SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2010-33

Maximum likelihood estimation for integrated diffusion processes

Fernando Baltazar-Larios and Michael Sørensen

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Maximum likelihood estimation for integrated diffusion processes

Fernando Baltazar-Larios*
Universidad Nacional Autónoma de México
IIMAS, A.P. 20-726
01000 Mexico, D.F.
Mexico
sheva7.fernando@gmail.com

Michael Sørensen†
University of Copenhagen and CREATES
Dept. of Mathematical Sciences
Universitetsparken 5
DK-2100 Copenhagen Ø
Denmark
michael@math.ku.dk

August 5, 2010

Abstract

We propose a method for obtaining maximum likelihood estimates of parameters in diffusion models when the data is a discrete time sample of the integral of the process, while no direct observations of the process itself are available. The data are, moreover, assumed to be contaminated by measurement errors. Integrated volatility is an example of this type of observations. Another example is ice-core data on oxygen isotopes used to investigate paleo-temperatures.

The data can be viewed as incomplete observations of a model with a tractable likelihood function. Therefore we propose a simulated EM-algorithm to obtain maximum likelihood estimates of the parameters in the diffusion model. As part of the algorithm, we use a recent simple method for approximate simulation of diffusion bridges. In simulation studies for the Ornstein-Uhlenbeck process and the CIR process the proposed method works well.

Key words: Diffusion bridge, discretely sampled diffusions, EM-algorithm, likelihood inference, measurement error, stochastic differential equation, stochastic volatility.

JEL codes: C22, C51.

*Fernando Baltazar-Larios acknowledges support from Consejo Nacional de Ciencia y Tecnologia, which financed a one year stay at the University of Copenhagen.

†Michael Sørensen acknowledges support from Center for Research in Econometric Analysis of Time Series, CREATES, www.creates.au.dk, funded by the Danish National Research Foundation and from the Danish Center for Accounting and Finance funded by the Danish Social Science Research Council.

1 Introduction.

We consider estimation for a general one-dimensional diffusion process $X = \{X_t\}_{t\geq 0}$. Likelihood based estimation (including Bayesian) for discretely observed diffusion processes has been investigated by Ozaki (1985), Pedersen (1995), Poulsen (1999), Elerian, Chib & Shephard (2001), Eraker (2001), Roberts & Stramer (2001), Aït-Sahalia (2002), Durham & Gallant (2002), Aït-Sahalia & Mykland (2003), Beskos et al. (2006) Aït-Sahalia (2008). Martingale estimating functions for discretely observed diffusions are reviewed in Sørensen (1997) and Sørensen (2010).

In this paper we consider maximum likelihood estimation in the situation where we do not observe the process X itself directly, but instead observe integrals of the process over disjoint time-intervals. These observations are, moreover, assumed to be contaminated by measurement errors. Integrated diffusion processes play an important role in finance as models for realized volatility, see e.g. Andersen et al. (2001b), Andersen et al. (2001a), Bollerslev & Zhou (2002), and Barndorff-Nielsen & Shephard (2002). These processes are also used for modelling purposes in fields of engineering and the sciences. An example is provided by the records of the concentration of oxygen isotopes in ice-core data from Greenland and Antarctica, see e.g. Ditlevsen, Ditlevsen & Andersen (2002). Such data are used to investigate the paleo-climate.

The likelihood function for a discretely sampled integrated diffusion with observation error is in almost all cases not explicitly available. Moreover, the integrated process is not a Markov process, so there is no easily calculated martingales. Therefore martingale estimating functions are not a feasible alternative, but prediction-based estimating function can be applied, see Sørensen (2000). In the present paper, we note instead that the data can be viewed as incomplete observations from a model with a tractable likelihood function. The full data set is a continuous time record of the diffusion process and the observation errors. We can therefore find maximum likelihood estimates by applying the Expectation-Maximization (EM) algorithm, see Dempster, Laird & Rubin (1977) and McLachlan & Krishnan (1997). To do so we need to calculate the conditional expectation of the log-likelihood function for the full model given the observations. We do this by simulating sample paths of the diffusion process given the data using ideas from Chib, Pitt & Shephard (2006). An essential step in doing this is to simulate a part of a sample path given the rest, which corresponds to simulation a diffusion bridge. This is done by applying the method for approximate diffusion bridge simulation recently proposed by Bladt & Sørensen (2009).

Parametric inference for integrated diffusion process has been considered by Gloter (2000), Bollerslev & Zhou (2002), Ditlevsen & Sørensen (2004), Gloter (2006), and Forman & Sørensen (2008). Nonparametric inference has been considered in Comte, Genon-Catalot & Rozenholc (2009).

In Section 2 the model of an integrated diffusion process with measurement error and its assumptions are presented. Section 3 contains the calculation of the likelihood function with full diffusion observation and the EM-algorithm. In Section 4 we present a method for simulation of a diffusion process conditional on integrals observed with measurement error. In Section 5 we consider the Ornstein-Uhlenbeck process in detail and a simulation study for this model is reported. A similar investigation for the CIR/square root process is presented in Section 6, where also stochastic volatility models are briefly discussed. Some concluding remarks are given in Section 7.

2 Model and data

We consider likelihood estimation the general one-dimensional diffusion process $X = \{X_t\}_{t\geq 0}$ given by the stochastic differential equation

$$dX_t = b(X_t;\psi)dt + \sigma(X_t;\psi)dW_t \tag{2.1}$$

where $W = \{W_t\}$ is a standard Wiener process, and where the drift and diffusion coefficients depend on an unknown p-dimensional parameter ψ belonging to the parameter set $\Psi \subseteq R^p$. We assume that the solution X is an ergodic, stationary diffusion with invariant measure with density function $\nu_\psi(x)$ ($X_0 \sim \nu_\psi$ is independent of W). We also assume that the stochastic differential equation has a unique weak solution, i.e. a solution exists and all solutions have identical finite-dimensional distributions; see e.g. Karatzas & Shreve (1991). It is well-known that sufficient conditions for these assumptions can be expressed in terms of the so-called scale function and speed measure; see e.g. Karlin & Taylor (1981).

In this paper we consider the situation where the process X has not been observed directly. Instead the data are integrals of X_t over intervals $[t_{i-1}, t_i]$ observed with measurement error, i.e.

$$Y_i = \int_{t_{i-1}}^{t_i} X_s ds + Z_i, \quad i = 1, \ldots, n, \tag{2.2}$$

where $Z_i \sim N(0, \tau^2)$, $i = 1, \ldots, n$ are mutually independent and independent of X. We assume that $t_0 = 0$, so the total interval of observation is $[0, t_n]$. Note that the variance of the measurement error, τ^2, is an extra unknown parameter. Thus we need to estimate the $p+1$-dimensional parameter $\theta = (\psi, \tau^2)$.

Conditionally on the sample path of X, the observations Y_i, $i = 1, \ldots, n$ are independent and normal distributed:

$$Y_i \,|\, X_t : t \in [0, t_n] \sim N\left(\int_{t_{i-1}}^{t_i} X_s ds, \tau^2\right), \tag{2.3}$$

We assume that the coefficients of the stochastic differential equation (2.1) satisfy the following conditions which we need in the following sections.

Condition 2.1 *The drift and diffusion coefficients of (2.1), $b(x;\psi)$ and $\sigma(x;\psi)$ satisfy that for all $\psi \in \Psi$*

- *$b(x;\psi)$ is continuously differentiable w.r.t. x*
- *$\sigma(x;\psi)$ is twice continuously differentiable w.r.t. x*
- *$\sigma(x;\psi) > 0$ for all x in the state space of X*

3 The likelihood function and the EM-Algorithm

We can think of the data set $Y = (Y_1, \ldots, Y_n)$ as an incomplete observation of a full data set given by the sample path X_t, $t \in [0, t_n]$ and the measurement errors $Z_1, \cdots, Z_n$, or equivalently X_t, $t \in [0, t_n]$ and $Y = (Y_1, \ldots, Y_n)$. Therefore likelihood based estimation can be done by means of the EM-algorithm or MCMC-methods. In this paper we concentrate on the EM-algorithm. We need to find the likelihood function for the full data set and the conditional expectation of this full log-likelihood function given the observations $Y = (Y_1, \ldots, Y_n)$.

3.1 Likelihood with full diffusion observation

The full observation of a diffusion sample path in the time interval $[0, t_n]$ is an element in the space C of continuous functions from $[0, t_n]$ to $\mathbb{R}$. We equip this space with the usual σ-algebra, $\mathcal{C}$, generated by the cylinder sets, and consider the probability measures induced on $(C, \mathcal{C})$ by the solutions to (2.1). These measures are in general singular because the diffusion coefficient depends on the parameter ψ. In order to obtain a likelihood function, we use the standard 1-1 transformation

$$h(x; \psi) = \int_{x^*}^{x} \frac{1}{\sigma(u; \psi)} du, \tag{3.1}$$

where x^* is some arbitrary element of the state space of X. By this parameter dependent transformation, we obtain a diffusion process with unit diffusion coefficient. Specifically, we obtain (by Ito's formula) that

$$U_t = h(X_t; \psi)$$

satisfies the stochastic differential equation

$$dU_t = \mu(U_t; \psi)dt + dW_t, \tag{3.2}$$

with

$$\mu(u; \psi) = \frac{b\left(h^{-1}(u; \psi); \psi\right)}{\sigma\left(h^{-1}(u; \psi); \psi\right)} - \frac{\sigma'\left(h^{-1}(u; \psi); \psi\right)}{2},$$

where σ' denotes the derivative of σ w.r.t. x. In (3.2) the diffusion coefficient does not depend on the parameters, so the probability measures induced on $(C, \mathcal{C})$ by the solution to (3.2) are equivalent and the likelihood function can be found.

We can express the observations Y_i in terms of the process U. By inserting $X_s = h^{-1}(U_s; \psi)$ in (2.2), we find that

$$Y_i = \int_{t_{i-1}}^{t_i} h^{-1}(U_s; \psi) ds + Z_i, \quad i = 1, \ldots, n.$$

Therefore we will think of the full dataset as $U_t,\ t \in [0, t_n]$ and $Y = (Y_1, \ldots, Y_n)$. Since conditionally on the sample path of U the observations Y_i, $i, \ldots, n$ are independent, we have that the likelihood of Y conditional on the sample path of U in $[0, t_n]$ is

$$L(Y_1, \ldots, Y_n \,|\, U_t, t \in [0, t_n]) = \prod_{i=1}^{n} \phi(Y_i; \int_{t_{i-1}}^{t_i} h^{-1}(U_s; \psi) ds, \tau^2) \tag{3.3}$$

where $\phi(u; a_1, a_2)$ denotes the density of the normal distribution with mean a_1 and variance a_2 evaluated at u.

Let P_ψ be the probability measure induced by $U = \{U_t\}_{t \in [0, t_n]}$ on $(C, \mathcal{C})$, i.e. the probability measure with respect to which the coordinate process has the same distribution as U, and let Q be the Wiener measure on $(C, \mathcal{C})$. We assume that the coefficient μ satisfies conditions ensuring that the Girsanov theorem holds so that we have the Radon-Nykodym derivative

$$\frac{dP_\psi}{dQ}(B) = \exp\left\{\int_0^{t_n} \mu(B_t;\psi)dB_t - \frac{1}{2}\int_0^{t_n} \mu^2(B_t;\psi)dt\right\}; \tag{3.4}$$

see e.g. Liptser & Shiryaev (1977), Jacod & Shiryaev (1987) or Øksendal (1998).

The evaluation of $\frac{dP_\psi}{dQ}$ is difficult because of the Ito integral term. To simplify the likelihood function, we apply the transformation

$$a(x;\psi) = \int^x \mu(u;\psi)du$$

(any antiderivative of μ), which under Condition 2.1 is twice continuously differentiable. By Ito's formula

$$\int_0^{t_n} \mu(B_t)dB_t = a(B_{t_n};\psi) - a(B_0;\psi) - \frac{1}{2}\int_0^{t_n} \mu'(B_t;\psi)dt,$$

where μ' denotes the derivative of $\mu(u;\psi)$ w.r.t. u. We can now write the likelihood function (3.4) as

$$\frac{dP_\psi}{dQ}(B) = \exp\left\{a(B_{t_n};\psi) - a(B_0;\psi) - \frac{1}{2}\int_0^{t_n} [\mu(B_t;\psi)^2 + \mu'(B_t;\psi)]dt\right\}.$$

By combining this expression and (3.3), we see that the log-likelihood function for θ based on the full data set $U_t,\ t \in [0, t_n]$ and $Y = (Y_1, \ldots, Y_n)$ is given by

$$\begin{aligned} \log L(\theta; Y_1, \ldots, Y_n, U_t, t \in [0, t_n]) \quad &= \quad \sum_{i=1}^n \log \phi(Y_i; \int_{t_{i-1}}^{t_i} h^{-1}(U_s;\psi)ds, \tau^2) \qquad (3.5) \\ &+ a(U_{t_n};\psi) - a(U_0;\psi) - \frac{1}{2}\int_0^{t_n} \Big(\mu(U_t;\psi)^2 + \mu'(U_t;\psi)\Big)\, dt. \end{aligned}$$

3.2 EM Algorithm.

We can now apply the EM-algorithm to the full log-likelihood function (3.5) to obtain the maximum likelihood estimate of the parameter θ.

As the initial value for the algorithm, let $\hat\theta$ be any value of the parameter vector $\theta = (\psi, \tau^2) \in \Psi \times (0, \infty)$. Then the EM-algorithm works as follow.

1. **E-STEP.**

   Generate M sample paths of the diffusion process X, $X^{(k)}, k = 1, \ldots, M$, conditional on the observations $Y_1, \ldots, Y_n$ using the parameter value $\hat\theta = (\hat\psi, \hat\tau^2)$, and calculate

   $$g(\theta) = \frac{1}{M - M_0} \sum_{k=M_0+1}^{M} \log L(\theta; Y_1, \ldots, Y_n, h(X_t^{(k)}; \hat\psi), t \in [0, t_n]),$$

   for a suitable burn-in period M_0 and M sufficiently large.

2. **M-STEP.**

   $\hat\theta = \text{argmax}\ g(\theta).$

3. Go to 1.

To implement this algorithm, the main issue is how to generate sample paths of X conditionally on $Y_1, \ldots, Y_n$, where the relation between the Y_is and X is given by (2.2). The algorithm must produce a sequence $X^{(k)}, k = 1, \ldots, M$, that is sufficiently mixing to ensure that $g(\theta)$ approximates the conditional expectation of the full log-likelihood function (3.5) given the data. This problem is discussed at the next section.

4 Conditional diffusion process simulation.

In this section we present a method for generating a sample from

$$\{X_t; t \in [0, t_n]\} | (Y_1, \ldots, Y_n)$$

for a given value of the parameter vector θ, i.e. for simulating the diffusion X conditional on the observations $Y = (Y_1, \ldots, Y_n)$ of integrals of X over subintervals $[t_{j-1}, t_j]$, $j = 1, \ldots, n$ perturbed by measurement errors. This can be done by means of a Metropolis-Hastings algorithm, see e.g. Chib & Greenberg (1995) or Gilks, Richardson & Spiegelhalter (1996). However, if the sample path in the entire time interval $[0, t_n]$ is updated in one step, the rejection probability is typically very large. Therefore it is more efficient to randomly divide the time interval into subintervals and update the sample path in each of the subintervals conditional on the rest of the sample path. This corresponds to simulating a (conditional) diffusion bridge in each subinterval (except the end-intervals). The method outlined in this section is a modification of the method in Chib, Pitt & Shephard (2006), where we use the algorithm for approximate diffusion bridge simulation proposed by Bladt & Sørensen (2009).

In the following the parameter value $\theta = (\psi, \tau^2)$ is fixed.

Algorithm 1

1. Generate an initial unrestricted stationary sample path, $\{X_t^{(0)} : t \in [0, t_n]\}$, of the diffusion given by (2.1) using for instance the Milstein scheme or one of the other methods in Kloeden & Platen (1999).

2. Set $l = 1$.

3. Generate a sample path $\{X_t^{(l)} : t \in [0, t_n]\}$ conditional on Y by updating the subsets of the sample path:

   (a) Randomly split the time interval from 0 to t_n in K blocks, and write these subsampling times as
   $$0 = \tau_0 \leq \tau_1 \leq \ldots \leq \tau_K = t_n,$$
   where each τ_i is one of the end-points of the integration intervals, t_j, $j = 0, \ldots, n$. Let $Y_{\{k\}}$ denote the collection of all observations Y_j for which $\tau_{k-1} < t_j \leq \tau_k$.

   (b) Draw $X_0^{(l)}$ from the stationary distribution, ν_ψ, and simulate the conditional subpath
   $$\{X_t^{(l)} : t \in [\tau_{k-1}, \tau_k]\} \,|\, Y_{\{k\}}, X_{\tau_{k-1}}^{(l)}, X_{\tau_k}^{(l-1)} \tag{4.1}$$

for $k = 1, \ldots, K-1$. Finally, simulate a sample path from

$$\{X_t^{(l)} : t \in [\tau_{K-1}, \tau_K]\} \,|\, Y_{\{K\}}, X_{\tau_{K-1}}^{(l)}.$$

4. l=l+1.

5. Go to 3.

To implement of this algorithm, the main issue is how to sample variables of the type (4.1), which is a non-linear diffusion bridge. We use the method for approximate diffusion bridge simulation proposed by Bladt & Sørensen (2009). The idea (in the case of a diffusion bridge in the time interval $[0, 1]$) is to let one diffusion process move forward from time zero out of one given point, a, until it meets another diffusion process that independently moves backwards from time one out of another given point, b. Conditional on the event that the two diffusions intersect, the process constructed in this way is an approximation to a realization of a diffusion bridge between a and b. The diffusions can be simulated by means of simple procedures like the Euler scheme or the Milstein scheme, see Kloeden & Platen (1999). The method is therefore very easy to implement. The resulting sample path is an approximation to a diffusion bridge in the sense that it has the distribution of a diffusion bridge from a to b conditional on the event that the bridge is hit by an independent diffusion with stochastic differential equation (2.1) and initial distribution with density $p_1(b, \cdot)$. Simulation studies in Bladt & Sørensen (2009) indicate that the approximation is very good for bridges between points that are likely to appear on a sample path of the diffusion, which is the type of bridges that are relevant to this paper.

Alternative methods that provide exact diffusion bridges have been proposed by Beskos, Papaspiliopoulos & Roberts (2006) and Beskos, Papaspiliopoulos & Roberts (2007). When the drift and diffusion coefficients satisfy certain boundedness conditions, this algorithm is relatively simple, but under weaker condition it is more complex. A simulation study in Bladt & Sørensen (2009) indicates that for the method which we use here, the CPU-time is linear in the length of the interval where the diffusion bridge is defined, whereas for the method in Beskos, Papaspiliopoulos & Roberts (2006), the CPU time increases exponentially with the interval length. This is an advantage of the method in Bladt & Sørensen (2009) in the present context. MCMC algorithms for simulation of diffusion bridges were proposed by Roberts & Stramer (2001), Durham & Gallant (2002), and Chib, Pitt & Shephard (2006).

To generate the random subintervals in step 3 (a) of Algorithm 1, we use the following algorithm, where the number of integration subintervals $[t_{j-1}, t_j]$ included in one of the random subintervals is a Poisson distributed random number plus 1. The draws in the algorithm are independent. First choose the expectation of the Poisson distribution, $\lambda \geq 1$.

Algorithm 2

1. Draw $k_1 \sim Poisson(\lambda) + 1$: if $k_1 \geq n$ set $k_1 = n$, $K = 1$ and stop, otherwise set $i = 2$.

2. Draw $k_i \sim Poisson(\lambda) + 1$, if $\sum_{j=1}^{i} k_j \geq n$ set $k_i = n$, $K = i$ and stop, else set $i = i + 1$ and repeat 2.

Finally define $\tau_i = t_{k_i}$, $i = 1, \ldots, K$.

We have discussed how to simulate diffusion bridges, but we need diffusion bridges conditional on the data Y. Sample paths of the conditional bridges (4.1) can be obtained by the following Metropolis-Hastings algorithm. By a (t, a, s, b)–bridge, we mean a diffusion bridge in the time interval $[t, s]$ with $X_t = a$ and $X_s = b$. After a burn-in period the following algorithm will output samples from a $(\tau_{k-1}, a, \tau_k, b)$–bridge conditional on $Y_{\{k\}}$, the data in $(\tau_{k-1}, \tau_k]$. To formulate the algorithm we need to specify that the end-point τ_{k-1} is equal to t_j, and that there are n_k observations in the interval $(\tau_{k-1}, \tau_k]$, namely, $Y_{j+1}, \ldots, Y_{j+n_k}$.

Algorithm 3

1. Simulate a $(\tau_{k-1}, a, \tau_k, b)$–bridge, $X^{(0)}$, and set $l = 1$.
2. Propose a new sample paths by simulating a $(\tau_{k-1}, a, \tau_k, b)$–bridge, $X^{(l)}$.
3. Accept the proposed diffusion bridge with probability
$$\min\left(1, \prod_{i=1}^{n_k} \frac{\phi(Y_{j+i}; \int_{t_{j+i-1}}^{t_{j+i}} X_s^{(l)} ds, \tau^2)}{\phi(Y_{j+i}; \int_{t_{j+i-1}}^{t_{j+i}} X_s^{(l-1)} ds, \tau^2)}\right).$$
Otherwise set $X^{(l)} = X^{(l-1)}$.
4. Set $l = l + 1$ and go to 2.

As previously, $\phi(x; \mu, \tau^2)$ denotes the density function of the normal distribution with mean μ and variance τ^2.

5 The Ornstein-Uhlenbeck process: a simulation study.

In this section we apply the method developed above to the Ornstein-Uhlenbeck process, which is a solution of the stochastic differential equation

$$dX_t = -\alpha X_t dt + \sigma dW_t, \tag{5.1}$$

where $\alpha > 0$ and $\sigma > 0$ are unknown parameters to be estimated, and W is a standard Wiener process. We investigate the bias of the estimators in a simulation study.

5.1 The likelihood and the EM-algorithm

The transformation (3.1) is here given by

$$h(x; \sigma) = \frac{x}{\sigma},$$

so $h^{-1}(x; \sigma) = \sigma x$. Hence $U_t = h(X_t; \sigma) = X_t/\sigma$, solves the stochastic differential equation

$$dU_t = -\alpha U_t dt + dW_t.$$

We have $\mu(u;\alpha,\sigma) = -\alpha u$, so

$$a(u;\alpha,\sigma) = -\frac{1}{2}\alpha u^2.$$

Thus the full log-likelihood function (3.5) is given by

$$\begin{aligned} &\log L(\theta; Y_1,\ldots,Y_n, U_t, t\in[0,t_n]) \\ &= \sum_{i=1}^{n}\log\phi\left(Y_i;\sigma\int_{t_{i-1}}^{t_i}U_s ds,\tau^2\right) + \frac{\alpha}{2}(U_0^2 - U_{t_n}^2 + t_n) - \frac{\alpha^2}{2}\int_0^{t_n}U_t^2 dt, \end{aligned} \tag{5.2}$$

where $\theta = (\alpha,\sigma,\tau^2)$.

Now, the EM algorithm works as follow.

E-STEP

The objective function $g(\theta)$ is for the Ornstein-Uhlenbeck process given by

$$\begin{aligned} g(\theta) &= -\frac{1}{2\tau^2(M-M_0)}\sum_{k=M_0+1}^{M}\sum_{i=1}^{n}(Y_i - \sigma\int_{t_{i-1}}^{t_i}U_t^{(k)}dt)^2 - \frac{n}{2}\log(2\pi\tau^2) + \frac{\alpha}{2}t_n \\ &+ \frac{\alpha}{2(M-M_0)}\sum_{k=M_0+1}^{M}((U_0^{(k)})^2 - (U_{t_n}^{(k)})^2) - \frac{\alpha^2}{2(M-M_0)}\sum_{k=M_0+1}^{M}\int_0^{t_n}(U_t^{(k)})^2 dt. \end{aligned}$$

Here $U_t^{(k)} = X_t^{(k)}/\hat{\sigma}$, where $X_t^{(k)}$ is the k-th sample path of the process X simulated conditionally on the data Y using the Algorithms 1 – 3 with the parameter value obtained in the previous step $(\hat{\alpha},\hat{\sigma},\hat{\tau}^2)$.

M-STEP

The maximum $\hat{\theta}$ is obtained as the solution to the following system of equations

$$\frac{\partial g(\theta)}{\partial\alpha} = \frac{1}{2}t_n + \frac{\sum_{k=M_0+1}^{M}\left[(U_0^{(k)})^2 - (U_{t_n}^{(k)})^2\right]}{2(M-M_0)} - \frac{\alpha\sum_{k=M_0+1}^{M}\int_0^{t_n}(U_t^{(k)})^2 dt}{M-M_0} = 0, \tag{5.3}$$

$$\frac{\partial g(\theta)}{\partial\sigma} = \frac{\sum_{k=M_0+1}^{M}\sum_{i=1}^{n}(Y_i - \sigma\int_{t_{i-1}}^{t_i}U_t^{(k)}dt)(\int_{t_{i-1}}^{t_i}U_t^{(k)}dt)}{\tau^2(M-M_0)} = 0 \tag{5.4}$$

and

$$\frac{\partial g(\theta)}{\partial\tau^2} = \frac{\sum_{k=M_0+1}^{M}\sum_{i=1}^{n}(Y_i - \sigma\int_{t_{i-1}}^{t_i}U_t^{(k)}dt)^2}{2\tau^4(M-M_0)} - \frac{n}{2\tau^2} = 0. \tag{5.5}$$

From (5.3) we have

$$\hat{\alpha} = \frac{t_n(M-M_0) + \sum_{k=M_0+1}^{M}\left[(U_0^{(k)})^2 - (U_{t_n}^{(k)})^2\right]}{2\sum_{k=M_0+1}^{M}\int_0^{t_n}(U_t^{(k)})^2 dt},$$

and from (5.4)

$$\hat{\sigma} = \frac{\sum_{k=M_0+1}^{M}\sum_{i=1}^{n}Y_i\int_{t_{i-1}}^{t_i}U_t^{(k)}dt}{\sum_{k=M_0+1}^{M}\sum_{i=1}^{n}(\int_{t_{i-1}}^{t_i}U_t^{(k)}dt)^2}. \tag{5.6}$$

Now inserting $\hat{\sigma}$ given by (5.6) in (5.5) we obtain

$$\hat{\tau}^2 = \frac{(M-M_0)(\sum_{i=1}^n Y_i^2)\left[\sum_{k=M_0+1}^M \sum_{i=1}^n (\int_{t_{i-1}}^{t_i} U_t^{(k)} dt)^2\right] - \left[\sum_{k=M_0+1}^M \sum_{i=1}^n Y_i \int_{t_{i-1}}^{t_i} U_t^{(k)} dt\right]^2}{n(M-M_0)\sum_{k=M_0+1}^M \sum_{i=1}^n (\int_{t_{i-1}}^{t_i} U_t^{(k)} dt)^2}.$$

The Hessian matrix of $g(\theta)$ evaluated at $\hat{\theta}$ is negative define, so $\hat{\theta}$ is maximum.

5.2 A simulation study

In this section we present the result of a small simulation study, in which we simulated 1000 datasets and for each of them obtained estimates by means of the EM-algorithm proposed in the present paper. Each data set was obtained by simulating a sample path of length 1500 with initial distribution $X_0 \sim N(0, \sigma^2/(2\alpha))$, and then calculating data Y_i, $i = 1, \ldots, 1500$ by (2.2) with $t_i = i$, $i = 0 \ldots, n$. The parameter values were $\alpha = 0.1$, $\sigma = 0.5$ and $\tau^2 = 1.25$.

The EM-algorithm was run with $M = 10000$ and $M_0 = 1000$ and for three different values of λ, namely $\lambda = 10, 20, 30$. The average of the estimates obtained for the 1000 dataset are given in Table 5.1. The bias is small, and is overall most satisfactory for $\lambda = 20$.

λ	α	σ	τ^2
10	0.106	0.523	1.229
20	0.101	0.507	1.235
30	0.084	0.458	1.252

Table 5.1: Average of parameter estimates obtained from 1000 simulated datasets with parameter values $\alpha = 0.1$, $\sigma = 0.5$ and $\tau^2 = 1.25$.

6 The CIR process and a stochastic volatility model

In this section we apply our method to the CIR process, which solve

$$dX_t = (\alpha - \beta X_t)dt + \sigma\sqrt{X_t}dW_t, \tag{6.1}$$

where $X_0 > 0$, $\alpha > 0$, $\beta > 0$, $\sigma > 0$ and W is a standard Brownian motion. If $2\alpha > \sigma^2$, the CIR process is strictly positive. Otherwise it can reach the boundary 0 in finite time with positive probability, but if the boundary is made instantaneously reflecting the process stays non-negative. In both cases, the stationary distribution is $\Gamma(2\alpha/\sigma^2, \sigma^2/\beta)$.

The CIR process plays important roles in financial mathematics, where it has been used to describe the evolution of interest rates (Cox, Ingersoll, Jr. & Ross (1985)), and to model the volatility in the Heston (1993) model. In the latter model, the dynamics of the logarithm of the price, P_t, of a financial asset is given by

$$dP_t = (\kappa + \nu X_t)dt + \sqrt{X_t}dB_t,$$

where the volatility process X is given by (6.1), and where the standard Brownian motion B may possibly be correlated with the Brownian motion W in (6.1). If high frequency observations of the asset price are available at the time points $j\delta$, $j = 0, \ldots, N$, then the integrated volatility over longer time intervals of length $\Delta = m\delta$ (e.g. hours or days)

$$\int_{(i-1)\Delta}^{i\Delta} X_t dt,$$

$i = 1, \ldots, n = [N/m]$, can be estimated by the quadratic variation/realized volatility

$$V_i = \sum_{j=(i-1)m+1}^{im} (P_{j\delta} - P_{(j-1)\delta})^2$$

$i = 1, \ldots, n$. We can therefore estimate the parameters α, β, σ in the volatility process (6.1) by treating the realized volatilities V_i, $i = 1, \ldots, n$, as observations Y_i of the type (2.2) with X given by (6.1) and $t_i = i\Delta$.

In the simulation study below, we investigate the bias of the estimators obtained by our methods for data Y_i, $i = 1, \ldots, n$ of the type (2.2) with X given by the CIR process (6.1).

6.1 The likelihood and the EM-algorithm

We apply the transformation (3.1), which for the CIR process is

$$h(x; \sigma) = \frac{2\sqrt{x}}{\sigma},$$

with $h^{-1}(x; \sigma) = \sigma^2 x^2/4$. Hence $U_t = h(X_t; \sigma) = 2\sqrt{X_t}/\sigma$, solves the stochastic differential equation

$$dU_t = \mu(U_t; \alpha, \beta, \sigma)dt + dW_t,$$

where

$$\mu(u; \alpha, \beta, \sigma) = \frac{4\alpha - \beta\sigma^2 u^2 - \sigma^2}{2\sigma^2 u},$$

so that

$$a(u; \alpha, \beta, \sigma) = \log(u)\left(\frac{2\alpha}{\sigma^2} - \frac{1}{2}\right) - \frac{\beta u^2}{4}.$$

For the CIR model the full log-likelihood function (3.5) is given by

$$\begin{aligned}
&\log L(\theta; Y_1, \ldots, Y_n, U_t, t \in [0, t_n]) \\
&= \sum_{i=1}^{n} \log\phi\left(Y_i; \tfrac{1}{4}\sigma^2 \int_{t_{i-1}}^{t_i} (U_s)^2 ds, \tau^2\right) + \left(\frac{2\alpha}{\sigma^2} - \frac{1}{2}\right)\log\left(\frac{U_{t_n}}{U_0}\right) + \frac{\beta}{4}(U_0^2 - U_{t_n}^2) \\
&\quad + \frac{\alpha\beta t_n}{\sigma^2} - \frac{\beta^2}{8}\int_0^{t_n} U_t^2 dt + \left(\frac{2\alpha}{\sigma^2} - \frac{2\alpha^2}{\sigma^4} - \frac{3}{8}\right)\int_0^{t_n} U_t^{-2} dt,
\end{aligned}$$

where $\theta = (\alpha, \beta, \sigma, \tau^2)$.

Now, we can specify the EM algorithm.

E-STEP

The objective function $g(\theta)$ is for the CIR process given by

$$\begin{aligned} g(\theta) &= -\frac{1}{2\tau^2(M-M_0)} \sum_{k=M_0+1}^{M} \sum_{i=1}^{n} \left(Y_i - \tfrac{1}{4}\sigma^2 \int_{t_{i-1}}^{t_i} (U_t^{(k)})^2 dt \right)^2 - \frac{n}{2}\log(2\pi\tau^2) + \frac{\alpha\beta t_n}{\sigma^2} \\ &+ \frac{2\alpha\sigma^{-2} - \frac{1}{2}}{M-M_0} \sum_{k=M_0+1}^{M} \log\left(\frac{U_{t_n}^{(k)}}{U_0^{(k)}}\right) + \frac{\beta}{4(M-M_0)} \sum_{k=M_0+1}^{M} ((U_0^{(k)})^2 - (U_{t_n}^{(k)})^2) \\ &+ \frac{2\alpha\sigma^2 - 2\alpha^2}{\sigma^4(M-M_0)} \sum_{k=M_0+1}^{M} \int_0^{t_n} (U_t^{(k)})^{-2} dt - \frac{3}{8(M-M_0)} \sum_{k=M_0+1}^{M} \int_0^{t_n} (U_t^{(k)})^{-2} dt \\ &- \frac{\beta^2}{8(M-M_0)} \sum_{k=M_0+1}^{M} \int_0^{t_n} (U_t^{(k)})^2 dt. \end{aligned}$$

Here $U_t^{(k)} = 2\sqrt{X_t^{(k)}}/\hat{\sigma}$, where $X_t^{(k)}$ is the k-th sample path of the process X simulated conditionally on the data Y using the Algorithms 1 – 3 with the parameter value obtained in the previous step $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \hat{\tau}^2)$.

M-STEP

The maximum $\hat{\theta}$ is obtained as the solution to the following system of equations

$$\frac{\partial g(\theta)}{\partial \alpha} = \frac{2}{\sigma^2(M-M_0)} \sum_{k=M_0+1}^{M} \log\left(\frac{U_{t_n}^{(k)}}{U_0^{(k)}}\right) + \frac{\beta t_n}{\sigma^2} + \frac{2\sigma^2 - 4\alpha}{\sigma^4(M-M_0)} \sum_{k=M_0+1}^{M} \int_0^{t_n} (U_t^{(k)})^{-2} dt = 0,$$

$$\frac{\partial g(\theta)}{\partial \beta} = \frac{1}{4(M-M_0)} \sum_{k=M_0+1}^{M} ((U_0^{(k)})^2 - (U_{t_n}^{(k)})^2) + \frac{\alpha t_n}{\sigma^2} - \frac{\beta}{4(M-M_0)} \sum_{k=M_0+1}^{M} \int_0^{t_n} (U_t^{(k)})^2 dt = 0,$$

$$\begin{aligned} \frac{\partial g(\theta)}{\partial \sigma} &= \frac{\sigma}{2\tau^2(M-M_0)} \sum_{k=M_0+1}^{M} \sum_{i=1}^{n} \left(Y_i - \tfrac{1}{4}\sigma^2 \int_{t_{i-1}}^{t_i} (U_t^{(k)})^2 dt \right) \int_{t_{i-1}}^{t_i} (U_t^{(k)})^2 dt - \frac{2\alpha\beta t_n}{\sigma^3} \\ &- \frac{4\alpha}{\sigma^3(M-M_0)} \sum_{k=M_0+1}^{M} \log\left(\frac{U_{t_n}^{(k)}}{U_0^{(k)}}\right) + \frac{4\alpha(2\alpha - \sigma^2)}{\sigma^5(M-M_0)} \sum_{k=M_0+1}^{M} \int_0^{t_n} (U_t^{(k)})^{-2} dt = 0, \end{aligned}$$

and

$$\frac{\partial g(\theta)}{\partial \tau^2} = \frac{1}{2\tau^2(M-M_0)} \sum_{k=M_0+1}^{M} \sum_{i=1}^{n} \left(Y_i - \tfrac{1}{4}\sigma^2 \int_{t_{i-1}}^{t_i} (U_t^{(k)})^2 dt \right)^2 - \frac{n}{2\tau^2} = 0.$$

The solution $\hat{\theta}$ is given by

$$\begin{aligned} \hat{\alpha} &= \frac{C_5(2C_2C_4 + 2C_1C_4 + C_3 t_n)}{C_6(C_4C_2 - t_n^2)} \\ \hat{\beta} &= \frac{C_3C_2 + 2t_n(C_2 + C_1)}{C_4C_2 - t_n^2} \end{aligned}$$

$$\hat{\sigma} = \sqrt{4C_5/C_6}$$
$$\hat{\tau}^2 = \frac{C_7 C_6 - C_5^2}{n(M-M_0)C_6},$$

where the values of the constants C_i are

$$C_1 = \frac{1}{M-M_0} \sum_{k=M_0+1}^{M} \log\left(U_{t_n}^{(k)}/U_0^{(k)}\right)$$
$$C_2 = \sum_{k=M_0+1}^{M} \int_{t_0}^{t_n} (U_t^{(k)})^{-2} dt$$
$$C_3 = \sum_{k=M_0+1}^{M} ((U_0^{(k)})^2 - (U_{t_n}^{(k)})^2)$$
$$C_4 = \sum_{k=M_0+1}^{M} \int_0^{t_n} (U_t^{(k)})^2 dt$$
$$C_5 = \sum_{k=M_0+1}^{M} \sum_{i=1}^{n} Y_i \int_{t_{i-1}}^{t_i} (U_t^{(k)})^2 dt$$
$$C_6 = \sum_{k=M_0+1}^{M} \sum_{i=1}^{n} (\int_{t_{i-1}}^{t_i} (U_t^{(k)})^2 dt)^2$$
$$C_7 = \sum_{k=M_0+1}^{M} \sum_{i=1}^{n} Y_i.$$

6.2 A simulation study

Here we present a simulation study for the integrated CIR-model. We simulated 1500 datasets, and for each of them obtained estimates by means of our EM-algorithm. Each data set was obtained by simulating a sample path of length 1500 with initial distribution $X_0 \sim \Gamma(2\alpha/\sigma^2, \sigma^2/\beta)$, and then calculating data Y_i, $i = 1, \ldots, 1500$ by (2.2) with $t_i = i$, $i = 0. \ldots, n$. The parameter values were $\alpha = 0.5$, $\beta = 0.2$, $\sigma = 0.5$ and $\tau^2 = 1.25$.

The EM-algorithm was run with $M = 10000$ and $M_0 = 1000$ for three values of λ. The average of the estimates obtained for the 1500 dataset are given in Table 6.1. Also for the CIR model the bias is small.

λ	α	β	σ	τ^2
30	0.4802	0.2056	0.4787	1.2432
20	0.4727	0.2043	0.4698	1.2406
10	0.4587	0.1965	0.4609	1.2287

Table 6.1: Average of parameter estimates obtained from 1500 simulated datasets with parameter values $\alpha = 0.1$, $\sigma = 0.5$ and $\tau^2 = 1.25$.

7 Concluding remarks

We have presented an EM-algorithm for obtaining maximum likelihood estimates of parameters in diffusion models when the data is a discrete time sample of the integral of the diffusion process contaminated by measurement errors, while no direct observations of the process itself are available. This was done by viewing the data as an incomplete observation, where the full data set includes a continuous time record of the diffusion process.

It is not difficult to generalize the method presented in this paper to the situation, where the diffusion process is integrated w.r.t. a more general measure than the Lebesgue measure considered in this paper. This would allow analysis of e.g. weighted averages of diffusion processes, and discrete time observation would be a particular case. Note also that a Gibbs sampler could easily be set up in close analogy to the EM-algorithm used in the present paper. This would be much closer to the approach in Chib, Pitt & Shephard (2006).

References

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

Aït-Sahalia, Y. (2008). "Closed-form likelihood expansions for multivariate diffusions". *Ann. Statist.*, 36:906–937.

Aït-Sahalia, Y. & Mykland, P. (2003). "The effects of random and discrete sampling when estimating continuous-time diffusions". *Econometrica*, 71:483–549.

Andersen, T. G.; Bollerslev, T.; Diebold, F. X. & Ebens, H. (2001a). "The distribution of realized stock return volatility". *J. Finan. Econ.*, 61:43–76.

Andersen, T. G.; Bollerslev, T.; Diebold, F. X. & Labys, P. (2001b). "The distribution of exchange rate volatility". *J. Amer. Statist. Assoc.*, 96:42–55.

Barndorff-Nielsen, O. E. & Shephard, N. (2002). "Econometric analysis of realized volatility and its use in estimating stochastic volatility models". *J. R. Statist. Soc B*, 64:253–280.

Beskos, A.; Papaspiliopoulos, O. & Roberts, G. O. (2006). "Retrospective exact simulation of diffusion sample paths with applications". *Bernoulli*, 12:1077–1098.

Beskos, A.; Papaspiliopoulos, O. & Roberts, G. O. (2007). "A new factorization of diffusion measure with view towards simulation". To be published.

Beskos, A.; Papaspiliopoulos, O.; Roberts, G. O. & Fearnhead, P. (2006). "Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes". *J. R. Statist.Soc. B*, 68:333–382.

Bladt, M. & Sørensen, M. (2009). "Simple simulation of diffusion bridges with application to likelihood inference for diffusions". Working paper, Dept. Math. Sciences, Univ. of Copenhagen.

Bollerslev, T. & Zhou, H. (2002). "Estimating stochastic volatility diffusion using conditional moments of integrated volatility". *Journal of Econometrics*, 109:33–65.

Chib, S. & Greenberg, E. (1995). "Understanding the Metropolis-Hastings algorithm". *The American Statistician*, 49:327–335.

Chib, S.; Pitt, M. K. & Shephard, N. (2006). "Likelihood based inference for diffusion driven state space models". Working paper.

Comte, F.; Genon-Catalot, V. & Rozenholc, Y. (2009). "Nonparametric adaptive estimation for integrated diffusions". *Stochastic processes and their applications*, 119:811–834.

Cox, J. C.; Ingersoll, Jr., J. E. & Ross, S. A. (1985). "A Theory of the Term Structure of Interest Rates". *Econometrica*, 53:385–407.

Dempster, A. P.; Laird, N. M. & Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm (with discussion)". *J. Roy. Statist. Soc. B*, 39:1 – 38.

Ditlevsen, P. D.; Ditlevsen, S. & Andersen, K. K. (2002). "The fast climate fluctuations during the stadial and interstadial climate states". *Annals of Glaciology*, 35:457–462.

Ditlevsen, S. & Sørensen, M. (2004). "Inference for observations of integrated diffusion processes". *Scand. J. Statist.*, 31:417–429.

Durham, G. B. & Gallant, A. R. (2002). "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes". *J. Business & Econom. Statist.*, 20:297–338.

Elerian, O.; Chib, S. & Shephard, N. (2001). "Likelihood inference for discretely observed non-linear diffusions". *Econometrica*, 69:959–993.

Eraker, B. (2001). "MCMC Analysis of Diffusion Models with Application to Finance". *Journal of Business and Economic Statistics*, 19:177–191.

Forman, J. L. & Sørensen, M. (2008). "The Pearson diffusions: A class of statistically tractable diffusion processes". *Scand. J. Statist.*, 35:438–465.

Gilks, W. R.; Richardson, S. & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Gloter, A. (2000). "Parameter estimation for a discrete sampling of an integrated Ornstein-Uhlenbeck process". *Statistics*, 35:225–243.

Gloter, A. (2006). "Parameter estimation for a discretely observed integrated diffusion process". *Scand. J. Statist.*, 33:83–104.

Heston, S. L. (1993). "A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options". *Review of Financial Studies*, 6:327–343.

Jacod, J. & Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, New York.

Karatzas, I. & Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus.* Springer, New York.

Karlin, S. & Taylor, H. M. (1981). *A Second Course in Stochastic Processes.* Academic Press, Orlando.

Kloeden, P. E. & Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations.* 3rd revised printing. Springer-Verlag, New York.

Liptser, R. S. & Shiryaev, A. N. (1977). *Statistics of Random Processes.* Springer, New York.

McLachlan, G. J. & Krishnan, T. (1997). *The EM algorithm and Extensions.* Wiley, New York.

Øksendal, B. (1998). *Stochastic Differential Equations.* Springer, New York.

Ozaki, T. (1985). "Non-linear time series models and dynamical systems". In Hannan, E. J.; Krishnaiah, P. R. & Rao, M. M., editors, *Handbook of Statistics, Vol. 5*, pages 25–83. Elsevier Science Publishers.

Pedersen, A. R. (1995). "A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations". *Scand. J. Statist.*, 22:55–71.

Poulsen, R. (1999). "Approximate maximum likelihood estimation of discretely observed diffusion processes". Working Paper 29, Centre for Analytical Finance, Aarhus.

Roberts, G. O. & Stramer, O. (2001). "On inference for partially observed nonlinear diffusion models using Metropolis-Hastings algorithms". *Biometrika*, 88:603–621.

Sørensen, M. (1997). "Estimating functions for discretely observed diffusions: a review". In Basawa, I. V.; Godambe, V. P. & Taylor, R. L., editors, *Selected Proceedings of the Symposium on Estimating Functions*, pages 305–325. Hayward: Institute of Mathematical Statistics. IMS Lecture Notes – Monograph Series, Vol. 32.

Sørensen, M. (2000). "Prediction-Based Estimating Functions". *Econometrics Journal*, 3:123–147.

Sørensen, M. (2010). "Estimating functions for diffusion-type processes". In Kessler, M.; Lindner, A. & Sørensen, M., editors, *Statistical Methods for Stochastic Differential Equations*. Chapman & Hall. Forthcoming.

Research Papers 2010

CREATES
Center for Research in Econometric Analysis of Time Series

2010-19: Jeroen V.K. Rombouts and Lars Stentoft: Multivariate Option Pricing with Time Varying Volatility and Correlations

2010-20: Charlotte Christiansen: Intertemporal Risk-Return Trade-off in

2010-21: Marco Aiolfi, Carlos Capistrán and Allan Timmermann : Forecast Combinations

2010-22: Ivan Nourdin, Giovanni Peccati and Mark Podolskij: Quantitative Breuer-Major Theorems

2010-23: Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Bootstrapping Density-Weighted Average Derivatives

2010-24: Søren Johansen and Morten Ørregaard Nielsen: Likelihood inference for a fractionally cointegrated vector autoregressive model

2010-25: Tom Engsted and Bent Nielsen: Testing for rational bubbles in a co-explosive vector autoregression

2010-26: Robinson Kruse: On European monetary integration and the persistence of real effective exchange rates

2010-27: Sanne Hiller and Robinson Kruse: Milestones of European Integration: Which matters most for Export Openness?

2010-28: Robinson Kruse: Forecasting autoregressive time series under changing persistence

2010-29: Nikolaus Hautsch and Mark Podolskij: Pre-Averaging Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence

2010-30: Martin M. Andreasen: Non-linear DSGE Models and The Central Difference Kalman Filter

2010-31: Morten Ørregaard Nielsen and Per Frederiksen: Fully Modified Narrow-Band Least Squares Estimation of Weak Fractional Cointegration

2010-32: Mogens Bladt and Michael Sørensen: Simple simulation of diffusion bridges with application to likelihood inference for diffusions

2010-33: Fernando Baltazar-Larios and Michael Sørensen: Maximum likelihood estimation for integrated diffusion processes