Estimation for Discretely Observed Diffusions using Transform Functions

Leah Kelly[1], **Eckhard Platen**[1] and **Michael Sørensen**[2]

April 29, 2003

Abstract.

This paper introduces a new estimation technique for discretely observed diffusion processes. Transform functions are applied to transform the data to obtain good and easily calculated estimators of both the drift and diffusion coefficients. Consistency and asymptotic normality of the resulting estimators is investigated. Power transforms are used to estimate the parameters of affine diffusions, for which explicit estimators are obtained.

View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

[1]University of Technology Sydney, School of Finance & Economics and School of Mathematical Sciences, PO Box 123, Broadway, NSW, 2007, Australia
[2]University of Copenhagen, Department of Applied Mathematics and Statistics, Universitetspraken 5, DK-2100 Copenhagen Ø, Denmark

1 Introduction

There exists a rich literature on the parameter estimation for diffusion processes. Estimation of continuously observed diffusions is well studied, see, for instance, Kutoyants (1984), Prakasa Rao (1999) or Liptser & Shiryaev (2001). When the diffusion process is observed continuously, *maximum likelihood estimation* results in estimators that are consistent, asymptotically normal and efficient.

However, in practice it is usually only possible to observe the diffusion process at discrete time points. Inference for discretely observed diffusions can be based, for instance, on an approximation to the likelihood function. If the transition densities of the diffusion are known explicitly, then the likelihood function can be used to estimate the parameters. For the resulting maximum likelihood estimators the properties of consistency and asymptotic normality have been studied in Dacunha-Castelle & Florens-Zmirou (1986). When the transition densities are unknown, a possible approach is to approximate the log-likelihood function based on continuous observations. This technique has the problem that the estimators that result are inconsistent if the time between observations is fixed, see Florens-Zmirou (1989). This problem can be solved by suitable modifications, see Bibby & Sørensen (1995) and H. Sørensen (2001). If the time between observations is sufficiently small, this method works for drift parameters, see Kloeden et al. (1996).

In lieu of this, there has been a great deal of research into alternative methods for the estimation of discretely observed diffusions, many of which are reviewed in Prakasa Rao (1999). One main strand of research has been to develop approximations to the transition density and hence to the likelihood function. Pedersen (1995) and Brandt & Santa-Clara (2001) independently derived a simulation based method for approximating the likelihood function. A sequence of approximating transition densities is constructed that converges to the true transition density. Based on these transition densities a sequence of likelihood functions is used to approximate the true likelihood. Elerain, Chib & Shepard (2001) and Eraker (2001) also developed a simulation based estimation method using a Bayesian Markov Chain Monte Carlo (MCMC) technique. Poulsen (1999) discussed an approximate maximum likelihood technique, which involves solving the Kolmogorov forward equation, see Karatzas & Shreve (1991), to obtain approximations to the transition densities and hence the likelihood function. Aït-Sahalia (2002) used an approximation to the likelihood function based on Hermitian expansions to estimate the transition density. Jensen & Poulsen (1999) compared a large number of techniques used to approximate transition densities. The techniques they considered included simulation-based methods, binomial approximations, numerical solutions of the Kolmogorov forward equation and Hermitian expansions. Their results indicate that the Hermitian expansion technique of Aït-Sahalia (2002) performs best when speed and efficiency considerations are included.

An alternative approach is to use estimating functions, which are functions of both the parameter and the observed data, to derive the estimators. Bibby & Sørensen (1995) studied martingale estimating functions obtained from the derivative of the continuous time log-likelihood function by correcting for the discretization bias by subtracting its compensator. The resulting estimating function, known as a linear martingale estimating function, depends on the conditional moments of the diffusion process. Quadratic martingale estimating functions involving also second order conditional moments were obtained from a Gaussian approximation to the likelihood function by Bibby & Sørensen (1996), while Kessler & Sørensen (1999) considered estimating functions based on eigenfunctions, for which the conditional moments are explicit. Sørensen (2000) considered more general estimating functions, known as prediction-based estimating functions, where conditional moments are approximated by expressions involving only unconditional moments. This type of estimating function is a useful alternative to martingale estimating functions when the observed process is non-Markovian, as in stochastic volatility models, for example. Christensen, Poulsen & Sørensen (2001) compared optimal martingale estimating functions and the approximate maximum likelihood method mentioned earlier for the estimation of the parameters in a model of the short rate to techniques such as the generalized method of moments, see Hansen (1982), and indirect inference, see Gouriéroux, Monfort & Renault (1993) and Gallant & Tauchen (1996). It was found that optimal martingale estimating functions and the approximate maximum likelihood method reduce bias, true standard errors and bias in estimated standard errors when compared to the aforementioned techniques. As discussed in Heyde (1997), it can be more advantageous to work with estimating functions than the estimators themselves. Reasons for this include that estimating functions are invariant under one-to-one transforms of the data and these functions can be combined more simply than the estimators themselves. For example, in Bibby (1994) martingale estimating functions are combined to estimate parameters in both the drift and the diffusion coefficient. Surveys of recent results on estimating function theory are given in Heyde (1997), Sørensen (1997) and Bibby, Jacobsen & Sørensen (2003).

In a number of papers, for example Dorogovcev (1976), Prakasa Rao (1988), Florens-Zmirou (1989) and Kessler (1997), contrast functions based on approximations to the conditional moments have been proposed to estimate parameters in diffusion models. The approach in the present paper is somewhere between this method and that of martingale estimating functions, but firmly based on the foundations of estimating function theory.

Our objective is to obtain a simple yet general estimation method, which provides more flexibility in the estimation of discretely observed diffusion processes via the use of *transform functions*. Unlike some of the aforementioned techniques, particular information about the conditional and unconditional moments of the diffusion process are not needed. Section 2 introduces the transform function method. Section 3 discusses the asymptotics of the resulting estimators. Section 4 reviews results about affine diffusions and Section 5 applies the technique using a power transform function. An illustration of the methodology is given in Section 6.

2 Transform Function for a Diffusion Process

Consider a class of one-dimensional diffusion processes defined by the following stochastic differential equation (SDE)

$$dX_t = b(t, X_t; \theta)dt + \sigma(t, X_t; \theta)dW_t \tag{2.1}$$

for $t \geq 0$. The initial value $X_0 = x_0$ is assumed to be $\mathcal{A}_0$-measurable. Here W denotes a standard Wiener process given on the filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P_\theta)$, where the filtration $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t\geq 0}$ satisfies appropriate conditions, see Karatzas & Shreve (1991) or Jacod & Shiryaev (2003). We assume that the SDE (2.1) has a unique solution for all parameter values θ in a given open subset $\Theta \subseteq \mathbb{R}^p$, $p \in \{1, 2, \dots\}$, see Kloeden & Platen (1999). The drift and diffusion coefficient functions $b(\cdot, \cdot; \theta) : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ and $\sigma(\cdot, \cdot; \theta) : [0, \infty) \times \mathbb{R} \to \mathbb{R}$, respectively, are assumed to be known with the exception of the parameter vector $\theta = (\theta^1, \dots, \theta^p)^\top \in \Theta$. Throughout $A^\top$ denotes transposition of A.

It is our aim to estimate the unknown parameter vector θ from observations of the diffusion process $X = \{X_t, t \geq 0\}$. For simplicity, an equidistant time discretization with observation times τ_n, where $0 = \tau_0 < \tau_1 < \dots < \tau_n < \tau_{n+1} < \dots$, is assumed to be such that the time step size $\Delta = \tau_n - \tau_{n-1} \in (0, 1)$. For $t \geq 0$ we introduce the integer n_t as the largest integer n for which τ_n does not exceed t, that is

$$n_t = \max\{n \in \{0, 1, \dots\} : \tau_n \leq t\} = \left[\frac{t}{\Delta}\right], \tag{2.2}$$

where $[x]$ denotes the integer part of the real number x.

To provide sufficient flexibility for our estimation approach we consider, at the observation times $\tau_0, \tau_1, \tau_2, \dots$, the original data

$$X_{\tau_0}, X_{\tau_1}, X_{\tau_2}, \dots \tag{2.3}$$

and the transformed data

$$U(\tau_0, X_{\tau_0}; \lambda_i), U(\tau_1, X_{\tau_1}; \lambda_i), U(\tau_2, X_{\tau_2}; \lambda_i), \dots \tag{2.4}$$

for $i \in \{1, 2, \dots, p\}$. Here $U(\cdot, \cdot; \cdot) : [0, \infty) \times \mathbb{R} \times \Lambda \to \mathbb{R}$ is a smooth real valued function with respect to $t \in [0, \infty)$ and $x \in \mathbb{R}$, where $\Lambda \subseteq \mathbb{R}$.

The function $U(\cdot, \cdot; \lambda_i)$ for $i \in \{1, 2, \dots, p\}$, is called the *ith transform function*, and is used to transform the data in a manner that allows us to obtain good estimates of the unknown parameters. In principle, for each $i \in \{1, 2, \ldots, p\}$, a different function could be used to estimate the parameters. For fixed $\lambda_i \in \Lambda$ we obtain, by the Itô formula, the following SDE for the transformed data

$$dU(t, X_t; \lambda_i) = L_\theta^0 U(t, X_t; \lambda_i)dt + L_\theta^1 U(t, X_t; \lambda_i)dW_t \tag{2.5}$$

for $t \in [0, \infty)$. Here we have used the operators

$$L_\theta^0 u(t,x) = \left(\frac{\partial}{\partial t}u(t,x) + b(t,x;\theta)\frac{\partial}{\partial x}u(t,x) + \frac{1}{2}\sigma^2(t,x;\theta)\frac{\partial^2}{\partial x^2}u(t,x)\right) \tag{2.6}$$

and

$$L_\theta^1 u(t,x) = \sigma(t,x;\theta)\frac{\partial}{\partial x}u(t,x). \tag{2.7}$$

For $n \in \{1, 2, \ldots\}$, $i \in \{1, 2, \ldots, p\}$ and $\lambda_i \in \Lambda$ we introduce the normalized difference

$$D_{\lambda_i,n,\Delta} = \frac{1}{\tau_n - \tau_{n-1}}(U(\tau_n, X_{\tau_n}; \lambda_i) - U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)) \tag{2.8}$$

and the normalized squared increment

$$Q_{\lambda_i,n,\Delta} = \frac{1}{\tau_n - \tau_{n-1}}(U(\tau_n, X_{\tau_n}; \lambda_i) - U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))^2. \tag{2.9}$$

By a truncated Wagner-Platen expansion, see Kloeden & Platen (1999), the increment of U in (2.8) and (2.9) can be expressed in terms of multiple stochastic integrals. Thus, we obtain

$$\begin{aligned}
&U(\tau_n, X_{\tau_n}; \lambda_i) - U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) \\
&\quad = L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)(W_{\tau_n} - W_{\tau_{n-1}}) + L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)(\tau_n - \tau_{n-1}) \\
&\quad + L_\theta^1 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)\frac{1}{2}((W_{\tau_n} - W_{\tau_{n-1}})^2 - (\tau_n - \tau_{n-1})) \\
&\quad + L_\theta^0 L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)\frac{(\tau_n - \tau_{n-1})^2}{2} \\
&\quad + L_\theta^1 L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)\int_{\tau_{n-1}}^{\tau_n}\int_{\tau_{n-1}}^{s} dz dW_s \\
&\quad + L_\theta^0 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)\int_{\tau_{n-1}}^{\tau_n}\int_{\tau_{n-1}}^{s} dW_z ds \\
&\quad + R_{\lambda_i,n,\theta}(\tau_n, \tau_{n-1}, X_{\tau_n}),
\end{aligned} \tag{2.10}$$

where $R_{\lambda_i,n,\theta}(\tau_n, \tau_{n-1}, X_{\tau_n})$ is the corresponding remainder term, as follows from Kloeden & Platen (1999). The expansion (2.10) can also be obtained by application of the Itô formula to U and then repeated to $L_\theta^0 U$ and $L_\theta^1 U$. The first term in (2.10) has mean zero and is the leading term of the expansion. Note that the third, fifth and sixth terms also have mean zero but are of a higher order than the first term. In (2.10) the order of the first term is $\sqrt{\Delta}$, that of the second and third term is Δ and that of the fourth term is Δ^2. The order of the fifth and sixth term is $\Delta^{\frac{3}{2}}$. The remainder term has mean and variance of order Δ^3.

Using (2.8) and (2.9) we can construct estimating functions exploiting the structure of the first and second term of the above increment (2.10). To do this define $F_n(\theta) = (F_n^{(1)}(\theta)^\top, F_n^{(2)}(\theta)^\top)^\top$ with $F_n^{(j)}(\theta)^\top = (F_{1,n}^{(j)}(\theta), \dots, F_{q,n}^{(j)}(\theta))$, $j \in \{1,2\}$ where $F_{i,n}^{(1)}(\theta) = D_{\lambda_i,n,\Delta} - L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)$, $i \in \{1,2,\dots,q\}$ and $F_{i,n}^{(2)}(\theta) = Q_{\lambda_i,n,\Delta} - (L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))^2$, $i \in \{1,2,\dots,q\}$ for $\theta \in \Theta$ and suitably chosen values of $\lambda_i \in \Lambda$, $i \in \{1,2,\dots,q\}$. It is not necessary that the number of λ_i's is the same for $F_n^{(1)}(\theta)$ and $F_n^{(2)}(\theta)$ and they need not have the same value for the two functions. This assumption simplifies the exposition.

A class of estimating functions is then given by

$$K(\theta, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} M(\theta)\, F_n(\theta), \tag{2.11}$$

where the $p \times 2q$ matrix valued function $M(\theta) = M(\theta, \tau_{n-1}, X_{\tau_{n-1}}, \Delta)$ is free to be chosen appropriately. Throughout the paper the dependence of a weighting matrix and its elements on τ_{n-1}, $X_{\tau_{n-1}}$ and Δ will be suppressed. The estimating function $K(\theta_0, t, \Delta)$, where θ_0 is taken to be the true parameter value, has expectation of order Δ. Thus, when the observation interval Δ is sufficiently small, the expectation of $K(\theta_0, t, \Delta)$ is approximately zero. Essentially, the approach adopted here is to approximate the conditional moments of the transformed diffusion process using the expansion in (2.10). This is similar to the approach in Kessler (1997), where closed form approximations for the first two conditional moments are derived and used to construct a contrast estimator for parameters in the drift and diffusion functions.

The estimating function (2.11) is slightly biased. To determine the optimal weighting matrix $M^*(\theta)$, we consider the unbiased estimating function $K^\circ(\theta, t, \Delta)$ obtained by compensating $K(\theta, t, \Delta)$ such that

$$K^\circ(\theta, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} M(\theta)\, (F_n(\theta) - \bar{F}_n(\theta)). \tag{2.12}$$

Here $\bar{F}_n(\theta) = E_\theta(F_n(\theta) | X_{\tau_{n-1}})$ is the compensator for $F_n(\theta)$ and is of order Δ. The optimal choice for the weighting matrix $M(\theta)$ in the unbiased estimating equation (2.12), in the sense of Godambe & Heyde (1987), can be derived using the method outlined in Heyde (1997). For details regarding the case of diffusion processes, see Sørensen (1997), where the optimal weighting matrix is given by

$$M^*(\theta) = B^*(\theta)\, V^*(\theta)^{-1}, \tag{2.13}$$

where $V^*(\theta)$ is the $2q \times 2q$ conditional covariance matrix

$$V^*(\theta) = V^*(\theta, \tau_{n-1}, X_{\tau_{n-1}}) = E_\theta\left((F_n(\theta) - \bar{F}_n(\theta))(F_n(\theta) - \bar{F}_n(\theta))^\top \middle| X_{\tau_{n-1}} \right)$$

and $B^*(\theta) = (B^{*(1)}(\theta), B^{*(2)}(\theta))$ with $B^{*(k)}(\theta)$, $k \in \{1,2\}$, denoting the $p \times q$ matrix where the (i,j)th entry is

$$B^{*(k)}(\theta) = B^{*(k)}(\theta, \tau_{n-1}, X_{\tau_{n-1}})_{i,j} = E_\theta \left(\frac{\partial}{\partial \theta^i} \left[F_{j,n}^{(k)}(\theta) - \bar{F}_{j,n}^{(k)}(\theta) \right] \Bigg| X_{\tau_{n-1}} \right).$$

The values of the λ_i's should be chosen in such a way that the conditional covariance matrix $V^*(\theta)$ is invertible.

Keeping only the leading terms, we obtain

$$\tilde{K}(\theta, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} \tilde{M}(\theta)\, F_n(\theta), \tag{2.14}$$

where $\tilde{M}(\theta) = B(\theta)V(\theta)^{-1}$. Here $B(\theta) = (B^{(1)}(\theta),\ B^{(2)}(\theta))$, where the (i,j)th entry of the $p \times q$ matrices $B^{(1)}(\theta)$ and $B^{(2)}(\theta)$ are

$$B^{(1)}(\theta)_{i,j} = \frac{\partial}{\partial \theta^i} L_\theta^0\, U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) \tag{2.15}$$

and

$$B^{(2)}(\theta)_{i,j} = \frac{\partial}{\partial \theta^i} \left(L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) \right)^2, \tag{2.16}$$

respectively. Moreover,

$$V(\theta) = \left\{ \begin{array}{cc} V^{11}(\theta) & V^{12}(\theta) \\ V^{21}(\theta) & V^{22}(\theta) \end{array} \right\},$$

where the (i,j)th entry of the $q \times q$ matrices $V^{11}(\theta)$, $V^{22}(\theta)$ and $V^{12}(\theta)$ are

$$V^{11}(\theta)_{i,j} = L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j), \tag{2.17}$$

$$V^{22}(\theta)_{i,j} = 2 \left[L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) \right]^2, \tag{2.18}$$

and

$$\begin{aligned} V^{12}(\theta)_{i,j} &= 2 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) \\ &\quad + L_\theta^1 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) \left[L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) \right]^2, \end{aligned} \tag{2.19}$$

while $V^{21}(\theta) = (V^{12}(\theta))^\top$.

The weighting matrix $\tilde{M}(\theta)$ in (2.14) is optimal in what is referred to as the fixed sample sense, see Godambe & Heyde (1987) and Heyde (1997). This means, the weighting matrix results in an estimating function (2.14), that is, to the order of approximation used, closest within the class of estimating functions of the form (2.11) to the corresponding, usually unknown, score function. In this sense, this gives the most efficient estimator for a fixed number of observations within the class of estimators considered. Under appropriate assumptions, for example, if the diffusion process is ergodic, it can be proved that a fixed sample optimal martingale estimating function is also asymptotically optimal, see Heyde (1997). Asymptotic optimality results in an estimator that has the smallest asymptotic confidence intervals within the class of estimators considered. The estimating functions proposed in this paper are approximations of martingale estimating functions to the order Δ.

For given transform functions U, with parameters $\lambda_i \in \Lambda$, $i \in \{1, ..., q\}$, we have now obtained a p-dimensional estimating equation

$$\tilde{K}(\theta, t, \Delta) = 0$$

for $t \geq \tau_1$, see (2.14). Assuming that the resulting system of p equations has a unique solution, we obtain for the particular SDE (2.1) an estimator $\hat{\theta} = (\hat{\theta}^1, ..., \hat{\theta}^p)^\top$ for the parameter vector θ. Note that the vector of estimators, $\hat{\theta}$, depends on t, Δ, $\lambda_1, ..., \lambda_p$ and the observed data. Appropriate values of λ_i for $i \in \{1, 2, ..., p\}$ can be found by exploiting asymptotic properties of the estimating functions as described in the next section. The choice of the λ_i for $i \in \{1, 2, ..., p\}$ determines the resulting system of equations.

A simpler, although less efficient, estimation procedure can be used when the parameter θ can be written as $\theta = (\alpha, \beta)$. Here it is assumed that the p_1-dimensional parameter α appears only in the drift coefficient, while the diffusion coefficient depends only on the p_2-dimensional parameter β. In this case we first estimate β by solving $\tilde{H}(\hat{\beta}, t, \Delta) = 0$, where

$$\tilde{H}(\beta, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} B^{(2)}(\beta)\, V^{22}(\beta)^{-1} F_n^{(2)}(\beta) \tag{2.20}$$

with the $p_2 \times q$ matrix $B^{(2)}(\beta)$ given by

$$B^{(2)}(\beta)_{i,j} = \frac{\partial}{\partial \beta^i} \left(L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) \right)^2, \tag{2.21}$$

and $V^{22}(\beta) = V^{22}(\theta)$ given by (2.18). Note that $\tilde{H}(\beta, t, \Delta)$ does not depend on α. Next estimate α by solving $\tilde{G}(\hat{\alpha}, \hat{\beta}, t, \Delta) = 0$, where $\hat{\beta}$ is the estimator of β previously obtained, and

$$\tilde{G}(\alpha, \beta, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} B^{(1)}(\alpha, \beta) V^{11}(\alpha, \beta)^{-1} F_n^{(1)}(\alpha, \beta) \tag{2.22}$$

with the $p_1 \times q$ matrix $B^{(1)}(\alpha, \beta)$ given by

$$B^{(1)}(\alpha, \beta)_{i,j} = \frac{\partial}{\partial \alpha^i} L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j). \tag{2.23}$$

and $V^{11}(\alpha,\beta) = V^{11}(\theta)$ given by (2.17). The estimating functions $\tilde{G}(\alpha,\beta,t,\Delta)$ and $\tilde{H}(\beta,t,\Delta)$ are, to the order of approximation used, optimal within the classes

$$G(\alpha,\beta,t,\Delta) = \frac{1}{n_t}\sum_{n=1}^{n_t} M^{(1)}(\alpha,\beta)\, F_n^{(1)}(\alpha,\beta)$$

and

$$H(\beta,t,\Delta) = \frac{1}{n_t}\sum_{n=1}^{n_t} M^{(2)}(\beta) F_n^{(2)}(\beta)$$

for estimating α and β, respectively. The optimal martingale estimating function for the form (2.12) is the optimal combination of the optimal martingale estimating functions to which (2.20) and (2.22) are approximations, see Heyde (1997) and Bibby (1994).

3 Asymptotics

The proposed transform function method is designed to encompass both stationary and nonstationary diffusion processes. Despite this, it is advantageous to analyze the asymptotic behaviour and bias of the parameter estimates for some given class of real valued diffusion processes. We assume in this section that X is ergodic and described by the SDE (2.1) with time homogeneous coefficient functions

$$b(t,x;\theta) = b(x,\theta) \tag{3.1}$$

and

$$\sigma(t,x;\theta) = \sigma(x;\theta) \tag{3.2}$$

for $t \geq 0$, $x \in \mathbb{R}$ and $\theta \in \Theta$. We use as state space the interval (ℓ, r) where $-\infty \leq \ell < r \leq \infty$. For given parameter vector $\theta \in \Theta$, the density of the scale measure $s : (\ell, r) \to [0,\infty)$ is given by the expression

$$s(x;\theta) = \exp\left(-2\int_{y_0}^{x} \frac{b(y;\theta)}{\sigma^2(y;\theta)}dy\right) \tag{3.3}$$

for $x \in (\ell, r)$ with some reference value $y_0 \in (\ell, r)$. If the following two conditions

$$\int_{y_0}^{r} s(x;\theta)ds = \int_{\ell}^{y_0} s(x;\theta)ds = \infty \tag{3.4}$$

and

$$\int_{\ell}^{r} \frac{1}{s(x;\theta)\sigma^2(x;\theta)}dx < \infty \tag{3.5}$$

are satisfied, then it is well known that X is ergodic with stationary density

$$\bar{p}(x,\theta) = \frac{C(\theta)}{\sigma^2(x;\theta)} \exp\left(2\int_{y_0}^{x} \frac{b(u;\theta)}{\sigma^2(u;\theta)}du\right) \tag{3.6}$$

for $x \in (\ell, r)$ and $\theta \in \Theta$. The constant $C(\theta)$ results from the normalization condition

$$\int_{\ell}^{r} \bar{p}(x,\theta)dx = 1. \tag{3.7}$$

To prove the existence, consistency and asymptotic normality of the estimators we introduce the following conditions and notation where, essentially, we follow Sørensen (1999). We denote by θ_0 the true parameter value, where θ_0 is an interior point of Θ. The true probability measure is denoted by P_{θ_0}. Further, let $p(\Delta, x, y; \theta_0)$ be the true transition density of the observed diffusion process X for a transition from x to y over a time period of length $\Delta > 0$. Throughout the remainder of this section we take Δ to be fixed. We consider estimating functions of the form

$$G_t(\theta) = \frac{1}{n_t}\sum_{n=1}^{n_t} g(\Delta, X_{\tau_{n-1}}, X_{\tau_n}; \theta), \tag{3.8}$$

for $t \geq 0$ and where G_t and $g = (g_1, g_2, \ldots, g_p)^\top$ are p-dimensional. Furthermore, we assume that X is stationary and impose the condition that X is geometrically α-mixing. For a definition of this concept, see, for instance, Doukhan (1994). For a given one-dimensional, ergodic diffusion process X there are a number of relatively simple criteria ensuring α-mixing with exponentially decreasing mixing coefficients. We cite the following straightforward and rather weak set of conditions used in Genon-Catalot, Jeantheau & Laredo (2000) on the coefficients b and σ that are sufficient to ensure geometric α-mixing of X.

Condition 3.1

(i) The function b is continuously differentiable and σ is twice continuously differentiable with respect to $x \in (\ell, r)$, $\sigma(x;\theta_0) > 0$ for all $x \in (\ell, r)$, and there exists a constant $K > 0$ such that $|b(x;\theta_0)| \leq K(1 + |x|)$ and $\sigma^2(x;\theta_0) \leq K(1 + x^2)$ for all $x \in (\ell, r)$.

(ii) $\sigma(x;\theta_0)\bar{p}(x,\theta_0) \to 0$ as $x \downarrow \ell$ and $x \uparrow r$.

(iii) $\frac{1}{\gamma(x;\theta_0)}$ has a finite limit as $x \downarrow \ell$ and $x \uparrow r$, where

$$\gamma(x;\theta_0) = \partial_x \sigma(x;\theta_0) - \frac{2b(x;\theta_0)}{\sigma(x;\theta_0)}.$$

Each pair of neighboring observations $(X_{\tau_{n-1}}, X_{\tau_n})$ has the joint probability density

$$q_{\theta_0}^{\Delta}(x,y) = \bar{p}(x;\theta_0)p(\Delta,x,y;\theta_0)$$

on $(\ell,r)^2$. For a function $f:(\ell,r)^2 \to \mathbb{R}$, where we assume that the following integral exists, we introduce the functional

$$q_{\theta_0}^{\Delta}(f) = \int_{\ell}^{r}\int_{\ell}^{r} f(x,y)p(\Delta,x,y;\theta_0)\bar{p}(x;\theta_0)dy\,dx.$$

For our purposes, we cannot assume that the estimating function in (3.8) is unbiased. Instead we make the following assumption.

Condition 3.2 *There exists a unique parameter value $\bar{\theta}$ that is an interior point of Θ such that*

$$q_{\theta_0}^{\Delta}\left(g(\Delta,\bar{\theta})\right) = 0.$$

We can now impose our conditions on the estimating function (3.8), see Barndorff-Nielsen & Sørensen (1994) and Sørensen (1999).

Condition 3.3

(i) The function $g_i(\Delta,x,y;\cdot):\Theta\to\mathbb{R}$ is twice continuously differentiable with respect to $\theta\in\Theta$ for all $x,y\in(\ell,r)$, and $i\in\{1,2,\dots,p\}$.

(ii) The function $g_i(\Delta,\cdot,\cdot;\theta):(\ell,r)\times(\ell,r)\to\mathbb{R}$ is such that there exists a $\delta>0$ with $q_{\theta_0}^{\Delta}(g_i(\Delta,\theta)^{2+\delta})<\infty$ for all $\theta\in\Theta$ and $i\in\{1,2,\dots,p\}$.

(iii) For the partial derivatives $\frac{\partial}{\partial\theta^j}g_i(\Delta,x,y;\theta)$ and $\frac{\partial^2}{\partial\theta^j\partial\theta^k}g_i(\Delta,x,y;\theta)$, $i,j,k\in\{1,2,\dots,p\}$, there exists for every $\theta^*\in\Theta$ a neighborhood $\mathcal{N}(\theta^*)\subset\Theta$ of θ^* and a non-negative random variable $L(\theta^*)$ with $E_{\theta_0}(L(\theta^*))<\infty$ such that $|\frac{\partial}{\partial\theta^j}g_i(\Delta,x,y;\theta)|\le L(\theta^*)$ and $|\frac{\partial^2}{\partial\theta^j\partial\theta^k}g_i(\Delta,x,y;\theta)|\le L(\theta^*)$ for all $\theta\in\mathcal{N}(\theta^*)$, $(x,y)\in(\ell,r)^2$, and $i,j,k\in\{1,2,\dots,p\}$.

(iv) The $p\times p$ matrix

$$A(\theta_0,\bar{\theta}) = \left\{q_{\theta_0}^{\Delta}\left(\frac{\partial}{\partial\theta^j}g_i(\Delta,\cdot,\cdot;\bar{\theta})\right)\right\}_{i,j=1}^{p}$$

is invertible.

Theorem 3.4 *Suppose Conditions 3.2 and 3.3 are satisfied. Then for every $t>\Delta$, there exists an estimator $\hat{\theta}_{n_t}$ that solves the estimating equation $G_t(\hat{\theta}_{n_t})=0$*

with a probability tending to one as $n_t \to \infty$. Moreover, we have the limit in P_{θ_0}-probability

$$\lim_{n_t \to \infty} \hat{\theta}_{n_t} \overset{P_{\theta_0}}{=} \bar{\theta}$$

and under P_{θ_0} the limit in distribution

$$\lim_{n_t \to \infty} \sqrt{n_t}(\hat{\theta}_{n_t} - \bar{\theta}) \overset{d}{=} R,$$

where

$$R \sim N\left(0, A(\theta_0, \bar{\theta})^{-1} v(\theta_0, \bar{\theta})(A(\theta_0, \bar{\theta})^{-1})^\top\right)$$

is a p-dimensional, zero mean Gaussian distributed random variable with covariance matrix $A(\theta_0, \bar{\theta})^{-1} v(\theta_0, \bar{\theta})(A(\theta_0, \bar{\theta})^{-1})^\top$, where

$$\begin{aligned} v(\theta_0, \bar{\theta}) &= q_{\theta_0}^{\Delta}(g(\Delta, \bar{\theta}) g(\Delta, \bar{\theta})^\top) \\ &+ \sum_{k=1}^{\infty} \left\{ E_{\theta_0}\left(g(\Delta, X_{\tau_0}, X_{\tau_1}; \theta) g(\Delta, X_{\tau_k}, X_{\tau_{k+1}}; \theta)^\top \right) \right. \\ &\qquad \left. + E_{\theta_0}\left(g(\Delta, X_{\tau_k}, X_{\tau_{k+1}}; \theta) g(\Delta, X_{\tau_0}, X_{\tau_1}; \theta)^\top \right) \right\}. \end{aligned} \tag{3.9}$$

Under the conditions imposed, the covariances in the infinite sum in (3.9) tend to zero exponentially fast as $k \to \infty$, so the sum converges quickly and can usually be well approximated by a finite sum with relatively few terms.

The theorem can be proved in complete analogy with the proof of Theorem 3.6 in Sørensen (1999). The only difference is that the martingale limit theory used in that paper must be replaced by results for α-mixing processes because the estimating functions considered here are not martingales. Let us briefly outline the necessary limit results. By the ergodic theorem one has

$$\lim_{n_t \to \infty} \frac{1}{n_t} \sum_{n=1}^{n_t} g(\Delta, X_{\tau_{n-1}}, X_{\tau_n}; \theta) \overset{a.s.}{=} q_{\theta_0}^{\Delta}(g(\Delta, \theta)). \tag{3.10}$$

Note from Condition 3.2 the limit is zero for $\theta = \bar{\theta}$.

Under the conditions imposed, we have the following central limit theorem:

$$\lim_{n_t \to \infty} \frac{1}{\sqrt{n_t}} \sum_{n=1}^{n_t} g(\Delta, X_{\tau_{n-1}}, X_{\tau_n}; \bar{\theta}) \overset{d}{=} \tilde{F} \tag{3.11}$$

under P_{θ_0}, where

$$\tilde{F} \sim N(0, v(\theta_0, \bar{\theta}))$$

denotes a p-dimensional Gaussian random variable with mean zero and covariance matrix $v(\theta_0, \bar{\theta})$ given by (3.9), provided that the matrix $v(\theta_0, \bar{\theta})$ is strictly positive definite. This follows from Theorem 1 in Section 1.5 of Doukhan (1994) by application of the Cramér-Wold device. The condition that X is geometrically α-mixing is actually stronger than what is needed for the central limit theorem to hold. Minimal, but more technical, conditions can be found in Doukhan, Massart & Rio (1994).

It is clearly desirable to use estimating functions for which $\bar{\theta}$ is close to the true parameter value θ_0. Let us discuss this for the estimating function $\tilde{G}$ given by (2.22). In this case it follows from (2.10) that the leading term in an expansion in powers of Δ of $q^{\Delta}_{\theta_0}(g(\Delta, \theta))$, from Condition 3.2, is $\frac{1}{2}\Delta m(\theta_0, \theta)$, where

$$m_i(\theta_0, \theta) = E_{\theta_0}\left(B_i^{(1)}(\alpha, \beta)\, V_i^{11}(\alpha, \beta)^{-1}\, L_\theta^0 L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)\right). \tag{3.12}$$

We could choose the transform function U and the value of λ_i to make $m_i(\theta_0, \theta_0)$ as small as possible in an attempt to minimize the bias of the estimating function. However, we can expect to achieve a good approximation to $\bar{\theta}$ by solving the equation $m(\theta_0, \theta) = 0$ with respect to θ. By an expansion of m in θ around θ_0, we find that

$$\theta_0 - \bar{\theta} \simeq \left(\frac{\partial m(\theta_0, \theta_0)}{\partial \theta^\top}\right)^{-1} m(\theta_0, \theta_0). \tag{3.13}$$

Here $\frac{\partial m(\theta_0, \theta)}{\partial \theta^\top}$ denotes the $p \times p$-matrix, where the (i, j)th entry is $\frac{\partial m_i(\theta_0, \theta)}{\partial \theta^j}$. To reduce the distance between $\bar{\theta}$ and θ_0, it therefore seems appropriate to choose the transform function U and the values of $\lambda_1, \ldots, \lambda_p$ in a way that makes the right-hand side of (3.13) as small as possible. The estimating functions $\tilde{H}$ and $\tilde{K}$ can be treated in a similar manner, however, the resulting expressions for $m(\theta_0, \theta)$ are more complicated.

A different asymptotic scenario could have been considered, namely that Δ goes to zero sufficiently fast as n_t tends to infinity, see, for instance, Prakasa Rao (1988) and Kessler (1997). In such a scenario the estimators proposed here would be consistent. However, the kind of asymptotics studied in this section shows more clearly the advantage of the transform function approach.

4 Affine Diffusions

We now introduce a specific class of *affine diffusions* that will aid us in highlighting the features of the methodology proposed above. Consider the affine SDE for the *shifted square root process*

$$dX_t = (\theta^1 + \theta^2 X_t)dt + \sqrt{\theta^3 + \theta^4 X_t}\, dW_t \tag{4.1}$$

for $t \geq 0$, where the drift function $b(t, x; \theta) = \theta^1 + \theta^2 x$ is affine, as is the squared diffusion coefficient function $\sigma^2(t, x; \theta) = \theta^3 + \theta^4 x$. In the following, the parameter vector $\theta = (\theta^1, \theta^2, \theta^3, \theta^4)^\top \in \mathbb{R}^4$ shall be chosen such that the process $X = \{X_t, t \geq 0\}$ is ergodic. This happens when either

$$\theta^4 = 0, \qquad \theta^2 < 0 \qquad \text{and} \qquad \theta^3 > 0 \tag{4.2}$$

or

$$\theta^4 > 0, \qquad \theta^2 < 0 \qquad \text{and} \qquad \frac{2}{\theta^4}\left(\theta^1 - \frac{\theta^2\theta^3}{\theta^4}\right) \geq 1. \tag{4.3}$$

In the first case, the Ornstein-Uhlenbeck process, the process X lives on the whole real line and the stationary distribution is Gaussian with mean $-\frac{\theta^1}{\theta^2}$ and variance $-\frac{\theta^3}{(2\theta^2)}$. In the latter case the process X lives on the interval $E = (y_0, \infty)$ with $y_0 = -\frac{\theta^3}{\theta^4}$. The stationary density for such an ergodic affine diffusion is of the form

$$\bar{p}(x) = \frac{\left(\frac{-2\theta^2}{\theta^4}\right)^{\frac{2}{\theta^4}\left(\theta^1 - \frac{\theta^2\theta^3}{\theta^4}\right)} \left(x + \frac{\theta^3}{\theta^4}\right)^{\frac{2}{\theta^4}\left(\theta^1 - \frac{\theta^2\theta^3}{\theta^4}\right) - 1} \exp\left(\frac{2\theta^2}{\theta^4}\left(x + \frac{\theta^3}{\theta^4}\right)\right)}{\Gamma\left(\frac{2}{\theta^4}\left(\theta^1 - \frac{\theta^2\theta^3}{\theta^4}\right)\right)} \tag{4.4}$$

for $x \in E = (y_0, \infty)$, where $\Gamma(\cdot)$ denotes the Gamma function. In this case the stationary mean is

$$\int_{y_0}^{\infty} x\bar{p}(x)dx = -\frac{\theta^1}{\theta^2} \tag{4.5}$$

and the stationary second moment has the form

$$\int_{y_0}^{\infty} x^2\bar{p}(x)dx = -\frac{(2\theta^1 + \theta^4)\theta^1 - \theta^3\theta^2}{2(\theta^2)^2}. \tag{4.6}$$

Note that the stationary density in (4.4) is a shifted Gamma distribution.

5 Power Transform Function

To illustrate the transform function method for affine diffusions we need to specify a class of transform functions. Let us consider the *power transform function*, which is one of the most tractable transforms. We set

$$U(t, x; \lambda) = x^\lambda \tag{5.1}$$

for $t \in [0, \infty)$, $x \in E$ with $\lambda > 0$. Setting $\alpha = (\theta^1, \theta^2)$ and $\beta = (\theta^3, \theta^4)$ for constant weighting functions

$$B^{(1)}(\alpha, \beta)V^{11}(\alpha, \beta)^{-1} = 1 \tag{5.2}$$

and

$$B^{(2)}(\beta)V^{22}(\beta)^{-1} = 1, \tag{5.3}$$

we obtain from (2.20) and (2.22), the estimating functions

$$\tilde{H}_i(\beta, t, \Delta) = \frac{1}{n_t}\sum_{n=1}^{n_t} F_{i,n}^{(2)}(\beta) \tag{5.4}$$

with

$$F_{i,n}^{(2)}(\beta) = \left(Q_{\lambda_i,n,\Delta} - (L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))^2\right)$$

for $i \in \{3, 4\}$ and

$$\tilde{G}_i(\alpha, \beta, \Delta) = \frac{1}{n_t}\sum_{n=1}^{n_t} F_{i,n}^{(1)}(\alpha, \beta), \tag{5.5}$$

where

$$F_{i,n}^{(1)}(\alpha, \beta) = (D_{\lambda_i,n,\Delta} - L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))$$

for $i \in \{1, 2\}$. For the affine diffusions we have by (2.6)

$$L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) = (\theta^1 + \theta^2 X_{\tau_{n-1}})\lambda_i X_{\tau_{n-1}}^{\lambda_i - 1} + \frac{1}{2}(\theta^3 + \theta^4 X_{\tau_{n-1}})\lambda_i(\lambda_i - 1)X_{\tau_{n-1}}^{\lambda_i - 2} \tag{5.6}$$

and by (2.7)

$$L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) = \lambda_i X_{\tau_{n-1}}^{\lambda_i - 1}\sqrt{\theta^3 + \theta^4 X_{\tau_{n-1}}}. \tag{5.7}$$

We obtain from (2.9), (2.20), (5.4) and (5.7) the estimating function

$$\tilde{H}_i(\beta, t, \Delta) = A_\Delta^{0,1,0}(\lambda_i) - \theta^3(\lambda_i)^2 A_\Delta^{2(\lambda_i - 1)} - \theta^4(\lambda_i)^2 A_\Delta^{2\lambda_i - 1} \tag{5.8}$$

for $i \in \{3, 4\}$ for two different values λ_3, $\lambda_4 > 0$. Here we have used the notation

$$A_\Delta^{r,k,j}(\lambda_i) = \frac{1}{n_t}\sum_{n=1}^{n_t}(X_{\tau_{n-1}})^r(Q_{\lambda_i,n,\Delta})^k(D_{\lambda_i,n,\Delta})^j \tag{5.9}$$

and $A_\Delta^r = A_\Delta^{r,0,0}(\lambda_i)$, which refers to an equidistant time discretization of step size Δ.

Similarly, we obtain from (2.8), (2.22), (5.5) and (5.6) the estimating function

$$\tilde{G}_i(\alpha, \beta, t, \Delta) = C_\Delta(\lambda_i, \theta^3, \theta^4) - \theta^1\lambda_i A_\Delta^{\lambda_i - 1} - \theta^2\lambda_i A_\Delta^{\lambda_i} \tag{5.10}$$

with

$$C_\Delta(\lambda_i, \theta^3, \theta^4) = A_\Delta^{0,0,1}(\lambda_i) - \frac{\lambda_i(\lambda_i - 1)}{2}(\theta^3 A_\Delta^{\lambda_i - 2} + \theta^4 A_\Delta^{\lambda_i - 1}) \tag{5.11}$$

for $i \in \{1, 2\}$ for two different values $\lambda_1,\ \lambda_2 > 0$.

It follows from (2.10), properties of multiple stochastic integrals, and the existence of all moments of positive order for X that

$$\begin{aligned}
\lim_{n_t\to\infty} E(\tilde{H}_i(\beta, t, \Delta)) &= \lim_{n_t\to\infty} E\left(\frac{1}{n_t}\sum_{n=1}^{n_t}\frac{1}{\Delta}\Big[(L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))^2\Delta^2\right.\\
&+ (L_\theta^1 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))^2\frac{\Delta^2}{2}\\
&+ L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) L_\theta^1 L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)\frac{\Delta^2}{2}\\
&+ \left. L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) L_\theta^0 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)\frac{\Delta^2}{2}\Big]\right)\\
&+ \Delta^2 R_1(\theta, t, \Delta, \lambda_i)\\
&= \Delta\int_{y_0}^{\infty}\Big((L_\theta^0 U(1, y; \lambda_i))^2 + \frac{1}{2}(L_\theta^1 L_\theta^1 U(1, y; \lambda_i))^2\\
&+ \frac{1}{2}L_\theta^1 U(1, y; \lambda_i) L_\theta^1 L_\theta^0 U(1, y; \lambda_i)\\
&+ \frac{1}{2}L_\theta^1 U(1, y; \lambda_i) L_\theta^0 L_\theta^1 U(1, y; \lambda_i)\Big)\,\bar{p}(y)dy\\
&+ \Delta^2 R_2(\theta, 1, \Delta, \lambda_i).
\end{aligned} \tag{5.12}$$

Here $\bar{p}(y)$ is given by (4.4), and $R_j(\theta, t, \Delta, \lambda_i)$, for $j \in \{1, 2\}$ and $i \in \{3, 4\}$ are some finite functions. Similarly, we obtain

$$\lim_{n_t\to\infty} E(\tilde{G}_i(\alpha, \beta, t, \Delta)) = \Delta\int_{y_0}^{\infty}\left(\frac{1}{2}(L_\theta^0 L_\theta^0 U(1, y; \lambda_i))\right)\bar{p}(y)dy + \Delta^{\frac{3}{2}} R_4(\theta, 1, \Delta, \lambda_i), \tag{5.13}$$

where $R_4(\theta, t, \Delta, \lambda_i)$, for $i \in \{1, 2\}$ are finite functions.

Note that if the time between observations Δ tends to zero, then the expectation of the functions $\tilde{H}_i$ and $\tilde{G}_i$ will approach zero. By setting the estimating functions to zero we obtain the linear system of four estimating equations

$$0 = A_\Delta^{0,1,0}(\lambda_i) - \hat{\theta}^3(\lambda_i)^2 A_\Delta^{2(\lambda_i - 1)} - \hat{\theta}^4(\lambda_i)^2 A_\Delta^{2\lambda_i - 1} \tag{5.14}$$

for $i \in \{3, 4\}$ and

$$0 = C_\Delta(\lambda_i, \hat{\theta}^3, \hat{\theta}^4) - \hat{\theta}^1\lambda_i A_\Delta^{\lambda_i - 1} - \hat{\theta}^2\lambda_i A_\Delta^{\lambda_i} \tag{5.15}$$

for $i \in \{1, 2\}$.

As discussed in a previous section, the values of λ_i, $i \in \{1,2,3,4\}$ should ideally be chosen such that the estimator bias is minimized. Here we will choose them based on a more practical consideration. Intuitively, the values chosen for the λ_i, $i \in \{1,2,3,4\}$ should remain small, since large values of λ_i would result in transform functions that produce unstable estimating functions with terms that may increase rapidly over time. Furthermore, simple explicit solutions of the system of equations (5.14) and (5.15) can be obtained by choosing small integer values. A convenient choice is $\lambda_1 = \lambda_3 = 1$, $\lambda_2 = \lambda_4 = 2$. Using these values for λ_i for $i \in \{1,2,3,4\}$ we obtain the following four equations for the estimators,

$$\begin{aligned} \hat\theta^1 + A^1_\Delta \hat\theta^2 &= C_\Delta(1, \hat\theta^3, \hat\theta^4) \\ A^1_\Delta \hat\theta^1 + A^2_\Delta \hat\theta^2 &= \frac{1}{2} C_\Delta(2, \hat\theta^3, \hat\theta^4) \\ \hat\theta^3 + A^1_\Delta \hat\theta^4 &= A^{0,1,0}_\Delta(1) \\ A^2_\Delta \hat\theta^3 + A^3_\Delta \hat\theta^4 &= \frac{1}{4} A^{0,1,0}_\Delta(2). \end{aligned} \tag{5.16}$$

This system has the explicit solution

$$\begin{aligned} \hat\theta^1 &= C_\Delta(1, \hat\theta^3, \hat\theta^4) - A^1_\Delta \hat\theta^2 \\ \hat\theta^2 &= \frac{\frac{1}{2} C_\Delta(2, \hat\theta^3, \hat\theta^4) - A^1_\Delta C_\Delta(1, \hat\theta^3, \hat\theta^4)}{A^2_\Delta - (A^1_\Delta)^2} \\ \hat\theta^3 &= A^{0,1,0}_\Delta(1) - A^1_\Delta \hat\theta^4 \\ \hat\theta^4 &= \frac{\frac{1}{4} A^{0,1,0}_\Delta(2) - A^2_\Delta A^{0,1,0}_\Delta(1)}{A^3_\Delta - A^1_\Delta A^2_\Delta}. \end{aligned} \tag{5.17}$$

Here we have derived explicit expressions for the estimators of the given class of affine diffusions using power transform functions. The illustrated transform function method can be extended to other classes of diffusions including nonergodic and multi-dimensional diffusions. If no explicit solution of the system of estimating equations is available, then numerical solution techniques can be applied to identify their solution.

6 Example

To illustrate the practical applicability of the proposed transform function method, we consider an example for the above affine diffusion given in (4.1). Sample paths of this process were simulated using two different simulation schemes. Firstly, a *Wagner-Platen order* 1.5 *strong scheme*, see Kloeden & Platen (1999), was used.

The scheme has the form

$$
\begin{aligned}
X_{\tau_n} &= X_{\tau_{n-1}} + (\theta^1 + \theta^2 X_{\tau_{n-1}})\Delta + \sqrt{\theta^3 + \theta^4 X_{\tau_{n-1}}}\Delta W_n \\
&+ \frac{1}{2}\theta^2(\theta^1 + \theta^2 X_{\tau_{n-1}})\Delta^2 + \frac{1}{4}\theta^4(\Delta W_n^2 - \Delta) + \theta^2\sqrt{\theta^3 + \theta^4 X_{\tau_{n-1}}}\Delta Z_n \\
&+ \left(\frac{\theta^4(\theta^1 + \theta^2 X_{\tau_{n-1}})}{2\sqrt{\theta^3 + \theta^4 X_{\tau_{n-1}}}} + \frac{-\theta_4^2}{8\sqrt{\theta^3 + \theta^4 X_{\tau_{n-1}}}}\right)(\Delta W_n \Delta - \Delta Z_n), \qquad (6.18)
\end{aligned}
$$

where $\Delta W_n = \sqrt{\Delta}\epsilon_{1_n}$ and $\Delta Z_n = \frac{\Delta^{\frac{3}{2}}}{2}(\epsilon_{1_n} + \frac{\epsilon_{2_n}}{\sqrt{3}})$. Here ϵ_{1_n} and ϵ_{2_n} are independent, standard Gaussian random variables. Additionally, the *balanced implicit scheme*, introduced by Milstein, Platen & Schurz (1998), was also used to simulate the affine diffusion. This scheme has the form

$$
X_{\tau_n} = X_{\tau_{n-1}} + (\theta^1 + \theta^2 X_{\tau_{n-1}})\Delta + \sqrt{\theta^3 + \theta^4 X_{\tau_{n-1}}}\Delta W_n + C_n(X_{\tau_{n-1}} - X_{\tau_n}) \qquad (6.19)
$$

with ΔW_n as above. Here the function C_n was chosen to be

$$
C_n = (\theta^3 + \theta^4)|\Delta W_n|. \qquad (6.20)
$$

The affine diffusion was simulated with 20,000 steps over the time period $[0, T]$, with the parameters set to $\theta^1 = 0.01$, $\theta^2 = -0.01$, $\theta^3 = 0.01$, $\theta^4 = 0.01$, $T = 20$ and $X(0) = 1$, see Figure 1. There was no significant difference in the paths obtained when simulated by the different numerical schemes. The parameters were then estimated from the path shown in Figure 1 by application of the estimators given in (5.17), using every tenth observation only in the estimation procedure. Thus $\Delta = 0.009$, so we expect the estimator to be consistent.

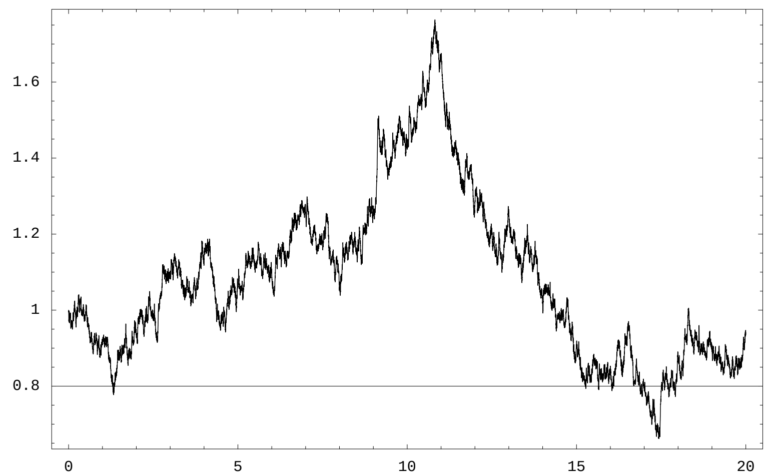

Figure 1: Sample path of the affine diffusion with $\theta^1 = 0.01$, $\theta^2 = -0.01$, $\theta^3 = 0.01$ and $\theta^4 = 0.01$.

The evolution of the estimators through time is shown in Figures 2 and 3. There is substantial variability of the drift estimators as they evolve over time. The estimates of the diffusion parameters are relatively stable as can be seen from Figure 3.

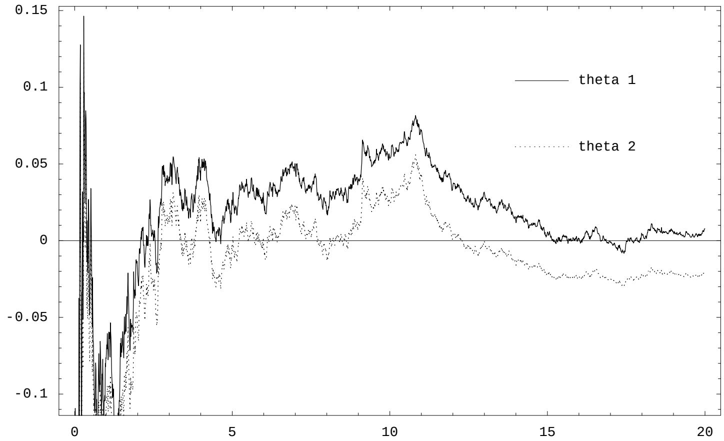

Figure 2: Estimates of the drift parameters $\hat{\theta}^1$ and $\hat{\theta}^2$.

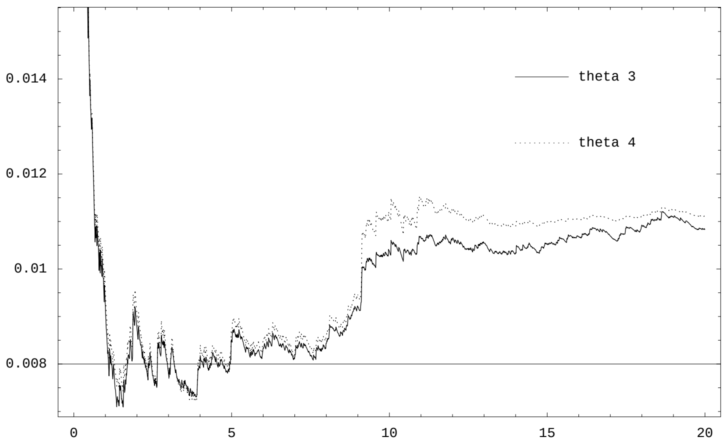

Figure 3: Estimates for the diffusion parameters $\hat{\theta}^3$ and $\hat{\theta}^4$.

To study the variability of the estimators, results from the estimation of $1,000$ simulated paths are shown in Table 1. As before, we have used every tenth observation and for comparison also every twentieth and fiftieth observation for the estimation where $\Delta = 0.019$ and $\Delta = 0.049$, respectively. The corresponding number of observations used are given in the first column of Table 1. The mean and standard deviation of the estimators are given in the corresponding columns of Table 1. It is clear that, on average, the estimates are reasonably accurate. We see that the accuracy of the estimators for the drift parameters increases as the time between observations is decreased. Additionally, there is a clear decrease in the variance of the estimates for the parameters of the diffusion coefficient. Note that the standard deviation for the estimates of the drift parameters is significantly greater than that of the estimates of the diffusion parameters. This phenomenon is usually seen for diffusion processes sampled at a high frequency, and is caused by the large amount of information contained in the sample path about the diffusion coefficient. The quadratic variation also reflects such information.

n_t	$\hat{\theta}^1$	$\hat{\theta}^2$	$\hat{\theta}^3$	$\hat{\theta}^4$
2220	0.0101	-0.02087	0.0080	0.0081
	(0.0286)	(0.0394)	(0.0006)	(0.0009)
1050	0.0102	-0.0291	0.0090	0.0090
	(0.0301)	(0.0411)	(0.0010)	(0.0012)
408	0.0102	-0.0292	0.0097	0.0096
	(0.0309)	(0.0420)	(0.0016)	(0.0018)

Table 1: Mean and standard deviation of estimators for the drift and diffusion coefficients when $T = 20$.

n_t	$\hat{\theta}^1$	$\hat{\theta}^2$	$\hat{\theta}^3$	$\hat{\theta}^4$
1110	0.0104	-0.0192	0.0086	0.0088
	(0.0407)	(0.0482)	(0.0009)	(0.0008)
525	0.0109	-0.0194	0.0094	0.0091
	(0.0421)	(0.0506)	(0.0014)	(0.0018)
204	0.0110	-0.0198	0.0096	0.0096
	(0.0442)	(0.0521)	(0.0019)	(0.0023)

Table 2: Mean and standard deviation of estimators for the drift and diffusion coefficients when $T = 10$.

The estimation of the affine diffusion process described by (4.1) was repeated for a wide range of other parameter settings with similar outcomes indicating that the derived estimation procedure is reasonably robust.

Additionally, the estimation was performed for the shorter observation period $T = 10$. The diffusion process in (4.1) was simulated with $10,000$ steps with the parameter settings as above. In order for the time between observations to be unchanged, the number of simulated points was simply halved. Results using the estimating equations in (5.17) for $1,000$ simulated paths are given in Table 2. It is clear that there is little change in the accuracy of the estimates for the diffusion coefficients. The standard deviation and mean of the estimators of the diffusion parameters are relatively unchanged. However, there was an increase in the standard deviation for the drift estimates. The need for a long observation interval to reliably estimate the drift coefficients reflects the reliance of the estimation technique on the ergodic theorem. Simulation results indicate that to accurately estimate the drift coefficient a much longer observation period than either $T = 10$ or $T = 20$ is required.

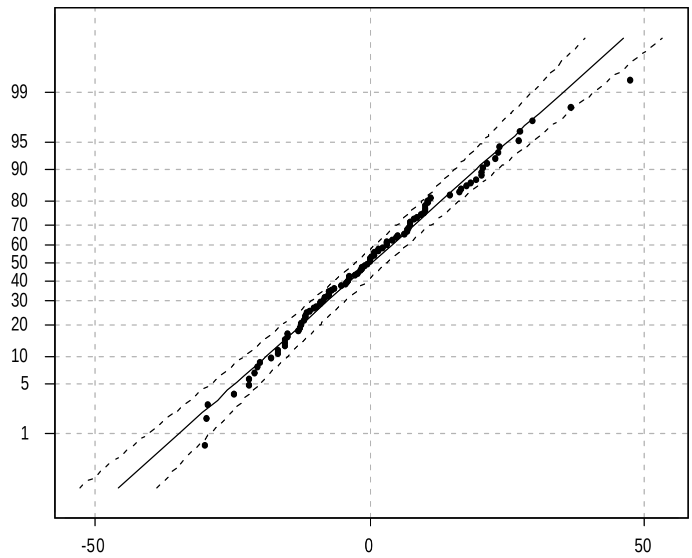

Figure 4: Gaussian quantile plot for $\hat{\theta}^1$.

The estimators in Figures 2 and 3 and Tables 1 and 2 appear to have some bias. In particular the diffusion coefficient parameters are underestimated. In theory, this could be rectified by simply using the compensated estimating functions in (2.12). However, when Δ is small, the effect in practice on the estimated parameter values is minimal due to the presence of Δ in the compensators. Alternatively, it may be possible to eliminate the leading error term, which is a factor of the time between observations, by combining estimating functions in a way that the resulting expectation is small. This was tested and the results suggest that such bias reduced estimators overcompensate the estimators for the bias in the given example.

Another, rather simple way of correcting for the bias in the method is to simulate artificially the affine diffusion with the biased parameter estimates. By correcting for the observed bias, new estimates are obtained that can be used for an improved simulation. One can repeat this procedure until the biased parameter estimates are reasonably matched by the simulated results. The parameter estimates of the artificially simulated diffusion may then be interpreted as good proxies for the

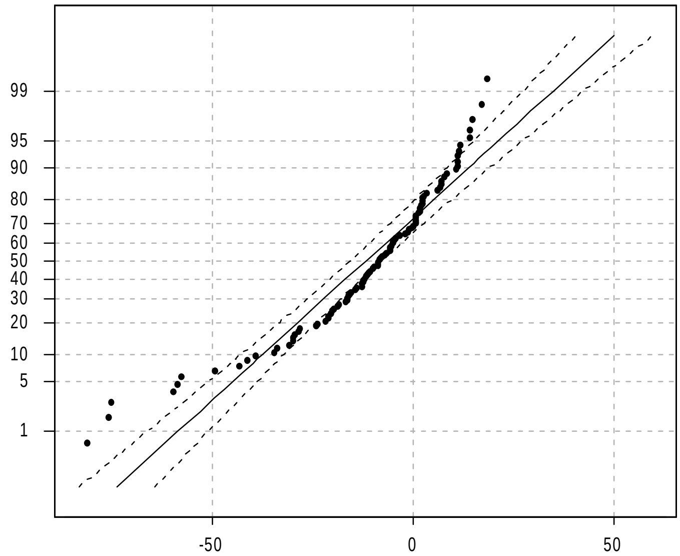

Figure 5: Gaussian quantile plot for $\hat{\theta}^2$.

The normality of the estimators is illustrated in Gaussian quantile plots shown in Figures 4, 5, 6 and 7 for the case $T = 20$. The quantile plots, which also show 95% confidence intervals, indicate that the distributions of the parameter estimates $\hat{\theta}^1$, $\hat{\theta}^3$ and $\hat{\theta}^4$ are close to Gaussian. The distribution of $\hat{\theta}^2$ exhibits some larger deviations from a Gaussian distribution. These deviations are also reflected in the estimates given in Tables 1 and 2.

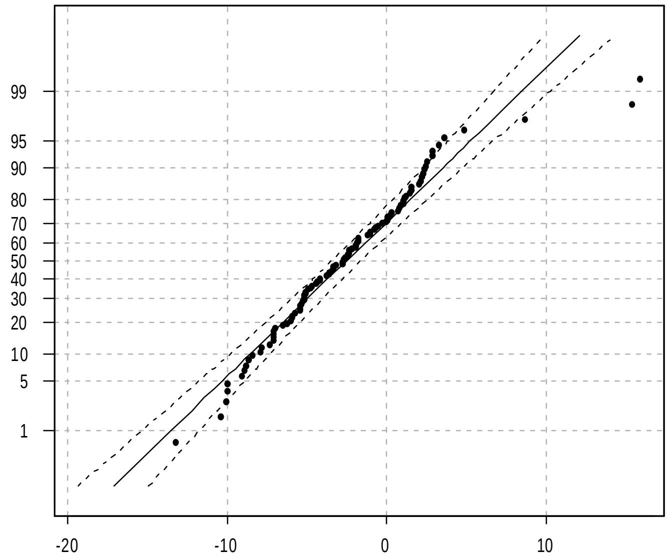

Figure 6: Gaussian quantile plot for $\hat{\theta}^3$.

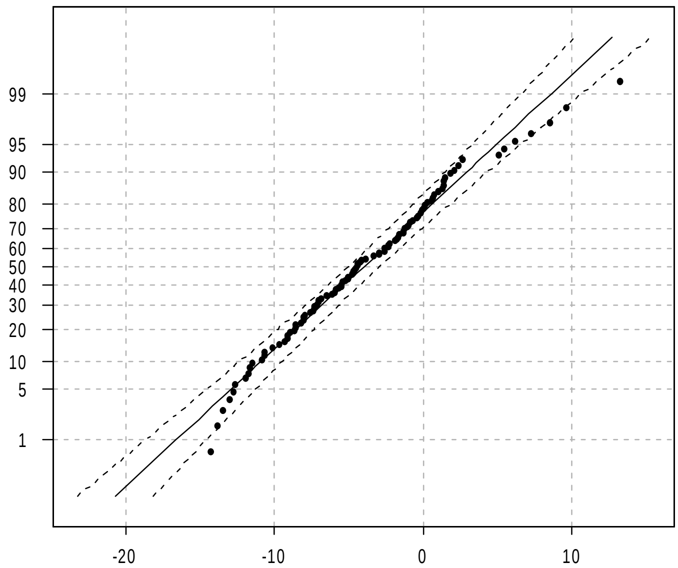

Figure 7: Gaussian quantile plot for $\hat{\theta}^4$.

7 Conclusion

We have proposed a simple and rather general estimation technique that is capable of estimating the drift and the diffusion coefficient functions of discretely observed diffusions. The transform function method presented has the advantage that it is easy to implement and does not need either explicit expressions for moments, conditional moments or transition densities of the original process nor for the transformed observations. For the case of affine diffusions the method has been demonstrated.

Acknowledgement

The research of Michael Sørensen on this paper started during a visit at the University of Technology, Sydney and was later supported by the European Commission through the Research Training Network DYNSTOCH under the Human Potential Programme, by MaPhySto, The Centre for Mathematical Physics and Stochastics, funded by a grant from The Danish National Research Foundation, and by the Centre for Analytical Finance and the Danish Mathematical Finance Network, both financed by the Danish Social Science Research Council.

References

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

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

Bibby, B. M. (1994). Optimal combination of martingale estimating functions for discretely observed diffusions. Technical report, University of Aarhus, Department of Theoretical Statistics. Research Report No. 298.

Bibby, B. M., M. Jacobsen, & M. Sørensen (2003). Estimating functions for discretely sampled diffusion-type models. In Y. Ait-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Economics.* Amsterdam: North Holland. Forthcoming.

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

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

Brandt, M. & P. Santa-Clara (2001). Simulated likelihood estimation of diffusions with application to exchange rate dynamics in incomplete markets. *J. Financial Economics*. Forthcoming.

Christensen, B. J., R. Poulsen, & M. Sørensen (2001). Optimal inference in diffusion models of the short rate of interest. Technical report, University of Aarhus, Centre for Analytical Finance. Working Paper No. 102.

Dacunha-Castelle, D. & D. Florens-Zmirou (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **19**(4), 263–284.

Dorogovcev, A. J. (1976). The consistency of an estimate of a parameter of a stochastic differential equation. *Theory Probab. Math. Statist.* **10**, 73–82.

Doukhan, P. (1994). *Mixing, Properties and Examples.* Springer.

Doukhan, P., P. Massart, & E. Rio (1994). The functional central limit theorem for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.* **30**, 63–82.

Elerain, O., S. Chib, & N. Shepard (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica* **69**, 959–993.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.* **19**(2), 177–191.

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Stochastics* **20**(4), 547–557.

Gallant, A. R. & G. Tauchen (1996). Which moments to match? *Econometric Theory* **12**, 657–681.

Genon-Catalot, V., T. Jeantheau, & C. Laredo (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6**, 1051–1079.

Godambe, V. P. & C. C. Heyde (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **55**, 231–244.

Gouriéroux, C., A. Monfort, & E. Renault (1993). Indirect inference. *J. Appl. Econometrics* **8**, 85–118.

Hansen, L. P. (1982). Large sample properties of the generalized method of moments. *Econometrica* **50**, 1029–1054.

Heyde, C. C. (1997). *Quasi-Likelihood and Its Application. A General Approach to Optimal Parameter Estimation.* Springer Ser. Statist. Springer.

Jacod, J. & A. N. Shiryaev (2003). *Limit Theorems for Stochastic Processes* (2nd ed.). Springer.

Jensen, B. & R. Poulsen (1999). A comparison of approximation techniques for transition densities of diffusion processes. Technical report, University of Aarhus, Centre for Analytical Finance. Working Paper No. 38.

Karatzas, I. & S. E. Shreve (1991). *Brownian Motion and Stochastic Calculus* (2nd ed.). Springer.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211–229.

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

Kloeden, P. E. & E. Platen (1999). *Numerical Solution of Stochastic Differential Equations*, Volume 23 of *Appl. Math.* Springer. Third corrected printing.

Kloeden, P. E., E. Platen, H. Schurz, & M. Sørensen (1996). On effects of discretization on estimators of drift parameters for diffusion processes. *J. Appl. Probab.* **33**, 1061–1076.

Kutoyants, Y. (1984). *Parameter Estimation for Stochastic Processes.* Helderman Verlag Berlin.

Liptser, R. & A. Shiryaev (2001). *Statistics of Random Processes: II. Applications* (2nd, revised and expanded ed.), Volume 6 of *Appl. Math.* Springer.

Milstein, G. N., E. Platen, & H. Schurz (1998). Balanced implicit methods for stiff stochastic systems. *SIAM J. Numer. Anal.* **35**(3), 1010–1019.

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

Poulsen, R. (1999). Approximate maximum likelihood estimation of discretely observed diffusion processes. Technical report, University of Aarhus, Centre for Analytical Finance. Working Paper No. 29.

Prakasa Rao, B. L. S. (1988). Statistical inference from sampled data for stochastic processes. *Contemp. Math.* **80**, 249–284.

Prakasa Rao, B. L. S. (1999). *Statistical Inference for Diffusion Type Processes.* Arnold Publishers.

Sørensen, H. (2001). Discretely observed diffusions: Approximation of the continuous-time score function. *Scand. J. Statist.* **28**, 113–121.

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

Sørensen, M. (1999). On asymptotics of estimating functions. *Brazilian J. Probab. Statist.* **13**, 111–136.

Sørensen, M. (2000). Prediction based estimating functions. *J. Econometrics* **3**, 123–147.