Statistical Inference for Integrated Diffusion Processes

Sørensen, Michael
University of Copenhagen, Department of Mathematical Sciences
Universitetsparken 5
DK-2100 Copenhagen Ø, Denmark
E-mail: michael@math.ku.dk

We present estimation methods for data that are modelled as integrals of a diffusion process over disjoint time-intervals. An example of an application is realized volatility in financial econometrics, see e.g. Bollerslev & Zhou (2002). Such models are also used in fields of engineering and of the sciences. The analysis of the records of the concentration of oxygen isotopes in ice-core data from Greenland and Antarctica in an example, see e.g. Ditlevsen, Ditlevsen & Andersen (2002). Ice-core data are used to investigate the paleo-climate.

Let X be a d-dimensional diffusion process given as the solution to the stochastic differential equation

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

where σ is a $d \times d$-matrix and W a d-dimensional standard Wiener process. We consider partial observations of the type

$$Y_i = \int_{t_{i-1}}^{t_i} k(X_s)ds + Z_i, \quad i = 1, \ldots, n, \tag{2}$$

where $t_0 = 0 < t_1 < t_2 < \ldots < t_n$, where k is a real-valued function, and where the d-dimensional measurement errors Z_i are independent and identically distributed and independent of X. Typical examples of the function k are $k(x) = x_1$ or $k(x) = x_1 + \cdots + x_d$, where x_i denotes the ith coordinate of x. The observations form a non-Markovian stochastic process, which complicates statistical inference.

Estimation based on observations that are integrals of a diffusion ($d = 1$, $k(x) = x$) with no measurement error was studied by Gloter (2000), Bollerslev & Zhou (2002), Ditlevsen & Sørensen (2004), and Gloter (2006), while maximum likelihood estimation in the case of measurement errors was investigated by Baltazar-Larios & Sørensen (2010).

Prediction-based estimating functions

First we present an approach using prediction-based estimating functions, which were introduced in Sørensen (2000). A contemporary review can be found in Sørensen (2011a). For a general review of estimating function techniques for diffusion-type models, see Sørensen (2011b).

Suppose that we have observed the random variables $Y_1, \ldots, Y_n$ that form a stationary stochastic process, the distribution of which is parametrised by $\Theta \subseteq \mathbb{R}^p$. Assume that $E_\theta(Y_i^{2m}) < \infty$ for all $\theta \in \Theta$ for some $m \in \mathbb{N}$. For each $i = r+1, \ldots, n$ and $j = 1, \ldots, m$ let the class $\{Z_{jk}^{(i-1)} \mid k = 1, \ldots, q_j\}$ be a subset of the random variables $\{Y_{i-\ell}^{\kappa} \,|\, \ell = 1, \ldots, r, \kappa = 0, \ldots, j\}$, where $Z_{j1}^{(i-1)}$ is always equal to 1. We wish to predict Y_i^j by means of linear combinations of the $Z_{jk}^{(i-1)}$-s for each of the values of i and j listed above and then to use suitable linear combinations of the prediction errors to estimate the parameter θ. Let $\mathcal{P}_{i-1,j}$ denote the space of predictors of Y_i^j, i.e. the space of square integrable random variables spanned by $Z_{j1}^{(i-1)}, \ldots, Z_{jq_j}^{(i-1)}$. The elements of $\mathcal{P}_{i-1,j}$ are of the form $a^T Z_j^{(i-1)}$, where $a^T = (a_1, \ldots, a_{q_j})$ and $Z_j^{(i-1)} = (Z_{j1}^{(i-1)}, \ldots, Z_{jq_j}^{(i-1)})^T$ are q_j-dimensional vectors. In this paper vectors are column vectors and T denotes transposition.

We will use estimating functions of the type

$$G_n(\theta) = \sum_{i=r+1}^{n} \sum_{j=1}^{m} \Pi_j^{(i-1)}(\theta) \left[Y_i^j - \hat{\pi}_j^{(i-1)}(\theta) \right] \tag{3}$$

where $\Pi_j^{(i-1)}(\theta)$ is a p-dimensional data dependent vector of weights, the coordinates of which belong to $\mathcal{P}_{i-1,j}$, and where $\hat{\pi}_j^{(i-1)}(\theta)$ is the minimum mean square error predictor of Y_i^j in $\mathcal{P}_{i-1,j}$, which is the usual L_2-projection of Y_i^j onto $\mathcal{P}_{i-1,j}$. Define $C_j(\theta)$ as the covariance matrix of $(Z_{j2}^{(r)}, \ldots, Z_{jq_j}^{(r)})^T$ for the parameter value θ and $b_j(\theta) = (\mathrm{Cov}_\theta(Z_{j2}^{(r)}, Y_{r+1}^j), \ldots, \mathrm{Cov}_\theta(Z_{jq_j}^{(r)}, Y_{r+1}^j))^T$. Then we have

$$\hat{\pi}_j^{(i-1)}(\theta) = \hat{a}_j(\theta)^T Z_j^{(i-1)}$$

where $\hat{a}_j(\theta)^T = (\hat{a}_{j1}(\theta), \hat{a}_{j*}(\theta)^T)$ with $\hat{a}_{j*}(\theta)^T = (\hat{a}_{j2}(\theta), \ldots, \hat{a}_{jq_j}(\theta))$ defined by

$$\hat{a}_{j*}(\theta) = C_j(\theta)^{-1} b_j(\theta) \tag{4}$$

and

$$\hat{a}_{j1}(\theta) = E_\theta(Y_1^j) - \sum_{k=2}^{q_j} \hat{a}_{jk}(\theta) E_\theta(Z_{jk}^{(r)}). \tag{5}$$

Thus to find $\hat{\pi}_j^{(i-1)}(\theta)$, $j = 1, \ldots, m$, we need to calculate moments of the form

$$E_\theta(Y_1^\kappa Y_k^j), \quad 0 \le \kappa, j \le m, \quad k = 1, \ldots, r. \tag{6}$$

If the original diffusion model (1) is exponentially ρ-mixing, then the observations inherit this property. In such cases r will usually not need to be chosen particularly large. If Y_i^j is restricted to have mean zero, we need not include a constant in the space of predictors, i.e. we need only the space spanned by $Z_{j2}^{(i-1)}, \ldots, Z_{jq_j}^{(i-1)}$. In many situations $m = 2$ with $Z_{jk}^{(i-1)} = Y_{i-k}$, $k = 1, \ldots, r, j = 1, 2$ and $Z_{2k}^{(i-1)} = Y_{i+r-k}^2$, $k = r+1, \ldots, 2r$, will be a reasonable choice. Including predictors in the form of lagged terms $Y_{i-k}Y_{i-k-l}$ for a number of lags l's might also be of relevance.

The choice of the weights $\Pi_j^{(i-1)}(\theta)$ in (3) for which the asymptotic variance of the estimators is minimized is the Godambe optimal prediction-based estimating function, that was derived in Sørensen (2000). The optimal estimating function of the type (3) can be written in the form

$$G_n^*(\theta) = A^*(\theta) \sum_{i=r+1}^{n} H^{(i)}(\theta), \tag{7}$$

where

$$H^{(i)}(\theta) = Z^{(i-1)} \left(F(Y_i) - \hat{\pi}^{(i-1)}(\theta) \right), \tag{8}$$

with $F(x) = (x, x^2, \ldots, x^m)^T$, $\hat{\pi}^{(i-1)}(\theta) = (\hat{\pi}_1^{(i-1)}(\theta), \ldots \hat{\pi}_m^{(i-1)}(\theta))^T$ and

$$Z^{(i-1)} = \begin{pmatrix} Z_1^{(i-1)} & 0_{q_1} & \cdots & 0_{q_1} \\ 0_{q_2} & Z_2^{(i-1)} & \cdots & 0_{q_2} \\ \vdots & \vdots & & \vdots \\ 0_{q_m} & 0_{q_m} & \cdots & Z_m^{(i-1)} \end{pmatrix}. \tag{9}$$

Here 0_{q_j} denotes the q_j-dimensional zero-vector. Finally,

$$A^*(\theta) = \partial_\theta \hat{a}(\theta)^T \bar{C}(\theta) \bar{M}(\theta)^{-1}, \tag{10}$$

with

$$\bar{M}(\theta) = E_\theta\Big(H^{(r+1)}(\theta)H^{(r+1)}(\theta)^T\Big) + \sum_{k=1}^{\infty}\Big[E_\theta\Big(H^{(r+1)}(\theta)H^{(r+1+k)}(\theta)^T\Big) + E_\theta\Big(H^{(r+1+k)}(\theta)H^{(r+1)}(\theta)^T\Big)\Big], \tag{11}$$

$$\bar{C}(\theta) = E_\theta\left(Z^{(i-1)}(Z^{(i-1)})^T\right), \tag{12}$$

and

$$\hat{a}(\theta)^T = \left(\hat{a}_1(\theta)^T, \ldots, \hat{a}_m(\theta)^T\right), \tag{13}$$

where $\hat{a}_j(\theta)$ is given by (4) and (5). A necessary condition that the moments in (11) exist is that $E_\theta(Y_i^{4m}) < \infty$ for all $\theta \in \Theta$. For (7) to be optimal we need that the matrix $\partial_\theta \hat{a}(\theta)^T$ has full rank. The matrix $\bar{M}(\theta)$ is invertible under weak conditions.

For exponentially mixing diffusions (11) can in practice often be truncated so that only relatively few terms need to be calculated. In practice, it is usually also a good idea to replace $A^*(\theta)$ by $A^*(\bar{\theta}_n)$, where $\bar{\theta}_n$ is a $\sqrt{n}$-consistent estimator of θ. This has the advantages that (11) need only be calculated once and that a simpler estimating equation is obtained, while the asymptotic variance of the estimator is unchanged. The estimator $\bar{\theta}_n$ can, for instance, be obtained from an estimating function similar to (7), where $A^*(\theta)$ has been replaced by a suitable simple matrix independent of θ, but such that the estimating equation has a solution. A simple possibility is to use the first p coordinates of $H^{(i)}(\theta)$, where p is the dimension of the parameter. In order to calculate (11), we need mixed moments of the form

$$E_\theta[Y_1^{k_1}Y_i^{k_2}Y_j^{k_3}Y_\ell^{k_4}], \quad 1 \le i \le j \le \ell \quad k_1+k_2+k_3+k_4 \le 4m \tag{14}$$

where k_i, $i = 1, \ldots, 4$ are non-negative integers.

To find the moments (6), we use that by the binomial formula,

$$\begin{aligned} E_\theta(Y_1^{k_1}Y_\ell^{k_2}) &= E_\theta\left((B_1+Z_1)^{k_1}(B_\ell+Z_\ell)^{k_2}\right) \\ &= \sum_{i_1=0}^{k_1}\sum_{i_2=0}^{k_2}\binom{k_1}{i_1}\binom{k_2}{i_2}E_\theta(B_1^{i_1}B_\ell^{i_2})E_\theta(Z_1^{k_1-i_1})E_\theta(Z_\ell^{k_2-i_2}) \end{aligned}$$

(with a slight modification for $\ell = 1$), where

$$B_i = \int_{t_{i-1}}^{t_i} k(X_s)ds.$$

The distribution of the measurement error Z_i can depend on components of the unknown parameter θ. The moments $E_\theta(B_1^{i_1}B_\ell^{i_2})$ can be found by

$$E\left(B_1^{i_1}B_\ell^{i_2}\right) = \int_A E\left(k(X_{v_1})\cdots k(X_{v_{i_1}})k(X_{u_1})\cdots k(X_{u_{i_2}})\right) du_{i_2}\cdots du_1\, dv_{i_1}\cdots dv_1,$$

where $1 \le \ell$ and $A = [0\,,\, t_1]^{i_1} \times [t_{\ell-1}\,,\, t_\ell]^{i_2}$. Thus we need to calculate mixed moments of the type $E(k(X_{t_1})\cdots k(X_{t_\ell}))$. Such mixed moments can be determined by simulation. A mixed moment of the form (14) can similarly be expressed as an integral of expectations of the type $E(k(X_{t_1})\cdots k(X_{t_\ell}))$. Sometimes these mixed moments and their integral can be found explicitly such that an explicit optimal estimating function is available. This is, for instance, the case for the Pearson diffusions, see Forman & Sørensen (2008), when $d = 1$ and $k(x) = x$. For these diffusions, $E(X_{t_1}\cdots X_{t_\ell})$ depends on $t_1, \ldots, t_\ell$ through sums and products of exponential functions.

Example 1 Consider observations where $d = 1$ and $k(x) = x$, where the diffusion process X is the square root process

$$dX_t = -\beta(X_t - \alpha)dt + \tau\sqrt{X_t}dW_t, \;\; X_0 > 0,$$

and where there are no measurement errors. We will find a prediction-based estimating function with $F(x) = (x, x^2)^T$ and with predictors given by $\pi_1^{(i-1)} = a_{1,0} + a_{1,1}Y_{i-1}$ and $\pi_2^{(i-1)} = a_{2,0}$. Then the minimum mean square error predictors are

$$\begin{aligned}\hat{\pi}_1^{(i-1)}(Y_{i-1};\theta) &= \mu\,(1 - a(\beta)) + a(\beta)Y_{i-1},\\ \hat{\pi}_2^{(i-1)}(\theta) &= \alpha^2 + \alpha\tau^2\beta^{-3}\Delta^{-2}(e^{-\beta\Delta} - 1 + \beta\Delta)\end{aligned}$$

with

$$a(\beta) = \frac{(1 - e^{-\beta\Delta})^2}{2(\beta\Delta - 1 + e^{-\beta\Delta})}.$$

The optimal prediction-based estimating function is

$$\sum_{i=1}^{n}\begin{pmatrix}1\\ Y_{i-1}\\ 0\end{pmatrix}[Y_i - \hat{\pi}_1^{(i-1)}(Y_{i-1};\theta)] + \sum_{i=1}^{n}\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}[Y_i^2 - \hat{\pi}_2^{(i-1)}(\theta)],$$

from which we obtain the estimators

$$\begin{aligned}\hat{\alpha} &= \frac{1}{n}\sum_{i=1}^{n}Y_i + \frac{a(\hat{\beta})Y_n - Y_1}{(n-1)(1 - a(\hat{\beta}))}\\ \sum_{i=2}^{n}Y_{i-1}Y_i &= \hat{\alpha}(1 - a(\hat{\beta}))\sum_{i=2}^{n}Y_{i-1} + a(\hat{\beta})\sum_{i=2}^{n}Y_{i-1}^2\\ \hat{\tau}^2 &= \frac{\hat{\beta}^3\Delta^2\sum_{i=2}^{n}(Y_i^2 - \hat{\alpha}^2)}{(n-1)\hat{\alpha}(e^{-\hat{\beta}\Delta} - 1 + \hat{\beta}\Delta)}.\end{aligned}$$

The estimators are explicit apart from $\hat{\beta}$, which can easily be found numerically by solving a non-linear equation in one variable. For details, see Ditlevsen & Sørensen (2004). ◯

An interesting more general case is that of hypoelliptic stochastic differential equations, where one or more components are not directly affected by the Wiener process and hence are smooth. If a smooth component is observed at discrete time points, then we obtain data of the type considered in this paper. Hypoelliptic stochastic differential equations are, for instance, used to model molecular dynamics, see e.g. Pokern, Stuart & Wiberg (2009). A simple example is the stochastic harmonic oscillator

$$\begin{aligned}dX_{1,t} &= -(\beta_1 X_{1,t} + \beta_2 X_{2,t})\,dt + \gamma dW_t\\ dX_{2,t} &= X_{1,t}\,dt,\end{aligned}$$

$\beta_1, \beta_2, \gamma > 0$, where the position of the oscillator, X_2, is observed at discrete time points.

Maximum likelihood estimation

The likelihood function for a discretely sampled integrated diffusion with observation errors is in almost all cases not explicitly available. In this section an EM-algorithm for the case $d = 1$ proposed by Baltazar-Larios & Sørensen (2010) is outlined. It is based on the fact 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). To do so we need to calculate the conditional expectation of the log-likelihood function for the full model given the observations. This is done by simulating sample paths of the diffusion process given the data using ideas from Chib, Pitt & Shephard (2006).

In the following we assume that the measurement errors, Z_i are normal distributed with mean zero and variance τ^2. Then, conditionally on the sample path of X, the observations Y_i, $i = 1, \ldots, n$ are independent and normal distributed:

$$Y_i \,|\, \{X_t : t \in [0, t_n]\} \sim N\left(\int_{t_{i-1}}^{t_i} X_s ds, \tau^2\right), \tag{15}$$

The probability measures corresponding to the full observation of a diffusion sample path in the time interval $[0, t_n]$ are in general singular because the diffusion coefficient depends on the parameter θ. In order to obtain a likelihood function, we use the standard 1-1 transformation

$$h(x;\theta) = \int_{x^*}^{x} \frac{1}{\sigma(u;\theta)} du, \tag{16}$$

where x^* is some arbitrary element of the state space of X. By Ito's formula

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

satisfies the stochastic differential equation

$$dU_t = \mu(U_t;\theta)dt + dW_t, \tag{17}$$

with

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

where σ' denotes the derivative of σ w.r.t. x.

It follows from (15) that the likelihood of Y conditional on the sample path of U in $[0, t_n]$ is

$$L(\theta; Y_1, \ldots, Y_n \,|\, U_t, t \in [0, t_n]) = \prod_{i=1}^{n} \phi(Y_i; \int_{t_{i-1}}^{t_i} h^{-1}(U_s;\theta)ds, \tau^2) \tag{18}$$

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. By Girsanov's formula, 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

$$\begin{aligned} \log L(\theta; Y_1, \ldots, Y_n, U_t, t \in [0, t_n]) \quad &= \quad \sum_{i=1}^{n} \log \phi(Y_i; \int_{t_{i-1}}^{t_i} h^{-1}(U_s;\theta)ds, \tau^2) \\ &\quad + a(U_{t_n};\theta) - a(U_0;\theta) - \frac{1}{2}\int_0^{t_n} \left(\mu(U_t;\theta)^2 + \mu'(U_t;\theta)\right) dt, \end{aligned} \tag{19}$$

where

$$a(u;\theta) = \int^{u} \mu(x;\theta)dx.$$

We can now apply the EM-algorithm to the full log-likelihood function (19) 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. 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 $\hat{\theta}$, and calculate

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

   for a suitable burn-in period M_0 and for 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). 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 (19) given the data. This can be done by means of a Metropolis-Hastings algorithm. 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). For details of the algorithm, see Baltazar-Larios & Sørensen (2010).

An essential step in the algorithm is to simulate a diffusion bridge (not conditional on the observations). This is done by applying the method for approximate diffusion bridge simulation proposed by Bladt & Sørensen (2009). This method is particularly useful and accurate in relatively long intervals, which is important in the present application. The main idea of the technique (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 Milstein scheme, see Kloeden & Platen (1999). The method is therefore very easy to implement.

An alternative method that provide exact diffusion bridges has been proposed by Beskos, Papaspiliopoulos & Roberts (2006). When the the drift and diffusion coefficients satisfy certain boundedness conditions, this algorithm is relatively simple, but under weaker condition it is more complex. Bladt & Sørensen (2009) showed that the computational complexity of their method is linear in the length of the interval where the diffusion bridge is defined, while a simulation study indicated that for the method in Beskos, Papaspiliopoulos & Roberts (2006), the CPU time increases exponentially with the interval length.

Example 2 Consider the Ornstein-Uhlenbeck process, which is a solution of the stochastic differential equation

$$dX_t = -\alpha X_t d_t + \sigma dW_t,$$

where $\alpha > 0$ and $\sigma > 0$. Here U_t solves the stochastic differential equation

$$dU_t = -\alpha U_t d_t + dW_t,$$

and the full log-likelihood function (19) is given by

$$\begin{aligned}(20)\qquad &\log L(\theta; Y_1, \ldots, Y_n, U_t, t \in [0, t_n]) \\ &= \sum_{i=1}^{n} \log \phi\left(Y_i; \sigma \int_{t_{i-1}}^{t_i} U_s ds, \tau^2\right) + \frac{\alpha}{2}(U_0^2 - U_{t_n}^2 + t_n) - \frac{\alpha^2}{2}\int_0^{t_n} U_t^2 dt,\end{aligned}$$

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

In the M-step the estimator $\hat{\theta}$ is given by

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

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

and

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

○

REFERENCES (RÉFERENCES)

Baltazar-Larios, F. & Sørensen, M. (2010). "Maximum likelihood estimation for integrated diffusion processes". In Chiarella, C. & Novikov, A., editors, *Contemporary Quantitative Finance: Essays in Honour of Eckhard Platen*, pages 407–423. Springer.

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

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

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.

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

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.

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

Pokern, Y.; Stuart, A. M. & Wiberg, P. (2009). "Parameter estimation for partially observed hypoelliptic diffusions". *J. Roy. Statist. Soc. B*, 71:49–73.

Sørensen, M. (2000). "Prediction-based estimating functions". *Econometrics Journal*, 3:123–147.

Sørensen, M. (2011a). "Prediction-based estimating functions: review and new developments". To appear in *Brazilian Journal of Probability and Statistics*.

Sørensen, M. (2011b). "Estimating functions for diffusion-type processes". To appear in Kessler, M.; Lindner, A. & Sørensen, M., editors, *Statistical Methods for Stochastic Differential Equations*. Chapman & Hall.

RÉSUMÉ (ABSTRACT) — optional

Methods are presented for analysing data that are integrals of a diffusion process, i.e. the solution to a stochastic differential equation, observed with measurement error. First a relatively simple methodology is outlined: prediction-based estimating function. Then we present in detail a computationally more demanding method for obtaining maximum likelihood estimates. The data can be viewed as incomplete observations from a model with a tractable likelihood function. Therefore a simulated EM-algorithm is used to obtain maximum likelihood estimates of the model parameters. An essential part of the algorithm is a recent simple method for approximate simulation of diffusion bridges, which is used to simulate the full hidden data given the observations.