Transitional Densities of Diffusion Processes: *A New Approach to Solving the Fokker-Planck Equation*

A.S. HURN, J.I. JEISMAN, AND K.A. LINDSAY

A.S. HURN
is a professor of econometrics in the School of Economics and Finance, Queensland University of Technology in Brisbane, Australia.
s.hurn@qut.edu.au

J.I. JEISMAN
is a research fellow in the School of Economics and Finance, Queensland University of Technology in Brisbane, Australia.
j.jeisman@qut.edu.au

K.A. LINDSAY
is a professor of mathematics in the Department of Mathematics, University of Glasgow in Glasgow, Scotland.
kal@maths.gla.ac.uk

Computing the transitional probability density function (PDF) of diffusion processes is an important problem in finance. One popular way to obtain the transitional PDF is to solve the Fokker-Planck equation numerically. This is not straightforward because the delta function initial condition and zero-flux boundary conditions are difficult to implement within the numerical scheme. By reformulating the problem in terms of the transitional cumulative distribution function (CDF), it is shown that these conditions are handled easily. The transitional PDF is subsequently computed by numerical differentiation of the transitional CDF and used to construct a likelihood function in the usual way.

In a recent article Jensen and Poulsen [2002] argue persuasively that computing the transitional probability density function (PDF) of diffusion processes is an important problem in financial modeling. They cite two reasons for this, namely, the need to provide maximum-likelihood estimators of the parameters of diffusions and the need to price contingent claims based on risk neutral densities. If the diffusion is a Markov process, the transitional PDF is a solution of the Fokker-Planck equation. Unfortunately, closed-form solutions for this equation are not usually available[1] and therefore numerical approximation is required. The two methods recommended by Jensen and Poulsen, based on their comparative study, are the Hermite polynomial approximation to the transitional PDF (Aït-Sahalia [2002]) and the numerical solution of the Fokker-Planck equation using a finite-difference algorithm.

This article focuses exclusively on the finite-difference approach to solving the Fokker-Planck equation and is motivated by the fact that the version of the algorithm implemented by Jensen and Poulsen may not be robust. The problem stems from the fact that the Fokker-Planck equation must be solved with an initial condition (a Dirac delta function) that is not representable within a numerical framework. Jensen and Poulsen suggest approximating the initial condition with a Normal distribution derived from the first order Euler approximation of the stochastic differential equation (SDE). The location and dispersion of this distribution depend on the current estimates of the parameters of the SDE and the size of the time interval, Δt. Although this approximation works well for simulated data, it is often problematic when using real data. The problem arises when the optimisation routine proposes parameter values for which the initial approximation of the transitional PDF is distinguishable from zero at just a few (or even none) of the spatial nodes. This occurrence undermines the accuracy of the estimation procedure and may cause it to fail completely, a situation that is encountered in an Application to monthly LIBOR data in this article. Using a very fine

discretisation of state space can potentially avert the problem, but at the expense of making the procedure impractical in terms of the required computational effort.[2] Moreover, there is no way of knowing *a priori* how fine a grid will be required to prevent the problem from occurring.

The central contribution made by this article is to develop a natural evolution of the finite-difference scheme advocated by Jensen and Poulsen which improves its robustness. This is achieved by reformulating the problem in terms of the transitional cumulative distribution function (CDF). The result of this reformulation is that the delta function initial condition is replaced by a step function, which has an exact numerical representation. Furthermore, this approach delivers an additional benefit through the simplification of the boundary conditions. Simulation experiments and a practical example demonstrate that the method is numerically robust and that it delivers accurate estimates of transitional PDF by numerical differentiation of the transitional CDF.

MAXIMUM LIKELIHOOD FRAMEWORK

Consider the general one-dimensional, time-homogeneous SDE

$$dX = \mu(X;\theta)\,dt + \sqrt{g(X;\theta)}\,dW \tag{1}$$

where X is a stochastic Markov process, $\mu(x;\theta)$ and $g(x;\theta)$ are respectively the instantaneous drift and instantaneous variance of X, dW is the differential of the Wiener process and θ is a vector of parameters to be estimated. Furthermore, suppose that $X_0, \ldots, X_N$ is a sample of $(N + 1)$ observations of the state variable at the times $t_0, \ldots, t_N$. The negative log-likelihood function of the sample is

$$-\log \mathcal{L} = -\log f_0(X_0|\theta) - \sum_{k=0}^{N-1} \log f(X_{k+1}|X_k;\theta)$$

where $f_0(X_0 \mid \theta)$ is the density of the initial state and $f(X_{k+1} \mid X_k;\theta)$ is the value of the transitional density at (X_{k+1}, t_{k+1}) for a process starting at (X_k, t_k) and evolving to (X_{k+1}, t_{k+1}) in accordance with Equation (1). The Markovian property of Equation (1) ensures that the transitional density of X at time t_{k+1} depends only on X_k.

The aim of ML estimation is to minimise the negative log-likelihood function with respect to the parameter vector θ. ML estimation relies on the fact that the transitional PDF $f(x, t) = f(x, t \mid X_k, t_k; \theta)$ of X at time t is the solution of the Fokker-Planck equation

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial x}\left(\frac{1}{2}\frac{\partial(g(x;\theta)f)}{\partial x} - \mu(x;\theta)f\right) \tag{2}$$

with suitable initial and boundary conditions. For a process with state space $[a, b]$ starting at $x = X_k$ at time t_k, the appropriate initial condition is

$$f(x, t_k) = \delta(x - X_k) \tag{3}$$

where δ is the Dirac delta function, and the boundary conditions required to conserve unit density within this state space are the zero probability flux conditions

$$\lim_{x\to a^+}\left(\mu f - \frac{1}{2}\frac{\partial(gf)}{\partial x}\right) = 0 \quad \lim_{x\to b^-}\left(\mu f - \frac{1}{2}\frac{\partial(gf)}{\partial x}\right) = 0 \tag{4}$$

This article is concerned with an equivalent statement of the problem posed by Equations (2–4) in terms of the transitional CDF, $F(x, t)$, which is defined in terms of the transitional PDF $f(x, t)$ by

$$F(x,t) = \int_a^x f(u,t)\,du \tag{5}$$

When expressed in terms of $F(x, t)$, Equation (2) takes the form

$$\frac{\partial^2 F}{\partial x\,\partial t} = \frac{\partial}{\partial x}\left(\frac{1}{2}\frac{\partial}{\partial x}\left(g\frac{\partial F}{\partial x}\right) - \mu\frac{\partial F}{\partial x}\right) \tag{6}$$

which can be integrated with respect to x to give

$$\frac{\partial F}{\partial t} = \frac{1}{2}\frac{\partial}{\partial x}\left(g\frac{\partial F}{\partial x}\right) - \mu\frac{\partial F}{\partial x} + C(t) \tag{7}$$

where $C(t)$ is an arbitrary function of integration. The boundary conditions for this equation require that $F(a, t) = 0$

and $F(b, t) = 1$ which in turn require that $C(t) = 0$. Therefore $F(x, t)$ satisfies the partial differential equation

$$\frac{\partial F}{\partial t} = \frac{1}{2}\frac{\partial}{\partial x}\left(g(x;\theta)\frac{\partial F}{\partial x}\right) - \mu(x;\theta)\frac{\partial F}{\partial x} \qquad (8)$$

with Dirichlet boundary conditions $F(a, t) = 0$ and $F(b, t) = 1$. The initial condition $F(x, t_k)$ for a transition from (X_k, t_k) is constructed from the definition (5) and the properties of the delta function to obtain

$$F(x, t_k) = \begin{bmatrix} 0 & x < X_k \\ 1/2 & x = X_k \\ 1 & x > X_k \end{bmatrix} \qquad (9)$$

One important advantage of this approach is that the delta function initial condition required in the computation of transitional PDF is now replaced by a step function initial condition in the computation of the transitional CDF. The latter has a precise numerical representation whereas the delta function (3) must be approximated. The method of finite differences will now be used to illustrate the numerical implementation of the maximum likelihood procedures.

FINITE-DIFFERENCE PROCEDURE

The implementation of the finite-difference procedure in this article is based on the discretisation of state space into n intervals of size $\Delta x = (b - a)/n$ by the $(n + 1)$ uniformly spaced nodes $x_0, \ldots, x_n$ where $x_p = a + p\Delta x$ and the discretisation of the time interval $[t_k, t_{k+1}]$ into m uniform subintervals of duration $\Delta t = (t_{k+1} - t_k)/m$ by the time points $t_{k,0}, t_{k,1}, \ldots, t_{k,m}$ where $t_{k,q} = t_k + q\Delta t$.

Transitional PDF Specification

Let $f_p^{(q)} = f(x_p, t_{k,q})$ be the value of the transitional PDF at $(x_p, t_{k,q})$ then integration of Equation (2) over $[t_{k,q}, t_{k,q+1}]$ gives

$$f(x,t_{k,q+1}) - f(x,t_{k,q}) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(g(x;\theta)\int_{t_{k,q}}^{t_{k,q+1}} f(x,t)\,dt\right) - \frac{\partial}{\partial x}\left(\mu(x;\theta)\int_{t_{k,q}}^{t_{k,q+1}} f(x,t)\,dt\right) \qquad (10)$$

Let the auxiliary variables

$$\phi_p = \int_{t_{k,q}}^{t_{k,q+1}} f(x_p, t)\,dt$$

be defined, then in the usual notation, Equation (10) has finite difference approximation

$$f_p^{(q+1)} - f_p^{(q)} = \frac{g_{p+1}\phi_{p+1} - 2g_p\phi_p + g_{p-1}\phi_{p-1}}{2(\Delta x)^2} - \frac{\mu_{p+1}\phi_{p+1} - \mu_{p-1}\phi_{p-1}}{2(\Delta x)}$$

The terms in this equation are now regrouped to give

$$f_p^{(q+1)} - f_p^{(q)} = \left(\frac{g_{p-1} + \mu_{p-1}\Delta x}{2(\Delta x)^2}\right)\phi_{p-1} - \frac{g_p}{(\Delta x)^2}\phi_p + \left(\frac{g_{p+1} - \mu_{p+1}\Delta x}{2(\Delta x)^2}\right)\phi_{p+1}$$

The trapezoidal quadrature is now used to approximate ϕ_p by the formula

$$\phi_p = \frac{\Delta t}{2}\left(f_p^{(q+1)} + f_p^{(q)}\right) + O(\Delta t)^3$$

so that the final finite difference representation of Equation (2) is

$$\begin{aligned} &-\frac{r}{4}(g_{p-1} + \mu_{p-1}\Delta x)f_{p-1}^{(q+1)} + \left(1 + \frac{r}{2}g_p\right)f_p^{(q+1)} \\ &-\frac{r}{4}(g_{p+1} - \mu_{p+1}\Delta x)f_{p+1}^{(q+1)} \\ &= \frac{r}{4}(g_{p-1} + \mu_{p-1}\Delta x)f_{p-1}^{(q)} + \left(1 - \frac{r}{2}g_p\right)f_p^{(q)} \\ &+\frac{r}{4}(g_{p+1} - \mu_{p+1}\Delta x)f_{p+1}^{(q)} \end{aligned} \qquad (11)$$

where $r = \Delta t/(\Delta x)^2$ is the Courant number. The procedure used to construct Equation (11) is essentially the Crank-Nicolson algorithm, and it is well known that this algorithm is consistent and exhibits robust numerical properties.

Suppose that the solution is sought in the finite interval $[x_0, x_n]$. For many SDEs of type (1), the sample

space is the semi-infinite interval $[0, \infty]$ so that the drift and diffusion specifications will often satisfy $g(x_0) = 0$ and $\mu(x_0) \geq 0$. Under these conditions the boundary condition at $x = x_0$ is equivalent to the condition $f(x_0, t) = 0$, that is, no density can accumulate at the boundary $x = x_0$. However, no equivalent simplification exists for the boundary condition at $x = x_n$, where x_n is now suitably large, but finite.[3] The derivation of this condition is now described.

The backward-difference representation of the boundary condition (4) at $x = x_n$ is

$$\frac{1}{2}\left(\frac{3g_n f_n^{(q)} - 4g_{n-1} f_{n-1}^{(q)} + g_{n-2} f_{n-2}^{(q)}}{2\Delta x}\right) - \mu_n f_n^{(q)} + O(\Delta x)^2 = 0 \quad (12)$$

These terms are regrouped and the truncation error ignored to obtain

$$(3g_n - 4\mu_n \Delta x) f_n^{(q)} - 4g_{n-1} f_{n-1}^{(q)} + g_{n-2} f_{n-2}^{(q)} = 0 \quad (13)$$

This boundary condition is now used at $t_{k,q}$ and $t_{k,q+1}$ to eliminate $f_n^{(q)}$ and $f_n^{(q+1)}$ respectively from Equation (11) evaluated at $p = n-1$. The final result is

$$P f_{n-2}^{(q+1)} - (Q - R) f_{n-1}^{(q+1)} = -P f_{n-2}^{(q)} + (Q + R) f_{n-1}^{(q)} \quad (14)$$

where

$$P = g_{n-2}(3\mu_n \Delta x - 2g_n) - \mu_{n-2}(3g_n - 4\mu_n \Delta x)\Delta x$$

$$Q = 4g_{n-1}\left(\mu_n \Delta x - \frac{g_n}{2}\right)$$

$$R = \frac{4}{r}(3g_n - 4\mu_n \Delta x)$$

When it is not possible to assume that $f(x_0, t) = 0$ the lower boundary condition can be derived using a similar procedure. The result is an identical expression to Equation (14) but with the subscripts n, $n - 1$, and $n - 2$ replaced by 0, 1, and 2 respectively, and the negative sign between the two terms in P changed to a positive sign.

Assuming that $f(x_0, t) = 0$, the final specification of the numerical problem starts with equation

$$\left(1 + \frac{r}{2} g_1\right) f_1^{(q+1)} - \frac{r}{4}(g_2 - \mu_2 \Delta x) f_2^{(q+1)} = \left(1 - \frac{r}{2} g_1\right) f_1^{(q)} + \frac{r}{4}(g_2 - \mu_2 \Delta x) f_2^{(q)} \quad (15)$$

which is the particularisation of the general Equation (11) at x_1 taking account of the requirement $f(x_0, t) = 0$, followed by $(n - 3)$ equations with general form (11) in which the index p takes values from $p = 2$ to $p = n - 2$, followed finally by (14). Together these equations form a tri-diagonal system to be solved for the transitional PDF at time $t_{k,q+1}$ given the PDF at $t_{k,q}$. Note that the tri-diagonal system described by Equations (11), (14), and (15) is solved for the transitional PDF at nodes $x_1, \ldots, x_{n-1}$. Here the transitional density at x_0 is known *a priori* to be zero, and the final transitional PDF at x_n may be obtained by means of relation (13).

As has already been remarked, the initial condition for the solution of the Fokker-Planck equation is a delta function and is therefore not representable within the framework of the finite-difference method. Jensen and Poulsen suggest that this difficulty can be circumvented by starting the finite-difference algorithm at $t_{k,1}$ based on the assumption that the transitional PDF at this time is well approximated by the Normal distribution with mean value $X_k + \mu(X_k; \theta)\Delta t$ and variance $g(X_k; \theta)\Delta t$.

This is a good suggestion, but it suffers from an important drawback. Once Δt is chosen and the initial state is known, the diffusion occurring over the time interval Δt from the true initial condition determines the size of the interval of state space over which the transitional PDF is significantly different from zero. The resolution, Δx, of state space must now be chosen to be sufficiently small so as to guarantee that a reasonable number of nodes (say a dozen) lie within this interval of non-zero transitional PDF. Typically Δt and Δx are pre-programmed properties of the numerical procedure. If the optimisation procedure chooses a value of θ for which $g(X_k; \theta)$ is small, there may be an inadequate number of nodes in the region of state space where the density takes a value significantly different from zero. In this case, the algorithm may either fail or return very inaccurate estimates of the transitional PDF. There is no simple cure for this problem other than to make the resolution of state space so fine as to render the entire algorithm infeasible. Of course, this is a less serious problem with simulated data

where appropriate starting values are immediately available and the optimal parameters for any given simulation are likely to be near these starting values.

Transitional CDF Specification

The finite difference representation of Equation (8) is constructed by noting that the equation can be re-expressed in the form

$$\frac{\partial F}{\partial t} = \frac{1}{4}\left[\frac{\partial^2 (gF)}{\partial x^2} + g\frac{\partial^2 F}{\partial x^2} - F\frac{\partial^2 g}{\partial x^2}\right] - \mu\frac{\partial F}{\partial x} \qquad (16)$$

The motivation for this manipulation stems from the fact that central-difference expressions for second differences are readily available. The procedure used to derive Equation (11) from Equation (2) via Equation (10) is repeated for Equation (16). The calculation is routine and so the details are suppressed. If $F_p^{(q)} = F(x_p, t_{k,q})$ denotes the value of the CDF at $(x_p, t_{k,q})$, then the finite-difference approximation of Equation (16) is

$$\begin{aligned}
&-[g_{p-1} + g_p + 2\mu_p \Delta x]F_{p-1}^{(q+1)} \\
&\quad + \left[\frac{8}{r} + (g_{p-1} + 2g_p + g_{p+1})\right]F_p^{(q+1)} \\
&\quad - [g_p + g_{p+1} - 2\mu_p \Delta x]F_{p+1}^{(q+1)} \\
&\quad = [g_{p-1} + g_p + 2\mu_p \Delta x]F_{p-1}^{(q)} \\
&\quad + \left[\frac{8}{r} - (g_{p-1} + 2g_p + g_{p+1})\right]F_p^{(q)} \\
&\quad + [g_p + g_{p+1} - 2\mu_p \Delta x]F_{p+1}^{(q)}
\end{aligned} \qquad (17)$$

However the boundary conditions assert that $F_0^{(q)} \equiv 0$ and $F_n^{(q)} \equiv 1$, and therefore Equations (17) can be expressed in matrix form

$$T_L \mathbf{F}^{(q+1)} = T_R \mathbf{F}^{(q)} + \mathbf{B}^{(q)}$$

where T_L and T_R are tri-diagonal matrices of dimension $(n-1) \times (n-1)$, $\mathbf{B}$ is a constant vector of dimension $(n-1)$ which differs from the zero vector only in its last entry and $\mathbf{F}^{(q)}$ is the $(n-1)$ dimensional vector containing the values of the transitional CDF at the (internal) nodes $x_1, \ldots, x_{n-1}$ at time $t_{k,q}$. Together these equations describe how the transitional CDF at time t_{k+1} can be obtained. The value of the transitional PDF at (X_{k+1}, t_{k+1}) is estimated by numerical differentiation of the transitional CDF at the nodes to the left and right of X_{k+1} followed by linear interpolation of these values to find the required transitional density.

Unlike the PDF approach, the CDF approach can be initialised in two ways. Equation (9) gives a theoretical initial condition for the transitional CDF at time $t_{k,0}$ for which there is no equivalent statement in the PDF approach. Alternatively, the Jensen and Poulsen approximation can be used to initialise the transitional CDF approach at time $t_{k,1}$ by drawing on the properties of the CDF of the Normal distribution. The CDF approach so initialised can be directly compared with the equivalent PDF approach. This approach does not, however, suffer from the robustness issues associated with the use of the Jensen and Poulsen approximation in the PDF approach. For example, if the Jensen and Poulsen approximation of the initial transitional PDF is such that it is almost entirely localised between two spatial nodes and hence indistinguishable from zero at all of the nodes, the PDF approach will fail. By contrast, the equivalent approximation of the initial transitional CDF, obtained by passing the Jensen and Poulsen approximation to the Normal CDF, is a step function, which the CDF approach can handle routinely.

LOG-LIKELIHOOD COMPUTATION

In the context of the parameter estimation problem, the efficacy of the new procedure depends on the accuracy with which the log-likelihood of the data can be computed. This section outlines a simulation experiment to compare the accuracy of log-likelihood computation based on the PDF approach with that obtained by using the same finite-difference configuration to construct the transitional CDF. The experiment involves the simulation of a CIR process and the comparison of the log-likelihood of the resultant sample computed in a variety of ways with values obtained from the closed-form expression for the CIR transitional density. In order to ensure a fair comparison between the PDF and CDF approaches, the CDF approach will be run with both step function and Jensen and Poulsen initial conditions.

The details of the simulation exercise are now described. For each estimation procedure, 2,000 repetitions of the calculation of the log-likelihood for samples containing $N = 500$ transitions were run. The samples were generated by integrating the stochastic differential equation

$$dX = \alpha(\beta - X)dt + \sigma\sqrt{X}\,dW$$

with $\alpha = 0.2$, $\beta = 0.08$ and $\sigma = 0.1$, using Milstein's scheme with 1,000 time steps of size 0.001 between observations. The maximum relative error, mean absolute relative error, and mean squared relative error were calculated from the 2,000 repetitions and are presented in Exhibit 1.

Three main conclusions can be drawn from this simulation exercise. First, the CDF approach with the Jensen and Poulsen initial condition provides the best performance and is superior to the equivalent PDF approach for all combinations of Δx and Δt. Second, the CDF approach with the theoretical step function initial condition can be implemented successfully in practice and performs with credit. It is more accurate than the PDF approach for a coarse discretisation of state space (e.g., $\Delta x = 0.005$), although this result reverses as the spatial discretisation becomes finer. A final result worthy of passing comment occurs when the CDF approach starting with the step function initial condition is used with a fine spatial discretisation and crude temporal discretisation, for example, Exhibit 1 with $\Delta x = 0.001$ and $\Delta t = 0.02$. Under these circumstances, the method performs relatively poorly, suggesting that it is important to ensure that sufficient time steps are allowed in the integration phase of the calculation to allow diffusion to smooth the discontinuous initial condition.

EXHIBIT 1

Measures of Relative Error in the Calculation of Log-Likelihood from 2000 Simulations of the Process $dX = \alpha(\beta - X)dt + \sigma\sqrt{X}\,dW$ with Parameters $\alpha = 0.2$, $\beta = 0.08$ and $\sigma = 0.10$

		Measures of relative error		
	Δx	Maximum Absolute	Mean Absolute	Mean Squared
PDF Approach				
	0.005	9.266×10^{-3}	1.949×10^{-3}	5.822×10^{-6}
$\Delta t = 0.02$	0.002	5.177×10^{-3}	1.441×10^{-4}	7.603×10^{-8}
	0.001	1.736×10^{-3}	2.364×10^{-5}	5.189×10^{-9}
	0.005	1.672×10^{-2}	5.159×10^{-3}	3.457×10^{-5}
$\Delta t = 0.01$	0.002	5.579×10^{-3}	3.423×10^{-4}	2.352×10^{-7}
	0.001	2.331×10^{-3}	3.700×10^{-5}	8.492×10^{-9}
CDF Approach with step initial condition				
	0.005	4.370×10^{-3}	1.078×10^{-3}	1.744×10^{-6}
$\Delta t = 0.02$	0.002	1.612×10^{-3}	3.376×10^{-4}	1.809×10^{-7}
	0.001	6.613×10^{-2}	1.157×10^{-3}	2.940×10^{-5}
	0.005	4.369×10^{-3}	1.078×10^{-3}	1.744×10^{-6}
$\Delta t = 0.01$	0.002	1.613×10^{-3}	3.376×10^{-4}	1.810×10^{-7}
	0.001	9.806×10^{-4}	1.663×10^{-4}	4.325×10^{-8}
CDF Approach with **JP** initial condition				
	0.005	2.744×10^{-3}	4.071×10^{-4}	2.565×10^{-7}
$\Delta t = 0.02$	0.002	1.722×10^{-3}	7.362×10^{-5}	1.468×10^{-8}
	0.001	9.682×10^{-4}	2.014×10^{-5}	1.624×10^{-9}
	0.005	2.809×10^{-3}	4.154×10^{-4}	2.652×10^{-7}
$\Delta t = 0.01$	0.002	1.701×10^{-3}	7.277×10^{-5}	1.441×10^{-8}
	0.001	9.797×10^{-4}	1.950×10^{-5}	1.614×10^{-9}

As an additional check on the efficacy of the CDF approaches a second simulation exercise was undertaken. In this experiment, the data generated for the first simulation exercise were used to estimate the parameters of the underlying CIR model by the PDF and CDF methods. The mean of the parameter estimates and mean computation times over the 2,000 samples are presented in Exhibit 2. Given the small errors recorded in Exhibit 1 in estimating the likelihoods, it is not surprising that all of the approaches considered in the article deliver what are effectively identical parameter estimates. The computation times reveal that the CDF approaches require no significant increase in computational effort. Furthermore, it is worth noting that the additional robustness of the CDF approach with Jensen and Poulsen initial condition means that this procedure can safely be run with coarser discretisations of state, which will lead to computational savings.

APPLICATION TO MONTHLY LIBOR DATA

There is a large literature documenting the link between the instantaneous short-term interest rate and the pricing of zero-coupon bonds for various terms to maturity. The classic one-factor model of the term structure is based on the assumption that the instantaneous interest rate $X(t)$ evolves in accordance with

$$dX = \alpha(\beta - X)dt + \sigma X^{\gamma} dW \qquad (18)$$

where dW is the differential of the Wiener process and α (speed of adjustment parameter), β (the mean interest rate), σ (volatility control), and γ (the so-called levels effect) are parameters to be estimated. Many models in the finance literature are special cases of this equation. For example, the case $\beta = 0$, $\gamma = 1$ corresponds to geometric Brownian motion; $\gamma = 0$ has been treated by Vasicek [1977]; $\gamma = 1$ has been treated by Brennan and Schwartz [1979]; and, of course, $\gamma = 1/2$ is the Cox, Ingersoll, and Ross [1985] model discussed previously. Empirical evidence (Chan et al. [1992]; Pagan et al. [1995]) suggests that γ should be estimated rather than imposed. Notwithstanding any theoretical shortcomings in this general approach, estimating the parameters of Equation (18) is a worthwhile and challenging task with which to illustrate the method described here.

A reasonable approximation to the unobservable instantaneous short-term interest rate is the one-month London Interbank Offer Rate (LIBOR) observed at weekly intervals from 4 January, 1999 to 19 August, 2005 (345 observations). The data are plotted in Exhibit 3 with the estimated long-term mean interest rate superimposed. The period is mainly one of low interest rates with a maximum of 6.77% and a minimum of 1.09%.

Point estimates and standard errors for the parameters of the model in Equation (18) are reported in Exhibit 4 for two different estimation procedures. The first is discrete maximum likelihood (DML) where estimation is based on the Euler discretisation of equation (18) in an identical manner to that discussed by Jensen and Poulsen. Although DML provides biased estimates due to the discretisation, these estimates are simple to compute and useful for comparative purposes. The second estimation

EXHIBIT 2

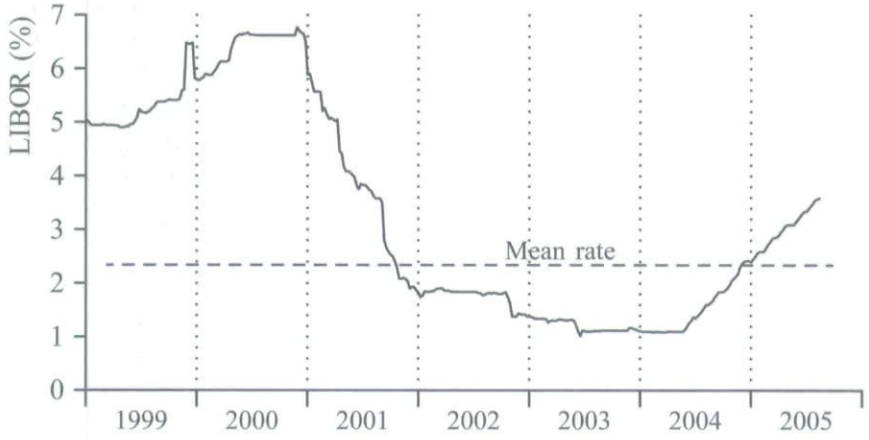
Mean Parameter Estimates from 2,000 Simulations of the Process $dX = \alpha(\beta - X)dt + \sigma\sqrt{X}\,dW$ with Parameters $\alpha = 0.2$, $\beta = 0.08$, and $\sigma = 0.10$ using the Method of Finite Differences with Spatial Resolution $\Delta x = 0.002$ and Temporal Resolution $\Delta t = 0.02$

Standard errors of the parameter estimates are shown in parentheses.

Estimation procedure	Mean Parameter Estimates α	β	σ	Mean Computation Times (*sec*)
PDF approach	0.2095 (0.0330)	0.0802 (0.0063)	0.1001 (0.0034)	14.19
CDF approach with step initial condition	0.2099 (0.0333)	0.0802 (0.0063)	0.1000 (0.0035)	14.27
CDF approach with **JP** initial condition	0.2096 (0.0332)	0.0802 (0.0063)	0.0999 (0.0034)	14.33

EXHIBIT 3

Evolution of Monthly London Interbank Offer Rate



method is the proposed transitional CDF approach with the Jensen and Poulsen initial condition described earlier. Interestingly enough, for the chosen discretisation of state space, it proved impossible to implement the equivalent PDF approach successfully. This is because the optimisation procedure occasionally tried values for the volatility control parameter, σ, for which the given resolution of state space was inadequate. In this instance this problem was sufficient to cause the optimisation procedure to fail. Of course, choosing a finer resolution of state space may have cured the problem, but this would render the procedure uncompetitive with respect to the CDF algorithm.

The first important conclusion to emerge from Exhibit 4 is that the DML estimates are poorly resolved and some way away from the ML estimates obtained from a continuous-time estimation procedure. This provides further support for the argument of Jensen and Poulsen that good estimates of transitional densities are an important aspect of modern finance. The long-term mean interest rate, β, is estimated to be 2.35%, which is below the unconditional mean of the sample. This probably is a result of stronger mean reversion at higher interest rates due to the fact that the volatility increases approximately quadratically with the interest rate. The ML estimate of the levels effect parameter, γ, is very well resolved. The estimated value of 0.95 is well above the value of 1/2 suggested by Cox et al. [1985]. Indeed, fitting the CIR model to these data, even using the known closed-form expression for the transitional PDF, proved problematic as β was estimated to be zero with a very large standard error. This highlights the fact that many popular SDEs in finance may not provide a good description of the actual process being modeled and thus robust estimation procedures are essential.

EXHIBIT 4

Parameter Estimates of the One-Factor Term Structure Equation

Standard errors of the parameter estimates are shown in parentheses.

Method	Parameters	α	β	σ	γ
DML	$\Delta t = 1.0$	0.0030 (0.0031)	0.0177 (0.0105)	0.0258 (0.0023)	0.9851 (0.0263)
CDF	$\Delta x = 0.0005$, $\Delta t = 0.001$	0.0178 (0.0016)	0.0234 (0.0001)	0.0136 (0.0002)	0.9491 (0.0017)

CONCLUSION

This article has introduced a robust modification of the traditional usage of the Fokker-Planck equation in the maximum-likelihood estimation of the parameters of stochastic differential equations. Instead of solving the Fokker-Planck equation for the transitional PDF, the approach proposed here reformulates the problem in terms of the transitional CDF. The technique is illustrated with reference to the method of finite differences, but it is an analytical procedure and its usefulness, therefore, extends to other numerical algorithms. The method is easier to implement than one based on the direct solution of the Fokker-Planck equation, first because the initial condition is more amenable to numerical work, and second, because the original gradient-like boundary conditions are replaced by Dirichlet conditions in the modification. A Monte Carlo simulation experiment based on the CIR process indicates that the method is generally more accurate and more robust than the traditional method. This conclusion is reinforced by a parameter-estimation exercise using short-term interest rate data.

ENDNOTES

[1]The most common exceptions in financial modeling are geometric Brownian motion, the square root or CIR process (Cox, Ingersoll, and Ross [1985]), and the Ornstein-Uhlenbeck (OU) process, often associated with Vasicek [1977].

[2]One potential way in which the benefits of using a very fine grid can be obtained without sacrificing computational efficiency is through the use of a numerical scheme in the spirit of adaptive mesh models (Figlewski and Gao [1999]) for option pricing. In these models, a fine grid is used in regions of interest in the state space, and a coarser grid is used elsewhere. In the context of this article, this suggests that a useful area for further research is to analyse the effectiveness of alternative numerical schemes, such as finite-element and spectral methods.

[3]In the applications here, the upper limit of the finite interval, x_n, is chosen as the maximum of the sample plus the range of the sample.

REFERENCES

Aït-Sahalia, Y. "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach." *Econometrica*, 70 (2002), pp. 223–262.

Brennan, M., and E. Schwartz. "A Continuous-Time Approach to the Pricing of Bonds." *Journal of Banking and Finance*, 3 (1979), pp. 133–155.

Chan, K., G. Karolyi, F. Longstaff, and A. Sanders. "An Empirical Comparison of Alternative Models of the Short Term Interest Rate." *Journal of Finance*, 52 (1992), pp. 1209–1227.

Cox, J., J. Ingersoll, and S. Ross. "A Theory of the Term Structure of Interest Rates." *Econometrica*, 53 (1985), pp. 385–407.

Figlewski, S., and B. Gao. "The Adaptive Mesh Model: A New Approach to Efficient Option Pricing." *Journal of Financial Economics*, 53 (1999), pp. 313-351.

Jensen, B. and R. Poulsen. "Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques." *Journal of Derivatives*, 9 (2002), pp. 18–32.

Pagan, A., A. Hall, and V. Martin. "Modelling the Term Structure." In G. Maddala and C. Rao eds., *Statistical Methods in Finance, Handbook of Statistics*, 13 (1995), Amsterdam: Elsevier Science.

Vasicek, O. "An Equilibrium Characterization of the Term Structure." *Journal of Financial Economics*, 5 (1977), pp. 177–188.