Stochastic Differential Mixed-Effects Models

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

[1] Department of Mathematical Sciences, University of Copenhagen, Denmark.
email: umberto@math.ku.dk, susanne@math.ku.dk
[2] Biomathematics Laboratory (BioMatLab) IASI - CNR, Rome, Italy;
email: andrea.degaetano@gmx.net

Published on *Scandinavian Journal of Statistics* 37(1), 67–90, 2010.
doi:10.1111/j.1467-9469.2009.00665.x

Abstract

Stochastic differential equations have shown useful to describe random continuous time processes. Biomedical experiments often imply repeated measurements on a series of experimental units and differences between units can be represented by incorporating random effects into the model. When both system noise and random effects are considered, *stochastic differential mixed-effects models* ensue. This class of models enables the simultaneous representation of randomness in the dynamics of the phenomena being considered and variability between experimental units, thus providing a powerful modeling tool with immediate applications in biomedicine and pharmacokinetic/pharmacodynamic studies. In most cases the likelihood function is not available, and thus maximum likelihood estimation of the unknown parameters is not possible. Here we propose a computationally fast approximated maximum likelihood procedure for the estimation of the non-random parameters and the random effects. The method is evaluated on simulations from some famous diffusion processes and on real datasets.

Keywords: biomedical applications; Brownian motion with drift; CIR process; closed-form transition density expansion; Gaussian quadrature; geometric Brownian motion; maximum likelihood estimation; Ornstein-Uhlenbeck process; random parameters; stochastic differential equations.

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 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 noise), both linear and non-linear (Lindstrom and Bates (1990), Breslow and Clayton (1993), Davidian and Giltinan (1995), Vonesh and Chinchilli (1997), McCulloch and Searle (2001), Diggle et al. (2002), Kuhn and Lavielle (2005), Guedj et al. (2007), Wang (2007)), and standard software for model fitting is available, e.g. Beal et al. (1999), Pinheiro and Bates (2002), the `R` package by Pinheiro et al. (2007), Lavielle et al. (2007) and the SAS `NLMIXED` procedure. 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, except for the references discussed below. The problem is that estimating parameters in SDE models is not straightforward, except for few 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 these authors regret that system noise is not considered since it is difficult to estimate. In Overgaard et al. (2005) and Tornøe et al. (2005) an SDE model with log-normally distributed random effects and a constant diffusion term is treated. In Ditlevsen and De Gaetano (2005a) the likelihood function for a simple SDE model with normally distributed random effects is calculated explicitly, but generally the likelihood function is unavailable. Recently Donnet and Samson (2008) developed an estimation method based on a stochastic EM algorithm for fitting SDE with mixed-effects. However, from a computational point of view, the proposed methods are time-consuming. Eventually, as SDE models are more commonly applied to biomedical data (e.g. Lansky et al. (2004); Andersen and Højbjerre (2005); Picchini et al. (2006); Ditlevsen and De Gaetano (2005b); Ditlevsen et al. (2007); Overgaard et al. (2007)), there will be an increasing need for developing a general theory for parameter estimation including mixed-effects.

In the present work a computationally efficient estimation method for the parameters of an SDE model incorporating random parameters is proposed: these models may be called *stochastic differential mixed-effects models* (SDMEMs). By using the proposed methodology on repeated measurements from different units (e.g. subjects) it is not necessary to fit the individual data separately, but a single estimation procedure is used to fit the overall data simultaneously. We consider SDMEMs whose drift and diffusion terms can depend linearly or nonlinearly on state variables and random effects following any sufficiently well-behaved continuous distribution (although discrete distributions can also be considered), 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 transition density. Various ways have been proposed to approximate the transition density: (i) solving numerically the Kolmogorov partial differential equations satisfied by the transition density (Lo (1988)); (*ii*) deriving a closed-form Hermite expansion to the transition density (Aït-Sahalia (2008, 2002b)); (*iii*) or 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 et al. (2003); Nicolau (2002)), and this is known as "simulated maximum likelihood" (SML). More recently a method using exact simulation has been proposed by Beskos et al. (2006). Each of these techniques have been successfully implemented by the aforementioned authors, but they also have limitations. Aït-Sahalia (2002a) notes that methods (*i*) and (*iii*) 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 (*ii*) above, while accurate and fast, may be difficult to apply.

We choose to employ the transition density approximation method suggested in Aït-Sahalia (2002b, 2008) for time-homogeneous SDEs, since it is fast and accurate among the available methods (Durham and Gallant (2002), Jensen and Poulsen (2002)). Attention is restricted to time-homogeneous SDEs and the generalization to time-inhomogeneous SDEs can be obtained according to Egorov et al. (2003), see Picchini, Ditlevsen and De Gaetano (2008) for an application of the time-inhomogeneous case. The likelihood function is calculated by numerically integrating the approximated conditional likelihood with respect to the random parameters using Gaussian quadrature rules and the parameters of the SDMEM are estimated by (approximated) maximum likelihood.

The method is evaluated by simulations of a Brownian motion with drift (or equivalently a log-transformed Geometric Brownian Motion), of the Ornstein-Uhlenbeck (OU) and the Cox-Ingersoll-Ross (CIR) process. The estimates are close to the true parameter values, only using moderate values of M (the number of experimental units) and n (the number of observations for a given experimental unit), relevant for most biomedical applications. Finally, two applications with real data are presented. In one of these the parameters of the SDMEM were estimated in a few minutes using simultaneously nearly two million observations from a neuronal experiment, by means of a single common PC. In conclusion, the method is an efficient computational method for fitting SDMEMs.

The paper is organized as follows. Section 2 introduces the SDMEMs, the observation scheme and the necessary notation. Section 3 includes the main tools for the parameter estimation of SDMEMs, i.e. introduces the likelihood function for a SDMEM and some approximations when the expression of the exact likelihood function cannot be obtained. Section 4 is devoted to the application of the estimation method presented in Section 3 to simulated datasets; implementation issues are also discussed. Section 5 presents two applications of the estimation method to real datasets. Section 6 summarizes the results of the paper and discusses the advantages and limitations of the method that is introduced. An appendix containing technical results closes the paper.

2 Formulation of Stochastic Differential Mixed-Effects Models

Consider a one-dimensional continuous process X_t evolving in M different experimental units (e.g. subjects) randomly chosen from a theoretical population: a SDMEM is defined as

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

where X_t^i is the value of the process at time $t \geq t_0^i$ in the ith unit and $X_0^i = X_{t_0^i}^i$; $\theta \in \Theta \subseteq \mathbb{R}^p$ is a p-dimensional fixed effects parameter (the same for the entire population) and $b^i \in B \subseteq \mathbb{R}^q$ is a q-dimensional random effects parameter (subject specific) with components $(b_1^i, ..., b_q^i)$; each component b_l^i may follow a different distribution $(l = 1, ..., q)$. The joint density function of the vector b^i is denoted $p_B(b^i \mid \Psi)$, which is parametrized by an r-dimensional parameter $\Psi \in \Upsilon \subseteq \mathbb{R}^r$, and thus Ψ collects all the parameters specifying the marginal distributions of the components $\{b_l^i\}$ of b^i as well as the covariance structure between the b_l^i's. The W_t^i are standard Brownian motions. The W_t^i and b^j are assumed mutually independent for all $1 \leq i, j \leq M$. Finally, X_0^i is assumed deterministic and equal to a known real constant x_0^i. The drift and the diffusion coefficient functions $\mu(\cdot) : E \times \Theta \times B \to \mathbb{R}$ and $\sigma(\cdot) : E \times \Theta \times B \to \mathbb{R}^+$ are assumed known up to the parameters, and are assumed sufficiently regular to ensure a unique weak solution (Øksendal (2007)), where $E \subseteq \mathbb{R}$ denotes the state space of X_t^i. Model (1) assumes that in each of the M subjects the evolution of X follows a common functional form, and differences between subjects are due to different realizations of the Brownian motion paths $\{W_t^i\}_{t \geq t_0^i}$ and of the random parameters b^i. Thus, in (1) the dynamics within a generic subject i are expressed via an (Itô) SDE model driven by Brownian motion, and the introduction of a parameter randomly varying among subjects allows for the explanation of the variability between the M subjects.

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 $\{t_0^i, t_1^i, \ldots, t_{n_i}^i\}$, $i = 1, ..., M$. Let $\underline{x}^i$ be the vector of responses for subject i, $\underline{x}^i = (x_0^i, \ldots, x_{n_i}^i)$, where $x^i(t_j^i) = x_j^i$, and let $\underline{x} = (\underline{x}^1, ..., \underline{x}^M)$ be the N-dimensional total response vector, $N = \sum_{i=1}^M (n_i + 1)$. Write $\Delta_j^i = t_j^i - t_{j-1}^i$ for the time-distance between the observations x_{j-1}^i and x_j^i.

We wish to estimate (θ, Ψ) using simultaneously all the data in $\underline{x}$, i.e. the individual data $\underline{x}^i$ are not fitted separately. Thus, for the moment, the specific values of the b^i's are not of interest, but only the identification of the vector-parameter Ψ characterizing their distribution. The problem of estimating the random effects b^i's will be considered in Section 3.2.

Note that it is straightforward to extend this setting to a multidimensional process X_t.

3 Maximum likelihood estimation in SDMEMs

The marginal density of $\underline{x}^i$ is obtained by integrating 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 that W_t^i and b^j are independent. This yields the likelihood function

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

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 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 (multi)normal, but $p_B(\cdot)$ could be any density function subject to mild regularity conditions. Solving the integral in (3) yields the marginal likelihood of the parameters, independent of the random effects b^i; by maximizing (3) with respect to θ and Ψ the corresponding maximum likelihood estimators (MLE) $\hat{\theta}$ and $\hat{\Psi}$ are obtained. Notice that it is possible to consider random effects having discrete distributions: in that case the integral becomes a sum and can be easily computed when the transition density p_X is known.

In simple cases the integral (3) can be solved, and explicit estimating equations for the MLE can be found, see Example 1. However, in general it is not possible to explicitly solve the integral, i.e. when: (i) $p_X(x^i_j, \cdot|x^i_{j-1}, \cdot)$ is known but the integral cannot be solved analytically, and (ii) $p_X(x^i_j, \cdot|x^i_{j-1}, \cdot)$ is unknown. In (i) the integral has to be numerically evaluated. In (ii) first $p_X(x^i_j, \cdot|x^i_{j-1}, \cdot)$ is approximated, then the integral is numerically solved. In situation (ii) we propose to approximate the transition density in closed-form, using a Hermite expansion as suggested in Aït-Sahalia (2002b, 2008), see Section 3.3.

3.1 Likelihood approximation

The MLE obtained by maximizing (4) has in most cases the usual good properties (see e.g. Dacunha-Castelle and Florens-Zmirou (1986)), but requires the transition densities, which are usually unknown. In particular, we assume that the MLE is a unique maximum of the continuous likelihood function. Assume an approximation $Q_K(\underline{x}^i|b^i, \theta) = \prod_{j=1}^{n_i} q_K(x^i_j, \Delta^i_j|x^i_{j-1}, b^i, \theta)$ to (4), and substitute it for the unknown conditional likelihood in (3), obtaining a sequence of approximations to the likelihood function

$$L^{(K)}(\theta, \Psi) = \prod_{i=1}^{M} \int_B Q_K(\underline{x}^i|b^i, \theta)\, p_B(b^i|\Psi)\, db^i. \tag{5}$$

By maximizing (5) with respect to (θ, Ψ) approximated MLE $\hat{\theta}^{(K)}$ and $\hat{\Psi}^{(K)}$ are obtained. In general, the integral in (5) does not have a closed form solution, and therefore efficient numerical integration methods are needed. General purpose approximation methods for one- or multi-dimensional integrals, irrespective of the random effects distribution, are available (e.g. Fröberg (1985), Krommer and Ueberhuber (1998)) within several software packages, though the complexity of the problem grows fast when increasing the dimension of B.

The literature devoted to nonlinear mixed-effects models (NLME) contains different approximate methods, with varying degrees of accuracy and computational complexity: e.g. in Lindstrom and Bates (1990) the likelihood of a NLME is approximated with the likelihood of a linear mixed-effects model; further approaches approximate the likelihood of a NLME using Laplacian and Gaussian quadrature approximation (see Pinheiro and Bates (1995, 2002), McCulloch and Searle (2001) and references therein); recent advances are considered in Pinheiro and Chao (2006) in the framework of generalized linear mixed models. In Section 3.1.1 the special case of a normally distributed random effect is treated; in Section 3.1.2 the general case of a random effect following any sufficiently well-behaved continuous distribution is considered.

3.1.1 A normally distributed random effect

Consider the following integral

$$\int_{-\infty}^{+\infty} h(u) e^{-u^2} du, \tag{6}$$

where $h(\cdot) \in C^{2R}(\mathbb{R})$, i.e. $h(\cdot)$ is $2R$ times continuously differentiable, for R a positive integer. It can be solved using Gaussian-Hermite quadrature (e.g. Fröberg (1985), Krommer and Ueberhuber (1998)), which is a Gaussian interpolatory quadrature formula approximating (6) as:

$$\int_{-\infty}^{+\infty} h(u) e^{-u^2} du \simeq \sum_{r=1}^{R} h(z_r) w_r$$

using R evaluation points z_r (nodes) and weights w_r defined by

$$z_r = r\text{th zero of } H_R(u) \tag{7}$$

$$w_r = \frac{2^{R-1} R! \sqrt{\pi}}{R^2 [H_{R-1}(z_r)]^2} \tag{8}$$

with an approximation error

$$E_R = \frac{R!\sqrt{\pi}}{2^R (2R)!} \frac{d^{2R}}{du^{2R}} h(u)\big|_{u=c} \qquad \text{for some } c \in \mathbb{R}. \tag{9}$$

Here $H_R(\cdot)$ is the Hermite polynomial of degree R. If $h(\cdot)$ is polynomial of degree at most $2R-1$, the Gauss-Hermite quadrature gives the exact value of the integral (Krommer and Ueberhuber (1998)).

Consider a one-dimensional ($q = 1$) normally distributed random effect $b^i \sim \mathcal{N}(0, \eta^2)$, so that (5) is the product of M one-dimensional integrals and $\Psi = \eta^2$. Define $u^i = b^i/(\sqrt{2}\eta)$, then (5) becomes

$$\begin{aligned} L^{(K)}(\theta, \eta^2) &= \prod_{i=1}^{M} \int_{-\infty}^{+\infty} \prod_{j=1}^{n_i} q_K(x_j^i, \Delta_j^i | x_{j-1}^i, \sqrt{2}\eta u^i, \theta) \frac{e^{-{u^i}^2}}{\sqrt{\pi}} du^i \\ &= \prod_{i=1}^{M} \int_{-\infty}^{+\infty} h_K^i(u^i) e^{-{u^i}^2} du^i \end{aligned}$$

where

$$h_K^i(u^i) = \prod_{j=1}^{n_i} q_K(x_j^i, \Delta_j^i | x_{j-1}^i, \sqrt{2}\eta u^i, \theta)/\sqrt{\pi}. \tag{10}$$

Thus, assuming $h_K^i(\cdot) \in C^{2R}(\mathbb{R})$ for all $i = 1, \ldots, M$ and using Gaussian-Hermite quadrature,

$$L^{(K)}(\theta, \eta^2) \simeq L^{(K,R)}(\theta, \eta^2) = \prod_{i=1}^{M} \sum_{r=1}^{R} h_K^i(z_r) w_r \tag{11}$$

where z_r and w_r are given by (7) and (8). An approximated MLE of (θ, η^2) is then given by $(\hat{\theta}^{(K,R)}, (\hat{\eta}^{(K,R)})^2) = \arg\min_{\theta,\eta^2}(-\log L^{(K,R)}(\theta, \eta^2))$.

Notice that using a Gaussian interpolatory quadrature formula (e.g. Gauss-Legendre, Gauss-Laguerre, Gauss-Hermite, Gauss-Jacobi) the approximation of an integral on the interval $[a, b]$ converges to the exact value when $R \to \infty$, $h(\cdot) \in C([a, b])$ and $[a, b]$ is a bounded interval, see Krommer and Ueberhuber (1998, p. 139).

3.1.2 A random effect following a continuous distribution

In this Section we consider the general case of a random effect b^i having density p_B (not necessarily Gaussian), with certain conditions on existence of moments. In Golub and Welsch (1969) a Gaussian quadrature integration method for any non-negative measure is suggested: in particular, Fernandes and Atchley (2006) report explicit formulae for the cases of Normal, Gamma, log-Normal, Student's t, inverse Gamma, Beta and Fisher's F distributions, covering a large class of problems commonly encountered in e.g. biomathematics/biostatistics.

Consider the following integral

$$\int_B h(u)\omega(u)du$$

where $h(\cdot) \in C^{2R}(B)$ for some chosen R and $\omega(\cdot)$ is a density function with support B fulfilling

$$\mathbb{E}(U^{2R}) < \infty \tag{12}$$

for $U \sim \omega(u)$. Then

$$\int_B h(u)\omega(u)du \;\; \simeq \;\; \sum_{r=1}^{R} h(z_r)w_r \tag{13}$$

with an approximation error E_R given by (Fröberg (1985) p. 290)

$$E_R = \frac{1}{(2R)!}\frac{d^{2R}}{du^{2R}}h(u)\big|_{u=c} \cdot \int_B \omega(y)[\pi(y)]^2 dy \tag{14}$$

for some $c \in B$, where $\pi(y) = \prod_{r=1}^{R}(y - z_r)$. The integral in (14) is finite under (12) and $E_R \to 0$ when $R \to \infty$ if B is bounded. The z_r's are the eigenvalues of a tridiagonal matrix J, defined by

$$J = \begin{pmatrix} \alpha_0 & \sqrt{\beta_1} & & & & \mathbf{0} \\ \sqrt{\beta_1} & \alpha_1 & \sqrt{\beta_2} & & & \\ & \sqrt{\beta_2} & \ddots & \ddots & & \\ & & \ddots & \ddots & \sqrt{\beta_{R-2}} & \\ & & & \sqrt{\beta_{R-2}} & \alpha_{R-2} & \sqrt{\beta_{R-1}} \\ \mathbf{0} & & & & \sqrt{\beta_{R-1}} & \alpha_{R-1} \end{pmatrix}$$

where the α_r's and the β_r's are specific to the distribution $\omega(\cdot)$, and $w_r = q_{r,1}^2$, where $q_{r,1}$ is the first component of the normalized eigenvector q_r of J. In Fernandes and Atchley (2006) the α_r's and β_r's are explicitly given for some important distributions $\omega(\cdot)$. The approximation (13) is exact whenever h is a polynomial of degree $2R-1$ or less. See Example 3 for an exponentially distributed random effect and Section 5.1 for a lognormally distributed effect.

Define $\omega(b^i) = p_B(b^i|\Psi)$ and

$$h_K^i(b^i) = \prod_{j=1}^{n_i} q^{(K)}(x_j^i, \Delta_j^i | x_{j-1}^i, b^i, \theta). \tag{15}$$

Assuming that

$$h_K^i(b^i) \in C^{2R}(B) \quad \text{and} \quad \mathbb{E}\left(b^{i^{2R}}\right) < \infty \tag{16}$$

the likelihood (5) is approximated by

$$L^{(K,R)}(\theta,\Psi)=\prod_{i=1}^{M}\sum_{r=1}^{R}h_K^i(z_r)w_r, \tag{17}$$

and $(\hat{\theta}^{(K,R)},\hat{\Psi}^{(K,R)})=\arg\min_{\theta,\Psi}(-\log L^{(K,R)}(\theta,\Psi))$ is an approximated MLE of (θ,Ψ).

3.2 Random effects estimation

The random parameters b^i are estimated in the standard way from mixed-effects theory by

$$b^{i(K,R)}=\arg\min_{b^i}\left(-\sum_{j=1}^{n_i}\log q^{(K)}(x_j^i,\Delta_j^i|x_{j-1}^i,b^i,\hat{\theta}^{(K,R)})\right),\qquad i=1,...,M \tag{18}$$

where the estimate of θ has been plugged in. See Section 5.2 for an application.

3.3 Closed-form transition density expansion

Here we review the transition density expansion of a one-dimensional time-homogeneous SDE as suggested in Aït-Sahalia (2002b) and adapt to the case of a SDMEM. An extension to time-inhomogeneous SDEs is given in Egorov et al. (2003). A generalization to multidimensional SDEs and references for additional extensions are given in Aït-Sahalia (2008). Consider the following one-dimensional time-homogeneous SDMEM for a generic subject i:

$$dX_t^i=\mu(X_t^i,\theta,b^i)dt+\sigma(X_t^i,\theta,b^i)dW_t^i,\qquad X_0^i=x_0^i. \tag{19}$$

To approximate $p_X(x_j^i,\Delta_j^i|x_{j-1}^i,b^i,\theta)$ we assume $\mu(\cdot)$ and $\sigma(\cdot)$ infinitely differentiable in X_t^i and three times continuously differentiable in θ and b^i for all $X_t^i\in E$ and $(\theta,b^i)\in\Theta\times B$; we also assume the existence of a constant c such that $\sigma(X_t^i,\theta,b^i)>c>0$ for all $X_t^i\in E$ and $(\theta,b^i)\in\Theta\times B$. Weaker conditions on the diffusion coefficient close to the boundary of the state space can be considered, e.g. at 0 for positive diffusions so that also the Cox-Ingersoll-Ross model is covered; see Aït-Sahalia (2002b) for further details. For a generic SDE the Lamperti transform $\gamma(\cdot)$ is defined by

$$Y_t\equiv\gamma(X_t)=\int^{X_t}\frac{du}{\sigma(u;\theta)} \tag{20}$$

where the lower bound of integration is an arbitrary point in the interior of E, and the resulting process Y_t is the solution of an SDE with diffusion term constantly equal to one and drift term given by

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

Using such transformation the transition density of X_t^i is approximated by

$$p_X^{(S)}(x_j^i,\Delta_j^i|x_{j-1}^i,b^i,\theta)\;=\;\sigma(z_j^i,\theta,b^i)^{-1}\Delta_j^{i\,-1/2}\phi(z_j^i)\sum_{s=0}^{S}\eta_Z^{(s)}(\Delta_j^i,\gamma(x_{j-1}^i);\theta,b^i)H_s(z_j^i)$$

where $\phi(\cdot)$ is the standard normal density function, $z_j^i = (\gamma(x_j^i) - \gamma(x_{j-1}^i))/\sqrt{\Delta_j^i}$, and H_s is the s'th Hermite polynomial. The coefficients $\eta_Z^{(s)}$ are given by the moments

$$\eta_Z^{(s)}(\Delta_j^i, \gamma(x_{j-1}^i); \theta, b^i) = \frac{1}{s!}\int_{-\infty}^{\infty} H_s(z_j^i) p_Z(\Delta_j^i, z_j^i | \gamma(x_{j-1}^i), b^i, \theta) dz_j^i \tag{21}$$

where $p_Z(\cdot)$ is the transition density of the transformed variable $Z_{t+\Delta} = (\gamma(X_{t+\Delta}) - \gamma(X_t))/\sqrt{\Delta}$. Following Theorem 1 in Aït-Sahalia (2002b) the $p_X^{(S)}$ converges uniformly in (θ, b^i) to the true transition density p_X when $S \to \infty$.

If the conditional moments (21) cannot be calculated explicitly (which is often the case), a Taylor series expansion in the time steps Δ_j^i can be used. The logarithm of the transition density can then be expanded in closed form using an order $S = \infty$ Hermite series, and approximated by a Taylor expansion up to order K, obtaining the explicit sequence:

$$\begin{aligned} \ln p_X^{(K)}(x_j^i, \Delta_j^i | x_{j-1}^i, b^i, \theta) &= -\frac{1}{2}\ln(2\pi\Delta_j^i) - \frac{1}{2}\ln(\sigma^2(x_j^i, \theta, b^i)) + \frac{C_Y^{(-1)}(\gamma(x_j^i)|\gamma(x_{j-1}^i))}{\Delta_j^i} \\ &+ \sum_{k=0}^{K} C_Y^{(k)}(\gamma(x_j^i)|\gamma(x_{j-1}^i)) \frac{\Delta_j^{i\,k}}{k!} \end{aligned} \tag{22}$$

where $\Delta_j^{i\,k}$ is Δ_j^i raised to the power of k. The coefficients $C_Y^{(k)}$ are given in Appendix A.

4 Implementation issues and numerical applications

Trajectories of the Geometric Brownian Motion, the OU and the CIR process perturbed with random effects were simulated. Data points from the trajectories were retrieved and on the obtained datasets the parameters were estimated. The main goals were to check the feasibility and effectiveness of the estimation procedure, and that acceptable results can be obtained for small sample sizes (say $M = 10, ..., 50$ subjects and $n = 10, ..., 50$ observations collected on each subject). Applications with real data are given in Section 5.

It has been shown that $K = 1$ or 2 is often sufficient to obtain a good approximation to the transition density (Aït-Sahalia (2008, 2002b), Egorov et al. (2003)). We use either $K = 1$ or 2 order density expansion depending on the model. In particular, for the Geometric Brownian Motion, $K = 1$ gives the exact density. All the integrals are numerically evaluated using Gaussian quadrature with $R = 40$: though $R = 20$ is usually considered enough for a good degree of approximation (McCulloch and Searle, 2001, p. 272). The coefficients $C_Y^{(k)}$ are given in Appendix A (in general, the $C_Y^{(k)}$ can be calculated using a symbolic calculus software).

Parametric bootstrap was performed to obtain means and 95% confidence intervals (CI) of the parameter estimates. For each SDMEM one thousand 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) MLE were obtained. For each parameter, the sample mean and the empirical 95% CI (from the 2.5th to the 97.5th percentile) from the 1000 obtained estimates are reported in Tables 1–5 together with measures of symmetry (skewness and kurtosis). To overcome numerical problems in the optimization procedure, due to very large or very small values returned by the product of densities (e.g. (10)) for the current parameter values, it might be necessary to use a package for arbitrary/variable precision computation. We used the package by Barrowes (2007) in our MATLAB program.

Finally, we want to stress the usefulness of the closed-form density expansion to approximate p_X. Using simulated maximum likelihood approaches (see the Introduction) the numerical simulation of thousands of trajectories of the process may be required in each step of an optimization algorithm, which is computationally expensive. Using the closed-form density expansion, simulating process trajectories is not required, e.g. in our examples the parameter estimates were all obtained within one minute (depending on the size (M, n) of the problem) using a MATLAB program on a 3.0 GHz Intel Pentium IV with 512 MB of RAM.

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

Consider a SDMEM of the Geometric Brownian motion

$$dX_t^i = (\beta + \beta^i)X_t^i dt + \sigma X_t^i dW_t^i, \qquad X_0^i = x_0^i, \quad i = 1, ..., M \tag{23}$$

which is relevant e.g. in pharmacokinetics for the metabolism of a compound in plasma following first order kinetics where $\beta < 0$, or as a growth model, e.g. the initial growth of bacterial or tumor cell populations, where $\beta > 0$. The transformed process $Z_t^i = \log(X_t^i)$ gives the SDMEM:

$$dZ_t^i = (\beta + \beta^i - \sigma^2/2)dt + \sigma dW_t^i, \qquad Z_0^i = z_0^i, \quad i = 1, ..., M \tag{24}$$

and we assume $\beta^i \sim \mathcal{N}(0, \eta^2)$. In this simple example $b^i = \beta^i$, $\theta = (\beta, \sigma^2)$ and $\Psi = \eta^2$. We wish to estimate $(\beta, \sigma^2, \eta^2)$ given the observations $\underline{z} = (\underline{z}^1, ..., \underline{z}^M)$ from model (24). Note that no stationary solution exists.

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

$$\begin{aligned}\log L(\theta, \Psi) \quad = \quad & \frac{M}{2}\log\left(\frac{\sigma^2}{\eta^2}\right) - \frac{N-M}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{M}\log\left(\Delta^{i^{n_i}}\left(T^i + \frac{\sigma^2}{\eta^2}\right)\right) \\ & - \frac{\sum_{i,j}\frac{1}{\Delta_j^i}(z_j^i - z_{j-1}^i - \alpha\Delta_j^i)^2 - \sum_i (z_{n_i}^i - z_0^i - \alpha T^i)^2 \left(T^i + \frac{\sigma^2}{\eta^2}\right)^{-1}}{2\sigma^2}\end{aligned} \tag{25}$$

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 $T^i = \sum_{j=1}^{n_i}\Delta_j^i$.

Assume equidistant observations and the same number of observations per subject, i.e. $\Delta_j^i = \Delta$ and $n_i = n$ for all $1 \le i \le M$, $1 \le j \le n_i$. The MLE are then given by (Ditlevsen and De Gaetano (2005a)):

$$\hat{\sigma}^2 \quad = \quad \frac{\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{n}(z_j^i - z_{j-1}^i - \hat{\alpha}\Delta)^2 - \frac{\Delta}{MT}\sum_{i=1}^{M}(z_n^i - z_0^i - \hat{\alpha}T)^2}{T - \Delta} \tag{26}$$

$$\hat{\eta}^2 \quad = \quad \frac{\frac{1}{MT}\left[\sum_{i=1}^{M}(z_n^i - z_0^i - \hat{\alpha}T)^2 - \sum_{i=1}^{M}\sum_{j=1}^{n}(z_j^i - z_{j-1}^i - \hat{\alpha}\Delta)^2\right]}{T - \Delta} \tag{27}$$

$$\hat{\beta} \quad = \quad \hat{\alpha} + \frac{\hat{\sigma}^2}{2} \tag{28}$$

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

Table 1: Example 1, exact MLE and 95% empirical CI from simulations of model (23).

Parameter values β	σ^2	η^2		$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\eta}^2$
					$M = 10$, $n = 50$	
-0.2	0.2	0.02	Mean [95% CI]	-0.132 [-0.199, -0.061]	0.199 [0.176, 0.224]	0.008 [0.001, 0.020]
			Skewness	0.142	0.271	0.887
			Kurtosis	2.686	2.942	3.975
					$M = 50$, $n = 10$	
-0.2	0.2	0.02	Mean [95% CI]	-0.200 [-0.249, -0.154]	0.200 [0.174, 0.227]	0.019 [0.012, 0.028]
			Skewness	-0.059	0.143	0.337
			Kurtosis	3.095	3.238	3.182
					$M = 10$, $n = 50$	
-0.02	0.02	0.02	Mean [95% CI]	-0.032 [-0.105, 0.040]	0.020 [0.018, 0.023]	0.014 [0.005, 0.026]
			Skewness	-0.104	0.315	0.491
			Kurtosis	2.764	3.408	3.165
					$M = 50$, $n = 10$	
-0.02	0.02	0.02	Mean [95% CI]	-0.020 [-0.062, 0.024]	0.020 [0.017, 0.023]	0.020 [0.012, 0.028]
			Skewness	-0.011	0.143	0.393
			Kurtosis	3.051	3.238	3.218

Table 2: Example 1, MLE and 95% empirical CI, from simulations of model (23), solving the integral numerically.

Parameter values β	σ^2	η^2		$\hat{\beta}^{(40)}$	$(\hat{\sigma}^{(40)})^2$	$(\hat{\eta}^{(40)})^2$
					$M = 10$, $n = 50$	
-0.2	0.2	0.02	Mean [95% CI]	-0.133 [-0.200, -0.061]	0.198 [0.176, 0.224]	0.008 [0.001, 0.020]
			Skewness	0.136	0.272	0.850
			Kurtosis	2.693	2.944	3.810
					$M = 50$, $n = 10$	
-0.2	0.2	0.02	Mean [95% CI]	-0.200 [-0.248, -0.155]	0.200 [0.174, 0.227]	0.020 [0.012, 0.028]
			Skewness	-0.056	0.142	0.438
			Kurtosis	3.088	3.233	3.737
					$M = 10$, $n = 50$	
-0.02	0.02	0.02	Mean [95% CI]	-0.021 [-0.042, -0.001]	0.020 [0.018, 0.022]	0.014 [0.006, 0.026]
			Skewness	-1.681	0.299	0.702
			Kurtosis	21.419	3.419	3.569
					$M = 50$, $n = 10$	
-0.02	0.02	0.02	Mean [95% CI]	-0.021 [-0.045, 0.002]	0.021 [0.018, 0.023]	0.015 [0.010, 0.022]
			Skewness	-0.041	0.135	0.717
			Kurtosis	3.678	3.223	3.397

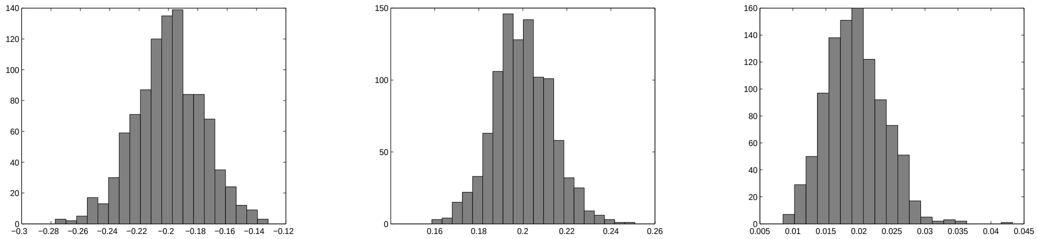
The exact estimators (26)–(28) can be used as a test of the estimation method. Here $C_Y^{(k)}(\cdot) = 0$ for all $k \geq 2$, and the order $K = 1$ density expansion results in the exact transition density, see Appendix A for details. Thus, the exact MLE are compared with the approximated estimators, 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, 1000 data sets were generated from (23) and the parameters were estimated using (26)–(28) (see Table 1), and using (11) (see Table 2).

In all simulations $X_0^i = 100$ for all i and $T = 100$. Comparing Table 1 and 2 shows that the numerical integration of the integral is accurate. The true parameter values are well identified when M is larger than 10, though σ results well identified also in the case $M = 10$: these results were expected, since M is the sample size of draws from the distribution of β^i. From the empirical distribution of the approximated estimates (Figure 1) it seems reasonable to assume an asymptotic normal distribution of the estimates.

Example 2: Ornstein-Uhlenbeck process with one random effect

Consider a SDMEM of the OU process:

$$dX_t^i = \left(-\frac{X_t^i}{\tau} + \mu + \mu^i\right)dt + \sigma dW_t^i; \qquad X_0^i = x_0^i = 0, \quad i = 1, ..., M \tag{29}$$



(a) $\hat{\beta}^{(40)}$ (b) $(\hat{\sigma}^{(40)})^2$ (c) $(\hat{\eta}^{(40)})^2$

Figure 1: Geometric Brownian Motion, histograms of $\hat{\beta}^{(40)}$, $(\hat{\sigma}^{(40)})^2$ and $(\hat{\eta}^{(40)})^2$ for the case $(\beta, \sigma^2, \eta^2) = (-0.2, 0.2, 0.02)$ with $(M, n) = (50, 10)$.

where $\mu \in \mathbb{R}$, $\tau > 0$ and $\sigma > 0$. The OU process is the simplest mean-reverting SDE, and has been widely used e.g. in neuronal modeling, biology, physics, engineering and finance. The parametrization is chosen as is customary in neuronal modeling. Assume $\mu^i \sim \mathcal{N}(0, \eta^2)$. Here $b^i = \mu^i$ and we want to estimate $\theta = (\mu, \tau, \sigma)$ and $\Psi = \eta^2$ given a set of observations $\underline{x}$ from (29). The conditional mean and variance of X_t^i are

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

and the transition density is normal and given by

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

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

$$\begin{aligned} L(\theta, \Psi) &= \prod_{i=1}^{M} \Bigg\{ (\pi\sigma^2\tau)^{-n_i/2} \left(\prod_{j=1}^{n_i} (1 - e^{-2\Delta_j^i/\tau})^{-1/2} \right) (2\pi\eta^2)^{-1/2} \qquad (30) \\ &\times \int_{\mathbb{R}} \exp\left\{ \sum_{j=1}^{n_i} \left[- \frac{\big(x_j^i - x_{j-1}^i e^{-\Delta_j^i/\tau} - (\mu + \mu^i)\tau(1 - e^{-\Delta_j^i/\tau})\big)^2}{\sigma^2 \tau (1 - e^{-2\Delta_j^i/\tau})} \right] - \frac{\mu^{i2}}{2\eta^2} \right\} d\mu^i \Bigg\}. \end{aligned}$$

We have no closed-form solution to this integral, so exact estimators of θ and Ψ are unavailable. We first consider a Gauss-Hermite integration approach with $R = 40$, the resulting estimators are denoted with $(\hat{\theta}^{(R)}, \hat{\Psi}^{(R)})$. Secondly, we ignore that the exact transition density expression is available, and we compute the approximated estimator $(\hat{\theta}^{(K,R)}, \hat{\Psi}^{(K,R)})$ by approximating in closed-form the transition density with $K = 2$. The estimation results, obtained on 1000 simulated data sets generated by (29) using the Euler-Maruyama scheme with integration stepsize of 0.01 (Kloeden and Platen (1992)), are reported in Table 3 and Table 4 for $(\hat{\theta}^{(R)}, \hat{\Psi}^{(R)})$ and $(\hat{\theta}^{(K,R)}, \hat{\Psi}^{(K,R)})$, respectively. For both strategies $n_i = n$ for all i and $T = 100$.

Table 3: Example 2, approximated MLE and 95% empirical CI from simulations of model (29), using the exact transition density.

Parameter values μ	τ	σ	η^2		$\hat{\mu}^{(40)}$	$\hat{\tau}^{(40)}$	$\hat{\sigma}^{(40)}$	$(\hat{\eta}^{(40)})^2$
					$M=10$, $n=50$			
1	10	1	1	Mean [95% CI]	1.023 [0.831, 1.209]	9.846 [8.423, 11.687]	0.989 [0.925, 1.050]	1.266 [0.374, 2.711]
				Skewness	0.307	0.332	-0.062	0.856
				Kurtosis	4.827	3.957	3.008	3.885
					$M=50$, $n=10$			
1	10	1	1	Mean [95% CI]	1.010 [0.838, 1.191]	9.824 [8.700, 11.061]	0.955 [0.893, 1.021]	1.097 [0.676, 1.654]
				Skewness	0.071	0.149	0.089	0.487
				Kurtosis	2.462	3.104	3.051	3.465
					$M=10$, $n=50$			
2	12	0.1	0.25	Mean [95% CI]	2.033 [1.844, 2.237]	11.902 [11.101, 12.761]	0.147 [0.124, 0.166]	0.330 [0.104, 0.692]
				Skewness	0.101	0.219	-0.303	0.856
				Kurtosis	2.410	3.139	2.830	4.006
					$M=50$, $n=10$			
2	12	0.1	0.25	Mean [95% CI]	2.045 [1.858, 2.223]	11.770 [11.192, 12.353]	0.283 [0.257, 0.305]	0.290 [0.177, 0.422]
				Skewness	-0.014	-0.004	-0.353	0.390
				Kurtosis	2.323	2.760	3.566	3.435

Table 4: Example 2, approximated MLE (95% empirical CI), from simulations of model (29), using an order $K=2$ density expansion.

Parameter values μ	τ	σ	η^2		$\hat{\mu}^{(2,40)}$	$\hat{\tau}^{(2,40)}$	$\hat{\sigma}^{(2,40)}$	$(\hat{\eta}^{(2,40)})^2$
					$M=10$, $n=50$			
1	10	1	1	Mean [95% CI]	1.019 [0.804, 1.257]	10.043 [8.307, 11.954]	1.000 [0.934, 1.063]	0.853 [0.417, 1.421]
				Skewness	0.239	0.285	-0.060	0.740
				Kurtosis	3.498	3.439	2.928	4.459
					$M=50$, $n=10$			
1	10	1	1	Mean [95% CI]	0.926 [0.739, 1.112]	11.930 [11.029, 12.972]	0.999 [0.932, 1.064]	0.578 [0.382, 0.839]
				Skewness	0.028	0.281	0.006	0.734
				Kurtosis	2.736	2.993	3.057	3.947
					$M=50$, $n=50$			
1	10	1	1	Mean [95% CI]	1.000 [0.820, 1.185]	10.306 [9.531, 11.182]	1.000 [0.971, 1.029]	0.718 [0.477, 1.025]
				Skewness	0.164	0.194	0.062	0.574
				Kurtosis	2.851	2.892	3.106	3.329
					$M=10$, $n=50$			
2	12	0.1	0.25	Mean [95% CI]	1.999 [1.889, 2.106]	12.082 [11.914, 12.249]	0.111 [0.103, 0.120]	0.072 [0.029, 0.162]
				Skewness	-0.054	-0.007	0.418	1.631
				Kurtosis	2.740	2.934	3.445	6.908
					$M=50$, $n=10$			
2	12	0.1	0.25	Mean [95% CI]	1.936 [1.772, 2.122]	13.411 [13.322, 13.517]	0.386 [0.360, 0.413]	0.142 [0.085, 0.227]
				Skewness	0.142	0.289	-0.022	1.006
				Kurtosis	3.091	3.000	2.827	5.116
					$M=50$, $n=50$			
2	12	0.1	0.25	Mean [95% CI]	2.005 [1.920, 2.093]	12.087 [12.008, 12.170]	0.111 [0.107, 0.116]	0.045 [0.030, 0.070]
				Skewness	0.062	0.191	0.326	1.055
				Kurtosis	2.987	3.042	3.388	4.806

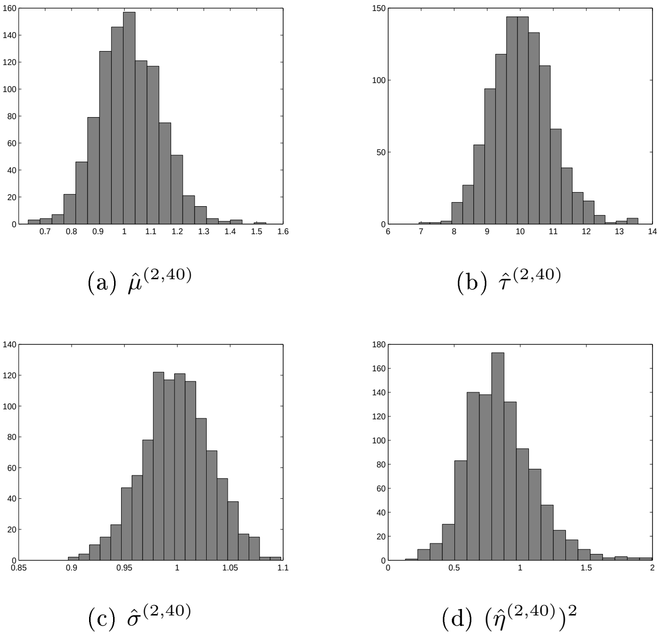
Tables 3 and 4 show that except for η^2 the true parameter values seem correctly identified using both the likelihood (31) and the corresponding order $K = 2$ approximation, though n should be larger than 10 to get satisfactory results. Therefore also $(M, n) = (50, 50)$ is considered in Table 4. The empirical distribution of the approximated estimates (Figure 2) seems to be reasonably close to a normal distribution.

Example 3: The CIR process with one random effect

Consider a SDMEM of the Cox-Ingersoll-Ross process (CIR) given by

$$dX_t^i = (-X_t^i + \mu + \mu^i)dt + \frac{\sigma\sqrt{X_t^i}}{\sqrt{\mu+\mu^i}}dW_t^i, \qquad X_0^i = x_0^i > 0, \quad i = 1, ..., M \tag{31}$$

with $\mu + \mu^i > 0$, $\sigma > 0$ and $2((\mu+\mu^i)/\sigma)^2 \geq 1$. For fixed i, the process is ergodic and its stationary distribution is a Gamma distribution with shape parameter $2((\mu+\mu^i)/\sigma)^2$ and scale parameter $\sigma^2/(2(\mu+\mu^i))$. Feller (1951) proposed it as a model for population growth,



(a) $\hat{\mu}^{(2,40)}$ (b) $\hat{\tau}^{(2,40)}$

(c) $\hat{\sigma}^{(2,40)}$ (d) $(\hat{\eta}^{(2,40)})^2$

Figure 2: OU process, histograms of $\hat{\mu}^{(2,40)}$, $\hat{\tau}^{(2,40)}$, $\hat{\sigma}^{(2,40)}$ and $(\hat{\eta}^{(2,40)})^2$ for the case $(\mu, \tau, \sigma, \eta^2) = (1, 10, 1, 1)$ with $(M, n) = (10, 50)$.

and it has been commonly applied in neuronal modeling under the name of a Feller process (e.g. Ditlevsen and Lansky (2006) and references therein). It was introduced in mathematical finance as a model of the short-term interest rate by Cox et al. (1985).

Assume μ^i exponentially distributed with mean $\lambda > 0$, so $\theta = (\mu, \sigma)$, $b^i = \mu^i$, $\Psi = \lambda$. Thus

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

where

$$p(\mu^i|\lambda) = \frac{1}{\lambda} e^{-\frac{\mu^i}{\lambda}}, \quad \mu^i > 0.$$

The integral is solved using (13) with R abscisses and coefficients $\alpha_{i_1} = \lambda(1+2i_1)$, $\beta_{i_2} = \lambda^2 i_2^2$, $i_1 = 0, 1, ..., R-1$, $i_2 = 1, 2, ..., R-1$ (Fernandes and Atchley (2006)). The estimation results, obtained on 1000 simulated data sets generated by (31) using the Milstein scheme with integration stepsize of 0.01 (Kloeden and Platen (1992)) are reported in Table 5 for $K = 2$ and $R = 40$. In all simulations $n_i = n$, $X_0^i = 1$ for all i and $T = 100$.

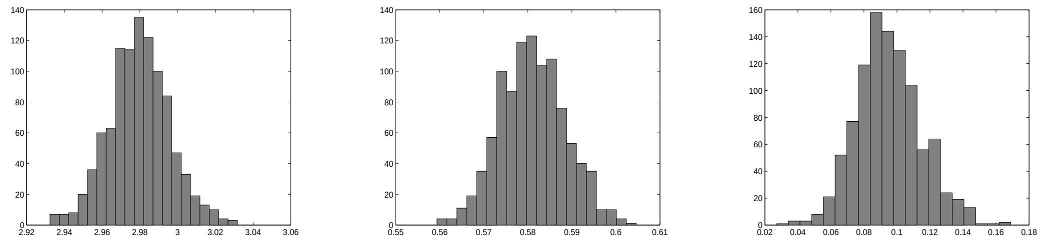
Here μ seems correctly identified, σ is overestimated and λ underestimated. The diffusion part of the SDMEM depends on the random effect, and this is a likely complication for the parameter estimation. However, the empirical distribution of the approximated estimates (Figure 3) seems to be reasonably close to a normal distribution.

5 Applications

In this Section we consider two applications to real data: a small data set ($M = 5$ experiments with $n = 7$ observations each) and a large data set ($M = 312$ experiments with n of the order of thousands for each experiment, see also Picchini, Ditlevsen, De Gaetano and Lansky (2008)).

Table 5: Example 3, approximated MLE (95% empirical CI) from simulations of model (31), using an order $K = 2$ density expansion.

Parameter values				$\hat{\mu}^{(2,40)}$	$\hat{\sigma}^{(2,40)}$	$\hat{\lambda}^{(2,40)}$
μ	σ	λ				
					$M = 10$, $n = 50$	
3	0.5	0.1	Mean [95% CI]	2.990 [2.919, 3.063]	0.582 [0.548, 0.615]	0.086 [0, 0.195]
			Skewness	0.148	-0.029	0.736
			Kurtosis	2.868	2.886	5.083
					$M = 50$, $n = 50$	
3	0.5	0.1	Mean [95% CI]	2.980 [2.948, 3.014]	0.581 [0.567, 0.596]	0.095 [0.061, 0.136]
			Skewness	0.065	0.111	0.265
			Kurtosis	3.253	2.861	3.328
					$M = 10$, $n = 50$	
1	0.1	0.2	Mean [95% CI]	0.989 [0.825, 1.066]	0.110 [0.103, 0.119]	0.129 [0.071, 0.220]
			Skewness	-2.492	0.744	0.901
			Kurtosis	12.470	5.515	3.674
					$M = 50$, $n = 50$	
1	0.1	0.2	Mean [95% CI]	1.008 [0.968, 1.034]	0.111 [0.107, 0.115]	0.110 [0.075, 0.180]
			Skewness	-3.601	0.148	1.605
			Kurtosis	31.676	2.815	6.968



(a) $\hat{\mu}^{(2,40)}$ (b) $\hat{\sigma}^{(2,40)}$ (c) $\hat{\lambda}^{(2,40)}$

Figure 3: CIR process, histograms of $\hat{\mu}^{(2,40)}$, $\hat{\sigma}^{(2,40)}$ and $\hat{\lambda}^{(2,40)}$ for the case $(\mu, \sigma, \lambda) = (3, 0.5, 0.1)$ with $(M, n) = (50, 50)$.

5.1 Orange trees growth

In Pinheiro and Bates (2002, Sections 8.1.1 and 8.2.1), data from a study on the growth of orange trees are analyzed by means of deterministic nonlinear mixed-effects models using the method proposed in Lindstrom and Bates (1990). The data are available in the `Orange` dataset provided in the `nlme` `R` package (Pinheiro et al. (2007); R Development Core Team (2007)). This is a balanced design consisting of seven measurements of the circumference of five orange trees. The following logistic model was proposed in Pinheiro and Bates (2002) to study the relationship between the circumference $X^{i,j}$ [mm], measured on the i-th tree at age t_{ij} [days], and the age ($i = 1, ..., 5$ and $j = 1, ..., 7$):

$$X^{i,j} = \frac{\phi_1}{1 + \exp(-(t_{ij} - \phi_2)/\phi_3)} + \varepsilon_{ij} \tag{32}$$

with ϕ_1 [mm], ϕ_2 [days] and ϕ_3 [days] all positive, and i.i.d. measurement errors $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. The parameter ϕ_1 represents the asymptotic circumference, ϕ_2 is the time at which $X = \phi_1/2$ (the inflection point of the logistic model) and ϕ_3 is the time-distance between the inflection point and the point where $X = \phi_1/(1 + e^{-1})$.

Then model (32) was enlarged by adding a zero mean normally distributed random effect with constant variance to each structural parameter (ϕ_1, ϕ_2, ϕ_3). The authors showed that the enlarged model lead to over-parametrization and they concluded that only the random effect $\phi_1^i \sim \mathcal{N}(0, \eta^2)$ for ϕ_1 was necessary. Thus, the model is

$$X^{i,j} = \frac{\phi_1 + \phi_1^i}{1 + \exp(-(t_{ij} - \phi_2)/\phi_3)} + \varepsilon_{ij} \tag{33}$$

and they obtained the estimates $\hat{\phi}_1 = 191.1$ $[159.4, 222.7]$, $\hat{\phi}_2 = 722.6$ $[653.7, 791.4]$, $\hat{\phi}_3 = 344.2$ $[291.0, 397.3]$, $\hat{\eta} = 31.5$ $[16.7, 59.4]$, and the residual standard deviation is estimated to $\hat{\sigma}_\varepsilon = 7.8$ $[6.1, 10.1]$. The dynamical model corresponding to (33) for the ith tree and ignoring the error is given by the ODE

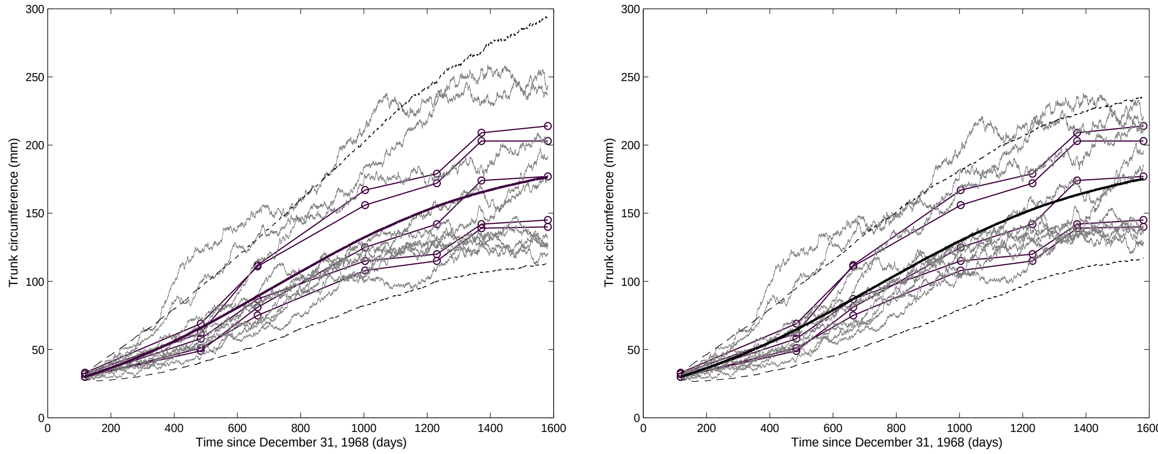
$$\frac{dX_t^i}{dt} = \frac{1}{\phi_3(\phi_1 + \phi_1^i)} X_t^i(\phi_1 + \phi_1^i - X_t^i), \qquad X_0^i = x_0^i, \quad t \geq t_0^i$$

with ϕ_2 appearing only in the deterministic initial condition $X_0^i = X_{t_0^i}^i = \phi_1/(1 + \exp[(\phi_2 - t_0^i)/\phi_3])$, where $t_0^i = 118$ days for all the trees. Since X_t^i and ϕ_1 are strictly positive we considered a lognormally distributed random effect and a state dependent diffusion coefficient, leading to the SDMEM

$$dX_t^i = \frac{1}{\phi_3(\phi_1 + \phi_1^i)} X_t^i(\phi_1 + \phi_1^i - X_t^i)dt + \sigma\sqrt{X_t^i}dW_t^i, \qquad X_0^i = x_0^i, \tag{34}$$

$$\phi_1^i \sim LN(\mu, \eta^2), \qquad \mu \in \mathbb{R}, \ \eta \in \mathbb{R}^+ \tag{35}$$

where σ has units $[(\text{mm/days})^{1/2}]$. Thus $M = 5$, $n_i = n = 7$, $\theta = (\phi_1, \phi_3, \sigma)$, $b^i = \phi_1^i$ and $\Psi = (\mu, \eta)$. We used an order $K = 2$ approximation to the likelihood and the integral was solved using the quadrature rule (13) with $R = 40$ abscisses and coefficients $\alpha_{i_1} = \exp\{\mu + \eta^2(2i_1 - 1)/2\}[\exp(\eta^2(i_1 + 1)) + \exp(i_1\eta^2) - 1]$, $\beta_{i_2} = \exp\{2\mu + (3i_2 - 2)\eta^2\}[\exp(i_2\eta^2) - 1]$, $i_1 = 0, 1, ..., R - 1$, $i_2 = 1, 2, ..., R - 1$, see Fernandes and Atchley (2006).



(a) (b)

Figure 4: (a): Measured circumferences ('o'; points corresponding to the same tree are connected by lines) and fit of the SDMEM (34)-(35) (lognormally distributed random effects): empirical mean curve of the SDMEM (bold solid line), 95% empirical confidence curves (dashed lines) and ten simulated trajectories (grey). (b): The same as in panel (a) but using normally distributed random effects.

The estimates were $\phi_1^{(2,40)} = 108.9$ $[54.1, 163.7]$, $\phi_3^{(2,40)} = 342.2$ $[269.9, 414.4]$, $\sigma^{(2,40)} = 0.084$ $[0.062, 0.106]$, $\mu^{(2,40)} = 4.267$ $[3.154, 5.379]$, $\eta^{(2,40)} = 0.586$ $[0.415, 0.758]$. For ease of comparison with the deterministic model results, it may be of interest to obtain an estimate for ϕ_2 also. By plugging the estimates into the expression $X_0^i = (\phi_1 + \mathbb{E}(\phi_1^i))/(1 + \exp[(\phi_2 - t_0^i)/\phi_3])$ we obtain $\phi_2^{(2,40)} = 700.0$ (using e.g. $X_0^i = 30$). The estimates cannot be directly compared, because in (34)-(35) lognormally distributed random effects are considered instead of the normal ones. Moreover, (33) models the measurement error but does not allow for stochastic fluctuations in the dynamical process, whereas no measurement error is considered in model (34)-(35). In (34)-(35) the estimated mean and standard deviation of ϕ_1^i are given by $\exp(\mu^{(2,40)} + {\eta^{(2,40)}}^2/2) = 84.66$ and $[(\exp({\eta^{(2,40)}}^2) - 1)\exp(2 \cdot \mu^{(2,40)} + {\eta^{(2,40)}}^2)]^{1/2} = 54.2$, respectively. In model (33) the random variable $\phi_1 + \phi_1^i$ has estimated mean 191.1 and standard deviation 31.5, whereas for the SDMEM $\phi_1 + \phi_1^i$ has estimated mean 194.5 and standard deviation 54.2.

The fit of the estimated SDMEM is given in Figure 4(a), which reports the data, the empirical mean of 5000 simulated trajectories from (34)-(35), generated with the Milstein scheme (Kloeden and Platen (1992)) using a stepsize of 0.4, the empirical 95% confidence bands and ten example trajectories. For each simulated trajectory a different realization of ϕ_1^i was drawn from the lognormal distribution $LN(\mu, \eta^2)$ with $\mu = 4.267$ and $\eta = 0.586$.

To compare with the results obtained by Lindstrom and Bates, now assume $\phi_1^i \sim \mathcal{N}(0, \eta^2)$. Under this model estimates are $\phi_1^{(2,40)} = 194.8$ $[158.5, 231.1]$, $\phi_3^{(2,40)} = 356.0$ $[270.2, 441.8]$, $\sigma^{(2,40)} = 0.088$ $[0.064, 0.113]$, $\eta^{(2,40)} = 28.17$ $[0.29, 56.04]$ and $\phi_2^{(2,40)} = 724.46$ (determined using $X_0^i = 30$ as in the previous case), which are close to the estimates obtained with the deterministic model. The model fit is given in Figure 4(b).

5.2 Stochastic leaky integrate-and-fire neuronal model

The stochastic leaky integrate-and-fire (LIF) neuronal models are common theoretical tools for studying properties of real neuronal systems. In Picchini, Ditlevsen, De Gaetano and Lansky (2008) the stochastic LIF model is extended allowing for a noise source determining slow fluctuations in the neuronal signal. This is achieved by adding a random variable to one of the parameters characterizing the neuronal input. The data consist of a 500 seconds recording of the membrane potential of a single auditory neuron from a guinea pig measured every 0.15 ms. When the membrane potential crosses a certain threshold the neuron fires, i.e. it produces a rapid electrical signal whereafter the potential resets to the resting value. Only the membrane potential values recorded between firings (inter-spike intervals, ISIs) are considered, and thus the data consist of $M = 312$ ISIs, which can be regarded as independent realizations of the same stochastic process, where n_i, the number of observations in the ith ISI, varies from few hundreds to several thousands, with a total number of observations equal to $N = \sum_{i=1}^{M} n_i \simeq 1.8 \cdot 10^6$. The OU SDMEM (29) is considered for the dynamics of the membrane potential X_t^i (see Picchini, Ditlevsen, De Gaetano and Lansky (2008) for details) with units volt $[V]$ for X_t^i, seconds $[s]$ for τ, $[V/\sqrt{s}]$ for σ, and $[V/s]$ for μ, μ^i and η. Now x_0^i can be different from zero.

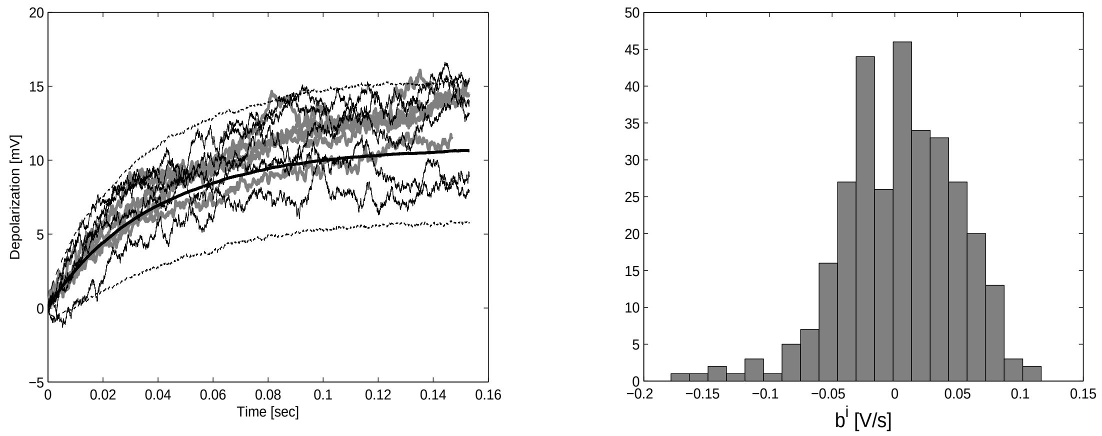
The estimates (95% CI) obtained with $(K, R) = (2, 40)$ are $\hat{\mu}^{(2,40)} = 0.494$ $[0.483, 0.506]$, $\hat{\tau}^{(2,40)} = 0.0210$ $[0.0206, 0.0215]$, $\hat{\sigma}^{(2,40)} = 0.0135$ $[0.0135, 0.0135]$, $\hat{\eta}^{(2,40)} = 0.072$ $[0.069, 0.075]$, which are within physiological plausible values. However, the data fit was not completely satisfactory, probably caused by not considering changes in τ depending on the time elicited since last spike. We repeated the maximum likelihood estimation after having fixed the value of τ to 0.039 s, as obtained in Lansky et al. (2006) by their regression method based on the first moment of the stochastic LIF model, which may be more robust to model misspecification. The estimates with τ fixed are $\hat{\mu}^{(2,40)} = 0.278$ $[0.273, 0.282]$, $\hat{\sigma}^{(2,40)} = 0.0135$ $[0.0135, 0.0135]$ and $\hat{\eta}^{(2,40)} = 0.041$ $[0.038, 0.045]$. These last results are in agreement with the regression estimates obtained in Lansky et al. (2006), where individual analyses were performed on each ISI. Their medians of the estimates were 0.285 V/s for μ and 0.0135 V/$\sqrt{\text{s}}$ for σ.

The fit obtained by fixing τ is given in Figure 5(a), reporting only five observed trajectories from the 312 ISIs, the empirical mean of 2000 trajectories simulated from model (29) according to the Euler-Maruyama scheme using the estimated parameters, the empirical 95% confidence bands of the 2000 trajectories and five simulated trajectories. For each simulated trajectory a different realization of μ^i was drawn from $\mathcal{N}(0, \eta^2)$, with $\eta = 0.0414$.

The estimates of the M random effects μ^i were obtained for each ISI from (18); the histogram of the estimates is given in Figure 5(b) with sample mean and standard deviation given by 0.0036 V/s and 0.0467 V/s, respectively. The empirical distribution seems to be close to normal, and the empirical mean and standard deviation are close to zero and to $\eta = 0.0414$ respectively, as they should be. Finally, to evaluate if the random effect on μ is statistically significant, the hypothesis $H_0 : \eta = 0$ was tested against $H_1 : \eta > 0$ in a likelihood ratio test. H_0 was rejected with $p < 0.001$, and thus we conclude that the SDMEM (29) describes the data better than the corresponding SDE model.

6 Conclusions

An approximated maximum likelihood estimation method for the parameters of mixed-effects models defined by stochastic differential equations has been proposed. The estimation method



(a) (b)

Figure 5: (a): Observations from five of the 312 ISIs (grey), empirical mean curve of 2000 trajectories of the stochastic process defined by model (29) (bold line) with their 95% confidence bands (dashed lines) and five simulated trajectories. (b): Histogram of the random effects $(\hat{\mu}^i)^{(2)}$ when τ is fixed to 0.039 s.

can be applied to models having random effects following any well-behaved distribution and can be extended to multidimensional SDMEMs. A sequence of approximations $p_X^{(K)}$ of the transition densities is constructed in closed-form, then the (approximated) likelihood can be calculated using suitable order R Gaussian quadrature schemes, available for many distributions of practical interest. For SDMEMs more complex than the ones considered here, the likelihood approximation can be obtained by taking advantage of any software with symbolic calculus capabilities.

Simulation results with $K = 1$ or 2 and $R = 40$ are promising, and can be achieved using moderate values of M (the number of experimental units, e.g. the number of subjects) and n (the number of observations for a given experimental unit). Satisfactory results are obtained even when the time-distance between observations Δ is not small, but see Stramer and Yan (2007) for possible drawbacks in the approximation provided by the transition density closed-form expansion method when Δ is "not small enough". This is relevant for applications where large data sets are unavailable, e.g. in biomedical applications, where Mixed-Effects theory is broadly applied.

When considering previously published estimation methods for SDEs with random parameters, a major drawback for their practical application is the requirement for a substantial amount of computational resources. Instead the proposed method is fast and it is possible to handle large datasets, as in Section 5.2, where few minutes are required using a MATLAB program on a single common PC (3.0 GHz Intel Pentium IV with 512 MB of RAM), therefore enabling practitioners to fit a SDMEM on their data rapidly.

In the examples we considered a simple additive relationship between a population parameter α and a random effect α^i: i.e. α and α^i entered the SDMEM as $\alpha + \alpha^i$. However, also nonlinear relations between α and α^i can be handled.

The method suffers some limitations, e.g. it may be difficult (though theoretically possible, see Aït-Sahalia (2008)) to obtain the transition density expansion for some multidimensional SDMEM systems with irreducible or non-commutative noise (Kloeden and Platen (1992, p.

348)). Moreover, it may be difficult to numerically evaluate the integral in (3) and (5) with multiple random effects, i.e. when $b^i \in B \subseteq \mathbb{R}^q$ with q larger than 2, and efficient numerical algorithms are needed. Some references are the review paper by Cools (2002), Krommer and Ueberhuber (1998) and references therein, or one of the several monographies on Monte Carlo methods (e.g. Ripley (2006)). In the mixed-effects framework the amount of literature devoted to the evaluation of q-dimensional integrals is large, e.g. the review by Davidian and Giltinan (2003), Pinheiro and Bates (1995), McCulloch and Searle (2001), Pinheiro and Chao (2006).

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

Acknowledgments

Work supported by a grant from the Danish Natural Science Research Council to S. Ditlevsen.

References

Aït-Sahalia, Y.: 2002a, Comment on G. Durham, A. Gallant, Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes, *J. Bus. Econom. Statist.* **20**, 317–321.

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

Aït-Sahalia, Y.: 2008, Closed-form likelihood expansion for multivariate diffusions, *Ann. Stat.* **36**(2), 906–937.

Andersen, K. and Højbjerre, M.: 2005, A population-based bayesian approach to the minimal model of glucose and insulin homeostasis, *Stat. Med.* **24**(15), 2381–2400.

Barrowes, B.: 2007, Multiple precision toolbox for MATLAB. `http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectI%d=6446&objectType=File`.

Beal, S., Sheiner, L. and Boeckmann, A.: 1999, *NONMEM User's Guide*, San Francisco: Division of Pharmacology, University of California.

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

Brandt, M. and Santa-Clara, P.: 2002, Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets, *J. Financ. Econ.* **63**, 161–210.

Breslow, N. and Clayton, D.: 1993, Approximate inference in generalized linear mixed models, *J. Am. Stat. Assoc.* **88**(421), 9–25.

Cools, R.: 2002, Advances in multidimensional integration, *J. Comput. Appl. Math.* **149**, 1–12.

Cox, J., Ingersoll, J. and Ross, S.: 1985, A theory of the term structure of interest rate, *Econometrica* **53**, 385–407.

Dacunha-Castelle, D. and Florens-Zmirou, D.: 1986, Estimation of the coefficients of a diffusion from discrete observations, *Stochastics* **19**, 263–284.

Davidian, M. and Giltinan, D.: 1995, *Nonlinear models for repeated measurement data*, Chapman and Hall.

Davidian, M. and Giltinan, D.: 2003, Nonlinear models for repeated measurements: an overview and update, *J. Agr. Biol. Envir. St.* **8**, 387–419.

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

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

Ditlevsen, S. and De Gaetano, A.: 2005b, Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers, *Bull. Math. Bio.* **67**(3), 547–561.

Ditlevsen, S. and Lansky, P.: 2006, Estimation of the input parameters in the Feller neuronal model, *Phys. Rev. E* **73**, Art. No. 061910.

Ditlevsen, S., Yip, K.-P., Marsh, D. and Holstein-Rathlou, N.-H.: 2007, Parameter estimation of the feedback gain in a stochastic model of renal hemodynamics: differences between spontaneously hypertensive rats and Sprague-Dawley rats, *Am. J. Physiol-Renal.* **292**, 607–616.

Donnet, S. and Samson, A.: 2008, Parametric inference for mixed models defined by stochastic differential equations, *ESAIM: P&S* **12**, 196–218.

Durham, G. and Gallant, A.: 2002, Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes, *J. Bus. Econ. Stat.* **20**(3), 297–316.

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

Feller, W.: 1951, Diffusion processes in genetics, *in* J. Neyman (ed.), *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, University of California Press, Berkeley, pp. 227–246.

Fernandes, A. and Atchley, W.: 2006, Gaussian quadrature formulae for arbitrary positive measures, *Evolutionary Bioinformatics Online* **2**, 261–269.

Fröberg, C.: 1985, *Numerical mathematics*, The Benjamin/Cummings Publishing Company, Inc.

Golub, G. and Welsch, J.: 1969, Calculation of Gauss quadrature rules, *Math. Comp.* **23**, 221–230.

Guedj, J., Thiébaut, R. and Commenges, D.: 2007, Maximum likelihood estimation in dynamical models of HIV, *Biometrics* **63**(4), 1198–1206.

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

Jelliffe, R., Schumitzky, A. and Guilder, M. V.: 2000, Population pharmacokinetics/pharmacodynamics modeling: Parametric and nonparametric methods, *Ther. Drug Monit.* **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. and Platen, E.: 1992, *Numerical solution of stochastic differential equations*, Springer.

Krommer, A. and Ueberhuber, C.: 1998, *Computational integration*, Society for Industrial and Applied Mathematics.

Kuhn, E. and Lavielle, M.: 2005, Maximum likelihood estimation in nonlinear mixed effects models, *Comput. Stat. Data An.* **49**(4), 1020–1038.

Lansky, P., Lanska, V. and Weiss, M.: 2004, A stochastic differential equation model for drug dissolution and its parameters, *J. Control. Release* **100**, 267–274.

Lansky, P., Sanda, P. and He, J.: 2006, The parameters of the stochastic leaky integrate-and-fire neuronal model, *J. Comput. Neurosci.* **21**, 211–223.

Lavielle, M., Mesa, H. and the Monolix Group: 2007, MONOLIX (MOdèles NOn LInéaires à effets miXtes). `http://www.monolix.org`.

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, *Economet. Theor.* **4**, 231–247.

McCulloch, C. and Searle, S.: 2001, *Generalized, linear and mixed models*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc.

Nicolau, J.: 2002, A new technique for simulating the likelihood of stochastic differential equations, *Economet. J.* **5**, 91–103.

Øksendal, B.: 2007, *Stochastic differential equations: an introduction with applications*, sixth edn, Springer.

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

Overgaard, R. V., Holford, N., Rytved, K. A. and Madsen, H.: 2007, PKPD model of interleukin-21 effects on thermoregulation in monkeys, *Pharmaceutical Research* **24**(2), 298–309.

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

Picchini, U., Ditlevsen, S. and De Gaetano, A.: 2006, Modeling the euglycemic hyperinsulinemic clamp by stochastic differential equations, *J. Math. Biol.* **53**(5), 771–796.

Picchini, U., Ditlevsen, S. and De Gaetano, A.: 2008, Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics, *Math. Med. Biol.* **25**(2), 141–155.

Picchini, U., Ditlevsen, S., De Gaetano, A. and Lansky, P.: 2008, Parameters of the diffusion leaky integrate-and-fire neuronal model for a slowly fluctuating signal, *Neural Comput.* **20**(11), 2696–2714.

Pinheiro, J. and Bates, D.: 1995, Approximations of the log-likelihood function in the nonlinear mixed-effects model, *J. Comput. Graph. Stat.* **4**(1), 12–35.

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

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and the R Core Team: 2007, *The* `nlme` *package*, The R Project for Statistical Computing. `http://www.r-project.org/`.

Pinheiro, J. and Chao, E.: 2006, Efficient Laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models, *J. Comput. Graph. Stat.* **15**(1), 58–81.

R Development Core Team: 2007, *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org`.

Ripley, B.: 2006, *Stochastic simulation*, Wiley.

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

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

Stramer, O. and Yan, J.: 2007, On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation, *J. Comput. Graph. Stat.* **16**(3), 672–691.

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, *Pharm. Res.* **22**(8), 1247–1258.

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

Wang, J.: 2007, EM algorithms for nonlinear mixed effects models, *Comput. Stat. Data An.* **51**(6), 3244–3256.

A Appendix

Here we report the explicit expressions for the coefficients of the log-density expansion as suggested in Aït-Sahalia (2008). The coefficients for the Geometric Brownian Motion, the OU, the CIR and the orange trees growth SDMEMs are given.

Density expansion coefficients

For given values y_j and y_{j-1} of the Y process (20) the coefficients of the log-density expansion (22) 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

In model (23) is $Y_t^i = \gamma(X_t^i) = \log(X_t^i)/\sigma$, so $\mu_Y(Y_t^i) = (\beta + \beta^i)/\sigma - \sigma/2$ and for given values y_j^i and y_{j-1}^i of Y_t^i, 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

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

OU process with one random effect: order $K = 2$ density expansion coefficients

In model (29) is $Y_t^i = \gamma(X_t^i) = X_t^i/\sigma$ so $\mu_Y(Y_t^i) = -Y_t^i/\tau + \rho$, where $\rho = (\mu + \mu^i)/\sigma$, and for given values y_j^i and y_{j-1}^i of Y_t^i, 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 $\tilde{C}^{(k)}(x_j^i | x_{j-1}^i) = C_Y^{(k)}(\frac{x_j^i}{\sigma} | \frac{x_{j-1}^i}{\sigma}), \qquad k = 0, 1, 2.$

CIR process with one random effect: order $K = 2$ density expansion coefficients

In model (31) is $Y_t^i = \gamma(X_t^i) = 2\sqrt{(\mu + \mu^i)X_t^i}/\sigma$ so $\mu_Y(Y_t^i) = (-{Y_t^i}^2 + 4\rho^2 - 1)/(2Y_t^i)$, where $\rho = (\mu + \mu^i)/\sigma$, and for given values y_j^i and y_{j-1}^i of Y_t^i, we have

$$\begin{aligned} C_Y^{(0)}(y_j^i | y_{j-1}^i) &= \frac{1}{4}({y_{j-1}^i}^2 - {y_j^i}^2) + \frac{(4\rho^2 - 1)}{2} \log\left(\frac{y_j^i}{y_{j-1}^i} \right) \\ C_Y^{(1)}(y_j^i | y_{j-1}^i) &= -\frac{48\rho^4 + y_j^i {y_{j-1}^i}^3 - 24 y_j^i y_{j-1}^i \rho^2 + {y_j^i}^2 {y_{j-1}^i}^2 + {y_j^i}^3 y_{j-1}^i - 48\rho^2 + 9}{24 y_j^i y_{j-1}^i} \\ C_Y^{(2)}(y_j^i | y_{j-1}^i) &= -\frac{{y_j^i}^2 {y_{j-1}^i}^2 + 48\rho^4 - 48\rho^2 + 9}{24 {y_j^i}^2 {y_{j-1}^i}^2} \end{aligned}$$

and

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

where $\tilde{C}^{(k)}(x_j^i|x_{j-1}^i) = C_Y^{(k)}\left(\frac{2\sqrt{(\mu+\mu^i)x_j^i}}{\sigma}\Big|\frac{2\sqrt{(\mu+\mu^i)x_{j-1}^i}}{\sigma}\right), \qquad k = 0, 1, 2.$

Orange trees growth: order $K = 2$ density expansion coefficients

In model (34) is $Y_t^i = \gamma(X_t^i) = 2\sqrt{X_t^i}/\sigma$ so $\mu_Y(Y_t^i) = Y_t^i(\phi_1 + \phi_1^i - \sigma^2 {Y_t^i}^2/4)/(2\phi_3(\phi_1+\phi_1^i)) - 1/(2Y_t^i)$, and for given values y_j^i and y_{j-1}^i of Y_t^i, we have

$$C_Y^{(0)}(y_j^i|y_{j-1}^i) = -\frac{\sigma^2({y_j^i}^4 - {y_{j-1}^i}^4)}{32\phi_3(\phi_1+\phi_1^i)} + \frac{({y_j^i}^2 - {y_{j-1}^i}^2)}{4\phi_3} - \frac{1}{2}\log\left(\frac{y_j^i}{y_{j-1}^i}\right)$$

$$\begin{aligned} C_Y^{(1)}(y_j^i|y_{j-1}^i) = & -\frac{\sigma^4\left({y_j^i}^6 + {y_j^i}^5 y_{j-1}^i + {y_j^i}^4 {y_{j-1}^i}^2 + {y_j^i}^3 {y_{j-1}^i}^3 + {y_j^i}^2 {y_{j-1}^i}^4 + y_j^i {y_{j-1}^i}^5 + {y_{j-1}^i}^6\right)}{896\phi_3^2(\phi_1+\phi_1^i)^2} \\ & + \frac{\sigma^2(10\phi_3({y_j^i}^2 + y_j^i y_{j-1}^i + {y_{j-1}^i}^2) + 3({y_j^i}^5 + {y_j^i}^2 y_{j-1}^i + {y_{j-1}^i}^3))}{240\phi_3^2(\phi_1+\phi_1^i)} \\ & - \frac{9\phi_3^2 + y_j^i y_{j-1}^i({y_j^i}^2 + y_j^i y_{j-1}^i + {y_{j-1}^i}^2)}{24 y_j^i y_{j-1}^i \phi_3^2} \end{aligned}$$

$$\begin{aligned} C_Y^{(2)}(y_j^i|y_{j-1}^i) = & -\frac{\sigma^4(5({y_j^i}^4 + {y_{j-1}^i}^4) + 8y_j^i y_{j-1}^i({y_j^i}^2 + {y_{j-1}^i}^2) + 9{y_j^i}^2{y_{j-1}^i}^2)}{896\phi_3^2(\phi_1+\phi_1^i)^2} \\ & + \frac{\sigma^2\left(9({y_j^i}^2 + {y_{j-1}^i}^2) + 12 y_j^i y_{j-1}^i + 10\phi_3\right)}{240\phi_3^2(\phi_1+\phi_1^i)} - \frac{({y_j^i}^2{y_{j-1}^i}^2 + 9\phi_3^2)}{24{y_j^i}^2{y_{j-1}^i}^2\phi_3^2} \end{aligned}$$

and

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

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