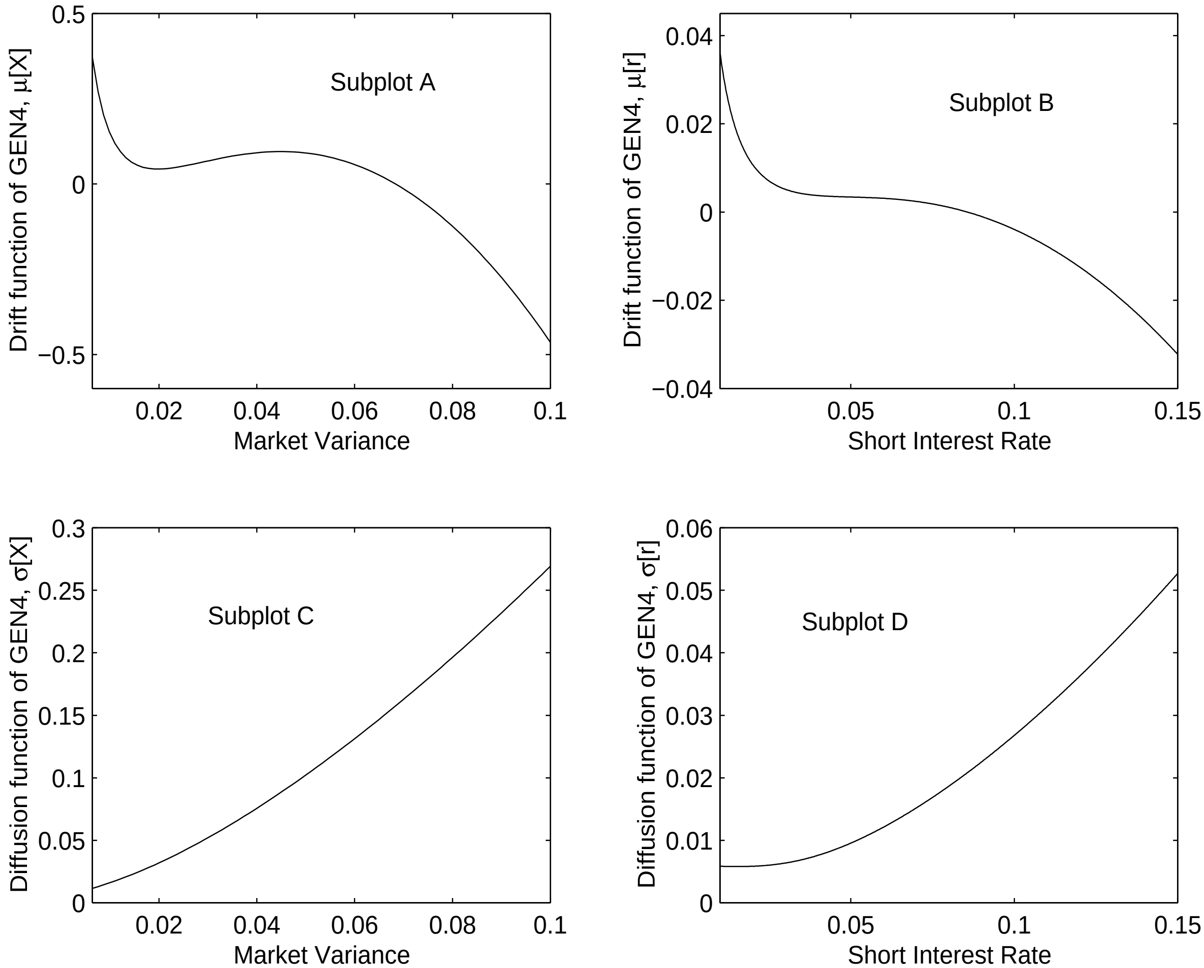

Figure 1: **Drift and diffusion functions for market variance and short interest rate for the GEN4 model.** Subplots A and B plot the drift function, $\mu[.]$, of market variance and interest-rate, respectively. We plot the diffusion function, $\sigma[.]$ in Subplots C and D. All calculations for the market variance process are based on the maximum-likelihood estimation results reported in Table 2. The daily parameter estimates for the interest-rate process are taken from the simulated maximum-likelihood approach of Durham (2003): $\alpha_0 = -3.3157$, $\alpha_1 = 0.7328$, $\alpha_2 = -0.0503$, $\alpha_3 = 6.2555$, $\beta_0 = 0.3667$, $\beta_1 = -0.0275$, $\beta_2 = 0.003$, and $\beta_3 = 3.3732$. The plotted $\mu[r]$ and $\sigma[r]$ are scaled by 100 to maintain consistency with the interest-rate series, which is in decimal form.