Parameter Estimation in Stochastic Differential Mixed-Effects Models

Umberto Picchini[1,2] Andrea De Gaetano[1] Susanne Ditlevsen[3]

[1] BioMatLab, IASI - National Research Council, Rome, Italy;
[2] Department of Statistics, Probability and Applied Statistics, University of Rome "La Sapienza";
[3] Department of Biostatistics, University of Copenhagen, Denmark.

Abstract

Stochastic differential equation (SDE) models have shown useful to describe continuous time processes, e.g. a physiological process evolving in an individual. Biomedical experiments often imply repeated measurements on a series of individuals or experimental units and individual differences can be represented by incorporating random effects into the model. When both system noise and individual differences are considered, *stochastic differential mixed effects* models ensue. In most cases the likelihood function is not available, and thus maximum likelihood estimation is not possible. Here we propose to approximate the unknown likelihood function by first approximating the conditional transition density of the diffusion process given the random effects by a Hermite expansion, as suggested by Aït-Sahalia (2001, 2002), and then numerically integrate the obtained conditional likelihood with respect to the random effects. The approximated maximum likelihood estimators are evaluated on simulations from the Ornstein-Uhlenbeck process and Geometric Brownian motion.

Keywords: Approximate maximum likelihood; closed-form transition density expansion; Hermite expansion; random effects; Ornstein-Uhlenbeck process; Geometric Brownian motion; diffusion processes; stochastic differential equations; biomedical applications.

1 Introduction

Studies in which repeated measurements are taken on a series of individuals or experimental animals play an important role in biomedical research. It is often reasonable to assume that responses follow the same model form for all experimental subjects, but model parameters vary randomly among individuals. The increasing popularity of Mixed-Effects models lies in their ability to model total variation, splitting it into its within- and between-individual components. This often leads to more precise estimation of population parameters, which is especially useful in pharmacokinetic/pharmacodynamic (PK/PD) modeling, where enhanced precision of estimation translates into considerable savings both in resources and in human or animal discomfort.

Dynamical biological processes are usually modeled by means of systems of deterministic differential equations (ordinary (ODE), partial (PDE), or delay (DDE)). These however do not account for the noisy components of the system dynamics often present in biological systems. System error (or system noise) represents the cumulative effect on the actual state of the system of a host of mechanisms which cannot be individually included in the model description (like hormonal oscillations, variations of the stress level, variable muscular activity etc.). Noise in the differential equations describing the behavior of the system requires an extension to the class of stochastic differential equation (SDE) models.

The theory for Mixed-Effects models is well developed for deterministic models (without system error), both linear and non-linear (Lindstrom and Bates (1990), Breslow and Clayton (1993), Vonesh and Chinchilli (1997), Diggle *et al.* (2002)), and standard software for model fitting is available, see e.g. Pinheiro and Bates (2002) and references therein. Early and important references in the pharmacokinetic field are Sheiner and Beal (1980, 1981). On the other hand, to our knowledge there is practically no theory at present for SDE models with random effects. The problem here is that estimating parameters in SDE models is not straightforward, except for simple cases. A natural approach would be likelihood inference, but the transition densities of the process are rarely known, and thus it is usually not possible to write the likelihood function explicitly. In Jelliffe *et al.* (2000) methods for PK/PD population modeling are reviewed, but the authors regret that system noise is not considered since it is difficult to estimate. In Overgaard *et al.* (2005) and Tornøe *et al.* (2005) a SDE model with log-normal distributed random effects and a constant diffusion term is treated, but this constrains the class of models to be SDEs with additive noise. In Ditlevsen and De Gaetano (2005) the likelihood function for a simple SDE model with random effects is calculated explicitly, but generally the likelihood function is unavailable. Eventually, as SDE models are more commonly applied to biomedical data, there will be an increasing need for developing a general theory for parameter estimation including mixed-effects.

In the present work an estimation method for the parameters of an SDE model incorporating random effects is proposed: these models may be called *stochastic differential mixed effects* (SDME) models. We consider SDME models whose drift and diffusion terms can depend linearly or nonlinearly on state variables and random effects following any continuous distribution, and an approximation to the likelihood function is computed. The likelihood can seldom be obtained in closed form since it involves explicit knowledge of the process transition density, which is often unavailable, and thus exact parameter estimators are also unavailable. It is therefore necessary to approximate the transition density numerically. To our knowledge, three ways have been proposed to do this:

1. solving numerically the Kolmogorov partial differential equations satisfied by the transition density (Lo (1988));

2. deriving a closed-form Hermite expansion to the transition density (Aït-Sahalia (2001, 2002));

3. simulating the process in order to Monte-Carlo-integrate the transition density (e.g. Pedersen (1995), Brandt and Santa-Clara (2002), Durham and Gallant (2002), Hurn and Lindsay (1999), Hurn *et al.* (2003), Nicolau (2002)): this methodology is known as *simulated maximum likelihood* (SML).

Each of these three techniques has been successfully implemented, but each has its limitations. Aït-Sahalia (2002) notes that methods 1 and 3 above are computationally intense and poorly accurate. Conversely, Durham and Gallant (2002) build on their importance sampling ideas in order to improve the performance of Pedersen's (1995) (or equivalently Brandt and Santa-Clara's (2002)) method, and point out that method 2 above, while accurate and fast, is only available for a small number of models.

We choose to employ the transition density approximation method suggested in Aït-Sahalia (2001, 2002) for time-homogeneous SDE, since it is the fastest and the most accurate among the available methods (Durham and Gallant (2002), Jensen and Poulsen (2002)). This is a desirable condition to make the parameter estimation procedure proposed here effective and reliable. We thus derive an approximation to the likelihood function and estimate the parameters of a SDME model by (approximated) maximum likelihood.

In this paper attention is restricted to time-homogeneous SDEs, but the method proposed can be applied to more general multidimensional SDME models, since also for time-inhomogeneous SDEs (which depend directly on time t, not only through the process values) the transition density can be expanded in closed form (Egorov *et al.* (2003)).

Evidence of the accuracy of the estimation method is given by simulation results, where exact and approximated parameter estimates are compared for SDME models of a Brownian motion with drift, of Geometric Brownian Motion and of the Ornstein-Uhlenbeck process. The estimates obtained are close to the true parameter values, and this result is achieved using moderate values of M (the number of experimental units, e.g. the number of subjects) and n (the number of observation for a given experimental unit). This is relevant for applications of these methods in situations where large data sets are unavailable, e.g. in biomedical applications, where Mixed-Effects theory is broadly applied.

2 Stochastic Differential Mixed-Effects Models

Consider a d-dimensional (Itô) SDE model for some continuous process evolving in M different experimental units (e.g. subjects) randomly chosen from a theoretical population:

$$dX_t^i = \mu(X_t^i, \theta, b^i)dt + \sigma(X_t^i, \theta, b^i)\, dW_t^i, \qquad X_0^i = x_0^i \qquad i = 1, \ldots, M \tag{1}$$

where $\theta \in \Theta \subseteq \mathbb{R}^p$ is a p-dimensional *fixed effects* parameter (the same for the entire population) and $b^i \equiv b^i(\Psi) \in B \subseteq \mathbb{R}^q$ is a q-dimensional *random effects* parameter (subject specific) whose density function in the population p_B is parametrized by an r-dimensional parameter $\Psi \in \Upsilon \subseteq \mathbb{R}^r$. The W_t^i are standard $(L \times 1)$-dimensional Brownian motions. The $W_t^{i,l}$ and b^j are assumed mutually independent for all $1 \leq i, j \leq M$, $1 \leq l \leq L$, and independent of X_0^i. The drift and the diffusion coefficient functions $\mu(\cdot) : E \times \Theta \times B \to \mathbb{R}^d$ and $\sigma(\cdot) : E \times \Theta \times B \to \mathbb{S}$ are assumed known up to the parameters, and are assumed sufficiently regular to ensure a unique solution (Øksendal (2000)), where $E \subseteq \mathbb{R}^d$ denote the state space of X_t^i and $\mathbb{S}$ denotes the set of the $d \times L$ positive definite matrices.

Assume that the distribution of X_t^i given (b^i, θ) and $X_s^i = x_s$, $s < t$, has a strictly positive density w.r.t. the Lebesgue measure on E, which we denote by

$$x \to p_X(x, t - s | x_s, b^i, \theta) > 0, \;\; x \in E. \tag{2}$$

Assume that subject i is observed at $(n_i + 1)$ discrete time points $(t_0^i, t_1^i, \ldots, t_{n_i}^i)$ for each coordinate k of the process $(k = 1, ..., d;\ i = 1, ..., M)$. Let $\underline{x}^i$ be the $d(n_i + 1)$-dimensional vector

containing the model (1) responses for the i'th subject, $\underline{x}^i = (x_0^{i,1}, \ldots, x_{n_i}^{i,1}, ..., x_0^{i,d}, \ldots, x_{n_i}^{i,d})$, where $x^{i,k}(t_j^i) = x_{t_j^i}^{i,k} = x_j^{i,k}$, and let $\underline{x} = (\underline{x}^1, ..., \underline{x}^M)$ be the N-dimensional total response vector, $N = \sum_{i=1}^M d(n_i + 1)$. Write $t_j^i - t_{j-1}^i = \Delta_j^i$ for the distance between observation $j-1$ and j for subject i.

We wish to estimate (θ, Ψ) given $\underline{x}$, and we call (1) a *stochastic differential mixed-effects* (SDME) model.

3 Maximum Likelihood Estimation in SDME Models

To obtain the marginal density of $\underline{x}^i$, we integrate the conditional density of the data given the non-observable random effects b^i with respect to the marginal density of the random effects, using the fact that $W_t^{i,l}$ and b^j are independent ($1 \leq i, j \leq M$, $1 \leq l \leq L$). This yields the likelihood

$$L(\theta, \Psi) \;=\; \prod_{i=1}^M p(\underline{x}^i|\theta, \Psi) \;=\; \prod_{i=1}^M \int_B p_{\underline{X}}(\underline{x}^i|b^i, \theta)\, p_B(b^i|\Psi)\, db^i \tag{3}$$

where $p(\cdot)$, $p_{\underline{X}}(\cdot)$ and $p_B(\cdot)$ are density functions. Notice that $p(\underline{x}^i|\cdot)$ and $p_{\underline{X}}(\underline{x}^i|\cdot)$ are in general different: the former being the density of $\underline{x}^i$ given (θ, Ψ), and the latter being the product of the transition densities for a given realization of the random effects and for a given θ :

$$p_{\underline{X}}(\underline{x}^i|b^i, \theta) \;=\; \prod_{j=1}^{n_i} p_X(x_j^i, \Delta_j^i|x_{j-1}^i, b^i, \theta), \tag{4}$$

where the transition densities $p_X(\cdot)$ are as in (2). The distribution of the random effects is often assumed to be normal, but $p_B(\cdot)$ could be any density function. Solving the integral in (3) yields the marginal likelihood of the parameters, independent of the random effects b^i; by maximizing the resulting expression (3) with respect to θ and Ψ we obtain the corresponding maximum likelihood estimators $\hat{\theta}$ and $\hat{\Psi}$.

In simple cases we can find an explicit expression for the likelihood function, and even find explicit estimating equations for the maximum likelihood estimators (see Example 1). However, in general it is not possible to find an explicit solution for the integral, and thus exact maximum likelihood estimators are unavailable, i.e. when: (i) $p_X(x_j^i, \cdot|x_{j-1}^i, \cdot)$ is known but we are unable to analytically solve the integral, and (ii) $p_X(x_j^i, \cdot|x_{j-1}^i, \cdot)$ is unknown. In (i) we have to numerically evaluate the integral to obtain an approximation of the likelihood (3) and then, by maximizing the resulting expression, approximate maximum likelihood estimators are obtained. In (ii) we can approximate $p_X(x_j^i, \cdot|x_{j-1}^i, \cdot)$, then numerically solve the integral in (3) and get the corresponding approximated maximum likelihood estimators.

In situation (ii) there exist several strategies to approximate the density $p_X(x_j^i, \cdot|x_{j-1}^i, \cdot)$, e.g. by simulating a large number of process sample paths (e.g. Pedersen (1995), Brandt and Santa-Clara (2002), Nicolau (2002), Hurn and Lindsay (1999)), or by solving numerically the Kolmogorov partial differential equations satisfied by the transition density (Lo (1988)). However, these techniques are computationally expensive. We propose to approximate the transition density as suggested in Aït-Sahalia (2001, 2002), where the approximation is obtained in closed-form, using a Hermite expansion as reviewed in section 4.1. Then, using this

expression, the likelihood function is approximated, thus deriving approximated maximum likelihood estimators of θ and Ψ, as suggested in section 4.2.

4 Closed-form transition density expansion and likelihood approximation

4.1 Transition density expansion for SDE models

Here we review the transition density expansion of a scalar ($d = L = 1$) time-homogeneous SDE as suggested in Aït-Sahalia (2002). A generalization to multidimensional SDEs can be found in Aït-Sahalia (2001), and we adapt the more compact notation used in the multidimensional case to the one-dimensional situation. The extension to time-inhomogeneous processes (i.e. the SDE depends directly on t, not only through the state variable X_t) is given in Egorov *et al.* (2003). In the remaining of this section we drop the reference to θ when not necessary, that is, we write $f(x)$ instead of $f(x, \theta)$ for a given function f.

Consider the following 1-dimensional time-homogeneous SDE

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \qquad X(t_0) = x_0 \tag{5}$$

where we want to approximate $p_X(x_j, \Delta_j|x_{j-1})$, the conditional density of X_{t_j} given $X_{j-1} = x_{j-1}$, where $\Delta_j = t_j - t_{j-1}$. Under mild regularity conditions (Aït-Sahalia (2001)) the logarithm of the transition density can be expanded in closed form using an order $J = +\infty$ Hermite series, and approximated by a Taylor expansion up to order K:

$$\begin{aligned} \ln p_X^{(K)}(x_j, \Delta_j|x_{j-1}) &= -\frac{1}{2}\ln(2\pi\Delta_j) - \frac{1}{2}\ln(\sigma^2(x_j)) + \frac{C_Y^{(-1)}(\gamma(x_j)|\gamma(x_{j-1}))}{\Delta_j} \\ &+ \sum_{k=0}^{K} C_Y^{(k)}(\gamma(x_j)|\gamma(x_{j-1}))\frac{\Delta_j^k}{k!}. \end{aligned} \tag{6}$$

The coefficients $C_Y^{(k)}$ are given in the appendix and $\gamma(\cdot)$ is the Lamperti transform, defined by

$$Y_t \equiv \gamma(X_t) = \int^{X_t} \frac{du}{\sigma(u)} \tag{7}$$

where the lower bound of integration is an arbitrary point in the interior of E (i.e. the constant of integration is irrelevant). Then Y_t is the solution to the SDE

$$dY_t = \mu_Y(Y_t)dt + dW_t$$

where $\mu_Y(\cdot)$ is given by

$$\mu_Y(Y_t) = \frac{\mu(\gamma^{-1}(Y_t))}{\sigma(\gamma^{-1}(Y_t))} - \frac{1}{2}\frac{\partial\sigma}{\partial x}(\gamma^{-1}(Y_t)).$$

4.2 Likelihood approximation for SDME models

For scalar time-homogeneous SDME models, the coefficients $C_Y^{(k)}$ can be obtained in the same way as suggested in section 4.1, by considering (θ, b^i), Δ_j^i and (x_j^i, x_{j-1}^i) instead of θ,

Δ_j and (x_j, x_{j-1}), respectively. Once the coefficients are obtained, an expression for $p_X^{(K)}$ is available and it is possible to approximate the likelihood of (θ, Ψ) for the SDME model (1) by substituting the unknown transition density in (4) with its approximation, thus obtaining a sequence of approximations to the likelihood function

$$L^{(K)}(\theta, \Psi) = \prod_{i=1}^{M} \int_B p_{\underline{X}}^{(K)}(\underline{x}^i | b^i, \theta)\, p_B(b^i | \Psi)\, db^i, \tag{8}$$

where

$$p_{\underline{X}}^{(K)}(\underline{x}^i | b^i, \theta) \;=\; \prod_{j=1}^{n_i} p_X^{(K)}(x_j^i, \Delta_j^i | x_{j-1}^i, b^i, \theta) \tag{9}$$

and $p_X^{(K)}$ is given by equation (6). By maximizing (8) with respect to (θ, Ψ), we obtain the corresponding approximated maximum likelihood estimators $\theta^{(K)}$ and $\Psi^{(K)}$.

The method can be extended to time-inhomogeneous and/or multidimensional SDME models, by extensions of the density expansion method, which are given by Aït-Sahalia (2001) for the multidimensional time-homogeneous case, and by Egorov *et al.* (2003) for the one-dimensional time-inhomogeneous case.

5 Implementation issues and numerical applications

This section reports applications of our estimation method to some famous SDE models that we perturb with random effects: Brownian motion with drift, the Geometric Brownian Motion and the Ornstein-Uhlenbeck process. The main goals are to show the feasibility and effectiveness of the proposed estimation method for SDME models, and to show that accurate results can be obtained when using a "reasonable" and "realistic" data-set, i.e. when handling a limited amount of data (say $M = 10, ..., 50$ subjects and $n = 10, ..., 50$ observations collected on each subject), instead of considering large data-sets that are often unavailable, especially in biomedical applications.

For numerical optimization reasons, the approximated estimators are always obtained by minimizing the negative log-likelihood function, e.g. when using the density expansion method we minimize

$$-\log L^{(K)}(\theta, \Psi) = -\sum_{i=1}^{M} \log \int_B p_{\underline{X}}^{(K)}(\underline{x}^i | b^i, \theta)\, p_B(b^i | \Psi)\, db^i, \tag{10}$$

and we denote with $(\theta^{(K)}, \Psi^{(K)})$ the resulting estimator

$$(\theta^{(K)}, \Psi^{(K)}) = \arg\min_{\theta, \Psi}(-\log L^{(K)}(\theta, \Psi)).$$

It has been shown that $K = 1$ or 2 (Aït-Sahalia, 1999, 2001, 2002) is often sufficient to approximate the transition density to obtain accurate estimates. We use either $K = 1$ or 2 order density expansion depending on the model, which seems to be sufficient for the considered applications (in particular, for the Brownian motion with drift and the Geometric Brownian Motion, a $K = 1$ order expansion gives the *exact* density expression). The coefficients $C_Y^{(k)}$ for the considered models are given in appendix (notice that, in general, the $C_Y^{(k)}$ can be

calculated using a symbolic calculus software). The integral appearing in (10) is numerically evaluated using the trapezoidal rule on a grid of two-hundred b^i values, except for Example 3 where B is assumed to be a square grid of 100×100 values.

For each example, parametric bootstrap was performed (Efron and Tibshirani (1993)) to obtain means of the parameters estimates and their 95% confidence intervals. More specifically, for each SDME model two hundred data sets, of dimensions $n \times M$ each, were generated using different sets of parameters and different values of M and n, and the corresponding (exact and/or approximated) parameter estimates were obtained. For each parameter, the sample mean and the empirical 95% confidence intervals (from the 2.5th to the 97.5th percentile) from the obtained estimates are reported in Table 1–5 together with measures of symmetry (skewness and kurtosis). In Example 1 the exact expression for the log-likelihood function of the SDME model is available, so we can graphically compare the shape of the surface of a chosen profile-loglikelihood with the corresponding surface obtained by numerical integration, as given in Figure 1. The same comparison is conducted in Example 2, but here only the transition density is known whereas the exact expression of the log-likelihood is unavailable in closed form, so we compare the surface of a profile of the logarithm of expression (19) (where the integral is numerically evaluated) with the surface of its $K = 2$ order approximation, as given in Figure 2.

Finally, we want to stress the usefulness of using the method considered in section 4.1 to approximate p_X. In fact, using e.g. the SML-like approaches (see the Introduction), for each iteration of an optimization algorithm maximizing (3), the numerical simulation of thousands of trajectories of the process can be required to approximate $p_X(x_j^i, \Delta_j^i | x_{j-1}^i, \cdot)$. Then, expression (4) must be evaluated and the integral in (3) must be numerically computed for the given subject. Finally, repeating the procedure for all the M subjects, we get the likelihood approximation for the current iteration of the optimization algorithm. From a computational point of view this is a highly expensive procedure, essentially because of the necessary large number of simulations of trajectories. Worse still, the larger the dimensions of θ and Ψ the slower the optimization procedure convergence, and obviously the computational time increases for large values of M and n. Instead, using the closed-form density expansion, simulating process trajectories is not required, and the likelihood approximation (8) can be evaluated more rapidly (Jensen and Poulsen (2002)).

Example 1: Brownian Motion with drift and Geometric Brownian Motion with one random effect

Consider a Brownian motion with drift:

$$dZ_t = (\beta - \sigma^2/2)dt + \sigma dW_t, \qquad Z_0 = z_0,$$

with solution

$$Z_t = Z_0 + (\beta - \sigma^2/2)t + \sigma W_t. \tag{11}$$

Assume an experiment is conducted on M different subjects. We are interested in estimating the parameters in the population, but expect individual differences in the processes, and would therefore consider a random effect in β, which leads to the SDME model:

$$dZ_t^i = (\beta + \beta^i - \sigma^2/2)dt + \sigma dW_t^i, \qquad Z_0^i = z_0^i, \quad i = 1, ..., M$$

and we assume $\beta^i \sim N(0, \sigma_\beta^2)$. The latter model has solution given by

$$Z_t^i = Z_0^i + (\beta + \beta^i - \sigma^2/2)t + \sigma W_t^i, \quad i = 1, ..., M. \tag{12}$$

In this simple example we have $b^i = \beta^i$, $\theta = (\beta, \sigma^2)$ and $\Psi = \sigma_\beta^2$. We wish to estimate $(\beta, \sigma^2, \sigma_\beta^2)$ given a set $\underline{z} = (\underline{z}^1, ..., \underline{z}^M)$ of observations from model (12).

The log-likelihood function is (Ditlevsen and De Gaetano (2005))

$$\begin{aligned} \log L(\theta, \Psi) &= \frac{M}{2}\log\left(\frac{\sigma^2}{\sigma_\beta^2}\right) - \frac{N-M}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{M}\log\left((\Delta^i)^{n_i}\left(T^i + \frac{\sigma^2}{\sigma_\beta^2}\right)\right) \\ &\quad - \frac{\sum_{i,j}\frac{1}{\Delta_j^i}(y_j^i - y_{j-1}^i - \alpha\Delta_j^i)^2 - \sum_i (y_{n_i}^i - y_0^i - \alpha T^i)^2\left(T^i + \frac{\sigma^2}{\sigma_\beta^2}\right)^{-1}}{2\sigma^2} \end{aligned} \tag{13}$$

where, for ease of notation, we define $\alpha = \beta - \sigma^2/2$, $\Delta^i = \left(\prod_{j=1}^{n_i}\Delta_j^i\right)^{\frac{1}{n_i}}$ and $T^i = \sum_{j=1}^{n_i}\Delta_j^i$. The last sum is simply the length of the observation interval for the i'th subject.

Assume equidistant observations and that each subject has the same number of observations, that is, assume $\Delta_j^i = \Delta$ and $n_i = n$ for all $1 \le i \le M$, $1 \le j \le n_i$. The maximum likelihood estimators are given by (Ditlevsen and De Gaetano (2005)):

$$\hat{\sigma}^2 = \frac{\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{n}(z_j^i - z_{j-1}^i - \hat{\alpha}\Delta)^2 - \frac{\Delta}{MT}\sum_{i=1}^{M}(z_n^i - z_0^i - \hat{\alpha}T)^2}{T - \Delta} \tag{14}$$

$$\hat{\sigma}_\beta^2 = \frac{\frac{1}{MT}\left[\sum_{i=1}^{M}(z_n^i - z_0^i - \hat{\alpha}T)^2 - \sum_{i=1}^{M}\sum_{j=1}^{n}(z_j^i - z_{j-1}^i - \hat{\alpha}\Delta)^2\right]}{T - \Delta} \tag{15}$$

$$\hat{\beta} = \hat{\alpha} + \frac{\hat{\sigma}^2}{2} \tag{16}$$

where $\hat{\alpha} = \sum_{i=1}^{M}(z_n^i - z_0^i)/(MT)$ and $T = T^i = n\Delta$.

Now consider the transformed process $X_t = \exp(Z_t)$, which leads to a SDME model of the Geometric Brownian motion

$$dX_t^i = (\beta + \beta^i)X_t^i dt + \sigma X_t^i dW_t^i, \qquad X_0^i = x_0^i, \quad i = 1, ..., M$$

with $\beta^i \sim N(0, \sigma_\beta^2)$ and Itô solution

$$X_t^i = X_0^i \exp((\beta + \beta^i - \sigma^2/2)t + \sigma W_t^i), \qquad i = 1, ..., M. \tag{17}$$

The process is relevant e.g. in pharmacokinetics for the metabolism of a compound in plasma following first order kinetics where we expect $\beta < 0$, or as a growth model, e.g. the initial growth of bacterial or tumor cell populations, where we expect $\beta > 0$. See e.g. Braumann (2002) for generalizations of this model.

The exact estimators (14)–(16) can be used as a benchmark to test the effectiveness of the estimation method. In this example $C_Y^{(k)}(\cdot) = 0$ for all $k \ge 2$, and thus the order $K = 1$ density expansion results in the *exact* transition density of the process, see the appendix for details. We therefore compare the exact maximum likelihood estimators with the approximated estimators, the only difference being that the integral in (3) is solved analytically or numerically.

For different sets of parameter values and for different choices of M and n, 200 data sets were generated from (12) and the parameters were estimated using (14)–(16) (see Table 1). Then, 200 data sets were generated from (17) and the approximated estimators were obtained by minimization of the numerical solution of (10); results from the latter approach are reported in Table 2. In all simulations we fixed $Z_0^i = \log(100)$, $X_0^i = 100$ for all i and $T = 100$. From Table 1 it is seen that the true parameter values are well identified using the exact maximum likelihood estimators (14)-(16) and, in particular, the cases $(M, n) = (50, 10)$ produce better estimates of β and σ_β than the cases $(M, n) = (10, 50)$ as expected, since M is the sample size of draws from the distribution of β. The same apply to the approximated estimates in Table 2, but here the cases $(M, n) = (10, 50)$ produce much worse estimates whereas cases $(M, n) = (50, 10)$ produce estimates comparable in quality to the exact ones. Obviously, the approximated estimators suffer the bias induced by the numerical integration in expression (10), and a finer integration grid (see section 5) should improve the performance of the method at the cost of increasing computational time. In all cases σ is well determined and does not seem affected by the numerical integration.

Finally, in Figure 1 the contour plots of the shapes of the profiles of the exact log-likelihoods (13) (for fixed $\sigma_\beta^2 = 0.02$) and the shapes of the corresponding approximations are compared for different values of M and n: the exact log-likelihood is conditioned on observations generated from model (12) with $(\beta, \sigma^2, \sigma_\beta^2) = (-0.2, 0.2, 0.02)$, whereas the approximated log-likelihood is conditioned on observations generated from model (17) with the same parameter values. By looking at Figure 1, we see that the exact and approximated surfaces are quite similar, and the approximation improves for increasing values of M. Differences in contour values are imputable to the models used to generate observations: model (12) for the exact log-likelihood and model (17) for the corresponding approximation. This implies proportional surfaces with similar shapes: differences in the shape are due to the numerical evaluation of the integral in (3).

Example 2: Ornstein-Uhlenbeck process with one random effect

Consider the Ornstein-Uhlenbeck process, defined by the following scalar SDE ($d = L = 1$)

$$dX_t = \left(-\frac{X_t}{\tau} + \mu\right)dt + \sigma dW_t; \qquad X_0 = x_0 = 0$$

where $\mu \in \mathbb{R}$, $\tau > 0$ and $\sigma > 0$ (see later for a different parametrization). This model is the simplest mean-reverting SDE, and has been widely used e.g. in neuronal modeling, biology, physics, engineering and finance, see e.g. Ditlevsen *et al.* (2005).

Consider the following SDME model

$$dX_t^i = \left(-\frac{X_t^i}{\tau} + \mu + \mu^i\right)dt + \sigma dW_t^i; \qquad X_0^i = x_0^i = 0, \quad i = 1, ..., M \tag{18}$$

and assume $\mu^i \sim N(0, \sigma_\mu^2)$. Here $b^i = \mu^i$ and we want to estimate $\theta = (\mu, \tau, \sigma)$ and $\Psi = \sigma_\mu^2$ given a set of observations $\underline{x}$ from model (18).

The conditional mean and variance of the X_t^i process are

$$\begin{aligned} \mathbb{E}(X_t^i | X_0^i = x_0, \mu, \tau, \sigma, \mu^i) &= x_0 e^{-t/\tau} + (\mu + \mu^i)\tau(1 - e^{-t/\tau}) \\ Var(X_t^i | X_0^i = x_0, \mu, \tau, \sigma, \mu^i) &= \frac{\sigma^2 \tau}{2}(1 - e^{-2t/\tau}) \end{aligned}$$

and the transition density is normal and given by

$$\begin{aligned} p_X(x_j^i, \Delta_j^i | x_{j-1}^i, \mu, \tau, \sigma, \mu^i) &= \left\{ \pi\sigma^2\tau\big(1 - e^{(-2\Delta_j^i/\tau)}\big) \right\}^{-1/2} \\ &\times \exp\left(- \frac{\big(x_j^i - x_{j-1}^i e^{-\Delta_j^i/\tau} - (\mu + \mu^i)\tau(1 - e^{-\Delta_j^i/\tau})\big)^2}{\sigma^2\tau(1 - e^{-2\Delta_j^i/\tau})} \right). \end{aligned}$$

Thus, the likelihood of (θ, Ψ) is given by

$$\begin{aligned} L(\theta, \Psi) &= \prod_{i=1}^{M} \Bigg\{ (\pi\sigma^2\tau)^{-n_i/2} \prod_{j=1}^{n_i} (1 - e^{-2\Delta_j^i/\tau})^{-1/2} (2\pi\sigma_\mu^2)^{-1/2} \\ &\times \int_{\mathbb{R}} \exp\Bigg\{ \sum_{j=1}^{n_i} \Bigg[- \frac{\big(x_j^i - x_{j-1}^i e^{-\Delta_j^i/\tau} - (\mu + \mu^i)\tau(1 - e^{-\Delta_j^i/\tau})\big)^2}{\sigma^2\tau(1 - e^{-2\Delta_j^i/\tau})} \Bigg] \\ &- \frac{(\mu^i)^2}{2\sigma_\mu^2} \Bigg\} d\mu^i \Bigg\}. \end{aligned} \tag{19}$$

We have no closed-form solution to the integral in (19), so exact estimators of θ and Ψ are unavailable. We first consider a numerical integration approach, and the resulting estimators are denoted with $(\tilde{\theta}, \tilde{\Psi}) = \arg\min_{\theta,\Psi}(-\log L(\theta, \Psi))$. As a second attempt, we ignore the fact that the exact transition density expression is already available, and we compute the approximated estimator $(\theta^{(K)}, \Psi^{(K)})$ by approximating in closed-form the transition density of model (18) with $K = 2$. The estimation results, obtained on 200 artificial data sets generated by (18) using the Euler-Maruyama scheme with integration stepsize of 0.01 (Kloeden and Platen (1992)), are reported in Table 3 and Table 4 for the first and the second estimation approach respectively. For both the strategies we fixed $n_i = n$ for all i and $T = 100$. From Tables 3 and 4 we notice that the true parameter values are correctly identified using both the likelihood (19) and the corresponding order $K = 2$ approximation, though in the second approach we notice that n should be larger than 10 in order to get satisfactory results.

Surface shapes of the log-likelihood profiles (for fixed $(\sigma, \sigma_\mu^2) = (1, 1)$) are reported in Figure 2, and compare the numerical evaluation of the logarithm of expression (19) with the corresponding order $K = 2$ expansion, both conditioned on observations generated from model (18) with $(\mu, \tau, \sigma, \sigma_\mu^2) = (1, 10, 1, 1)$. The comparison is satisfactory and, as suggested above, values of n larger than 10 produce better results.

Many readers will be more familiar with a different parametrization of the Ornstein-Uhlenbeck process, i.e.

$$dX_t = -\beta(X_t - \alpha)dt + \sigma dW_t; \qquad X_0 = x_0 = 0$$

with $(\alpha, \beta, \sigma) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+$. The relations between μ, τ, α and β are obviously given by $\tau = 1/\beta$ and $\mu = \alpha\beta$. If we consider the following SDME model

$$dX_t^i = -\beta(X_t^i - \alpha - \alpha^i)dt + \sigma dW_t^i; \qquad X_0 = x_0 = 0, \quad i = 1, ..., M$$

with $\alpha^i \sim N(0, \sigma_\mu^2/\beta^2)$, we have $\mu^i = \alpha^i\beta$. Thus, it is straightforward to obtain the coefficients of the transition density expansion with respect to this parametrization, by substituting $\tau = 1/\beta$, $\mu = \alpha\beta$ and $\mu^i = \alpha^i\beta$ into the expressions given in appendix.

Example 3: Ornstein-Uhlenbeck process with two random effects

Reconsider the Ornstein-Uhlenbeck model with both μ and τ perturbed by random effects μ^i and τ^i, respectively. The following SDME model results

$$dX_t^i = \left(-\frac{X_t^i}{\tau+\tau^i} + \mu + \mu^i\right)dt + \sigma dW_t^i; \qquad X_0^i = x_0^i = 0, \quad i = 1, ..., M \tag{20}$$

where $\mu^i \sim N(0, \sigma_\mu^2)$ and τ^i has exponential pdf with parameter $\lambda > 0$. The latter distribution is chosen to ensure that $\tau + \tau^i > 0$. Assume μ^i and $\tau^{i'}$ independent for any $i, i' = 1, ..., M$. Here $b^i = (\mu^i, \tau^i)$ and we want to estimate $\theta = (\mu, \tau, \sigma)$ and $\Psi = (\sigma_\mu^2, \lambda)$ given a set of observations $\underline{x}$ from model (20).

Now we only consider the estimation approach based on the transition density expansion, i.e. we optimize (10), where

$$\int_B p_{\underline{X}}^{(K)}(\underline{x}^i|b^i, \theta) p_B(b^i|\Psi) db^i = \int_{-\infty}^{+\infty} \int_0^{+\infty} p_{\underline{X}}^{(K)}(\underline{x}^i|\mu^i, \tau^i, \theta) p(\mu^i|\sigma_\mu^2) p(\tau^i|\lambda) d\mu^i d\tau^i,$$

and

$$p(\mu^i|\sigma_\mu^2) = \frac{\exp(-(\mu^i)^2/(2\sigma_\mu^2))}{\sigma_\mu\sqrt{2\pi}}, \qquad p(\tau^i|\lambda) = \lambda \exp(-\lambda\tau^i).$$

The estimation results with $K = 2$, obtained on 200 artificial data sets generated by (20) using the Euler-Maruyama scheme with stepsize 0.01, are reported in Table 5. We fixed $n_i = n$ for all i and $T = 100$. Also in this example estimates are satisfactory, especially for increasing n values; only the true λ value is not well identified in the case $(\mu, \tau, \sigma, \sigma_\mu^2, \lambda) = (2, 12, 3, 0.5, 6)$. That is natural considering the large variance of the exponential distribution. However, it has to be noticed that, for ease of computations in the bootstrap procedure, a coarse grid has been chosen for the numerical integration of the likelihood (see section 5), so it is likely that better results can be achieved using a finer grid.

6 Conclusions

In the present work an approximated maximum likelihood estimator for the parameters of stochastic differential mixed-effects models has been proposed. SDE models incorporating random effects have been considered in few recent works (Overgaard *et al.* (2005); Tornøe *et al.* (2005); Ditlevsen and De Gaetano (2005)) focused on models with constant diffusion and normal or log-normal distributed random effects. The proposed estimation method can be applied to models having non-constant and non-linear diffusion term, with random effects following any continuous distribution and can be extended to multidimensional SDMEs. The method is based on the construction of a sequence of approximations $L^{(K)}$ to the true likelihood function L, which is obtained by expanding the process transition densities in closed-form to order K, thus obtaining an expression which can be rapidly evaluated. For SDME models more complex than the ones here considered, the likelihood approximation can be obtained by taking advantage of any software with symbolic calculus capabilities.

Simulation results for the considered models show that the estimates obtained by minimizing $-\log L^{(K)}$, with $K = 1$ or 2, are close to the true parameter values, and this result can be achieved using moderate values of M (the number of experimental units, e.g. the

number of subjects) and n (the number of observation for a given experimental unit). This is relevant for applications in situations where large data sets are unavailable, e.g. in biomedical applications, where Mixed-Effects theory is broadly applied.

The method suffers some limitations, e.g. it may be difficult (though theoretically possible, see Aït-Sahalia (2001)) to obtain the transition density expansion for some multidimensional SDME systems with *irreducible* or *non-commutative* noise (Kloeden and Platen (1992)). Moreover, it may be difficult to numerically evaluate the integral in (3) when the dimension of B increases, and efficient numerical algorithms are needed. Finally, the models used e.g. in biomedical applications are often more complicated than the simple examples illustrated here, and it is still needed to see the applicability of the method in more realistic settings.

In conclusion, we propose a parameter estimation method for SDE models incorporating random effects, which at least for the models considered here is reliable and effective and can be easily applied using commonly available computational resources. We believe that such a class of models will undergo increasing popularity, since it combines the nice features of the Mixed-Effects theory (total variation is split in within-subject and between-subject variation) with the possibility of considering random variability into the within-subject process dynamics, thus providing a very flexible modeling approach.

Acknowledgments

Work supported by grants from the Danish Medical Research Council and the Lundbeck Foundation to S. Ditlevsen.

References

Aït-Sahalia, Y. (1999). Transition densities for interest rates and other nonlinear diffusions. *The Journal Of Finance*, **54**(4), 1361–1395.

Aït-Sahalia, Y. (2001). Closed-form likelihood expansion for multivariate diffusions. Technical report, Princeton University.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, **70**(1), 223–262.

Brandt, M. W. and Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, **63**, 161–210.

Braumann, C. A. (2002). Variable effort harvesting models in random environments: generalization to density-dependent noise intensities. *Mathematical Biosciences*, **177-178**, 229–245.

Breslow, N. and Clayton, D. (1993). Approximate inference in generalized linear mixed models. *JASA*, **88**(421), 9–25.

Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press.

Ditlevsen, S. and De Gaetano, A. (2005). Mixed effects in stochastic differential equations models. *REVSTAT - Statistical Journal*, **3**(2), 137–153.

Ditlevsen, S., Yip, K.-P., and Holstein-Rathlou, N.-H. (2005). Parameter estimation in a stochastic model of the tubuloglomerular feedback mechanism in a rat nephron. *Math. Biosci.*, **194**, 49–69.

Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business and Economic Statistics*, **20**(3), 297–316.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall.

Egorov, A., Li, H., and Xu, Y. (2003). Maximum likelihood estimation of time-inhomogeneous diffusions. *Journal of Econometrics*, **114**, 107–139.

Hurn, A. and Lindsay, K. (1999). Estimating the parameters of stochastic differential equations. *Mathematics And Computers In Simulation*, **48**, 373–384.

Hurn, A., Lindsay, K., and Martin, V. (2003). On the efficacy of simulated maximum likelihood for estimating the parameters of stochastic differential equations. *Journal of Time Series Analysis*, **24**(1), 45–63.

Jelliffe, R., Schumitzky, A., and Guilder, M. V. (2000). Population pharmacokinetics/pharmacodynamics modeling: Parametric and nonparametric methods. *Therapeutic Drug Monitoring*, **22**, 354–365.

Jensen, B. and Poulsen, R. (2002). Transition densities of diffusion processes: numerical comparison of approximation techniques. *Journal of Derivatives*, **9**, 1–15.

Kloeden, P. E. and Platen, E. (1992). *Numerical solution of stochastic differential equations*. Springer.

Lindstrom, M. and Bates, D. (1990). Nonlinear mixed-effects models for repeated measures data. *Biometrics*, **46**, 673–687.

Lo, A. (1988). Maximum likelihood estimation of generalized Ito processes with discretely-sample data. *Econometric Theory*, **4**, 231–247.

Nicolau, J. (2002). A new technique for simulating the likelihood of stochastic differential equations. *Econometrics Journal*, **5**, 91–103.

Øksendal, B. (2000). *Stochastic differential equations: an introduction with applications*. Springer, second edition.

Overgaard, R., Jonsson, N., Tornøe, C., and Madsen, H. (2005). Non-linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm. *Journal of Pharmacokinetics and Pharmacodynamics*, **32**, 85–107.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, **22**(1), 55–71.

Pinheiro, J. and Bates, D. (2002). *Mixed-effects models in S and S-PLUS*. Springer-Verlag, New York.

Sheiner, L. and Beal, L. (1980). Evaluation of methods for estimationg population pharmacokinetic parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. *J. Pharmacokinet. Biopharm.*, **8**(6), 553–571.

Sheiner, L. and Beal, L. (1981). Evaluation of methods for estimationg population pharmacokinetic parameters. II. Biexponential model and experimental pharmacokinetic data. *J. Pharmacokinet. Biopharm.*, **9**(5), 635–651.

Tornøe, C., Overgaard, R., Agersø, H., Nielsen, H. A., Madsen, H., and Jonsson, E. N. (2005). Stochastic differential equations in NONMEM: implementation, application, and comparison with ordinary differential equations. *Pharmaceutical Research*, **22**(8), 1247–1258.

Vonesh, E. and Chinchilli, V. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. Marcel Dekker, New York.

Appendix

Here we report the explicit expressions for the coefficients of the log-density expansion as suggested in Aït-Sahalia (2001). Moreover, the coefficients are reported for both the Geometric Brownian Motion and the Ornstein-Uhlenbeck SDME models.

Density expansion coefficients

For given values y_j and y_{j-1} of the Y process (7) the coefficients of the log-density expansion (6) are given by

$$C_Y^{(-1)}(y_j|y_{j-1}) = -\frac{1}{2}(y_j - y_{j-1})^2$$

$$C_Y^{(0)}(y_j|y_{j-1}) = (y_j - y_{j-1}) \int_0^1 \mu_Y(y_{j-1} + u(y_j - y_{j-1}))du$$

and, for $k \geq 1$,

$$C_Y^{(k)}(y_j|y_{j-1}) = k \int_0^1 G_Y^{(k)}(y_{j-1} + u(y_j - y_{j-1})|y_{j-1})u^{k-1}du.$$

The functions $G_Y^{(k)}$ are given by

$$G_Y^{(1)}(y_j|y_{j-1}) = -\frac{\partial \mu_Y(y_j)}{\partial y_j} - \mu_Y(y_j)\frac{\partial C_Y^{(0)}(y_j|y_{j-1})}{\partial y_j} + \frac{1}{2}\frac{\partial^2 C_Y^{(0)}(y_j|y_{j-1})}{\partial y_j^2} + \frac{1}{2}\left(\frac{\partial C_Y^{(0)}(y_j|y_{j-1})}{\partial y_j}\right)^2$$

and for $k \geq 2$

$$\begin{aligned} G_Y^{(k)}(y_j|y_{j-1}) &= -\mu_Y(y_j)\frac{\partial C_Y^{(k-1)}(y_j|y_{j-1})}{\partial y_j} + \frac{1}{2}\frac{\partial^2 C_Y^{(k-1)}(y_j|y_{j-1})}{\partial y_j^2} \\ &\quad + \frac{1}{2}\sum_{h=0}^{k-1}\binom{k-1}{h}\frac{\partial C_Y^{(h)}(y_j|y_{j-1})}{\partial y_j}\frac{\partial C_Y^{(k-1-h)}(y_j|y_{j-1})}{\partial y_j}. \end{aligned}$$

Geometric Brownian Motion: order $K = 1$ density expansion coefficients

For model (17) we have:

$$Y_t = \gamma(X_t) = \frac{\log(X_t)}{\sigma},$$

then

$$\mu_Y(Y_t) = \frac{\beta + \beta^i}{\sigma} - \frac{\sigma}{2}$$

and for given values y_j^i and y_{j-1}^i of the Y process, we have

$$\begin{aligned}
C_Y^{(0)}(y_j^i|y_{j-1}^i) &= (y_j^i - y_{j-1}^i)\left(\frac{\beta + \beta^i}{\sigma} - \frac{\sigma}{2}\right) = \frac{\log(x_j^i) - \log(x_{j-1}^i)}{\sigma^2}\left(\beta + \beta^i - \frac{\sigma^2}{2}\right) \\
C_Y^{(1)}(y_j^i|y_{j-1}^i) &= -\frac{1}{2\sigma^2}\left(\beta + \beta^i - \frac{\sigma^2}{2}\right)^2 \\
C_Y^{(k)}(y_j^i|y_{j-1}^i) &= 0, \qquad k \geq 2
\end{aligned}$$

which yields the exact transition density

$$\begin{aligned}
p_X^{(1)}(x_j^i, \Delta_j^i|x_{j-1}^i) &= \frac{1}{x_j^i\sqrt{2\pi\sigma^2\Delta_j^i}} \exp\left(-\frac{\left(\log(x_j^i) - \log(x_{j-1}^i) - (\beta + \beta^i - \frac{\sigma^2}{2})\Delta_j^i\right)^2}{2\sigma^2\Delta_j^i}\right) \\
&= p_X(x_j^i, \Delta_j^i|x_{j-1}^i).
\end{aligned}$$

Ornstein-Uhlenbeck process with one random effect: order $K = 2$ density expansion coefficients

For model (18) we have:

$$Y_t = \gamma(X_t) = X_t/\sigma$$

then

$$\mu_Y(Y_t) = -Y_t/\tau + \rho,$$

where $\rho = (\mu + \mu^i)/\sigma$, and for given values y_j^i and y_{j-1}^i of the Y process, we have

$$\begin{aligned}
C_Y^{(0)}(y_j^i|y_{j-1}^i) &= (y_j^i - y_{j-1}^i)\left(\rho - \frac{y_j^i + y_{j-1}^i}{2\tau}\right) \\
C_Y^{(1)}(y_j^i|y_{j-1}^i) &= \frac{3\tau - (y_j^i)^2 - y_j^i y_{j-1}^i - (y_{j-1}^i)^2 + 3\rho\tau(y_j^i + y_{j-1}^i) - 3\rho^2\tau^2}{6\tau^2} \\
C_Y^{(2)}(y_j^i|y_{j-1}^i) &= -\frac{1}{6\tau^2}
\end{aligned}$$

and

$$\begin{aligned}
p_X^{(2)}(x_j^i, \Delta_j^i|x_{j-1}^i) &= \frac{1}{\sqrt{2\pi\sigma^2\Delta_j^i}} \exp\left(-\frac{(x_j^i - x_{j-1}^i)^2}{2\sigma^2\Delta_j^i} + \tilde{C}^{(0)}(x_j^i|x_{j-1}^i) + \tilde{C}^{(1)}(x_j^i|x_{j-1}^i)\Delta_j^i\right. \\
&+ \left.\frac{(\Delta_j^i)^2}{2}\tilde{C}^{(2)}(x_j^i|x_{j-1}^i)\right)
\end{aligned}$$

where $\tilde{C}^{(k)}(x_j^i|x_{j-1}^i) = C_Y^{(k)}(\frac{x_j^i}{\sigma}|\frac{x_{j-1}^i}{\sigma})$, $k = 0, 1, 2$.

Ornstein-Uhlenbeck process with two random effects: order $K = 2$ density expansion coefficients

For model (20) we have:

$$Y_t = \gamma(X_t) = X_t/\sigma$$

then

$$\mu_Y(Y_t) = -Y_t/(\tau + \tau^i) + \rho,$$

where $\rho = (\mu + \mu^i)/\sigma$, and for given values y_j^i and y_{j-1}^i of the Y process, we have

$$\begin{aligned}
C_Y^{(0)}(y_j^i|y_{j-1}^i) &= (y_j^i - y_{j-1}^i)\left(\rho - \frac{y_j^i + y_{j-1}^i}{2(\tau + \tau^i)}\right) \\
C_Y^{(1)}(y_j^i|y_{j-1}^i) &= \frac{3(\tau + \tau^i) - (y_j^i)^2 - y_j^i y_{j-1}^i - (y_{j-1}^i)^2 + 3\rho(\tau + \tau^i)(y_j^i + y_{j-1}^i) - 3\rho^2(\tau + \tau^i)^2}{6(\tau + \tau^i)^2} \\
C_Y^{(2)}(y_j^i|y_{j-1}^i) &= -\frac{1}{6(\tau + \tau^i)^2}
\end{aligned}$$

and

$$\begin{aligned}
p_X^{(2)}(x_j^i, \Delta_j^i|x_{j-1}^i) &= \frac{1}{\sqrt{2\pi\sigma^2\Delta_j^i}} \exp\left\{-\frac{(x_j^i - x_{j-1}^i)^2}{2\sigma^2\Delta_j^i} + \frac{(x_j^i - x_{j-1}^i)}{\sigma}\left(\rho - \frac{(x_j^i + x_{j-1}^i)}{2\sigma(\tau + \tau^i)}\right)\right. \\
&+ \left. \tilde{C}^{(1)}(x_j^i|x_{j-1}^i)\Delta_j^i + \frac{(\Delta_j^i)^2}{2}\tilde{C}^{(2)}(x_j^i|x_{j-1}^i)\right\}
\end{aligned}$$

where $\tilde{C}^{(k)}(x_j^i|x_{j-1}^i) = C_Y^{(k)}(\frac{x_j^i}{\sigma}|\frac{x_{j-1}^i}{\sigma})$, $k = 1, 2$.

Parameter values β	σ^2	σ^2_β		$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\sigma}^2_\beta$
				$M = 10$, $n = 50$		
-0.2	0.2	0.02	Mean [95% CI]	-0.203 [-0.291, -0.112]	0.201 [0.173, 0.222]	0.018 [0.005, 0.038]
			Skewness	0.065	-0.301	0.590
			Kurtosis	2.780	3.540	3.048
				$M = 50$, $n = 10$		
-0.2	0.2	0.02	Mean [95% CI]	-0.198 [-0.245, -0.152]	0.199 [0.171, 0.226]	0.019 [0.012, 0.029]
			Skewness	0.070	-0.025	0.394
			Kurtosis	2.890	2.630	2.717
				$M = 10$, $n = 50$		
-0.02	0.02	0.02	Mean [95% CI]	-0.023 [-0.103, 0.061]	0.020 [0.017, 0.022]	0.018 [0.006, 0.036]
			Skewness	0.093	-0.301	0.600
			Kurtosis	2.688	3.540	3.057
				$M = 50$, $n = 10$		
-0.02	0.02	0.02	Mean [95% CI]	-0.018 [-0.062, 0.022]	0.020 [0.017, 0.023]	0.019 [0.012, 0.028]
			Skewness	-0.051	-0.025	0.302
			Kurtosis	2.909	2.630	2.675

Table 1: Brownian Motion with drift: exact maximum likelihood estimates (and 95% empirical confidence intervals) from simulations of model (11).

Parameter values β	σ^2	σ^2_β		$\beta^{(1)}$	$(\sigma^{(1)})^2$	$(\sigma^{(1)}_\beta)^2$
				$M = 10$, $n = 50$		
-0.2	0.2	0.02	Mean [95% CI]	-0.135 [-0.197, -0.083]	0.200 [0.176, 0.221]	0.008 [0.001, 0.019]
			Skewness	-0.069	0.161	0.887
			Kurtosis	2.761	2.295	4.329
				$M = 50$, $n = 10$		
-0.2	0.2	0.02	Mean [95% CI]	-0.198 [-0.247, -0.158]	0.199 [0.171, 0.226]	0.019 [0.011, 0.028]
			Skewness	-0.209	-0.025	0.066
			Kurtosis	2.817	2.630	2.488
				$M = 10$, $n = 50$		
-0.02	0.02	0.02	Mean [95% CI]	-0.038 [-0.101, -10^{-4}]	0.020 [0.017, 0.022]	0.014 [0.005, 0.026]
			Skewness	-0.564	0.072	0.482
			Kurtosis	2.580	3.772	2.958
				$M = 50$, $n = 10$		
-0.02	0.02	0.02	Mean [95% CI]	-0.020 [-0.061, -10^{-4}]	0.020 [0.017, 0.023]	0.019 [0.012, 0.028]
			Skewness	-0.715	-0.023	0.164
			Kurtosis	2.938	2.632	2.539

Table 2: Geometric Brownian Motion: maximum likelihood estimates (and 95% empirical confidence intervals), from simulations of model (17), solving the integral numerically.

Parameter values								
μ	τ	σ	σ_μ^2		$\tilde{\mu}$	$\tilde{\tau}$	$\tilde{\sigma}$	$\tilde{\sigma}_\mu^2$
					$M = 10$, $n = 50$			
1	10	1	1	Mean [95% CI]	0.980 [0.380, 1.576]	10.084 [8.085, 12.082]	0.990 [0.919, 1.047]	0.915 [0.304, 1.935]
				Skewness	0.104	0.240	-0.268	0.982
				Kurtosis	2.594	3.025	3.694	4.257
					$M = 50$, $n = 10$			
1	10	1	1	Mean [95% CI]	1.019 [0.693, 1.317]	9.943 [8.852, 10.949]	0.947 [0.875, 1.022]	0.991 [0.553, 1.471]
				Skewness	-0.117	-0.105	-0.047	0.395
				Kurtosis	2.802	2.760	2.795	3.423
					$M = 10$, $n = 50$			
2	12	0.1	0.25	Mean [95% CI]	2.021 [1.783, 2.213]	11.996 [11.840, 12.168]	0.099 [0.091, 0.104]	0.240 [0.076, 0.444]
				Skewness	-0.385	-0.213	-0.395	0.698
				Kurtosis	4.896	3.534	3.665	4.456
					$M = 50$, $n = 10$			
2	12	0.1	0.25	Mean [95% CI]	2.018 [1.869, 2.212]	11.995 [11.917, 12.068]	0.094 [0.088, 1.100]	0.251 [0.154, 0.370]
				Skewness	-0.536	-0.011	-0.065	0.268
				Kurtosis	3.570	2.205	2.902	2.871

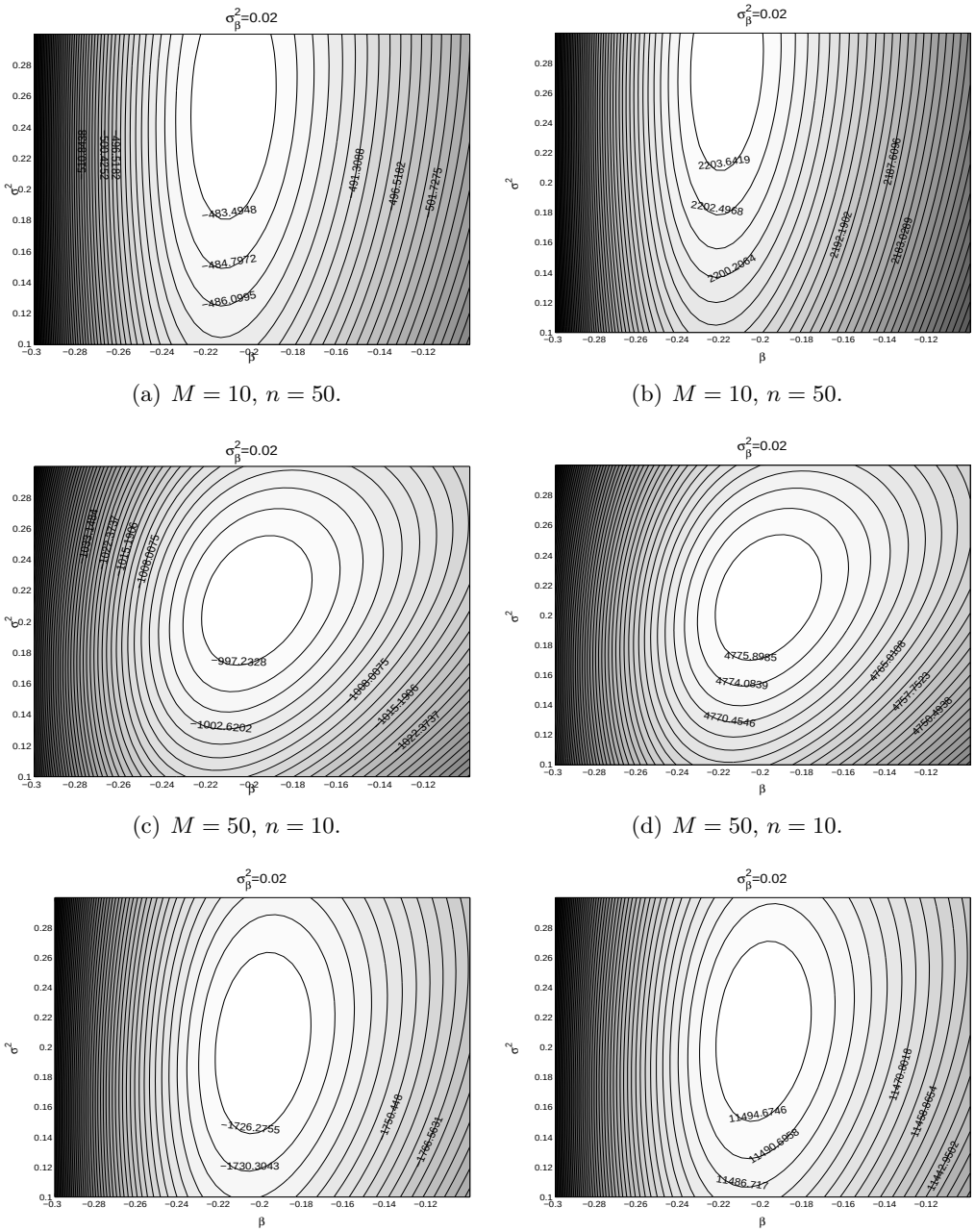
Table 3: Ornstein-Uhlenbeck process: approximated maximum likelihood estimates (and 95% empirical confidence intervals) from simulations of model (18), using the exact transition density.

Parameter values								
μ	τ	σ	σ_μ^2		$\mu^{(2)}$	$\tau^{(2)}$	$\sigma^{(2)}$	$(\sigma_\mu^{(2)})^2$
					$M = 10$, $n = 50$			
1	10	1	1	Mean [95% CI]	0.972 [0.377, 1.562]	10.182 [8.196, 12.174]	1.000 [0.928, 1.057]	0.898 [0.298, 1.898]
				Skewness	0.105	0.250	-0.268	0.972
				Kurtosis	2.597	3.033	3.680	4.214
					$M = 50$, $n = 10$			
1	10	1	1	Mean [95% CI]	0.866 [0.585, 1.130]	11.820 [10.887, 12.727]	0.994 [0.920, 1.058]	0.711 [0.403, 1.041]
				Skewness	-0.101	0.052	-0.160	0.309
				Kurtosis	2.848	2.736	2.742	3.026
					$M = 50$, $n = 50$			
1	10	1	1	Mean [95% CI]	1.006 [0.691, 1.283]	10.077 [9.221, 10.848]	1.000 [0.970, 1.028]	0.962 [0.556, 1.362]
				Skewness	-0.084	0.042	-0.112	0.327
				Kurtosis	2.972	2.789	2.832	2.963
					$M = 10$, $n = 50$			
2	12	0.1	0.25	Mean [95% CI]	2.017 [1.761, 2.248]	12.058 [11.901, 12.230]	0.106 [0.099, 0.111]	0.239 [0.075, 0.475]
				Skewness	-0.180	-0.240	-0.222	0.618
				Kurtosis	3.729	3.529	3.012	3.039
					$M = 50$, $n = 10$			
2	12	0.1	0.25	Mean [95% CI]	1.813 [1.671, 1.945]	13.366 [13.294, 13.433]	0.375 [0.349, 0.400]	0.197 [0.123, 0.284]
				Skewness	-0.076	-0.012	-0.031	0.226
				Kurtosis	2.904	2.179	2.838	2.626
					$M = 50$, $n = 50$			
2	12	0.1	0.25	Mean [95% CI]	2.010 [1.854, 2.134]	12.057 [11.977, 12.133]	0.106 [0.103, 0.108]	0.245 [0.151, 0.344]
				Skewness	-0.573	0.008	-0.172	0.042
				Kurtosis	3.272	2.361	2.923	2.396

Table 4: Ornstein-Uhlenbeck process: approximated maximum likelihood estimates (and 95% empirical confidence intervals), from simulations of model (18), using an order $K = 2$ density expansion.

Parameter values										
μ	τ	σ	σ_μ^2	λ		$\mu^{(2)}$	$\tau^{(2)}$	$\sigma^{(2)}$	$(\sigma_\mu^{(2)})^2$	$\lambda^{(2)}$
						$M = 10$, $n = 50$				
5	10	1	1	1	Mean [95% CI]	4.944 [4.336, 5.617]	10.029 [8.448, 11.638]	1.002 [0.931, 1.053]	1.101 [0.261, 2.230]	1.165 [$1.983 \cdot 10^{-6}$, 2.507]
					Skewness	0.196	0.234	-0.366	0.427	0.168
					Kurtosis	2.754	2.999	3.454	2.798	2.732
						$M = 50$, $n = 10$				
5	10	1	1	1	Mean [95% CI]	4.427 [4.132, 4.735]	11.655 [10.939, 12.683]	1.336 [1.266, 1.398]	1.131 [0.711, 1.552]	1.019 [0.001, 1.630]
					Skewness	-10^{-4}	0.772	-0.135	0.152	-0.596
					Kurtosis	2.926	4.251	2.594	2.631	3.318
						$M = 50$, $n = 50$				
5	10	1	1	1	Mean [95% CI]	4.973 [4.627, 5.309]	9.924 [9.317, 10.511]	1.001 [0.971, 1.029]	1.208 [0.736, 1.671]	1.261 [0.729, 1.975]
					Skewness	-0.084	-0.090	-0.112	0.119	0.472
					Kurtosis	3.012	3.362	2.803	2.859	3.279
						$M = 10$, $n = 50$				
2	12	3	0.5	6	Mean [95% CI]	2.027 [1.363, 2.820]	11.330 [5.086, 25.388]	3.003 [2.786, 3.182]	0.583 [0.001, 1.510]	8.509 [$5.276 \cdot 10^{-6}$, 19.897]
					Skewness	0.284	1.441	-0.281	0.665	0.071
					Kurtosis	2.809	5.114	3.642	3.309	2.195
						$M = 50$, $n = 10$				
2	12	3	0.5	6	Mean [95% CI]	1.904 [1.633, 2.162]	11.520 [8.555, 16.126]	2.960 [2.748, 3.170]	0.640 [0.317, 1.061]	9.646 [4.682, 14.757]
					Skewness	-0.108	1.049	-0.162	0.388	-0.204
					Kurtosis	2.519	4.995	2.454	2.982	3.848
						$M = 50$, $n = 50$				
2	12	3	0.5	6	Mean [95% CI]	2.061 [1.728, 2.373]	9.841 [7.123, 14.127]	2.999 [2.905, 3.088]	0.686 [0.330, 1.058]	9.510 [5.939, 14.336]
					Skewness	0.047	0.621	-0.101	0.152	0.354
					Kurtosis	2.594	3.293	2.838	2.650	3.050

Table 5: Ornstein-Uhlenbeck process: approximated maximum likelihood estimates (and 95% empirical confidence intervals), from simulations of model (20), using an order $K = 2$ density expansion.



(a) $M = 10$, $n = 50$.

(b) $M = 10$, $n = 50$.

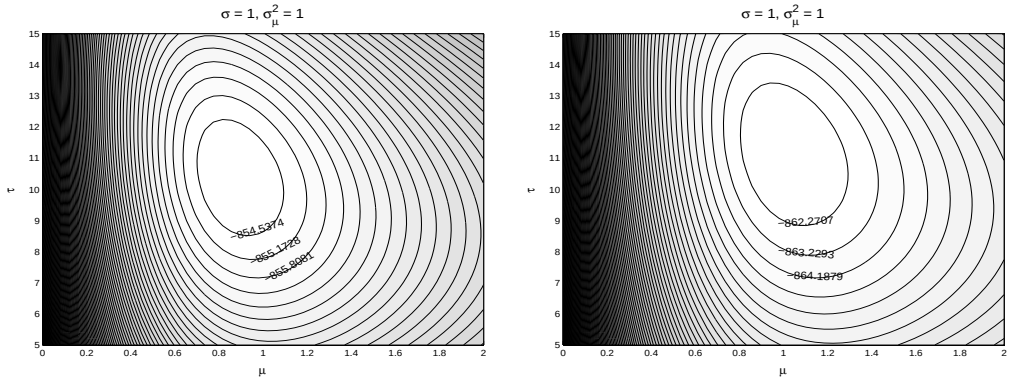
(c) $M = 50$, $n = 10$.

(d) $M = 50$, $n = 10$.

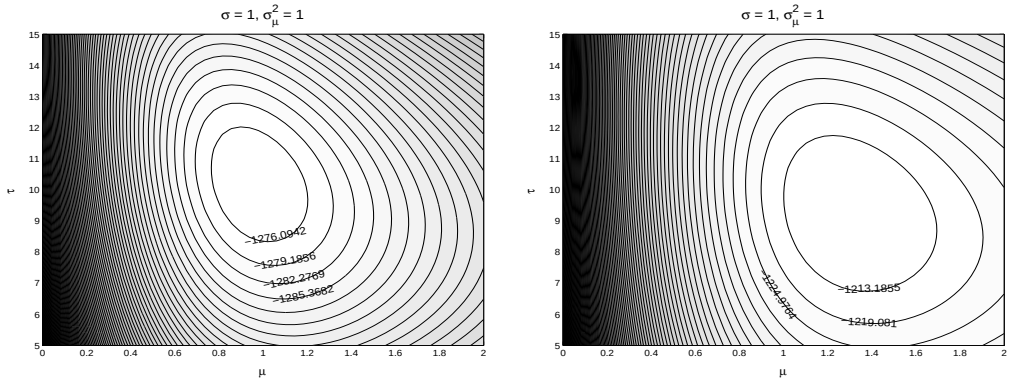
(e) $M = 50$, $n = 26$.

(f) $M = 50$, $n = 26$.

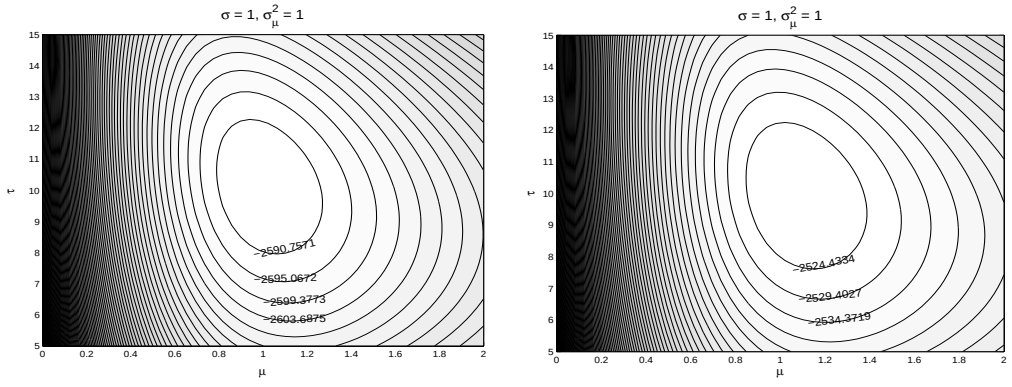
Figure 1: Example 1 - Contour plots of the exact log-likelihood profiles (left panels) and the corresponding approximations by numerical integration (right panels) for fixed $\sigma_\beta^2 = 0.02$, given observations generated from model (12) (left) and (17) (right) with $(\beta, \sigma^2, \sigma_\beta^2) = (-0.2, 0.2, 0.02)$.



(a) $M = 10$, $n = 50$. (b) $M = 10$, $n = 50$.



(c) $M = 50$, $n = 10$. (d) $M = 50$, $n = 10$.



(e) $M = 50$, $n = 26$. (f) $M = 50$, $n = 26$.

Figure 2: Example 2 - Contour plots of the profiles of the numerically approximated log-likelihood (19) (left panels) and the corresponding order $K = 2$ approximations (right panels) for fixed $\sigma = 1$ and $\sigma_\mu^2 = 1$, given observations generated from model (18) with $(\mu, \tau, \sigma, \sigma_\mu^2) = (1, 10, 1, 1)$.