Bayesian Inference in Reducible Stochastic Differential Equations

Ludovic GIET[1] Michel LUBRANO[2]

October 6, 2004

Abstract

The linear Ornstein-Ulenbeck diffusion model is too simple to describe the movement of short term interest rates. However diffusions with a non-linear drift and volatility function have no closed form likelihood function which make inference either classical or Bayesian very problematic. A vast range of approximation were proposed in the literature. In this paper, we develop the idea of a non-linear diffusion model, which after transformation can be reduced to an Ornstein-Uhlenbeck. At the price of a constrained drift function, we get a model equipped with a closed form likelihood function. We test this class of models on the the US Federal fund rate data. We propose a Bayesian approach to compare the performance of various specification of the volatility function.

Keywords: Bayesian inference, continuous time, reducible diffusions, model evaluation, interest rate.
JEL Classification: C11, C13, C52, E43

[1]GREQAM, Marseille.
[2]GREQAM and CNRS, Marseille

1 Introduction

Empirical interest rate models usually involve a non-linear stochastic differential equation of the form $dr = \mu(r)dt + \sigma(r)dW$ where $\mu(r)$ is a non-linear drift function and $\sigma(r)$ a non-linear volatility function. This writing means that the interest rate r is a solution of the integral equation

$$r_t - r_0 = \int_0^t \mu(r_u)du + \int_0^t \sigma(r_u)dW_u. \tag{1}$$

The option of modelling interest rates using a continuous time approach is taken because valuating options is most of the time not possible in discrete time (see for instance Merton's derivation of the Black-Scholes formula). However, inference in continuous time processes is not a simple matter. Exact inference by maximum likelihood is in general possible only if the above integral equation possess an analytical solution implying the existence of an exact discretization. This happens only for a very limited class of processes: the Brownian and geometric Brownian motions and the Ornstein-Uhlenbeck process to speak quickly. The scarcity of exact solutions may find an explanation in the fact that a stochastic process which follows a SDE presents important parametric restrictions. Let us take the exemple of the Ornstein-Uhlenbeck process written as

$$dr_t = (\phi - \lambda r_t)dt + \sigma dW_t. \tag{2}$$

The exact discretization of this process is an AR(1) with an autoregressive coefficient equal to $\exp(-\lambda)$. This coefficient is all the time positive which means that the first order autocovariance of the process is strictly positive. Another exemple is the geometric Brownian motion which implies a log normal transition probability that imbeds a dependence between the drift and the volatility parameters. If we choose to consider a Euler discretization of the process, this type parametric constraint or dependence is broken. This explains first the possibility of a discretization biais, and second the fact that stationarity conditions are in general not the same for the model in continuous time and for its Euler discretisation (see Conley, Hansen, Luttmer, and Scheinkman (1997)).

The aim of this paper is to study a class of non-linear SDE for which there exists a non-linear transformation such that the transformed SDE is linear and possess an exact discretization. We shall obtain a non-linear SDE where we can rather freely choose the shape of the volatility function. But the corresponding drift function will be very much constrained. A similar approach is presented in Kloeden and Platen (1999), page 113 to define classes of SDE which are explicitly solvable, but to our knowledge this idea has not been applied for empirical modelling.

This approach has pro's and con's. The maximum likelihood approach is directly implementable, which mean that Bayesian inference is tractable at the cost of a numerical integration of moderate dimension. Model choice is also possible and will be necessary to chose the correct and interesting transformations to introduce. Basically this choice concerns only the shape of the volatility function $\sigma(r)$. The serious drawback is that there no degree of freedom on the shape of the drift function which is totaly determined by the shape of $\sigma(r)$ and the choice of the underlying linear model. We shall thus have to discuss if that cost is not too high and does not yield unrealistic drift functions.

The paper is organised as follows. Section 2 presents a class of reducible SDE and its exact Ornstein-Uhlenbeck discretization. Section 3 is devoted to the CEV model and its CEV-OU version. It presents the stationarity conditons. Section 4 discuss Bayesian inference and model choice. Section 5 compares various empirical model for the US short term interest rate. Section 6 concludes.

2 Reducible SDE

2.1 Linear SDE

A general linear SDE is noted as in Kloeden and Platen (1999)

$$dx_t = (a_1 x_t + a_2)dt + (b_1 x_t + b_2)dW_t. \tag{3}$$

This general writing allows us to recover some of the usual linear SDE having an exact discretisation. The geometric Brownian motion corresponds to $a_2 = b_2 = 0$ while the simple Brownian motion corresponds to $a_1 = b_1 = 0$. The Ornstein-Uhlenbeck process is obtained for $b_1 = 0$. The model of Brennan and Schwartz (1979) is obtained for $b_2 = 0$. The general specification (3) admits an exact discretization:

Theorem 1 *The exact discretization of the linear SDE (3) is*

$$x_t = \phi(t) \left[x_{t_0} + (a_2 - b_1 b_2) \int_{t_0}^{t} \frac{1}{\phi(u)} du + b_2 \int_{t_0}^{t} \frac{dW_u}{\phi(u)} \right] \tag{4}$$

where

$$\phi(t) = \exp \left[\left(a_1 - \frac{b_1^2}{2} \right) (t - t_0) - b_1 \left(W_t - W_{t_0} \right) \right] \tag{5}$$

and initial condition $\phi(0) = 1$.

Proof: see appendix A.

In practice Theorem 1 is not very useful. It gives the exact discretization for cases corresponding to either $b_1 = 0$ and $b_2 \neq 0$ (Ornstein-Uhlenbeck or Brownian motion) or $b_1 \neq 0$ and $a_2 = b_2 = 0$ (Geometric Brownian). In these polar cases, W_t never appears twice in the right hand side product (4) which is then easy to evaluate. If we now for instance relax the constraint $a_2 = 0$ in this last case, we have the model of Brennan and Schwartz (1979), the exact discretisation of which is

$$\begin{aligned} x_t &= \phi(t) \left[x_{t_0} + a_2 \int_{t_0}^{t} \frac{du}{\phi(u)} \right] \\ \phi(t) &= \exp \left(\left(a_1 - \frac{b_1^2}{2} \right) (t - t_0) + b_1 (W_t - W_{t_0}) \right). \end{aligned} \tag{6}$$

It is not evident to find an analytical solution to

$$\int \exp\left[-\left(a_1 - \frac{b_1^2}{2}\right)(u - t_0) - b_1(W_u - W_{t_0})\right] du. \tag{7}$$

Consequently, the most convenient linear SDE possessing an exact discretisation are the Ornstein-Uhlenbeck process and the geometric Brownian motion. Their associated likelihood functions are respectively normal and log-normal.

2.2 A class of admissible transformations

Let us consider the non-linear SDE in Y_t

$$dy_t = \mu(y_t)dt + \sigma_0\sigma(y_t)dW_t \tag{8}$$

where σ_0 is a scale parameter and $\sigma(y_t)$ is supposed to be normalised. Apart from this normalising constraint, the functional forms of the drift $\mu(y_t)$ and volatility $\sigma(y_t)$ function are left totaly unspecified. Let us call $D_Y = (\underline{y}, \overline{y})$ the domain of definition of the diffusion. We shall consider two cases: either D_Y covers the real line or D_Y is confined to $\mathbb{R}^+$.

We are looking for a transformation $x_t = U(y_t)$ such that x_t follows a SDE which has an exact discretisation. Applying Ito's lemma to find the SDE of $x_t = U(y_t)$ starting from (8), we have

$$dx_t = \left(\mu(y_t)U'(y_t) + \frac{1}{2}\sigma_0^2\sigma^2(y_t)U''(y_t)\right) dt + \sigma_0\sigma(y_t)U'(y_t)dW_t. \tag{9}$$

Of course it will not be possible to find such a transformation in the general case. We are thus faced to three questions. Under which conditions is it possible to find such a transformation, what is the transformation and finally what are the restrictions implied on $\mu(.)$ and $\sigma(.)$. We shall restrict our attention to the case where the transformed process x_t follows an Ornstein-Uhlenbeck (OU). The OU process is very convenient for modelling interest rates as it is a stationary process while the geometric Brownian motion is non-stationary.

2.3 Ornstein-Uhlenbeck solutions

An Ornstein-Uhlenbeck process is defined as

$$dx_t = (a_1x_t + a_2)dt + b_2dW_t. \tag{10}$$

With such a target, the shape of the transformation is found by identifying the volatility of the two SDE (10)-(9). It implies first that $\sigma_0 = b_2$ and second that

$$U'(y) = 1/\sigma(y). \tag{11}$$

Solving this differential equation gives

$$U(y) = \int^y \frac{1}{\sigma(u)}du + c \tag{12}$$

where c depends on the lower bound of integration when necessary. Due to the particular configuration of the OU process, the transformation $U(.)$ is reduced to a mere standardisation of the process as that used for instance in Ait-Sahalia (2002) or Durham and Gallant (2002) [1]. We have thus answered to question 2. This transformation exists iff $1/\sigma(y)$ is integrable (question 1).

We can now find the shape of the drift function by identifying the first member of (9) with that of the drift of the OU process and replacing $U(y)$ by its expression:

$$\mu(y) = \sigma(y)\left[\frac{1}{2}\sigma_0^2\sigma'(y) + a_1 U(y) + a_2\right]. \tag{13}$$

Consequently, we can find a non-linear reducible SDE by specifying a volatility function which has an integrable inverse and a drift function which verifies (13) (question 3). This is exactly equation (4.47, chapter 4) in Kloeden and Platen (1999) except for the additional terms a_2 and σ_0.[2]

Remarks:

- If $D_Y = (0, +\infty)$, D_X, the domain of definition of X has to cover the same range. Consequently, depending on the parameter configuration, we may have to consider $x = -U(y,\theta)$ instead of $x = U(y,\theta)$, so that $x = |U(y,\theta)|$. The implied drift is then transformed into

$$\mu(y) = \sigma(y)\left[\frac{1}{2}\sigma_0^2\sigma'(y) + a_1|U(y)| + a_2\mathrm{sign}(U(y))\right]. \tag{14}$$

- The constant of integration c in (12) may play a major role for certain models. It depends on the lower bound of integration z. Let us call $F(y)$ a primitive of $1/\sigma(y)$. The function $U(y)$ is defined as

$$U(y) = F(y) - F(z). \tag{15}$$

  We shall see below that for the CEV model one has to choose $z = 1$ to insure the continuity of U in the parameters.

We can now write the solution of this reducible SDE. Let us start from the solution of the OU SDE (10). When the two dates of interest are t and $t - \Delta$, and Δ is the discretisation step, this solution is

$$x_t = -\frac{a_2}{a_1}(1 - e^{a_1\Delta}) + e^{a_1\Delta}x_{t-\Delta} + b_2\left(\frac{e^{2a_1\Delta} - 1}{2a_1}\right)^{1/2}\epsilon_t, \tag{16}$$

where ϵ_t is a Gaussian white noise of zero mean and unit variance. We have just to apply the inverse transformation $U^{-1}(.)$ to obtain the exact discretization of the process in y_t as $x_t = U(y_t)$:

$$y_t = U^{-1}\left(-\frac{a_2}{a_1}(1 - e^{a_1\Delta}) + e^{a_1\Delta}U(y_{t-\Delta}) + b_2\left(\frac{e^{2a_1\Delta} - 1}{2a_1}\right)^{1/2}\epsilon_t\right). \tag{17}$$

[1] These authors do not consider the case $c \neq 0$. Durham and Gallant (2002) says that c is irrelevant. This is true when it is possible to get an analytical solution for (12). If (12) has to be computed numerically, c must compensate for the lower bound of integration.

[2] Kloeden and Platen (1999) consider a Langevin SDE instead of an Ornstein-Uhlenbeck SDE. We can go from one to the other by an affine transformation of the data.

The conditional density of $x_t|x_{t-1}$ is

$$p_x(x_t|x_{t-\Delta}) = \frac{1}{\sqrt{2\pi v^2}} \exp\left(-\frac{1}{2v^2}\left[x_t + \frac{a_2}{a_1}(1-e^{a_1\Delta}) - e^{a_1\Delta}x_{t-\Delta}\right]^2\right), \tag{18}$$

with $v^2 = b_2^2(e^{2a_1\Delta}-1)/(2a_1)$. The conditional density of $y_t|y_{t-\Delta}$ is found using the Jacobian of the transformation $J(y \to x) = U'(y) = 1/\sigma(y)$:

$$f_y(y_t|y_{t-\Delta}) = \frac{1}{\sigma(y_t)\sqrt{2\pi v^2}} \exp\left(-\frac{1}{2v^2}\left[U(y_t) + \frac{a_2}{a_1}(1-e^{a_1\Delta}) - e^{a_1\Delta}U(y_{t-\Delta})\right]^2\right). \tag{19}$$

The associated likelihood function has to be maximised numerically. If the Jacobian $1/\sigma'(y_t)$ were constant, this density would belong to the normal family. In most of the usual cases, the transformation $U(y_t)$ which involves an integral can be computed analytically (see examples below). However, for some non-standard specifications of the volatility, this computation has to be done numerically.

2.4 Alternative approaches

In order to achieve at feasible solutions in implementing the maximum likelihood estimator, Ait-Sahalia (1999) has chosen to approximate the true likelihood function of a non-linear SDE by an Hermite expansion around the Normal density. The first step of the method consists in reducing the SDE by a transformation which is identical to U and makes the volatility of the SDE scalar. With a scalar volatility, the true unknown likelihood is closer to the Normal density. The Hermite expansion produces very rapidly very long expression, several pages for the unconstrained CEV model.

As an alternative to this approach, we have proposed to constrain the drift function in such a way that the standardised diffusion $x_t = U(y_t)$ is an OU process having an explicit parametric transition density. If we turn to the original process y_t, the transition probability density is (19) in our case while it has the following formulation in Ait-Sahalia paper using our notations

$$f_y^K(y_t|y_{t-\Delta}, \theta) = \frac{1}{\sigma(y_t)} p_x^K(U(y_t)|U(y_{t-\Delta}), \theta). \tag{20}$$

K represents the order of the approximation while $p_x^K(x_t|x_{t-\Delta})$ represents the Hermite approximation of the true density of the reduced process around the Normal density. In our case $p_x^K(.)$ is simply the Normal density associated with the OU reduced process $x = U(y)$. The factor $1/\sigma(y)$ is common to both approaches.

The most common way of finding an approximate likelihood function is to discretise the original SDE using an Euler approximation:

$$y_t - y_{t-\Delta} = \mu(y_{t-\Delta}, \theta)\Delta + \sigma_0\sqrt{\Delta}\sigma(y_{t-\Delta}, \theta)\epsilon_t \qquad \epsilon \sim \mathrm{N}(0,1). \tag{21}$$

The corresponding transition probability

$$f_y(y_t|y_{t-\Delta},\theta) = f_N(y_t - y_{t-\Delta}|\mu(y_{t-\Delta},\theta)\Delta, \sigma_0^2\sigma^2(y_{t-\Delta},\theta)\Delta). \tag{22}$$

is a plain normal density at odds with (19) and (20). This approach suffers of course from a discretisation biais while the two other ones do not.

2.5 Uncovered models

The starting point of our approach is to choose the volatility function and to deduce the implied drift function. The reverse does not seem to be feasible. Let us for a while consider as a starting point the non-linear drift function

$$\mu(y) = c_0 + c_1 y + c_2 y^2 + c_3/y, \tag{23}$$

as promoted in Ait-Sahalia (1996). We have now to find a parametric form for $\sigma(y)$ such that the reduced model is an Ornstein-Uhlenbeck. This means solving the following differential equation in $\sigma(y)$

$$(\mu'(y) - a_1)\,\sigma(y) - \mu(y)\sigma'(y) - 0.5\sigma_0^2\sigma^2(y)\sigma''(y) = 0. \tag{24}$$

It does not seems easy to find an analytic solution to this second order non-linear differential equation. Consequently, the approach is rather flexible when starting from the volatility function. A wide range of specification is feasible. The approach consisting in starting from the drift function does not seem to be feasible.

3 Stationarity conditions

A usual OU process is stationary under very mild conditions, essentially $a_1 < 0$. As soon as we transform this process considering $x = U(y)$, it is not evident that the transformed process remains stationary. We shall consider in this section two alternative models of the volatility.

The constant elasticity volatility model has been introduced by Chan, Karolyi, Longstaff, and Sanders (1992) and is noted

$$dy = (c_1 y + c_2)dt + \sigma_0 y^\gamma dW. \tag{25}$$

These authors claim that it was their best fitting model. This model was further studied by Ait-Sahalia (1996) who promoted the use of a non-linear drift function to provide a better mean reversion. The same type of specification was also estimated by Conley, Hansen, Luttmer, and Scheinkman (1997) and Gallant and Tauchen (1998). We can thus consider this model and its variants as a convincing benchmark model. It has no exact discretisation. Its Euler discretisation is stationary only when $c_1 < 0$ and $\gamma \leq 1$ (see Broze, Scaillet, and Zakoïan (1995)), whereas the continuous time specification is stationary under much more general conditions (see Conley, Hansen, Luttmer, and Scheinkman (1997)) and in particular when $\gamma > 1$, value which is often encountered in empirical applications using US short term interest rates.

Pfann, Schotman, and Tschernig (1996) considerably improved Vasicek's model by allowing for a two regime volatility function. Volatility is supposed constant and equal to σ_1 in a regime of low interest rates and becomes equal to σ_2 in a regime corresponding to high interest rates in their discretised model:

$$\Delta y_t = \begin{cases} c_{11} y_{t-1} + c_{12} + \sigma_1 \epsilon_t & y_{t-1} < c \\ c_{21} y_{t-1} + c_{22} + \sigma_2 \epsilon_t & y_{t-1} \geq c. \end{cases} \tag{26}$$

We shall propose a continuous-time version of their model, using a smooth logistic transition function $F(y)$ at values in $[0, 1]$.

3.1 The CEV-OU model

The general CEV model we want to consider is

$$dy = \mu(y) + \sigma_0 y^\gamma dW. \tag{27}$$

In order to find an exact discretisation, let us impose that the transformed process $x = U(y)$ is an OU. The requested transformation $U(y)$ is

$$\begin{aligned} U(y) &= \int_1^y \frac{1}{u^\gamma} du = \frac{y^{1-\gamma} - 1}{|(1-\gamma)|} && \text{for } \gamma \neq 1 \\ &= \ln(y) && \text{for } \gamma = 1. \end{aligned} \tag{28}$$

Considering a lower bound of integration equal to 1 insures that the transformation is continuous in γ. Using (13), we have

$$\begin{aligned} \mu(y) &= y^\gamma \left[\frac{\sigma_0^2 \gamma}{2} y^{\gamma-1} + a_1 \frac{y^{1-\gamma} - 1}{|1-\gamma|} + a_2 \text{sign}(1-\gamma) \right] && \text{for } \gamma \neq 1 \\ \mu(y) &= y \left[\frac{\sigma_0^2}{2} + a_1 \ln y + a_2 \right] && \text{for } \gamma = 1 \end{aligned} \tag{29}$$

This drift function is thus very different from the linear drift function (25). It is both highly non-linear and parsimonious. We show in the next subsections that it induces mean reversion under fairly general conditions.

3.2 The ST-OU model

A general continuous-time model with a change in volatility similar to (26) can be

$$dy = \mu(y)dt + [\sigma_0(1 - F(y)) + \sigma_1 F(y)]dW, \tag{30}$$

where $F(y)$ is a smooth-transition function, whose specification is not yet chosen. It is convenient to factorize the volatility function as

$$\sigma(y) = 1 + (sk - 1)F(y) \tag{31}$$

where $sk = \sigma_1/\sigma_0$. Consequently

$$dy = \mu(y)dt + \sigma_0\sigma(y)dW. \tag{32}$$

The parameter σ_0 is free while sk is constrained to be greater than 1. We can find in the literature a wide class of smooth transition functions like for instance $\arctan(y)/\pi + 0.5$. However, for further computations the integrability of $1/\sigma(y)$ is highly desirable. The usual logistic function

$$F(y) = \frac{1}{1+\exp(-\gamma(y-c))} \tag{33}$$

fulfils this requirement. Parameter c is the threshold separating regimes of low and high volatility. Parameter γ monitors the speed of adjustment and has to be strictly positive for identification purposes. For $\gamma = 0$, we have a single regime and sk is not identified. The induced transformation is

$$U(y) = y - 1 - \left(\frac{sk-1}{sk\,\gamma}\right)\ln\frac{sk\,e^{\gamma y} + e^{\gamma c}}{sk\,e^{\gamma} + e^{\gamma c}}. \tag{34}$$

Contrary to the CEV-OU model, the transformation (34) is always continuous. The corresponding drift function has a rather long algebraic expression, but can easily be derived from (13).

In the limiting case of $\gamma \to \infty$, the smooth transition function (33) becomes the Dirac function defined as

$$F(y) = \mathbb{1}(y-c) = \begin{cases} 1 & \text{if } y > c \\ = 0 & \text{otherwise} \end{cases}$$

The transformation $U(y)$ given in (34) adopt a much simpler form with

$$U(y) = y - 1 - (y-c)(sk-1)/sk\mathbb{1}(y-c)$$

The corresponding drift function is linear with a break at $y = c$.

$$\mu(y) = [1 + (sk-1)\mathbb{1}_{y-c}][a_1(y - 1 - (y-c)\frac{sk-1}{sk}\mathbb{1}_{y-c}) + a_2]$$

3.3 General stationarity conditions for a scalar SDE

In any general SDE, it is always possible to define the transition density which is the density of y_s given y_t with $s > t$. The SDE is said to be stationary if the limit of this density is well defined for $s \to \infty$ and is by the way equal to the marginal density of the process. In order to explore under which condition a scalar SDE is stationary, we need to define the scale and speed densities. The scale density is defined as

$$s(y) = \exp\left\{-2\int^y \frac{\mu(v)}{\sigma^2(v)}dv\right\} \tag{35}$$

from which we define the scale function $S(y)$

$$S(y) = \int^{y} s(v)dv. \tag{36}$$

The speed density is

$$m(y) = \frac{1}{\sigma^2(y)} \exp\left\{2\int^{y} \frac{\mu(v)}{\sigma^2(v)}dv\right\} \tag{37}$$

and is in fact proportional to the marginal density of the observations.

A scalar diffusion process is stationary if three sufficient conditions are met:

1) The diffusion coefficient is strictly positive: $\sigma^2(y) > 0$.

2) The scale function diverges at both boundaries, which means that the process cannot reach its boundaries (non-exploding solution):

$$\int_{\underline{y}}^{y} s(v)dv = \infty \qquad \int_{y}^{\overline{y}} s(v)dv = \infty \quad \forall y \in D_Y.$$

3) The speed density is integrable on D_Y.

Under these conditions the stationary density is

$$\pi(y) = \frac{A}{\sigma^2(y)} \exp\left\{2\int^{y} \frac{\mu(v)}{\sigma^2(v)}dv\right\} \tag{38}$$

where A is such that $\pi(y)$ integrates to unity. These conditions can be found for instance in Ait-Sahalia (1996) or Rao (1999, page 178) or Lund (1999).

3.4 Stationarity conditions for a reducible OU model

In an OU reducible SDE model, the scale and speed densities have a natural expression:

$$\begin{aligned} s(y) &= \frac{1}{\sigma(y)} \exp\left(-\frac{a_1}{\sigma_0^2}U^2(y) - 2\frac{a_2}{\sigma_0^2}U(y)\right) \\ m(y) &= \frac{1}{\sigma_0^2\sigma(y)} \exp\left(\frac{a_1}{\sigma_0^2}U^2(y) + 2\frac{a_2}{\sigma_0^2}U(y)\right). \end{aligned} \tag{39}$$

It is worth noting that these expressions are correct whatever the sign of $U(y)$. A sufficient condition leading to the divergence of the scale function is that $a_1 < 0$ and $\lim U(y) = \pm\infty$ when y reaches its boundaries. It is evident from (39) that when $s(y)$ diverges, then $m(y)$ is integrable (because of the properties of the exponential). Consequently, we just have to study the behaviour of $U(y)$.

For the CEV model, $U(y)$, given in (28) diverges at both boundaries for $\gamma < 1$ and for $\gamma = 1$, provided the domain of definition is extended to $\mathbb{R}$ using a symmetric argument. However it does not diverge for $\gamma > 1$.

For the ST model, $U(y)$, as given in (34) is of the same order as y for any finite sk. Consequently, $U(y)$ diverges at both boundaries using the same symmetric argument as above.

We can conclude that the ST-OU model is stationary provided $a_1 < 0$. The CEV-OU model also requires that $a_1 < 0$, but imposes the additional condition $\gamma \leq 1$ to be stationary. We must note however that the last condition is only a sufficient condition. We can encounter particular configurations of the parameters where $\hat{\gamma} > 1$ while $\mu(\hat{\theta})$ presents an obvious mean reversion. See the empirical application below.

4 Likelihood inference

4.1 Ornstein Uhlenbeck processes

Bayesian inference in the usual Ornstein Uhlenbeck SDE is straightforward provided we reparametrise the model in the following way:

$$\left\{ \begin{array}{rcl} \mu & = & -(1 - e^{a_1 \Delta}) a_2 / a_1 \\ \rho & = & e^{a_1 \Delta} \\ v^2 & = & b_2^2 (e^{2a_1 \Delta} - 1)/(2a_1). \end{array} \right. \tag{40}$$

In this new parameterisation, the likelihood function (18) becomes identical to that of a linear regression model with

$$l(x, \mu, \rho, v^2) \propto v^{-T} \exp\left(-\frac{1}{2v^2} \sum [x_t - \mu - \rho x_{t-\Delta}]^2\right). \tag{41}$$

This transformation moreover provides interpretable coefficients. μ is related to the long term mean of the process, ρ is an autoregressive coefficient and v^2 is the variance of the error term. The stationarity condition $a_1 < 0$ is immediately translated in $\rho < 1$.

We must note that the original parameterisation contained the restriction $\rho > 0$. With the new parameterisation, that restriction is no longer automatic. It has to be imposed by an adequate prior. For instance, we can have

$$\pi(\mu, \rho, v^2) \propto 1/v^2 \mathbb{1}(\rho > 0),$$

where $\mathbb{1}(.)$ is the Dirac function. Using this prior, a textbook result (see e.g. Bauwens, Lubrano, and Richard (1999)) indicates that the posterior density of (μ, ρ) is a truncated Student and that of v^2 is an inverted gamma2. More precisely, let us define

$$x = [x_t] \qquad X = [1, x_{t-\Delta}] \qquad \beta' = [\mu, \rho],$$

and the sufficient statistics

$$\begin{cases} \hat{\beta} &= (X'X)^{-1}X'x \\ s^2 &= x'x - x'X(X'X)^{-1}X'x. \end{cases}$$

Then

$$\begin{cases} \pi(\beta|x) &= f_t(\beta|\hat{\beta}, X'X, s^2, T)\mathbb{I}(\rho > 0) \\ \pi(v^2|x) &= f_{i\gamma}(v^2|s^2, T). \end{cases}$$

Posterior draws can easily be obtained from these densities, rejecting negative draws for ρ. Draws in the original parameterisation are obtained by solving the system

$$\begin{cases} a_1 &= \Delta^{-1} \log \rho \\ a_2 &= -\dfrac{\mu}{\Delta(1-\rho)} \log \rho \\ b_2 &= v\Delta^{-1/2}\sqrt{\ln \rho^2/(\rho^2 - 1)}. \end{cases} \tag{42}$$

Further restrictions can be imposed: a positive long term mean implies rejecting negative draws for μ; stationarity means rejecting draws of ρ which are greater than 1.

4.2 Bayesian inference for the general model

The complete likelihood function corresponding to (19) is

$$l(y; \theta, \mu, \rho, v^2) \propto v^{-T} \prod \sigma(y_t, \theta)^{-1} \exp\left(-\frac{1}{2v^2}[U(y_t, \theta) - \mu - \rho U(y_{t-\Delta}, \theta)]^2\right), \tag{43}$$

where θ is the parameter involved in the transformation U. We note immediately that the previous change of parameterisation can again be suggested even if this likelihood function differs from (41) because of the presence of a Jacobian. This reparameterisation is convenient because, conditionally on θ, we recover the previous linear regression model. The previous semi-diffuse prior can be extended as

$$\pi(\theta, \mu, \rho, v^2) = \pi(\theta)\pi(\mu, \rho, v^2) \propto 1/v^2\mathbb{I}(\rho > 0) \tag{44}$$

Let us define

$$y(\theta) = [U(y_t, \theta)] \qquad Z(\theta) = [1, U(y_{t-\Delta}, \theta)]. \tag{45}$$

The conditional posterior densities of β and v^2 given θ have the same usual form. More precisely, defining the conditional sufficient statistics:

$$
\begin{aligned}
\hat{\beta}(\theta) &= (Z(\theta)'Z(\theta))^{-1}Z(\theta)'y(\theta) \\
s^2(\theta) &= y(\theta)'y(\theta) - y(\theta)'Z(\theta)(Z(\theta)'Z(\theta))^{-1}Z(\theta)'y(\theta),
\end{aligned} \tag{46}
$$

the conditional posterior densities of β and v^2 are truncated Student and inverted gamma2 densities with

$$
\begin{aligned}
\pi(\beta|y,\theta) &= f_t(\beta|\hat{\beta}(\theta), Z(\theta)'Z(\theta), s^2(\theta), T)\,\mathbb{1}(\rho > 0) \\
\pi(v^2|y,\theta) &= f_{i\gamma}(v^2|s^2(\theta), T).
\end{aligned} \tag{47}
$$

The marginal posterior density of θ belongs to an unknown family and is formed by the product of the Jacobian of the transformation $U(.)$ coming from the likelihood function and of the inverse of the integrating constant of the conditional posterior density of $\beta|\theta$. We have:

$$
\pi(\theta|y) = |Z(\theta)'Z(\theta)|^{-1/2} s^2(\theta)^{-(T-2)/2} \prod \sigma(y_t, \theta)^{-1} \pi(\theta). \tag{48}
$$

Posterior draws of θ can be obtained using a Griddy-Gibbs sampler (see Bauwens and Lubrano (1998)). Once these draws are stored, corresponding draws of β and v^2 are obtained directly from their conditional posterior densities, simply rejecting negative draws of $\rho(\theta)$.

4.3 The special case of the ST-OU model

In section 3.2, we have detailed the specification of a ST-OU model. The Bayesian treatment of this model contains some specific difficulties due to the presence of two regimes and to the shape of the transition function. To speak quickly, Lubrano (2000) shows that smooth transition models cannot be analysed without an informative prior. There is a problem of identification and a problem of integrability.

1. When $\gamma \to 0$, the smooth transition function $F(y)$ is constant and equal to $1/2$. The volatility function becomes

   $$\sigma_0\sigma(y,\theta) = \sigma_0(1 + (sk-1)/2).$$

   Consequently, sk is no longer identified.

2. When $\gamma \to +\infty$, the smooth transition function converges to the Dirac function which is equal to zero for $y < c$ and to 1 for $y \geq c$. In discrete time, we would have a sharp regime change in the volatility depending on the sign of $(y-c)$. In continuous time, it is less evident that this is a well defined model. But the associated likelihood function is well defined. Consequently, the region corresponding to $\gamma \to +\infty$ has a non-zero probability. The posterior density of γ is thus not integrable in the absence of a prior density that would put a zero weight on that region.

The two parameters that create problems, sk and γ are elements of θ in our notations. Under a non-informative prior, the posterior density of θ given in (48) is already non-standard. Thus we are more free for choosing the type of prior needed for sk and γ. Let us first consider the case of an informative prior on sk. This parameter is scale free. It represents the proportion of volatility increase in the second regime of high volatility. It has to be greater than 1. We can consider a translated inverted gamma2 such as

$$\varphi(sk|s_0,\nu) \propto (sk-1)^{-(\nu+2)/2} \exp -\frac{s_0}{2(sk-1)}$$

The domain of definition of this density is $[1,\infty]$. We get a non-informative over $[1,\infty]$ for $\nu = -2$, $s_0 > 0$ and a uniform over $[-\infty,\infty]$ for $\nu = -2$ and $s_0 = 0$. The expectation $(s_0)/(\nu-2)-1$ exists provided that $\nu > 2$.

The case of γ is similar to that of the degrees of freedom ν in regression models with Student errors. For $\nu \to \infty$, the Student density tends to the normal density. Geweke (1993) uses an exponential prior in order to get an integrable posterior for ν. However a prior with the same asymptotic magnitude as $\nu^{\epsilon+1}$, with $\epsilon > 0$, is sufficient since it allows the posterior density to decay to zero quickly enough at its right tail in order to be integrable. This led Bauwens and Lubrano (1998) to use a less radical prior such as the truncated Cauchy density which we can adopt here:

$$\pi(\gamma) = \begin{cases} (1+\gamma^2)^{-1} & \text{if } \gamma > 0 \\ 0 & \text{otherwise} \end{cases}$$

We have just to replace $\pi(\theta)$ by its new expression in (48):

$$\pi(\theta) \propto \pi(\gamma)\pi(sk)$$

4.4 Comparison with the Euler discretisation

The likelihood function associated with the discretised model (21) is

$$\begin{aligned} l(y;\theta,\beta,\sigma_0) \;\propto\; & \sigma_0^{-T} \prod(\sigma(y_{t-\Delta},\theta)\sqrt{\Delta})^{-1} \\ & \exp\left(-\frac{1}{2\sigma_0^2\Delta}[(y_t-y_{t-\Delta})/\sigma(y_{t-\Delta},\theta)-\mu(y_{t-\Delta})\Delta/\sigma(y_{t-\Delta},\theta)]^2\right). \end{aligned} \quad (49)$$

If the drift function $\mu(y_{t-\Delta},\beta)$ is linear in β [3], the conditional posterior density of β and σ_0^2, conditionally on θ has the same general form as above because (21) is nothing but an heteroskedastic regression model. Let us define

$$y(\theta) = [(y_t - y_{t-\Delta})/\sigma(y_{t-\Delta},\theta)] \qquad Z(\theta) = [\mu(y_{t-\Delta})/\sigma(y_{t-\Delta},\theta)]. \quad (50)$$

[3]Linearity in β is not a very restrictive condition as for instance the non-linear drift function adopted in Ait-Sahalia (1996) is simply non-linear in the variable, but linear in the parameter.

The discretised model can be rewritten as in matrix notations

$$y(\theta) = Z(\theta)\beta\Delta + \sigma_0\sqrt{\Delta}\epsilon. \tag{51}$$

Let us consider the diffuse prior

$$\pi(\theta, \beta, \sigma_0^2) \propto 1/\sigma_0^2 \tag{52}$$

and let us define $\hat{\beta}(\theta)$ and $s^2(\theta)$ as in (46). The conditional posterior density of β and σ_0^2 is

$$\begin{aligned} \pi(\beta|y, \theta) &= f_t(\beta|\hat{\beta}(\theta), Z(\theta)'Z(\theta), s^2(\theta), T) \\ \pi(\sigma_0^2|y, \theta) &= f_{i\gamma}(\sigma_0^2|s^2(\theta), T) \end{aligned} \tag{53}$$

while

$$\pi(\theta|y) \propto |Z(\theta)'Z(\theta)|^{-1/2} s^2(\theta)^{-(T-2)/2} \prod \sigma(y_{t-\Delta}, \theta)^{-1}. \tag{54}$$

While the posterior densities of the continuous time model and of the discrete time model belong to the same families, they are indexed by different hyper-parameters. They may get closer when $\Delta \to 0$, but a gap still remains because the two models cannot have the same drift function: the drift function of the discretised model is linear in β by assumption while the implicit drift function of the continuous time model is totaly non-linear, except in trivial cases. Moreover, the transformations $y(\theta)$ and $Z(\theta)$ in (45) and (50) are not defined in the same way.

4.5 Model choice

Once different models are estimated, we can compute Bayes factors and select the model which has the highest posterior probability. However Bayes factors are first not easy to compute most of the time and second they are very sensitive to the specification of the prior which has to be informative. Lubrano (2001) develops a Bayesian procedure which compares two densities: a non-parametric estimate of the data density $\hat{f}(y)$ and the marginal density of the observations implied by the model $f(y|\theta)$. The distance between the two densities is measured with a ϕ divergence such as the Hellinger distance. The square of the Hellinger distance between $\hat{f}(y)$ and $f(y|\theta)$ is defined as

$$D_H^2(\theta) = 2\left(1 - \int \sqrt{\hat{f}(y)\, f(y|\theta)}\, dy\right). \tag{55}$$

The nonparametric estimate $\hat{f}(y)$ is obtained over a predetermined grid. Conditionally on θ, $f(y|\theta)$ can be evaluated over the same grid so that the integral in (55) can be evaluated using a Simpson rule

$$\hat{D}_H^2(\theta) = 2\left(1 - \frac{d}{3}\sum_i \sqrt{\hat{f}(y_i)\, f(y_i|\theta)}\, w_i\right) \tag{56}$$

where d is the size of the increment and w_i a weight equal to 1 for the extremes, 2 for even points and 4 for odd points. Up to now, the result is conditional on θ. We can obtain draws of the distribution of $\hat{D}_H^2(\theta)$ if we compute (55) for each draw of θ coming from the posterior generator. The model which will have the distribution closest to zero will be the preferred model.

Let us define k as the index of the observations so that $t = k\Delta$. When the process is stationary, it is possible to find the marginal density of the observations[4] from (19) letting $k \to \infty$. We have

$$f(y|\theta) = \frac{1}{\sigma(y)\sqrt{2\pi\phi^2}} \exp\left(-\frac{1}{2\phi^2}[U(y) + a_2/a_1]^2\right)$$

with $\phi^2 = (-b_2^2/2a_1)$. The marginal density of the transformed process $U(y)$ is thus normal with mean $-a_2/a_1$ and variance ϕ^2.

We shall compare our different models which have an exact discretisation, but a constrained drift, to an Euler discretisation equipped with a linear drift

$$y_t - y_{t-\Delta} = (a_1 y_{t-\Delta} + a_2)\Delta + \sigma_0\sqrt{\Delta}\sigma(y_{t-\Delta}, \theta)\epsilon_t$$

Using a stationarity assumption and the law of iterated expectations, the marginal expectation of y is equal to $-a_2/a_1$. Let us define the transformation $g(y_t) = (y_t + a_2/a_1)/\sigma(y_{t-\Delta}, \theta)$. Then $g(y_t) \sim \mathrm{N}(0, \sigma_0^2\Delta/(1-(1+a_1\Delta)^2)$ provided $|1 + a_1\Delta| < 1$. Consequently, the marginal distribution of y is equal to

$$f(y|\theta) = \frac{g'(y)}{\sqrt{2\pi v^2}} \exp\left(-\frac{1}{2v^2} g(y)^2\right)$$

where $v^2 = \sigma_0^2\Delta/(1-(1+a_1\Delta)^2)$. We have thus all the necessary ingredients to generate draws from the posterior distribution of $D_H^2(\theta)$.

Let us suppose that we want to compare model A and model B. Model A will be preferred to model B if $D_{HA}^2(\theta) < D_{HB}^2(\theta)$. Consequently

$$\Pr(A \succ B) \simeq \frac{1}{N}\sum_i \mathbb{1}(D_{HA}^2(\theta_i) < D_{HB}^2(\xi_i))$$

where θ_i and ξ_i represent the i^{th} out of N draws of the posterior density of models A and B.

4.6 Classical goodness of fit tests

The above procedure can be seen as the counterpart of a series of goodness-of-fit tests that have been proposed in the literature. Bickel and Rosenblatt (1973) were the first to propose a goodness of fit test based on a measure of distance between a parametric density and a non-parametric estimate of that density. Their test was improved by Fan (1994). Both tests are based on an integrated squared error. Ait-Sahalia (1996) uses a mean integrated squared error. Beran (1977) and his followers prefer to use a ϕ divergence and more precisely the Hellinger distance. Under

[4] When the process is not stationary, it is possible to make calculations on the underlying OU which is stationary iff $a_1 < 0$.

the null hypothesis of adequation, these tests are in general asymptotically N(0,1). Let us detail the Fan's test

$$Tf = \frac{nh\,ISE - ||K||^2}{\sqrt{2h||K * K||^2 \times ||\hat{f}||^2}} \tag{57}$$

where K is the kernel used for estimating the empirical density, h the window size and n the sample size. $||.||$ is the L^2 norm. ISE is defined as

$$ISE = \int [\hat{f}(y) - f(y|\hat{\theta})]^2 dy$$

and

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - y_i}{h}\right).$$

5 Comparison of empirical models for the US short term interest rate

We have chosen to use a similar data set as Ait-Sahalia (1999) as a benchmark to compare various models of the interest rate. We have thus first chosen the monthly observations of the Federal fund rate between january 1963 and december 1998. This makes 432 monthly observations. The source of the data is the Web site of the Federal Reserve Bank of St Louis. Figure 1 shows that the series experienced large variations, mainly during the Volker period and that its density is far from normal. We shall see how our different models accommodate this departure from normality. We have chosen to complete this data set by weekly observations of the same rate and the same

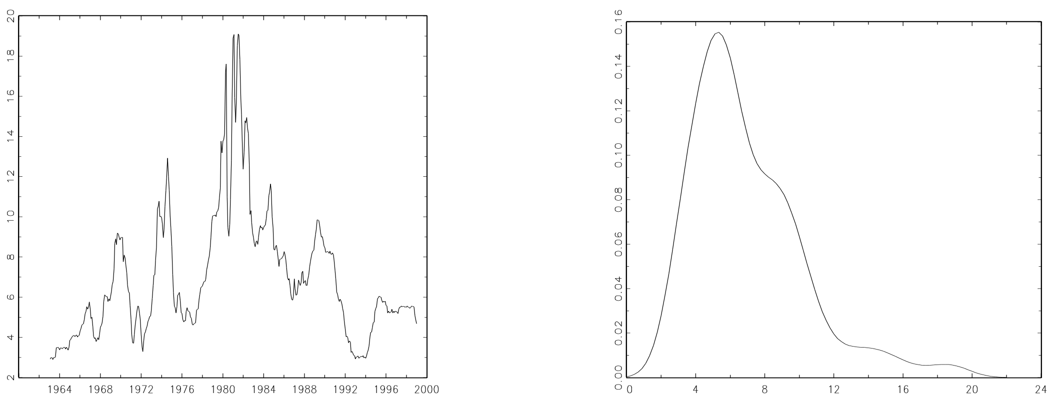

Figure 1: Federal Fund Rate, monthly frequency, 1963-1998

period. This makes now 1879 observations. Microstructure effects may not still appear at that frequency while weekly observations provide a reasonable yardstick to measure the effects of a discretisation biais when compared to results obtained with monthly data.

Table 1: Maximum likelihood estimates
Monthly US Federal Funds

	OU	OU-CEV	OU-ST	EU-CEV	EU-ST
a_1	-0.261 (0.12)	-0.177 (0.081)	-0.145 (0.076)	-0.0971 (0.094)	-0.0777 (0.084)
a_2	1.612 (0.79)	0.221 (0.10)	0.590 (0.33)	0.785 (0.44)	0.673 (0.46)
b_2	5.004 (0.34)	0.0072 (0.0019)	0.294 (0.20)	0.0075 (0.0020)	0.0444 (0.15)
γ		1.483 (0.070)	0.386 (0.10)	1.461 (0.069)	0.249 (0.099)
c			13.971 (2.03)		18.647 (8.41)
sk			20.110 (10.28)		96.40 (266.98)
D_H^2	0.0804	0.0713	0.0753	0.383	0.549
ISE	0.114	0.104	0.128	0.150	0.175
Fan's test	22.28	18.48	28.55	40.08	54.88

5.1 Classical Likelihood inference

We have selected three models: the simple OU, the OU-CEV and the OU-smooth transition volatility model. As a point of comparison, we propose an Euler discretisation with a two parameter linear drift for the CEV and the ST models. We first present inference results for the monthly data in Table 1. The first remarquable result is that we obtain for the OU-CEV the same estimated value for γ and its standard deviation as that obtained by Ait-Sahalia (1999) in Table VI with his expanded likelihood function for $K = 1$. Namely $\hat{\gamma} = 1.48\,(0.08)$ when we have $\hat{\gamma} = 1.483\,(0.070)$. The other parameters are more difficult to compare. The second result is that the values obtained for D_H^2 indicate that the Euler approximation does not give satisfactory results. In particular in the case of the smooth transition model, the parameter estimates obtained with the Euler approximation are not feasible. The threshold parameter c becomes too high and not enough observations are left in the high volatility regime.

Using weekly observations brings in more information. The fit of each model becomes much better, except for the simple linear OU model. For the smooth transition model, the volatility parameters are much nearer between the reducible model and its Euler approximation. The discretisation bias seems to be lower. We can conclude two things from these remarks: first we need a large data set to estimate a diffusion even with an exact discretisation; second, a non-linear model needs weekly observations to show up.

For a fixed critical level of 5%, Fan's test rejects all models. However, this test is misleading (as well as all classical tests) as the critical level should be a function of the sample size in order

Table 2: Maximum likelihood estimates
Weekly US Federal Funds

	OU	OU-CEV	OU-ST	EU-CEV	EU-ST
a_1	-0.325 (0.13)	-0.319 (0.13)	-0.294 (0.124)	-0.308 (0.142)	-0.258 (0.14)
a_2	1.985 (0.87)	0.635 (0.25)	1.289 (0.52)	2.167 (0.78)	1.837 (0.81)
b_2	6.390 (0.21)	0.126 (0.016)	1.187 (0.31)	0.140 (0.018)	1.040 (0.35)
γ		0.946 (0.034)	0.387 (0.075)	0.920 (0.033)	0.329 (0.063)
c			11.752 (0.93)		12.615 (1.17)
sk			6.694 (1.60)		7.856 (2.35)
D_H^2	0.0942	0.0232	0.0323	0.116	0.158
ISE	0.123	0.0724	0.0907	0.107	0.112
Fan's test	101.61	34.31	54.88	77.38	85.36

to avoid Lindley's paradox. We see from Tables 1 and 2 that the drop in the value of the ISE when going from monthly to weekly data does not manage to compensate for the increase in the sample size in (57).

We compare in Figure 2 the non-parametric estimate of the marginal density of the weekly observations combined with the parametric estimate of the stationary marginal density for the OU-CEV and OU-ST models. The two models manage roughly equivalently to mimic the non-parametric estimate, in particular in the tails. Of course they cannot reproduce the secondary humps; a mixture model would be necessary for that.

Figure 3 is particularly interesting. It compares the model implicit non-linear drift functions to a non-parametric estimate. The fact that the shape of the drift function is totaly determined by the shape of the volatility function does not appear to be a major drawback. In both models, the implicit drift functions are fairly realistic. They both display an increasing mean reversion despite the fact that they are highly parsimonious.

Figure 4 compares the two volatility functions to a non-parametric estimate. They both fit the data extremely well for low values of the interest rate, say till 10. For higher values, the CEV model underestimates the volatility while the smooth transition model gives a better account for it. The dropping shape of the non-parametric estimate of the volatility at the end of the range is certainly due to the lack of observations in that part of the distribution.

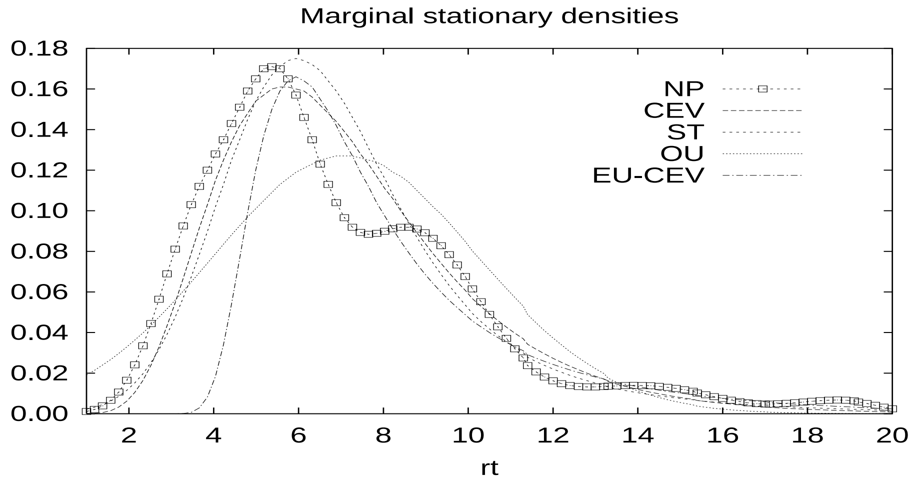

Figure 2: Federal Fund Rate, weekly frequency, 1963-1998, classical results

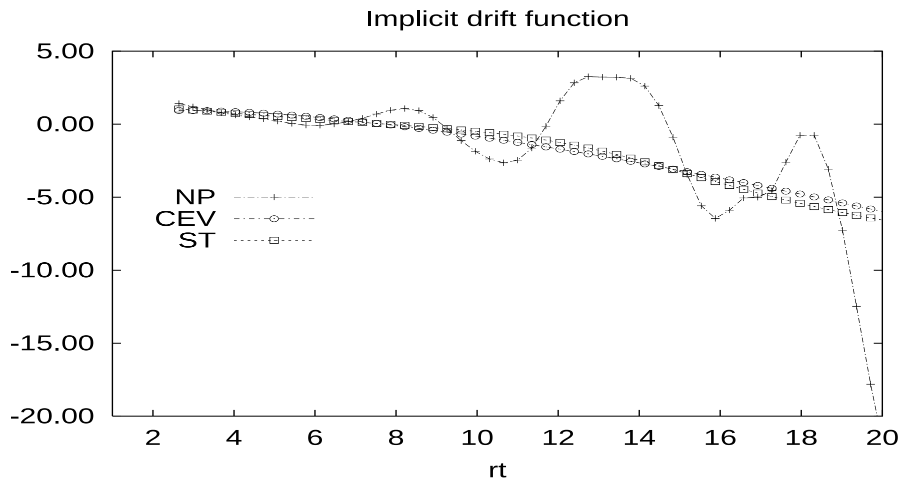

Figure 3: Federal Fund Rate, weekly frequency, 1963-1998

5.2 Bayesian model comparison

Classical inference gave a lot of interesting results concerning the efficiency of reducible empirical models. However, we were not able to formally test the fit of these models because of the large sample size. A Bayesian approach may give a better picture. We have conducted Bayesian inference for four models, leaving aside the Euler discretisation of the smooth transition model. We already know that this model experiences some problems. Bayesian inference for the reducible smooth transition models proved to be difficult. Even with an informative prior on γ and sk, we did not manage to get sensible results on these two data sets. As already noticed in Lubrano (2000), Bayesian inference in smooth transition models is made difficult by the fact that

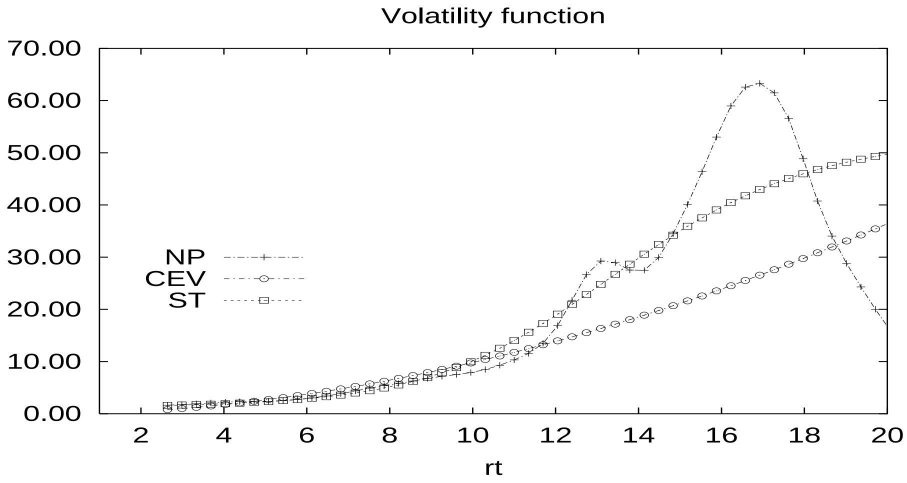

Figure 4: Federal Fund Rate, weekly frequency, 1963-1998

integrating on the threshold parameter already provides a kind of smoothing. In a model with an abrupt change of regime determined by the value of the threshold parameter c, the change of regime is sudden only if c is a point. In the Bayesian approach c takes a whole range of value depending on the spread of its posterior distribution. We thus decided to estimate a model with an abrupt change of regime. But even with this simplified model, we had to be informative on sk, saying that on average volatility of the high regime was 4 times that of the low regime with monthly observations and 5 times that of the low regime with weekly observations.

Posterior moments are reported in Tables 3 for monthly observations and Table 4 for weekly observations. They are obtained for 10000 draw of the simulator. For the OU model, the simulation is direct. For the OU-CEV and EU-CEV, we draw in the univariate posterior distribution of γ using a numerical inverse transform method and conditionally on that draw, we draw are back to the previous case. For the OU-ST, we have to draw in the bivariate posterior distribution of c and sk using the Griddy Gibbs sampler of Ritter and Tanner (1992). The correlation was very small. For each draw, we have computed the square of the Hellinger distance between the marginal stationary density and a non-parametric estimate of the sample density. These values were stored so that we could finally compare our four models for each frequency of observation. Table 5 reports the posterior probability that one model dominates another model. The OU-CEV model always dominates its Euler discretisation equiped with a linear drift, whatever the data frequency. This confirms the classical results: whatever the data frequency, it is better to work with an exact discretisation. With monthly observation, the simple OU is even better than the discretised CEV, showing that it is preferable to consider an exact model, than a discretised more realistic model. However, for weekly observations, the two models become equivalent.

With weekly observations, the OU-CEV model dominates all the other models. The two regime models is not at ease with this data set. It is found to be roughly equivalent to simpler models in many cases. It manages to dominate the Euler discretisation of the CEV model for monthly observations, but is strongly dominated by the CEV model for weekly observations.

Table 3: Bayesian Inference
Monthly US Federal Funds

	OU	OU-CEV	OU-ST	Eu-CEV
a_1	-0.265 (0.11)	-0.181 (0.079)	-0.221 (0.10)	-0.124 (0.075)
a_2	1.900 (0.88)	0.225 (0.10)	1.367 (0.67)	0.901 (0.37)
b_2	5.024 (0.35)	0.0074 (0.0020)	3.790 (0.30)	0.0075 (0.0021)
γ		1.485 (0.071)	-	1.471 (0.070)
c			16.27 (0.29)	
sk			3.706 (0.66)	
$\mathrm{E}(D_H^2)$	0.183	0.201	0.201	0.487
$D_H^2(\mathrm{E}(\theta)$	0.0735	0.0739	0.0712	0.332

Table 4: Bayesian Inference
Weekly US Federal Funds

	OU	OU-CEV	OU-ST	Eu-CEV
a_1	-0.329 (0.13)	-0.320 (0.12)	-0.311 (0.13)	-0.315 (0.13)
a_2	2.346 (0.99)	0.638 (0.25)	1.899 (0.84)	2.209 (0.74)
b_2	6.397 (0.21)	0.127 (0.017)	5.901 (0.20)	0.140 (0.018)
γ		0.945 (0.034)	-	0.922 (0.034)
c			17.32 (0.10)	
sk			1.896(0.19)	
$\mathrm{E}(D_H^2)$	0.170	0.104	0.175	0.201
$D_H^2(\mathrm{E}(\theta)$	0.0855	0.0223	0.0841	0.116

Table 5: Posterior probabilities of dominance

<	OU	OU-CEV	OU-ST	EU-CEV
OU	-	0.527	0.526	0.915
	-	0.216	0.497	0.537
OU-CEV	0.463	-	0.489	0.897
	0.784	-	0.788	0.801
OU-ST	0.474	0.511	-	0.894
	0.503	0.212	-	0.531
EU-CEV	0.085	0.101	0.106	-
	0.463	0.199	0.469	-

Posterior probability that the model indicated in the left column dominates the model indicated in the top line in the sense of a smaller Hellinger distance. The first line corresponds to monthly data and the second line to weekly data.

For both data frequency, it is found to be equivalent to the simple OU model. We observe that the posterior expectation of c is very high for both frequencies of observations : 16.27 and 17.32. These values are much higher than the MLE estimates. Consequently, very few observations are found in the high volatility regime.

6 Conclusion

In this paper, we have exploited the idea of reducible models in order to obtain a tractable expression for an exact likelihood function in continuous time. From our empirical application, a model naturally emerged: the constant elasticity of volatility model that can be reduced to an Orstein-Uhlenbeck. It managed to reproduce fairly well the density of the observation. Its induced drift and volatility functions compared well to their non-parametric estimates.

Computation time was very reasonable for the OU-CEV model: less than 10 seconds using a Gauss program on a recent PC. This result should be compared to the review of available estimation methods made by Durham and Gallant (2002). Our method ranges with the performance of that developed by Ait-Sahalia (1999). It is much more efficient than simulation methods.

References

AIT-SAHALIA, Y. (1996): "Testing continuous-time models of the spot interest rate," *Review of Financial Studies*, 9, 385–426.

——— (1999): "Transition densities for interest rate and other non-linear diffusions," *The Journal of Finance*, 54(4), 1391–1394.

——— (2002): "Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach," *Econometrica*, 70(1), 223–262.

BAUWENS, L., AND M. LUBRANO (1998): "Bayesian Inference on GARCH Models using the Gibbs Sampler," *Econometrics Journal*, 1, C23–C46.

BAUWENS, L., M. LUBRANO, AND J.-F. RICHARD (1999): *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.

BERAN, R. J. (1977): "Minimum Hellinger distance estimates for parametric models," *The Annals of Statistics*, 5, 445–463.

BICKEL, P., AND M. ROSENBLATT (1973): "On some global measures of the deviations of density function estimates," *The Annals of Statistics*, 1(6), 1071–1095.

BRENNAN, M., AND E. SCHWARTZ (1979): "A continuous time approach to the pricing of bonds," *Journal of banking and finance*, 3, 135–155.

BROZE, L., O. SCAILLET, AND J.-M. ZAKOÏAN (1995): "Testing for continuous-time models of the short term interest rate," *Journal of empirical Finance*, 2, 199–223.

CHAN, K., G. KAROLYI, F. LONGSTAFF, AND A. SANDERS (1992): "An empirical comparison of alternative models of the short-term interest rates," *Journal of Finance*, 47(3), 1209–1227.

CONLEY, T., L. P. HANSEN, E. LUTTMER, AND J. SCHEINKMAN (1997): "Short-term interest rates as subordinated diffusions," *The Review of Financial Studies*, 10(3), 525–577.

DURHAM, G., AND R. GALLANT (2002): "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes," *The Journal of Business and Economic Statistics*, 20(3), 297–316.

FAN, Y. (1994): "Testing a goodness-of-fit of a parametric density function by kernel method," *Econometric Theory*, 10(2), 316–356.

GALLANT, R., AND G. TAUCHEN (1998): "Reprojecting partially observed systems with an application to interest rate diffusions," *Journal of the American Statistical Asspciation*, 93(441), 10–24.

GEWEKE, J. (1993): "Bayesian treatment of the independent Student-t linear model," *Journal of Applied Econometrics*, 8, S19–S40.

KLOEDEN, P., AND E. PLATEN (1999): *Numerical solution of stochastic differential equations*, Applications of mathematics; Stochastic modelling and applied probability. Springer-Verlag, third edn.

LUBRANO, M. (2000): “Bayesian Analysis of Nonlinear Time Series with Models with a Threshold,” in *Non-linear Econometric Modelling*, ed. by W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim, and A. Würtz, pp. 79–118. Cambridge University Press, Cambridge, UK.

——— (2001): “Bayesian Model Choice for the Short Term US Interest Rate,” in *Bayesian methods with application to science, policy and official statistics*, ed. by E. I. George, pp. 341–350. Euro-stat, Luxemburg.

LUND, J. (1999): “Estimation of continuous-time models,” Ph.d. course, Copenhagen Business School, http://www.jesperlund.com/.

PFANN, G., P. SCHOTMAN, AND R. TSCHERNIG (1996): “Nonlinear Interest Rate Dynamics and Implications for the Term Structure,” *Journal of Econometrics*, 74(1), 149–176.

RAO, P. (1999): *Statistical Inference for Diffusion Type Processes*, Kendall’s Library of Statistics. Arnold, London.

RITTER, C., AND M. TANNER (1992): “Facilating the Gibbs sampler: the Gibbs stopper and the griddy Gibbs sampler,” *Journal of the American Statistical Association*, 87, 861–868.

A Proof of theorem 1

Proof 1: Let us consider $x_t/\phi(t)$, where $\phi(t)$ verifies the homogeneous SDE $d\phi(t) = a_1\phi(t)dt + b_1\phi(t)dW_t$ which is geometric Brownian motion. Consequently, an exact discretisation of $\phi(t)$ is (5). Applying Ito’s lemma, we can obtain the SDE of $1/\phi(t)$:

$$d\phi^{-1}(t) = \frac{(b_1^2 - a_1)}{\phi(t)}dt - \frac{b_1}{\phi(t)}dW_t$$

We have now to compute $d(x_t/\phi(t))$. As both x_t and $\phi(t)$ verify a SDE which involve the same Brownian increment, we have to use a modified version of the Ito’s lemma:

Lemma 1 *Let us consider $X_1(t)$ and $X_2(t)$ verifying*

$$dX_i(t) = \mu_i(\cdot)dt + \sigma_i(\cdot)dW(t), \quad i = 1, 2$$

and the fonction $U(X_1, X_2)$. Then

$$dU = \left[\sum_{k=1}^{2} \mu_k(\cdot)\frac{\partial U}{\partial x_k} + \sigma_1(\cdot)\sigma_2(\cdot)\right] dt + \left[\sum_{k=1}^{2} \sigma_k(\cdot)\frac{\partial U}{\partial x_k}\right] dW(t)$$

Defining $U = x_t/\phi(t)$, the application of this lemma gives after some manipulations:

$$d\left(\frac{x_t}{\phi(t)}\right) = \frac{(a_2 - b_1 b_2)}{\phi(t)}dt + \frac{b_2}{\phi(t)}dW_t$$

□