Simplified Estimating Functions for Diffusion Models with a High-dimensional Parameter

Bo Martin Bibby
Department of Mathematics and Physics
The Royal Veterinary and Agricultural University, Denmark

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

Abstract

We consider estimating functions for discretely observed diffusion processes of the following type: For one part of the parameter of interest we propose to use a simple and explicit estimating function of the type studied by Kessler (1996); for the remaining part of the parameter we use a martingale estimating function. Such an approach is particularly useful in practical applications when the parameter is high-dimensional. It is also often necessary to supplement a simple estimating function by another type of estimating function because only the part of the parameter on which the invariant measure depends can be estimated by a simple estimating function. Under regularity conditions the resulting estimators are consistent and asymptotically normal. Several examples are considered in order to demonstrate the idea of the estimating procedure. The method is applied to two data sets comprising wind velocities and stock prices. In one example we also propose a general method for constructing diffusion models with a prescribed marginal distribution which have a flexible dependence structure.

Key words: asymptotic normality, consistency, Cox-Ingersoll-Ross model, discretely observed diffusions, hyperbolic diffusions, Ornstein-Uhlenbeck process, pseudo likelihood, stochastic differential equations, stock prices, wind velocity.

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

1 Introduction

A satisfactory description of complex dynamical systems by means of stochastic differential equations often leads to a parametric model with a high-dimensional parameter. In this situation likelihood methods typically fail because the likelihood function is not explicitly known. However, martingale estimating functions provide a good alternative, see Bibby & Sørensen (1995), Kessler (1995), Kessler & Sørensen (1995), Bibby & Sørensen (1996), and Sørensen (1997a).

Optimal martingale estimating functions tend to be complex in cases with many parameters. Often simplifications and computer intensive methods are necessary in practice. See Bibby & Sørensen (1997) for an example of the problems encountered in connection with applying martingale estimating functions to a complex diffusion model for financial data.

A considerable simplification of the estimation procedure is obtained if one part of the parameter vector can be estimated well without involving the remaining part of the parameter, which must then be estimated using a martingale estimating function. An example of this is maximum pseudo likelihood estimation for ergodic diffusions; introduced by Kessler (1996). The pseudo likelihood function is obtained by pretending that the observations are independent and identically distributed following the invariant distribution. Thus the pseudo likelihood is the product of invariant densities evaluated at the data points. Often the invariant measure depends only on a part of the parameter vector that can then be estimated using the pseudo likelihood function. Kessler (1996) also introduced other types of simple and explicit estimating functions; see in addition Hansen & Scheinkman (1995). Typically only the part of the parameter vector on which the invariant measure depends can be estimated from the simple estimating functions, which therefore must be supplemented by, for instance, a martingale estimating function as discussed in this paper.

In Section 2 we define the class of diffusion processes that we will study. In Section 3 we propose the estimating procedure and give the asymptotic properties of the resulting estimators. Finally, in Section 4 we look at some examples, namely the Ornstein-Uhlenbeck process, the Cox-Ingersoll-Ross model, and a class of hyperbolic diffusions. We also consider two data sets: One consisting of wind velocities, the other of stock prices.

2 The general model

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

$$dX_t = b(X_t; \theta, \psi)dt + \sigma(X_t; \theta, \psi)dW_t, \quad X_0 = x_0, \tag{2.1}$$

where W is a standard Wiener process. We assume that the drift b and the diffusion coefficient σ are known apart from the parameters $(\theta, \psi) \in \Theta \times \Psi$, where $\Theta \subseteq \mathbb{R}^p$ while $\Psi \subseteq \mathbb{R}^q$. The functions b and σ are, for all $(\theta, \psi) \in \Theta \times \Psi$, assumed to be sufficiently smooth that a unique weak solution exists. The state space is denoted by $S = (l, r)$, where $-\infty \leq l < r \leq \infty$, and we assume that $\sigma(x; \theta, \eta) > 0$ for all $x \in S$ and all $(\theta, \psi) \in \Theta \times \Psi$.

The scale measure of the solution to (2.1) has density

$$s(x;\theta,\psi) = \exp\left[-2\int_{x^*}^{x} \frac{b(y,\theta,\psi)}{\sigma^2(y,\theta,\psi)}dy\right], \quad x \in (l,r), \tag{2.2}$$

with respect to the Lebesgue measure for some arbitrary, but fixed, $x^* \in (l,r)$. We restrict our attention to diffusions that fulfill the following condition.

Condition 2.1 *For all $(\theta,\psi) \in \Theta \times \Psi$,*

$$\int_{x^*}^{r} s(x;\theta,\psi)dx = \int_{l}^{x^*} s(x;\theta,\psi)dx = \infty.$$

Moreover, the speed measure of the solution to (2.1), which has density

$$m(x;\theta,\psi) = \frac{1}{s(x,\theta,\psi)\sigma^2(x,\theta,\psi)}, \quad x \in (l,r), \tag{2.3}$$

with respect to the Lebesgue measure, is assumed to be finite,

$$\mathcal{M}(\theta,\psi) = \int_{l}^{r} m(x;\theta,\psi)dx < \infty.$$

Under Condition 2.1, the process X is ergodic, and its invariant measure has density

$$\mu_{\theta,\psi}(x) = \frac{m(x;\theta,\psi)}{\mathcal{M}(\theta,\psi)}, \tag{2.4}$$

with respect to the Lebesgue measure on (l,r).

We suppose that we have observed the diffusion process X at n distinct time-points that we for the sake of simplicity assume to be equidistant. We denote the observations $X_\Delta,\ldots,X_{n\Delta}$.

3 The estimation method and asymptotic results

The problem is to estimate θ and ψ based on the observations $X_\Delta,\ldots,X_{n\Delta}$. As the likelihood function is typically not available we consider estimating functions of the form

$$G_n(\theta,\psi) = \begin{pmatrix} H_n(\theta) \\ K_n(\theta,\psi) \end{pmatrix}, \tag{3.1}$$

where

$$H_n(\theta) = \sum_{i=1}^{n} h(X_{i\Delta};\theta), \tag{3.2}$$

and

$$K_n(\theta,\psi) = \sum_{i=1}^{n} k(\Delta, X_{(i-1)\Delta}, X_{i\Delta};\theta,\psi). \tag{3.3}$$

An estimator for θ and ψ is then obtained by solving the equation $G_n(\theta, \psi) = 0$. We suppose that h in (3.2) is p-dimensional and satisfies

$$\mu_{\theta,\psi}(h(\theta)) = 0, \tag{3.4}$$

for all $(\theta, \psi) \in \Theta \times \Psi$. Here $\mu_{\theta,\psi}(h(\theta))$ is the integral of h with respect to the invariant measure. Estimating functions of the type in (3.2) were studied by Kessler (1996). If the density of the invariant measure only depends on θ then the function h can for instance be obtained by differentiating the logarithm of the density of the invariant measure μ_θ, see Kessler (1996). If h is chosen in this way, $H_n(\theta)$ is the score function that one obtains by pretending that the data are i.i.d. observations with distribution μ_θ, and the estimator of θ obtained by solving $G_n(\theta, \psi) = 0$ is the corresponding maximum pseudo likelihood estimator. Equation (3.4) holds provided that differentiation and integration can be interchanged in the integral $\int \partial_\theta \mu_\theta(x) dx$. A sufficient condition is that $\partial_\theta \mu_\theta(x)$ is locally dominated integrable with respect to the Lebesgue measure.

Let $y \mapsto p(\Delta, x, y; \theta, \psi)$ denote the density with respect to the Lebesgue measure of the transition distribution of X when the true parameter value is (θ, ψ), i.e. it is the conditional density of X_Δ given that $X_0 = x$. In (3.3), we suppose that k is q-dimensional and satisfies

$$\int k(\Delta, x, y; \theta, \psi) p(\Delta, x, y; \theta, \psi) dy = 0, \tag{3.5}$$

for all $x \in S$ and all $(\theta, \psi) \in \Theta \times \Psi$. This condition exactly amounts to assuming that $K_n(\theta, \psi)$ is a martingale estimating function with respect to the natural filtration.

In the proofs of asymptotic results about the estimators obtained from the estimating function (3.1), we need the ergodic theorem. Provided that $Q^\Delta_{\theta,\psi}(f) < \infty$, it holds under Condition 2.1 that

$$\frac{1}{n} \sum_{i=1}^{n} f(X_{(i-1)\Delta}, X_{i\Delta}) \to Q^\Delta_{\theta,\psi}(f), \tag{3.6}$$

in probability as $n \to \infty$ when (θ, ψ) are the true parameter values; see e.g. Billingsley (1961b). The probability measure $Q^\Delta_{\theta,\psi}$ is the two-dimensional invariant measure which has density

$$Q^\Delta_{\theta,\psi}(x, y) = \mu_{\theta,\psi}(x) p(\Delta, x, y; \theta, \psi), \tag{3.7}$$

with respect to the Lebesgue measure on S^2. Again $Q^\Delta_{\theta,\psi}(f)$ is the integral of f with respect to $Q^\Delta_{\theta,\psi}$. Condition 2.1 also implies that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} k(\Delta, X_{(i-1)\Delta}, X_{i\Delta}; \theta, \psi) \xrightarrow{\mathcal{D}} N\left(0, B_2(\theta, \psi)\right), \tag{3.8}$$

as $n \to \infty$ when (θ, ψ) are the true parameter values; see Billingsley (1961a). Here

$$B_2(\theta, \psi) = Q^\Delta_{\theta,\psi}\left(k(\theta, \psi) k(\theta, \psi)^T\right).$$

To ensure asymptotic normality of the estimating function H, which is not a martingale, we need to impose the following extra condition on our model.

Condition 3.1 *For all* $(\theta, \psi) \in \Theta \times \Psi$,

$$\min\{ \lim_{x \to l} u(x;\theta,\psi), \lim_{x \to r} u(x;\theta,\psi) \} > 0,$$

where

$$\begin{aligned} u(x;\theta,\psi) &= \tfrac{1}{2}[b(x;\theta,\psi)^2 v(x;\theta,\psi)^{-2} + b'(x;\theta,\psi)] \\ &\quad - v'(x;\theta,\psi) b(x;\theta,\psi)/v(x;\theta,\psi) + \tfrac{1}{8} v'(x;\theta,\psi)^2 - \tfrac{1}{4} v(x;\theta,\psi) v''(x;\theta,\psi). \end{aligned}$$

Here $v(x;\theta,\psi) = \sigma^2(x;\theta,\psi)$, *and a prime denotes differentiation with respect to* x.

Condition 3.1 implies, see e.g. Kessler (1996), that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_{i\Delta};\theta) \xrightarrow{\mathcal{D}} N\left(0, B_1(\theta,\psi)\right), \tag{3.9}$$

as $n \to \infty$ when (θ, ψ) are the true parameter values, and where

$$B_1(\theta,\psi) = Q^{\Delta}_{\theta,\psi}\left(a(\theta,\psi) a(\theta,\psi)^T\right), \tag{3.10}$$

with

$$a(x,y;\theta,\psi) = U_{\theta,\psi} h(y;\theta) - U_{\theta,\psi} h(x;\theta) + h(x;\theta). \tag{3.11}$$

Here $U_{\theta,\psi}$ denotes the potential of X, defined by

$$U_{\theta,\psi} f(x) = \sum_{i=0}^{\infty} E_{\theta,\psi}(f(X_{i\Delta}) | X_0 = x),$$

for $f \in L_0^2(\mu_{\theta,\psi})$, where $L_0^2(\mu_{\theta,\psi})$ is the set of real functions on (l, r) that are square integrable with expectation zero under $\mu_{\theta,\psi}$. The sum converges in $L^2(\mu_{\theta,\psi})$.

Condition 3.1 also implies convergence in distribution of $G_n(\theta,\eta)/\sqrt{n}$. Specifically,

$$\frac{1}{\sqrt{n}} G_n(\theta,\psi) \xrightarrow{\mathcal{D}} N(0, \Sigma(\theta,\psi)), \tag{3.12}$$

as $n \to \infty$ when (θ, ψ) are the true parameter values. The asymptotic covariance matrix is given by

$$\Sigma(\theta,\psi) = \begin{pmatrix} B_1(\theta,\psi) & C(\theta,\psi)^T \\ C(\theta,\psi) & B_2(\theta,\psi) \end{pmatrix}, \tag{3.13}$$

where the $q \times p$-matrix $C(\theta,\psi)$ is given by

$$C(\theta,\psi) = Q^{\Delta}_{\theta,\psi}\left(k(\theta,\psi) a(\theta,\psi)^T\right). \tag{3.14}$$

The weak convergence result (3.12) follows by applying the martingale central limit theorem for ergodic processes, Billingsley (1961a), to the martingale

$$\tilde{G}(\theta,\psi) = \begin{pmatrix} \sum_{i=1}^{n} a(\Delta, X_{(i-1)\Delta}, X_{i\Delta};\theta,\psi) \\ K_n(\theta,\psi) \end{pmatrix};$$

for details, see Kessler (1996).

It is usually difficult to find an explicit expression for the potential operator $U_{\theta,\psi}$ and hence for the asymptotic covariance matrix $\Sigma(\theta,\psi)$. In practice, one would therefore have to estimate $\Sigma(\theta,\psi)$ from the data. In Sørensen (1997b) this problem was studied in the case of pseudo likelihood with the invariant measure belonging to an exponential family.

To prove asymptotic results about our estimators we finally make the following assumption about our estimating function. In the rest of the section we denote the true parameter value by (θ_0,ψ_0). Moreover, h_i and k_i denotes the ith coordinate of h and k respectively.

Condition 3.2

(1) The function h is twice continuously differentiable with respect to θ for all x, and the function k is twice continuously differentiable with respect to (θ,ψ) for all x,y.

(2) The functions

$$\begin{aligned} x &\mapsto h_i(x;\theta), \quad i=1,\ldots,p, \\ x &\mapsto \partial_{\theta_j} h_i(x;\theta), \quad i,j=1,\ldots,p, \\ x &\mapsto \partial_{\theta_i}\partial_{\theta_j} h_l(x;\theta), \quad i,j,l=1,\ldots,p, \end{aligned}$$

are all locally dominated integrable with respect to μ_{θ_0,ψ_0}, and the functions $x \mapsto h_i(x;\theta)$, $i=1,\ldots,p$ are in $L^2(\mu_{\theta_0,\psi_0})$ for all $\theta\in\Theta$.

(3) The functions

$$\begin{aligned} (x,y) &\mapsto k_i(x,y;\theta,\psi), \quad i=1,\ldots,q, \\ (x,y) &\mapsto \partial_{\theta_j} k_i(x,y;\theta,\psi), \quad i=1,\ldots,q,\ j=1,\ldots,p, \\ (x,y) &\mapsto \partial_{\psi_j} k_i(x,y;\theta,\psi), \quad i,j=1,\ldots,q, \\ (x,y) &\mapsto \partial_{\theta_i}\partial_{\theta_j} k_l(x,y;\theta,\psi), \quad i,j=1,\ldots,p,\ l=1,\ldots,q, \\ (x,y) &\mapsto \partial_{\theta_i}\partial_{\psi_j} k_l(x,y;\theta,\psi), \quad i=1,\ldots,p,\ j,l=1,\ldots,q, \\ (x,y) &\mapsto \partial_{\psi_i}\partial_{\psi_j} k_l(x,y;\theta,\psi), \quad i,j,l=1,\ldots,q, \end{aligned}$$

are all locally dominated integrable with respect to $Q^{\Delta}_{\theta_0,\psi_0}$, and the functions $(x,y)\mapsto k_i(x,y;\theta,\psi)$, $i=1,\ldots,q$ are in $L^2(Q^{\Delta}_{\theta_0,\psi_0})$ for all $(\theta,\psi)\in\Theta\times\Psi$.

(4) The following $(p+q)\times(p+q)$-matrix is non-singular:

$$M(\theta_0,\psi_0) = \begin{pmatrix} A_1^{-1}(\theta_0,\psi_0) & 0 \\ D(\theta_0,\psi_0) & A_2^{-1}(\theta_0,\psi_0) \end{pmatrix}, \tag{3.15}$$

where the $p\times p$-matrix $A_1^{-1}(\theta_0,\psi_0)$ is given by

$$A_1^{-1}(\theta_0,\psi_0) = \left\{\mu_{\theta_0,\psi_0}\left(\partial_{\theta_j} h_i(\theta_0)\right)\right\}, \tag{3.16}$$

the $q\times q$-matrix $A_2^{-1}(\theta_0,\psi_0)$ is given by

$$A_2^{-1}(\theta_0,\psi_0) = \left\{Q^{\Delta}_{\theta_0,\psi_0}\left(\partial_{\psi_j} k_i(\theta_0,\psi_0)\right)\right\}, \tag{3.17}$$

and the $q\times p$-matrix $D(\theta_0,\psi_0)$ is given by

$$D(\theta_0,\psi_0) = \left\{Q^{\Delta}_{\theta_0,\psi_0}\left(\partial_{\theta_j} k_i(\theta_0,\psi_0)\right)\right\}. \tag{3.18}$$

Under the conditions imposed, the next theorem follows from general results in Sørensen (1998) about estimators obtained from estimating equations by arguments analogous to the proof of Theorem 3.6 in that paper; see also Bibby & Sørensen (1995).

Theorem 3.3 *Suppose Conditions 2.1, 3.1 and 3.2 are satisfied. Then for every n, an estimator $(\hat{\theta}_n, \hat{\psi}_n)$ exists that solves the estimating equation $G_n(\hat{\theta}_n, \hat{\psi}_n) = 0$ with a probability tending to one as $n \to \infty$. Moreover,*

$$\begin{pmatrix} \hat{\theta}_n \\ \hat{\psi}_n \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \theta_0 \\ \psi_0 \end{pmatrix},$$

as $n \to \infty$, and

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\psi}_n - \psi_0 \end{pmatrix} \xrightarrow{\mathcal{D}} N(0, V(\theta_0, \psi_0)),$$

where

$$V(\theta_0, \psi_0) = M(\theta_0, \psi_0)^{-1} \Sigma(\theta_0, \psi_0) (M(\theta_0, \psi_0)^{-1})^T,$$

or more specifically

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

where

$$\begin{aligned} V_{11} &= A_1 B_1 A_1^T, \\ V_{12} &= A_1 C^T A_2^T - A_1 B_1 A_1^T D^T A_2^T, \\ V_{21} &= V_{12}^T, \\ V_{22} &= A_2 B_2 A_2^T + A_2 D A_1 B_1 A_1^T D^T A_2^T - A_2 C A_1^T D^T A_2^T - A_2 D A_1 C^T A_2^T. \end{aligned} \tag{3.19}$$

For clarity we have suppressed the arguments (θ_0, ψ_0).

4 Examples

In this section we present some examples that illustrate the theory in the previous section.

Example 4.1 *The Ornstein-Uhlenbeck process.*

We consider the Ornstein-Uhlenbeck process, that is the solution to the stochastic differential equation given by

$$dX_t = \beta X_t dt + \sigma dW_t, \quad X_0 = x_0,$$

where $\beta < 0$ and $\sigma > 0$. The invariant distribution is a $N(0, \theta)$-distribution, where

$$\theta = -\frac{\sigma^2}{2\beta}.$$

If we pretend that the observations $X_\Delta, \ldots, X_{n\Delta}$ are independent and identically distributed according to the invariant distribution, we obtain a pseudo likelihood function which is the product of $N(0,\theta)$-densities evaluated at the data. By differentiating the corresponding pseudo log-likelihood function with respect to θ we get the following estimating function for θ,

$$H_n(\theta) = \sum_{i=1}^{n} (X_{i\Delta}^2 - \theta),$$

If we put $\psi = \sigma^2$, then the stochastic differential equation has the form

$$dX_t = -\frac{\psi}{2\theta} X_t dt + \sqrt{\psi} dW_t, \quad X_0 = x_0.$$

The parameter ψ can be estimated using the optimal quadratic martingale estimating function,

$$K_n(\theta,\psi) = \alpha \sum_{i=1}^{n} (X_{i\Delta}^2 + X_{(i-1)\Delta}^2) - (1+\alpha^2) \sum_{i=1}^{n} X_{i\Delta} X_{(i-1)\Delta} - n\alpha\theta(1-\alpha^2),$$

see Bibby & Sørensen (1996). Here

$$\alpha = e^{-\frac{\psi\Delta}{2\theta}}.$$

Using these estimating functions it is straightforward to verify that

$$\begin{aligned} A_1 &= -1, \\ A_2 &= -\frac{2}{\alpha\Delta(1+\alpha^2)}, \\ B_1 &= \frac{2\theta^2(1+\alpha^2)}{1-\alpha^2}, \\ B_2 &= \theta^2(1+\alpha^2)(1-\alpha^2)^2, \\ C &= 0, \\ D &= \frac{\psi\Delta}{2\theta}\alpha(1+\alpha^2) - \alpha(1-\alpha^2). \end{aligned}$$

For example, in order to find B_1 it is necessary to find $a(x,y;\theta,\psi)$ given by (3.11). To do this, note that $E_{\theta,\psi}(X_{i\Delta}^2 - \theta | X_0 = x) = \alpha^{2i}(x^2-\theta)$, which implies that $U_{\theta,\psi}h(x;\theta) = (x^2-\theta)/(1-\alpha^2)$, and hence that

$$a(x,y;\theta,\psi) = \frac{y^2 - \alpha^2 x^2}{1-\alpha^2} - \theta.$$

We get that

$$V(\theta_0,\psi_0) = \begin{pmatrix} \frac{2\theta_0^2(1+\alpha_0^2)}{1-\alpha_0^2} & \frac{2\theta_0\psi_0(1+\alpha_0^2)}{1-\alpha_0^2} - \frac{4\theta_0^2}{\Delta} \\ \frac{2\theta_0\psi_0(1+\alpha_0^2)}{1-\alpha_0^2} - \frac{4\theta_0^2}{\Delta} & \frac{2\psi_0^2(1+\alpha_0^2)}{1-\alpha_0^2} - \frac{8\theta_0\psi_0}{\Delta} + \frac{4\theta_0^2(1-\alpha_0^2)}{\Delta^2\alpha_0^2} \end{pmatrix}.$$

Comparing with the asymptotic variance for the maximum likelihood estimator for (θ,ψ) we note that our simplified estimation method in this case in fact produces an efficient estimator.

The estimator $\hat{\theta}_n$ is not equal to the maximum likelihood estimator, but of course asymptotically equivalent to it. In fact the maximum likelihood estimator for θ is given by

$$\hat{\theta}_n^{(mle)} = \frac{1}{n}\sum_{i=1}^n X_{(i-1)\Delta}^2 \frac{\left(\sum_{i=1}^n X_{(i-1)\Delta}^2\right)\left(\sum_{i=1}^n X_{i\Delta}^2\right) - \left(\sum_{i=1}^n X_{i\Delta}X_{(i-1)\Delta}\right)^2}{\left(\sum_{i=1}^n X_{(i-1)\Delta}^2\right)^2 - \left(\sum_{i=1}^n X_{i\Delta}X_{(i-1)\Delta}\right)^2}.$$

Example 4.2 *The Cox-Ingersoll-Ross process.*

We consider the diffusion model proposed by Cox, Ingersoll, Jr. & Ross (1985) to describe interest rate data. The process is the solution to

$$dX_t = (\alpha + \beta X_t)dt + \sigma\sqrt{X_t}dW_t, \quad X_0 = x_0,$$

where $\beta < 0$ and $\sigma > 0$. If we put $\psi = \sigma^2$ and

$$\theta = (\theta_1, \theta_2)^T = \left(\frac{2\alpha}{\sigma^2}, -\frac{2\beta}{\sigma^2}\right)^T,$$

then the invariant distribution is a $\Gamma(\theta_1, \theta_2)$-distribution.

Again pretending that $X_\Delta, \ldots, X_{n\Delta}$ are independent and identically distributed according to the invariant distribution results in the following estimating function (pseudo score function) for θ,

$$H_n(\theta) = \left(n(\log\theta_2 - F(\theta_1)) + \sum_{i=1}^n \log X_{i\Delta}\ ,\ n\frac{\theta_1}{\theta_2} - \sum_{i=1}^n X_{i\Delta}\right),$$

where F is the digamma function.

In terms of θ and ψ the stochastic differential equation describing the Cox-Ingersoll-Ross process has the form

$$dX_t = \tfrac{1}{2}\psi(\theta_1 - \theta_2 X_t)dt + \sqrt{\psi X_t}dW_t, \quad X_0 = x_0.$$

Using the approximately optimal linear martingale estimating function introduced in Bibby & Sørensen (1995) we have

$$K_n(\theta, \psi) = \sum_{i=1}^n \frac{\theta_1 - \theta_2 X_{(i-1)\Delta}}{\sqrt{X_{(i-1)\Delta}}}\left(X_{i\Delta} - \frac{1}{\theta_2}\left(\theta_1 - \left(\theta_1 - \theta_2 X_{(i-1)\Delta}\right)e^{-\frac{1}{2}\theta_2\Delta\psi}\right)\right).$$

This estimating functions actually yields an explicit estimate for ψ, namely,

$$\hat{\psi}_n = \frac{2}{\theta_2\Delta}\log\left(\frac{\sum_{i=1}^n \frac{(\theta_1 - \theta_2 X_{(i-1)\Delta})^2}{\sqrt{X_{(i-1)\Delta}}}}{\sum_{i=1}^n \frac{(\theta_1 - \theta_2 X_{(i-1)\Delta})(\theta_1 - \theta_2 X_{i\Delta})}{\sqrt{X_{(i-1)\Delta}}}}\right),$$

which exists provided the denominator is positive. However, in this case no explicit expression for the potential operator is known so efficiency properties of the estimators are hard to study.

In order to explore the estimation procedure for different values of Δ and n, we made a simulation study. We simulated 500 or 1000 observations from the Cox-Ingersoll-Ross process for values of Δ ranging from 0.5 and 3.0. For every combination of Δ and n we made 1000 simulations and in each case estimated the parameters using the proposed method. In Table 4.1 the empirical mean and standard error is given for the three parameter estimators.

For $\Delta = 3.0$ we got no estimate for ψ in 155 cases for $n = 500$ and in 72 cases for $n = 1000$. For all other combinations of Δ and n there were no problems with estimating ψ.

From Table 4.1 we see that the bias is small for all combinations of Δ and n. We also see that the accuracy, with which θ_1 and θ_2 are estimated, increases with Δ. The opposite is true for the accuracy of ψ. This is as could be expected since a large value of Δ corresponds to a situation where the observations are close to being independent.

n	Δ	θ_1	θ_2	ψ
500	0.5	20.2686 (1.8390)	2.0285 (0.1900)	1.0004 (0.0818)
500	1.0	20.1839 (1.4382)	2.0171 (0.1455)	1.0090 (0.1011)
500	2.0	20.0992 (1.2864)	2.0107 (0.1295)	1.0516 (0.2252)
500	3.0	20.1842 (1.2602)	2.0193 (0.1292)	0.8719 (0.4706)
1000	0.5	20.2250 (1.2984)	2.0218 (0.1340)	1.0024 (0.0584)
1000	1.0	20.0995 (1.0188)	2.0093 (0.1032)	1.0071 (0.0730)
1000	2.0	20.1553 (0.9531)	2.0158 (0.0971)	1.0117 (0.1269)
1000	3.0	20.0017 (0.8927)	1.9995 (0.0901)	0.9766 (0.3743)

Table 4.1: Empirical mean and standard error of the 1000 parameter estimates. Here the true parameter values are $\theta_1 = 20$, $\theta_2 = 2$, and $\psi = 1$. In all cases the initial value is $x_0 = 10$.

Example 4.3 *Modelling wind velocity.*

In september 1985 the wind velocity was measured on the beach at Ferring on the Danish West Coast. This was done using a sonic anemometer on a 30-meter mast. The three-dimensional wind velocity vector was measured with a 10-Hz frequency but we only consider the stream-wise component. For details about the experiment see Mikkelsen (1988) and Mikkelsen (1989). The data are given in Figure 4.1.
The log-histogram of the wind velocity data is given in Figure 4.2 along with a fit using a hyperbolic density function, see Barndorff-Nielsen (1977). The hyperbolic parameters were estimated by means of the program HYP, see Blæsild & Sørensen (1992). We propose to describe the data using a stationary diffusion process with hyperbolic marginal distribution. We will therefore briefly discuss a way of constructing a class of diffusions with a prescribed marginal distribution and which has a flexible dependence structure.

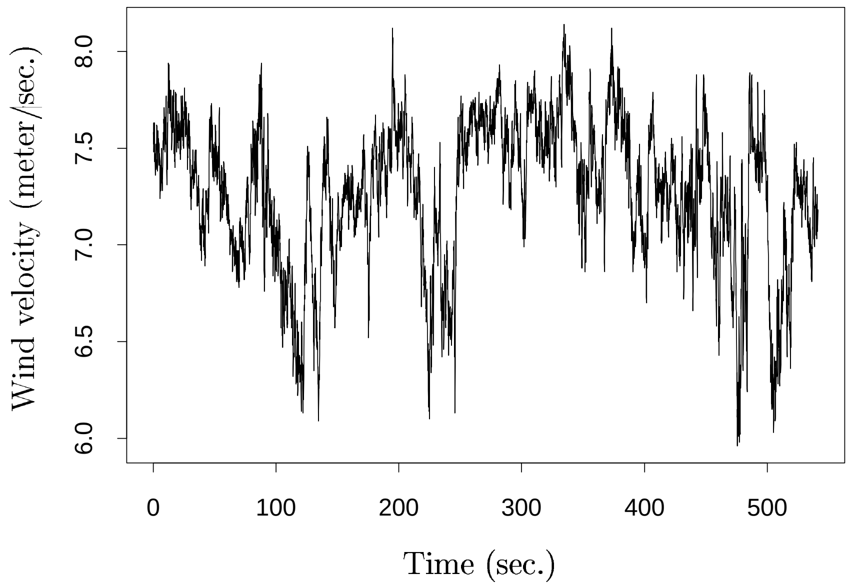

Figure 4.1: The stream-wise wind velocity component in meters per second plotted against time in seconds.

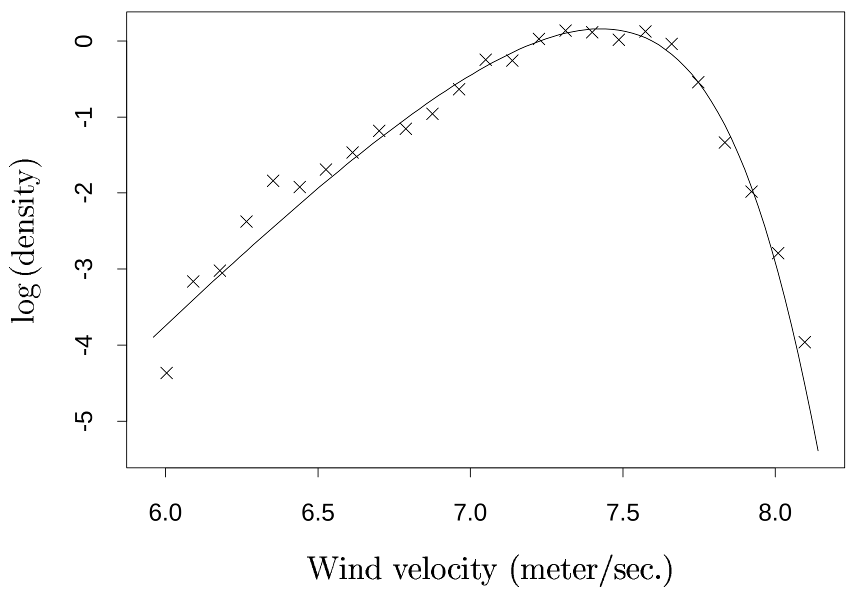

Figure 4.2: A log-histogram of the wind velocity data with a fitted curve corresponding to a hyperbolic density.

Let f and v be two continuously differentiable strictly positive real functions defined on the open interval (l, r), where $-\infty \leq l < r \leq \infty$. The function f is assumed to be integrable with respect to the Lebesgue measure on (l, r). Furthermore, suppose that

$$\int_{x^*}^{r} [v(x)f(x)]^{-1}dx = \int_{l}^{x^*} [v(x)f(x)]^{-1}dx = \infty, \tag{4.1}$$

where x^* is a fixed point in (l, r), and that v^{-1} is integrable on any compact subinterval of (l, r). Then the stochastic differential equation

$$dX_t = \tfrac{1}{2}v(X_t)[\log(f(X_t)v(X_t))]'dt + \sqrt{v(X_t)}dW_t, \tag{4.2}$$

where a prime denotes differentiation, has a unique ergodic Markovian weak solution with invariant measure proportional to $f(x)$. This follows from results in Engelbert & Schmidt (1981) after a transformation by the scale function.

We will now consider the class of diffusions obtained by the particular choice $v(x) = \sigma^2 f(x)^{-\gamma}$, where $\sigma^2 > 0$ and $\gamma \in [0, 1]$. Obviously, $f(x)^\gamma$ is integrable on any compact subinterval of (l, r). If (l, r) is the real line, it is easy to see that (4.1) is satisfied. Suppose f is unimodal and that the mode point m belongs to the interior of (l, r). The the diffusion has reversion towards m when $\gamma < 1$.

Since we want a diffusion on the real line with hyperbolic marginal distributions, we choose f proportional to the hyperbolic density function

$$f(x) = \exp\left[-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)\right], \tag{4.3}$$

where $\alpha > |\beta| \geq 0$ and $\delta > 0$. Thus we arrive at a diffusion model given by the stochastic differential equation

$$dX_t = \tfrac{1}{2}\sigma^2(1-\gamma)f(X_t)^{-\gamma}\left[\beta - \frac{\alpha(X_t-\mu)}{\sqrt{\delta^2 + (X_t-\mu)^2}}\right]dt + \sigma f(X_t)^{-\frac{1}{2}\gamma}dW_t. \tag{4.4}$$

If X_0 is a random variable independent of the Wiener process W and distributed with a density proportional to (4.3), then X is stationary with the same marginal distribution. If $X_0 = x_0$, the marginal distribution of X_t will converge in distribution as $t \to \infty$ to a distribution with density proportional to (4.3). Note that when $\gamma < 1$, the diffusion has reversion towards $\mu + \beta\delta/\sqrt{\alpha^2 - \beta^2}$. If $\gamma = 1$, the drift is zero, so only the diffusion term keeps the process stationary. Therefore, in the latter case, the trajectories can have "spikes" that go far away from the typical level of the process. Thus the full class where $\gamma \in [0, 1]$ spans the whole range from a process that is kept stationary purely by the effect of reversion ($\gamma = 0$) to a process where stationarity is ensured purely by a volatility effect. Between these extremes, both effects are present with varying weights.

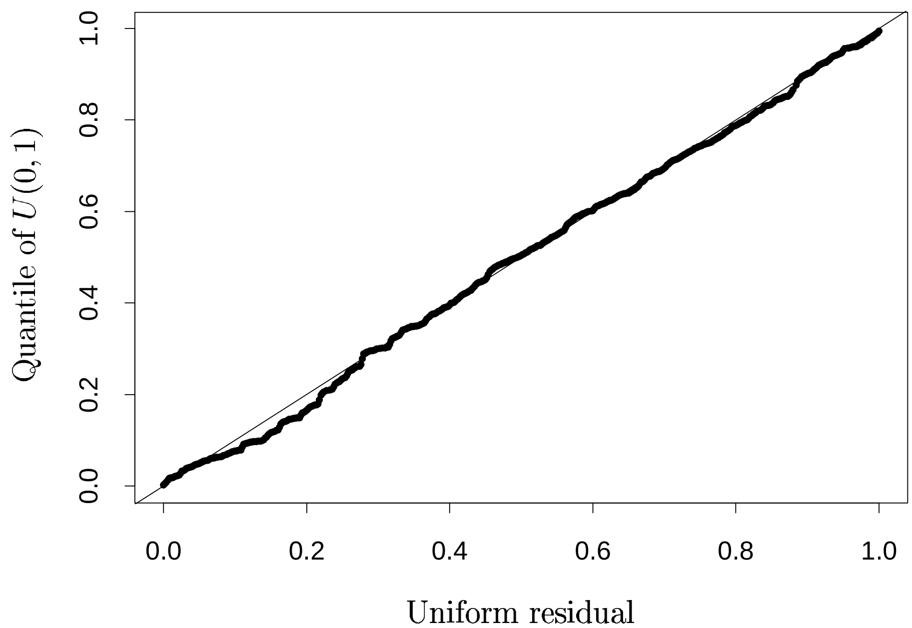

Figure 4.3: A plot of the uniform residuals for the fit of the hyperbolic diffusion model to the wind velocity data.

Let $\theta = (\alpha, \beta, \delta, \mu)^T$ and $\psi = (\sigma, \gamma)^T$. We estimate θ pretending that $X_\Delta, \ldots, X_{n\Delta}$ are independent and identically distributed according to the invariant distribution. In this case the program HYP produced the following estimates,

$$\begin{aligned} \hat{\alpha}_n &= 23.26 \\ \hat{\beta}_n &= -18.89 \\ \hat{\delta}_n &= 0.4790 \\ \hat{\mu}_n &= 8.0946 \end{aligned}$$

Using the optimal quadratic martingale estimating function, see Bibby & Sørensen (1996), in order to estimate ψ we get

$$\begin{aligned} \hat{\sigma}_n &= 0.1556 \\ \hat{\gamma}_n &= 0.1534 \end{aligned}$$

Figure 4.3 shows the uniform residuals (see Pedersen (1994)) associated with the parameter estimates. This figure gives no reason to doubt the model.

Example 4.4 *Modelling stock prices.*

We consider 1563 daily observations (weekends ignored) from the 2nd of October 1989 to the 29th of December 1995 of the price of VW-stocks. The data are from the Karlsruher Kapitalmarktdatenbank. Figure 4.4 shows the logarithm of the stock prices minus an estimated linear trend (de-trended data).

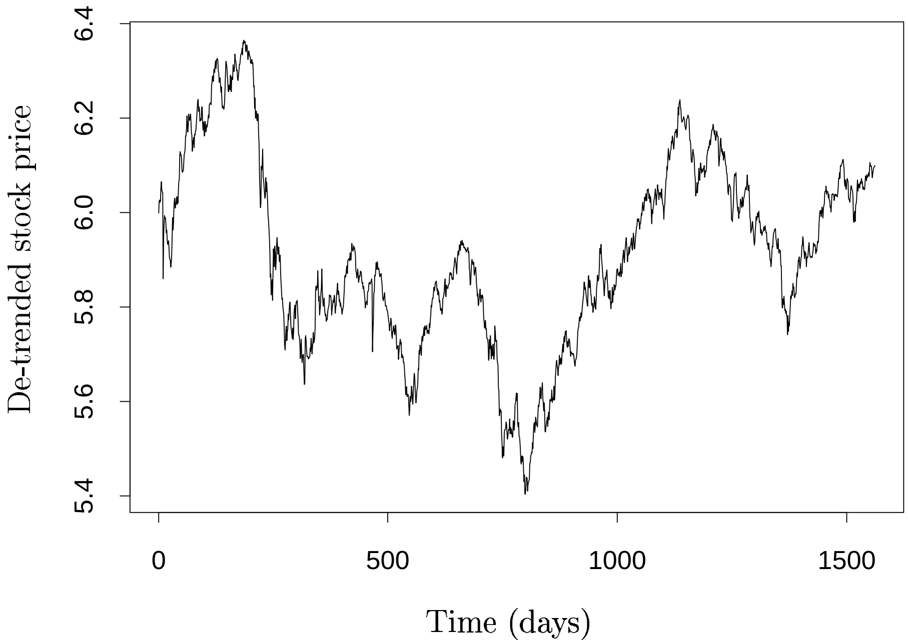

Figure 4.4: The transformed stock price data for VW plotted as a function of time in days.

In Figure 4.5 the log-histogram of the de-trended stock price data is given along with a fitted hyperbolic log-density function found by the HYP-program. In Bibby & Sørensen (1997) the diffusion process given by (4.4) was used to model de-trended stock prices in the special case where $\gamma = 1$. It is therefore of some interest to fit the full model (4.4) to stock price data. The parameter estimates found by HYP are

$$\begin{aligned} \hat{\alpha}_n &= 5.393 \\ \hat{\beta}_n &= 1.576 \\ \hat{\delta}_n &= 0.080 \\ \hat{\mu}_n &= 5.842 \end{aligned}$$

The estimates found by the optimal quadratic martingale estimating function are given by

$$\begin{aligned} \hat{\sigma}_n &= 0.0170 \\ \hat{\gamma}_n &= 0.0121 \end{aligned}$$

Figure 4.6 shows the uniform residuals associated with the parameter estimates. Also for these data the model looks reasonable based on Figure 4.6.

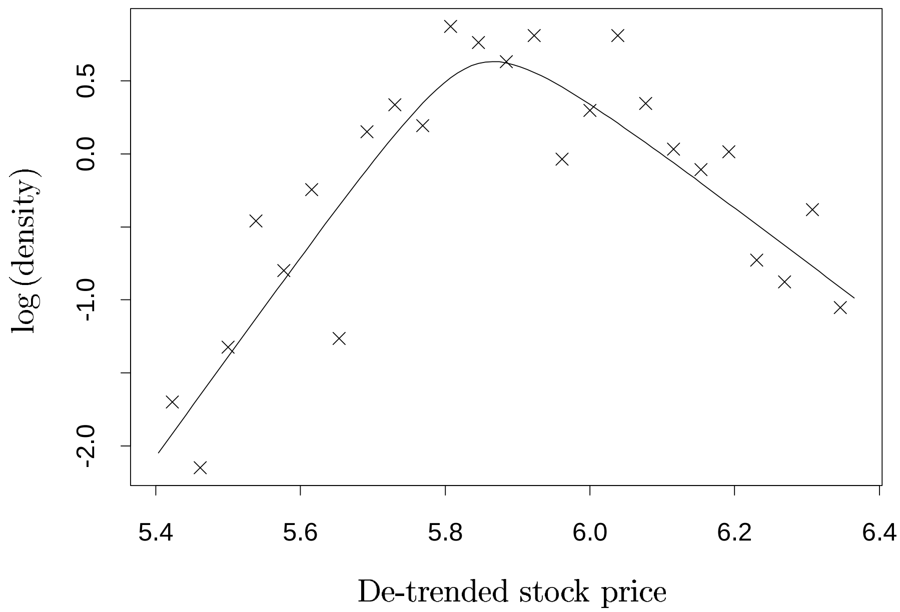

Figure 4.5: A log-histogram of the transformed stock price data with a fitted curve corresponding to a hyperbolic density.

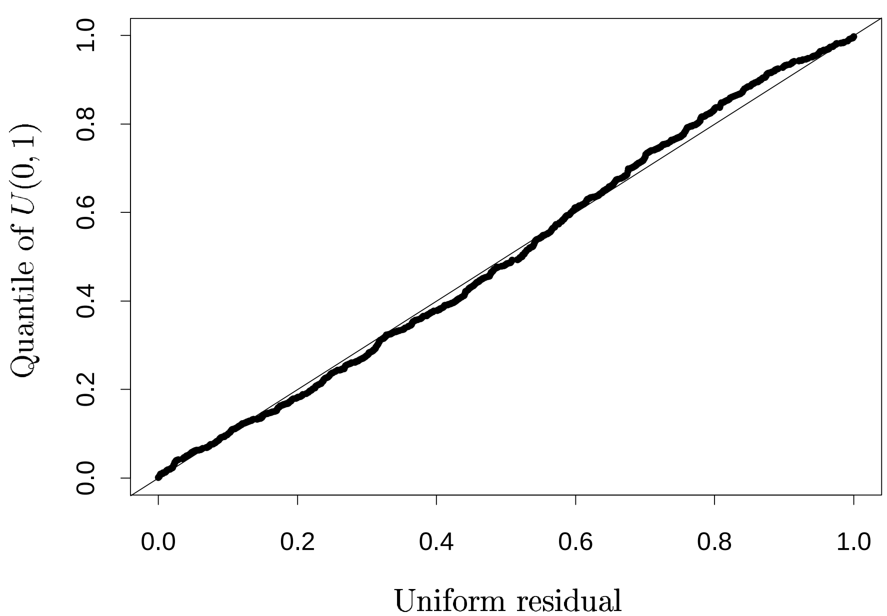

Figure 4.6: A plot of the uniform residuals for the fit of the hyperbolic diffusion model to the transformed stock price data.

References

Barndorff-Nielsen, O. E. (1977). "Exponentially Decreasing Distributions for the Logarithm of Particle Size". *Proc. Roy. Soc. London A*, 353:401–419.

Bibby, B. M. & Sørensen, M. (1995). "Martingale Estimation Functions for Discretely Observed Diffusion Processes". *Bernoulli*, 1(1/2):17–39.

Bibby, B. M. & Sørensen, M. (1996). "On Estimation for Discretely Observed Diffusions: A Review". *Theory of Stochastic Processe*, 2(18):49–56.

Bibby, B. M. & Sørensen, M. (1997). "A Hyperbolic Diffusion Model for Stock Prices". *Finance and Stochastics*, 1:25–41.

Billingsley, P. (1961a). "The Lindeberg-Lévy Theorem for Martingales". *Proc. Amer. Math. Soc.*, 12:788–792.

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

Blæsild, P. & Sørensen, M. K. (1992). "HYP - A Computer program for Analyzing Data by Means of the Hyperbolic Distribution". Research Report No. 248, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus.

Cox, J. C.; Ingersoll, Jr., J. E. & Ross, S. A. (1985). "A Theory of the Term Structure of Interest Rates". *Econometrica*, 53(2):385–407.

Engelbert, H. J. & Schmidt, W. (1981). "On the Behaviour of Certain Functionals of the Wiener Process and Applications to Stochastic Differential Equations". In *Stochastic Differential Systems. Lecture Notes in Control and Information Sciences No. 36*, pages 47–55, Berlin. Springer.

Hansen, L. P. & Scheinkman, J. A. (1995). "Back to the Future: Generating Moment Implications for Continuous Time Markov Processes". *Econometrica*, 63:767–804.

Kessler, M. (1995). "Quasi-Likelihood Inference for a Discrete Markov Chain". Preprint.

Kessler, M. (1996). "Simple and Explicit Estimating Functions for a Discretely Observed Diffusion Process". Research Report No. 336, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus.

Kessler, M. & Sørensen, M. (1995). "Estimating Equations Based on Eigenfunctions for a Discretely Observed Diffusion Process". Research Report No. 332, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus. To appear in *Bernoulli.*

Mikkelsen, H. E. (1988). "Turbulence in the Wake Zone of a Vegetated 2-Dimensional Dune". Geoskrifter 30, Geology Institute, University of Aarhus.

Mikkelsen, H. E. (1989). "Wind Flow and Sediment Transport over a Low Costal Dune". Geoskrifter 32, Geology Institute, University of Aarhus.

Pedersen, A. R. (1994). "Uniform Residuals for Discretely Observed Diffusion Processes". Research Report No. 292, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus.

Sørensen, M. (1997a). "Estimating Functions for Discretely Observed Diffusions: A Review". In Basawa, I. V.; Godambe, V. P. & Taylor, R. L., editors, *Selected Proceedings of the Symposium on Estimating Functions*, pages 305–325. IMS Lecture Notes - Monograph Series, Vol 32.

Sørensen, M. (1997b). "Exponential Family Inference for Diffusion Models". Research Report No. 383, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus.

Sørensen, M. (1998). "On Asymptotics of Estimating Functions". Preprint No. 6, Department of Theoretical Statistics, University of Copenhagen.