ML Estimation of the Parameters of SDEs by Numerical Solution of the Fokker-Planck Equation

A.S. Hurn[1], **J. Jeisman**[1] **and K.A. Lindsay**[2]

[1]School of Economics and Finance, Queensland University of Technology, Brisbane, 4001, Australia
[2]Department of Mathematics, University of Glasgow, University Gardens, Glasgow G12 8QW, Scotland

Keywords: *Stochastic differential equations, Finite difference, Cumulative distribution function*

ABSTRACT

Maximum likelihood (ML) estimates of the parameters of SDEs are consistent and asymptotically efficient, but unfortunately difficult to obtain if a closed form expression for the transitional probability density function (PDF) of the process is not available. One popular way to obtain the transitional PDF is to solve the Fokker-Planck equation numerically. However the treatment of the delta function initial condition and zero-flux boundary conditions, both of which are necessary to implement these numerical schemes, is not straightforward. 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.

Consider the general one-dimensional, time-homogeneous stochastic differential equation (SDE)

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

where X is a stochastic Markov process, $\mu(x;\boldsymbol{\theta})$ and $g(x;\boldsymbol{\theta})$ are respectively the instantaneous drift and instantaneous diffusion of X, dW is the differential of the Wiener process and $\boldsymbol{\theta}$ is a vector of parameters to be estimated. The aim of ML estimation is to minimise the negative log-likelihood function with respect to the parameter vector $\boldsymbol{\theta}$. ML estimation relies on the fact that the transitional density of X at time t is the solution of the Fokker-Planck equation

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

satisfying a delta function initial condition and gradient-like boundary conditions.

This paper is concerned with an equivalent statement of this problem in terms of the transitional CDF, $F(x,t)$, which is defined in terms of the transitional PDF, $f(x,t)$, by

$$F(x,t) = \int^{x} f(u,t)\,du\,.$$

When expressed in terms of $F(x,t)$, the Fokker-Planck equation takes the form

$$\frac{\partial F}{\partial t} = \frac{1}{2}\frac{\partial}{\partial x}\Big(g\frac{\partial F}{\partial x}\Big) - \mu\frac{\partial F}{\partial x}$$

with a step function initial condition and Dirichlet boundary conditions.

Both the PDF and CDF approaches to the solution of the Fokker-Planck equation are implemented using a finite-difference method. The latter is easier to implement than the former, because the initial condition is more amenable to numerical work and the gradient-like boundary conditions associated with the Fokker-Planck equation are replaced by Dirichlet boundary conditions in the modification.

The efficacy of the numerical solution is evaluated by means of two Monte Carlo exercises based on simulating the CIR 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 1000 time steps of size 0.001 between observations. The first experiment compares the PDF and CDF approaches in terms of the accuracy of log-likelihood computation, while the second simulation experiment involves the estimation of the parameters of the underlying CIR model.

The most significant finding is that the CDF approach using the step function initial condition can be implemented successfully in practice. By contrast, there is no equivalent result for the PDF-based procedure, because it is always necessary to approximate the initial condition. The main empirical result to emerge from the simulation exercises in this paper is that, given equivalent starting information, the CDF approach is always superior to the PDF approach in terms of the accuracy of likelihood evaluation.

1 INTRODUCTION

The problem of estimating the parameters of stochastic differential equations (SDEs) from discretely-observed time-series data has received much attention of late. If the state variable of an SDE satisfies the Markov property, the transitional probability density function (PDF) from the last known state encodes the history of the process. Furthermore, this transitional PDF is a solution of a partial differential equation known as the forward Kolmogorov or Fokker-Planck equation. Maximum-likelihood (ML) estimates of the parameters of the SDE therefore requires solution of this equation, a fact recognised in the context of financial econometrics by Lo (1988). If there is a closed-form expression for the transitional PDF, then obtaining the ML estimates of the parameters is straightforward. Unfortunately such an expression is usually unavailable[1] and therefore ML estimation requires numerical integration of the Fokker-Planck equation.

Most applications in financial econometrics ignore measurement error and so the appropriate initial condition for the Fokker-Planck equation is a delta function located at the last known state. A delta function is, however, not representable within a numerical scheme and therefore causes immediate difficulty for any numerical procedure to compute transitional density. The central idea of this paper is to reformulate the problem in terms of the transitional cumulative distribution function (CDF) for the Fokker-Planck equation instead of the transitional probability density function (PDF). The result of this reformulation is that the delta function initial condition in the traditional approach is replaced by a step function initial condition. Not only can this initial condition be represented within a numerical framework but this approach also delivers an additional benefit through the simplification of the boundary conditions.

The remainder of this paper consists of four sections. Section 2 provides a brief description of ML estimation of the parameters of SDEs. Section 3 outlines the finite-difference procedure for both the PDF and the CDF approaches. Section 4 details the results of a simulation experiment investigating the efficacy of the approaches and the concluding remarks are made in Section 5.

[1]The most common exceptions in financial modelling 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 MAXIMUM LIKELIHOOD FRAMEWORK

Consider the general one-dimensional, time-homogeneous stochastic differential equation (SDE)

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

where X is a stochastic Markov process, $\mu(x;\boldsymbol{\theta})$ and $g(x;\boldsymbol{\theta})$ are respectively the instantaneous drift and instantaneous diffusion of X, dW is the differential of the Wiener process and $\boldsymbol{\theta}$ is a vector of parameters to be estimated. Furthermore, suppose that $X_0, \cdots, X_T$ is a sample of $(T+1)$ observations of the state variable at the times $t_0, \cdots, t_T$. The negative log-likelihood function of the sample is

$$-\log f_0(X_0|\boldsymbol{\theta}) - \sum_{s=0}^{T-1} \log f(X_{s+1}|X_s;\boldsymbol{\theta})$$

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

The aim of ML estimation is to minimise the negative log-likelihood function with respect to the parameter vector $\boldsymbol{\theta}$. ML estimation relies on the fact that the transitional density of X at time t is the solution of the Fokker-Planck equation

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

satisfying a suitable initial condition and boundary conditions. Suppose, furthermore, that the state space of the problem is $[a, b]$ and the process starts at $x = X_s$ at time t_s. The initial condition is now

$$f(x, t_s) = \delta(x - X_s) \tag{3}$$

and the boundary conditions required to conserve unit density within this interval are

$$\begin{aligned} \lim_{x\to a^+}\left(\frac{1}{2}\frac{\partial(gf)}{\partial x} - \mu f\right) &= 0\,, \\ \lim_{x\to b^-}\left(\frac{1}{2}\frac{\partial(gf)}{\partial x} - \mu f\right) &= 0\,. \end{aligned} \tag{4}$$

This paper is concerned with an equivalent statement of this problem 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

$$\frac{\partial^2 F}{\partial x\,\partial t} = \frac{\partial}{\partial x}\Big[\,\frac{1}{2}\,\frac{\partial}{\partial x}\Big(g\frac{\partial F}{\partial x}\Big) - \mu\frac{\partial F}{\partial x}\,\Big] \qquad (6)$$

which can be integrated with respect to x to give

$$\frac{\partial F}{\partial t} = \Big[\,\frac{1}{2}\,\frac{\partial}{\partial x}\Big(g\frac{\partial F}{\partial x}\Big) - \mu\frac{\partial F}{\partial x}\,\Big] + C(t). \qquad (7)$$

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

$$\frac{\partial F}{\partial t} = \frac{1}{2}\,\frac{\partial}{\partial x}\Big(g\frac{\partial F}{\partial x}\Big) - \mu\frac{\partial F}{\partial x} \qquad (8)$$

with Dirichlet boundary conditions $F(a,t) = 0$ and $F(b,t) = 1$. The initial condition $F(x,t_s)$ for a transition from (X_s,t_s) is constructed from the definition (5) to obtain

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

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.

3 FINITE-DIFFERENCE PROCEDURE

The implementation of the finite-difference procedure in this paper is based on the discretisation of state space and time. Letting h and k denote respectively the units of discretisation of state space and time, the nodes of the finite-difference scheme are located at $x_p = ph$, where p is an integer satisfying $0 \le p \le n$, and the time interval $[t_s, t_{s+1}]$ is discretised into m uniform sub-intervals of duration k.

3.1 Transitional PDF specification

Let $f_p^{(q)} = f(x_p, t_s + qk)$ be the value of the transitional PDF at $(x_p, t_s + qk)$ where q is an integer taking values between 0 and m, then integration of equation (2) over $[t_s + qk, t_s + qk + k]$ gives

$$\begin{aligned} &f(x, t_s + qk + k) - f(x, t_s + qk) \\ &= \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\Big(g(x)\int_{t_s+qk}^{t_s+qk+k} f(x,t)\,dt\Big) \\ &\quad -\frac{\partial}{\partial x}\Big(\mu(x)\int_{t_s+qk}^{t_s+qk+k} f(x,t)\,dt\Big)\,. \end{aligned} \qquad (10)$$

Let the auxiliary variables

$$\phi_p = \int_{t_s+qk}^{t_s+qk+k} f(x_p,t)\,dt$$

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

$$\begin{aligned} f_p^{(q+1)} - f_p^{(q)} = \frac{g_{p+1}\phi_{p+1} - 2g_p\,\phi_p + g_{p-1}\phi_{p-1}}{2h^2} \\ - \frac{\mu_{p+1}\phi_{p+1} - \mu_{p-1}\phi_{p-1}}{2h}\,. \end{aligned}$$

The terms in this equation are now regrouped to give

$$\begin{aligned} f_p^{(q+1)} - f_p^{(q)} = \Big(\frac{g_{p-1} + h\mu_{p-1}}{2h^2}\Big)\,\phi_{p-1} - \frac{g_p}{h^2}\,\phi_p \\ + \Big(\frac{g_{p+1} - h\mu_{p+1}}{2h^2}\Big)\,\phi_{p+1}. \end{aligned}$$

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

$$\phi_p = \frac{k}{2}\Big(f_p^{(q+1)} + f_p^{(q)}\Big) + O(k^3)\,,$$

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

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

where $r = k/h^2$ 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.

Expression (11) forms the core of the finite-difference representation of equation (2). It suggest that the distribution of transitional density at any time is computed by solving a tri-diagonal system of equations given an initial distribution of transitional density and suitable boundary conditions. As has already been remarked, the required initial condition is a delta function and is therefore not representable within the framework of the finite-difference method.

Jensen and Poulsen (2002) suggest that this difficulty can be circumvented by starting the finite-difference algorithm with a specification of the distribution of transitional density at (t_s+k) based on the assumption that the transitional density at (t_s+k) can be expected to be well approximated by the normal distribution with mean value $X_s + k\,\mu(X_s;\boldsymbol{\theta})$ and variance $k\,g(X_s;\boldsymbol{\theta})$.

The treatment of the boundary conditions is more technical. Suppose that the solution is sought in the finite[2] 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. The derivation of this condition is now described.

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

$$\frac{1}{2}\Big(\frac{3\,g_n f_n^{(q)} - 4\,g_{n-1}\,f_{n-1}^{(q)} + g_{n-2}\,f_{n-2}^{(q)}}{2h}\Big) - \mu_n f_n^{(q)} + O(h^2) = 0\,. \tag{12}$$

These terms are regrouped and the truncation error ignored to obtain

$$\big(3g_n - 4h\mu_n\big) f_n^{(q)} - 4g_{n-1} f_{n-1}^{(q)} + g_{n-2} f_{n-2}^{(q)} = 0\,. \tag{13}$$

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

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

where

$$\begin{aligned} P &= g_{n-2}\big(3h\mu_n - 2g_n\big) - h\mu_{n-2}\,\big(3g_n - 4h\mu_n\big)\,, \\ Q &= 4g_{n-1}\big(h\mu_n - \frac{g_n}{2}\big)\,, \\ R &= \frac{4}{r}\,\big(3g_n - 4h\mu_n\big)\,. \end{aligned}$$

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

$$\Big(1 + \frac{r}{2}\,g_1\Big) f_1^{(q+1)} - \frac{r}{4}\Big(g_2 - h\mu_2\Big) f_2^{(q+1)} = \Big(1 - \frac{r}{2}\,g_1\Big) f_1^{(q)} + \frac{r}{4}\Big(g_2 - h\mu_2\Big) f_2^{(q)} \tag{15}$$

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 $(t_s + qk + k)$ given the PDF at $(t_s + qk)$. Note that the tri-diagonal system described by equations (11), (14) and (15) is solved for the transitional density at nodes $x_1, \cdots, x_{n-1}$. Here the transitional density at x_0 is known *a priori* to be zero, and the final transitional density at x_n may be obtained by means of relation (13).

3.2 Transitional CDF Specification

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

$$\frac{\partial F}{\partial t} = \frac{1}{4}\Big[\frac{\partial^2 (gF)}{\partial x^2} + g\frac{\partial^2 F}{\partial x^2} - F\frac{\partial^2 g}{\partial x^2}\Big] - \mu\frac{\partial F}{\partial x}\,. \tag{16}$$

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) in repeated for equation (16). The calculation is routine and so the details are suppressed. If $F_p^{(q)} = F(x_p, t_s + qk)$ denotes the value of the CDF at $(x_p, t_s + qk)$, then the finite-difference approximation of equation (16) is

$$\begin{aligned} &-\Big[g_{p-1} + g_p + 2h\mu_p\Big] F_{p-1}^{(q+1)} \\ &+\Big[\frac{8}{r} + (g_{p-1} + 2g_p + g_{p+1})\Big] F_p^{(q+1)} \\ &-\Big[g_p + g_{p+1} - 2h\mu_p\Big] F_{p+1}^{(q+1)} \\ &= \Big[g_{p-1} + g_p + 2h\mu_p\Big] F_{p-1}^{(q)} \\ &+\Big[\frac{8}{r} - (g_{p-1} + 2g_p + g_{p+1})\Big] F_p^{(q)} \\ &+\Big[g_p + g_{p+1} - 2h\mu_p\Big] F_{p+1}^{(q)}\,. \end{aligned} \tag{17}$$

[2] In this application, the upper limit of the finite interval, x_n, is chosen as the maximum of the sample plus the range of the sample.

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

$$T_{\mathrm{L}}\,\mathbf{F}^{(q+1)} = T_{\mathrm{R}}\,\mathbf{F}^{(q)} + \mathbf{B}^{(q)}$$

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 $x_1, \cdots, x_{n-1}$ at time $(t_s + qk)$.

In the practical implementation of this procedure, there is a natural initial condition given in equation (9), for which there is no equivalent statement in the transitional PDF formulation of the problem. Of course, the Jensen and Poulsen (2002) approximation can also be used in the transitional CDF approach by initialising $F_1, \cdots, F_{n-1}$ using the CDF of the normal distribution. The value of the transitional density function at (X_{s+1}, t_{s+1}) is estimated by numerical differentiation of the transitional CDF at the nodes to the left and right of X_{s+1} followed by linear interpolation of these values to find the required transitional density.

4 LOG-LIKELIHOOD COMPUTATION

The efficacy of the new procedure depends on the accuracy with which the log-likelihood of the data is computed. This section outlines a simulation experiment to compare the accuracy of log-likelihood computation based on the (traditional) 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-based approaches, the initial condition for the cumulative approach will be specified with and without the aid of the CDF of the normal distribution.

The details of the simulation exercise are now described. For each estimation procedure 2000 repetitions of the calculation of the log-likelihood for samples containing $T = 500$ observations 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 1000 time steps of size 0.001 between observations. The maximum relative error, mean absolute relative error and the mean squared relative error were calculated from the 2000 repetitions and are presented in Table 1.

The most significant finding is that the CDF approach using the theoretical step function initial condition in equation (9) can be implemented successfully in practice. By contrast, there is no equivalent result for the PDF-based procedure, because it is always necessary to approximate the initial condition. This approximate initial condition is is not as innocuous as it would appear, since a fine mesh of nodes in the vicinity of the initial condition is required to resolve the approximation. As this spatial resolution has to be maintained over the entire sample space there are obvious implications for the computational cost.

The main empirical result to emerge from this simulation exercise is that the CDF approach, using the cumulative normal initial condition, is always superior to the PDF approach for all combinations of h and k. The CDF approach based on the theoretical initial condition performs with credit and is more accurate than the probability density approach for a coarse discretisation[3] of state space (*e.g.* $h = 0.005$). This is most likely due to the poor resolution of the initial transitional density when using a coarse discretisation of state space, since under such circumstances the function is distinguishable from zero at a small number of nodes, or perhaps none at all for a particularly coarse discretisation. By contrast, the impact of a coarse discretisation of state space on the cumulative approach is to make the discontinuity in the initial step function less steep. As the discretisation of state space is refined, the PDF approach is more accurate than the CDF approach starting with the theoretical initial condition.

A final result worthy of passing comment occurs when the CDF approach starting with the theoretical initial condition is used with a fine discretisation of state space and crude discretisation of temporal space, for example, Table 1 with $h = 0.001$ and $k = 0.02$. 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.

As an additional check on the efficacy of the CDF method a second simulation exercise was undertaken. In this experiment, the data generated for the

[3]This may be important when a coarse discretisation of state space is necessary due to computational considerations, for example, if the problem requires working in two or more dimensions.

	h	Measures of relative error Maximum Absolute	Mean Absolute	Mean Squared
Probability Density Approach				
	0.005	9.266×10^{-3}	1.949×10^{-3}	5.822×10^{-6}
$k = 0.02$	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}
$k = 0.01$	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}
Cumulative Distribution Approach				
	0.005	4.370×10^{-3}	1.078×10^{-3}	1.744×10^{-6}
$k = 0.02$	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}
$k = 0.01$	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}
Cumulative Distribution Approach with equally informative initial condition				
	0.005	2.744×10^{-3}	4.071×10^{-4}	2.565×10^{-7}
$k = 0.02$	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}
$k = 0.01$	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}

Table 1. Measures of relative error in the calculation of log-likelihood for the CIR process using the various finite difference procedures and based on 2000 simulations.

Estimation procedure	Mean Parameter Estimates α	β	σ
Probability density function	0.2095	0.0802	0.1001
Cumulative distribution function	0.2099	0.0802	0.1000
Cumulative distribution function with equally informative initial condition	0.2096	0.0802	0.0999

Table 2. Mean parameter estimates from 2000 simulations of the process $dX = \alpha(\beta - X)dt + \sigma\sqrt{X}\,dW$ with parameters $\alpha = 0.2$, $\beta = 0.08$ and $\sigma = 0.10$ using the method of finite differences with spatial resolution $h = 0.002$ and temporal resolution $k = 0.02$.

first simulation exercise was used to estimate the parameters of the underlying CIR model by both the PDF and CDF methods. The mean of the parameter estimates over the 2000 samples are presented in Table 2. Given the small errors recorded in Table 1 in estimating the likelihoods, it is not surprising that all of the approaches considered in the paper deliver what are effectively identical parameter estimates.

5 CONCLUSION

This paper 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 propose 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 which seek to solve the Fokker-Planck equation. The method is easier to implement than one based on the direct solution of the Fokker-Planck equation, first because initial conditions are more amenable to numerical work, and second, because the gradient-like boundary conditions associated with the Fokker-Planck equation are replaced by Dirichlet conditions in the modification. A parameter estimation exercise for the CIR process indicates that the method is both more accurate and more robust than the traditional method.

6 REFERENCES

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

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

Lo, A.W. (1988). Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data. *Econometric Theory*, **4**, 231-247.

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