Available online at www.sciencedirect.com
ScienceDirect
Journal of Financial Economics 82 (2006) 227–249
JOURNAL OF Financial ECONOMICS
www.elsevier.com/locate/jfec

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

Gurdip Bakshi[a], Nengjiu Ju[b,*], Hui Ou-Yang[c]

[a]*Department of Finance, Smith School of Business, University of Maryland, College Park, MD 20742, USA*
[b]*School of Business and Management, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong*
[c]*Fuqua School of Business, Duke University, Durham, NC 27708, USA*

Received 18 August 2004; received in revised form 17 August 2005; accepted 23 September 2005
Available online 9 June 2006

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 on 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 rank-ordering of continuous-time stochastic volatility models, the absence or presence of nonlinearities in the drift function, and the desirability of pursuing more flexible diffusion function specifications.

JEL classification: G11; G12; G13; C15; C32; C52

Keywords: Continuous-time models; Maximum-likelihood estimation; Density approximation; Equity volatility; Market volatility dynamics

☆We are grateful for the feedback of the referee whose detailed comments have substantially improved this paper. We acknowledge useful comments by Yacine Aït-Sahalia, Doron Avramov, Charles Cao, Peter Carr, Ren-Raw Chen, Bob Dittmar, Steve Heston, Gautam Kaul, Mark Loewenstein, Dilip Madan, Bertrand Melenberg, Jun Pan, Matt Pritsker, Bill Schwert, Tyler Shumway, Guojun Wu, and Liuren Wu. Seminar participants at Michigan Business School, Rutgers University, 2004 European Finance Association meetings (Maastricht), and 2005 China International Conference in Finance provided several constructive suggestions. Computer codes used to implement the continuous-time models are available from the authors.
*Corresponding author. Tel.: +852 2358 8318; fax: +852 2358 1749.
E-mail address: nengjiu@ust.hk (N. Ju).

0304-405X/$ - see front matter
doi:10.1016/j.jfineco.2005.09.005

1. Introduction

Whether the end-goal is martingale pricing or maximum-likelihood estimation, theory invariably requires the knowledge of the transition density of the economic 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) proposes a method to approximate the transition density in a one-dimensional diffusion setting. Given its utility to the researcher in applied and theoretical work, the purpose of our 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. Based on 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 the rank-ordering of continuous-time volatility models, the presence of nonlinearities in the drift function, and 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 $\mathrm{d}X_t = \mu[X_t]\,\mathrm{d}t + \sigma[X_t]\,\mathrm{d}W_t$, for economic-variable X_t, extant density approximations hinge on transforming X_t to a unit-variance process via $\int^X \mathrm{d}u/\sigma[u]$ and then on inverting $\int^X \mathrm{d}u/\sigma[u]$. This requirement has proved analytically challenging for some economic models (see Bakshi and Ju, 2005). While developing the likelihood function of arbitrary multivariate diffusions, Aït-Sahalia (2003) constructs a Taylor series solution of the expansion coefficients. Such a procedure is capable of delivering a closed-form density approximation even when the multivariate diffusion is not reducible. A potential trade-off exists between the fully closed-form irreducible method and our approach. Second, in the enhanced-method of Aït-Sahalia (1999, 2002), the recursively defined coefficients that fulfill the forward and backward equation have a multidimensional integral dependence and are seldom tractable outside of the constant elasticity of variance diffusion class. The framework of our 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 $\int^X \mathrm{d}u/\sigma[u]$. 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. Despite the flurry of recent modeling efforts (see Andersen et al., 2003; Heston, 1993; Jones, 2003) no consensus has been reached on the dynamic evolution of equity 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 a reversal in the drift function at both high and low ends of the equity 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 stochastic volatility models (i.e., Heston, 1993) can be attributed to misspecified drift and diffusion coefficients. Overall, an equity market variance specification with nonlinear drift and diffusion function delivers the most desirable goodness-of-fit statistics, and this empirical result has wide-ranging implications for pricing and trading of risks associated with equity and volatility derivatives.

The rest of the paper proceeds as follows. Section 2 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 3 presents the density approximation for the encompassing model of Aït-Sahalia (1996). Section 4 is devoted to empirically evaluating continuous-time models of equity volatility. Section 5 summarizes our contributions and concludes.

2. Maximum-likelihood estimation of continuous-time models

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

$$\mathrm{d}X_t = \mu[X_t;\theta]\,\mathrm{d}t + \sigma[X_t;\theta]\,\mathrm{d}W_t, \tag{1}$$

where $\mu[X_t;\theta]$ and $\sigma[X_t;\theta]$ are, respectively, the coefficients of drift and diffusion, and θ represents the unknown parameter vector in an open-bounded set $\Theta \subset R^d$. 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) develops two analytical density approximations. Of particular interest are the enhanced formulae in Aït-Sahalia (1999, 2002), which correspond to the limit in which 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 $\sigma[X_t;\theta]$.

2.1. Enhanced method in Aït-Sahalia (1999, 2002)

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

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

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

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

Denoting $\phi[z] \equiv \mathrm{e}^{-z^2/2}/\sqrt{2\pi}$ 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 Kth term by

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

with $c_0[y|y_0,\theta] \equiv 1$. For $j \geqslant 1$, the recursive coefficients, $c_j[y|y_0;\theta]$, can be derived by solving

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

where $\lambda[y;\theta]$ is defined in terms of the drift function of $\mathrm{d}Y_t$ as

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

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

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

Several aspects of this methodology need discussion. For the expansion in Eq. (4) to converge, the X_t process in Eq. (1) must first be transformed to be sufficiently Gaussian. Based on the theoretical models adopted by Aït-Sahalia (1999), 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) and Durham (2003) point 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 (2005) relax the requirement that both Y_t and $\gamma^{-1}[y;\theta]$ be known analytically, and they 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 (2005) 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 Eq. (4) is superior if it can be transformed to make it apply to an equally wide class. The accuracy tests in Bakshi and Ju (2005) indicate that the approximation based on Eq. (4) is accurate with expansion coefficients as few as three and is substantially more reliable. The Hermite approach relies on expanding the density of Y_t around a standard normal, while the expansion in Eq. (4) forces the density function to satisfy the Kolmogorov forward and backward equations to the order Δ^K, resulting in greater accuracy.

The second feature of the methodology concerns the determination of the recursively defined $c_j[y|y_0;\theta]$. From the form of Eq. (5) it could 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. For example, Aït-Sahalia (1999) has solved the density function for models satisfying

$\mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1})\,\mathrm{d}t + \sigma X_t^{\rho}\,\mathrm{d}W_t$, 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 $\sigma[X_t;\theta]$. Suppose $\mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1})\,\mathrm{d}t + \sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}\,\mathrm{d}W_t$. For this particular model class the multidimensional integrals embedded in $c_j[y|y_0;\theta]$ remain unsolved, thereby eluding closed-form representations for $\{c_1[y|y_0;\theta], c_2[y|y_0;\theta], c_3[y|y_0;\theta], \ldots, c_K[y|y_0;\theta]\}$. The lack of analyticity of $c_j[y|y_0;\theta]$ is problematic as it has impaired the density-based inference of models with arbitrary $\mu[X;\theta]$ and $\sigma[X;\theta]$.

Citing this reason, Durham and Gallant (2002) and Brandt and Santa-Clara (2002) argue 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.

2.2. Central elements of the modification

The proposed approximation method 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 $\int^{X_t} \mathrm{d}u/\sigma[u;\theta]$ in practical applications. Our analytical contributions also ensure that each $c_j[y|y_0;\theta]$ contains at most one-dimensional integrals, not a set of complex multidimensional integrals.

2.2.1. 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 $\lambda[Y]$. For this family of continuous-time models, Eqs. (4), (5), and (7) constitute a conceptually simple and accurate approximation method. However, when Y_t or $\gamma^{-1}[Y;\theta]$ or both 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. *Let $\sigma'[X] \equiv \partial\sigma[X]/\partial X$ and define the function*

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

by analogy with Eq. (3). *The following density approximation components can be obtained in terms of $\mu[X]$ and $\sigma[X]$ of the original process X_t (suppressing the dependence on θ):*

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

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

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

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

and,

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

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

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

Proof. Make the change of variable $x=\gamma^{-1}[y]$. Based on the definition of $\gamma[X]$ given in Eq. (2), $\mathrm{d}y=\mathrm{d}x/\sigma[x]$. From the definition of $\lambda[y]$ in Eq. (6) and the new variable x comes Eq. (9). The chain rule of differentiation implies $\lambda'[y]=\partial\lambda[y]/\partial x\,\partial x/\partial y$ in Eq. (10). Similarly, Eq. (11) results. Noting that

$$\begin{aligned}\mu_Y[y] &= \mu[\gamma^{-1}[y]]/\sigma[\gamma^{-1}[y]] - \frac{\partial\sigma[\gamma^{-1}[y]]}{\partial x}\bigg/2\\ &= \mu[x]/\sigma[x]-\sigma'[x]/2\end{aligned}$$

and using $\mathrm{d}y=\mathrm{d}x/\sigma[x]$, we have the expression in Eq. (13). Using Eq. (9) and $\mathrm{d}y=\mathrm{d}x/\sigma[x]$ leads to Eqs. (14) and (15).

2.2.2. 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 Eq. (5) and $\lambda[y;\theta]$ defined in Eq. (6), the higher-order $c_j[y|y_0;\theta]$ is derived analytically with only one-dimensional integral dependence (suppressing the y_0 and θ arguments):*

$$c_1[y]=\frac{1}{y-y_0}\int_{y_0}^{y}\lambda[w]\,\mathrm{d}w, \tag{16}$$

$$c_2[y]=c_1^2[y]+\frac{1}{(y-y_0)^2}(\lambda[y]+\lambda[y_0]-2c_1[y]), \tag{17}$$

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

and

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

Proof. See Appendix A.

$c_1[y]$ involves computing two simple integrals: $y - y_0 = \int_{x_0}^{x} \mathrm{d}u/\sigma[u]$ and $\int_{y_0}^{y} \lambda[w]\,\mathrm{d}w$. Once $c_1[y]$ is obtained, $c_2[y]$ follows immediately and $c_3[y]$ merely requires $\int_{y_0}^{y} \lambda^2[w]\,\mathrm{d}w$. 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. Thus Proposition 2 achieves the crucial task of reducing the recursively defined multidimensional integrals in $c_j[y]$ 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

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

where $\{c_1[y|y_0;\theta], \ldots, c_4[y|y_0;\theta]\}$ are presented in Eqs. (16)–(19). Although some integrals still remain to be determined in our formulation in Eqs. (13)–(15) they are solely required for their numerical values. Compared with the recursively defined multidimensional integrals, the simplified $c_j[y]$'s are easier to evaluate and this connection is 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: $\log(p_X^{(K)}[\Delta, x|x_0;\theta]) \approx -\log(2\pi\sigma^2[X;\theta]\Delta)/2 - (\int_{x_0}^{x} \mathrm{d}u/\sigma[u])^2/(2\Delta) + \int_{x_0}^{x} f[u]\mathrm{d}u/\sigma[u] + \sum_{k=0}^{4}\mathscr{C}_k[\gamma[x]|\gamma[x_0];\theta]\Delta^k/k!$, where $\mathscr{C}_1 \equiv c_1$, $\mathscr{C}_2 \equiv c_2 - c_1^2$, $\mathscr{C}_3 \equiv c_3 - 3c_2c_1 + 2c_1^3$, and $\mathscr{C}_4 \equiv c_4 - 4c_3c_1 - 3c_2^2 + 12c_2c_1^2 - 6c_1^4$.

Aït-Sahalia (2003) extends the density approximation method in Aït-Sahalia (1999, 2002) to higher-dimensional diffusion processes. The problem of determining $c_j[y|y_0;\theta]$ is substantially harder in the multivariate setting, and we have been unable to work through the multivariate counterparts of Propositions 1 and 2. We direct the reader to Aït-Sahalia and Kimmel (2003) for closed-form likelihood expansions under affine multifactor models of the term structure of interest rates.

2.2.3. Density approximation when $x - x_0$ or $y - y_0$ is small or both

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 Taylor series approximation of the relevant integrals and renders the method analytical.

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

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

$$\varphi_i \equiv \partial^i(f[X_0]/\sigma[X_0])/\partial X_0^i, \text{ and} \tag{22}$$

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

When $x - x_0$, equivalently $y - y_0$, is small, the following Taylor expansions of the integrals $\int_{x_0}^{x} \mathrm{d}u/\sigma[u]$ and $\int_{x_0}^{x} f[u]\mathrm{d}u/\sigma[u]$ result:

$$D = y - y_0 = \int_{x_0}^{x} \frac{\mathrm{d}u}{\sigma[u]} = v_0 E + \frac{v_1 E^2}{2} + \frac{v_2 E^3}{6} + \frac{v_3 E^4}{24} + \frac{v_4 E^5}{120} + \frac{v_5 E^6}{720}, \tag{24}$$

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

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 Eqs. (63)–(66) *of Appendix B.*

Owing to the results in Eqs. (24)–(25) and Eqs. (63)–(68), the approximation for the transition density Eq. (20) inherits an analytical structure. The partial derivatives, v_i, φ_i, and λ_i, are compact and can be coded in any standard programming language for the estimation of a wide class of continuous-time models. The tractability of $c_j[y|y_0;\theta]$ 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 correspond to those in Aït-Sahalia (2003) when the irreducible multivariate method is applied to the univariate case. To be exact, Proposition 3 can be interpreted as the explicit Taylor expansion solution of the system of equations in Theorem 2 of Aït-Sahalia (2003) for univariate diffusions with arbitrary $\mu[X]$ and $\sigma[X]$. However, in the irreducible multivariate modeling case, the expansion coefficients do not generally afford the accuracy of integral representations and must be approximated by Taylor series in $x - x_0$ irrespective of whether $x - x_0$ is small or not. Meanwhile, the integral Eq. (5) and the reduced coefficients in our Proposition 2 hold for any x and x_0 and can be evaluated through efficient integration routines. Thus it may be preferable to use Proposition 2 to obtain the expansion coefficients, $c_j[y|y_0;\theta]$, when $x - x_0$ is not very small.

Proposition 3 becomes useful only when $x - x_0$ is very small (e.g., $x = x_0$). The reason is that when $x - x_0 = 0$ (or equivalently $y - y_0 = 0$), the coefficients in our Proposition 2 involve 0/0, which implies that when $x - x_0$ is very small the division of a very small number by another very small number can occur. In the present method, the development in Proposition 3 is intended to handle a tricky numerical situation, while in the irreducible multivariate setting one must resort to Taylor approximations. Thus it must be appreciated that a potential tradeoff exists between the fully closed-form irreducible method in Theorem 2 of Aït-Sahalia (2003) and the present scheme that retains simple one-dimensional integrals in Proposition 2 and must be evaluated numerically.

3. Density approximation for a parametric model class

Spurred by our characterizations in Propositions 1–3, this section applies the density approximation Eq. (20) to the following eight-parameter encompassing class of one-dimensional processes stemming from Aït-Sahalia (1996):

$$\mu[X;\theta] = \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1} \quad \text{and} \tag{26}$$

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

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

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

The distinctive property of this continuous-time process is that $\mu[X;\theta]$ in Eq. (26) accommodates a nonlinear drift, and $\sigma[X;\theta]$ in Eq. (27) implies stochastic elasticity of variance. Given these features, this model is labeled as SEV-ND. Under appropriate restrictions on θ, the SEV-ND model subsumes several theoretically appealing models for X_t that display constant elasticity of variance (hereby CEV) with nonlinear, linear, and constant drift, respectively, and the affine model (hereby AFF), as in

$$\text{SEV-ND}: \quad \mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1})\,\mathrm{d}t + \sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}\,\mathrm{d}W_t$$

$$\text{SEV-LD}: \quad \mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t)\,\mathrm{d}t + \sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}\,\mathrm{d}W_t$$

$$\text{SEV-CD}: \quad \mathrm{d}X_t = \alpha_0\,\mathrm{d}t + \sqrt{\beta_0 + \beta_1 X_t + \beta_2 x_t^{\beta_3}}\,\mathrm{d}W_t$$

$$\text{CEV-ND}: \quad \mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1})\,\mathrm{d}t + \beta_2 X_t^{\beta_3}\,\mathrm{d}W_t$$

$$\text{CEV-LD}: \quad \mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t)\mathrm{d}t + \beta_2 X_t^{\beta_3}\,\mathrm{d}W_t$$

$$\text{CEV-CD}: \quad \mathrm{d}X_t = \alpha_0\,\mathrm{d}t + \beta_2 X_t^{\beta_3}\,\mathrm{d}W_t$$

$$\text{AFF}: \mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t)\,\mathrm{d}t + \sqrt{\beta_0 + \beta_1 X_t}\,\mathrm{d}W_t.$$

To apply the enhanced method of Aït-Sahalia (1999, 2002) to the process in Eqs. (26)–(27) requires

$$Y_t \equiv \int^{X_t} \frac{\mathrm{d}u}{\sigma[u;\theta]} = \int \frac{\mathrm{d}u}{\sqrt{\beta_0 + \beta_1 u + \beta_2 u^{\beta_3}}},$$

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 SEV class is 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 Eq. (20) by defining $V_0 \equiv \beta_0 + \beta_1 X_0 + \beta_2 X_0^{\beta_3}$. Clearly, $v_0^2 V_0 = 1$, where $v_0 \equiv 1/\sqrt{\beta_0 + \beta_1 X + \beta_2 X^{\beta_3}}$. Straightforward successive differentiation of $v_0^2 V_0 = 1$ with respect to X_0 yields the first five derivatives of v_0 with respect to X_0:

$$v_1 = -v_0 V_1/(2V_0), \tag{29}$$

$$v_2 = -(3v_1 V_1 + v_0 V_2)/(2V_0), \tag{30}$$

$$v_3 = -(5v_2 V_1 + 4v_1 V_2 + v_0 V_3)/(2V_0), \tag{31}$$

$$v_4 = -(7v_3 V_1 + 9v_2 V_2 + 5v_1 V_3 + v_0 V_4)/(2V_0), \quad \text{and} \tag{32}$$

$$v_5 = -(9v_4 V_1 + 16v_3 V_2 + 14v_2 V_3 + 6v_1 V_4 + v_0 V_5)/(2V_0), \tag{33}$$

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 $\varphi_i = \partial^i(f[X_0]/\sigma[X_0])/\partial X_0^i$. $\varphi_0 = f[X_0]/\sigma[X_0] = \mu[X_0]/V_0 - V_1/(4V_0) = U_0/V_0$, 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,

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

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

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

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

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

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 Eqs. (26)–(27), the recursive nature of φ_i determines Eq. (25) of Proposition 3.

Finally, we characterize each $\lambda_i = \partial^i\lambda[y_0]/\partial y_0^i$ in Eqs. (67)–(68) as a function of X_0. Based on the calculations in Appendix C, each required λ_i is

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

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

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

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

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

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

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

where S_0 through S_7 are shown in Eqs. (69)–(76), and H_0 through H_6 are shown in Eqs. (85)–(91). 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 Eqs. (63)–(66) and φ_i in Eqs. (34)–(38) are employed for constructing $\int_{y_0}^{y}\mu_Y[w]\,\mathrm{d}w$ expansion in Eq. (25). Given the choice of $\mu[X;\theta]$ and $\sigma[X;\theta]$, we obtain the density approximation for SEV-ND through Eq. (20). Density functions for other continuous-time models are special cases.

Before moving 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 Eq. (20). To illustrate this aspect we pick two candidate continuous-time models by setting $\beta_0 \equiv 0$ in the AFF model (i.e., the square root model) and $\alpha_0 \equiv 0$ in the CEV-LD model. Each candidate stochastic process has an exact density that allows comparison with the approximate density. Guided by Aït-Sahalia (1999) and Bakshi and Ju (2005) 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 $\mathrm{MAXE} \equiv \max(|p^{\mathrm{exact}}[x|x_0] - p^{\mathrm{approx}}[x|x_0]|)$ and the maximum exact conditional density as $\max(p^{\mathrm{exact}}[x|x_0])$. The main finding is that our density approximation is accurate for both $\Delta = 1/12$ and 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 $\sum_{k=0}^{K} c_k[\gamma[x]|\gamma[x_0];\theta]\Delta^k/k!$: 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 reached machine precision. However, as would be expected, embedding each additional term in the approximation tends to make the method progressively more accurate.

4. 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, a consensus is growing that the square root specification is grossly misspecified (see Andersen et al., 2002; Bakshi et al., 2000; Bates, 2000; Elerian et al., 2001; Eraker et al., 2003; 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 shown rejection of square root volatility models: Does the drift of the volatility process admit departures from linearity? Are volatility models with general $\sigma[X]$ more properly specified from empirical standpoints? What is the empirical potential of variance processes in the Aït-Sahalia (1996) class Eqs. (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 the aforementioned questions we need a suitable proxy for market volatility, which is intrinsically unobservable. Among the possible choices at the daily frequency, the empirical literature has appealed to generalized autoregressive conditional heteroskedasticity volatility constructed from daily returns, cumulated squared intraday returns (Andersen et al., 2003), short-term near-money Black and Scholes implied volatility (Bakshi et al., 2000; Pan, 2002), and market volatility extracted from S&P 100 index option prices, VIX (Jones, 2003). For reasons outlined in Jones (2003), we adopt the forward-looking VIX volatility measure in our empirical work. Thus we are drawing conclusions about the desirability of stochastic volatility models based on the estimated dynamics of the VIX index.

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 Eq. (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.

x_0	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18
Panel A: $\mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t)\,\mathrm{d}t + \sqrt{\beta_1 X_t}\,\mathrm{d}W_t$, $t = 1/12$									
$\max(p\|x_0)(10^1)$	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: $\mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t)\,\mathrm{d}t + \sqrt{\beta_1 X_t}\,\mathrm{d}W_t$, $t = 1$									
$\max(p\|x_0)\ (10^1)$	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
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
Panel C: $\mathrm{d}X_t = \alpha_1 X_t + \beta_2 X_t^{\beta_3}\,\mathrm{d}W_t$, $t = 1/12$									
$\max(p\|x_0)(10^{-2})$	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: $\mathrm{d}X_t = \alpha_1 X_t + \beta_2 X_t^{\beta_3}\,\mathrm{d}W_t$, $t = 1$									
$\max(p\|x_0)\ (10^{-2})$	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

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 $\{t = i\Delta \mid i = 0, \ldots, n\}$.

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

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

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

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 of the models 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 impose $\beta_0 \equiv 0$. In this case, the AFF specification collapses to the square root model $\mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t)\,\mathrm{d}t + \sqrt{\beta_1 X_t}\,\mathrm{d}W_t$.

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 underlying market volatility 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 CEV-LD 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:

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

which is distributed $\chi^2[\dim[\theta_U] - \dim[\theta_R]]$. That the data favor 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.[1]

Analyzing the MLE results across CEV-CD, CEV-LD, and CEV-ND 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 nonlinearities in the drift of the variance process. Consider CEV-ND. 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 with the linear drift

[1]Demonstrating the computational superiority of the density approximation, the MLE computer code converged rapidly (in less than one minute on a one 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 et al. (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).

Table 2

Estimation results for market index variance

The encompassing model, SEV-ND, accommodates stochastic elasticity of variance and a nonlinear drift function of the type: $\mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1})\,\mathrm{d}t + \sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}\,\mathrm{d}W_t$. Nesting the constant drift, CEV-CD, and linear drift, CEV-LD, the constant elasticity of variance with nonlinear drift, CEV-ND, is $\mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1})\,\mathrm{d}t + \beta_2 X_t^{\beta_3}\,\mathrm{d}W_t$. The AFF model is $\mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t)\,\mathrm{d}t + \sqrt{\beta_0 + \beta_1 X_t}\,\mathrm{d}W_t$. 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. The daily data are sampled over the period of July 1, 1988 to January 10, 2000 (2,907 observations). We take $X_t = \mathrm{VIX}_t^2$ and the VIX series is scaled by one hundred to convert it into a decimal. The approximate density is analytical and displayed in Eq. (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 $\mathscr{L}[\theta] \equiv \sum_{i=1}^{n} \log\{p_X[\Delta, X_{i\Delta}|X_{(i-1)\Delta};\theta]\}$. The Akaike Information Criterion (AIC) is computed as $-2/n(\mathscr{L}[\theta] - \dim[\theta])$. Thus, a more properly specified model has higher $-(n/2)$ AIC. Likelihood ratio test statistic for comparing nested models is $\mathscr{L}^* \equiv -2 \times (\mathscr{L}[\theta_R] - \mathscr{L}[\theta_U]) \sim \chi^2[\mathrm{d}f]$, where d$f$ is the number of exclusion restrictions and the 95% criterion values are

df	1	2	3	4
$\chi^2[\mathrm{d}f]$	3.84	6.0	7.82	9.50

dim[θ] is reported in curly brackets below $\mathscr{L}$.

Model	$\mathscr{L}$	$-(n/2)$ 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)		
CEV-CD	12089.65	12086.65	0.0664					4.7825	1.2781
	{3}		(0.0114)					(0.1365)	(0.0214)
CEV-LD	12090.40	12086.40	0.0941	−1.4607				4.7046	1.2732
	{4}		(0.0255)	(1.1939)				(0.1365)	(0.0217)
CEV-ND	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)
SEV-CD	12093.21	12089.21	0.0654				0.0142	47.0436	2.8302
	{4}		(0.0117)				(0.0050)	(10.4937)	(0.1183)
SEV-LD	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)
SEV-ND	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)

model CEV-LD, which has a log-likelihood of 12090.40, adding two nonlinear drift parameters raises the log-likelihood to 12094.24. The resulting log-likelihood ratio statistic $\mathscr{L}^*$ 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 nonlinear 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 CEV-ND. The realized value of the test statistic $\mathscr{L}^*$ is 9.18, which can be compared with 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 CEV-CD specification. The small increase of $\mathscr{L}$ from 12089.65 to 12090.40 is insufficiently large to make $\mathscr{L}^* = 1.5$ statistically significant with one degree of freedom.

Ranging between 1.2732 and 1.2781, the exponent parameter β_3 is statistically significant in CEV-CD, CEV-LD, and CEV-ND 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 $\beta_3 \leqslant 1$).

The SEV class of variance processes shares the same drift and volatility structure as models in the CEV class except that the SEV embeds an additional linear-term in the $\sigma[X]$ specification. The implementation of SEV-CD, SEV-LD, and SEV-ND brings out several incremental facts about the behavior of the volatility processes. Comparing the log-likelihood and model estimates across CEV-CD and SEV-CD, CEV-LD and SEV-LD, and CEV-ND and SEV-ND 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 $\mathscr{L}^*$ 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 SEV class.

Because the estimated value of β_0 is zero in the SEV 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 SEV specifications increases faster than that in the CEV specifications. So, the SEV specifications accommodate greater volatility at both low and high variance levels. For the SEV specifications, the linear term $\beta_1 X$ becomes more pronounced than the nonlinear term $\beta_2 X^{\beta_3}$ for $X < (\beta_1/\beta_2)^{1/(\beta_3-1)}$. In the region $X < 0.0136$ in SEV-ND, the linear term $\beta_1 X$ is more heavily weighted than the nonlinear 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 SEV-CD, SEV-LD, and SEV-ND strengthen our earlier conclusions from CEV models that support the existence of nonlinearities in the drift function. The market variance data prefer SEV-ND over both SEV-CD and SEV-LD, while SEV-CD and SEV-LD are indistinguishable based on the log-likelihoods and log-likelihood ratio statistics. More precisely, the $\mathscr{L}^*$ is 11.72 between SEV-CD and SEV-ND and 10.48 between SEV-LD and SEV-ND. The estimated α_2 of -166.72 (-209.35) for the CEV-ND (SEV-ND) 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 nonlinearity of the drift function are robust phenomena in index volatility markets.

Elasticity parameters obtained from CEV and SEV models exhibit a mutually coherent mapping. The β_3 of SEV is slightly more than twice the CEV counterpart. The following interpretation holds for estimated β_3 of SEV-ND. With $\beta_3 = 2.8822$, the variance function $\sigma^2[X]$ is convex in variance, that is,

$$\frac{\partial \sigma^2[X;\theta]}{\partial X} = \beta_1 + \beta_2 \beta_3 X^{\beta_3 - 1} > 0 \quad \text{and} \quad \frac{\partial^2 \sigma^2[X;\theta]}{\partial X^2} = \beta_2 \beta_3 (\beta_3 - 1) X^{\beta_3 - 2} > 0.$$

The one-sided t-test overwhelmingly rejects the null hypothesis that $\beta_3 \leqslant 2$. 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, SEV-ND, exemplifies nonlinearity in the drift and diffusion function of the type described in Eqs. (26)–(27), it rank-orders first with $-n/2$ AIC of 12092.07, followed next by SEV-CD and SEV-LD with $-n/2$ AIC of 12089.21 and 12088.83, respectively. 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 SEV-ND admits a more complex representation, the nonlinearity parameters are vital to generating a more realistic time-evolution of market volatility.

To further analyze the structure of market volatility, Fig. 1 plots the drift, $\mu[X]$, and diffusion, $\sigma^2[X]$, function for the SEV-ND model and compares them with the short interest rate counterparts in Figs. 4B and 4C of Aït-Sahalia (1996). Accounting for differences in estimation methods and parameter values for θ in our Tables 2 and 4 in Aït-Sahalia (1996), this exercise nonetheless shows that market volatility and interest rate share a closely related parametric structure. First, the shapes of $\mu[X]$ and $\sigma^2[X]$ for market volatility and interest rate generally resemble one another. However, visual impression suggests that the mean-reversion in market volatility is relatively stronger especially in the tails. Second, the exponent parameter, β_3, is instrumental to the volatility specification of both interest rate and market variance. While Aït-Sahalia (1996) reports β_3 close to 2.0, the corresponding estimate from market volatility is 40% larger around 2.88. For an economic justification behind the shapes of $\mu[X]$ and $\sigma^2[X]$ for the SEV-ND model, we refer the reader to the discussions in Aït-Sahalia (1996).

5. 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 implications of estimating a much wider class of equity volatility models.

Our theoretical contribution in Propositions 1 and 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

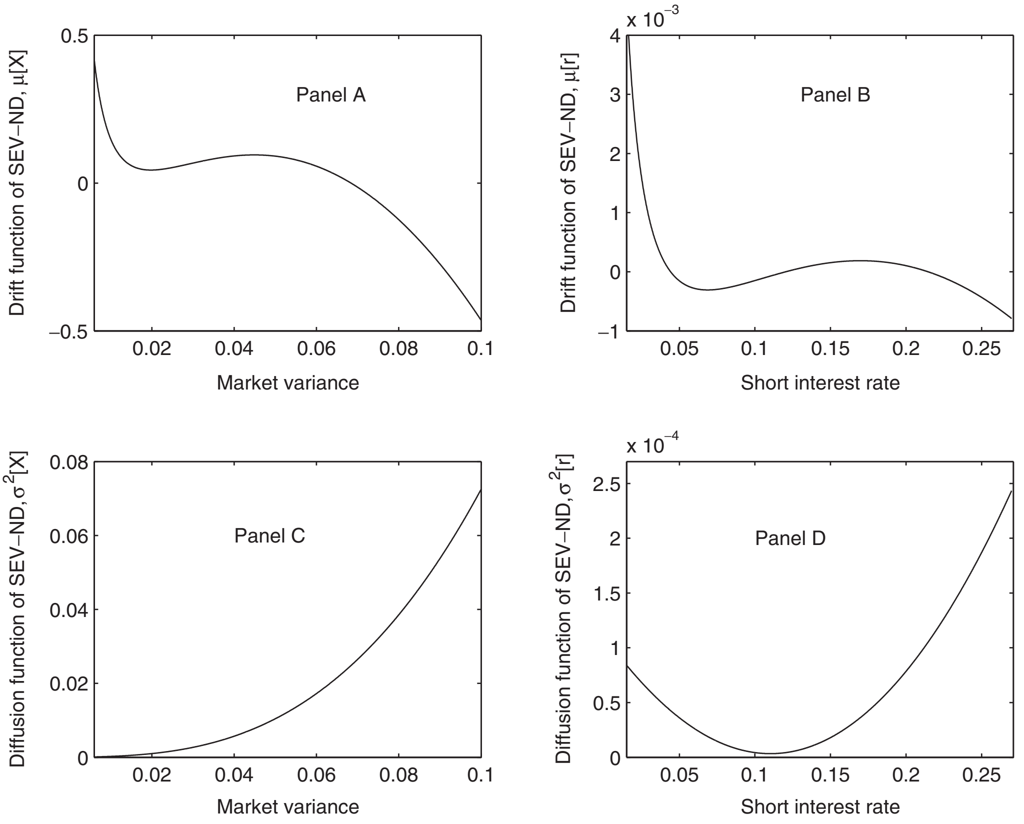

Fig. 1. Drift and diffusion functions for market variance and short interest rate for the SEV-ND model. Panels A and B, respectively plot the drift function, $\mu[.]$, when the market variance and the short interest rate each follow the SEV-ND process: $\mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^{-1})\,\mathrm{d}t + \sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}\,\mathrm{d}W_t$. We plot the corresponding diffusion function, $\sigma^2[.]$, in Panels C and D. All calculations for the market variance process are based on the maximum-likelihood estimation results reported in Table 2. The parameter estimates for the interest rate process are taken from Table 4 in Aït-Sahalia (1996): $\alpha_0 = -4.643 \times 10^{-3}$, $\alpha_1 = 4.333 \times 10^{-2}$, $\alpha_2 = -1.143 \times 10^{-1}$, $\alpha_3 = 1.304 \times 10^{-4}$, $\beta_0 = 1.108 \times 10^{-4}$, $\beta_1 = -1.883 \times 10^{-3}$, $\beta_2 = 9.681 \times 10^{-3}$, and $\beta_3 = 2.073$. Thus our interest rate Panels B and D coincide with Figs. 4B and C in Aït-Sahalia (1996).

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.

The method can be adapted to undertake risk measurements under nonlinear drift and diffusion forcing processes and for approximating the risk-neutral densities for valuing contingent claims when the characteristic function of the state-price density is unavailable (Bakshi and Madan, 2000; Duffie et al., 2000). Suppose the log stock price is $X_t \equiv \log(\mathscr{P}_t)$. Then Ito's lemma implies $\mathrm{d}X_t = (r - \frac{1}{2}\sigma[\mathrm{e}^X;\theta])\,\mathrm{d}t + \sigma[\mathrm{e}^X;\theta]\,\mathrm{d}W_t$. With an analytical risk-neutral density approximation $q_X^{(K)}[X]$, the European option price with strike price $\mathscr{K}$, maturity T, and interest rate r can be computed as $\mathrm{e}^{-rT}\int_{\log(\mathscr{K})}^{+\infty} (\mathrm{e}^{X_{t+T}} - \mathscr{K}) q_X^{(K)}[X_{t+T}]\,\mathrm{d}X_{t+T}$. The field of density approximation is amenable to analytical characterizations for estimation and pricing under a large class of nonstandard continuous-time models.

Appendix A. Proof of Proposition 2

To outline a proof of Proposition 2, consider the determination of $c_2[y]$:

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

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

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

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

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

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

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

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

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

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

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

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

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

It is straightforward to show that

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

and it holds that

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

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

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

Combining Eq. (55) and (59) we finally have

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

Now

$$\begin{aligned}\int_{y_0}^{y}c_2[w]\,\mathrm{d}w &= \int_{y_0}^{y}\frac{2}{(w-y_0)^2}\int_{y_0}^{w}(z-y_0)\left(\lambda[z]c_1[z]+\frac{1}{2}\frac{\partial^2 c_1[z]}{\partial z^2}\right)\mathrm{d}z\,\mathrm{d}w\\ &= \int_{y_0}^{y}\int_{y_0}^{y}\frac{2}{(w-y_0)^2}(z-y_0)\left(\lambda[z]c_1[z]+\frac{1}{2}\frac{\partial^2 c_1[z]}{\partial z^2}\right)1[z<w]\,\mathrm{d}z\,\mathrm{d}w\end{aligned}$$

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

Plugging Eq. (61) into Eq. (60) and rearranging results in

$$
\begin{aligned}
c_3[y] = {} & c_1^3[y] + \frac{3}{(y-y_0)^2}\left(c_1[y](\lambda[y]+\lambda[y_0]) - 3c_2[y]\right) \\
& + \frac{3}{(y-y_0)^3}\left(\frac{\lambda'[y]-\lambda'[y_0]}{2} + \int_{y_0}^{y}\lambda^2[w]\,\mathrm{d}w\right). \qquad (62)
\end{aligned}
$$

The proof of $c_4[y]$ is unwieldy and omitted to save on space (available from the authors).

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

$$
c_1[y|y_0;\theta] = \lambda_0 + \frac{\lambda_1 D}{2} + \frac{\lambda_2 D^2}{6} + \frac{\lambda_3 D^3}{24} + \frac{\lambda_4 D^4}{120} + \frac{\lambda_5 D^5}{720} + \frac{\lambda_6 D^6}{5040}, \qquad (63)
$$

$$
c_2[y|y_0;\theta] = c_1^2[y_0] + \frac{\lambda_2}{6} + \frac{\lambda_3 D}{12} + \frac{\lambda_4 D^2}{40} + \frac{\lambda_5 D^3}{180} + \frac{\lambda_6 D^4}{1008}, \qquad (64)
$$

$$
\begin{aligned}
c_3[y|y_0;\theta] = {} & c_1^3[y_0] + \left(\frac{\lambda_1^2}{4} + \frac{\lambda_0\lambda_2}{2} + \frac{\lambda_4}{40}\right) + \left(\frac{\lambda_1\lambda_2}{2} + \frac{\lambda_0\lambda_3}{4} + \frac{\lambda_5}{80}\right)D \\
& + \left(\frac{3\lambda_2^2}{20} + \frac{\lambda_1\lambda_3}{5} + \frac{3\lambda_0\lambda_4}{40} + \frac{\lambda_6}{280}\right)D^2 + \left(\frac{5\lambda_2\lambda_3}{48} + \frac{13\lambda_1\lambda_4}{240} + \frac{\lambda_0\lambda_5}{60}\right)D^3 \\
& + \left(\frac{23\lambda_3^2}{1344} + \frac{11\lambda_2\lambda_4}{420} + \frac{19\lambda_1\lambda_5}{1680} + \frac{\lambda_0\lambda_6}{336}\right)D^4 + \left(\frac{\lambda_3\lambda_4}{120} + \frac{\lambda_2\lambda_5}{192} + \frac{13\lambda_1\lambda_5}{6720}\right)D^5 \\
& + \left(\frac{43\lambda_4^2}{4320} + \frac{7\lambda_3\lambda_5}{4320} + \frac{13\lambda_2\lambda_6}{15120}\right)D^6 \quad \text{and} \qquad (65)
\end{aligned}
$$

$$
\begin{aligned}
c_4[y|y_0;\theta] = {} & c_1^4[y_0] + \left(\lambda_0\lambda_1^2 + \lambda_0^2\lambda_2 + \frac{3\lambda_2^2}{20} + \frac{\lambda_1\lambda_3}{5} + \frac{\lambda_0\lambda_4}{10} + \frac{\lambda_6}{280}\right) \\
& + \left(\frac{\lambda_3^2}{2} + 2\lambda_0\lambda_1\lambda_2 + \frac{\lambda_0^2\lambda_3}{2} + \frac{\lambda_2\lambda_3}{4} + \frac{3\lambda_1\lambda_4}{20} + \frac{\lambda_0\lambda_5}{20}\right)D \\
& + \left(\frac{11\lambda_1^2\lambda_2}{12} + \frac{3\lambda_0\lambda_2^2}{5} + \frac{4\lambda_0\lambda_1\lambda_3}{5} + \frac{23\lambda_3^2}{336} + \frac{19\lambda_2\lambda_4}{168} + \frac{3\lambda_1\lambda_5}{56} + \frac{\lambda_0\lambda_6}{70}\right)D^2
\end{aligned}
$$

$$
\begin{aligned}
&+\left(\frac{7\lambda_1\lambda_2^2}{15}+\frac{19\lambda_1^2\lambda_3}{60}+\frac{5\lambda_0\lambda_2\lambda_3}{12}+\frac{13\lambda_0\lambda_1\lambda_4}{60}+\frac{\lambda_3\lambda_4}{20}+\frac{\lambda_0^5\lambda_5}{30}+\frac{\lambda_2\lambda_5}{30}\right.\\
&\left.+\frac{11\lambda_1\lambda_6}{840}\right)D^3+\left(\frac{13\lambda_2^3}{180}+\frac{3\lambda_1\lambda_2\lambda_3}{10}+\frac{5\lambda_0\lambda_2\lambda_3}{12}+\frac{13\lambda_0\lambda_1\lambda_4}{60}+\frac{23\lambda_0\lambda_3^2}{336}\right.\\
&+\frac{19\lambda_2^2\lambda_4}{420}+\frac{11\lambda_0\lambda_2\lambda_4}{105}+\frac{11\lambda_1\lambda_6}{840}+\frac{23\lambda_4^2}{2880}+\frac{\lambda_0^2\lambda_6}{168}+\frac{19\lambda_0\lambda_1\lambda_5}{420}+\frac{19\lambda_3\lambda_5}{1440}\\
&\left.+\frac{37\lambda_2\lambda_6}{5040}\right)D^4+\left(\frac{\lambda_2^2\lambda_3}{15}+\frac{157\lambda_1\lambda_3^2}{3360}+\frac{181\lambda_1\lambda_2\lambda_4}{2520}+\frac{\lambda_0\lambda_3\lambda_4}{30}+\frac{79\lambda_1^2\lambda_5}{5040}\right.\\
&\left.+\frac{\lambda_0\lambda_2\lambda_5}{48}+\frac{7\lambda_4\lambda_5}{1800}+\frac{13\lambda_0\lambda_1\lambda_6}{1680}+\frac{137\lambda_3\lambda_6}{50400}\right)D^5+\left(\frac{9\lambda_2\lambda_3^2}{448}+\frac{391\lambda_2^2\lambda_4}{25200}\right.\\
&+\frac{79\lambda_1\lambda_3\lambda_4}{3600}+\frac{43\lambda_0\lambda_4^2}{10800}+\frac{139\lambda_1\lambda_2\lambda_5}{10800}+\frac{7\lambda_0\lambda_3\lambda_5}{1080}+\frac{71\lambda_5^2}{158400}\\
&\left.+\frac{13\lambda_0\lambda_2\lambda_6}{3780}+\frac{281\lambda_4\lambda_6}{369600}\right)D^6, \qquad (66)
\end{aligned}
$$

where $D = y - y_0$ is given in Eq. (24). In terms of $\mu[X_0]$ and $\sigma[X_0]$, each required λ_i can be recursively derived as

$$\lambda_0 = -(f^2[X_0] + f'[X_0]\sigma[X_0])/2 \quad \text{and} \qquad (67)$$

$$\lambda_i = \lambda'_{i-1}\sigma[X_0], \quad i = 1, 2, 3, 4, 5, 6, \qquad (68)$$

which completes the characterization of the density approximation.

Appendix C. Expressions for λ_j in the SEV-ND model

To develop the expressions for λ_j, let

$$S_0 = \sigma[X_0] = \sqrt{\beta_0 + \beta_1 X_0 + \beta_2 X_0^{\beta_3}}. \qquad (69)$$

Then $S_0^2 = \beta_0 + \beta_1 X_0 + \beta_2 X_0^{\beta_3} = V_0$. The derivatives of S_0 with respect to X_0 are

$$S_1 = V_1/(2S_0), \qquad (70)$$

$$S_2 = (V_2 - 2S_1^2)/(2S_0), \qquad (71)$$

$$S_3 = (V_3 - 6S_1S_2)/(2S_0), \qquad (72)$$

$$S_4 = (V_4 - 8S_1S_3 - 6S_2^2)/(2S_0), \qquad (73)$$

$$S_5 = (V_5 - 10S_1S_4 - 20S_2S_3)/(2S_0), \qquad (74)$$

$$S_6 = (V_6 - 12S_1S_5 - 30S_2S_4 - 20S_3^2)/(2S_0) \quad \text{and} \qquad (75)$$

$$S_7 = (V_7 - 14S_1S_6 - 42S_2S_5 - 70S_3S_4)/(2S_0), \qquad (76)$$

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

$$F_0 = U_0/S_0, \tag{77}$$

$$F_1 = (U_1 - F_0S_1)/S_0, \tag{78}$$

$$F_2 = (U_2 - 2F_1S_1 - F_0S_2)/S_0, \tag{79}$$

$$F_3 = (U_3 - 3F_2S_1 - 3F_1S_2 - F_0S_3)/S_0, \tag{80}$$

$$F_4 = (U_4 - 4F_3S_1 - 6F_2S_2 - 4F_1S_3 - F_0S_4)/S_0, \tag{81}$$

$$F_5 = (U_5 - 5F_4S_1 - 10F_3S_2 - 10F_2S_3 - 5F_1S_4 - F_0S_5)/S_0, \tag{82}$$

$$F_6 = (U_6 - 6F_5S_1 - 15F_4S_2 - 20F_3S_3 - 15F_2S_4 - 6F_1S_5 - F_0S_6)/S_0 \quad \text{and} \tag{83}$$

$$F_7 = (U_7 - 7F_6S_1 - 21F_5S_2 - 35F_4S_3 - 35F_3S_4 - 21F_2S_5 - 7F_1S_6 - F_0S_7)/S_0, \tag{84}$$

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

$$H_0 = -(f^2[X_0] + f'[X_0]\sigma[X_0])/2 = -(F_0^2 + F_1S_0)/2. \tag{85}$$

The first six derivatives of H_0 with respect to X_0 are

$$H_1 = -(2F_0F_1 + F_2S_0 + F_1S_1)/2, \tag{86}$$

$$H_2 = -(2F_1^2 + 2F_0F_2 + 2F_2S_1 + F_3S_0 + F_1S_2)/2, \tag{87}$$

$$H_3 = -(6F_1F_2 + 2F_0F_3 + F_4S_0 + 3F_3S_1 + 3F_2S_2 + F_1S_3)/2, \tag{88}$$

$$H_4 = -(6F_2^2 + 8F_1F_3 + 2F_0F_4 + F_5S_0 + 4F_4S_1 + 6F_3S_2 + 4F_2S_3 + F_1S_4)/2, \tag{89}$$

$$\begin{aligned} H_5 = & -(20F_2F_3 + 10F_1F_4 + 2F_0F_5 + F_6S_0 + 5F_5S_1 + 10F_4S_2 + 10F_3S_3 \\ & + 5F_2S_4 + F_1S_5)/2 \quad \text{and} \end{aligned} \tag{90}$$

$$\begin{aligned} H_6 = & -(20F_3^2 + 30F_2F_4 + 12F_1F_5 + 2F_0F_6 + F_7S_0 + 6F_6S_1 + 15F_5S_2 \\ & + 20F_4S_3 + 15F_3S_4 + 6F_2S_5 + F_1S_6)/2. \end{aligned} \tag{91}$$

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, Princeton, NJ.

Aït-Sahalia, Y., Kimmel, R., 2003. Estimating affine multifactor term structure models using closed-form likelihood expansions. Unpublished working paper. Princeton University, Princeton, NJ.
Andersen, T., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. Journal of Finance 57, 1239–1284.
Andersen, T., Bollerslev, T., Diebold, F., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.
Bakshi, G., Cao, C., Chen, Z., 2000. Pricing and hedging long-term options. Journal of Econometrics 94, 277–318.
Bakshi, G., Ju, N., 2005. A refinement to Aït-Sahalia (2002) maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Journal of Business 78, 2037–2052.
Bakshi, G., Madan, D., 2000. Spanning and derivative security valuation. Journal of Financial Economics 55, 205–238.
Bates, D., 2000. Post-87 crash fears in S&P 500 futures options. Journal of Econometrics 94, 181–238.
Brandt, M., 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., Pan, J., 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., 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., Chib, S., Shephard, N., 2001. Likelihood inference for discretely observed nonlinear diffusions. Econometrica 69, 959–993.
Eraker, B., Johannes, M., Polson, N., 2003. The impact of jumps in volatility and returns. Journal of Finance 58, 1269–1300.
Heston, S., 1993. A closed form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies 6, 327–343.
Jones, C., 2003. The dynamics of stochastic volatility: evidence from underlying and options markets. Journal of Econometrics 116, 181–224.
Li, M., Pearson, N., Poteshman, A., 2004. Conditional estimation of diffusion processes. Journal of Financial Economics 74, 31–66.
Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. Journal of Financial Economics 63, 3–50.