Small dispersion asymptotics for diffusion martingale estimating functions.

Michael Sørensen
Department of Theoretical Statistics
University of Copenhagen
and MaPhySto*

MPS-RR 2000-4
ISSN 1398-599X

Abstract

Martingale estimating functions provide a flexible and powerful framework for statistical inference about diffusion models based on discrete time observations. We supplement the standard results on large sample asymptotics by results on small dispersion asymptotics, which can be applied in situations where the noise term is sufficiently small, compared to the drift term, that a Gaussian approximation to the diffusion can be used. The theory, which is based on the stochastic Taylor expansion, covers proper likelihood inference too. It is remarkable that the martingale property of an estimating function also for small dispersion asymptotics ensures that estimators are consistent. A model from mathematical finance is considered in detail. For this example the range of applicability of the small dispersion asymptotics is investigated in a simulation study of the distribution of estimators.

Key Words: Asymptotic Normality; Consistency; Cox-Ingersoll-Ross model; Inference for Diffusion Processes; Likelihood Inference; Quasi Likelihood; Stochastic Taylor expansion.

*MaPhySto - Centre for Mathematical Physics and Stochastics, funded by a grant from The Danish National Research Foundation

1 Introduction

We consider statistical inference for a class of d-dimensional diffusion processes defined as the solutions of the following family of stochastic differential equations

$$dX_t = b(X_t, t; \theta)dt + \epsilon\sigma(X_t, t; \theta)dW_t, \quad X_0 = x_0, \tag{1.1}$$

where W is an m-dimensional standard Wiener process. We assume that the drift b (d-dimensional) and the diffusion coefficient σ (a $d \times m$-matrix) are known apart from the parameters θ and ϵ, of which θ varies in a subset Θ of $\mathbb{R}^p$, while $\epsilon \geq 0$.

We shall only be concerned with inference about θ. The type of asymptotics considered in this paper is when ϵ goes to zero, which can be applied in situations where the noise term $\epsilon\sigma dW$ is sufficiently small, compared to the drift term, that a Gaussian approximation to the diffusion can be used. It does not matter whether ϵ is known or not: The parameter ϵ can always be estimated by means of a quadratic estimating function; see Bibby and Sørensen (1996).

The type of data to be considered are observations of X at discrete time points: $X_{t_0}, X_{t_1}, \cdots, X_{t_n}$, $t_0 = 0 < t_1 < \cdots < t_n$. Data of this form is what is typically met in statistical practice. For such data inference based on martingale estimating functions has been considered by Bibby (1994), Pedersen (1994), Bibby and Sørensen (1995, 1996, 1998), Kessler (1995), Kessler and Sørensen (1999), Sørensen (1997), and Jacobsen (1998). An application to financial data was given in Bibby and Sørensen (1987), while Pedersen (1999) used the method to estimate the nitrous oxide emission rate from the soil surface. In these papers, large sample asymptotics, where the number of observations tends to infinity, was studied for ergodic diffusions. It is, however, useful to develop alternative types of asymptotics that can be applied when the number of observations is small, or when the diffusion process is not ergodic. The present paper is a contribution in this direction. Note that the small dispersion asymptotics investigated in this paper is immediately applicable to multidimensional diffusions, whereas large sample asymptotics is harder for multivariate diffusions than in the one-dimensional case. Very little work has been done in this direction. The problem is that the number of ways a diffusion process can behave asymptotically as the sample size goes to infinity is much larger in higher dimensions than in the one-dimensional case. A final virtue of the small dispersion asymptotics is that there is scope for higher order asymptotics. Under weak conditions there is an expansion of the process X to any order, of which only the first two terms are used in this paper. The necessary results on stochastic Taylor expansions can be found in Azencott (1982). For an approach based on Malliavin calculus, see Yoshida (1992, 1997).

Let $y \mapsto p_\epsilon(s,t,x,y;\theta)$, where $t > s$, denote the transition density of the Markov process X, i.e. the density of X_t given $X_s = x$ when θ and ϵ are the true parameter values. We shall, in line with the papers mentioned, consider martingale estimating function of the form

$$G_\epsilon(\theta) = \sum_{i=1}^{n} g_\epsilon(t_{i-1}, t_i, X_{t_{i-1}}, X_{t_i}; \theta), \tag{1.2}$$

where the function $g_\epsilon(s,t,x,y;\theta)$ satisfies the equation

$$\int g_\epsilon(s,t,x,y;\theta) p_\epsilon(s,t,x,y;\theta) dy = 0 \tag{1.3}$$

for all x, θ, s, and t with $s < t$. Define the σ-algebras $\mathcal{F}_i = \sigma(X_{t_1}, \cdots, X_{t_i})$, $i = 1, \cdots, n$. Then (1.3) implies that $\{\sum_{j=1}^{i} g_\epsilon(t_{j-1}, t_j, X_{t_{j-1}}, X_{t_j}; \theta)\}_{i=1}^{n}$ is a martingale with respect to $\{\mathcal{F}_i\}_{i=1}^{n}$ when θ and ϵ are the true parameter values. The score function, i.e. the derivative with respect to θ of the log-likelihood function,

$$U_n(\theta) = \sum_{i=1}^{n} \partial_\theta \log p_\epsilon(t_{i-1}, t_i, X_{t_{i-1}}, X_{t_i}; \theta), \tag{1.4}$$

where $\partial_\theta \log p_\epsilon$ denotes the vector of partial derivatives of $\log p_\epsilon$ with respect to θ, is of the form (1.2). Under weak regularity conditions, $g_\epsilon = \partial_\theta \log p_\epsilon$ satisfies (1.3), see e.g. Barndorff-Nielsen and Sørensen (1994). Therefore the results derived in the present paper also apply to likelihood inference. Likelihood inference for discretely observed diffusions was considered by Dacunha-Castelle and Florens-Zmirou (1986), Lo (1988), Pedersen (1995a,b), Santa-Clara (1997), Aït-Sahalia (1998, 1999) and Poulsen (1999); see also Billingsley (1961). Bayesian inference and related MCMC techniques were studied by Eraker (1997) and Elerian, Chib and Shephard (1998). Genon-Catalot (1990) used the stochastic Taylor expansion to construct a contrast function for discretely observed diffusions. She also studied small dispersion asymptotics for her estimators. Kutoyants (1984) considered small dispersion asymptotics for continuously observed diffusions.

In Section 2, we review results on the stochastic Taylor expansion and give a couple of results needed later. In Section 3, we prove asymptotic normality (as $\epsilon \to 0$) of martingale estimating functions and a result on existence, consistency and asymptotic normality of an estimator obtained from a martingale estimating function. It is remarkable that the martingale property (1.3) of an estimating function ensures consistency of estimators under small dispersion asymptotics as it does for large sample asymptotics. In Section 4, an example from mathematical finance is considered, and in a simulation study of the distribution of estimators for this model, the range of applicability of the small dispersion asymptotics is investigated.

2 Stochastic Taylor expansions

In this section we discuss a stochastic Taylor expansion of the solution X^ϵ of the stochastic differential equation (1.1). When we want to emphasize the dependence of the solution on ϵ, we write X^ϵ. The functions b and σ appearing in (1.1) are assumed to be sufficiently smooth that a unique weak solution exists for all $\theta \in \Theta$ and all $\epsilon > 0$. For $\epsilon = 0$ it is assumed that the deterministic equation has a unique solution for all $\theta \in \Theta$. Further, it is supposed that $b(x,t;\theta)$ and $\sigma(x,t;\theta)$ are twice continuously differentiable with respect to x for every t and θ, and that the partial derivatives up to order two with respect to x are continuous functions of t. It is assumed that the solution of (1.1) does not explode at a finite time for any $\theta \in \Theta$ or any $\epsilon \geq 0$. Finally, we assume that the state space is an open subset U of $\mathbb{R}^d$.

Under these conditions, we have the following first order stochastic Taylor expansion of X^ϵ

$$X_t^\epsilon = \xi_t^\theta + \epsilon Y_t^\theta + \epsilon^2 R_t^{\theta,\epsilon}; \tag{2.1}$$

see Azencott (1982). In (2.1) the d-dimensional deterministic function $\xi^\theta = (\xi_1^\theta, \cdots, \xi_d^\theta)^T$, where T denotes transposition, is the unique solution of the equation (1.1) with $\epsilon = 0$, i.e.

$$\frac{d\xi_t^\theta}{dt} = b(\xi_t^\theta, t; \theta), \tag{2.2}$$

with $\xi_0^\theta = x_0$. The d-dimensional process Y^θ is, when θ is the true parameter value, the Gaussian diffusion which solves

$$dY_t^\theta = \partial_{x^T} b(\xi_t^\theta, t; \theta) Y_t^\theta dt + \sigma(\xi_t^\theta, t; \theta) dW_t, \quad Y_0^\theta = 0, \tag{2.3}$$

where $\partial_{x^T} b$ denotes the $d \times d$ matrix $\{\partial_{x_j} b_i\}$ of partial derivatives of the coordinates of b with respect to the coordinates of x. The remainder term satisfies $\sup_{s \leq t} |\epsilon R_s^{\theta,\epsilon}| \to 0$ in probability as $\epsilon \to 0$ for all $t > 0$ when θ is the true parameter value. More precisely, there exist for every given $\delta > 0$ constants $r_0 > 0$ and $\epsilon_0 > 0$ such that $P_{\theta,\epsilon}(\sup_{s \leq t} |R_s^{\theta,\epsilon}| > r) \leq \delta$ when $r \geq r_0$ and $\epsilon \leq \epsilon_0$. Under additional conditions on the growths rate of b and σ and their derivatives as functions of x, we have that for every $\theta \in \Theta$ and $k \in \mathbb{N}$ there exists a constant $K_{\theta,k}$ such that $E_{\theta,\epsilon}(\sup_{s \leq t} |R_s^{\theta,\epsilon}|^k) < K_{\theta,k}$ for all $\epsilon \leq 1$; see Azencott (1982). We will assume this to be the case. A sufficient condition is the usual local Lipschitz and linear growth conditions on b and σ.

Let the $d \times d$ matrix function $H_\theta(t)$ be the solution of the matrix differential equation

$$\frac{dH_\theta}{dt}(t) = \partial_{x^T} b(\xi_t^\theta, t; \theta) H_\theta(t), \quad H_\theta(0) = I_d, \tag{2.4}$$

where I_d is the $d \times d$ identity matrix. Then the Gaussian process Y^θ is given by

$$Y_t^\theta = H_\theta(t) \int_0^t H_\theta(s)^{-1} \sigma(\xi_s^\theta, s; \theta) dW_s. \tag{2.5}$$

Therefore, the distribution of the nd-dimensional stochastic vector $Y = (Y_{t_1}^T, \cdots, Y_{t_n}^T)^T$ is

$$Y \sim N(0, \Gamma^\theta), \tag{2.6}$$

where Γ^θ is the $nd \times nd$ matrix consisting of n^2 $d \times d$ matrices Γ_{ij}^θ of the form

$$\Gamma_{ij}^\theta = H_\theta(t_i) \int_0^{t_i \wedge j} H_\theta(s)^{-1} \sigma(\xi_s^\theta, s; \theta) \sigma(\xi_s^\theta, s; \theta)^T (H_\theta(s)^{-1})^T ds \, H_\theta(t_j)^T,$$

for $i, j = 1, \cdots, n$.

In the time homogeneous case, where the function b does not depend on t, it is sometimes possible to find a relatively simple expression for $H_\theta(t)$. In the one-dimensional case ($d = 1$), it is well known that quite generally $H_\theta(t) = \exp\left[\int_0^t \partial_x b(\xi_s^\theta; \theta) ds\right]$. For a d-dimensional process, consider the special case where, for all i, the i'th coordinate of b depends on x only through the i'th coordinate of x. This implies that $\partial_{x^T} b = \text{diag}(\partial_{x_1} b_1, \cdots, \partial_{x_d} b_d)$. Here $\text{diag}(a_1, \cdots, a_d)$ denotes the diagonal matrix with diagonal elements $a_1, \cdots, a_d$. Suppose that $b_i(\xi_{i,t}^\theta; \theta) \neq 0$ for all $t \geq 0$ and $i = 1, \cdots, d$. Then

$$H_\theta(t) = \text{diag}\{b_1(\xi_{1,t}^\theta; \theta)/b_1(\xi_{1,0}^\theta; \theta), \cdots, b_d(\xi_{d,t}^\theta; \theta)/b_d(\xi_{d,0}^\theta; \theta)\}$$

obviously satisfies the differential equation (2.4). Note that under the conditions imposed $b_i(\xi_{i,t}^\theta; \theta)/b_i(\xi_{i,0}^\theta; \theta) > 0$. In the one-dimensional case, we always have that $H_\theta(t) = b(\xi_t^\theta; \theta)/b(\xi_0^\theta; \theta)$ provided only that $b(\xi_t^\theta; \theta) \neq 0$ for all $t \geq 0$. The condition that $b_i(\xi_{i,t}^\theta; \theta) \neq 0$ for all $t \geq 0$ is not a strong restriction. In fact, if a time point τ exists such that $b_i(\xi_{i,\tau}^\theta; \theta) = 0$, then it is easy to see that $\xi_{i,t}^\theta = 0$ for all $t \geq \tau$.

In the next section, the following lemma will play a crucial role. By $\xi_t^\theta(z)$ we denote the solution of (2.2) with the initial condition $\xi_0^\theta = z$, and by $Y_t^\theta(z)$ we denote the solution of (2.3) with ξ_t^θ replaced by $\xi_t^\theta(z)$.

Lemma 2.1 *Suppose that for every $\epsilon \in [0,1]$ the function $F_\epsilon : U \mapsto \mathbb{R}$ is twice differentiable and that there exist constants $C, k > 0$ such that $|\partial_{x_i} \partial_{x_j} F_\epsilon(x)| \leq C|x|^k$ for all $\epsilon \in [0,1]$ and for $i, j = 1, \cdots, d$. Then*

$$E_{\theta,\epsilon}(F_\epsilon(X_\Delta^\epsilon) | X_0^\epsilon = z) = F_\epsilon(\xi_\Delta^\theta(z)) + O(\epsilon^2) \tag{2.7}$$

for $\epsilon \in [0,1]$ and $\Delta > 0$.

Proof: The result follows from the expansion

$$F_\epsilon(X_\Delta^\epsilon) = F_\epsilon(\xi_\Delta^\theta(z)) + \epsilon \partial_{x^T} F_\epsilon(\xi_\Delta^\theta(z)) Y_\Delta^\theta(z) + \epsilon^2 A_{\Delta,\epsilon}^\theta,$$

where

$$\begin{aligned} A_{\Delta,\epsilon}^\theta &= \partial_{x^T} F_\epsilon(\xi_\Delta^\theta(z)) R_\Delta^{\theta,\epsilon} \\ &+ \tfrac{1}{2}(Y_\Delta^\theta(z) + \epsilon R_\Delta^{\theta,\epsilon})^T \frac{\partial^2 F_\epsilon}{\partial x \partial x^T}(\xi_\Delta^\theta(z) + \alpha\{\epsilon Y_\Delta^\theta(z) + \epsilon^2 R_\Delta^{\theta,\epsilon}\})(Y_\Delta^\theta(z) + \epsilon R_\Delta^{\theta,\epsilon}) \end{aligned}$$

for some $\alpha \in [0,1]$. Under the conditions imposed, $\sup_{\epsilon \le 1} E_{\theta,\epsilon}(|A_{\Delta,\epsilon}^\theta||X_0^\epsilon = z) < \infty$. The covariance structure of $Y^\theta(z)$ depends on z through $\xi^\theta(z)$, but its expectation is zero whatever the value of z is.

□

3 Martingale estimating functions

In this section we prove asymptotic normality of martingale estimating functions as $\epsilon \to 0$, and show that in this limit a consistent and asymptotically normal estimator exists with a probability tending to one. Asymptotic normality of G_ϵ is proved under the following condition.

Condition 3.1 *The function $(x,y) \mapsto g_\epsilon(s,t,x,y;\theta)$ is twice continuously differentiable for all $\epsilon \ge 0$, $t > s > 0$, and $\theta \in \Theta$, and all first and second order derivatives are continuous functions of $(\epsilon, x, y) \in [0,1] \times U \times U$. For every (s,t,x,θ), there exist positive constants $C(s,t,x,\theta)$ and $k(s,t,x,\theta)$ such that $|\partial_{y_i}\partial_{y_j} g_\epsilon(s,t,x,y;\theta)_l| \le C(s,t,x,\theta)|y|^{k(s,t,x,\theta)}$ for all $\epsilon \in [0,1]$, for $i,j = 1,\cdots,d$, and for $l = 1,\cdots,p$.*

Define the $p \times d$ matrix $\phi_i(\theta)$ by

$$\phi_i(\theta) = \partial_{x^T} g_0(t_i, t_{i+1}, \xi_{t_i}^\theta, \xi_{t_{i+1}}^\theta; \theta) + \partial_{y^T} g_0(t_{i-1}, t_i, \xi_{t_{i-1}}^\theta, \xi_{t_i}^\theta; \theta)$$

for $i = 1, \cdots, n-1$, and

$$\phi_n(\theta) = \partial_{y^T} g_0(t_{n-1}, t_n, \xi_{t_{n-1}}^\theta, \xi_{t_n}^\theta; \theta).$$

The estimating function $K_\epsilon(\theta) = M_\epsilon G_\epsilon(\theta)$, where M_ϵ is an invertible matrix dependent on ϵ, give the same estimators as $G_\epsilon(\theta)$, so there is a problem of choosing the right version of the estimating function. If the wrong normalization is chosen, it could, for instance, happen that $\partial_{x^T} g_0 = \partial_{y^T} g_0 = 0$, in which case the following theorem is not terribly interesting. This problem is, of course, not really of importance in practice, but is only relevant to the arguments in this section. We choose a version of $G_\epsilon(\theta)$ such that as many entries as possible of the $p \times nd$ matrix $\phi(\theta) = (\phi_1(\theta), \cdots, \phi_n(\theta))$ are not zero. This together with the continuity conditions at $\epsilon = 0$ imposed in Condition 3.1 usually determine a unique version of $G_\epsilon(\theta)$.

Theorem 3.2 *Under Condition 3.1,*

$$\epsilon^{-1} G_\epsilon(\theta) \to \sum_{i=1}^{n} \phi_i(\theta) Y_{t_i} \tag{3.1}$$

in probability as $\epsilon \to 0$ when θ is the true parameter value. In particular,

$$\epsilon^{-1} G_\epsilon(\theta) \to N(0, V_\theta) \tag{3.2}$$

in distribution as $\epsilon \to 0$, where $V_\theta = \phi(\theta)^T \Gamma^\theta \phi(\theta)$.

Proof: By expanding g_ϵ and inserting (2.1) we obtain

$$\begin{aligned}
& g_\epsilon(t_{i-1}, t_i, X^\epsilon_{t_{i-1}}, X^\epsilon_{t_i}; \theta) \\
& \quad = g_\epsilon(t_{i-1}, t_i, \xi^\theta_{t_{i-1}}, \xi^\theta_{t_i}; \theta) + \epsilon \partial_{x^T} g_\epsilon(t_{i-1}, t_i, \xi^\theta_{t_{i-1}}, \xi^\theta_{t_i}; \theta) Y^\theta_{t_{i-1}} \\
& \quad + \epsilon \partial_{y^T} g_\epsilon(t_{i-1}, t_i, \xi^\theta_{t_{i-1}}, \xi^\theta_{t_i}; \theta) Y^\theta_{t_i} + \epsilon^2 R^\epsilon_{g,i},
\end{aligned}$$

where the jth coordinate of $R^\epsilon_{g,i}$ is

$$\begin{aligned}
& \partial_{x^T} g_\epsilon(t_{i-1}, t_i, \xi^\theta_{t_{i-1}}, \xi^\theta_{t_i}; \theta)_j R^{\theta,\epsilon}_{t_{i-1}} + \partial_{y^T} g_\epsilon(t_{i-1}, t_i, \xi^\theta_{t_{i-1}}, \xi^\theta_{t_i}; \theta)_j R^{\theta,\epsilon}_{t_i} \\
& \quad + \tfrac{1}{2} Z^T_{1,i} \partial_x \partial_{x^T} g_\epsilon(t_{i-1}, t_i, \xi^\theta_{t_{i-1}} + \epsilon \alpha Z_{1,i}, \xi^\theta_{t_i}; \theta)_j Z_{1,i} \\
& \quad + \tfrac{1}{2} Z^T_{2,i} \partial_y \partial_{y^T} g_\epsilon(t_{i-1}, t_i, \xi^\theta_{t_{i-1}}, \xi^\theta_{t_i} + \epsilon \beta Z_{2,i}; \theta)_j Z_{2,i} \\
& \quad + \tfrac{1}{2} Z^T_{1,i} \partial_x \partial_{y^T} g_\epsilon(t_{i-1}, t_i, \xi^\theta_{t_{i-1}} + \epsilon \gamma Z_{1,i}, \xi^\theta_{t_i}; \theta)_j Z_{2,i}
\end{aligned}$$

for some $\alpha, \beta, \gamma \in [0, 1]$. Here $Z_{1,i} = Y^\theta_{t_{i-1}} + \epsilon R^{\theta,\epsilon}_{t_{i-1}}$ and $Z_{2,i} = Y^\theta_{t_i} + \epsilon R^{\theta,\epsilon}_{t_i}$. From the similar property of $R^{\theta,\epsilon}_t$, it follows that $\epsilon R^\epsilon_{g,i} \to 0$ in probability as $\epsilon \to 0$ when θ is the true parameter value. Moreover, $E_{\theta,\epsilon}(g_\epsilon(s, t, z, X^\epsilon_t; \theta) | X^\epsilon_s = z) = 0$, so by Lemma 2.1, we see that $g_\epsilon(s, t, z, \xi^\theta_{t-s}(z, s); \theta) = O(\epsilon^2)$, where $\xi^\theta_t(z, s)$ denotes the solution of the equation

$$\frac{d\xi^\theta_t(z, s)}{dt} = b(\xi^\theta_t(z, s), t + s; \theta)$$

with the initial condition $\xi^\theta_0(z, s) = z$. Since $\xi^\theta_{t-s}(\xi^\theta_s(x_0), s) = \xi^\theta_t(x_0)$, it follows that

$$\begin{aligned}
& \epsilon^{-1} g_\epsilon(t_{i-1}, t_i, X^\epsilon_{t_{i-1}}, X^\epsilon_{t_i}; \theta) \to \\
& \quad \partial_{x^T} g_0(t_{i-1}, t_i, \xi^\theta_{t_{i-1}}, \xi^\theta_{t_i}; \theta) Y^\theta_{t_{i-1}} + \partial_{y^T} g_0(t_{i-1}, t_i, \xi^\theta_{t_{i-1}}, \xi^\theta_{t_i}; \theta) Y^\theta_{t_i}
\end{aligned}$$

in probability as $\epsilon \to 0$, which implies (3.1) and (3.2). $\square$

Note that a crucial step in the proof was the application of Lemma 2.1 to prove that $g_\epsilon(t_{i-1}, t_i, \xi^\theta_{t_{i-1}}, \xi^\theta_{t_i}; \theta) = O(\epsilon^2)$. As is the case for large sample asymptotics, see e.g. Sørensen (1997), the martingale property of the estimating function $G_\epsilon(\theta)$ ensures that the asymptotic distribution of $G_\epsilon(\theta)$, properly normalized, has mean value zero when θ is the true parameter value.

Suppose $\partial_\theta g$ exists and is continuous for $(\epsilon, x, y) \in [0,1] \times U \times U$. Then if θ_0 is the true value of θ,

$$\partial_{\theta^T} G_\epsilon(\theta) \to A(\theta, \theta_0) \tag{3.3}$$

in probability as $\epsilon \to 0$, where

$$A(\theta, \tilde{\theta}) = \sum_{i=1}^{n} \partial_{\theta^T} g_0(t_{i-1}, t_i, \xi^{\tilde{\theta}}_{t_{i-1}}, \xi^{\tilde{\theta}}_{t_i}; \theta). \tag{3.4}$$

In order to obtain results about estimators, we need the following stronger condition.

Condition 3.3 *The derivatives $\partial_\theta g$, $\partial^2_\theta g$, $\partial_x \partial_\theta g$, $\partial_x \partial^2_\theta g$, $\partial_y \partial_\theta g$, $\partial_y \partial^2_\theta g$, $\partial_\epsilon \partial_\theta g$ and $\partial_\epsilon \partial^2_\theta g$ exist and are continuous for $(\epsilon, x, y, \theta) \in [0,1] \times U \times U \times \Theta$. The matrix $A(\theta) = A(\theta, \theta)$, with $A(\theta, \tilde{\theta})$ given by (3.4), is invertible for all $\theta \in \Theta$.*

Theorem 3.4 *Suppose the Conditions 3.1 and 3.3 hold. Then for every $\epsilon \leq 1$, there exists an estimator $\hat{\theta}_\epsilon$ that solves the estimating equation $G_\epsilon(\hat{\theta}_\epsilon) = 0$ with a probability tending to one as $\epsilon \to 0$. Moreover, if θ_0 denotes the true value of θ, and if $\theta_0 \in \operatorname{int} \Theta$, then*

$$\hat{\theta}_\epsilon \to \theta_0 \tag{3.5}$$

in probability as $\epsilon \to 0$, and

$$\epsilon^{-1}(\hat{\theta}_\epsilon - \theta_0) \to N(0, A(\theta_0)^{-1} V_{\theta_0} (A(\theta_0)^{-1})^T) \tag{3.6}$$

in distribution as $\epsilon \to 0$.

Proof: First note that the proofs of Theorem 2.3, Corollary 2.7 and Theorem 2.8 in Sørensen (1998) do not use the fact that the particular type of asymptotics considered in that paper is large sample asymptotics. The result also holds for other types of asymptotics such as the small dispersion asymptotics considered here. Therefore, Theorem 3.4 follows if we verify Condition 2.6 in Sørensen (1998) or rather a form of these conditions adapted in an obvious way to the small dispersion asymptotics.

Let the estimating function $G_\epsilon(\theta)$ be normalized as discussed earlier. That $\{\epsilon^{-1}G_\epsilon(\theta_0) : \epsilon \in (0,1]\}$ is stochastically bounded has already been established in Theorem 3.2, so it remains to prove that

$$\sup_{\theta \in M_{c,\epsilon}} |G_\epsilon(\theta)_i| \to 0, \tag{3.7}$$

$$\sup_{\theta \in M_{c,\epsilon}} |\partial_{\theta_j} G_\epsilon(\theta)_i - A(\theta_0)_{ij}| \to 0, \tag{3.8}$$

and

$$\sup_{\theta \in M_{c,\epsilon}} |\partial_{\theta_i}\partial_{\theta_j} G_\epsilon(\theta)_k - B^{(k)}(\theta_0)_{ij}| \to 0 \tag{3.9}$$

in probability as $\epsilon \to 0$ for $i,j,k = 1,\ldots,p$ and for all $c > 0$, when θ_0 is the true parameter value. Here

$$M_{c,\epsilon} = \{\theta \in \Theta : |\theta - \theta_0| \le c\epsilon\}$$

and

$$B^{(k)}(\theta_0) = \sum_{\nu=1}^{n} \partial_\theta \partial_{\theta^T} g_0(t_{\nu-1}, t_\nu, \xi^{\theta_0}_{t_{\nu-1}}, \xi^{\theta_0}_{t_\nu}; \theta_0).$$

The three convergence results can be proved in the same way, so we prove only (3.8). Since

$$\begin{aligned} &\sup_{\theta \in M_{c,\epsilon}} |\partial_{\theta_j} G_\epsilon(\theta)_i - A(\theta_0)_{ij}| \le \\ &\qquad \sup_{\theta \in M_{c,1}} |\partial_{\theta_j} G_\epsilon(\theta)_i - A(\theta,\theta_0)_{ij}| + \sup_{\theta \in M_{c,\epsilon}} |A(\theta,\theta_0)_{ij} - A(\theta_0)_{ij}|, \end{aligned}$$

and since $\theta \mapsto A(\theta, \theta_0)$ is continuous under the conditions imposed, (3.8) follows if we can prove that for every compact subset $K \subseteq \Theta$

$$\sup_{\theta \in K} |\partial_{\theta_j} G_\epsilon(\theta)_i - A(\theta,\theta_0)_{ij}| \to 0 \tag{3.10}$$

in probability as $\epsilon \to 0$, when θ_0 is the true value of θ. An expansion gives us

$$\begin{aligned} D_1^{\nu,\epsilon} &= \sup_{\theta \in K} |\partial_{\theta_j} g_\epsilon(t_{\nu-1}, t_\nu, X^\epsilon_{t_{\nu-1}}, X^\epsilon_{t_\nu}; \theta)_i - \partial_{\theta_j} g_\epsilon(t_{\nu-1}, t_\nu, \xi^{\theta_0}_{t_{\nu-1}}, \xi^{\theta_0}_{t_\nu}; \theta)_i| \\ &\le \epsilon \sup_{\theta \in K}(|\partial_x \partial_{\theta_j} g_\epsilon(t_{\nu-1}, t_\nu, \xi^{\theta_0}_{t_{\nu-1}} + \alpha\epsilon \tilde{R}^\epsilon_{t_{\nu-1}}, \xi^{\theta_0}_{t_\nu} + \epsilon \tilde{R}^\epsilon_{t_\nu}; \theta)_i||\tilde{R}^\epsilon_{t_{\nu-1}}| \\ &\quad + |\partial_y \partial_{\theta_j} g_\epsilon(t_{\nu-1}, t_\nu, \xi^{\theta_0}_{t_{\nu-1}}, \xi^{\theta_0}_{t_\nu} + \beta\epsilon\tilde{R}^\epsilon_{t_\nu}; \theta)||\tilde{R}^\epsilon_{t_\nu}|) \end{aligned}$$

for certain α and β in $[0,1]$. Here $\tilde{R}^\epsilon_t = Y^{\theta_0}_t + \epsilon R^{\theta_0,\epsilon}_t$. Fix $\delta > 0$. Then there exist $r > 0$ and $\epsilon_0 > 0$ such that

$$P\left(\sup_{s \le t_n} |\tilde{R}^\epsilon_s| \le r\right) > 1 - \delta$$

for $\epsilon \le \epsilon_0$. Define $\mathcal{A}_{\epsilon,r} = \{\sup_{s\le t_n} |\tilde{R}^\epsilon_s| \le r\}$ and $L^\nu_{\epsilon_0,r} = r \sup |\partial_x \partial_{\theta_j} g_\epsilon(t_{\nu-1}, t_\nu, \xi^{\theta_0}_{t_{\nu-1}} + x, \xi^{\theta_0}_{t_\nu} + y; \theta)_i| \vee |\partial_y \partial_{\theta_j} g_\epsilon(t_{\nu-1}, t_\nu, \xi^{\theta_0}_{t_{\nu-1}}, \xi^{\theta_0}_{t_\nu} + y; \theta)_i|$, where the supremum is over $(\epsilon, x, y, \theta) \in [0, r^{-1} \wedge \epsilon_0] \times [-1,1] \times [-1,1] \times K$. Then $D^{\nu,\epsilon}_1 \le \epsilon L^\nu_{\epsilon_0,r}$ on the set $\mathcal{A}_{\epsilon,r}$ provided that $\epsilon \le \epsilon_0 \wedge r^{-1}$. Moreover,

$$\sup_{\theta\in K} |\partial_{\theta_j} g_\epsilon(t_{\nu-1}, t_\nu, \xi^{\theta_0}_{t_{\nu-1}}, \xi^{\theta_0}_{t_\nu}; \theta)_i - \partial_{\theta_j} g_0(t_{\nu-1}, t_\nu, \xi^{\theta_0}_{t_{\nu-1}}, \xi^{\theta_0}_{t_\nu}; \theta)_i| \le \epsilon K_\nu$$

for $\epsilon \in [0,1]$, where

$$K_\nu = \sup_{(\theta,\epsilon)\in K\times[0,1]} |\partial_{\theta_j} \partial_\epsilon g_\epsilon(t_{\nu-1}, t_\nu, \xi^{\theta_0}_{t_{\nu-1}}, \xi^{\theta_0}_{t_\nu}; \theta)_i|.$$

Therefore,

$$\sup_{\theta\in K} |\partial_{\theta_j} G_\epsilon(\theta)_i - A(\theta,\theta_0)_{ij}| \le \epsilon \sum_{\nu=1}^n (L^\nu_{\epsilon_0,r} + K_\nu)$$

on $\mathcal{A}_{\epsilon,r}$ when $\epsilon \le \epsilon_0 \wedge r^{-1} \wedge 1$. This proves (3.10).

□

4 An example from mathematical finance

In this section we consider the Cox, Ingersoll and Ross (1985) model that is used in mathematical finance to model interest rates. The model is given by the stochastic differential equation

$$dX_t = -\beta(X_t - \alpha)dt + \epsilon\sqrt{X_t} dW_t, \tag{4.1}$$

where $\theta = (\alpha, \beta) \in (0,\infty)^2$ and $\epsilon > 0$. It is easily found that

$$\xi^\theta_t = x_0 e^{-\beta t} + \alpha(1 - e^{-\beta t}),$$

$$Y^\theta_t = \int_0^t e^{-\beta(t-s)} \sqrt{\xi^\theta_s} dW_s,$$

and

$$\Gamma^\theta_{ij} = e^{-\beta(t_i+t_j)} \beta^{-1} [x_0(e^{\beta t_{i\wedge j}} - 1) + \alpha(\tfrac{1}{2} e^{2\beta t_{i\wedge j}} - e^{\beta t_{i\wedge j}} + \tfrac{1}{2})].$$

Note that when t is large ξ^θ_t is close to α, so for the Gaussian approximation $Z_t = \xi^\theta_t + \epsilon Y^\theta_t$ to the Cox, Ingersoll and Ross model we have $dZ_t \doteq -\beta(Z_t - \alpha)dt + \epsilon\sqrt{\alpha} dW_t$, i.e. Z_t behaves like an Ornstein-Uhlenbeck process when t is large. In mathematical finance the latter model is known as the Vasiček model for interest rates.

A martingale estimating function that yields explicit estimators for α and β is

$$G_\epsilon(\alpha,\beta) = \begin{bmatrix} \sum_{i=1}^n X_{t_{i-1}}^{-1}[X_{t_i} - \xi_{\Delta_i}^\theta(X_{t_{i-1}})] \\ \\ \sum_{i=1}^n [X_{t_i} - \xi_{\Delta_i}^\theta(X_{t_{i-1}})] \end{bmatrix}, \tag{4.2}$$

where $\xi_t^\theta(z) = ze^{-\beta t} + \alpha(1-e^{-\beta t})$; see Bibby and Sørensen (1995). For this estimating function, we find that

$$\phi_i(\theta) = \begin{bmatrix} (\xi_{t_{i-1}}^\theta)^{-1} - e^{-\beta\Delta_{i+1}}/\xi_{t_i}^\theta \\ \\ 1 - e^{-\beta\Delta_{i+1}} \end{bmatrix},$$

for $i = 1, \cdots, n-1$,

$$\phi_n(\theta) = \begin{bmatrix} (\xi_{t_{n-1}}^\theta)^{-1} \\ \\ 1 \end{bmatrix},$$

and

$$A(\theta,\tilde\theta) = \sum_{i=1}^n \begin{bmatrix} (e^{-\beta\Delta_i}-1)/\xi_{t_{i-1}}^{\tilde\theta} & \Delta_i e^{-\beta\Delta_i}(1-\alpha/\xi_{t_{i-1}}^{\tilde\theta}) \\ \\ (e^{-\beta\Delta_i}-1) & \Delta_i e^{-\beta\Delta_i}(\xi_{t_{i-1}}^{\tilde\theta}-\alpha) \end{bmatrix}.$$

The range of applicability of the small dispersion asymptotics for this example is investigated for two values of the parameters α and β. In both cases 500 independent sample paths of 300 observations each were simulated with $\Delta_i = 0.1$ (equidistant observation times) for four values of ϵ and two values of the initial value x_0. The simulations were done using the Milstein scheme, see Kloeden and Platen (1992). For each sample path the estimators $\hat\alpha$ and $\hat\beta$ were calculated, and the resulting 500 values of the estimators were used to calculate the mean and the standard deviation of the estimators. The means should be compared to the true parameter values, while the standard deviations can be compared to the theoretical values given by the small dispersion asymptotics, i.e. by (3.6).

First the parameter values $\alpha = 5$ and $\beta = 2$ were considered (Table 4.1 and Table 4.2). The small dispersion asymptotics gives a very good approximation to the standard deviation of $\hat\alpha$ in all cases, while this type of asymptotics can clearly not be used to calculate the standard deviation of $\hat\beta$ when $\epsilon \geq 0.05$. The bias of $\hat\alpha$ is small in all cases, whereas $\hat\beta$ has a considerable bias when ϵ is not sufficiently small. The normality of the estimators is studied in Figure 4.1 and Figure 4.2, which show normal quantile plots of the 500 simulated estimator values for $x_0 = 5.2$ with $\epsilon = 0.01$ and $\epsilon = 0.005$, respectively. The lines represent the limiting normal distributions obtained by small dispersion asymptotics. The normal approximations are rather satisfactory apart from the small inaccuracies of the mean and standard deviation already seen in Table 4.1 and Table 4.2.

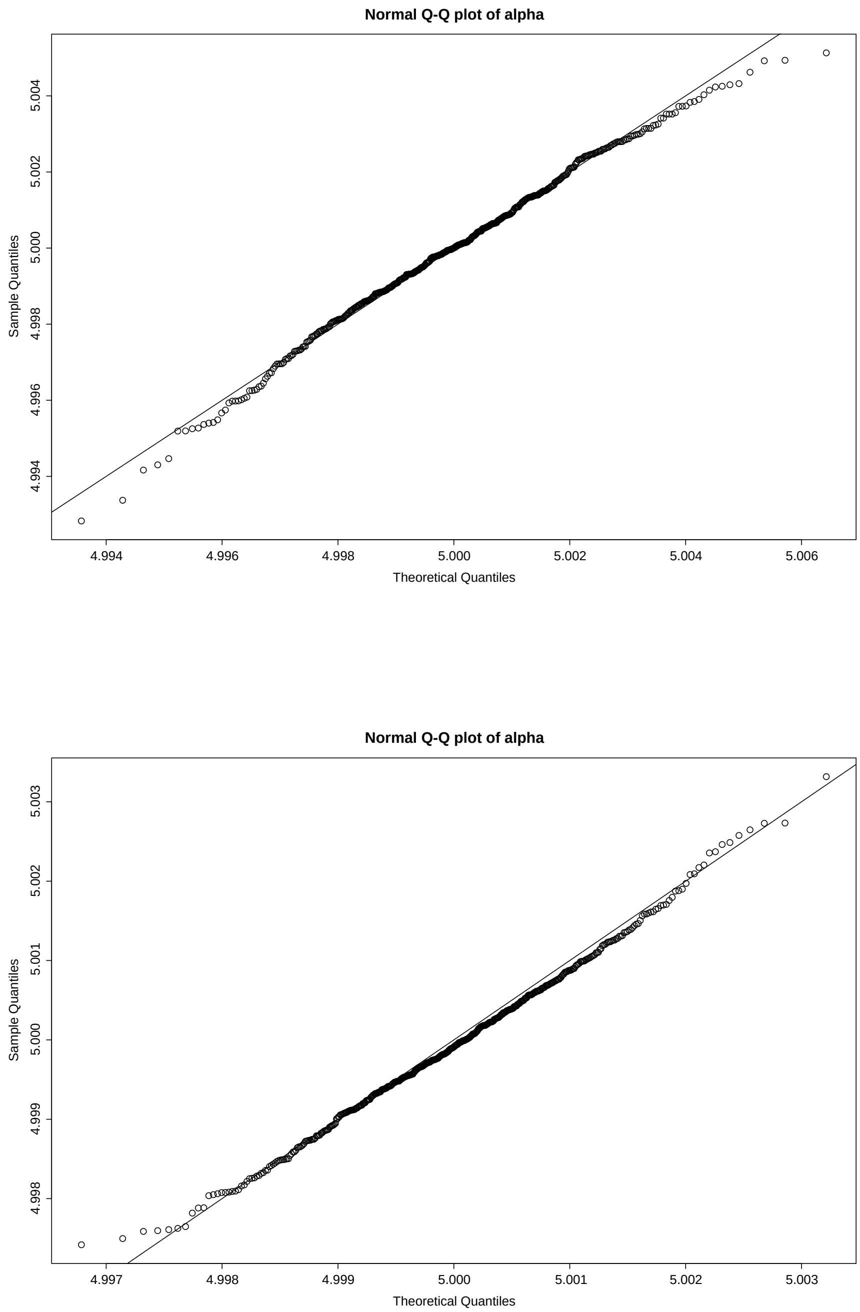
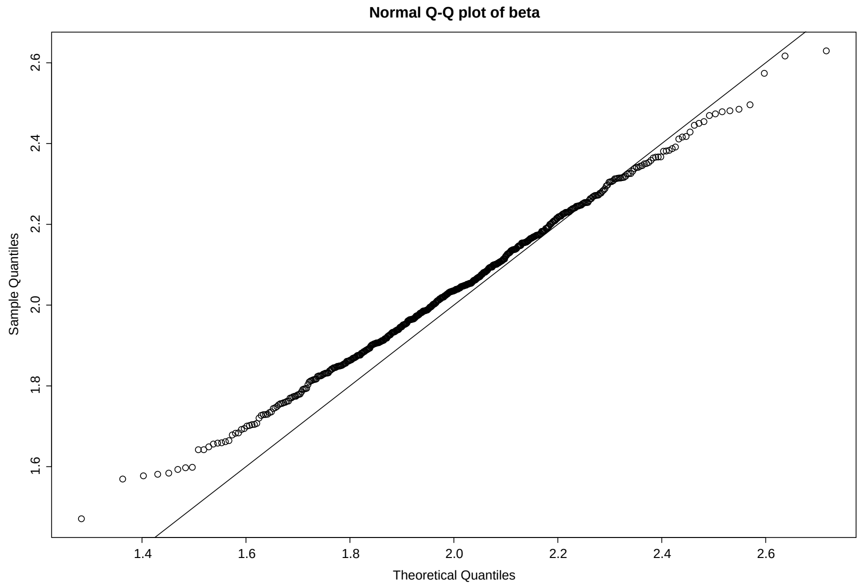


Figure 4.1: Normal quantile plots of the 500 simulated values of the estimator $\hat{\alpha}$ for $\alpha = 5$, $\beta = 2$ and $x_0 = 5.2$. In the upper plot $\epsilon = 0.01$ and in the lower $\epsilon = 0.005$. The lines represent the limiting normal distributions obtained by small dispersion asymptotics.



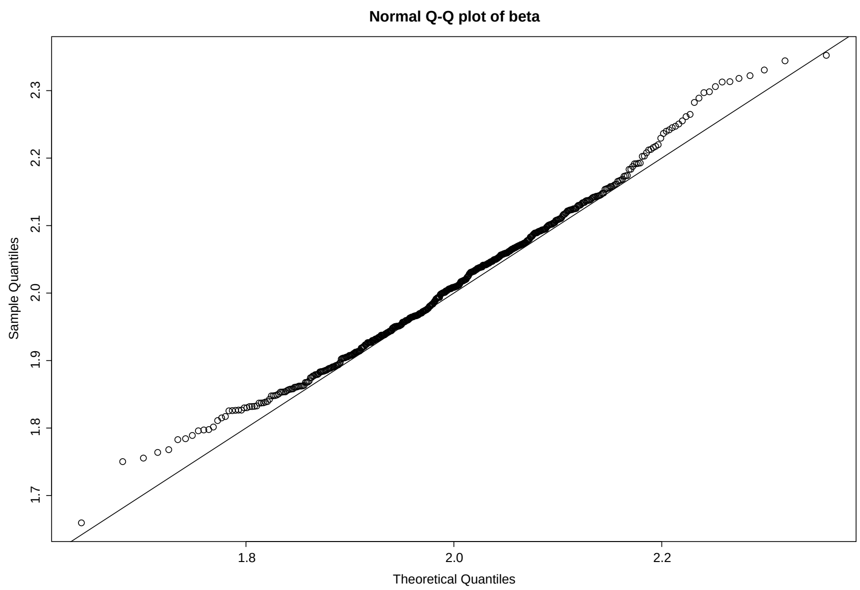


Figure 4.2: Normal quantile plots of the 500 simulated values of the estimator $\hat{\beta}$ for $\alpha = 5$, $\beta = 2$ and $x_0 = 5.2$. In the upper plot $\epsilon = 0.01$ and in the lower $\epsilon = 0.005$. The lines represent the limiting normal distributions obtained by small dispersion asymptotics.

x_0	ϵ	sim. mean	sim. s.d.	theoretical s.d.
5.2	0.1	4.9991	0.0208	0.0208
	0.05	4.9994	0.0104	0.0104
	0.01	5.0000	0.0021	0.0021
	0.005	5.0000	0.0010	0.0010
5.4	0.1	5.0008	0.0202	0.0208
	0.05	5.0004	0.0105	0.0104
	0.01	5.0000	0.0020	0.0021
	0.005	5.0000	0.0010	0.0010

Table 4.1: The mean and standard deviation of the estimator $\hat{\alpha}$ (determined from 500 independent simulated trajectories) and the value of the standard deviation given by (3.6) when $\alpha = 5$, $\beta = 2$, $n = 300$, and $\Delta = 0.1$.

x_0	ϵ	sim. mean	sim. s.d.	theoretical s.d.
5.2	0.1	2.1355	0.4354	2.3190
	0.05	2.0999	0.3989	1.1595
	0.01	2.0368	0.1972	0.2319
	0.005	2.0117	0.1161	0.1159
5.4	0.1	2.1453	0.4103	1.1739
	0.05	2.0962	0.3436	0.5869
	0.01	2.0183	0.1152	0.1174
	0.005	2.0037	0.0560	0.0587

Table 4.2: The mean and standard deviation of the estimator $\hat{\beta}$ (determined from 500 independent simulated trajectories) and the value of the standard deviation given by (3.6) when $\alpha = 5$, $\beta = 2$, $n = 300$, and $\Delta = 0.1$.

Next we choose $\alpha = 0.08$ and $\beta = 0.23$ (Table 4.3 and Table 4.4). These parameter values were taken from Chan, Karolyi, Longstaff and Sanders (1992), who fitted the Cox-Ingersoll-Ross model to the annualized one-month U.S. Treasury bill yield. Here $\Delta = 0.1$ corresponds to ten observations per year.

The small dispersion asymptotics gives a reasonable to good approximation to the standard deviation of $\hat{\alpha}$ in all cases, while again it gives reasonable values of the standard deviation of $\hat{\beta}$ only when $\epsilon < 0.05$. The bias of the estimators decreases with ϵ as one would expect. The estimator $\hat{\beta}$ is seriously biased in most cases. When modelling short term interest rates, a reasonable value of ϵ is probably between 0.05 and 0.1. The normality of the estimators is studied in Figures 4.3 and 4.4, which are normal quantile plots of the 500 estimator values for $x_0 = 0.1$ with $\epsilon = 0.01$ and $\epsilon = 0.005$, respectively. The

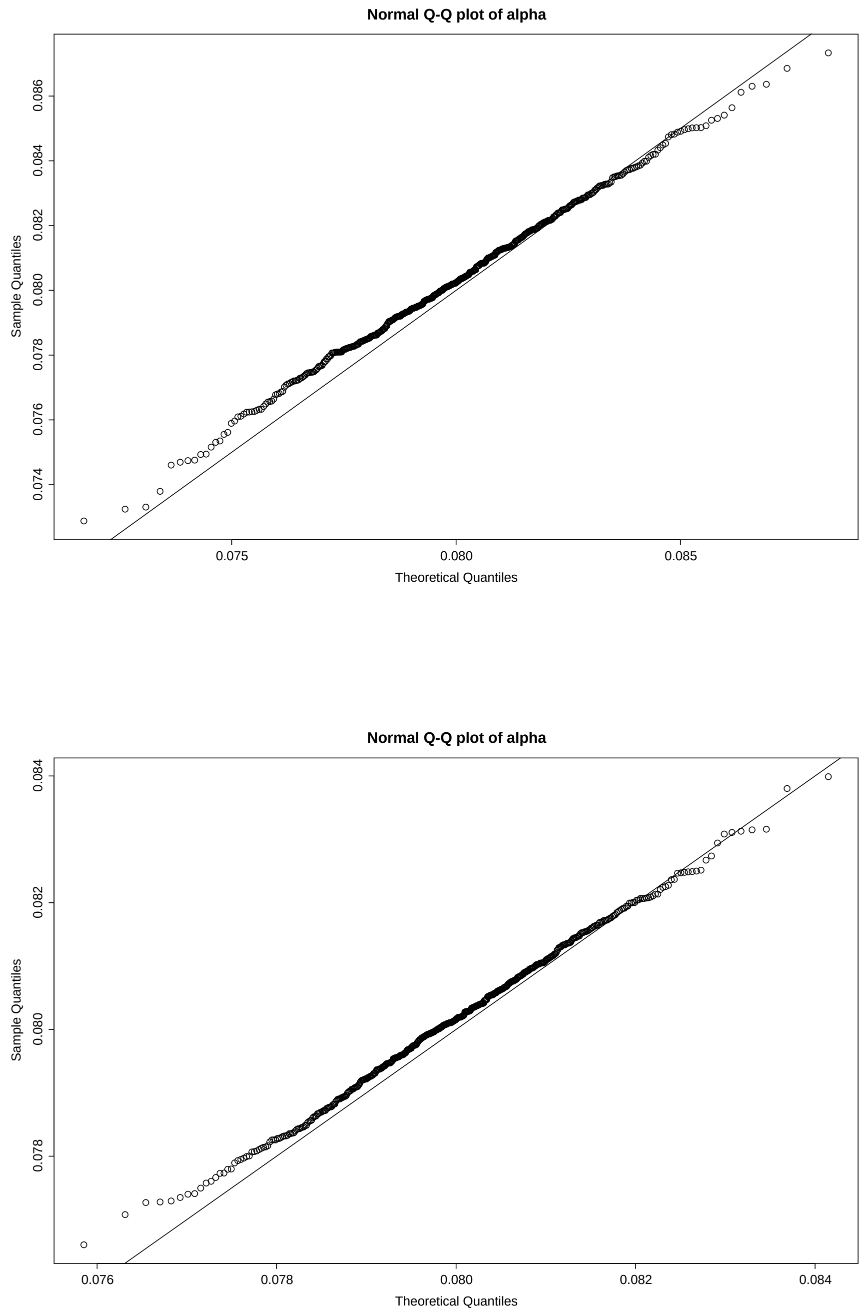


Figure 4.3: Normal quantile plots of the 500 simulated values of the estimator $\hat{\alpha}$ for $\alpha = 0.08$, $\beta = 0.23$ and $x_0 = 5.2$. In the upper plot $\epsilon = 0.01$ and in the lower $\epsilon = 0.005$. The lines represent the limiting normal distributions obtained by small dispersion asymptotics.

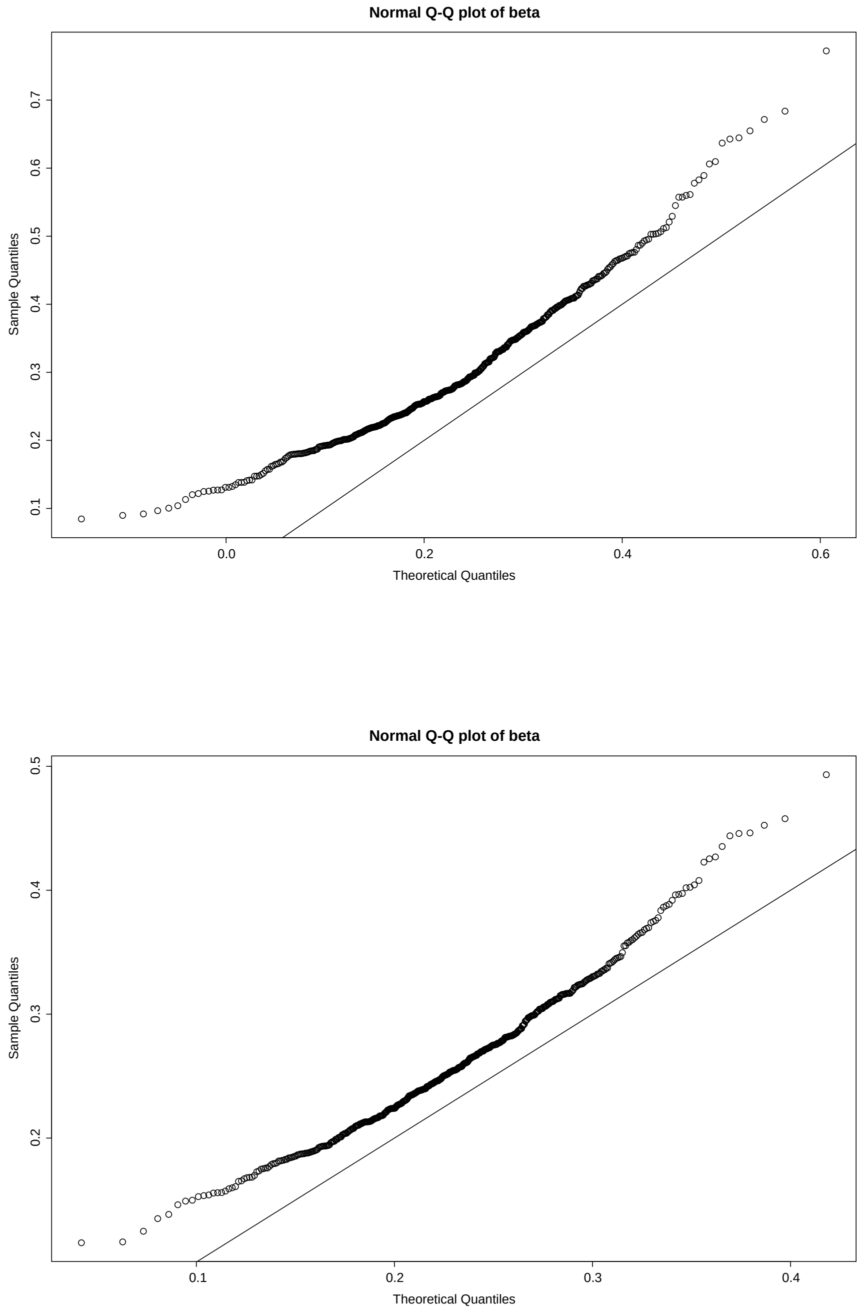


Figure 4.4: Normal quantile plots of the 500 simulated values of the estimator $\hat{\beta}$ for $\alpha = 0.08$, $\beta = 0.23$ and $x_0 = 5.2$. In the upper plot $\epsilon = 0.01$ and in the lower $\epsilon = 0.005$. The lines represent the limiting normal distributions obtained by small dispersion asymptotics.

x_0	ϵ	sim. mean	sim. s.d.	theoretical s.d.
0.10	0.1	0.0817	0.0247	0.0268
	0.05	0.0809	0.0106	0.0134
	0.01	0.0803	0.0024	0.0027
	0.005	0.0802	0.0013	0.0013
0.12	0.1	0.0816	0.0208	0.0270
	0.05	0.0813	0.0117	0.0135
	0.01	0.0802	0.0027	0.0027
	0.005	0.0801	0.0014	0.0014

Table 4.3: The mean and standard deviation of the estimator $\hat{\alpha}$ (determined from 500 independent simulated trajectories) and the value of the standard deviation given by (3.6) when $\alpha = 0.08$, $\beta = 0.23$, $n = 300$, and $\Delta = 0.1$.

x_0	ϵ	sim. mean	sim. s.d.	theoretical s.d.
0.10	0.1	0.3992	0.2000	1.2170
	0.05	0.3927	0.2003	0.6085
	0.01	0.3008	0.1121	0.1217
	0.005	0.2605	0.0626	0.0608
0.12	0.1	0.3774	0.1782	0.6442
	0.05	0.3508	0.1613	0.3221
	0.01	0.2597	0.0687	0.0644
	0.005	0.2386	0.0313	0.0322

Table 4.4: The mean and standard deviation of the estimator $\hat{\beta}$ (determined from 500 independent simulated trajectories) and the value of the standard deviation given by (3.6) when $\alpha = 0.08$, $\beta = 0.23$, $n = 300$, and $\Delta = 0.1$.

lines represent the limiting normal distributions obtained by small dispersion asymptotics. The normal approximation is not quite satisfactory for $\hat{\beta}$ when $\epsilon = 0.01$, and the inaccuracies of the mean and standard deviation already seen in Table 4.2 are clearly visible.

Acknowledgement.

Thanks are due to Anders Nielsen for assistance with the simulation study and to the Centre for Analytical Finance for supporting the research presented in this paper.

References

Aït-Sahalia, Y. (1998): Maximum likelihood estimation of discretely sampled diffusions: A closed-form approach. Working Paper, Princeton University.

Aït-Sahalia, Y. (1999): Transition densities for interest rate and other non-linear diffusions. *Journal of Finance* **54**, 1361 – 1395.

Azencott, R. (1982): Formule de Taylor stochastique et développement asymptotique d'intégrales de Feynmann. *Séminaire de Probabilités XVI; Supplément: Géométrie Différentielle Stochastique. Lecture Notes in Math.* **921**, 237–285. Springer Verlag, Berlin.

Barndorff-Nielsen, O.E. and Sørensen, M. (1994): A review of some aspects of asymptotic likelihood theory for stochastic processes. *Int. Statist. Rev.* **62**, 133–165.

Bibby, B.M. (1994): Optimal combinations of martingale estimating functions for discretely observed diffusion processes. Research Report No. 298, Dept. Theor. Statist., Univ. of Aarhus.

Bibby, B.M. and Sørensen, M. (1995): Martingale estimating functions for discretely observed diffusion processes. *Bernoulli* **1**, 17–39.

Bibby, B.M. and Sørensen, M. (1996): On estimation for discretely observed diffusions: A review. *Theory of Stochastic Processes* **2 (18)**, 49–56.

Bibby, B.M. and Sørensen, M. (1997): A hyperbolic diffusion model for stock prices. *Finance and Stochastics* **1**, 25–41.

Bibby, B.M. and Sørensen, M. (1998): Simplified estimating functions for diffusion models with a high-dimensional parameter. Preprint No. 98-10, Dept. of Theor. Statist. Univ. of Copenhagen. To appear in *Scand. J. Statist.*

Billingsley, P. (1961): *Statistical Inference for Markov Processes.* The University of Chicago Press, Chicago.

Chan, K.C., Karolyi, G.A., Longstaff, F.A. and Sanders, A.B. (1992): An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, **47**, 1209 – 1226.

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985): A theory of the term structure of interest rates. *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.

Elerian, O., Chib, S. and Shephard, N. (1998): Likelihood inference for discretely observed non-linear diffusions. Econometrics discussion paper 146, Nuffield College, Oxford.

Eraker, B. (1997): MCMC analysis of diffusion models with application to finance. Working Paper, Norwegian School of Economics and Business Administration, Bergen.

Genon-Catalot, V. (1990): Maximum contrast estimation for diffusion processes from discrete observations. *Statistics* **21**, 99–116.

Jacobsen, M. (1998): Discretely observed diffusions: Classes of estimating functions and small Δ-optimality. Preprint No. 98-11, Dept. of Theor. Statist. Univ. of Copenhagen. To appear in *Scand. J. Statist.*

Kessler, M. (1995): Martingale estimating functions for a Markov chain. Preprint.

Kessler, M. and Sørensen, M. (1999): Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **5**, 299 – 314.

Kloeden P.E. and Platen, E. (1992): *Numerical Solution of Stochastic Differential Equations.* Springer-Verlag, New York.

Kutoyants, Yu.A. (1984): *Parameter Estimation for Stochastic Processes.* Heldermann, Berlin.

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

Pedersen, A.R. (1994): Quasi-likelihood inference for discretely observed diffusion processes. Research Report No. 295, Dept. Theor. Statist., Univ. Aarhus.

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

Pedersen, A.R. (1995b): Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* **1**, 257–279.

Pedersen, A.R. (1999): Estimating the nitrous oxide emission rate from the soil surface by means of a diffusion model. To appear in *Scand. J. Statist.*

Poulsen, R. (1999): Approximate maximum likelihood estimation of discretely observed diffusion processes. Working Paper 29, CAF, Univ. of Aarhus.

Santa-Clara, P. (1997): Simulated likelihood estimation of diffusion with an application to the short term interest rate. Working Paper 12-97, UCLA.

Sørensen, M. (1997): Estimating functions for discretely observed diffusions: A review. In Basawa, I.V., Godambe, V.P. and Taylor, R.L. (eds.): *Selected Proceedings of the Symposium on Estimating Functions.* IMS Lecture Notes - Monograph Series, Vol. 32, 305–325.

Sørensen, M. (1998): On asymptotics of estimating functions. Preprint No. 98-6, Dept. of Theor. Statist. Univ. of Copenhagen.

Yoshida, N. (1992): Asymptotic expansion for statistics related to small diffusions. *J. Japan Statist. Soc.* **22**, 139 – 159.

Yoshida, N. (1997): Malliavin calculus and asymptotic expansion for martingales. *Probab. Theory Relat. Fields* **109**, 301 – 342.