THE JOURNAL OF FINANCE • VOL. LIV, NO. 4 • AUGUST 1999

Transition Densities for Interest Rate and Other Nonlinear Diffusions

YACINE AÏT-SAHALIA*

ABSTRACT

This paper applies to interest rate models the theoretical method developed in Aït-Sahalia (1998) to generate accurate closed-form approximations to the transition function of an arbitrary diffusion. While the main focus of this paper is on the maximum-likelihood estimation of interest rate models with otherwise unknown transition functions, applications to the valuation of derivative securities are also briefly discussed.

CONTINUOUS-TIME MODELING IN FINANCE, though introduced by Louis Bachelier's 1900 thesis on the theory of speculation, really started with Merton's seminal work in the 1970s. Since then, the continuous-time paradigm has proved to be an immensely useful tool in finance and more generally economics. Continuous-time models are widely used to study issues that include the decision to optimally consume, save, and invest, portfolio choice under a variety of constraints, contingent claim pricing, capital accumulation, resource extraction, game theory, and more recently contract theory. Many refinements and extensions are possible, but the basic dynamic model for the variable(s) of interest X_t is a stochastic differential equation,

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

where W_t is a standard Brownian motion and the drift μ and diffusion σ^2 are known functions except for an unknown parameter[1] vector θ in a bounded set $\Theta \subset R^d$.

One major impediment to both theoretical modeling and empirical work with continuous-time models of this type is the fact that in most cases little can be said about the implications of the dynamics in equation (1) for longer

*Department of Economics, Princeton University. *Mathematica* code to implement this method can be found at http://www.princeton.edu/~yacine. I am grateful to David Bates, René Carmona, Freddy Delbaen, Ron Gallant, Lars Hansen, Per Mykland, Peter C. B. Phillips, Peter Robinson, Angel Serrat, Suresh Sundaresan, and George Tauchen for helpful comments. Robert Kimmel provided excellent research assistance. This research was conducted during the author's tenure as an Alfred P. Sloan Research Fellow. Financial support from the NSF (Grant SBR-9996023) is gratefully acknowledged.

[1] Non- and semiparametric approaches, which do not constrain the functional form of the functions μ and/or σ^2 to be within a parametric class, have been developed (see Aït-Sahalia (1996a, 1996b) and Stanton (1997)).

time intervals. Though equation (1) fully describes the evolution of the variable X over each infinitesimal instant, one cannot in general characterize in closed form an object as simple (and fundamental for everything from prediction to estimation and derivative pricing) as the conditional density of $X_{t+\Delta}$ given the current value X_t. For a list of the rare exceptions, see Wong (1964). In finance, the well-known models of Black and Scholes (1973), Vasicek (1977), and Cox, Ingersoll, and Ross (1985) rely on these existing closed-form expressions. In this paper, I describe and implement empirically a method developed in a companion paper (Aït-Sahalia (1998)) which produces very accurate approximations *in closed form* to the unknown transition function $p_X(\Delta, x|x_0;\theta)$, the conditional density of $X_{t+\Delta} = x$ given $X_t = x_0$ implied by the model in equation (1).

These closed-form expressions can be useful for at least two purposes. First, they let us estimate the parameter vector θ by maximum-likelihood.[2] In most cases, we observe the process at dates $\{t = i\Delta | i = 0,\dots,n\}$, where $\Delta > 0$ is generally small, but fixed as n increases. For instance, the series could be weekly or monthly. Collecting more observations means lengthening the time period over which data are recorded, not shortening the time interval between successive existing observations.[3] Because a continuous-time diffusion is a Markov process, and that property carries over to any discrete subsample from the continuous-time path, the log-likelihood function has the simple form

$$\ell_n(\theta) \equiv n^{-1} \sum_{i=1}^{n} \ln\{p_X(\Delta, X_{i\Delta}|X_{(i-1)\Delta};\theta)\}. \tag{2}$$

With a given Δ, two methods are available in the literature to compute p_X numerically. They involve either solving numerically the Kolmogorov partial differential equation known to be satisfied by p_X (see, e.g., Lo (1988)), or simulating a large number of sample paths along which the process is sampled very finely (see Pedersen (1995), Honoré (1997), and Santa-Clara (1995)). Neither method however produces a closed-form expression to be maximized

[2] A large number of new approaches have been developed in recent years. Some theoretical estimation methods are based on the generalized method of moments (Hansen and Scheinkman (1995) and Bibby and Sørenson (1995)) and on nonparametric density-matching (Aït-Sahalia (1996a, 1996b)), others are based on nonparametric approximate moments (Stanton (1997)), simulations (Duffie and Singleton (1993), Gouriéroux, Monfort, and Renault (1993), Gallant and Tauchen (1998), and Pedersen (1995)), the spectral decomposition of the infinitesimal generator (Hansen, Scheinkman, and Touzi (1998) and Florens, Renault, and Touzi (1995)), random sampling of the process to generate moment conditions (Duffie and Glynn (1997)), or, finally, Bayesian approaches (Eraker (1997), Jones (1997), and Elerian, Chib, and Shephard (1998)).

[3] Discrete approximations to the stochastic differential equation (1) could be employed (see Kloeden and Platen (1992)): see Chan et al. (1992) for an example. As discussed by Merton (1980), Lo (1988), and Melino (1994), ignoring the difference generally results in inconsistent estimators, unless the discretization happens to be an exact one, which is tantamount to saying that p_X would have to be known in closed form.

over θ, and the calculations for all the pairs (x, x_0) must be repeated separately every time the value of θ changes. By contrast, the closed-form expressions in this paper make it possible to maximize the expression in equation (2) with p_X replaced by its closed-form approximation.

Derivative pricing provides a second natural outlet for applications of this methodology. Suppose that we are interested in pricing at date zero a derivative security written on an asset with price process $\{X_t | t \geq 0\}$, and with payoff function $\Psi(X_\Delta)$ at some future date Δ. For simplicity, assume that the underlying asset is traded, so that its risk-neutral dynamics have the form

$$dX_t/X_t = \{r - \delta\}dt + \sigma(X_t;\theta)dW_t, \tag{3}$$

where r is the riskfree rate and δ is the dividend rate paid by the asset—both constant again for simplicity.

It is well known that when markets are dynamically complete, the only price of the derivative security that is compatible with the absence of arbitrage opportunities is

$$P_0 = e^{-r\Delta}E[\Psi(X_\Delta)|X_0 = x_0] = e^{-r\Delta}\int_0^{+\infty}\Psi(x)p_X(\Delta, x|x_0;\theta)\,dx, \tag{4}$$

where p_X is the transition function (or risk-neutral density, or state-price density) induced by the dynamics in equation (3).

The Black–Scholes option pricing formula is the prime example of equation (4), when $\sigma(X_t;\theta) = \sigma$ is constant. The corresponding p_X is known in closed-form (as a lognormal density) and so the integral in equation (4) can be evaluated explicitly for specific payoff functions (see also Cox and Ross (1976)). In general, of course, no known expression for p_X is available and one must rely on numerical methods such as solving numerically the PDE satisfied by the derivative price, or Monte Carlo integration of equation (3). These methods are the exact parallels to the two existing approaches to maximum-likelihood estimation that I described earlier.

Here, given the sequence $\{\tilde{p}_X^{(K)} | K \geq 0\}$ of approximations to p_X, the valuation of the derivative security would be based on the explicit formula

$$P_0^{(K)} = e^{-r\Delta}\int_0^{+\infty}\Psi(x)\tilde{p}_X^{(K)}(\Delta, x|x_0;\theta)\,dx. \tag{5}$$

Formulas of the type given in equation (4) where the unknown p_X is replaced by another density have been proposed in the finance literature (see, e.g., Jarrow and Rudd (1982)). There is an important difference, however, between what I propose and the existing formulas: the latter are based on calculating the integral in equation (4) with an ad hoc density $\hat{p}_X$—typically adding free skewness and kurtosis parameters to the lognormal density, so as to allow for departures from the Black–Scholes formula. In doing so, these formulas ignore the underlying dynamic model specified in equation (3) for the asset price, whereas my method gives in closed form the option pricing formula (of order of precision corresponding to that of the approximation used) that corresponds to the given dynamic model in equation (3). Then one can, for instance, explore how changes in the specification of the volatility function $\sigma(x;\theta)$ affect the derivative price, which is obviously impossible when the specification of the density $\hat{p}_X$ to be used in equation (4) in lieu of p_X is unrelated to equation (3).

The paper is organized as follows. In Section I, I briefly describe the approach used in Aït-Sahalia (1998) to derive a closed-form sequence of approximations to p_X, give the expressions for the approximation, and describe its properties. In Section II, I study a number of interest rate models, some with unknown transition functions, and give the closed-form expressions of the corresponding approximations. Section III reports maximum-likelihood estimates for these models using the Federal funds rate, sampled monthly from 1963 through 1998. Section IV concludes, and a statement of the technical assumptions is in the Appendix.

I. Closed-Form Approximations to the Transition Function

A. Tail Standardization via Transformation to Unit Diffusion

The first step toward constructing the sequence of approximations to p_X consists of standardizing the diffusion function of X—that is, transforming X into another diffusion Y defined as

$$Y_t \equiv \gamma(X_t;\theta) = \int^{X_t} du/\sigma(u;\theta), \tag{6}$$

where any primitive of the function $1/\sigma$ may be selected.

Let $D_X = (\underline{x},\bar{x})$ denote the domain of the diffusion X. I will consider two cases, where $D_X = (-\infty,+\infty)$ or $D_X = (0,+\infty)$. The latter case is often relevant in finance, when considering models for asset prices or nominal interest rates. Moreover, the function σ is often specified in financial models in such a way that $\sigma(0;\theta) = 0$ and μ and/or σ violates the linear growth conditions near the boundaries. The assumptions in the Appendix allow for this behavior.

Because $\sigma > 0$ on the interior of the domain D_X, the function γ in equation (6) is increasing and thus invertible. It maps D_X into $D_Y = (\underline{y},\bar{y})$, the domain of Y. For a given model under consideration, I will assume that the parameter space Θ is restricted in such a way that D_Y is independent of θ in Θ. This restriction on Θ is inessential, but it helps keep the notation simple. Again, in finance, most, if not all cases, will have D_X and D_Y be either the whole real line $(-\infty,+\infty)$ or the half line $(0,+\infty)$.

By applying Itô's Lemma, Y has unit diffusion as desired:

$$dY_t = \mu_Y(Y_t;\theta)dt + dW_t, \tag{7}$$

where

$$\mu_Y(y;\theta) = \frac{\mu(\gamma^{-1}(y;\theta);\theta)}{\sigma(\gamma^{-1}(y;\theta);\theta)} - \frac{1}{2}\frac{\partial\sigma}{\partial x}(\gamma^{-1}(y;\theta);\theta). \tag{8}$$

Finally, note that it can be convenient to define Y_t instead as minus the integral in equation (6) if that makes $Y_t > 0$, for instance if $\sigma(x;\theta) = x^\rho$ and $\rho > 1$. For example, if $D_X = (0,+\infty)$ and $\sigma(x;\theta) = x^\rho$, then $Y_t = (1-\rho)X_t^{1-\rho}$ if $0 < \rho < 1$ (so $D_Y = (0,+\infty)$), $Y_t = \ln(X_t)$ if $\rho = 1$ (so $D_Y = (-\infty,+\infty)$), and $Y_t = (\rho - 1)X_t^{-(\rho-1)}$ if $\rho > 1$ (so $D_Y = (0,+\infty)$ again). In all cases, Y has unit diffusion; that is, $\sigma_Y^2(y;\theta) = 1$. When the transformation $Y_t \equiv \gamma(X_t;\theta) = -\int^{X_t} du/\sigma(u;\theta)$ is used, the drift $\mu_Y(y;\theta)$ in $dY_t = \mu_Y(Y_t;\theta)dt - dW_t$ is, instead of equation (8),

$$\mu_Y(y;\theta) = -\frac{\mu(\gamma^{-1}(y;\theta);\theta)}{\sigma(\gamma^{-1}(y;\theta);\theta)} + \frac{1}{2}\frac{\partial\sigma}{\partial x}(\gamma^{-1}(y;\theta);\theta). \tag{9}$$

The point of making the transformation from X to Y is that it is possible to construct an expansion for the transition density of Y. Of course, this would be of little interest because we only observe X, not the artificially introduced Y, and the transformation depends on the unknown parameter vector θ. However, the transformation is useful because one can obtain the transition density p_X from p_Y through the Jacobian formula

$$\begin{aligned}
p_X(\Delta,x|x_0;\theta) &= \frac{\partial}{\partial x}\,\mathrm{Prob}(X_{t+\Delta} \leq x | X_t = x_0;\theta) \\
&= \frac{\partial}{\partial x}\,\mathrm{Prob}(Y_{t+\Delta} \leq \gamma(x;\theta) | Y_t = \gamma(x_0;\theta);\theta) \\
&= \frac{\partial}{\partial x}\left[\int_{\underline{y}}^{\gamma(x;\theta)} p_Y(\Delta,y|\gamma(y_0;\theta);\theta)dy\right] \\
&= \frac{p_Y(\Delta,\gamma(x;\theta)|\gamma(x_0;\theta);\theta)}{\sigma(\gamma(x;\theta);\theta)}.
\end{aligned} \tag{10}$$

Therefore, there is never any need to actually transform the data $\{X_{i\Delta}, i = 0,\ldots,n\}$ into observations on Y (which depends on θ anyway). Instead, the transformation from X to Y is simply a device to obtain an approximation for p_X from the approximation of p_Y. Practically speaking, when the approximation for p_X has been derived once and for all as the Jacobian transform of that of Y, the process Y no longer plays any role.

B. Explicit Expressions for the Approximation

As shown in Aït-Sahalia (1998), one can derive an explicit expansion for the transition density of the variable Y based on a Hermite expansion of its density $y \mapsto p_Y(\Delta, y | y_0; \theta)$ around a Normal density function. The analytic part of the expansion of p_Y up to order K is given by

$$\tilde{p}_Y^{(K)}(\Delta, y | y_0; \theta) = \Delta^{-1/2} \phi\left(\frac{y - y_0}{\Delta^{1/2}}\right) \exp\left(\int_{y_0}^{y} \mu_Y(w;\theta) dw\right) \sum_{k=0}^{K} c_k(y | y_0; \theta) \frac{\Delta^k}{k!}, \tag{11}$$

where $\phi(z) \equiv e^{-z^2/2}/\sqrt{2\pi}$ denotes the $N(0,1)$ density function, $c_0(y | y_0; \theta) = 1$, and for all $j \geq 1$,

$$c_j(y | y_0; \theta) = j(y - y_0)^{-j} \int_{y_0}^{y} (w - y_0)^{j-1} \times \{\lambda_Y(w) c_{j-1}(w | y_0; \theta) + (\partial^2 c_{j-1}(w | y_0; \theta)/\partial w^2)/2\} \, dw, \tag{12}$$

where $\lambda_Y(y;\theta) \equiv -(\mu_Y^2(y;\theta) + \partial \mu_Y(y;\theta)/\partial y)/2$.

Tables I through V give the explicit expression of these coefficients for popular models in finance, which I discuss in detail in Section II. Before turning to these examples, a few general remarks are in order. The general structure of the expansion in equation (11) is as follows: The leading term in the expansion is Gaussian, $\Delta^{-1/2}\phi((y - y_0)/\Delta^{1/2})$, followed by a correction for the presence of the drift, $\exp(\int_{y_0}^{y} \mu_Y(w;\theta)\, dw)$, and then additional correction terms that depend on the specification of the function $\lambda_Y(y;\theta)$ and its successive derivatives. These correction terms play two roles: they account for the nonnormality of p_Y and they correct for the discretization bias implicit in starting the expansion with a Gaussian term with no mean adjustment and variance Δ (instead of $\text{Var}[Y_{t+\Delta} | Y_t]$, which is equal to Δ only in the first order).

In general, the function p_Y is not analytic in time. Therefore equation (11) must be interpreted strictly as the analytic part, or Taylor series. In particular, for given (y, y_0, θ) it will generally have a finite convergence radius in Δ. As we will see below, however, the series in equation (11) with $K = 1$ or 2 at most is very accurate for the values of Δ that one encounters in empirical work in finance.

The sequence of explicit functions $\tilde{p}_Y^{(K)}$ in equation (11) is designed to approximate p_Y. As discussed above, one can then approximate p_X (the object of interest) by using the Jacobian formula for the inverted change of variable $Y \to X$:

$$\tilde{p}_X^{(K)}(\Delta, x | x_0; \theta) \equiv \sigma(x;\theta)^{-1} \tilde{p}_Y^{(K)}(\Delta, \gamma(x;\theta) | \gamma(x_0;\theta); \theta). \tag{13}$$

The main objective of the transformation $X \to Y$ was to provide a method of controlling the size of the tails of the transition density. As shown in Aït-Sahalia (1998), the fact that Y has unit diffusion makes the tails of the density p_Y, in the limit where Δ goes to zero, similar in magnitude to those of a Gaussian variable. That is, the tails of p_Y behave like $\exp[-y^2/2\Delta]$ as is apparent from equation (11). However, the tails of the density p_X are proportional to $\exp[-\gamma(x;\theta)^2/2\Delta]$. So, for instance, if $\sigma(x;\theta) = 2\sqrt{x}$ then $\gamma(x;\theta) = \sqrt{x}$ and the right tail of p_X becomes proportional to $\exp[-x^2/2\Delta]$; this is verified by equation (13). Not surprisingly, this is the tail behavior for Feller's transition density in the Cox, Ingersoll, and Ross (1985) model. If now $\sigma(x;\theta) = x$, then $\gamma(x;\theta) = \ln(x)$ and the tails of p_X are proportional to $\exp[-\ln(x)^2/2\Delta]$: this is what happens in the log-Normal case (see the Black–Scholes model). In other words, while the leading term of the expansion in equation (11) for p_Y is Gaussian, the expansion for p_X will start with a deformed or "stretched" Gaussian term, with the specific form of the deformation given by the function $\gamma(x;\theta)$.

The sequence of functions in equation (11) solves the forward and backward Kolmogorov equations up to order Δ^K; that is,

$$\begin{cases} \dfrac{\partial \tilde{p}_Y^{(K)}}{\partial \Delta} + \dfrac{\partial}{\partial y}\{\mu_Y(y;\theta)\tilde{p}_Y^{(K)}\} - \dfrac{1}{2}\dfrac{\partial^2 \tilde{p}_Y^{(K)}}{\partial y^2} = O(\Delta^K) \\ \dfrac{\partial \tilde{p}_Y^{(K)}}{\partial \Delta} - \mu_Y(y_0;\theta)\dfrac{\partial \tilde{p}_Y^{(K)}}{\partial y_0} - \dfrac{1}{2}\dfrac{\partial^2 \tilde{p}_Y^{(K)}}{\partial y_0^2} = O(\Delta^K) \end{cases}. \tag{14}$$

The boundary behavior of the transition density $\tilde{p}_Y^{(K)}$ is similar to that of p_Y; under the assumptions made, $\lim_{y \to \underline{y} \text{ or } \bar{y}} p_Y = 0$. The expansion is designed to deliver an approximation of the density function $y \mapsto p_Y(\Delta, y | y_0; \theta)$ for a fixed value of conditioning variable y_0. Therefore, except in the limit where Δ becomes infinitely small, it is not designed to reproduce the limiting behavior of p_Y in the limit where y_0 tends to the boundaries.

Finally, note that the form of the expansion is compatible with the expression that arises out of Girsanov's Theorem in the following sense. Under the assumptions made, the process Y can be transformed by Girsanov's Theorem into a Brownian motion if $D_Y = (-\infty, +\infty)$, or into a Bessel process in dimension 3 if $D_Y = (0, +\infty)$. This gives rise to a formulation of p_Y in a form that involves the conditional expectation of the exponential of the integral of function of a Brownian Bridge (see Gihman and Skorohod (1972, Chap. 3) for the case where $D_Y = (-\infty,+\infty)$), or a Bessel Bridge if $D_Y = (0,+\infty)$. This conditional expectation term can either be expressed in terms of the conditional densities of the Brownian Bridge when $D_Y = (-\infty,+\infty)$ (see Dacunha-Castelle and Florens-Zmirou (1986)), or integrated by Monte Carlo simulation. Further discussion of these and other theoretical properties of the expansion is contained in Aït-Sahalia (1998).

II. Examples

A. Comparison of the Approximation to the Closed-Form Densities for Specific Models

In this section, I study the size of the approximation made when replacing p_X by $\tilde{p}_X^{(K)}$, in the case of typical examples in finance where p_X is known in closed form and sampling is at the monthly frequency. Since the performance of the approximation improves as Δ gets smaller, monthly sampling is taken to represent a worst-case scenario as the upper bound to the sampling interval relevant for finance. In practice, most continuous-time models in finance are estimated with monthly, weekly, daily, or higher frequency observations. The examples studied below reveal that including the term $c_2(y,y_0;\theta)$ generally provides an approximation to p_X which is better by a factor of at least 10 than what one obtains when only the term $c_1(y,y_0;\theta)$ is included. Further calculations show that each additional order produces additional improvements by an additional factor of at least 10.

I will often compare the expansion in this paper to the Euler approximation; the latter corresponds to a simple discretization of the continuous-time stochastic differential equation, where the differential equation (1) is replaced by the difference equation

$$X_{t+\Delta} - X_t = \mu(X_t;\theta)\Delta + \sigma(X_t;\theta)\sqrt{\Delta}\epsilon_{t+\Delta} \tag{15}$$

with $\epsilon_{t+\Delta} \sim N(0,1)$, so that

$$p_X^{\text{Euler}}(\Delta,x|x_0;\theta) = (2\pi\Delta\sigma^2(x_0;\theta))^{-1/2} \times \exp\{-(x - x_0 - \mu(x_0;\theta)\Delta)^2/2\Delta\sigma^2(x_0;\theta)\}. \tag{16}$$

Example 1 (Vasicek's Model): Consider the Ornstein–Uhlenbeck specification proposed by Vasicek (1977) for the short-term interest rate:

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t. \tag{17}$$

X is distributed on $D_X = (-\infty,+\infty)$ and has the Gaussian transition density

$$p_X(\Delta,x|x_0;\theta) = (\pi\gamma^2/\kappa)^{-1/2}\exp\{-(x - \alpha - (x_0 - \alpha)e^{-\kappa\Delta})^2\kappa/\gamma^2\}, \tag{18}$$

Table I

Explicit Sequence for the Vasicek Model

This table contains the coefficients of the density approximation for p_Y corresponding to the Vasicek model in Example 1, $dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t$. The terms in the expansion are evaluated by applying the formulas in equation (12). From equation (11), the $K = 0$ term in this expansion is $\tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)$, the $K = 1$ term is

$$\tilde{p}_Y^{(1)}(\Delta, y|y_0;\theta) = \tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)\{1 + c_1(y|y_0;\theta)\Delta\},$$

and the $K = 2$ term is

$$\tilde{p}_Y^{(2)}(\Delta, y|y_0;\theta) = \tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)\{1 + c_1(y|y_0;\theta)\Delta + c_2(y|y_0;\theta)\Delta^2/2\}.$$

Additional terms can be obtained in the same manner by applying equation (12) further. These computations and those of Tables II to V were all carried out in *Mathematica*.

$$\tilde{p}_Y^{(0)}(\Delta, y|y_0,\theta) = \frac{1}{\sqrt{\Delta}\sqrt{2\pi}} \exp\left[-\frac{(y-y_0)^2}{2\Delta} - \frac{y^2\kappa}{2} + \frac{y_0^2\kappa}{2} + \frac{y\alpha\kappa}{\sigma} - \frac{y_0\alpha\kappa}{\sigma}\right].$$

$$c_1(y|y_0,\theta) = -\frac{1}{6\sigma^2}\left(\kappa(3\alpha^2\kappa - 3(y+y_0)\alpha\kappa\sigma + (-3 + y^2\kappa + y y_0\kappa + y_0^2\kappa)\sigma^2)\right).$$

$$\begin{aligned} c_2(y|y_0,\theta) = \frac{1}{36\sigma^4}\big(&\kappa^2(9\alpha^4\kappa^2 - 18y\alpha^3\kappa^2\sigma + 3\alpha^2\kappa(-6 + 5y^2\kappa)\sigma^2 \\ &- 6y\alpha\kappa(-3 + y^2\kappa)\sigma^3 + (3 - 6y^2\kappa + y^4\kappa^2)\sigma^4 \\ &+ 2\kappa\sigma(-3\alpha + y\sigma)(3\alpha^2\kappa - 3y\alpha\kappa\sigma + (-3 + y^2\kappa)\sigma^2)y_0 \\ &+ 3\kappa\sigma^2(5\alpha^2\kappa - 4y\alpha\kappa\sigma + (-2 + y^2\kappa)\sigma^2)y_0^2 + 2\kappa^2\sigma^3(-3\alpha + y\sigma)y_0^3 \\ &+ \kappa^2\sigma^4 y_0^4)\big). \end{aligned}$$

where $\theta \equiv (\alpha, \kappa, \sigma)$ and $\gamma^2 \equiv \sigma^2(1 - e^{-2\kappa\Delta})$. In this case, we have that $Y_t = \gamma(X_t;\theta) = \sigma^{-1}X_t$ and $\mu_Y(y;\theta) = \kappa\alpha\sigma^{-1} - \kappa y$, so that $\lambda_Y(y;\theta) = \kappa/2 - \kappa^2(\alpha - \sigma y)^2/2\sigma^2$.

Table I reports the first two terms in the expansion for this model, obtained from applying the general formula in equation (11). More terms can be calculated in equation (12) one after the other: once $c_2(y|y_0;\theta)$ has been obtained, calculate $c_3(y|y_0;\theta)$, etc. Starting from the closed-form expression, one can show directly that these expressions indeed represent a Taylor series expansion for the closed-form density $p_X(\Delta, x|x_0;\theta)$.

Figure 1A plots the density p_X as a function of the interest rate value x for a monthly sampling frequency ($\Delta = 1/12$), evaluated at $x_0 = 0.10$ and for the parameter values corresponding to the maximum-likelihood estimator from the Federal funds data (see Table VI in Section IV below). Figure 1B plots

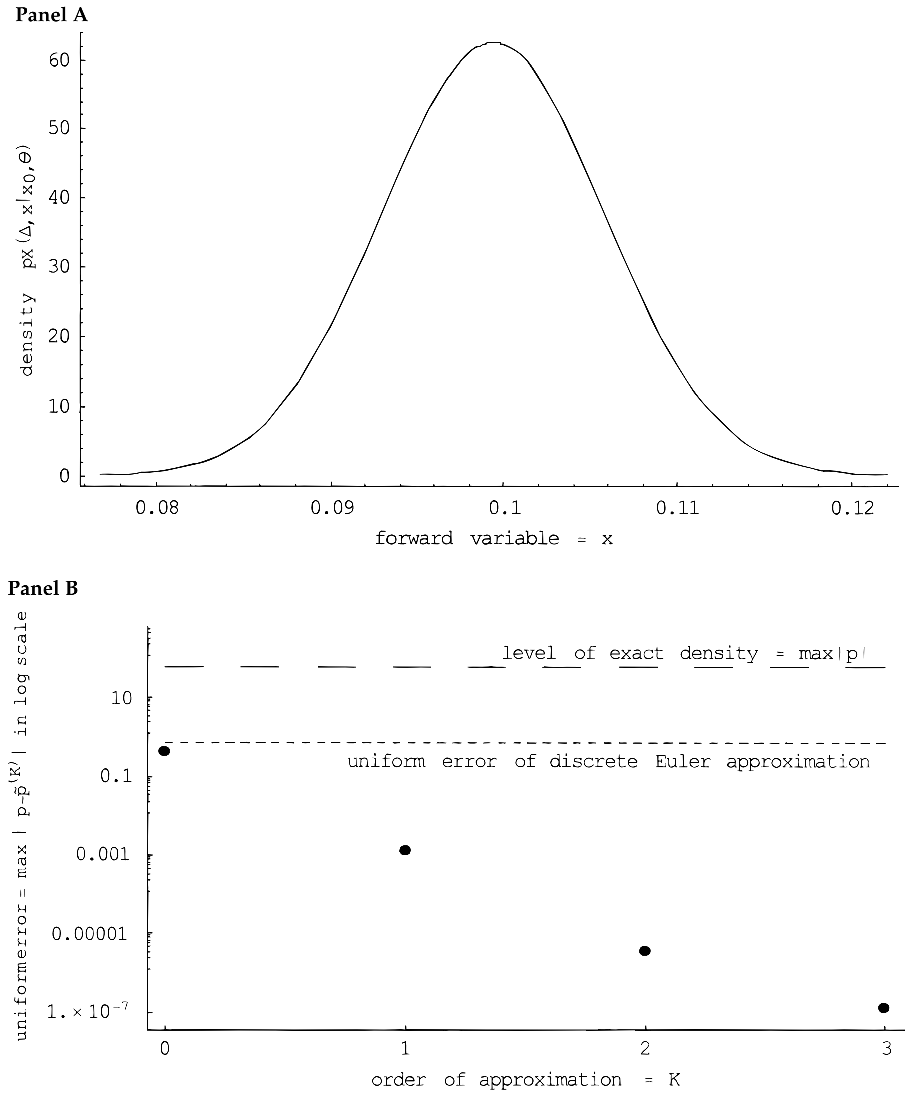


Figure 1. Exact conditional density and approximation errors for the Vasicek model. Figure 1A plots for the Vasicek (1977) model (see Example 1 and Table I) the closed-form conditional density $x \mapsto p_X(\Delta, x|x_0, \theta)$ as a function of x, with $x_0 = 10$ percent, monthly sampling ($\Delta = 1/12$) and θ replaced by the MLE reported in Table VI. Figure 1B plots the uniform approximation error $|p_X - \tilde{p}_X^{(K)}|$ for $K = 1$, 2, and 3, in log-scale, so that each unit on the y-axis corresponds to a reduction of the error by a multiplicative factor of 10. The error is calculated as the maximum absolute deviation between p_X and $\tilde{p}_X^{(K)}$ over the range ± 4 standard deviations around the mean of the density. Both the value of the exact conditional density at its peak and the uniform error for the Euler approximation p_X^{Euler} are also reported for comparison purposes. This figure illustrates the speed of convergence of the approximation. A lower sampling interval than monthly would provide an even faster convergence of the density approximation sequence.

the uniform approximation error $|p_X - \tilde{p}_X^{(K)}|$ for $K = 1$, 2, and 3, *in log-scale*. The error is calculated as the maximum absolute deviation between p_X and $\tilde{p}_X^{(K)}$ over the range ± 4 standard deviations around the mean of the density, and is also compared to the uniform error for the Euler approximation. The striking feature of the results is the speed of convergence to zero of the approximation error as K goes from 1 to 2 and from 2 to 3. In effect, one can approximate p_X (which is of order 10^{+1}) within 10^{-3} with the first term alone ($K = 1$) and within 10^{-7} with $K = 3$, even though the interest rate process is only sampled once a month. Similar calculations for a weekly sampling frequency ($\Delta = 1/52$) reveal that the approximation error gets smaller even faster for this lower value of Δ.

In other words, small values of K already produce extremely precise approximations to the true density, p_X, and the approximation is even more precise if Δ is smaller. Of course, the exact density being Gaussian, in this case the expansion, whose leading term is Gaussian, has fairly little "work" to do to approximate the true density. In the Ornstein–Uhlenbeck case, the expansion involves no correction for nonnormality, which is normally achieved through the change of variable X to Y; it reduces here to a linear transformation and therefore does not change the nature of the leading term in the expansion. Comparing the performance of the expansion to that of the Euler approximation in this model (where both have the correct Gaussian form for the density) reveals that the expansion is capable of correcting for the discretization bias involved in a discrete approximation, whereas the Euler approximation is limited to a first-order bias correction. In this case, the Euler approximation can be refined by increasing the precision of the conditional mean and variance approximations (see Huggins (1997)). Of course, discrete approximations to equation (1) of an order higher than equation (15) are available, but they do not lead to explicit *density* approximations since, compared to the Euler equation (15), they involve combinations of multiple powers of $\epsilon_{t+\Delta}$ (see, e.g., Kloeden and Platen (1992)).

Example 2 (The CIR Model): Consider Feller's (1952) square-root specification

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}\,dW_t, \tag{19}$$

proposed as a model for the short-term interest rate by Cox et al. (1985). X is distributed on $D_X = (0,+\infty)$ provided that $q \equiv 2\kappa\alpha/\sigma^2 - 1 \geq 0$. Its transition density is given by

$$p_X(\Delta, x|x_0;\theta) = ce^{-u-v}(v/u)^{q/2}I_q(2(uv)^{1/2}), \tag{20}$$

with $\theta \equiv (\alpha,\kappa,\sigma)$ all positive, $c \equiv 2\kappa/(\sigma^2\{1 - e^{-\kappa\Delta}\})$, $u \equiv cx_0e^{-\kappa\Delta}$, $v \equiv cx$, and I_q is the modified Bessel function of the first kind of order q. Here $Y_t = \gamma(X_t;\theta) = 2\sqrt{X_t}/\sigma$ and $\mu_Y(y;\theta) = (q + 1/2)/y - \kappa y/2$.

Table II

Explicit Sequence for the Cox–Ingersoll–Ross Model

This table contains the coefficients of the density approximation for p_Y corresponding to the Cox, Ingersoll, and Ross model in Example 2, $dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t$. The expansion for p_Y in this table applies also to the model proposed by Ahn and Gao (1988) (see Example 3). The terms in the expansion are evaluated by applying the formulas in equation (12). From equation (11), the $K = 0$ term in this expansion is $\tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)$, the $K = 1$ term is

$$\tilde{p}_Y^{(1)}(\Delta, y|y_0;\theta) = \tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)\{1 + c_1(y|y_0;\theta)\Delta\},$$

and the $K = 2$ term is

$$\tilde{p}_Y^{(2)}(\Delta, y|y_0;\theta) = \tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)\{1 + c_1(y|y_0;\theta)\Delta + c_2(y|y_0;\theta)\Delta^2/2\}.$$

Additional terms can be obtained in the same manner by applying equation (12) further.

$$\tilde{p}_X^{(0)}(\Delta, y|y_0,\theta) = \frac{1}{\sqrt{\Delta}\sqrt{2\pi}} \exp\left[-\frac{(y-y_0)^2}{2\Delta} - \frac{y^2\kappa}{4} + \frac{\kappa y_0^2}{4}\right] y^{-(1/2)+(2\alpha\kappa/\sigma^2)} y_0^{(1/2)-(2\alpha\kappa/\sigma^2)}.$$

$$c_1(y|y_0\theta) = -\frac{1}{24yy_0\sigma^4}\big(48\alpha^2\kappa^2 - 48\alpha\kappa\sigma^2 + 9\sigma^4 + y\kappa^2\sigma^2(-24\alpha + y^2\sigma^2)y_0 + y^2\kappa^2\sigma^4y_0^2 + y\kappa^2\sigma^4y_0^3\big).$$

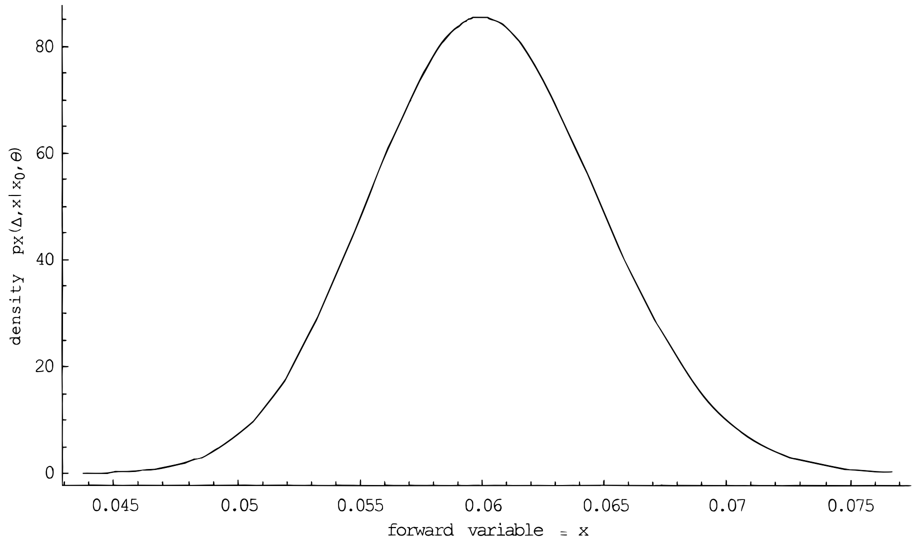
$$\begin{aligned} c_2(y|y_0\theta) = \frac{1}{576y^2y_0^2\sigma^8}\big(&9(256\alpha^4\kappa^4 - 512\alpha^3\kappa^3\sigma^2 + 224\alpha^2\kappa^2\sigma^4 + 32\alpha\kappa\sigma^6 - 15\sigma^8) \\ &+ 6y\kappa^2\sigma^2(-24\alpha + y^2\sigma^2)(16\alpha^2\kappa^2 - 16\alpha\kappa\sigma^2 + 3\sigma^4)y_0 \\ &+ y^2\kappa^2\sigma^4(672\alpha^2\kappa^2 - 48\alpha\kappa(2 + y^2\kappa)\sigma^2 + (-6 + y^4\kappa^2)\sigma^4)y_0^2 \\ &+ 2y\kappa^2\sigma^4(48\alpha^2\kappa^2 - 24\alpha\kappa(2 + y^2\kappa)\sigma^2 + (9 + y^4\kappa^2)\sigma^4)y_0^3 \\ &+ 3y^2\kappa^4\sigma^6(-16\alpha + y^2\sigma^2)y_0^4 + 2y^3\kappa^4\sigma^8y_0^5 + y^2\kappa^4\sigma^8y_0^6\big). \end{aligned}$$

The first two terms in the explicit expansion are given in Table II. When evaluated at the maximum-likelihood estimates from Fed funds data, the results reported in Figure 2 are very similar to those of Figure 1, again with an extremely fast convergence even for a monthly sampling frequency. The uniform approximation error is reduced to 10^{-5} with the first two terms, and 10^{-8} with the first three terms included.

Example 3 (Inverse of Feller's Square Root Model): In this example, I generate densities for Ahn and Gao's (1998) specification of the interest rate process as one over an auxiliary process that follows a Cox–Ingersoll–Ross specification. As a result of Itô's Lemma, the model's specification is

$$dX_t = X_t(\kappa - (\sigma^2 - \kappa\alpha)X_t)dt + \sigma X_t^{3/2}dW_t, \tag{21}$$

Panel A



Panel B

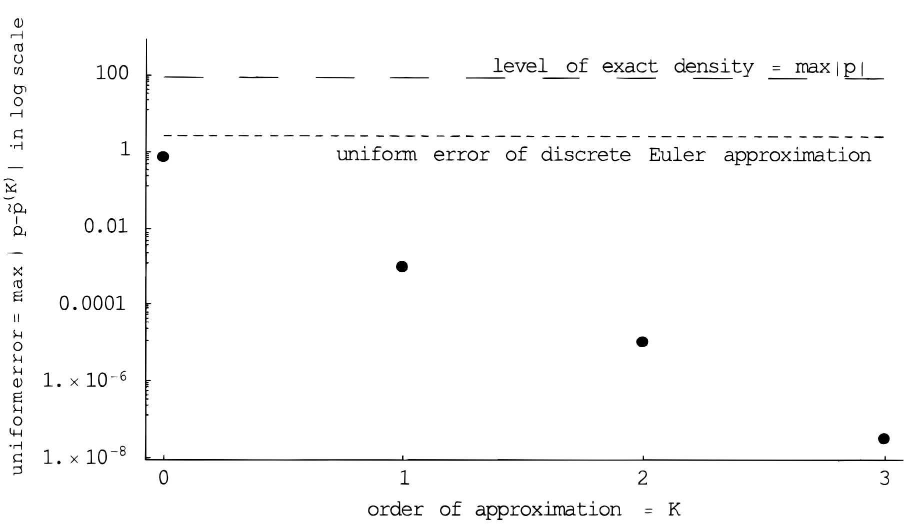


Figure 2. Exact conditional density and approximation errors for the Cox–Ingersoll–Ross model. Figure 2A plots for the CIR (1985) model (see Example 2 and Table II) the closed-form conditional density $x \mapsto p_X(\Delta, x|x_0, \theta)$ as a function of x, with $x_0 = 6$ percent, monthly sampling ($\Delta = 1/12$) and θ replaced by the MLE reported in Table VI. Figure 2B plots the uniform approximation error $|p_X - \tilde{p}_X^{(K)}|$ for $K = 1$, 2, and 3, in log-scale, so that each unit on the y-axis corresponds to a reduction of the error by a multiplicative factor of 10. The error is calculated as the maximum absolute deviation between p_X and $\tilde{p}_X^{(K)}$ over the range ± 4 standard deviations around the mean of the density. Both the value of the exact conditional density at its peak and the uniform error for the Euler approximation p_X^{Euler} are also reported for comparison purposes. This figure illustrates the speed of convergence of the approximation.

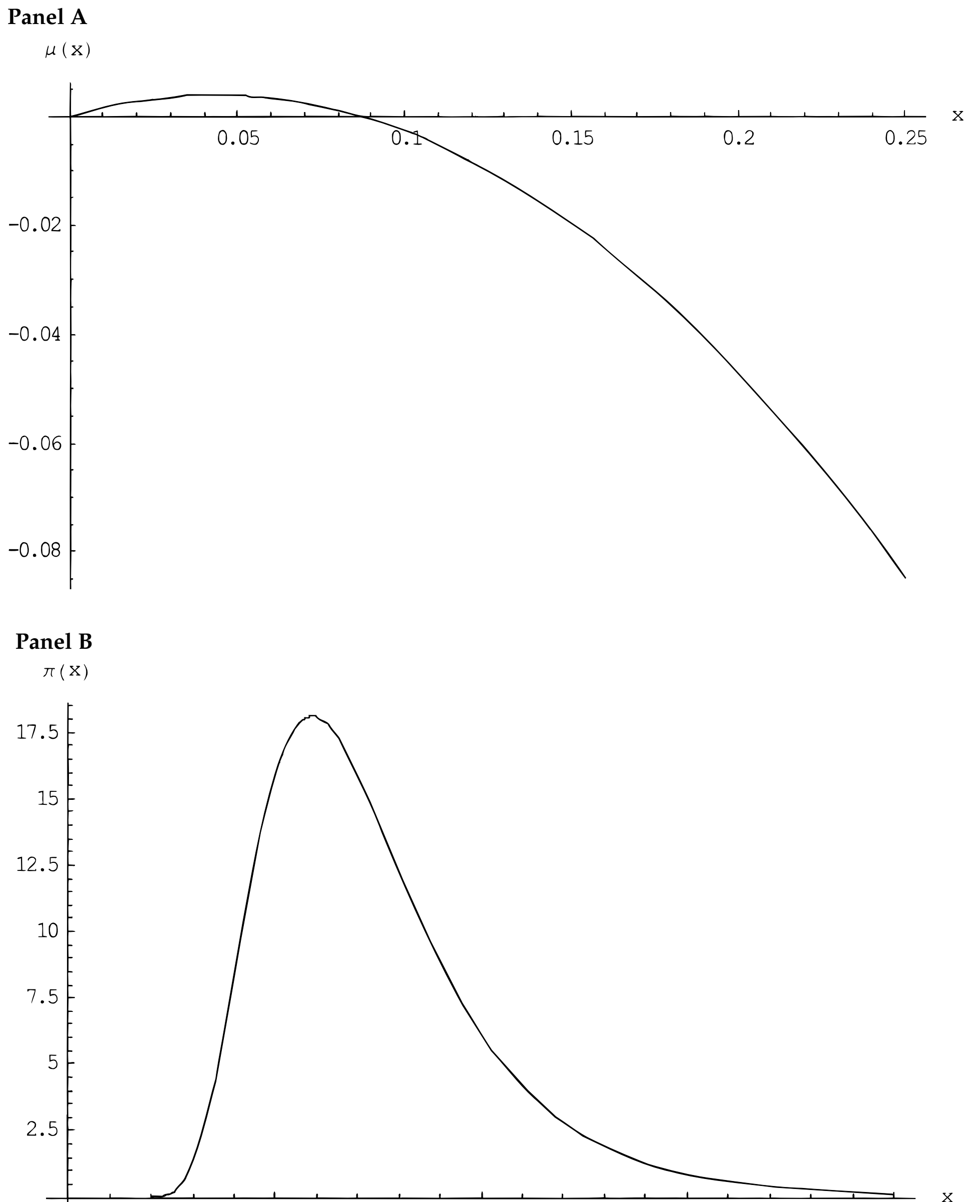
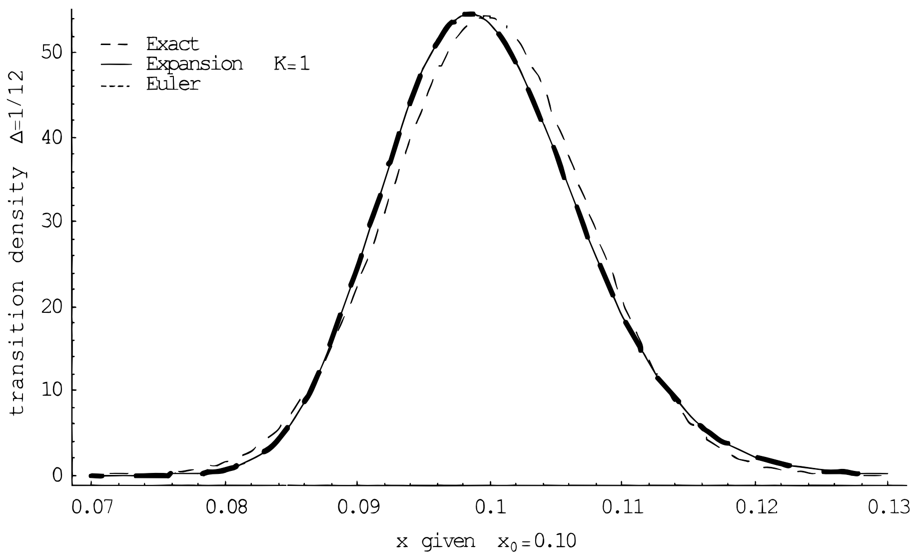


Figure 3. Drift, densities, and approximation errors for the inverse of Feller's process. Results for the model proposed by Ahn and Gao (1998) (see Example 3 and Table II) are reported: the drift $\mu(X_t,\theta) = X_t(\kappa - (\sigma^2 - \kappa\alpha)X_t)$ in Figure 3A, the marginal density $\pi(X_t,\theta)$ in Figure 3B, the exact and conditional density approximations, p_X, p_X^{Euler}, and $\tilde{p}_X^{(1)}$ as functions of the forward variable x, for $x_0 = 0.10$ in Figure 3C. The sampling frequency is monthly ($\Delta = 1/12$) and the parameter vector θ is evaluated at the MLE reported in Table VI. Figure 3D reports the uniform approximation error $|p_X - \tilde{p}_X^{(K)}|$ for $K = 1$, 2, and 3, in log-scale, as in Figures 1B and 2B.

Panel C



Panel D

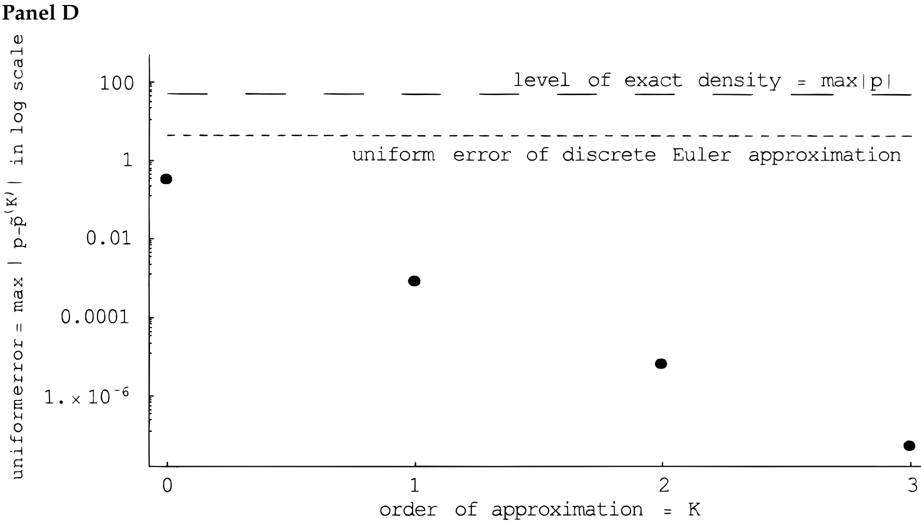


Figure 3. Continued

with closed-form transition density given by

$$p_X(\Delta, x | x_0; \theta) = (1/x^2) p_X^{\mathrm{CIR}}(\Delta, 1/x | 1/x_0; \theta), \tag{22}$$

where p_X^{CIR} is the density function given in equation (20). The expansion in equation (11) for p_Y is identical to that for the CIR model given in Table II (because the Y process is the same with the same transformed drift μ_Y and unit diffusion). To get back to an expansion for X, the change of variable $Y \to X$ however is different, and is now given by $Y_t = \gamma(X_t;\theta) = 2/(\sigma\sqrt{X_t})$; hence the expansion for p_X will naturally be different from that for the CIR model (it will now approximate the left-hand side of equation (22) rather than equation (20)).

Figure 3A reports the drift for this model, evaluated at the maximum-likelihood estimates from Table VI below. This model generates, in an environment where closed-form solutions are available, some of the effects documented empirically by Aït-Sahalia (1996b): almost no drift while the interest rate is in the middle of its range, strong mean-reversion when the interest rate gets large. Figure 3B plots the unconditional or marginal density, which is also the stationary density $\pi(x,\theta)$ for this process when the initial data point X_0 has π as its distribution. π is given by

$$\pi(y;\theta) \equiv \exp\left\{2\int^{y} \mu_Y(u;\theta)\,du\right\}\Bigg/\int_{\underline{y}}^{\bar{y}} \exp\left\{2\int^{v} \mu_Y(u;\theta)\,du\right\} dv. \tag{23}$$

Figure 3C compares the exact conditional density in equation (22), its Euler approximation, and the expansion with $K = 1$ for the conditioning interest rate $x_0 = 0.10$. It is apparent from the figure that including the first term alone is sufficient to make the exact and approximate densities fall on top of one another, whereas the Euler approximation is distinct. Finally, Figure 3D reports the uniform approximation error between the Euler approximation and the exact density on the one hand, and between the first three terms in the expansion and the exact density on the other. As can be seen from these figures, the expansion in equation (11) provides again a very accurate approximation to the exact density.

B. Density Approximation for Models with No Closed-Form Density

Of course, the usefulness of the method introduced in Aït-Sahalia (1998) lies largely in its ability to deliver explicit density approximations for models that do not have closed-form transition densities. The next two examples correspond to models recently proposed in the literature to describe the time series properties of the short-term interest rate, and the final example illustrates the applicability of the method to a double-well model where the stationary density is bimodal.

Example 4 (Linear Drift, CEV Diffusion): Chan et al. (1992) have proposed the specification

$$dX_t = \kappa(\alpha - X_t)\,dt + \sigma X_t^{\rho}\,dW_t, \tag{24}$$

with $\theta \equiv (\alpha,\kappa,\sigma,\rho)$. X is distributed on $(0,+\infty)$ when $\alpha > 0$, $\kappa > 0$, and $\rho > 1/2$ (if $\rho = 1/2$; see Example 2 for an additional constraint). This model does not admit a closed-form density unless $\alpha = 0$ (see Cox (1996)), which

Table III

Explicit Sequence for the Linear Drift, CEV Diffusion Model

This table contains the coefficients of the density approximation for p_Y corresponding to the Chan et al. (1992) model in Example 4, $dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^{\rho} dW_t$. The terms in the expansion are evaluated by applying the formulas in equation (12). From equation (11), the $K = 0$ term in this expansion is $\tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)$, the $K = 1$ term is

$$\tilde{p}_Y^{(1)}(\Delta, y|y_0;\theta) = \tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)\{1 + c_1(y|y_0;\theta)\Delta\}.$$

Additional terms can be obtained by applying equation (12) further.

$$\begin{aligned}\tilde{p}_Y^{(0)}(\Delta, y|y_0,\theta) = \frac{1}{\sqrt{\Delta}\sqrt{2\pi}} \exp\Bigg[-\frac{(y-y_0)^2}{2\Delta} + \kappa(\rho-1) \\ \times \big(y^2(2\rho-1) - 2y^{1+(\rho/(\rho-1))}\alpha(\rho-1)^{\rho/(\rho-1)}\sigma^{1/(\rho-1)} \\ + y_0(y_0 - 2\rho y_0 + 2\alpha(\rho-1)^{\rho/(\rho-1)}\sigma^{1/(\rho-1)}y_0^{\rho/(\rho-1)})\big)/(4\rho-2)\Bigg] \\ \times y^{\rho/(-2+2\rho)}y_0^{\rho/(2-2\rho)}.\end{aligned}$$

$$\begin{aligned}c_1(y|y_0,\theta) \text{ for } y \neq y_0 = \big(&-4y^4\kappa^2(\rho-1)^4(2-9\rho+9\rho^2)y_0 + 3\rho(4+20\rho+27\rho^2-9\rho^3) \\ &\times y_0 - 12y^2\kappa(\rho-1)^2(13\rho-27\rho^2+18\rho^3-2) \\ &\times y_0 + 24y^{3+(\rho/(\rho-1))}\alpha\kappa^2(\rho-1)^{4+(\rho/(\rho-1))}(3\rho-1)\sigma^{1/(\rho-1)} \\ &\times y_0 + 24y^{1+(\rho/(\rho-1))}\alpha\kappa(\rho-1)^{3+(1/(\rho-1))}(2-9\rho+9\rho^2)\sigma^{1/(\rho-1)} \\ &\times y_0 - 12y^{2+2(\rho/(\rho-1))}\alpha^2\kappa^2(\rho-1)^{5+(2/(\rho-1))}(3\rho-2)\sigma^{2/(\rho-1)} \\ &\times y_0 + y\big(3\rho(20\rho-27\rho^2+9\rho^3-4) \\ &\quad + 12\kappa(\rho-1)^2(13\rho-27\rho^2+18\rho^3-2)y_0^2 \\ &\quad + 4\kappa^2(\rho-1)^4(2-9\rho+9\rho^2)y_0^4 \\ &\quad - 24\alpha\kappa(\rho-1)^{3+(1/(\rho-1))}(2-9\rho+9\rho^2)\sigma^{1/(\rho-1)}y_0^{1+(\rho/(\rho-1))} \\ &\quad - 24\alpha\kappa^2(\rho-1)^{4+(\rho/(\rho-1))}(3\rho-1)\sigma^{1/(\rho-1)}y_0^{3+(\rho/(\rho-1))} \\ &\quad + 12\alpha^2\kappa^2(\rho-1)^{5+(2/(\rho-1))}(3\rho-2)\sigma^{2/(\rho-1)} \\ &\quad \times y_0^{2+(2\rho/(\rho-1))}\big)\big)/\big(24y(\rho-1)^2(3\rho-2)(3\rho-1)(y-y_0)y_0\big).\end{aligned}$$

$$\begin{aligned}c_1(y|y_0,\theta) \text{ for } y = y_0 = \frac{1}{8(\rho-1)^2y_0^2}\big(&(\rho-2)\rho - 4\kappa(\rho-1)^2(2\rho-1)y_0^2 - 4\kappa^2(\rho-1)^4y_0^4 \\ &+ 8\alpha\kappa(\rho-1)^{2+(1/(\rho-1))}\rho\sigma^{1/(-1+\rho)}y_0^{1+(\rho/(\rho-1))} \\ &+ 8\alpha\kappa^2(\rho-1)^{3+(\rho/(\rho-1))}\sigma^{1/(\rho-1)}y_0^{3+(\rho/(\rho-1))} \\ &- 4\alpha^2\kappa^2(\rho-1)^{4+(2/(\rho-1))}\sigma^{2/(\rho-1)}y_0^{2+(2\rho/(\rho-1))}\big).\end{aligned}$$

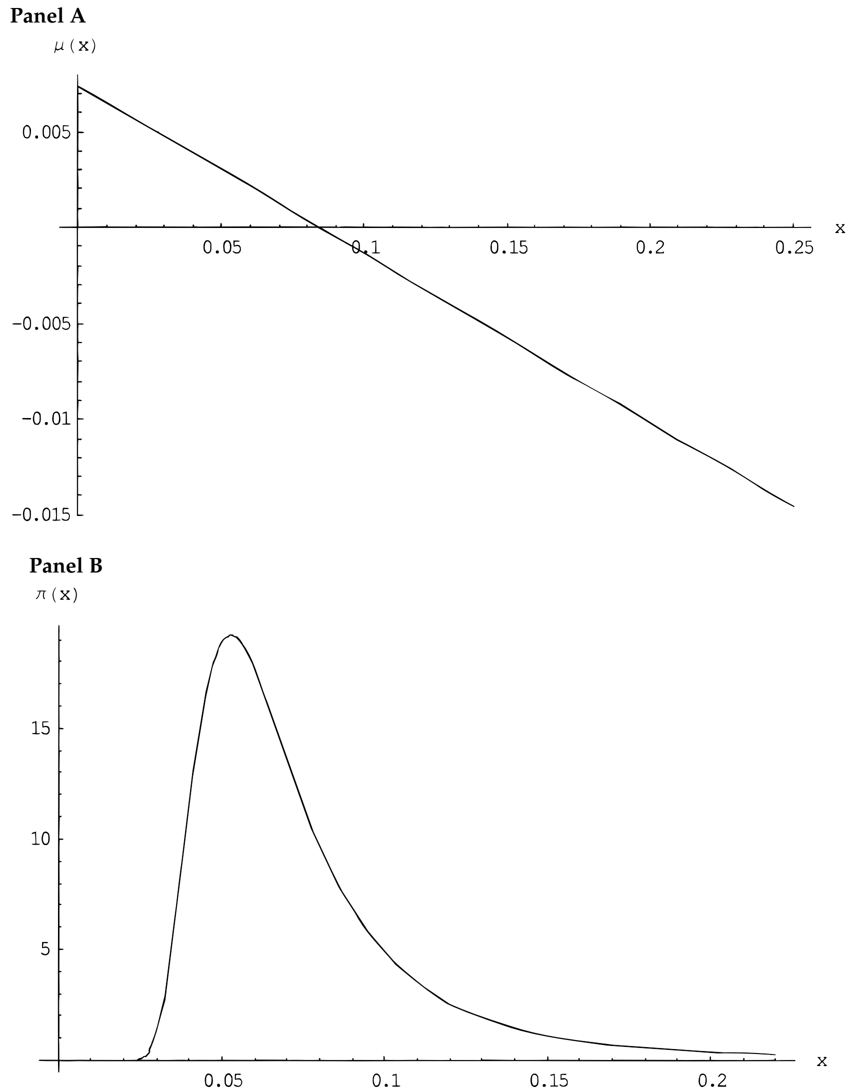
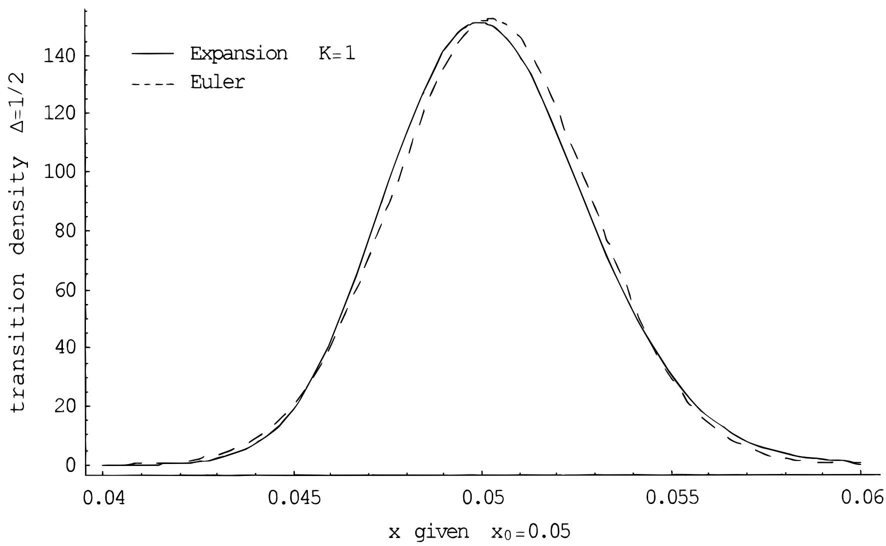


Figure 4. Conditional density approximations for the linear drift, CEV diffusion model. These figures plot for the linear drift, CEV diffusion model of Chan et al. (1992) (see Example 4 and Table III) the drift function, $\mu(X_t,\theta) = \kappa(\alpha - X_t)$ (Figure 4A), the marginal density $\pi(X_t,\theta)$ (Figure 4B), and the conditional density approximations p_X^{Euler} and $\tilde{p}_X^{(1)}$ as functions of the forward variable x, for two values of the conditioning variable x_0 in Figures 4C and 4D respectively. The sampling frequency is monthly ($\Delta = 1/12$) and the parameter vector θ is evaluated at the MLE reported in Table VI.

then makes it unrealistic for interest rates. I will concentrate on the case where $\rho > 1$, which corresponds to the empirically plausible estimate for U.S. interest rate data. The transformation from X to Y is given by $Y_t = \gamma(X_t;\theta) = X_t^{1-\rho}/\{\sigma(\rho - 1)\}$ and

Panel C



Panel D

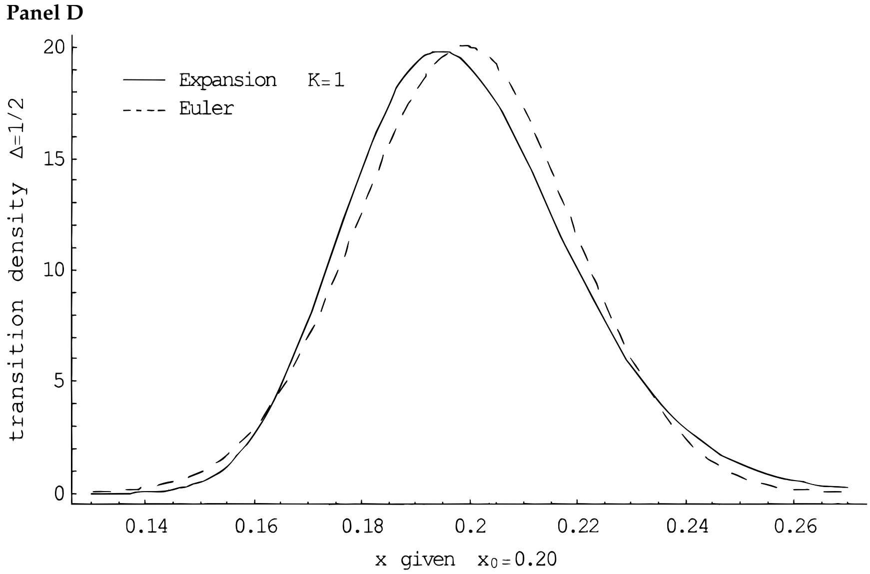


Figure 4. Continued.

$$\mu_Y(y;\theta) = \frac{\rho}{2(\rho-1)y} - \kappa(\rho-1)y + \alpha\kappa\sigma^{1/(\rho-1)}(\rho-1)^{\rho/(\rho-1)}y^{\rho/(\rho-1)}. \quad (25)$$

The first term in the expansion is given in Table III. The corresponding formulas can be derived analogously for the transformation $Y_t = \gamma(X_t;\theta) = X_t^{1-\rho}/\{\sigma(1-\rho)\}$, which is appropriate if $1/2 < \rho < 1$. I plot in Figure 4A the drift function corresponding to maximum-likelihood estimates (based on the expansion with $K = 1$, see Table VI below), in Figure 4B I plot the unconditional density, and in Figures 4C and 4D the conditional density approximations for monthly sampling at $x_0 = 0.05$ and 0.20, respectively.

Example 5 (Nonlinear Mean Reversion): The following model was designed to produce very little mean reversion while interest rate values remain in the middle part of their domain, and strong nonlinear mean reversion at either end of the domain (see Aït-Sahalia (1996b)):

$$dX_t = (\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2)\,dt + \sigma X_t^{\rho}\,dW_t, \tag{26}$$

with $\theta \equiv (\alpha_{-1},\alpha_0,\alpha_1,\alpha_2,\sigma,\rho)$. This model has been estimated empirically by Aït-Sahalia (1996b), Conley et al. (1997), and Gallant and Tauchen (1998) using a variety of empirical techniques. The new method in this paper makes it possible to estimate it using maximum likelihood. I again concentrate on the case where $\rho > 1$, and to save space I evaluate the formulas in Table IV for $\rho = 3/2$. This process has $D_X = (0,+\infty)$, $Y_t = \gamma(X_t;\theta) = 2/(\sigma\sqrt{X_t})$, and

$$\mu_Y(y;\theta) = \frac{3/2 - 2\alpha_2/\sigma^2}{y} - \frac{\alpha_1 y}{2} - \frac{\alpha_0\sigma^2 y^3}{8} - \frac{\alpha_{-1}\sigma^4 y^5}{32}. \tag{27}$$

Figure 5A plots the drift evaluated at the maximum-likelihood parameter estimates (corresponding to $K = 1$). Figure 5B plots the unconditional or marginal density of the process: in the specification test in Aït-Sahalia (1996b), this density is matched against a nonparametric kernel estimator. Figures 5C and 5D contain the conditional density approximations for $K = 1$, compared with the Euler approximation, for the two values $x_0 = 0.025$ and 0.20, respectively. As before, sampling is at the monthly frequency.

Example 6 (Double-Well Potential): In this example, I generate a bimodal stationary density through the specification

$$dX_t = (X_t - X_t^3)\,dt + dW_t. \tag{28}$$

This model is distributed on $D_X = (-\infty,+\infty)$. Since the model is already set in unit diffusion, no transformation is needed ($Y = X$).

Table V contains the first two terms of the expansion; Figure 6A plots its drift, Figure 6B its marginal density, and Figures 6C and 6D the transition density for $K = 2$, monthly sampling, and $x_0 = 0.0$ and 0.5, respectively, with $\Delta = 1/2$. As is apparent from the figures, the densities in this model exhibit strong nonnormality, which obviously cannot be captured by the Euler approximation of equation (16).

Table IV

Explicit Sequence for the Nonlinear Drift Model

This table contains the coefficients of the density approximation for p_Y corresponding to the model in Aït-Sahalia (1996b), Conley et al. (1997), and Tauchen (1997) given in Example 5, $dX_t = (\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2)\,dt + \sigma X_t^{\rho}\,dW_t$ with $\rho = 3/2$. The terms in the expansion are evaluated by applying the formulas in equation (12). From equation (11), the $K = 0$ term in this expansion is $\tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)$ and the $K = 1$ term is

$$\tilde{p}_Y^{(1)}(\Delta, y|y_0;\theta) = \tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)\{1 + c_1(y|y_0;\theta)\Delta\}.$$

Additional terms can be obtained by applying equation (12) further.

$$\tilde{p}_X^{(0)}(\Delta, y|y_0,\theta) = \frac{1}{\sqrt{\Delta}\sqrt{2\pi}} \exp\left[-\frac{(y-y_0)^2}{2\Delta} + \frac{1}{192}\big(\sigma^4(-y^6 + y_0^6)\alpha_{-1} - 6(y^2 - y_0^2)(\sigma^2(y^2 + y_0^2)\alpha_0 + 8\alpha_1)\big)\right] \times y^{(3/2)-(2\alpha_2/\sigma^2)} y_0^{-(3/2)+(2\alpha_2/\sigma^2)}.$$

$$\begin{aligned}
c_1(y|y_0\theta) = -\frac{1}{7096320 y\sigma^4 y_0}\big(&315 y\sigma^{12} y_0 (y^{10} + y^9 y_0 + y^8 y_0^2 + y^7 y_0^3 + y^6 y_0^4 + y^5 y_0^5 + y^4 y_0^6 \\
&+ y^3 y_0^7 + y^2 y_0^8 + y y_0^9 + y_0^{10})\alpha_{-1}^2 + 88 y\sigma^6 y_0 \alpha_{-1} \\
&\times \big(35\sigma^4(y^8 + y^7 y_0 + y^6 y_0^2 + y^5 y_0^3 + y^4 y_0^4 + y^3 y_0^5 \\
&+ y^2 y_0^6 + y y_0^7 + y_0^8) \\
&\times \alpha_0 + 36(-56y^4\sigma^2 - 56y^3\sigma^2 y_0 - 56y^2\sigma^2 y_0^2 - 56y\sigma^2 y_0^3 \\
&- 56\sigma^2 y_0^4 + 5y^6\sigma^2\alpha_1 + 5y^5\sigma^2 y_0\alpha_1 + 5y^4\sigma^2 y_0^2\alpha_1 \\
&+ 5y^3\sigma^2 y_0^3\alpha_1 + 5y^2\sigma^2 y_0^4\alpha_1 + 5y\sigma^2 y_0^5\alpha_1 \\
&+ 5\sigma^2 y_0^6\alpha_1 + 28y^4\alpha_2 + 28y^3 y_0\alpha_2 + 28y^2 y_0^2\alpha_2 \\
&+ 28 y y_0^3\alpha_2 + 28 y_0^4\alpha_2)\big) \\
&+ 528\big(15y\sigma^8 y_0(y^6 + y^5 y_0 + y^4 y_0^2 + y^3 y_0^3 + y^2 y_0^4 + y y_0^5 + y_0^6) \\
&\times \alpha_0^2 + 56y\sigma^4 y_0\alpha_0(-30y^2\sigma^2 - 30y\sigma^2 y_0 - 30\sigma^2 y_0^2 \\
&+ 3y^4\sigma^2\alpha_1 + 3y^3\sigma^2 y_0\alpha_1 \\
&+ 3y^2\sigma^2 y_0^2\alpha_1 + 3y\sigma^2 y_0^3\alpha_1 + 3\sigma^2 y_0^4\alpha_1 \\
&+ 20y^2\alpha_2 + 20y y_0\alpha_2 + 20y_0^2\alpha_2) \\
&+ 560(9\sigma^4 - 24y\sigma^4 y_0\alpha_1 + y^3\sigma^4 y_0\alpha_1^2 + y^2\sigma^4 y_0^2\alpha_1^2 \\
&+ y\sigma^4 y_0^3\alpha_1^2 - 48\sigma^2\alpha_2 + 24y\sigma^2 y_0\alpha_1\alpha_2 + 48\alpha_2^2)\big)\big).
\end{aligned}$$

Panel A

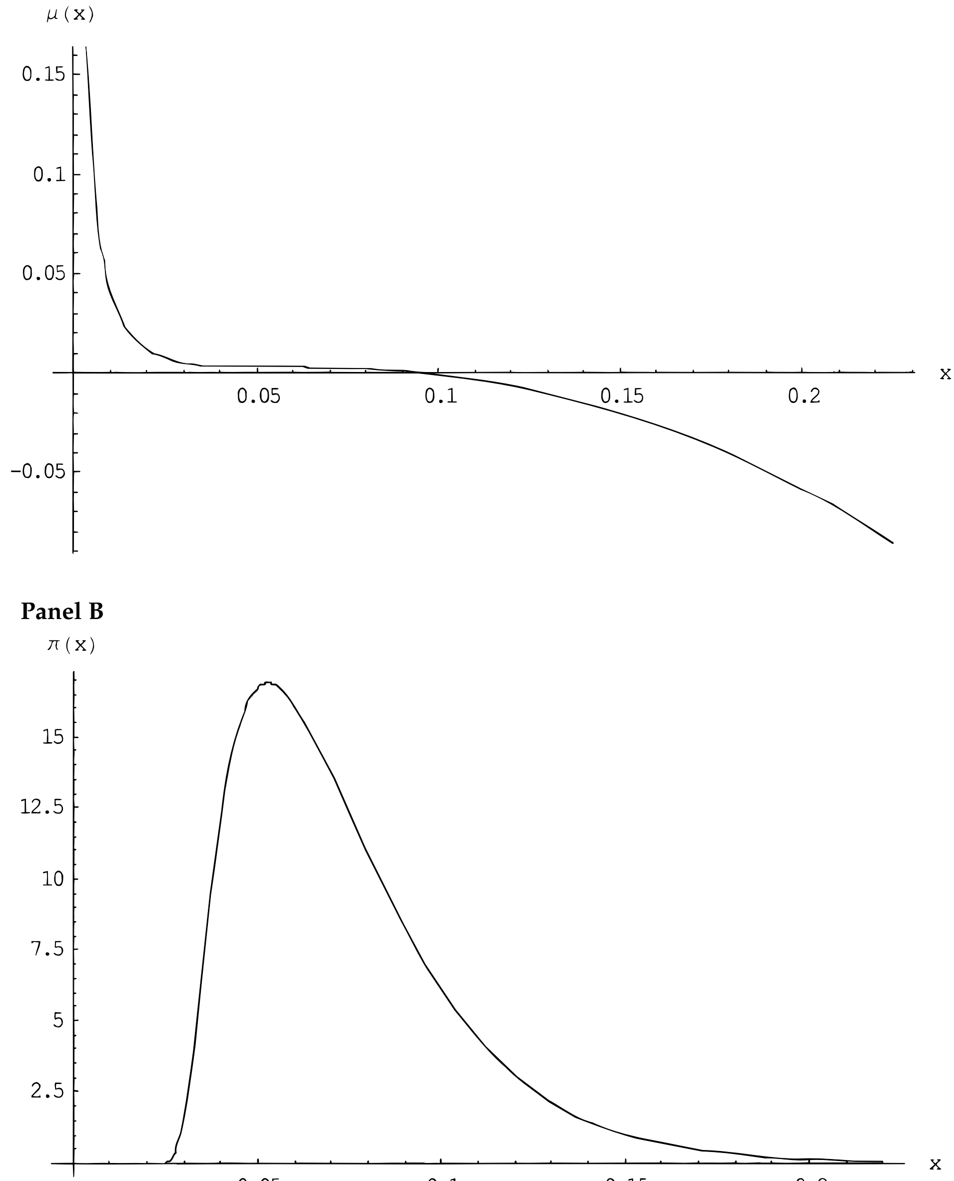
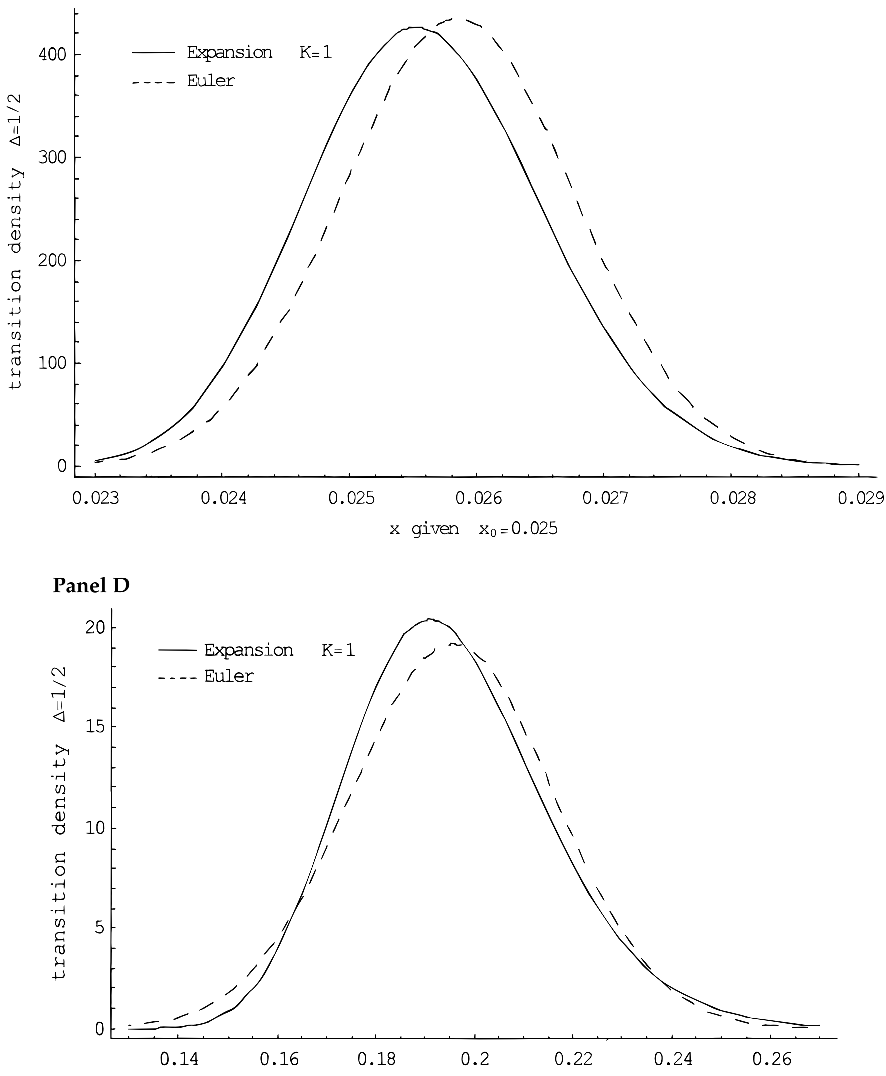


Figure 5. Drift and density approximations for the nonlinear drift model. These figures report results for the nonlinear drift model of Aït-Sahalia (1996b) (also estimated by Conley et al. (1997) and Gallant and Tauchen (1998)) described in Example 5 and Table IV. Figure 5A plots the drift function, $\mu(X_t,\theta) = \alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2$ and Figure 5B the marginal density $\pi(X_t,\theta)$. This model does not have a closed-form solution for p_X. Figures 5C and 5D plot the conditional density approximations p_X^{Euler} and $\tilde{p}_X^{(1)}$ as functions of the forward variable x, for two different values of the conditioning variable x_0. The sampling frequency is monthly ($\Delta = 1/12$) and the parameter vector θ is evaluated at the MLE reported in Table VI.

Panel C



Panel D

Figure 5. Continued

III. The Estimation of Interest Rate Diffusions

A. The Data and Maximum-Likelihood Estimates

To calculate approximate maximum-likelihood estimates, I maximize the approximate log-likelihood function

Table V

Explicit Sequence for the Double-Well Model

This table contains the coefficients of the density approximation for p_Y corresponding to the model in Example 6, $dX_t = (X_t - X_t^3)\,dt + dW_t$. The terms in the expansion are evaluated by applying the formulas in equation (12). From equation (11), the $K = 0$ term in this expansion is $\tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)$, the $K = 1$ term is

$$\tilde{p}_Y^{(1)}(\Delta, y|y_0;\theta) = \tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)\{1 + c_1(y|y_0;\theta)\Delta\},$$

and the $K = 2$ term is

$$\tilde{p}_Y^{(2)}(\Delta, y|y_0;\theta) = \tilde{p}_Y^{(0)}(\Delta, y|y_0;\theta)\{1 + c_1(y|y_0;\theta)\Delta + c_2(y|y_0;\theta)\Delta^2/2\}.$$

Additional terms can be obtained in the same manner by applying equation (12) further.

$$\tilde{p}_X^{(0)}(\Delta, y|y_0,\theta) = \frac{1}{\sqrt{\Delta}\sqrt{2\pi}}\exp\left[-\frac{(y-y_0)^2}{2\Delta} + \frac{y^2}{2} - \frac{y^4}{4} - \frac{y_0^2}{2} + \frac{y_0^4}{4}\right].$$

$$\begin{aligned} c_1(y|y_0,\theta) = \frac{1}{210}\big(&-105 + 70y^2 + 42y^4 - 15y^6 + (70y + 42y^3 - 15y^5)y_0 + (70 + 42y^2 - 15y^4)y_0^2 \\ &+ (42y - 15y^3)y_0^3 + (42 - 15y^2)y_0^4 - 15yy_0^5 - 15y_0^6\big). \end{aligned}$$

$$\begin{aligned} c_2(y|y_0,\theta) = \frac{1}{44100}\big(&25725 + 11760y^2 - 19670y^4 + 9030y^6 - 336y^8 - 1260y^{10} + 225y^{12} \\ &+ 2y(10290 - 12110y^2 + 7455y^4 - 336y^6 - 1260y^8 + 225y^{10})y_0 \\ &+ 3(3920 - 7490y^2 + 6930y^4 - 336y^6 - 1260y^8 + 225y^{10})y_0^2 \\ &+ 2y(-12110 + 10395y^2 + 378y^4 - 2520y^6 + 450y^8)y_0^3 \\ &+ 5(-3934 + 4158y^2 + 504y^4 - 1260y^6 + 225y^8)y_0^4 \\ &+ 6y(2485 + 126y^2 - 1050y^4 + 225y^6)y_0^5 \\ &+ 21(430 - 48y^2 - 300y^4 + 75y^6)y_0^6 + 6y(-112 - 840y^2 + 225y^4)y_0^7 \\ &+ 3(-112 - 1260y^2 + 375y^4)y_0^8 + 180y(-14 + 5y^2)y_0^9 \\ &+ 45(-28 + 15y^2)y_0^{10} + 450yy_0^{11} + 225y_0^{12}\big). \end{aligned}$$

$$\ell_n^{(K)}(\theta) \equiv n^{-1}\sum_{i=1}^{n}\ln\{\tilde{p}_X^{(K)}(\Delta, X_{i\Delta}|X_{(i-1)\Delta};\theta)\} \tag{29}$$

(with the convention that $\ln(\alpha) = -\infty$ if $\alpha < 0$) over θ in Θ. This results in an estimator $\hat{\theta}_n^{(K)}$, which, as shown in Aït-Sahalia (1998), is close to the exact (but uncomputable in practice) maximum-likelihood estimator $\hat{\theta}_n$.

The data consist of monthly sampling of the Fed funds rate between January 1963 and December 1998 (see Figure 7). The source for the data is the H-15 Federal Reserve Statistical Release (Selected Interest Rate Series).

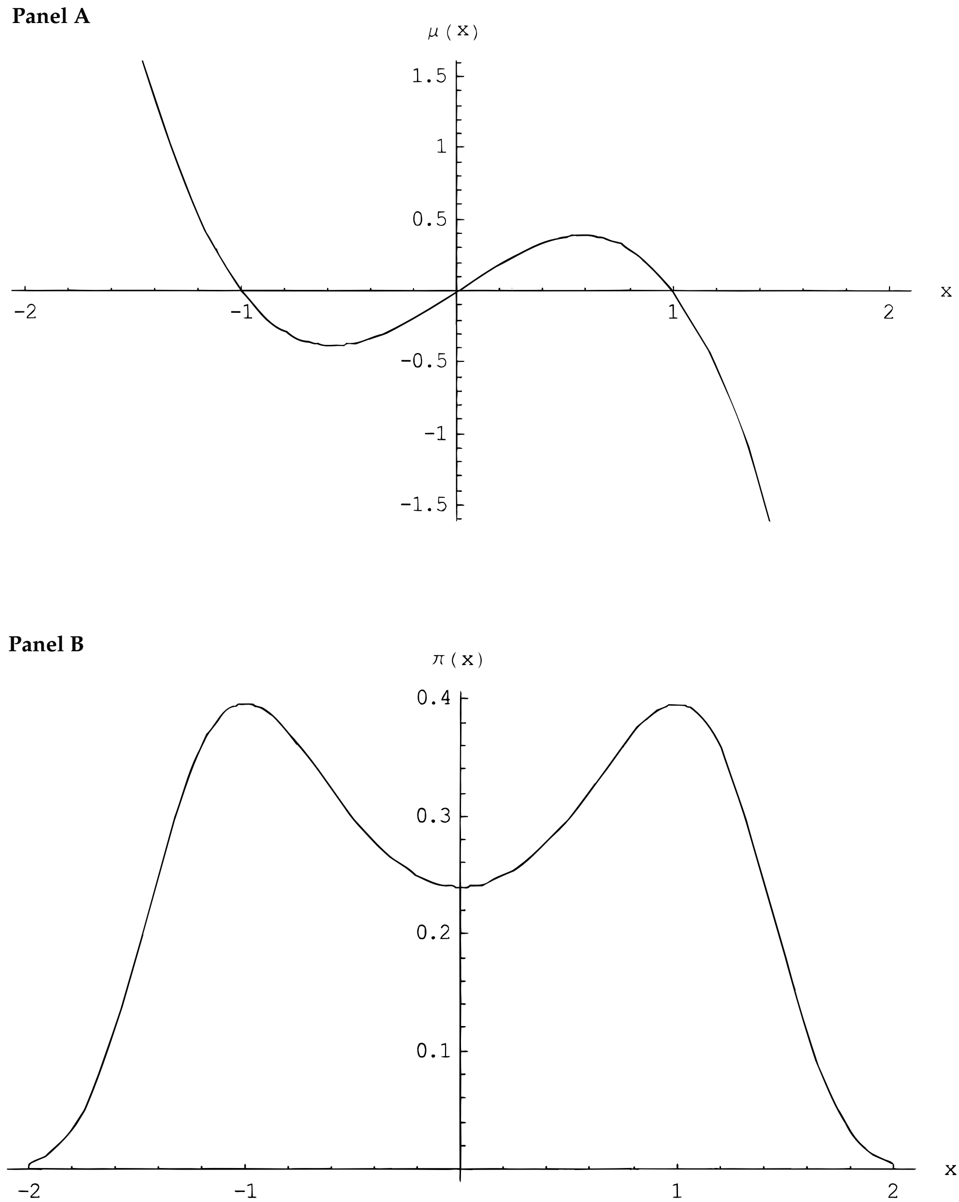


Figure 6. Drift and densities for the double-well model. Results for the double-well model of Example 6 and Table V are reported. The drift function $\mu(x) = x - x^3$ (Figure 6A) is such that the process avoids staying near 0 and is attracted to either -1 or $+1$, a fact reflected by the bimodality of the marginal density $\pi(x)$ in Figure 6B. This model does not have closed-form solutions for p_X. Figures 6C and 6D plot the conditional density approximations p_X^{Euler} and $\tilde{p}_X^{(2)}$ as functions of the forward variable x, for two different values of the conditioning variable x_0 with $\Delta = 1/2$. As is clear from these figures, the Euler approximation cannot reflect the substantial nonnormality captured by the density approximation of this paper. Figure 6E plots the conditional density surface, $(x, x_0) \mapsto \tilde{p}_X^{(2)}(\Delta, x|x_0, \theta)$ for $\Delta = 1/2$, and θ replaced by the MLE.

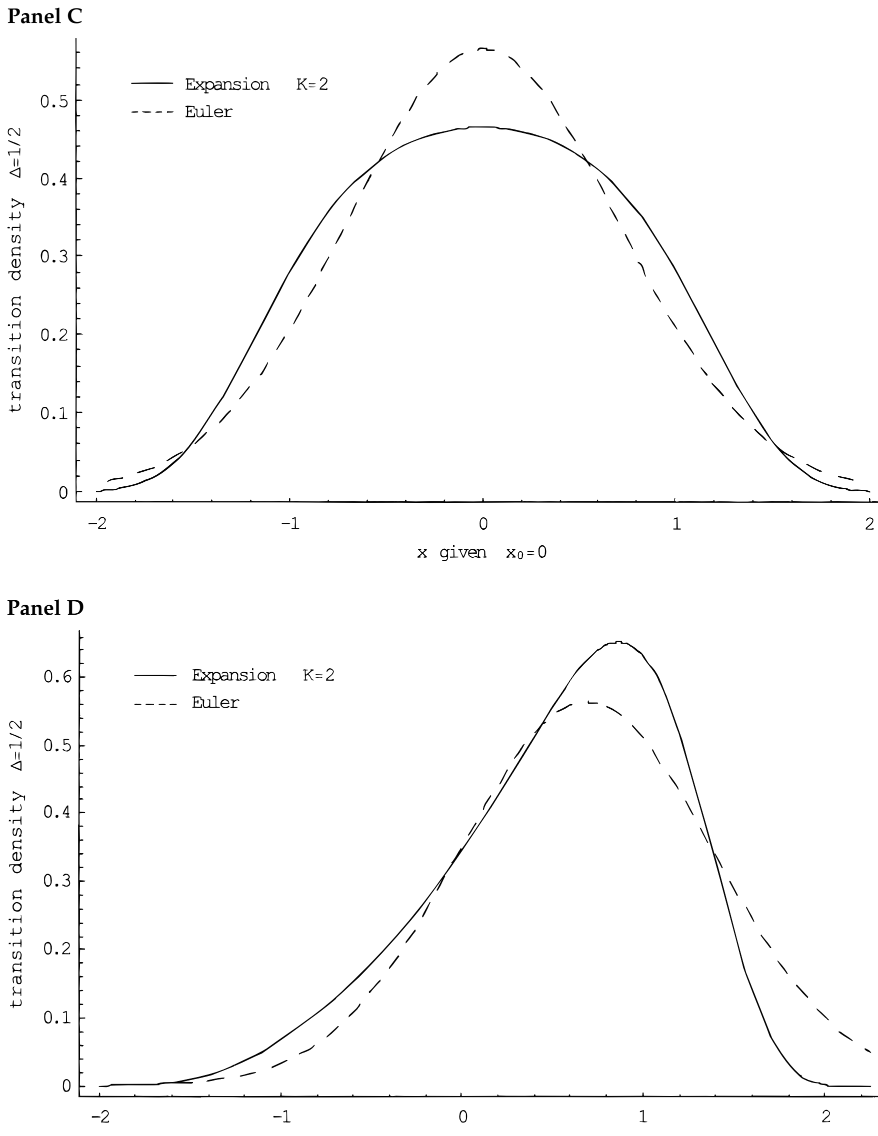


Figure 6. Continued.

Though the Fed funds rate series exhibits strong microstructure effects at the daily frequency (due for instance to the second Wednesday settlement effect; see Hamilton (1996)), these effects are largely mitigated at the monthly frequency. On the other hand, this rate represents one of the closest possible proxies for what is meant by an "instantaneous" short rate in theoretical models. Since the method in this paper does not rely on the sampling inter-

Panel E

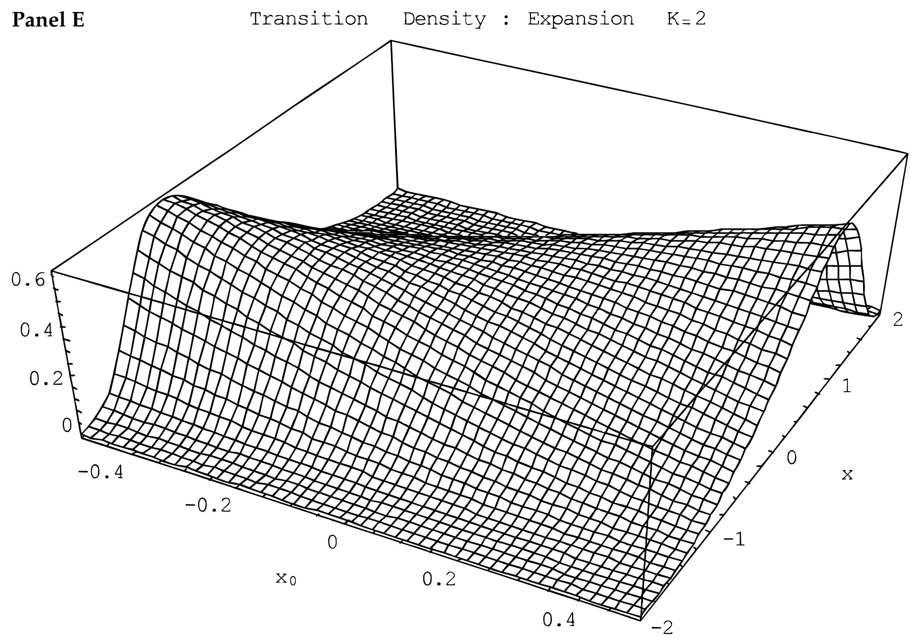


Figure 6. Continued

val being small, the trade-off between a larger sampling interval and the virtual absence of microstructure effects seems worthwhile. Of course, the implicit (unrealistic) assumption is made that a single diffusion specification can represent the evolution of the short rate for the entire period. Naturally, nothing prevents the estimation from being conducted on a shorter time period at the expense of reducing the sample size. One advantage of the long time series used here is that it contains different episodes of U.S. interest rate history, such as the Volcker period, as well as the low interest rate environments that preceded it and followed it. It is therefore interesting to see how different models would accommodate these different regimes.

The results for the five models of Examples 1 to 5 compared to the Euler approximation and, when available (Examples 1 to 3), the true log-likelihood, are reported in Table VI. The last column of the table reports the asymptotic standard deviations for the estimated parameters, derived as explained below.

The results in Table VI confirm those of Section II: the expansion used with $K = 1$ or 2 produces estimates $\hat{\theta}_n^{(K)}$ that are very close to $\hat{\theta}_n$. It is interesting to note that because the models evaluated at the true parameter values often display very little drift (hence their near unit root behavior), and because interest rates are not particularly volatile, the fitted densities over a one-month interval are often fairly close to a Gaussian density. In other words, for these data, Δ = one month is a "small" time interval. Hence, the Euler approximation performs relatively well in this specific context (except

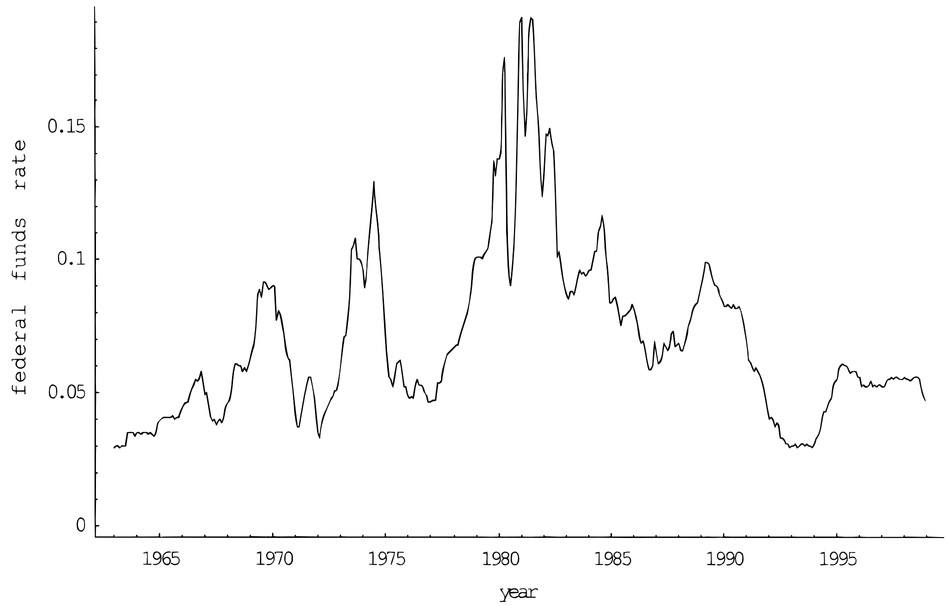


Figure 7. Federal funds rate, monthly frequency, 1963–1998. This figure plots the time series of the Federal funds data used for the estimation of the parameters in Table VI.

in the nonlinear drift model of Example 5, where the estimated parameters can be off by as much as 30 percent (although the standard deviation in this case is large), in the inverse Feller process of Example 3 where they are off by 5 to 10 percent, and in the Chan et al. (1992) specification of Example 4 where the drift parameters are off by 10 percent).

B. Estimation of the Asymptotic Variance and How Many Terms to Include

I consider here only the situation where the process admits a stationary distribution. For the more general case, see Aït-Sahalia (1998). The asymptotic variance of the maximum-likelihood estimator is given by the inverse of Fisher's Information Matrix, which is the lowest possible achievable variance among the competing estimators discussed in the Introduction.

Define $L(\theta) \equiv \ln(p_X(\Delta, X_\Delta | X_0; \theta))$, the $d \times 1$ vector $\dot{L}(\theta) \equiv \partial L(\theta)/\partial\theta$, and the $d \times d$ matrix $\ddot{L}(\theta) \equiv \partial^2 L(\theta)/\partial\theta\partial\theta^T$, where ^{T} denotes transposition. We have that

$$n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, i(\theta_0)^{-1}), \tag{30}$$

where Fisher's Information Matrix is

$$i(\theta) \equiv E[\dot{L}(\theta)\dot{L}(\theta)^T] = -E[\ddot{L}(\theta)]. \tag{31}$$

Table VI

Maximum-Likelihood Estimates for the Monthly Federal Funds Data, 1963–1998

This table reports the MLE for the parameters of five interest rate models estimated using the Fed funds data, monthly from January 1963 through December 1998. The estimates are calculated using the Euler approximation, the density approximation of this paper with $K = 1$, and, when the transition density is available in closed-form (Examples 1, 2 and 3), the expansion with $K = 2$ and the true density. In the table, "$\ln L$" refers to the maximized value of the log-likelihood. The formulas for the density expansion can be found in the respective tables indicated in the third column. The asymptotic standard errors in the last column are computed from equation (30), with Fisher's Information Matrix in equation (31) replaced by the sample averages evaluated at the second derivative of the log-likelihood expansion with $K = 1$, and confirmed with the average of the first derivative squared.

Model	Example Number	Density Expansion Table	Figure	Parameter Estimates: Euler	Parameter Estimates: Expansion $K = 1$	Parameter Estimates: Expansion $K = 2$	Parameter Estimates: True Density	Asymptotic Standard Error
$dX_t = \kappa(\alpha - X_t)\,dt + \sigma\,dW_t$	1	I	1	$\alpha = 0.0717$	$\alpha = 0.0719$	$\alpha = 0.0717$	$\alpha = 0.0717$	α: 0.014
				$\kappa = 0.258$	$\kappa = 0.257$	$\kappa = 0.261$	$\kappa = 0.261$	κ: 0.12
				$\sigma = 0.02213$	$\sigma = 0.02237$	$\sigma = 0.02237$	$\sigma = 0.02237$	σ: 0.00078
				$\ln L = 3.634$	$\ln L = 3.634$	$\ln L = 3.634$	$\ln L = 3.634$	
$dX_t = \kappa(\alpha - X_t)\,dt + \sigma\sqrt{X_t}\,dW_t$	2	II	2	$\alpha = 0.0732$	$\alpha = 0.0742$	$\alpha = 0.0742$	$\alpha = 0.0721$	α: 0.016
				$\kappa = 0.145$	$\kappa = 0.189$	$\kappa = 0.189$	$\kappa = 0.219$	κ: 0.10
				$\sigma = 0.06521$	$\sigma = 0.06658$	$\sigma = 0.06658$	$\sigma = 0.06665$	σ: 0.0023
				$\ln L = 3.917$	$\ln L = 3.918$	$\ln L = 3.918$	$\ln L = 3.918$	
$dX_t = X_t(\kappa - (\sigma^2 - \kappa\alpha)X_t)\,dt + \sigma X_t^{3/2}\,dW_t$	3	II	3	$\alpha = 15.019$	$\alpha = 15.157$	$\alpha = 15.150$	$\alpha = 15.141$	α: 2.9
				$\kappa = 0.177$	$\kappa = 0.181$	$\kappa = 0.182$	$\kappa = 0.182$	κ: 0.1
				$\sigma = 0.8059$	$\sigma = 0.8211$	$\sigma = 0.8211$	$\sigma = 0.8211$	σ: 0.03
				$\ln L = 4.171$	$\ln L = 4.158$	$\ln L = 4.158$	$\ln L = 4.158$	
$dX_t = \kappa(\alpha - X_t)\,dt + \sigma X_t^{\rho}\,dW_t$	4	III	4	$\alpha = 0.0808$	$\alpha = 0.0844$			α: 0.05
				$\kappa = 0.0972$	$\kappa = 0.0876$			κ: 0.11
				$\sigma = 0.7224$	$\sigma = 0.7791$			σ: 0.16
				$\rho = 1.46$	$\rho = 1.48$			ρ: 0.08
				$\ln L = 4.172$	$\ln L = 4.159$			
$dX_t = (\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2)\,dt$ $+ \sigma X_t^{3/2}\,dW_t$	5	IV	5	$\alpha_{-1} = 0.00107$	$\alpha_{-1} = 0.000693$			α_{-1}: 0.002
				$\alpha_0 = -0.0517$	$\alpha_0 = -0.0347$			α_0: 0.09
				$\alpha_1 = 0.877$	$\alpha_1 = 0.676$			α_1: 1.3
				$\alpha_2 = -4.604$	$\alpha_2 = -4.059$			α_2: 6.4
				$\sigma = 0.8047$	$\sigma = 0.8214$			σ: 0.03
				$\ln L = 4.173$	$\ln L = 4.160$			

Note that it is necessary that the transition function p_X not be uniformly flat in the direction of any one of the parameters θ_m, $m = 1,\ldots,d$, otherwise $\partial p_X(\Delta, x|x_0;\theta)/\partial\theta_m \equiv 0$ for all (x, x_0) and the model cannot be identified. In other words, no parameter entering the likelihood function can be redundant. The asymptotic standard deviations from equation (30) are reported in the last column of Table VI for the interest rate models estimated above, with the expected values in equation (31) replaced by the sample averages evaluated at the MLE.

Test statistics can be derived. Suppose that the model is given by equation (1) and that we wish to test $H_0: \theta = \theta_0$ against the two-sided alternative $H_a: \theta \neq \theta_0$. The likelihood ratio test statistic evaluated behaves under H_0 as:

$$2\{\ell_n(\hat{\theta}_n) - \ell_n(\theta_0)\} \xrightarrow{d} \chi^2_d. \tag{32}$$

Distributional results can also be obtained for tests of a nested model that only allows for $\bar{d}$ free parameters from the d parameters in θ, and one can also consider Rao's efficient score statistic, which depends only on the restricted estimator $\bar{\theta}_n$, and Wald's test statistic, which depends only on the unrestricted estimator $\hat{\theta}_n$.

In all the results above, one can then replace $\hat{\theta}_n$ (respectively $\bar{\theta}_n$) by $\hat{\theta}_n^{(K)}$ (respectively $\bar{\theta}_n^{(K)}$). As the examples above have shown, it is not necessary to go much beyond $K = 2$ in the relevant financial examples to estimate the true density with a high degree of precision. More generally, to select an appropriate K at which to stop adding terms to the expansion, the following approach can be adopted: take K large enough so that the *approximation error* made in replacing p_X by $\tilde{p}_X^{(K)}$ is smaller than the *sampling error* due to the random character of the data, by a predetermined factor.

That is, in

$$||\hat{\theta}_n^{(K)} - \theta_0|| \leq ||\hat{\theta}_n^{(K)} - \hat{\theta}_n|| + ||\hat{\theta}_n - \theta_0|| \tag{33}$$

we can estimate the asymptotic standard variance of $\hat{\theta}_n$ about θ_0 by equation (30). By Chebyshev's Inequality, one can then bound the second term on the right-hand-side of equation (33). We can then stop considering higher order approximations at an order K such that the distance between the two successive estimates $\hat{\theta}_n^{(K)}$ and $\hat{\theta}_n^{(K-1)}$ is an order of magnitude smaller than the distance between $\hat{\theta}_n$ and θ_0. In practice, this is unlikely to make much of a difference and in most cases one can safely restrict attention to the first two terms, $K = 1$ and $K = 2$.

IV. Conclusion

This paper has demonstrated how to obtain very accurate closed-form approximations to the respective transition densities of a variety of models commonly used to represent the dynamics of the short-term interest rate.

Applications to derivative pricing, consisting of obtaining pricing formulas for any underlying price process, have been briefly outlined and will be developed in future work. Finally, an extension of these results to multivariate diffusions will be investigated.

Appendix: Regularity Conditions

ASSUMPTION 1 (Smoothness of the coefficients): *The functions* $\mu(x;\theta)$ *and* $\sigma(x;\theta)$ *are infinitely differentiable in x in* D_X*, and twice continuously differentiable in* θ *in the parameter space* $\Theta \subset R^d$.

ASSUMPTION 2 (Nondegeneracy of the diffusion):

1. *If* $D_X = (-\infty, +\infty)$*, there exists a constant c such that* $\sigma(x;\theta) > c > 0$ *for all* $x \in D_X$ *and* $\theta \in \Theta$.
2. *If* $D_X = (0, +\infty)$*, I allow for the possible local degeneracy of* σ *at* $x = 0$*: If* $\sigma(0;\theta) = 0$*, then there exist constants* ξ_0*,* $\omega \geq 0$*,* $\rho \geq 0$ *such that* $\sigma(x;\theta) \geq \omega x^{\rho}$ *for all* $0 < x < \xi_0$ *and* $\theta \in \Theta$*. Away from 0,* σ *is nondegenerate; that is, for each* $\xi > 0$*, there exists a constant* c_ξ *such that* $\sigma(x;\theta) \geq c_\xi > 0$ *for all* $x \in [\xi\ +\infty)$ *and* $\theta \in \Theta$.

Assumption 3 below restricts the behavior of the function μ_Y and its derivatives near the boundaries of D_Y. It is formulated in terms of the function μ_Y for reasons of convenience, but the equivalent formulation directly in terms of the original functions μ and σ can be obtained from equation (8). Recall that $\lambda_Y(y;\theta) \equiv -(\mu_Y^2(y;\theta) + \partial\mu_Y(y;\theta)/\partial y)/2$.

ASSUMPTION 3 (Boundary behavior): *For all* $\theta \in \Theta$*,* $\mu_Y(y;\theta)$*,* $\partial\mu_Y(y;\theta)/\partial y$*, and* $\partial^2\mu_Y(y;\theta)/\partial y^2$ *have at most exponential growth near the infinity boundaries and* $\lim_{y \to \underline{y} \text{ or } \bar{y}} \lambda_Y(y;\theta) < +\infty$.

1. *Left Boundary:*
   i. *If* $\underline{y} = 0^+$*, there exist constants* ϵ_0*,* κ*,* α *such that for all* $0 < y \leq \epsilon_0$ *and* $\theta \in \Theta$*,* $\mu_Y(y;\theta) \geq \kappa y^{-\alpha}$ *where either* $\alpha > 1$ *and* $\kappa > 0$*, or* $\alpha = 1$ *and* $\kappa \geq 1$.
   ii. *If* $\underline{y} = -\infty$*, there exist constants* $E_0 > 0$ *and* $K > 0$ such that for all $y \leq -E_0$ *and* $\theta \in \Theta$*,* $\mu_Y(y;\theta) \geq Ky$.
2. *Right Boundary: If* $\bar{y} = +\infty$*, there exist constants* $E_0 > 0$ *and* $K > 0$ *such that for all* $y \geq E_0$ *and* $\theta \in \Theta$*,* $\mu_Y(y;\theta) \leq Ky$.

The following remarks can help demonstrate the generality of these assumptions:

1. The upper bound $\lim_{y \to \underline{y} \text{ or } \bar{y}} \lambda_Y(y;\theta) < +\infty$ does not restrict λ_Y from going to $-\infty$ near the boundaries.
2. Similarly, Assumption 3 does not preclude μ_Y from going to $-\infty$ very fast near $\bar{y}$, and similarly, from going to $+\infty$ very fast near $\underline{y}$. Assumption 3 only restricts how large μ_Y can grow if it has the "wrong" sign;

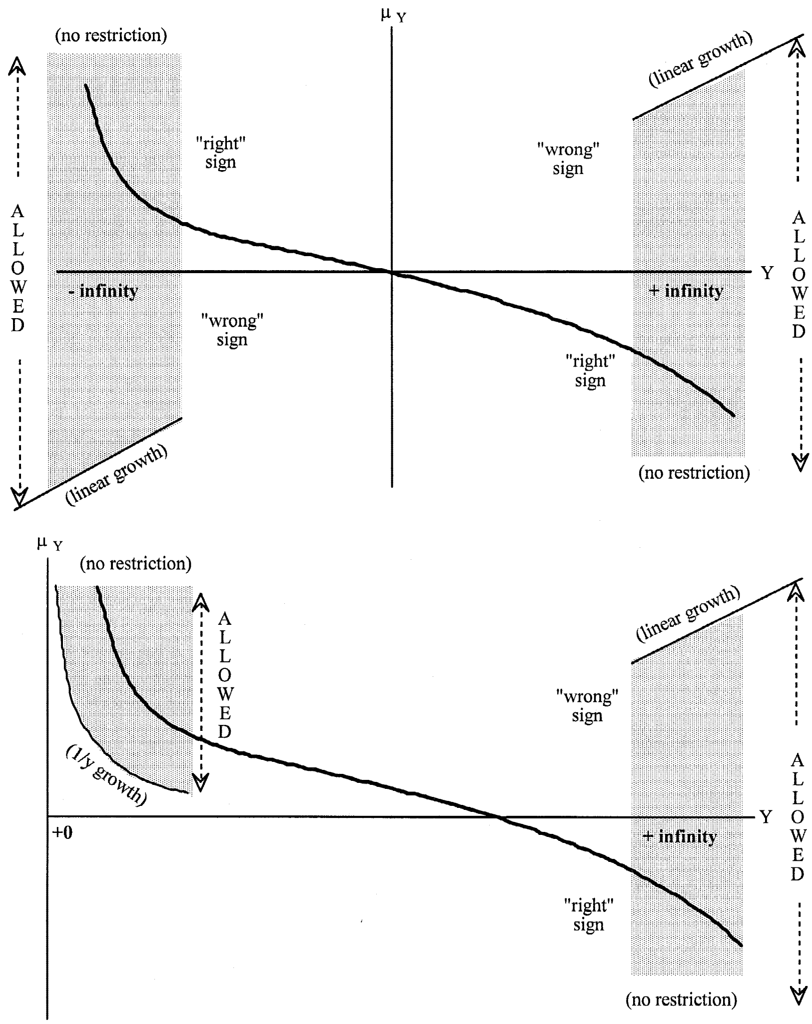


Figure A1. Growth conditions for the drift $\mu_Y(Y;\theta)$. This figure translates graphically the assumptions made in the Appendix regarding the shape of the function μ_Y. The admissible shape of the function is substantially less restricted than under the standard growth conditions. In particular, I only restrict the growth of μ_Y when it has the "wrong" sign (positive near $+\infty$, negative near $-\infty$).

that is, if it is positive near $\bar{y}$ and negative near $\underline{y}$ then linear growth is at the maximum possible growth rate. If μ_Y has the "right" sign then the process is being pulled back away from the boundary and I do not restrict how fast mean reversion occurs (up to an exponential rate for technical reasons). The admissible behavior of the drift function μ_Y under these assumptions is summarized in Figure A1.

3. The constraints on the behavior of the function μ_Y are essentially the best possible. For example, if μ_Y has the "wrong" sign near an infinity boundary, and grows faster than linearly, then Y explodes in finite time. Near a zero boundary at 0^+, if there exist $\kappa > 0$ and $\alpha < 1$ such that $\mu_Y(y;\theta) \leq ky^{-\alpha}$ in a neighborhood of 0^+ then 0 and negative values become attainable.
4. I can now fully characterize the boundary behavior of the diffusion Y implied by the assumptions made: if $+\infty$ is a boundary then it is natural if, near $+\infty$, $|\mu_Y(y;\theta)| \leq Ky$ and entrance if $\mu_Y(y;\theta) \leq -Ky^{\beta}$ for some $\beta > 1$. If $-\infty$ is a boundary then it is natural if, near $-\infty$, $|\mu_Y(y;\theta)| \leq K|y|$ and entrance if $\mu_Y(y;0) \geq K|y|^{\beta}$ for some $\beta > 1$. If 0^+ is a boundary, then it is entrance.

   Both entrance and natural boundaries are unattainable (see Feller (1952) or Karlin and Taylor (1981, Sec. 15.6) for the definition of boundaries). Natural boundaries can neither be reached in finite time, nor can the diffusion be started from there. Entrance boundaries, such as 0^+, cannot be reached starting from an interior point in $D_Y = (0,+\infty)$, but it is possible for Y to begin there. In that case, the process moves quickly away from 0 and never returns there. Typically, economic intuition says little about how the process would behave if it were to start at the boundary, or whether that is even possible, and hence it is sensible to allow both types of boundary behavior.
5. Assumption 3 neither requires nor implies that the process is stationary. When *both* boundaries of the domain D_Y are entrance boundaries then the process is necessarily stationary with unconditional (marginal) density,

$$\pi(y;\theta) \equiv \exp\left\{2\int^{y} \mu_Y(u;\theta)\,du\right\} \Big/ \int_{\underline{y}}^{\bar{y}} \exp\left\{2\int^{v} \mu_Y(u;\theta)\,du\right\} dv, \tag{A.1}$$

   provided that the initial random variable Y_0 is itself distributed with the same density π. When at least one of the boundaries is natural, stationarity is neither precluded nor implied. For instance, both an Ornstein–Uhlenbeck process, where $\mu_Y(y;\theta) = \kappa(\alpha - y)$, and a standard Brownian motion, where $\mu_Y(y;\theta) = 0$, satisfy the assumptions made, and both have natural boundaries at $-\infty$ and $+\infty$. Yet the former process is stationary, due to mean reversion, while the latter (null recurrent) is not.

Finally, the following assumption is needed for the purpose of maximizing the log-likelihood function only, not for the purpose of constructing the density expansion in equation (11).

Assumption 4 (Strengthening of Assumption 2 in the limiting case where $\alpha = 1$ and the diffusion is degenerate at 0): *Recall the constant* ρ *in Assumption 2(2), and the constants* α *and* κ *in Assumption 3(1.i). If* $\alpha = 1$, *then either* $\rho \geq 1$ *with no restriction on* κ, *or* $\kappa \geq 2\rho/(1 - \rho)$ *if* $0 < \rho < 1$. *If* $\alpha > 1$, *no restriction is required.*

REFERENCES

Ahn, Dong-Hyun, and Bin Gao, 1998, A parametric nonlinear model of term structure dynamics, Working paper, University of North Carolina at Chapel Hill.

Aït-Sahalia, Yacine, 1996a, Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, 527–560.

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

Aït-Sahalia, Yacine, 1998, Maximum-likelihood estimation of discretely sampled diffusions: A closed-form approach, Working paper, Princeton University.

Bibby, Bo M., and Michael Sørenson, 1995, Martingale estimation functions for discretely observed diffusion processes, *Bernoulli* 1, 17–39.

Black, Fisher, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–654.

Chan, K. C., G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209–1227.

Conley, Timothy G., Lars P. Hansen, Erzo G. J. Luttmer, and José A. Scheinkman, 1997, Short-term interest rates as subordinated diffusions, *Review of Financial Studies* 10, 525–578.

Cox, John C., 1996, The constant elasticity of variance option pricing model, *The Journal of Portfolio Management*, Special issue.

Cox, John C., John E. Ingersoll, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–407.

Cox, John C., and Stephen A. Ross, 1976, The valuation of options for alternative stochastic processes, *Journal of Financial Economics* 3, 145–166.

Dacunha-Castelle, Didier, and Danielle Florens-Zmioru, 1986, Estimation of the coefficients of a diffusion from discrete observations, *Stochastics* 19, 263–284.

Duffie, Darrell, and Peter Glynn, 1997, Estimation of continuous-time Markov processes sampled at random time intervals, Working paper, Stanford University.

Duffie, Darrell, and Kenneth Singleton, 1993, Simulated moments estimation of Markov models of asset prices, *Econometrica* 61, 929–952.

Elerian, Ola, Sidartha Chib, and Neil Shephard, 1998, Likelihood inference for discretely observed non-linear diffusions, Working paper, Oxford University.

Eraker, Bjorn, 1997, MCMC analysis of diffusion models with application to finance, Working paper, Norwegian School of Economics, Bergen.

Feller, William, 1952, The parabolic differential equations and the associated semi-groups of transformations, *Annals of Mathematics* 55, 468–519.

Florens, Jean-Pierre, Eric Renault, and Nizar Touzi, 1995, Testing for embeddability by stationary scalar diffusions, *Econometric Theory*, forthcoming.

Gallant, A. Ronald, and George Tauchen, 1998, Reprojecting partially observed systems with an application to interest rate diffusions, *Journal of the American Statistical Association* 93, 10–24.

Gihman, I. I., and A. V. Skorohod, 1972, *Stochastic Differential Equations* (Springer-Verlag, New York).

Gouriéroux, Christian, Alain Monfort, and Eric Renault, 1993, Indirect inference, *Journal of Applied Econometrics* 8, S85–S118.

Hamilton, James D, 1996, The daily market for Federal funds, *Journal of Political Economy* 104, 26–56.

Hansen, Lars P., and José A. Scheinkman, 1995, Back to the future: Generating moment implications for continuous time Markov processes, *Econometrica* 63, 767–804.

Hansen, Lars P., José A. Scheinkman, and Nizar Touzi, 1998, Identification of scalar diffusions using eigenvectors, *Journal of Econometrics* 86, 1, 1–32.

Honoré, Peter, 1997, Maximum-likelihood estimation of non-linear continuous-time term structure models, Working paper, Aarhus University.

Huggins, Douglas J., 1997, Estimation of a diffusion process for the U.S. short interest rate using a semigroup pseudo-likelihood, Ph.D. Dissertation, University of Chicago.

Jarrow, Robert, and Andrew Rudd, 1982, Approximate option valuation for arbitrary stochastic processes, *Journal of Financial Economics* 10, 347–349.

Jones, Christopher S., 1997, Bayesian analysis of the short-term interest rate, Working paper, The Wharton School, University of Pennsylvania.

Karlin, Samuel, and Howard M. Taylor, 1981, *A Second Course in Stochastic Processes* (Academic Press, New York).

Kloeden, Peter E. and Eckhardt Platen, 1992, *Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, New York).

Lo, Andrew W., 1988, Maximum likelihood estimation of generalized Itô processes with discretely sampled data, *Econometric Theory* 4, 231–247.

Melino, Angelo, 1994, Estimation of continuous-time models in finance; in Christopher S. Sims, ed.: *Advances in Econometrics, Sixth World Congress*, Vol. II (Cambridge University Press, Cambridge, England).

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Pedersen, Asger R., 1995, A new approach to maximum-likelihood estimation for stochastic differential equations based on discrete observations, *Scandinavian Journal of Statistics* 22, 55–71.

Santa-Clara, Pedro, 1995, Simulated likelihood estimation of diffusions with an application to the short term interest rate, Working paper, UCLA.

Stanton, Richard, 1997, A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* 52, 1973–2002.

Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177–188.

Wong, Eugene, 1964, The construction of a class of stationary Markov processes; in R. Bellman, ed.: *Stochastic Processes in Mathematical Physics and Engineering*, Proceedings of Symposia in Applied Mathematics, 16 (American Mathematical Society, Providence, RI).