Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2007-28

The Pearson diffusions: A class of statistically tractable diffusion processes

Julie Lyng Forman and Michael Sørensen

School of Economics and Management
University of Aarhus
Building 1322, DK-8000 Aarhus C
Denmark

Aarhus School of Business
University of Aarhus
Handelshøjskolen
Aarhus Universitet

UNIVERSITY OF COPENHAGEN

The Pearson diffusions: A class of statistically tractable diffusion processes.

Julie Lyng Forman
University of Copenhagen

Michael Sørensen*
University of Copenhagen and CREATES

September 27, 2007

Abstract

The Pearson diffusions is a flexible class of diffusions defined by having linear drift and quadratic squared diffusion coefficient. It is demonstrated that for this class explicit statistical inference is feasible. Explicit optimal martingale estimating functions are found, and the corresponding estimators are shown to be consistent and asymptotically normal. The discussion covers GMM, quasi-likelihood, and non-linear weighted least squares estimation too, and it is discussed how explicit likelihood or approximate likelihood inference is possible for the Pearson diffusions. A complete model classification is presented for the ergodic Pearson diffusions. The class of stationary distributions equals the full Pearson system of distributions. Well-known instances are the Ornstein-Uhlenbeck processes and the square root (CIR) processes. Also diffusions with heavy-tailed and skew marginals are included. Special attention is given to a skew t-type distribution. Explicit formulae for the conditional moments and the polynomial eigenfunctions are derived. The analytical tractability is inherited by transformed Pearson diffusions, integrated Pearson diffusions, sums of Pearson diffusions, and stochastic volatility models with Pearson volatility process. For the non-Markov models explicit optimal prediction based estimating functions are found and shown to yield consistent and asymptotically normal estimators.

Key words: eigenfunction, ergodic diffusion, integrated diffusion, martingale estimating function, likelihood inference, mixing, optimal estimating function, Pearson system, prediction based estimating function, quasi likelihood, spectral methods, stochastic differential equation, stochastic volatility.

JEL codes: C22, C51.

*Michael Sørensen acknowledges support from Center for Research in Econometric Analysis of Time Series, CREATES, www.creates.au.dk, funded by the Danish National Research Foundation and from the Danish Center for Accounting and Finance funded by the Danish Social Science Research Council.

1 Introduction.

In applications of diffusions the Ornstein-Uhlenbeck process and the square-root process (a.k.a. the CIR process) are often used, more because of their tractability than because they fit the data particularly well. The aim of this paper is to communicate that these two diffusion processes belong to a versatile class of tractable diffusion models, which we call the Pearson diffusions. For these diffusion models moments and conditional moments can be calculated explicitly. Moreover, the optimal martingale estimating functions based on eigenfunctions of the generator, introduced by Kessler & Sørensen (1999), can be found explicitly. Thus statistical inference using this method is straightforward. Recently, Sørensen (2007) has proved that optimal martingale estimating functions give estimators that are efficient in a high frequency asymptotics and therefore often offer a simpler alternative to maximum likelihood estimation with high efficiency. This is certainly the case for many financial data for which the speed of mean reversion is slow so that, for instance, daily observations can be considered high frequency sampling. Because of the versatility of the Pearson class of diffusions, it is often reasonable to use a Pearson diffusion as statistical model instead of another similar, but less tractable, diffusion process. By choosing a Pearson diffusion model and using an optimal martingale estimating function an estimation method can be achieved that is both of high efficiency and easy to implement. Other estimation methods based on conditional moments like the generalized method of moments, quasi-likelihood and non-linear weighted least squares estimation are also easy to use and are covered in the discussion as particular cases. Likelihood or approximate likelihood inference is more complicated to implement, but is relatively simple for the Pearson diffusions compared to most other diffusion models. Parameter estimation is also easy for some diffusion-type models obtained using the Pearson diffusions as building blocks such as transformations and sums of Pearson diffusions, integrated Pearson diffusions, and Pearson stochastic volatility models. Most of these models are non-Markovian processes, for which we derive explicit optimal prediction-based estimating functions, see Sørensen (2000).

We shall use the term Pearson diffusion for any stationary solution of a stochastic differential equation specified by a mean reverting linear drift and a squared diffusion coefficient which is a second order polynomial of the state. The motivation is that when a stationary solution exists, then its invariant density belongs to the Pearson system, Pearson (1895). In fact, the class of stationary distributions equals the full Pearson system of distributions. The class of Pearson diffusions is thus highly flexible and therefore suited for many different applications. Just like the Pearson densities the diffusions can be positive, negative, real valued, or bounded, symmetric or skewed, and heavy- or light-tailed. We give special attention to the Pearson diffusion with type IV marginals (the type IV Pearson distribution is a skewed kind of t-distribution). This process has received little attention in the literature, see however Nagahara (1996), and has a noteworthy potential in, for instance, financial applications because of its skew and heavy-tailed marginal distribution. The basic reason for the tractability of the Pearson diffusions is that the generator maps polynomials into polynomials of (at most) the same degree.

Most of the Pearson diffusions were derived and studied from a probabilistic viewpoint by Wong (1964) using a different approach and with another aim. In particular, he did not consider the nice statistical properties of the Pearson diffusions on which the present paper focuses. The Pearson diffusions form a subset of the class of jump-diffusions investigated by Zhou (2003), who derived formulae for conditional polynomial moments and used these for generalized method of moments estimation. Most of the Pearson diffusions are among the diffusion models studied in Bibby, Skovgaard & Sørensen (2005), where no attention was, however, given to statistical inference. The affine diffusions form a subclass of the Pearson diffusions. Affine processes have attracted a lot of attention because for these an expression is available for the characteristic function of the transition distribution which implies a number of explicit estimators, see e.g. Singleton (2001), Chacko & Viceira (2003) and Carrasco et al. (2002). General affine processes were studied thoroughly by Duffie, Filipović & Schachermayer (2003). Note, however, that in the context of the one-dimensional processes without jumps considered in this paper, the class of affine processes consist only of a small subset of the Pearson diffusions, namely the Ornstein-Uhlenbeck process and the CIR-process and affine transformations of these. Meddahi (2001), Meddahi (2002b), and Meddahi (2002a) has proposed and studied an eigenfunction approach to stochastic volatility modelling, where the volatility process is a weighted sum of eigenfunctions of a diffusion process applied to that process (or a similar construction for two processes). In this context he considered the Pearson diffusions.

The paper is organized as follows. In Section 2 we give a complete classification of the Pearson diffusions and demonstrate their tractability. We show that all Pearson diffusions have polynomial eigenfunctions that can be found explicitly. It is also demonstrated that transformations of Pearson diffusions are similarly tractable so that estimation is easy for such models too. Likelihood inference and statistical inference based on conditional moments are studied in Section 3. Optimal martingale estimating functions based on eigenfunctions of the generator are investigated in detail including asymptotic results. This discussion also covers other estimation methods such as the generalized method of moments, quasi-likelihood and non-linear weighted least squares estimation. A simulation study investigates the efficiency of the estimators. In Section 4 we explicitly find optimal prediction-based estimating functions for integrated Pearson diffusions, for sums of Pearson diffusions and for stochastic volatility models where the volatility process is a Pearson diffusion or a sum of Pearson diffusions. Also asymptotics for these models are considered. Proofs of the asymptotic results are given in the appendix.

2 The Pearson diffusions.

A Pearson diffusion is a stationary solution to a stochastic differential equation of the form

$$dX_t = -\theta(X_t - \mu)dt + \sqrt{2\theta(aX_t^2 + bX_t + c)}dB_t, \tag{2.1}$$

where $\theta > 0$, and where a, b and c are such that the square root is well defined when X_t is in the state space. The parameters of (2.1) are referred to as the canonical parameterisation: $\theta > 0$ is a scaling of time that determines how fast the diffusion moves. The parameters μ, a, b, and c determine the state space of the diffusion as well as the shape of the invariant distribution. In particular, μ is the mean of the invariant distribution.

Let us first briefly outline, why the stationary density of the diffusion (2.1) belongs to the Pearson system. The scale and speed densities of the diffusion (2.1) are

$$s(x) = \exp\left(\int_{x_0}^{x} \frac{u-\mu}{au^2+bu+c}du\right) \quad \text{and} \quad m(x) = \frac{1}{s(x)(ax^2+bx+c)}$$

where x_0 is a fixed point such that $ax_0^2 + bx_0 + c > 0$. Let (l, r) be an interval such that $ax^2 + bx + c > 0$ for all $x \in (l, r)$. A unique ergodic weak solution to (2.1) with values in the interval $(l, r) \ni x_0$ exists if and only if $\int_{x_0}^{r} s(x)dx = \infty$, $\int_{l}^{x_0} s(x)dx = \infty$, and $\int_l^r m(x)dx < \infty$. Its invariant distribution has density proportional to the speed density, $m(x)$. Since

$$\frac{dm(x)}{dx} = -\frac{(2a+1)x - \mu + b}{ax^2+bx+c} m(x),$$

we see that when a stationary solution to (2.1) exists, the invariant distribution belongs to the Pearson system, which is defined as the class of probability densities obtained by solving a differential equation of this form. If $\int_{x_0}^{r} s(x)dx < \infty$, the boundary l can with positive probability be reached in finite time. In this case a solution for which the invariant distribution has density proportional to the speed density is obtained if the boundary l is made instantaneously reflecting. Similarly for the other boundary, r.

2.1 Classification of the stationary solutions.

In the following we present a full classification of the ergodic Pearson diffusions. Needless to say, the squared diffusion coefficient must be positive on the state space of the diffusion. We consider six cases according to whether the squared diffusion coefficient is constant, linear, a convex parabola with either zero, one or two roots, or a concave parabola with two roots. The classification problem can be reduced by first noting that the Pearson class of diffusions is closed under translations and scale-transformations. To be specific, if $(X_t)_{t\geq 0}$ is an ergodic Pearson diffusion, then so is $(\tilde{X}_t)_{t\geq 0}$ where $\tilde{X}_t = \gamma X_t + \delta$. The parameters of the stochastic differential equation (2.1) for $(\tilde{X}_t)_{t\geq 0}$ are $\tilde{a} = a$, $\tilde{b} = b\gamma - 2a\delta$, $\tilde{c} = c\gamma^2 - b\gamma\delta + a\delta^2$, $\tilde{\theta} = \theta$, and $\tilde{\mu} = \gamma\mu + \delta$.
Hence, up to translation and transformation of scale the ergodic Pearson diffusions can take the following forms. Note that we consider scale transformations in a general sense where multiplication by -1 is allowed, so that to each case of a diffusion with state space $(0, \infty)$ there corresponds a diffusion with state space $(-\infty, 0)$. Note also that the enumeration of cases does not correspond to the types of the Pearson system.

Case 1: $\sigma^2(x) = 2\theta$**.**
For all $\mu \in \mathbb{R}$ there exists a unique ergodic solution to (2.1). It is an Ornstein-Uhlenbeck process, and the invariant distribution is the normal distribution with mean μ and variance 1. In the finance literature this model is sometimes referred to as the Vasiček model.

Case 2: $\sigma^2(x) = 2\theta x$**.**
A unique ergodic solution to (2.1) on the interval $(0, \infty)$ exists if and only if $\mu > 1$. The invariant distribution is the gamma distribution with scale parameter 1 and shape parameter μ. In particular μ is the mean of the invariant distribution. If $0 < \mu \leq 1$, the boundary 0 can with positive probability be reached at a finite time point, but if the boundary is made instantaneously reflecting, we obtain a stationary process for which

the invariant distribution is the gamma distribution with scale parameter 1 and shape parameter μ. The process goes back to Feller (1951), who introduced it as a model of population growth. It is often referred to as the square-root process. In the finance literature it is often refereed to as the CIR-process; Cox, Ingersoll & Ross (1985).

Case 3: $a > 0$ **and** $\sigma^2(x) = 2\theta a(x^2 + 1)$**.**
The scale and speed densities are given by $s(x) = (x^2+1)^{\frac{1}{2a}} \exp(-\frac{\mu}{a}\tan^{-1} x)$ and $m(x) = (x^2+1)^{-\frac{1}{2a}-1}\exp(\frac{\mu}{a}\tan^{-1} x)$. Hence, for all $a > 0$ and all $\mu \in \mathbb{R}$ a unique ergodic solution to (2.1) exists on the real line. If $\mu = 0$ the invariant distribution is a scaled t-distribution with $\nu = 1 + a^{-1}$ degrees of freedom and scale parameter $\nu^{-\frac{1}{2}}$. If $\mu \neq 0$ the invariant distribution is skew and has tails decaying at the same rate as the t-distribution with $1 + a^{-1}$ degrees of freedom. A fitting name for this distribution is the skew t-distribution. It is also known as Pearson's type IV distribution. In either case the mean is μ and the invariant distribution has moments of order k for $k < 1 + a^{-1}$. With its skew and heavy tailed marginal distribution, the class of diffusions with $\mu \neq 0$ is potentially very useful in many applications, e.g. finance. It was studied and fitted to the Nikkei 225 index, the TOPIX index and the Standard and Poors 500 index by Nagahara (1996) using the local linearization method of Ozaki (1985). The skew t-distribution with mean zero, ν degrees of freedom, and skewness parameter ρ has (unnormalized) density

$$f(z) \propto \{(z/\sqrt{\nu} + \rho)^2 + 1\}^{-(\nu+1)/2} \exp\left\{\rho(\nu-1)\tan^{-1}\left(z/\sqrt{\nu}+\rho\right)\right\}, \qquad (2.2)$$

which is the invariant density of the diffusion $Z_t = \sqrt{\nu}(X_t - \rho)$ with $\nu = 1 + a^{-1}$ and $\rho = \mu$. An expression for the normalizing constant when ν is integer valued was derived in Nagahara (1996). By the transformation result above, the corresponding stochastic differential equation is

$$dZ_t = -\theta Z_t dt + \sqrt{2\theta(\nu-1)^{-1}\{Z_t^2 + 2\rho\nu^{\frac{1}{2}} Z_t + (1+\rho^2)\nu\}}dB_t. \qquad (2.3)$$

For $\rho = 0$ the invariant distribution is the t-distribution with ν degrees of freedom. Figure 2.1 shows the density for a range of ρ values.

Case 4: $a > 0$ **and** $\sigma^2(x) = 2\theta a x^2$**.**
The scale and speed densities are $s(x) = x^{\frac{1}{a}}\exp(\frac{\mu}{ax})$ and $m(x) = x^{-\frac{1}{a}-2}\exp(-\frac{\mu}{ax})$. The integrability conditions hold if and only if $\mu > 0$. Hence, for all $a > 0$ and all $\mu > 0$ a unique ergodic solution to (2.1) exists on the positive halfline. The invariant distribution is an inverse gamma distribution with shape parameter $1 + \frac{1}{a}$ and scale parameter $\frac{a}{\mu}$. In particular the mean is μ and the invariant distribution has moments of order k for $k < 1 + \frac{1}{a}$. This process is sometimes referred to as the GARCH diffusion model.

Case 5: $a > 0$ **and** $\sigma^2(x) = 2\theta a x(x + 1)$**.**
The scale and speed densities are $s(x) = (1+x)^{\frac{\mu+1}{a}} x^{-\frac{\mu}{a}}$ and $m(x) = (1+x)^{-\frac{\mu+1}{a}-1} x^{\frac{\mu}{a}-1}$. The integrability conditions hold if and only if $\frac{\mu}{a} \geq 1$. Hence, for all $a > 0$ and all $\mu \geq a$ a unique ergodic solution to (2.1) exists on the positive halfline. The invariant distribution is a scaled F-distribution with $\frac{2\mu}{a}$ and $\frac{2}{a} + 2$ degrees of freedom and scale parameter $\frac{\mu}{1+a}$. In particular the mean is μ and the invariant distribution has moments of order k for $k < 1 + \frac{1}{a}$. If $0 < \mu < 1$, the boundary 0 can with positive probability be reached

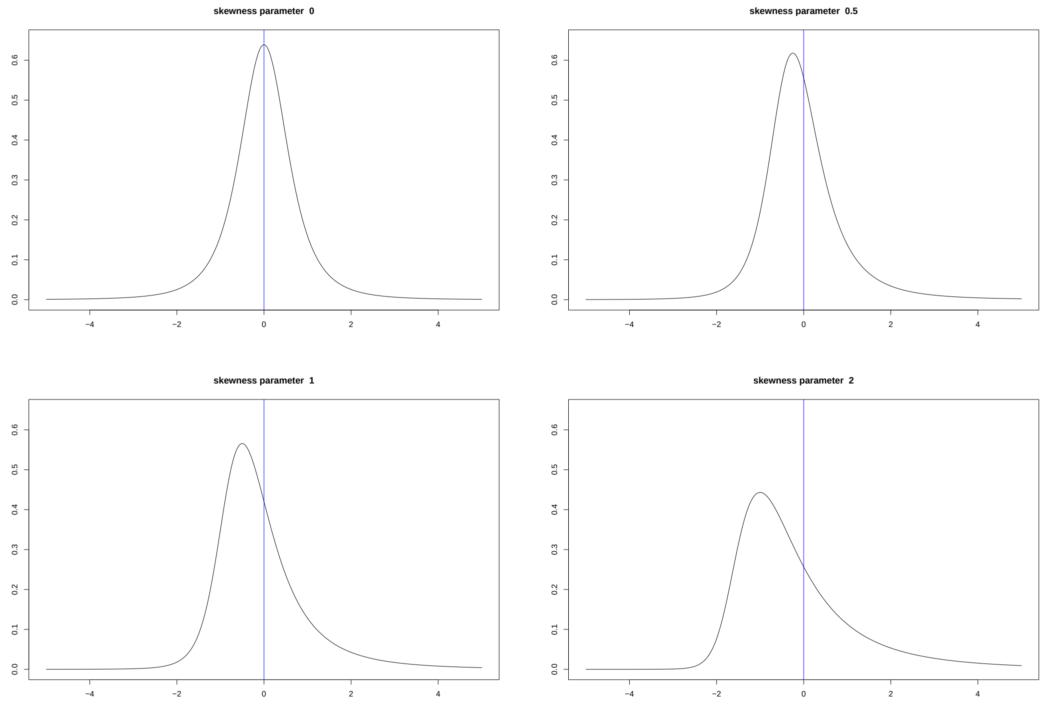

Figure 2.1: Densities of skew t-distributions (Pearson type IV distributions) with zero mean for $\rho = 0$, 0.5, 1, and 2 respectively.

at a finite time point, but if the boundary is made instantaneously reflecting, a stationary process is obtained for which the invariant distribution is the indicated F-distribution.

Case 6: $a < 0$ **and** $\sigma^2(x) = 2\theta a x(x-1)$**.**
The scale and speed densities are $s(x) = (1-x)^{\frac{1-\mu}{a}} x^{\frac{\mu}{a}}$ and $m(x) = (1-x)^{-\frac{1-\mu}{a}-1} x^{-\frac{\mu}{a}-1}$. The integrability conditions hold if and only if $\frac{\mu}{a} \leq -1$ and $\frac{1-\mu}{a} \leq -1$. Hence, for all $a < 0$ and all $\mu > 0$ such that $\min(\mu, 1-\mu) \geq -a$ a unique ergodic solution to (2.1) exists on the interval $(0,1)$. The invariant distribution is a Beta distribution with shape parameters $\frac{\mu}{-a}$, $\frac{1-\mu}{-a}$. In particular the mean is μ. If $0 < \mu < -a$, the boundary 0 can with positive probability be reached at a finite time point, but if the boundary is made instantaneously reflecting, a stationary process is obtained with the indicated Beta distribution as invariant distribution. Similar remarks apply to the boundary 1 when $0 < 1-\mu < -a$. These diffusions are often referred to as the Jacobi diffusions because the related eigenfunctions are Jacobi polynomials, see below. The model was used (after a position and scale transformation) by De Jong, Drost & Werker (2001) (with $\mu = \frac{1}{2}$) and Larsen & Sørensen (2007) to model the logarithm of exchange rates in a target zone. Multivariate Jacobi diffusions were considered by Gourieroux & Jasiak (2006).

2.2 Mixing and moments.

Common to the stationary solutions of (2.1) is that they are *ergodic and ρ-mixing with exponentially decaying mixing coefficients*. This follows from Genon-Catalot, Jeantheau

& Laredo (2000) theorem 2.6 by the fact that the drift is linear, see Hansen, Scheinkman & Touzi (1998), section 5. If the marginal distribution has finite second order moment, the linear drift implies, moreover, that the *autocorrelation function* is given by

$$r(t) = \mathrm{Cor}(X_s, X_{s+t}) = e^{-\theta t}$$

see for instance Bibby, Skovgaard & Sørensen (2005). Another important and appealing feature is that explicit expressions of the marginal and conditional moments can be found. We saw in subsection 2.1 that $E(|X_t|^\kappa) < \infty$ if and only if $a < (\kappa - 1)^{-1}$. Thus if $a \leq 0$ all moments exist, while for $a > 0$ only the moments satisfying that $\kappa < a^{-1} + 1$ exist. In particular, the expectation always exists. By Ito's formula

$$dX_t^n = -\theta n X_t^{n-1}(X_t - \mu)dt + \theta n(n-1)X_t^{n-2}(aX_t^2 + bX_t + c)dt + nX_t^{n-1}\sigma(X_t)dB_t, \quad (2.4)$$

and if $E(X_t^{2n})$ is finite, i.e. if $a < (2n-1)^{-1}$, the integral of the last term is a martingale. Thus, the *moments* of the invariant distribution satisfy

$$E(X_t^n) = a_n^{-1}\{b_n \cdot E(X_t^{n-1}) + c_n \cdot E(X_t^{n-2})\} \quad (2.5)$$

where $a_n = n\{1-(n-1)a\}\theta$, $b_n = n\{\mu+(n-1)b\}\theta$, and $c_n = n(n-1)c\theta$ for $n = 0, 1, 2, \ldots$. Initial conditions are given by $E(X_t^0) = 1$, and $E(X_t) = \mu$.

Example 2.1 Equation (2.5) allows us to find the moments of the skewed t-distribution, in spite of the fact that the normalising constant of the density (2.2) is unknown. In particular, for the diffusion (2.3), $E(Z_t^2) = \mathrm{Var}(Z_t) = \frac{(1+\rho^2)\nu}{\nu-2}$,

$$E(Z_t^3) = \frac{4\rho(1+\rho^2)\nu^{\frac{3}{2}}}{(\nu-3)(\nu-2)}, \quad E(Z_t^4) = \frac{24\rho^2(1+\rho^2)\nu^2 + 3(\nu-3)(1+\rho^2)^2\nu^2}{(\nu-4)(\nu-3)(\nu-2)}.$$

Recall that the mean of Z_t is zero. △

The *conditional moments* $q_n(x,t) = E(X_t^n | X_0 = x)$ satisfy the recursive system of first order linear differential equations

$$\frac{d}{dt}q_n(x,t) = -a_n q_n(x,t) + b_n q_{n-1}(x,t) + c_n q_{n-2}(x,t).$$

This follows from (2.4), again under the condition that the $2n$'th moment is finite. Solving for the initial condition $q_n(x,0) = x^n$ yields

$$q_n(x,t) = x^n e^{-a_n t} + b_n I_{n-1}(a_n, x, t) + c_n I_{n-2}(a_n, x, t)$$

where $I_\eta(\alpha, x, t) = \exp(-\alpha t)\int_0^t e^{\alpha s} q_\eta(x,s)ds$. Using once more the recursion, we get

$$I_\eta(\alpha) = \frac{x^\eta\{e^{-a_\eta t} - e^{-\alpha t}\} + b_\eta\{I_{\eta-1}(a_\eta) - I_{\eta-1}(\alpha)\} + c_\eta\{I_{\eta-2}(a_\eta) - I_{\eta-2}(\alpha)\}}{\alpha - a_\eta}.$$

To calculate $I_1(\alpha, x, t)$ we use that $I_0(\alpha, x, t) = \alpha^{-1}\{1 - e^{-\alpha t}\}$ as $q_0(x,t) = 1$ and that $c_1 = 0$. We see that $q_n(x,t)$ is a polynomial of order n in x for any fixed t. A somewhat easier derivation of this result comes by means of the eigenfunctions considered below.

2.3 Eigenfunctions.

Recall that for a diffusion process

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t$$

the generator is the second order differential operator

$$L = b(x)\frac{d}{dx} + \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2}.$$

A function h is an eigenfunction if there exist a positive number $\lambda > 0$, an eigenvalue, such that $Lh = -\lambda h$. Under mild regularity conditions, see e.g. Kessler & Sørensen (1999), it follows from Ito's formula that

$$E(h(X_t)|X_0 = x) = e^{-\lambda t}h(x). \tag{2.6}$$

This relationship can be used to construct martingale estimating functions. In case of the Pearson diffusions that have a linear drift and a quadratic squared diffusion coefficient, the generator maps polynomial into polynomials. It is therefore natural to search for eigenfunctions among the polynomials

$$p_n(x) = \sum_{j=0}^{n} p_{n,j}x^j.$$

The polynomial $p_n(x)$ is an eigenfunction if an eigenvalue $\lambda_n > 0$ exist satisfying that $\theta(ax^2 + bx + c)p_n''(x) - \theta(x - \mu)p_n'(x) = -\lambda_n p_n(x)$, i.e.

$$\sum_{j=0}^{n}\{\lambda_n - a_j\}p_{n,j}x^j + \sum_{j=0}^{n-1} b_{j+1}p_{n,j+1}x^j + \sum_{j=0}^{n-2} c_{j+2}p_{n,j+2}x^j = 0.$$

where $a_j = j\{1 - (j-1)a\}\theta$, $b_j = j\{\mu + (j-1)b\}\theta$, and $c_j = j(j-1)c\theta$ for $j = 0, 1, 2, \ldots$. Without loss of generality, we assume $p_{n,n} = 1$. Thus, equating the coefficients we find that the eigenvalue is given by $\lambda_n = a_n = n\{1-(n-1)a\}\theta$. If further we define $p_{n,n+1} = 0$, then the coefficients $\{p_{n,j}\}_{j=0,\ldots,n-1}$ solve the linear system

$$(a_j - a_n)p_{n,j} = b_{j+1}p_{n,j+1} + c_{j+2}p_{n,j+2} \tag{2.7}$$

Equation (2.7) is equivalent to a simple recursive formula if $a_n - a_j \neq 0$ for all $j = 0, 1, \ldots, n-1$. Note that $a_n - a_j = 0$ if and only if there exists an integer $n-1 \leq m < 2n-1$ such that $a = m^{-1}$ and $j = m - n + 1$. In particular, $a_n - a_j = 0$ cannot occur if $a < (2n-1)^{-1}$. It is important to notice that λ_n is positive if and only if $a < (n-1)^{-1}$. This is exactly the condition ensuring that $p_n(x)$ is integrable with respect to the invariant distribution. If the stronger condition $a < (2n-1)^{-1}$ is satisfied, the first n eigenfunctions belong to the space of functions that are square integrable with respect to the invariant distribution, and they are orthogonal with respect to the usual inner product in this space. The space of functions that are square integrable with respect to the invariant distribution (or a subset of this space) is often taken as the domain of the generator.

From equation (2.6) the conditional moments can be derived. Sufficient conditions that (2.6) holds is that the drift and diffusion coefficients, b and σ, are of linear growth and that the eigenfunction h is of polynomial growth, see e.g. Kessler & Sørensen (1999). These conditions are clearly satisfied here. Thus,

$$E(X_t^n \,|\, X_0 = x) = e^{-a_n t} \sum_{j=0}^{n} p_{n,j} x^j - \sum_{j=0}^{n-1} p_{n,j} E(X_t^j | X_0 = x). \tag{2.8}$$

For any fixed t the conditional expectation is a polynomial of order n in x the coefficients of which are linear combinations of $1, e^{-\lambda_1 t}, \ldots, e^{-\lambda_n t}$. Let $\lambda_0 = 0$ and

$$E(X_t^n \,|\, X_0 = x) = q_n(x,t) = \sum_{k=0}^{n} q_{n,k}(t) x^k = \sum_{k=0}^{n} \sum_{l=0}^{n} q_{n,k,l} \cdot e^{-\lambda_l t} \cdot x^k. \tag{2.9}$$

Initially $q_0(x,t) = 1$. From the above it follows that $q_{n,n}(t) = e^{-a_n t}$ and for $k = 0, \ldots, n-1$,

$$q_{n,k}(t) = p_{n,k} e^{-a_n t} - \sum_{j=k}^{n-1} p_{n,j} q_{j,k}(t). \tag{2.10}$$

In particular, $q_{n,k,n} = p_{n,k}$ and $q_{n,k,l} = -\sum_{j=l}^{n-1} p_{n,j} q_{j,k,l}$ for $l = 0, \ldots, n-1$.
For the diffusions of form (2.1) with $a \leq 0$ there are infinitely many polynomial eigenfunctions. In these cases the eigenfunctions are well-known families of orthogonal polynomials. In case 1, where the marginal distribution is the normal distribution, the eigenfunctions are the Hermite polynomials. In case 2, with gamma marginals, the eigenfunctions are the Laguerre polynomials, and finally in case 6, where the marginals are Beta-distributions, the eigenfunctions are Jacobi polynomials (on the interval $(0,1)$). For these cases all moments of the marginal distribution exists.
In the remaining cases, 3, 4, and 5, $a > 0$ which implies that there is only a finite number of polynomial eigenfunctions. The number is the integer part of $1 + a^{-1}$, which is also the order of the highest finite moment of the marginal distribution. In these cases the marginal distributions are the inverse gamma distributions, the F-distributions, and the skew (and symmetric) t-distributions, respectively. Wong (1964) showed that the spectrum of these diffusions is not discrete. The spectrum has a continuous part with eigenvalues that are larger than those in the discrete part of the spectrum. The polynomials associated with the inverse gamma distribution are known as the Bessel polynomials. It is of some historical interest that Hildebrandt (1931) derived the polynomials above from the viewpoint of Gram-Charlier expansions associated with the Pearson system. Some special cases had previously been derived by Romanovsky (1924).

Example 2.2 The skew t-diffusion (2.3) has the eigenvalues $\lambda_n = n(\nu - n)(\nu - 1)^{-1}\theta$ for $n < \nu$. The four first eigenfunctions are $p_1(z) = z$,

$$\begin{aligned}
p_2(z) &= z^2 - \frac{4\rho\nu^{\frac{1}{2}}}{\nu - 3} z - \frac{(1+\rho^2)\nu}{\nu - 2}, \\
p_3(z) &= z^3 - \frac{12\rho\nu^{\frac{1}{2}}}{\nu - 5} z^2 + \frac{24\rho^2\nu + 3(1+\rho^2)\nu(\nu - 5)}{(\nu - 5)(\nu - 4)} z + \frac{8\rho(1+\rho^2)\nu^{\frac{3}{2}}}{(\nu - 5)(\nu - 3)},
\end{aligned}$$

and

$$\begin{aligned} p_4(z) &= z^4 - \frac{24\rho\nu^{\frac{1}{2}}}{\nu-7}z^3 + \frac{144\rho^2\nu - 6(1+\rho^2)\nu(\nu-7)}{(\nu-7)(\nu-6)}z^2 \\ &+ \frac{8\rho(1+\rho^2)\nu^{\frac{3}{2}}(\nu-7) + 48\rho(1+\rho^2)\nu^{\frac{3}{2}}(\nu-6) - 192\rho^3\nu^{\frac{3}{2}}}{(\nu-7)(\nu-6)(\nu-5)}z \\ &+ \frac{3(1+\rho^2)^2\nu(\nu-7) - 72\rho^2(1+\rho^2)\nu^2}{(\nu-7)(\nu-6)(\nu-4)}, \end{aligned}$$

provided that $\nu > 4$. Conditional moments are readily obtained from equation (2.8). The most simple cases are $E(Z_t|Z_0 = z) = ze^{-\theta t}$ and

$$E(Z_t^2|Z_0 = z) = e^{-\frac{2\nu-4}{\nu-1}\theta t}z^2 + \frac{4\rho\nu^{\frac{1}{2}}}{\nu-3}(e^{-\theta t} - e^{-\frac{2\nu-4}{\nu-1}\theta t})z + \frac{(1+\rho^2)\nu}{\nu-2}(1 - e^{-\frac{2\nu-4}{\nu-1}\theta t}).$$

These formulae are used in Examples 4.1 and 4.4 below. △

2.4 Transformations.

For any diffusion obtained from a solution to (2.1) by a twice differentiable and invertible transformation T, the eigenfunctions of the generator are $p_n\{T^{-1}(x)\}$, which have the same eigenvalues as the original eigenfunctions p_n. Thus the estimation methods discussed below can be used for the much broader class of diffusions obtained by such transformations. Their stochastic differential equations can, of course, be found by Ito's formula. We will just give a couple of examples.

Example 2.3 For the Jacobi-diffusion (case 6) with $\mu = -a = \frac{1}{2}$, i.e.

$$dX_t = -\theta(X_t - \tfrac{1}{2})dt + \sqrt{\theta X_t(1-X_t)}dW_t$$

the invariant distribution is the uniform distribution on $(0,1)$ for any $\theta > 0$. For any strictly increasing and twice differentiable distribution function F we therefore have a class of diffusions given by $Y_t = F^{-1}(X_t)$ or

$$dY_t = -\theta\frac{(F(Y_t) - \frac{1}{2})f(Y_t)^2 + \frac{1}{2}F(Y_t)\{1-F(Y_t)\}}{f(Y_t)^3}dt + \frac{\theta F(Y_t)\{1-F(Y_t)\}}{f(Y_t)}dW_t,$$

which has invariant distribution with density $f = F'$. A particular example is the logistic distribution

$$F(x) = \frac{e^x}{1+e^x} \quad x \in \mathbb{R},$$

for which

$$dY_t = -\theta\left\{\sinh(x) + 8\cosh^4(x/2)\right\}dt + 2\sqrt{\theta}\cosh(x/2)dW_t.$$

If the same transformation $F^{-1}(y) = \log(y/(1-y))$ is applied to the general Jacoby diffusion (case 6), then we obtain

$$dX_t = -\theta\left\{1 - 2\mu + (1-\mu)e^x - \mu e^{-1} - 8a\cosh^4(x/2)\right\}dt + 2\sqrt{-a\theta}\cosh(x/2)dW_t,$$

a diffusion for which the invariant distribution is the generalized logistic distribution with density

$$f(x) = \frac{e^{\alpha x}}{(1+e^x)^{\alpha+\beta} B(\alpha,\beta)}, \quad x \in \mathbb{R},$$

where $\alpha = -(1-\mu)/a$, $\beta = \mu/a$ and B denotes the Beta-function. This distribution was introduced and studied in Barndorff-Nielsen, Kent & Sørensen (1982). △

Example 2.4 Let again X be a general Jacobi-diffusion (case 6). If we apply the transformation $T(x) = \sin^{-1}(2x-1)$ to X_t we obtain the diffusion

$$dY_t = -\rho \frac{\sin(Y_t) - \varphi}{\cos(Y_t)} dt + \sqrt{-a\theta/2} dW_t,$$

where $\rho = \theta(1 + a/4)$ and $\varphi = (2\mu - 1)/(1 + a/4)$. The state space is $(-\pi/2, \pi/2)$. The model was proposed and studied in Kessler & Sørensen (1999) for $\varphi = 0$, where the drift is $-\rho \tan(x)$. The general asymmetric version was proposed in Larsen & Sørensen (2007) as a model for exchange rates in a target zone. △

3 Estimation for Pearson diffusions.

Suppose $\{Y_i\}_{i=0,1,\dots,n}$ is a sequence of observations from an ergodic Pearson diffusion made at the time points $t_i = i\Delta$ for $i = 0, \dots, n$, and that we wish to estimate a parameter ψ belonging to the parameter space $\Psi \subset \mathbb{R}^d$. The parameter ψ might be the parameter (θ, μ, a, b, c) of the full class of Pearson diffusions, or it might be a subclass, e.g. a class corresponding to one of the Pearson types. In this section we discuss estimations methods that are simpler for Pearson diffusion that for general diffusions. We do not consider methods for which no simplification is achieved by using the Pearson diffusions.

3.1 Maximum likelihood estimation.

For the Ornstein-Uhlenbeck process it is well-known that the transition is Gaussian with a simple expression for the first and second conditional moments. For the CIR process it is equally well-known that the transition density is a non-central χ^2-distribution which can be expressed in terms of a modified Bessel function. Thus exact likelihood inference is relatively easy for these models. For a diffusion with a discrete spectrum representation of the transition density exists in terms of the eigenfunctions, see Karlin & Taylor (1981),

$$\pi(\Delta, x, y) = m(y) \sum_{j=1}^{\infty} e^{-\Delta\lambda_j} p_j(x) p_j(y) c_j. \tag{3.1}$$

Here $y \mapsto \pi(\Delta, x, y)$ is the transition density, i.e. the conditional density of $X_{t+\Delta}$ given that $X_t = x$, p_j is the jth eigenfunction with eigenvalue λ_j, and $c_j^{-1} = \int_\ell^r p_j(x)^2 m(x) dx$. For the Ornstein-Uhlenbeck process this is just a classical formula for Hermite polynomials (Mehler's formula) that yields the well-known Gaussian transition density. For the CIR process it is a classical expansion of the modified Bessel function. Zhou (2001) pointed out that in this case the series can be interpreted as a Poisson-mixture of gamma-distribution and found that good numerical performance was obtained by using the first 100 terms in the series provided that the diffusion coefficient was not too small. Approximate likelihood inference is similarly feasible by truncating the expansion of the transition density for the Jacobi diffusion.

The remaining Pearson diffusion have only finitely many discrete eigenvalues and the spectrum comprises a continuous part. For the diffusions with a symmetric t-distribution or an inverse Gamma distribution, Wong (1964) gave a spectral expression of the transition density with two parts: a finite sum similar to (3.1) and an integral over the continuous spectrum involving eigenfunctions corresponding to this part of the spectrum. The latter eigenfunctions are quite complicated, so a possible approach to approximate likelihood inference would be to use the sum involving only the polynomial eigenfunctions. It is however a problem that the number of terms, and hence the accuracy of the approximation, depends on the value of the parameter a. For the Pearson diffusion with a classical symmetric t-distribution as stationary distribution Wong (1964) pointed out that the eigenfunctions for both parts of the spectrum simplify considerably for even degrees of freedom. Nagahara (1996) used the simplified version of Wong's expression for the transition density to obtain maximum likelihood estimates with the parameter ν restricted to be even.

For invertible transformations of Pearson diffusions expression for the likelihood function or the approximate likelihood function can be obtained by the transformation theorem.

A number of general techniques for doing likelihood inference for discretely observed diffusions are available, see e.g. Beskos et al. (2006) and the references in that paper. These methods do not simplify for the Pearson diffusion, and will therefore not be considered in the present paper.

3.2 Estimation based on conditional moments.

We have seen that likelihood inference is, at least approximately, feasible for the Pearson diffusions. However, much simpler estimators can be obtained by using that explicit expressions are available for the conditional polynomial moments. When the sampling frequency is not too small, these estimators have, if properly chosen, an efficiency close to that of the maximum likelihood estimator.

If the Pearson diffusion has moments of order N, then the N first eigen-polynomials $p_1(\cdot,\psi),\ldots,p_N(\cdot,\psi)$ are well defined. Thus, we can apply a martingale estimating function of the type introduced by Kessler & Sørensen (1999),

$$G_n(\psi) = \sum_{i=1}^{n}\sum_{j=1}^{N} \alpha_j(Y_{i-1},\psi)\{p_j(Y_i,\psi) - e^{-\lambda_j(\psi)\Delta} p_j(Y_{i-1},\psi)\} \tag{3.2}$$

where $\alpha_1,\ldots,\alpha_N$ are weight functions and $\lambda_1(\psi),\ldots,\lambda_N(\psi)$ are the eigenvalues. Written on matrix form the associated estimating equation take the form

$$G_n(\psi) = \sum_{i=1}^{n} \alpha(Y_{i-1},\psi) h(Y_{i-1},Y_i,\psi) = 0. \tag{3.3}$$

where α is the $d \times N$ weight matrix and $h_j(x, y, \psi) = p_j(y, \psi) - e^{-\lambda_j(\psi)\Delta} p_j(x, \psi)$, $j = 1, \ldots, N$. In order to apply a cental limit theorem to prove asymptotic normality of estimators, we will later assume that the diffusion has finite moment of order $2N$.

We shall focus on the optimal estimating function of the type (3.3). Optimality is in the sense of Godambe & Heyde (1987), which means that the weight matrix α is chosen to minimize the asymptotic variance of the related estimator. An account of the general theory of optimal estimating functions can be found in Heyde (1997). For other choices of weight functions we refer to Bibby, Jacobsen & Sørensen (2004). Other estimators based on the conditional polynomial moments have been proposed in the literature. The generalized method of moments (GMM) by Hansen (1982) is often used in the econometric literature. The way it is usually implemented, see e.g. Campbell, Lo & MacKinlay (1997), yields an estimator with a lower efficiency than the optimal martingale estimating function, unless instruments are chosen in such a way that the estimator is equal to the estimator obtained from the optimal martingale estimating function, see Christensen & Sørensen (2007). Thus optimal GMM is covered by the following discussion. Another popular method is quasi-maximum-likelihood (in the econometric sense), where the transition density is approximated by a Gaussian density with the exact first and second order conditional moments inserted. It is obviously easy to use this method for the Pearson diffusions because the conditional moments are explicitly available. However, the quasi-score function obtained by differentiation of the log-quasi-likelihood is of the form (3.3) with $N = 2$ and is in fact an approximation to the optimal martingale estimating function that is very good when the sampling frequency is high, see Bibby, Jacobsen & Sørensen (2004). Hence also quasi-maximum-likelihood estimation is covered by a discussion of (3.3). Finally the estimators obtained from (3.3) can be expressed as non-linear weighted least squares estimators, a method that is thus also covered by our discussion.

As discussed in the previous subsection, the transition probabilities of an ergodic diffusion have series expansions in terms of the eigenfunctions of the generator. As the expansion (3.1) depends mainly on the first eigenfunctions, the optimally weighted martingale estimating function can be interpreted as an approximation to the score function. In fact the optimal martingale estimating function is the L_2 projection of the score function onto the set of square integrable martingale estimating functions given by the various selections of weights as was proved by Kessler (1996), see also Sørensen (1997). In fact, Sørensen (2007) has proved that optimal martingale estimating functions give estimators that are efficient in a high frequency asymptotics provided that $N \geq 2$. For financial data the speed of mean-reversion is usually so slow that the sampling frequence need not be particularly high for the estimators to have a high efficiency. Calculations of asymptotic variances in Larsen & Sørensen (2007) indicate that for the weakly observations considered in that paper the estimators based on the two first eigenfunctions for the Jacobi diffusion are close to being efficient. Also the simulation study below demonstrates the high efficiency of the estimators obtained from (3.3).

3.3 Optimal martingale estimating function.

A feature of the Pearson diffusions that makes relatively efficient inference easy is that the optimal weights in the sense of Godambe & Heyde (1987) are simple and explicit. Assume that the Pearson diffusion is ergodic and has moments of order $2N$. In particular, $a <$

$(2N-1)^{-1}$. Further assume that the mapping $\psi \mapsto \tau = (\theta, \mu, a, b, c)$ is differentiable. Then the optimal weights for the martingale estimating function (3.2) are given by proposition 3.1 of Kessler & Sørensen (1999) as

$$\alpha^\star(x,\psi) = -S(x,\psi)^T \cdot V(x,\psi)^{-1} \tag{3.4}$$

where ^{T} denotes transposition and

$$\begin{aligned} S_{j,k}(x,\psi) &= -E_\psi\{\partial_{\psi_k} p_j(Y_i,\psi)|Y_{i-1}=x\} + \partial_{\psi_k}\{e^{-\lambda_j(\psi)\Delta} p_j(x,\psi)\} \\ V_{j,k}(x,\psi) &= E_\psi\{p_j(Y_i,\psi)p_k(Y_i,\psi)|Y_{i-1}=x\} - e^{-\{\lambda_j(\psi)+\lambda_k(\psi)\}\Delta} p_j(x,\psi)p_k(x,\psi). \end{aligned}$$

Note that the indicated conditions imply that S and V are well defined. The proof that V is invertible is implicitly given as part of the proof of Theorem 3.1 below. Moreover, the formula defining the optimal weights can be made explicit by means of the recursive formula (2.8) and (2.7) of Section 2. Note that

$$\begin{aligned} V_{j,k}(x,\psi) &= \sum_{j'=0}^{j}\sum_{k'=0}^{k} p_{j,j'}(\psi)p_{k,k'}(\psi)q_{j'+k'}(x,\Delta,\psi) - e^{-(\lambda_j(\psi)+\lambda_k(\psi))\Delta} p_j(x,\psi)p_k(x,\psi) \\ S_{j,k}(x,\psi) &= p_j(x,\psi)e^{-\lambda_j(\psi)\Delta}\partial_{\psi^T}\lambda_j(\psi) + \sum_{j'=0}^{j}\{q_{j'}(x,\Delta,\psi) - e^{-\lambda_j(\psi)\Delta}x^{j'}\}\partial_{\psi^T}p_{j,j'}(\psi). \end{aligned}$$

where $q_j(x,t,\psi) = E_\psi(X_t^j|X_0=x)$ is specified by equations (2.8) and (2.9). Hence, the j,k'th element of $V(x,\psi)$ is a polynomial $v_{j,k}(x) = \sum_{l=0}^{j+k} v_{j,k,l}x^l$ with coefficients given by

$$v_{j,k,l} = \sum_{j'=0}^{j}\sum_{k'=0}^{k} p_{j,j'}p_{k,k'} \cdot (q_{j'+k',l}(\Delta) - e^{-(\lambda_j+\lambda_k)\Delta} I_{\{j'+k'=l\}}),$$

where $I_{\{j'+k'=l\}}$ denotes the indicator function. Similarly, the j,k'th element of $S(x,\psi)$ is the j'th order polynomial $s_{j,k}(x) = \sum_{l=0}^{j} s_{j,k,l}x^l$ the coefficients of which are

$$s_{j,k,l} = e^{-\lambda_j\Delta}(p_{j,l}\partial_{\psi_k}\lambda_j - \partial_{\psi_k}p_{j,l}) + \sum_{j'=0}^{l} \partial_{\psi_k}p_{j,j'} \cdot q_{j',l}(\Delta).$$

It is important to notice that the derivatives $d_{j,l} = \partial_{\psi^T}p_{j,l}$ satisfy the recursion

$$d_{j,l} = \frac{b_{l+1}d_{j,l+1} + c_{l+2}d_{j,l+2} + p_{j,l}\partial_{\psi^T}(a_l - a_j) + p_{j,l+1}\partial_{\psi^T}b_{l+1} + p_{j,l+2}\partial_{\psi^T}c_{l+2}}{a_l - a_j}$$

for $l = j-1, j-2, \ldots, 0$ where initially $d_{j,j} = d_{j,j+1} = 0$.

In practice, it is often a good idea to replace the weight matrix $\alpha^\star(x,\psi)$ by

$$\tilde{\alpha}_n(x) = \alpha^\star(x,\tilde{\psi}_n), \tag{3.5}$$

where $\tilde{\psi}_n$ is a $\sqrt{n}$-consistent estimator of ψ. For instance $\tilde{\psi}_n$ could be an estimator obtained from (3.3) for some simple choice of the weight matrix α independent of ψ. The resulting estimating equations are much easier to solve numerically because of the simpler dependence on θ and because the weight matrix need only be evaluated once for every observation. Moreover, replacing the weights by estimates does not affect the asymptotic distribution of the estimator so there is no loss of efficiency (see Theorem 3.1 below).

3.4 Asymptotic theory.

The optimally weighted martingale estimating function (3.2) provides consistent and asymptotically normal estimators of the parameters of a Pearson diffusion under mild regularity conditions. In what follows ψ_0 denotes the true parameter value.

Theorem 3.1 *Suppose that the following hold true:*

R0: *The Pearson diffusion is ergodic and has moments of order $2N$ where $N \geq 2$.*

R1: *ψ_0 belongs to the interior of Ψ.*

R2: *The mapping $\psi \mapsto \tau = (\theta, \mu, a, b, c)$ is differentiable and $\partial_\psi \tau(\psi_0)$ has full rank d.*

Then with probability tending to one as $n \to \infty$ there exist a solution $\hat{\psi}_n$ to the estimating equation (3.3) with weights specified by either (3.4) or (3.5) such that $\hat{\psi}_n$ converges to ψ_0 in probability and

$$\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, W(\psi_0)^{-1})$$

where $W(\psi_0) = E_{\psi_0}\{S(Y_i, \psi_0)^T V(Y_i, \psi_0)^{-1} S(Y_i, \psi_0)\}$.

The proof of Theorem 3.1 is given in the appendix. Condition **R0** ensures that the eigenfunctions are well defined and that $h_1, \ldots, h_N$ have finite variance so that $G_n(\psi_0)$ is indeed a martingale. In fact, **R0** implies that $G_n(\psi_0)$ is a square integrable martingale.

Example 3.2 For the skewed t-diffusion with parameter $\psi = (\theta, \nu, \rho)$ the canonical parameter is

$$(\theta, \mu, a, b, c) = \left(\theta, 0, \frac{1}{\nu - 1}, \frac{2\rho\nu^{\frac{1}{2}}}{\nu - 1}, \frac{(1 + \rho^2)\nu}{\nu - 1}\right)$$

and

$$\frac{\partial \tau}{\partial \psi^T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{(\nu-1)^2} & \frac{\rho}{\nu^{\frac{1}{2}}(\nu-1)} - \frac{2\rho\nu^{\frac{1}{2}}}{(\nu-1)^2} & \frac{1+\rho^2}{(\nu-1)} - \frac{\nu(1+\rho^2)}{(\nu-1)^2} \\ 0 & 0 & 0 & \frac{2\nu^{\frac{1}{2}}}{\nu-1} & \frac{2\nu\rho}{\nu-1} \end{pmatrix}$$

which has full rank three. Hence, consistent and asymptotically normal estimators are obtained by means of the optimally weighted martingale estimating function under the further assumption that $\nu_0 > 2N$. △

3.5 A simulation study.

In this section we present a small simulation study in order to compare the variance of the optimally weighted martingale estimator based on the two first eigenfunctions to that of the maximum likelihood estimator. We will refer to the former estimator as the optimal martingale estimator. Save for the case of the Ornstein-Uhlenbeck process, calculation of the maximum likelihood estimator is complicated and no expression is available for the Fisher information. Therefore we exploit that Gobet (2002) has shown that discretely observed diffusion processes are locally asymptotically normal in the high frequency limit where $\Delta \to 0$ and $n\Delta \to \infty$ and has given an expression for the Fisher information. Instead of finding the finite sample variance of the maximum likelihood estimator by a time consuming simulation study, we compare the finite sample variance of the optimal martingale estimator to the inverse high frequency Fisher information for sampling frequencies that are not too low.

3.5.1 The Ornstein-Uhlenbeck process.

For the Ornstein-Uhlenbeck process the score function is a quadratic martingale estimating function, so the optimal martingale estimating function based on the two first eigenfunctions is necessarily equal to the score function and hence yields the maximum likelihood estimator. Therefore in this case the estimator is efficient and no simulation study seems necessary. However, we intend to compare finite sample variances to a high frequency limit result, which may underestimate the efficiency of the optimal martingale estimator. To get a feeling for what deviations to expect, we make a similar comparison for the maximum likelihood estimator for the Ornstein-Uhlenbeck process. The results are reported in Table 3.1 where also bias is considered. We see that the finite sample size variances are larger than the limit values by from a few up to 10 per cent depending on the sampling frequency.

		Mean		Variance		
Δ	Parameter	True	MLE	MLE	Asymptotic	Ratio
1	θ	0.1	0.1002	$2.192 \cdot 10^{-5}$	$2 \cdot 10^{-5}$	1.096
1	ρ^2	0.2	0.2000	$8.841 \cdot 10^{-6}$	$8 \cdot 10^{-6}$	1.105
0.5	θ	0.1	0.1005	$4.273 \cdot 10^{-5}$	$4 \cdot 10^{-5}$	1.068
0.5	ρ^2	0.2	0.2001	$8.386 \cdot 10^{-6}$	$8 \cdot 10^{-6}$	1.048
0.2	θ	0.1	0.1010	$1.032 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	1.032
0.2	ρ^2	0.2	0.2001	$8.239 \cdot 10^{-6}$	$8 \cdot 10^{-6}$	1.030
0.1	θ	0.1	0.1020	$2.109 \cdot 10^{-4}$	$2 \cdot 10^{-4}$	1.055
0.1	ρ^2	0.2	0.2000	$8.087 \cdot 10^{-6}$	$8 \cdot 10^{-6}$	1.011

Table 3.1: Empirical means and variances of the maximum likelihood estimator for the Ornstein-Uhlenbeck process $dX_t = -\theta X_t dt + \rho dB_t$ based on 10,000 independent samples of $\{X_{i\Delta}\}_{i=1,\ldots,n}$ with sample size $n = 10,000$ and varying sampling frequencies.

3.5.2 The t-diffusion.

As an example we consider the t-diffusion given by

$$dX_t = -\theta X_t dt + \sqrt{\alpha X_t^2 + \beta}\, dB_t$$

where $\theta, \alpha, \beta > 0$. The invariant distribution is the scaled t-distribution with $\nu = 2\theta/\alpha + 1$ degrees of freedom and scale parameter $\gamma = \sqrt{\beta/(2\theta + \alpha)}$. The diffusion has moment of order k if $k < 2\theta/a + 1$. The variance is given by $\gamma^2 \nu/(\nu - 2)$ whenever $\alpha < 2\theta$. The high-frequency Fisher information is given by

$$\begin{pmatrix} \int_0^\infty \frac{(x^2\pi(x)}{\alpha x^2+\beta}dx & 0 & 0 \\ 0 & \frac{1}{2}\int_0^\infty \frac{x^4\pi(x)}{(\alpha x^2+\beta)^2}dx & \frac{1}{2}\int_0^\infty \frac{x^2\pi(x)}{(\alpha x^2+\beta)^2}dx \\ 0 & \frac{1}{2}\int_0^\infty \frac{x^2\pi(x)}{(\alpha x^2+\beta)^2}dx & \frac{1}{2}\int_0^\infty \frac{\pi(x)}{(\alpha x^2+\beta)^2}dx \end{pmatrix}$$

where π denotes the invariant density $\pi(x) = B(\nu/2, 1/2)^{-1}(\nu\gamma^2)^{\nu/2}(x^2 + \nu\gamma^2)^{-(\nu+1)/2}$. The integrals can readily be found explicitly.

For our simulation study we have chosen the parameter values $\theta = 0.1$, $\alpha = 0.025$, and $\beta = 0.175$. This ensures that the diffusion is well defined and ergodic. Further the implied variance is equal to one so that the diffusion is comparable to the Ornstein-Uhlenbeck process considered previously. The invariant distribution is $\sqrt{9/7} \cdot T(9)$ and has finite moment of order eight but not of order nine. Results for $\Delta = 1$ are reported in Table 3.2. As expected the optimal martingale estimator seems to be less efficient than the maximum likelihood estimator, but not by much.

	Mean		Variance		
Parameter	True	OME	OME	Asymptotic	Ratio
θ	0.1	0.1009	$2.818 \cdot 10^{-5}$	$2.250 \cdot 10^{-5}$	1.252
α	0.025	0.0253	$8.340 \cdot 10^{-6}$	$7.403 \cdot 10^{-6}$	1.127
β	0.175	0.1768	$1.254 \cdot 10^{-5}$	$1.099 \cdot 10^{-5}$	1.140

Table 3.2: Empirical means and variances of the optimal martingale estimator (OME) for the t-diffusion $dX_t = -\theta X_t dt + \sqrt{\alpha X_t^2 + \beta} dB_t$ based on 5,000 independent samples of $\{X_{i\Delta}\}_{i=1,\ldots,n}$ with sample size $n = 10,000$ and $\Delta = 1$.

4 Derived diffusion-type models.

The Pearson diffusion processes can be used as building blocks to obtain more general diffusion-type models. In what follows we consider inference for integrated diffusions, sums of diffusions, and stochastic volatility models. These derived processes are not Markovian. Therefore explicit martingale estimating functions are no longer available. In stead we suggest to base the statistical inference on prediction based estimation functions, introduced in Sørensen (2000). We will demonstrate that such estimating functions can be found explicitly for models based on Pearson diffusions. We start by briefly reviewing the method of prediction based estimating functions.

4.1 Prediction based estimating functions.

Here we focus on estimating functions based on prediction of powers of the observations of the process. Suppose that we have observed the random variables $Y_1, \ldots, Y_n$ that form a stationary stochastic process the distribution of which is parametrised by $\Psi \subseteq \mathbb{R}^d$. Assume that $E_\psi(Y_i^{2m}) < \infty$ for all $\psi \in \Psi$ for some $m \in \mathbb{N}$. For each $i = r + 1, \ldots, n$ and $j = 1, \ldots, m$ let the class $\{Z_{jk}^{(i-1)} \mid k = 1, \ldots, q_j\}$ be a subset of the random variables

$\{Y_{i-\ell}^{\kappa} \,|\, \ell = 1, \ldots, r, \kappa = 0, \ldots, j\}$, where $Z_{j1}^{(i-1)}$ is always equal to 1. We wish to predict Y_i^j by means of linear combinations of the $Z_{jk}^{(i-1)}$-s for each of the values of i and j listed above and then to use suitable linear combinations of the prediction errors to estimate the parameter ψ. Let $\mathcal{P}_{i-1,j}$ denote the space of predictors of Y_i^j, i.e. the space of square integrable random variables spanned by $Z_{j1}^{(i-1)}, \ldots, Z_{jq_j}^{(i-1)}$. The elements of $\mathcal{P}_{i-1,j}$ are of the form $a^T Z_j^{(i-1)}$, where $a^T = (a_1, \ldots, a_{q_j})$ and $Z_j^{(i-1)} = (Z_{j1}^{(i-1)}, \ldots, Z_{jq_j}^{(i-1)})^T$ are q_j-dimensional vectors.

We will use estimating functions of the type

$$G_n(\psi) = \sum_{i=r+1}^{n} \sum_{j=1}^{m} \Pi_j^{(i-1)}(\psi) \left[Y_i^j - \breve{\pi}_j^{(i-1)}(\psi) \right] \tag{4.1}$$

where $\Pi_j^{(i-1)}(\psi)$ is a d-dimensional data dependent vector of weights, the coordinates of which belong to $\mathcal{P}_{i-1,j}$, and where $\breve{\pi}_j^{(i-1)}(\psi)$ is the minimum mean square error predictor of Y_i^j in $\mathcal{P}_{i-1,j}$, which is the usual L_2-projection of Y_i^j onto $\mathcal{P}_{i-1,j}$. When ψ is the true parameter value, we define $C_j(\psi)$ as the covariance matrix of $(Z_{j2}^{(r)}, \ldots, Z_{jq_j}^{(r)})^T$ and $b_j(\psi) = (\mathrm{Cov}_\psi(Z_{j2}^{(r)}, Y_{r+1}^j), \ldots, \mathrm{Cov}_\psi(Z_{jq_j}^{(r)}, Y_{r+1}^j))^T$. Then we have

$$\breve{\pi}_j^{(i-1)}(\psi) = \hat{a}_j(\psi)^T Z_j^{(i-1)}$$

where $\hat{a}_j(\psi)^T = (\hat{a}_{j1}(\psi), \hat{a}_{j*}(\psi)^T)$ with $\hat{a}_{j*}(\psi)^T = (\hat{a}_{j2}(\psi), \ldots, \hat{a}_{jq_j}(\psi))$ defined by

$$\hat{a}_{j*}(\psi) = C_j(\psi)^{-1} b_j(\psi) \tag{4.2}$$

and

$$\hat{a}_{j1}(\psi) = E_\psi(Y_1^j) - \sum_{k=2}^{q_j} \hat{a}_{jk}(\psi) E_\psi(Z_{jk}^{(r)}). \tag{4.3}$$

Thus to find $\breve{\pi}_j^{(i-1)}(\psi)$, $j = 1, \ldots, m$, we need to calculate moments of the form

$$E_\psi(Y_1^\kappa Y_k^j), \quad 0 \leq \kappa \leq j \leq m, \quad k = 1, \ldots, r. \tag{4.4}$$

Once we have calculated these moments, the vector of coefficients $\hat{a}_j$ can easily be found by means of the m-dimensional Durbin-Levinson algorithm applied to $\{(Y_i, Y_i^2, \ldots, Y_i^m)\}_{i \in \mathbb{N}}$, see Brockwell & Davis (1991). The non-Markovian diffusion-type models considered in this paper inherit the exponential ρ-mixing property from the Pearson diffusions. Therefore constants $K > 0$ and $\lambda > 0$ exist such that $\left|\mathrm{Cov}_\psi(Y_1^j, Y_k^j)\right| \leq K e^{-\lambda k}$ (λ is typically the smallest speed of mean reversion of the involved Pearson diffusions). Therefore r will usually not need to be chosen particularly large. If Y_i^j is restricted to have mean zero, we need not include a constant in the space of predictors, i.e. we need only the space spanned by $Z_{j2}^{(i-1)}, \ldots, Z_{jq_j}^{(i-1)}$.

In many situations $m = 2$ with $Z_{jk}^{(i-1)} = Y_{i-k}$, $k = 1, \ldots, r, j = 1, 2$ and $Z_{2k}^{(i-1)} = Y_{i+r-k}^2$, $k = r+1, \ldots, 2r$, will be a reasonable choice. In this case the minimum mean square error predictor of Y_i can be found using the Durbin-Levinson algorithm for real processes, while the predictor of Y_i^2 can be found by applying the two-dimensional Durbin-Levinson

algorithm to the process (Y_i, Y_i^2).
Including predictors in the form of lagged terms $Y_{i-k}Y_{i-k-l}$ for a number of lags l's might also be of relevance. These terms enter into the least squares estimator of Forman (2005), which produces good estimates for a sum of Ornstein-Uhlenbeck processes.

The choice of the weights $\Pi_j^{(i-1)}(\psi)$ in (4.1) for which the asymptotic variance of the estimators is minimized is the Godambe optimal prediction-based estimating function, that was derived in Sørensen (2000). An account of the theory of optimal estimating functions can be found in Heyde (1997). The optimal estimating function of the type (4.1) can be written in the form

$$G_n^*(\psi) = A_n^*(\psi) \sum_{i=r+1}^{n} H^{(i)}(\psi), \tag{4.5}$$

where

$$H^{(i)}(\psi) = Z^{(i-1)} \left(F(Y_i) - \hat{\pi}^{(i-1)}(\psi) \right), \tag{4.6}$$

with $F(x) = (x, x^2, \ldots, x^m)^T$, $\hat{\pi}^{(i-1)}(\psi) = (\hat{\pi}_1^{(i-1)}(\psi), \ldots \hat{\pi}_m^{(i-1)}(\psi))^T$ and

$$Z^{(i-1)} = \begin{pmatrix} Z_1^{(i-1)} & 0_{q_1} & \cdots & 0_{q_1} \\ & & & \\ 0_{q_2} & Z_2^{(i-1)} & \cdots & 0_{q_2} \\ \vdots & \vdots & & \vdots \\ 0_{q_m} & 0_{q_m} & \cdots & Z_m^{(i-1)} \end{pmatrix}. \tag{4.7}$$

Here 0_{q_j} denotes the q_j-dimensional zero-vector. Finally,

$$A_n^*(\psi) = \partial_\psi \hat{a}(\psi)^T \bar{C}(\psi) \bar{M}_n(\psi)^{-1}, \tag{4.8}$$

with

$$\bar{M}_n(\psi) = E_\psi\left(H^{(r+1)}(\psi)H^{(r+1)}(\psi)^T\right) + \tag{4.9}$$
$$\sum_{k=1}^{n-r-1} \frac{(n-r-k)}{(n-r)} \left[E_\psi\left(H^{(r+1)}(\psi)H^{(r+1+k)}(\psi)^T\right) + E_\psi\left(H^{(r+1+k)}(\psi)H^{(r+1)}(\psi)^T\right)\right],$$

$$\bar{C}(\psi) = E_\psi \left(Z^{(i-1)} (Z^{(i-1)})^T \right), \tag{4.10}$$

and

$$\hat{a}(\psi)^T = \left(\hat{a}_1(\psi)^T, \ldots, \hat{a}_m(\psi)^T \right), \tag{4.11}$$

where $\hat{a}_j(\psi)$ is given by (4.2) and (4.3). A necessary condition that the moments in (4.9) exist is that $E_\psi(Y_i^{4m}) < \infty$ for all $\psi \in \Psi$. For (4.5) to be optimal we need that the matrix $\partial_\psi \hat{a}(\psi)^T$ has full rank. The matrix $\bar{M}_n(\psi)$ is always invertible.

Because the processes considered below inherit the exponential ρ-mixing property from the Pearson diffusions, there exist constants $K > 0$ and $\lambda > 0$ such that the absolute values of all entries in the expectation matrices in the sum in (4.9) are dominated by $K\mathrm{e}^{-\lambda(k-r-1)}$ when $k > r$. Therefore, the sum in (4.9) can in practice often be truncated so that fewer moments need to be calculated. The matrix $\bar{M}_n(\psi)$ can also be approximated by a truncated version of the limiting matrix

$$\begin{aligned}\bar{M}(\psi) \;=\; & E_\psi\Big(H^{(r+1)}(\psi)H^{(r+1)}(\psi)^T\Big) + \\ & \sum_{k=1}^{\infty}\Big[E_\psi\Big(H^{(r+1)}(\psi)H^{(r+1+k)}(\psi)^T\Big) \;+\; E_\psi\Big(H^{(r+1+k)}(\psi)H^{(r+1)}(\psi)^T\Big)\Big],\end{aligned} \tag{4.12}$$

obtained for $n \to \infty$. In practice, it is usually also a good idea to replace $A_n^*(\psi)$ by $A_n^*(\bar{\psi}_n)$, where $\bar{\psi}_n$ is a $\sqrt{n}$-consistent estimator of ψ (and similarly for approximations to $A_n^*(\psi)$). This has the advantages that (4.9) or (4.12) need only be calculated once and that a simpler estimating equation is obtained, while the asymptotic variance of the estimator is unchanged. The estimator $\bar{\psi}_n$ can, for instance, be obtained from an estimating function similar to (4.5), where $A_n^*(\psi)$ has been replaced by a suitable simple matrix independent of ψ, but such that the estimating equation has a solution. Usually it is enough to use the first d coordinates of $H^{(i)}(\psi)$, where d is the dimension of the parameter. In order to calculate (4.9) or (4.12), we need mixed moments of the form

$$E_\psi[Y_1^{k_1}Y_{t_1}^{k_2}Y_{t_2}^{k_3}Y_{t_3}^{k_4}], \quad 1 \le t_1 \le t_2 \le t_3 \quad k_1+k_2+k_3+k_4 \le 4m \tag{4.13}$$

where k_i, $i = 1, \ldots, 4$ are non-negative integers. In the following subsections we demonstrate that in three diffusion-type models derived from Pearson diffusions, explicit expressions can be found for the necessary moments, (4.4) and (4.13). Thus the optimal prediction-based estimating functions are explicit.

4.2 Integrated Pearson diffusions.

Let X be a stationary Pearson diffusion, i.e. a solution to (2.1). Suppose that the diffusion has not been observed directly, but that the data are

$$Y_i = \frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta} X_s\, ds, \qquad i = 1, \ldots, n \tag{4.14}$$

for some fixed Δ. Such observations can be obtained if the process X is observed after passage through an electronic filter. Another example is provided by ice-core records. The isotope ratio $^{18}O/^{16}O$ in the ice, measured as an average in pieces of ice, each piece representing a time interval with time increasing as a function of the depth, is a proxy for paleo-temperatures. The variation of the paleo-temperature can be modelled by a stochastic differential equation, and it is natural to model the ice-core data as an integrated diffusion process, see Ditlevsen, Ditlevsen & Andersen (2002). A third example is the realized volatility of financial econometrics, see e.g. Bollerslev & Zhou (2002). Estimation based on data of the type (4.14) was considered by Gloter (2000), Bollerslev & Zhou (2002), Ditlevsen & Sørensen (2004), and Gloter (2006). Since X is stationary, the random variables Y_i, $i = 1, \ldots, n$ form a stationary process with the same mixing properties as X, i.e. it is exponentially mixing. However, the observed process is not Markovian, so martingale estimating functions are not available in a tractable form, but explicit prediction-based estimating functions can be found.

Suppose that $4m$'th moment of X_t is finite. The moments (4.4) and (4.13) can be calculated by

$$E\left[Y_1^{k_1}Y_{t_1}^{k_2}Y_{t_2}^{k_3}Y_{t_3}^{k_4}\right] = \frac{\int_A E[X_{v_1}\cdots X_{v_{k_1}}X_{u_1}\cdots X_{u_{k_2}}X_{s_1}\cdots X_{s_{k_3}}X_{r_1}\cdots X_{r_{k_4}}]\,d\mathbf{t}}{\Delta^{k_1+k_2+k_3+k_4}}$$

where $1 \le t_1 \le t_2 \le t_3$, $A = [0\,,\,\Delta]^{k_1} \times [(t_1-1)\Delta\,,\,t_1\Delta]^{k_2} \times [(t_2-1)\Delta\,,\,t_2\Delta]^{k_3} \times [(t_3-1)\Delta\,,\,t_3\Delta]^{k_4}$, and $d\mathbf{t} = dr_{k_4}\cdots dr_1\,ds_{k_3}\cdots ds_1\,du_{k_2}\cdots du_1\,dv_{k_1}\cdots dv_1$. The domain of integration can be reduced considerably by symmetry arguments, but here the point is that we need to calculate moments of the type $E(X_{t_1}^{\kappa_1}\cdots X_{t_k}^{\kappa_k})$, where $t_1 < \cdots < t_k$. Since $E(X_t^n \,|\, X_0 = x)$ is a polynomial in x given by (2.9), it follows that we can find the needed moments iteratively

$$E(X_{t_1}^{\kappa_1}\cdots X_{t_k}^{\kappa_k}) = \sum_{j=1}^{\kappa_k} q_{\kappa_k,j}(t_k - t_{k-1})E(X_{t_1}^{\kappa_1}\cdots X_{t_{k-1}}^{\kappa_{k-1}+j}),$$

where $q_{\kappa_k,j}(t_k - t_{k-1})$ is given by (2.10). The coefficient depends on time through an exponential function, so $E(X_{t_1}^{\kappa_1}\cdots X_{t_k}^{\kappa_k})$ depends on $t_1,\ldots,t_k$ through sums and products of exponential functions. Therefore the integral above can be explicitly calculated.

Example 4.1 *Integrated skew t-diffusion.* Let us see how this works by calculating an optimal estimating function for the integrated skew t-diffusion (2.3). To simplify the exposition we consider the simple case where $m = 2$, $Z_{1,1}^{(i-1)} = Y_{i-1}$, $Z_{2,1}^{(i-1)} = 1$, and $Z_{2,2}^{(i-1)} = Y_{i-1}^2$ (i.e. $q_1 = r = 1, q_2 = 2$). The estimating equations take the form

$$G_n(\theta,\rho,\nu) = \sum_{i=2}^{n}\begin{bmatrix} Y_{i-1}Y_i - \beta_1 Y_{i-1}^2 \\ Y_i^2 - \sigma^2(1-\beta_2) - \beta_2 Y_{i-1}^2 \\ Y_{i-1}^2Y_i^2 - \sigma^2(1-\beta_2)Y_{i-1}^2 - \beta_2 Y_{i-1}^4 \end{bmatrix} = 0, \qquad (4.15)$$

with $\sigma^2 = \mathrm{Var}(Y_{i-1})$ and $\beta_j = \mathrm{Cov}(Y_{i-1}^j, Y_i^j)\cdot\mathrm{Var}(Y_{i-1}^j)^{-1}$ for $j = 1, 2$. In particular,

$$\sigma^2 = \frac{2\nu(1+\rho^2)}{\nu-2}\cdot\left\{\frac{1}{\theta\Delta} - \frac{1-e^{-\theta\Delta}}{(\theta\Delta)^2}\right\}, \quad \beta_1 = \frac{(1-e^{-\theta\Delta})^2}{2(\theta\Delta - 1 + e^{-\theta\Delta})}.$$

In order to get an explicit expression of β_2 let $f_n(x) = x^{-n}(1-e^{-x})$, then

$$\mathrm{Cov}(Y_{i-1}^2, Y_i^2) = 4\gamma_1(\lambda\Delta - \theta\Delta)^{-2}\{f_1(\lambda\Delta) - f_1(\theta\Delta)\}^2 + 4\gamma_2(\theta\Delta)^{-2}\{1-(1+\theta\Delta)f_1(\theta\Delta)\}^2$$

where $\lambda = \frac{2\theta(\nu-2)}{\nu-1}$, $\gamma_1 = \frac{(3\nu^3-10\nu^2-4\nu)\rho^2\sigma^2}{(\nu-4)(\nu-3)^2} + \frac{3\nu\sigma^2}{\nu-4} - \sigma^4$, and $\gamma_2 = \frac{16\nu\rho^2\sigma^2}{(\nu-3)^2}$. Likewise,

$$\begin{aligned}\mathrm{Var}(Y_{i-1}^2) &= 24\gamma_1[(\lambda\Delta-\theta\Delta)^{-2}\{f_2(\theta\Delta) - f_2(\lambda\Delta)\} + (\theta\Delta)^{-2}\{(\lambda\Delta)^{-1} + (\lambda\Delta-\theta\Delta)^{-1}\}] \\ &- 24\gamma_1(\theta\Delta)^{-1}(\theta\Delta-\lambda\Delta)^{-1}(2+\theta\Delta)f_2(\theta\Delta) \\ &+ 12\gamma_2(\theta\Delta)^{-2}[1 + 6(\theta\Delta)^{-1} - \{(\theta\Delta)^2 + 4\theta\Delta + 6\}f_2(\theta\Delta)] \\ &+ 12\sigma^4(\theta\Delta)^{-2}\{1 - 6(\theta\Delta)^{-1} + (2\theta\Delta+6)f_2(\theta\Delta)\} - 4\sigma^4(\theta\Delta)^{-2}\{1-f_1(\theta\Delta)\}^2.\end{aligned}$$

Solving equation (4.15) for β_1, β_2, and σ^2 we get

$$\hat\beta_1 = \frac{\frac{1}{n-1}\sum_{i=2}^n Y_{i-1}Y_i}{\frac{1}{n-1}\sum_{i=2}^n Y_{i-1}^2}, \quad \hat\beta_2 = \frac{\frac{1}{n-1}\sum_{i=2}^n Y_{i-1}^2Y_i^2 - (\frac{1}{n-1}\sum_{i=2}^n Y_{i-1}^2)(\frac{1}{n-1}\sum_{i=2}^n Y_i^2)}{\frac{1}{n-1}\sum_{i=2}^n Y_{i-1}^4 - (\frac{1}{n-1}\sum_{i=2}^n Y_{i-1}^2)^2}$$

and

$$\hat{\sigma}^2 = \tfrac{1}{1-\hat{\beta}_2}\tfrac{1}{n-1}\sum_{i=2}^n Y_{i-1}^2 + \tfrac{\hat{\beta}_2}{1-\hat{\beta}_2}\tfrac{1}{n-1}\sum_{i=2}^n Y_i^2.$$

Hence, if $0 < \hat{\beta}_1 < 1$, which happens eventually with probability one, the expression of β_1 yields a unique estimate $\hat{\theta} > 0$ satisfying

$$2\hat{\beta}_1\{\hat{\theta}\Delta - (1 - e^{-\hat{\theta}\Delta})\} - (1 - e^{-\hat{\theta}\Delta})^2 = 0.$$

The remaining equations are solved by substituting $\hat{\rho}(\nu)^2 = \frac{\nu-2}{2\nu}\{(\hat{\theta}\Delta)^{-1} - f_2(\hat{\theta}\Delta)\}^{-1} - 1$ into the equation $\beta_2(\hat{\theta}, \hat{\rho}(\nu)^2, \nu) = \hat{\beta}_2$, which has to be solved numerically. To estimate the sign of ρ, note that for instance $E(Y_1^3) = \frac{24\sqrt{\nu}\rho\sigma^2}{\nu-3} \cdot (\theta\Delta)^{-2}\{2 - (2 + \theta\Delta)f_1(\theta\Delta)\}$ has the same sign as ρ. △

4.3 Sums of diffusions.

The simple exponentially decreasing autocorrelation function of the Pearson diffusions is too simple in some applications, but we can obtain a much richer autocorrelation structure by considering sums of Pearson diffusions:

$$Y_t = X_{1,t} + \ldots + X_{M,t} \tag{4.16}$$

$$dX_{i,t} = -\theta_i(X_{i,t} - \mu_i) + \sigma_i(X_{i,t})dB_{i,t}, \quad i = 1, \ldots, M, \tag{4.17}$$

where $\theta_1, \ldots, \theta_M > 0$ and $B_1, \ldots, B_M$ are independent Brownian motions. The diffusion coefficients $\sigma_1, \ldots, \sigma_M$ are of the form of a Pearson diffusion (2.1). Suppose all $X_{i,t}$ have finite second moment. Then the autocorrelation function of Y is

$$\rho(t) = \phi_1 \exp(-\theta_1 t) + \ldots + \phi_M \exp(-\theta_M t) \tag{4.18}$$

with

$$\phi_i = \frac{\mathrm{Var}(X_{i,t})}{\mathrm{Var}(X_{1,t}) + \cdots + \mathrm{Var}(X_{M,t})}.$$

Thus $\phi_1 + \ldots + \phi_M = 1$. The expectation of Y_t is $\mu_1 + \cdots + \mu_M$. Sums of diffusions with a pre-specified marginal distribution of Y were considered by Bibby & Sørensen (2003), Bibby, Skovgaard & Sørensen (2005) and Forman (2005). Here we specify instead the distributions of the $X_{i,t}$'s, which implies that the models are simpler to handle. Sums of Ornstein-Uhlenbeck processes driven by Lévy processes were introduced and studied in Barndorff-Nielsen, Jensen & Sørensen (1998). An autocorrelation function of the form (4.18) fits turbulence data well, see Barndorff-Nielsen, Jensen & Sørensen (1990) and Bibby, Skovgaard & Sørensen (2005).

Example 4.2 *Sum of Ornstein-Uhlenbeck processes.* If $\sigma_i^2(x) = 2\theta_i c_i$, the stationary distribution of Y_t is a normal distribution with mean $\mu_1 + \cdots + \mu_M$ and variance $c_1^2 + \cdots + c_M^2$. △

Example 4.3 *Sum of CIR processes.* If $\sigma_i^2(x) = 2\theta_i bx$ and $\mu_i = \alpha_i b$, then the stationary distribution of Y_t is a Gamma-distribution with shape parameter $\alpha_1 + \cdots + \alpha_M$ and scale parameter b. The weights in the autocorrelation function are $\phi_i = \alpha_i/(\alpha_1 + \cdots + \alpha_M)$. △

In the other cases of Pearson diffusions, the class of marginal distributions is not closed under convolution, so the stationary distribution of Y_t is not in the Pearson class and is, in fact, not any of the standard distributions. It has recently been proven that the sum of two t-distributions with odd degrees of freedom is a finite mixture (over degrees of freedom) of scaled t-distributions, see Berg & Vignat (2006). In the case of the Jacobi-diffusions it might be preferable to consider Y_t/M to obtain again a process with state space $(0, 1)$.

A sum of diffusions is not a Markov process, so also for this type of model we use prediction-based estimating functions rather than martingale estimating functions. Suppose that the process Y has been observed at the time points $t_i = \Delta i$, $i = 1, \ldots, n$. The necessary moments of the form (4.4) and (4.13) can, provided they exist, be obtained from the mixed moments of the Pearson diffusions because by the multinomial formula we find, for instance

$$E(Y_{t_1}^{\kappa} Y_{t_2}^{\nu}) = \sum\sum \binom{\kappa}{\kappa_1, \ldots, \kappa_M} \binom{\nu}{\nu_1, \ldots, \nu_M} E(X_{1,t_1}^{\kappa_1} X_{1,t_2}^{\nu_1}) \ldots E(X_{M,t_1}^{\kappa_M} X_{M,t_2}^{\nu_M})$$

where

$$\binom{\kappa}{\kappa_1, \ldots, \kappa_M} = \frac{\kappa!}{\kappa_1! \cdots \kappa_M!}$$

is the multinomial coefficient, and where the first summation is over $0 \leq \kappa_1, \ldots, \kappa_M$ such that $\kappa_1 + \ldots \kappa_M = \kappa$ and the second summation is the same just for the ν's. The higher order mixed moments of the form (4.13) can be found by a similar formula with four sums and four multinomial coefficients. Such formulae may appear daunting, but are easy to programme. Mixed moments of the form $E(X_{t_1}^{\kappa_1} \cdots X_{t_k}^{\kappa_k})$ can be calculated iteratively as explained in Subsection 4.2.

Example 4.4 *Sum of two skew t-diffusions.* If, for i=1,2, $\sigma_i^2(x) = 2\theta_i(\nu_i - 1)^{-1}\{x^2 + 2\rho\sqrt{\nu_i}x + (1+\rho^2)\nu\}$, the stationary distribution of $X_{i,t}$ is a skew t-diffusion. The distribution of Y_t is a convolution of skew t-diffusions,

$$\mathrm{Var}(Y) = (1+\rho^2)\left(\frac{\nu_1}{\nu_1 - 2} + \frac{\nu_2}{\nu_2 - 2}\right),$$

and $\phi_i = \nu_i(\nu_i - 2)^{-1}/\{\nu_1(\nu_1-2)^{-1} + \nu_2(\nu_2-2)^{-1}\}$. To simplify the exposition we assume that the correlation parameters θ_1, θ_2, ϕ_1, and ϕ_2 are known or have been estimated in advance (the least squares estimator of Forman (2005) applies and so does the predictions based estimating function with $m = 1$, $Z_{1,k}^{(i-1)} = Y_{i-k}$, $k = 1, \ldots, r$). We will find the optimal estimating function in the simple case where predictions of Y_i^2 are made based on $Z_{1,1}^{(i-1)} = 1$ and $Z_{1,2}^{(i-1)} = Y_{i-1}$. The estimating equations take the form

$$G_n(\theta, \rho, \nu) = \sum_{i=2}^{n} \begin{bmatrix} Y_i^2 - \sigma^2 - \beta_{21}Y_{i-1} \\ Y_{i-1}Y_i^2 - \sigma^2 Y_{i-1} - \beta_{21}Y_{i-1}^2 \end{bmatrix} = 0, \tag{4.19}$$

with $\sigma^2 = \mathrm{Var}(Y_{i-1})$ and $\beta_{21} = \mathrm{Cov}(Y_{i-1}, Y_i^2) \cdot \mathrm{Var}(Y_{i-1})^{-1}$. To be specific

$$\sigma^2 = (1+\rho^2)\left\{\frac{\nu_1}{\nu_1 - 2} + \frac{\nu_2}{\nu_2 - 2}\right\}, \quad \beta_{21} = 4\rho\left\{\frac{\sqrt{\nu_1}}{\nu_1 - 3}\phi_1 e^{-\theta_1\Delta} + \frac{\sqrt{\nu_2}}{\nu_2 - 3}\phi_2 e^{-\theta_2\Delta}\right\}.$$

Solving equation (4.19) for β_{21} and σ^2 we get

$$\begin{aligned}\hat{\beta}_{21} &= \frac{\frac{1}{n-1}\sum_{i=2}^n Y_{i-1}Y_i^2 - (\frac{1}{n-1}\sum_{i=2}^n Y_{i-1})(\frac{1}{n-1}\sum_{i=2}^n Y_i^2)}{\frac{1}{n-1}\sum_{i=2}^n Y_{i-1}^2 - (\frac{1}{n-1}\sum_{i=2}^n Y_{i-1})^2}, \\ \hat{\sigma}^2 &= \tfrac{1}{n-1}\sum_{i=2}^n Y_i^2 + \hat{\beta}_{21}\tfrac{1}{n-1}\sum_{i=2}^n Y_{i-1}.\end{aligned}$$

In order to estimate ρ we restate β_{21} as

$$\beta_{21} = \sqrt{32(1+\rho^2)} \cdot \rho \cdot \left\{ \frac{\sqrt{9(1+\rho^2) - \phi_1\sigma^2}}{3(1+\rho^2) - \phi_1\sigma^2}\phi_1 e^{-\theta_1\Delta} + \frac{\sqrt{9(1+\rho^2) - \phi_2\sigma^2}}{3(1+\rho^2) - \phi_2\sigma^2}\phi_2 e^{-\theta_2\Delta} \right\}$$

and insert $\hat{\sigma}^2$ for σ^2. Thus, we get a one-dimensional estimating equation, $\beta_{21}(\theta, \phi, \hat{\sigma}^2, \rho) = \hat{\beta}_{21}$, which can be solved numerically. Finally by inverting $\phi_i = \frac{1+\rho^2}{\sigma^2}\frac{\nu_i}{\nu_i - 2}$ we find the estimates $\hat{\nu}_i = \frac{2\phi_i\hat{\sigma}^2}{\phi_i\hat{\sigma}^2 - (1+\hat{\rho}^2)}$, $i = 1, 2$. △

A more complex model is obtained if the observations are integrals of Y in analogy with the previous subsection:

$$Z_i = \frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta} Y_s\, ds = \frac{1}{\Delta}\left(\int_{(i-1)\Delta}^{i\Delta} X_{1,t}ds + \cdots + \int_{(i-1)\Delta}^{i\Delta} X_{M,t}ds\right), \tag{4.20}$$

$i = 1, \ldots, n$. Also here the moments of form (4.4) and (4.13) can be found explicitly because each of the observations Z_i is a sum of processes of the type considered in the previous subsection. To calculate $E(Z_1^{k_1} Z_{t_1}^{k_2} Z_{t_2}^{k_3} Z_{t_3}^{k_4})$, first apply the multinomial formula to express this quantity in terms of moments of the form $E(Y_{j,1}^{\ell_1} Y_{j,t_1}^{\ell_2} Y_{j,t_2}^{\ell_3} Y_{j,t_3}^{\ell_4})$, where

$$Y_{j,i} = \frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta} X_{j,s}\, ds.$$

Now proceed as in Subsection 4.2.

4.4 Stochastic volatility models.

A stochastic volatility model is a generalization of the Black-Scholes model for the logarithm of an asset price $dX_t = (\kappa + \beta\sigma^2)dt + \sigma dW_t$, that takes into account the empirical finding that the volatility σ^2 varies randomly over time:

$$dX_t = (\kappa + \beta v_t)dt + \sqrt{v_t}dW_t. \tag{4.21}$$

Here the volatility v_t is a stochastic process that cannot be observed directly. If the data are observations at the time points Δi, $i = 0, 1, 2, \ldots, n$, then the returns $Y_i = X_{i\Delta} - X_{(i-1)\Delta}$ can be written in the form

$$Y_i = \kappa\Delta + \beta S_i + \sqrt{S_i}A_i, \tag{4.22}$$

where

$$S_i = \int_{(i-1)\Delta}^{i\Delta} v_t dt, \tag{4.23}$$

and where the A_i's are independent, standard normal distributed random variables. Here we consider the case where v is a sum of independent Pearson diffusions with state-space $(0, \infty)$ (the cases 2, 4 and 5). Barndorff-Nielsen & Shephard (2001) demonstrated that an autocorrelation function of the type (4.18) fits empirical autocorrelation functions of volatility well, while an autocorrelation function like that of a single Pearson diffusion is too simple to obtain a good fit. Stochastic volatility models where the volatility process is a sum of independent CIR-processes (Pearson diffusions with gamma marginals) were considered by Bollerslev & Zhou (2002) and Bibby & Sørensen (2003). Meddahi (2001) and Meddahi (2002a) studied models where the volatility process is a more complicated function of one or two diffusion processes. We assume that v and W are independent, so that the sequences $\{A_i\}$ and $\{S_i\}$ are independent.

By the multinomial formula we find that

$$E\left(Y_1^{k_1} Y_{t_1}^{k_2} Y_{t_2}^{k_3} Y_{t_3}^{k_4}\right) =$$

$$\sum K_{k_{11},\ldots,k_{43}} E(S_1^{k_{12}+k_{13}/2} S_{t_1}^{k_{22}+k_{23}/2} S_{t_2}^{k_{32}+k_{33}/2} S_{t_3}^{k_{42}+k_{43}/2}) E(A_1^{k_{13}}) E(A_{t_1}^{k_{23}}) E(A_{t_2}^{k_{33}}) E(A_{t_3}^{k_{43}}),$$

where the sum is over all non-negative integers k_{ij}, $i = 1, 2, 3, 4$, $j = 1, 2, 3$ such that $k_{i1} + k_{i2} + k_{i3} = k_i$ $(i = 1, 2, 3, 4)$, and where

$$K_{k_{11},\ldots,k_{43}} = \binom{k_1}{k_{11}, k_{12}, k_{13}} \binom{k_2}{k_{21}, k_{22}, k_{23}} \binom{k_3}{k_{31}, k_{32}, k_{33}} \binom{k_4}{k_{41}, k_{42}, k_{43}} (\kappa\Delta)^{k_{\cdot 1}} \beta^{k_{\cdot 2}}$$

with $k_{\cdot j} = k_{1j} + k_{2j} + k_{3j} + k_{4j}$. The moments $E(A_i^{k_{i3}})$ are the well-known moments of the standard normal distribution. When k_{i3} is odd, these moments are zero. Thus we only need to calculate the mixed moments of the form $E(S_1^{\ell_1} S_{t_1}^{\ell_2} S_{t_2}^{\ell_3} S_{t_3}^{\ell_4})$, where $\ell_1, \ldots, \ell_4$ are integers. However, when the volatility process is a sum of independent Pearson diffusions, S_i of the same form as Z_i in (4.20) (apart from $1/\Delta$), so we can proceed as in the previous section. Thus also for the stochastic volatility models defined in terms of Pearson diffusions we can explicitly find the optimal estimating function based on prediction of powers of returns.

4.5 Asymptotics.

In this subsection we give a result on the asymptotic distribution of the estimators obtained from prediction based estimating functions for the model types discussed above. We assume that the estimating function has the form

$$G_n(\psi) = A(\psi) \sum_{i=r+1}^{n} H^{(i)}(\psi) \tag{4.24}$$

with $H^{(i)}(\psi)$ given by (4.6), and that it is based on predicting powers up to m of the observations. The observations are either Y_i given by (4.14), (4.16) or (4.22) or Z_i given by (4.20). We denote the true value of ψ by ψ_0. The following theorem is proved in the appendix.

Theorem 4.5 *Assume that the underlying Pearson diffusions have finite $(4m+\epsilon)$'th moment ($a < (4m-1+\epsilon)^{-1}$) for some $\epsilon > 0$. Suppose, moreover, that $A(\psi)$ is twice continuously differentiable, and that the matrices $A(\psi)$, $A(\psi)\bar{M}(\psi)A(\psi)^T$ and $\partial_{\psi^T}\hat{a}$ have full rank, d, where $\hat{a}(\psi)$ is given by (4.11) and $\bar{M}(\psi)$ by (4.12). Then with probability tending to one as $n \to \infty$ there exists a solution $\hat{\psi}_n$ to the estimating equation $G_n(\psi) = 0$ such that $\hat{\psi}_n$ converges to ψ_0 in probability and*

$$\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, W^{-1}(\psi_0)V(\psi_0)(W^{-1}(\psi_0))^T\right),$$

where $W(\psi_0) = A(\psi_0)\bar{C}(\psi_0)\partial_{\psi^T}\hat{a}(\psi_0)$ and $V(\psi_0) = A(\psi_0)\bar{M}(\psi_0)A(\psi_0)^T$ with $\bar{C}(\psi)$ given by (4.10). For the optimal matrix $A^*(\psi) = \partial_\psi \hat{a}(\psi)^T \bar{C}(\psi)\bar{M}(\psi)^{-1}$, the asymptotic covariance matrix of $\hat{\psi}_n$ simplifies to

$$\left[\partial_\psi \hat{a}(\psi_0)^T \bar{C}(\psi_0)\bar{M}(\psi_0)^{-1}\bar{C}(\psi_0)\partial_{\psi^T}\hat{a}(\psi_0)\right]^{-1}.$$

A Proofs and general asymptotics.

Theorem 3.1 can be established using standard asymptotic techniques. Regularity conditions to ensure the existence of a consistent and asymptotically normal sequence of solutions to a general martingale estimating equation of form (3.3) can be found in Sørensen (1999). In case estimates are inserted for the parameter in the weights, as in (3.5), we need somewhat stronger conditions. The following result is taken from Jacod & Sørensen (2007).

Theorem A.1 *Suppose that $\{Y_i\}_{i\in\mathbb{N}_0}$ is a stationary ergodic process with state-space D and that*

$$G_n(\psi) = \sum_{i=1}^{n} \alpha(Y_{i-1},\psi)h(Y_{i-1},Y_i,\psi)$$

is a martingale estimating function such that the following holds.

A1: *The true parameter ψ_0 belongs to the interior of Ψ.*

A2: *For all ψ in a neighborhood of ψ_0 each of the variables $\alpha(Y_{i-1},\psi_0)h(Y_{i-1},Y_i,\psi)$ is P_{ψ_0}-integrable and $\alpha(Y_{i-1},\psi_0)h(Y_i,Y_{i-1},\psi_0)$ is square integrable.*

A3: *The mappings $\psi \mapsto \alpha(x,\psi)$ and $\psi \mapsto h(x,y,\psi)$ are continuously differentiable in a neighborhood of ψ_0 for all $x, y \in D$.*

A4: *For all ψ, ψ' in a neighborhood of ψ_0 each of the entries of $\partial_{\psi_k}\alpha(Y_{i-1},\psi)h(Y_{i-1},Y_i,\psi_0)$, $\alpha(Y_{i-1},\psi_0)\partial_{\psi_k}h(Y_{i-1},Y_i,\psi)$, and $\partial_{\psi_k}\alpha(Y_{i-1},\psi)\partial_{\psi_{k'}}h(Y_{i-1},Y_i,\psi')$ $(k,k'=1,\ldots,d)$ is dominated by a P_{ψ_0}-integrable function.*

A5: *The $d \times d$ matrix $W(\psi_0) = E_{\psi_0}\left\{\partial_{\psi^T}\left[\alpha(Y_{i-1},\psi)h(Y_{i-1},Y_i,\psi)\right]\right\}$ is invertible.*

Then with probability tending to one as $n \to \infty$ the estimating equation $G_n(\psi) = 0$ has a solution, $\hat{\psi}_n$, satisfying that $\hat{\psi}_n \to \psi_0$ in probability and

$$\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, W(\psi_0)^{-1}V(\psi_0)W(\psi_0)^{-1})$$

where $V(\psi_0) = E_{\psi_0}\{\alpha(Y_{i-1},\psi_0)h(Y_{i-1},Y_i,\psi_0)h(Y_{i-1},Y_i,\psi_0)^T\alpha(Y_{i-1},\psi_0)^T\}$*. The same result holds for the estimating function*

$$\tilde{G}_n(\psi) = \sum_{i=1}^{n} \alpha(Y_{i-1},\tilde{\psi}_n)h(Y_{i-1},Y_i,\psi),$$

where $\tilde{\psi}_n$ *is a* $\sqrt{n}$*-consistent estimator of* ψ*.*

Proof of Theorem 3.1:
Preliminarily we demonstrate that $V(x,\psi)$ is positive definite for all (x,ψ) and that the smallest eigenvalue is bounded away from zero uniformly in (x,ψ) when ψ belongs to a compact subset $\Psi_0 \subset \Psi$. Clearly $V(x,\psi)$ is positive semidefinite for all (x,ψ). Moreover, for $z \in \mathbb{R}^N$ it holds that $z^TV(x,\psi)z = 0$ if and only if

$$\sum_{j=1}^{N} z_j\{p_j(y,\psi) - e^{-\lambda_j(\psi)\Delta}p_j(x,\psi)\} = 0$$

for almost every y with respect to the conditional distribution of Y_i given $Y_{i-1} = x$ under ψ. However, the above is a polynomial in y and thus cannot equal zero almost surely unless the order is zero. As $p_j(y,\psi)$ is a j'th order polynomial with leading coefficient $p_{j,j} = 1$ we deduce that that $z^TV(x,\psi)z = 0$ if and only if $z = 0$. Hence, $V(x,\psi)$ is positive definite. By continuity the smallest eigenvalue

$$\varepsilon_1\{V(x,\psi)\} = \inf\{z^TV(x,\psi)z : |z| = 1\}$$

is bounded away from zero on compact subsets of $\Psi \times \mathcal{X}$ where $\mathcal{X}$ is the state space. To make the bound valid for all $x \in \mathcal{X}$ we need only check that it holds as $|x| \to \infty$. To this end note that $z^TV(x,\psi)z$ is a non-zero polynomial in x of order at most $2N$ the coefficient of which are given as continuous functions of z and ψ. If a sequence (x_n,z_n,ψ_n) were to exist such that $z_n^TV(x_n,\psi_n)z_n \to 0$, $|x_n| \to \infty$, $|z_n| = 1$, and $\{\psi_n\} \subset \Psi_0$, then we would find an accumulation point (z_0,ψ_0) such that $z_0^TV(x_n,\psi_0)z_0 \to 0$ although $z_0^TV(x,\psi_0)z_0$ defines a non-zero polynomial in x. By contradiction we conclude that $\inf\{\varepsilon_1\{V(x,\psi)\} : x \in \mathcal{X}, \psi \in \Psi_0\} > 0$.

As to the regularity conditions, **A1** holds true by assumption, and **A3** follows from **R2** as $\alpha^\star(x,\cdot)$ and $h(y,x,\cdot)$ are continuously differentiable with respect to the canonical parameter.
In order to check the integrability condition **A2** let Ψ_0 be a compact neighbourhood of ψ_0 and denote by $||B|| = \max_{j,k}|B_{j,k}|$ the max-norm of a matrix. A diagonalization argument shows that

$$||V(x,\psi)^{-1}|| \leq \frac{N^2}{C_1(\Psi_0)}$$

for all x and all $\psi \in \Psi_0$ where $C_1(\Psi_0)$ is the lower bound on the smallest eigenvalue of $V(x,\psi)$ on $\mathcal{X} \times \Psi_0$. Thus, by continuity of the coefficients a constant $C_2(\Psi_0)$ exist such that

$$\begin{aligned}|\alpha^\star(Y_{i-1},\psi_0)h(Y_i,Y_{i-1},\psi)| &\leq d \cdot N \cdot ||S(Y_{i-1},\psi_0)|| \cdot ||V(Y_{i-1},\psi_0)^{-1}|| \cdot |h(Y_i,Y_{i-1},\psi)| \\ &\leq C_2(\Psi_0)(1 + Y_{i-1}^{2N} + Y_i^{2N})\end{aligned}$$

for all $\psi \in \Psi_0$. The latter is integrable by **R0**. Further we note that

$$E_{\psi_0}\{\alpha^\star(Y_{i-1},\psi_0)h(Y_i,Y_{i-1},\psi_0)h(Y_i,Y_{i-1},\psi_0)^T\alpha^\star(Y_{i-1},\psi_0)^T\} \\ = E_{\psi_0}\{S(Y_{i-1},\psi_0)^T V(Y_{i-1},\psi_0)^{-1}S(Y_{i-1},\psi_0)\},$$

which by **R0** is finite because

$$||S(x,\psi_0)^T V(x,\psi_0)^{-1}S(x,\psi_0)|| \leq N^2 \cdot ||S(x,\psi_0)||^2 \cdot ||V(x,\psi_0)^{-1}|| \leq C_3(\psi_0)(1+x^{2N})$$

for some constant $C_3(\psi_0)$.
Similar bounds can be established for the derivatives of **A4**.
Finally, let us check that **A5** holds true. Clearly, $W(\psi_0)$ is negative semidefinite as

$$W(\psi_0) = -E_{\psi_0}\{S(Y_{i-1},\psi_0)^T V(Y_{i-1},\psi_0)^{-1}S(Y_{i-1},\psi_0)\}.$$

Let $z \in \mathbb{R}^d$ be such that $z^T W(\psi_0)z = 0$. The task is to demonstrate that $z = 0$. As $V(x,\psi_0)$ is positive definite for all x the assumption is that $S(x,\psi_0)z = 0$ for almost every x. We assume without loss of generality that $\psi = \tau = (\theta,\mu,a,b,c)$ is the canonical parameter. The general case follows readily as

$$S(x,\psi_0) = S(x,\tau_0) \cdot \partial_{\psi^T}\tau(\psi_0)$$

where by **R3** $\partial_{\psi^T}\tau(\psi_0)$ has full rank d. Hence, the assumption is

$$E_{\tau_0}(\partial_{\tau^T}\{p_j(Y_i,\tau_0) - e^{-\lambda_j(\tau_0)\Delta}p_j(Y_{i-1},\tau_0)\} \cdot z|Y_{i-1} = x) = 0$$

for $j = 1,\ldots,N$ and almost every x. The first equation reads

$$z_1(x-\mu_0)\Delta e^{-\theta_0\Delta} + z_2(e^{-\theta_0\Delta} - 1) \overset{\text{a.e.x}}{=} 0$$

which only holds true if $z_1 = z_2 = 0$. As $N \geq 2$ at least one more equation is available, namely

$$z_3 S_{2,3}(x,\tau_0) + z_4 S_{2,4}(x,\tau_0) + z_5 S_{2,5}(x,\tau_0) = 0$$

where

$$\begin{aligned} S_{2,3}(x,\tau_0) &= -2\theta\Delta e^{-2(1-a)\theta\Delta}p_2(x,\tau_0) + \frac{4(\mu+b)}{(2a-1)^2}(e^{-2(1-a)\theta\Delta} - e^{-\theta\Delta})x \\ &\quad -\frac{4\mu(\mu+b)}{(2a-1)^2}(1-e^{-\theta\Delta}) + \left\{\frac{\mu(\mu+b)(4a-3)}{(2a-1)^2(a-1)^2} + \frac{c}{(a-1)^2}\right\}(e^{-2(1-a)\theta\Delta} - 1), \\ S_{2,4}(x,\tau_0) &= \frac{2}{2a-1}(e^{-\theta\Delta} - e^{-2(1-a)\theta\Delta})x + \frac{2\mu}{2a-1}(1-e^{-\theta\Delta}) + \frac{\mu(1-e^{-2(1-a)\theta\Delta})}{(2a-1)(a-1)}, \\ S_{2,5}(x,\tau_0) &= \frac{1}{a-1}(1-e^{-2(1-a)\theta\Delta}) \end{aligned}$$

from which we deduce that $z_3 = z_4 = z_5$. □

Proof of Theorem 4.5:
The result follows from Theorem 6.2 in Sørensen (2000). We just need to check the

conditions of that theorem. First, we note that the observations are exponentially α-mixing. In the cases of integrated Pearson diffusions and sums of Pearson diffusions, this follows immediately from the exponential α-mixing of the Pearson diffusions. That the sequence of observations of a stochastic volatility model with exponentially α-mixing volatility process is exponentially α-mixing, was proven in Sørensen (2000) in the case $\kappa = \beta = 0$. The proof holds in the more general case too, see also the more general result in Genon-Catalot, Jeantheau & Laredo (2000). Secondly, we need that $H^{(i)}(\psi)$, given by (4.6), has finite $(2+\delta)$'th moment for some $\delta > 0$. In the case of an integrated Pearson diffusion this follows from Jensen's inequality and Fubini's theorem:

$$E_\psi\left(\left|\int_0^\Delta X_s ds\right|^{2m(2+\delta)}\right) \leq \int_0^\Delta E_\psi\left(|X_s|^{2m(2+\delta)}\right) ds = \Delta E_\psi\left(|X_0|^{2m(2+\delta)}\right) < \infty.$$

In the case of a sum of Pearson diffusions, it follows from Minkowski's inequality that the $(2+\delta)$'th moment of $H^{(i)}(\psi)$ is finite. For Pearson stochastic volatility models, Minkowski's inequality shows that it is sufficient that the integrated volatility process has finite $(4m+\epsilon)$'th moment, and integrated Pearson diffusion were considered above. Finally, it follows from (2.5) that the (finite) moments of a Pearson diffusion are twice continuously differentiable, so that $\hat{a}$ is twice continuously differentiable, cf. (4.2) and (4.3). Now all conditions of Theorem 6.2 in Sørensen (2000) have been shown to hold.

□

Acknowledgements

We are grateful to the referees for their many valuable and helpful comments that have very significantly improved the paper.

References

Barndorff-Nielsen, O. E.; Jensen, J. L. & Sørensen, M. (1990). "Parametric Modelling of Turbulence". *Phil. Trans. R. Soc. Lond. A*, 332:439–455.

Barndorff-Nielsen, O. E.; Jensen, J. L. & Sørensen, M. (1998). "Some Stationary Processes in Discrete and Continuous Time". *Advances in Applied Probability*, 30:989–1007.

Barndorff-Nielsen, O. E.; Kent, J. & Sørensen, M. (1982). "Normal variance-mean mixtures and z-distributions". *International Statistical Review*, 50:145–159.

Barndorff-Nielsen, O. E. & Shephard, N. (2001). "Non-Gaussian Ornstein-Uhlenbeck-Based Models and some of their Uses in Financial Econometrics (with discussion)". *Journal of the Royal Statistical Society* **B**, 63:167–241.

Berg, C. & Vignat, C. (2006). "Linearization coefficients of Bessel polynomials and properties of Student *t*-distributions". To appear in *Constr. Approx.*, published online DOI:10.1007/s00365-006-0643-6.

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

Bibby, B. M.; Jacobsen, M. & Sørensen, M. (2004). “Estimating functions for Discretely Sampled Diffusion-type Models”. AMS-preprint 4.

Bibby, B. M.; Skovgaard, I. M. & Sørensen, M. (2005). “Diffusion-type Models with given Marginals and Autocorrelation Function”. *Bernoulli*, 11:191–220.

Bibby, B. M. & Sørensen, M. (2003). “Hyperbolic processes in finance”. In Rachev, S., editor, *Handbook of Heavy Tailed Distributions in Finance*, pages 211–248. Elsevier Science.

Bollerslev, T. & Zhou, H. (2002). “Estimating stochastic volatility diffusion using conditional moments of integrated volatility”. *Journal of Econometrics*, 109:33–65.

Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer-Verlag, New York.

Campbell, J. Y.; Lo, A. W. & MacKinlay, A. C. (1997). *The Econometrics of Financial Markets.* Princeton University Press, Princeton.

Carrasco, M.; Chernov, M.; Florens, J.-P. & Ghysels, E. (2002). “Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions”. Working paper, University of North Carolina.

Chacko, G. & Viceira, L. M. (2003). “Spectral GMM estimation of continuous-time processes”. *Journal of Econometrics*, 116:259–292.

Christensen, B. J. & Sørensen, M. (2007). “Optimal inference in dynamic models with conditional moment restrictions”. In preparation.

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

De Jong, F.; Drost, F. C. & Werker, B. J. M. (2001). “A jump-diffusion model for exchange rates in a target zone”. *Statistica Neerlandica*, 55:270–300.

Ditlevsen, P. D.; Ditlevsen, S. & Andersen, K. K. (2002). “The fast climate fluctuations during the stadial and interstadial climate states”. *Annals of Glaciology*, 35:457–462.

Ditlevsen, S. & Sørensen, M. (2004). “Inference for observations of integrated diffusion processes”. *Scand. J. Statist.*, 31:417–429.

Duffie, D.; Filipović, D. & Schachermayer, W. (2003). “Affine processes and applications in finance”. *Annals of Applied Probability*, 13:984–1053.

Feller, W. (1951). “Diffusion processes in genetics”. In Neyman, J., editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statisitcs and Probability*, pages 227–246. University of California Press.

Forman, J. L. (2005). "Least squares estimation for correlation parameters with applications to sums of Ornstein-Uhlenbeck type processes". AMS preprint 3.

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

Gloter, A. (2000). "Parameter estimation for a discrete sampling of an integrated Ornstein-Uhlenbeck process". *Statistics*, 35:225–243.

Gloter, A. (2006). "Parameter estimation for a discretely observed integrated diffusion process". *Scand. J. Statist.*, 33:83–104.

Gobet, E. (2002). "LAN property for ergodic diffusions with discrete observations". *Ann. I. H. Poincaré PR*, 38:711–737.

Godambe, V. P. & Heyde, C. C. (1987). "Quasi likelihood and optimal estimation". *International Statistical Review*, 55:231–244.

Gourieroux, C. & Jasiak, J. (2006). "Multivariate Jacobi process and with application to smooth transitions". *Journal of Econometrics*, 131:475–505.

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

Hansen, L. P.; Scheinkman, J. A. & Touzi, N. (1998). "Spectral methods for identifying scalar diffusions". *Journal of Econometrics*, 86:1–32.

Heyde, C. C. (1997). *Quasi-Likelihood and Its Application.* Springer-Verlag, New York.

Hildebrandt, E. H. (1931). "Systems of polynomials connected with the Charlier expansions and the Pearson differential and difference equations". *Ann. Math. Statist.*, 2:379–439.

Jacod, J. & Sørensen, M. (2007). "Asymptotic statistical theory for stochastic processes: A review.". Preprint, Department of Applied Mathematics and Statistics, University of Copenhagen. In preparation.

Karlin, S. & Taylor, H. M. (1981). *A Second Course in Stochastic Processes.* Academic Press.

Kessler, M. (1996). *Estimation paramétrique des coefficients d'une diffusion ergodique à partir d'observations discrètes.* Ph.d. thesis, Laboratoire de Probabilités, Université Paris VI.

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

Larsen, K. S. & Sørensen, M. (2007). "A diffusion model for exchange rates in a target zone". *Mathematical Finance*, 17:285–306.

Meddahi, N. (2001). "An eigenfunction approach for volatility modelling". Working paper, Département de sciences économiques, CIRANO.

Meddahi, N. (2002a). "Moments of continuous time stochastic volatility". Working paper, Département de sciences économiques, CIRANO.

Meddahi, N. (2002b). "Theoretical comparison between integrated and realized volatility". *Journal of Applied Econometrics*, 17:479–508.

Nagahara, Y. (1996). "Non-Gaussian distribution for stock returns and related stochastic differential equation". *Financial Engineering and the Japanese Markets*, 3:121–149.

Ozaki, T. (1985). "Non-linear time series models and dynamical systems". In Hannan, E. J.; Krishnaiah, P. R. & Rao, M. M., editors, *Handbook of Statistics, Vol. 5*, pages 25–83. Elsevier Science Publishers.

Pearson, K. (1895). "Contributions to the Mathematical Theory of Evolution II. Skew Variation in Homogeneous Material". *Philosophical Transactions of the Royal Society of London. A*, 186:343–414.

Romanovsky, V. (1924). "Generalization of some types of the frequency curves of Professor Pearson". *Biometrika*, 16:106–117.

Singleton, K. J. (2001). "Estimation of affine asset pricing models using the empirical characteristic function". *Journal of Econometrics*, 102:111–141.

Sørensen, M. (1997). "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. Hayward: Institute of Mathematical Statistics. IMS Lecture Notes – Monograph Series, Vol. 32.

Sørensen, M. (1999). "On asymptotics of Estimating Functions". *Brazilian Journal of Probability and Statistics*, 13:111–136.

Sørensen, M. (2000). "Prediction-Based Estimating Functions". *Econometrics Journal*, 3:123–147.

Sørensen, M. (2007). "Efficient martingale estimating functions for discretely sampled diffusions". In preparation.

Wong, E. (1964). "The Construction of a Class of Stationary Markoff Processes". In Bellman, R., editor, *Stochastic Processes in Mathematical Physics and Engineering*, pages 264–276. American Mathematical Society, Rhode Island.

Zhou, H. (2001). "Finite sample properties of EMM, GMM, QMLE and MLE for a square-root interest rate diffusion model". *Journal of Computational Finance*, 5:89–122.

Zhou, H. (2003). "Itô conditional moment generator and the estimation of short-rate processes". *Journal of Financial Econometrics*, 1:250–271.

Research Papers 2007

Center for Research in Econometric Analysis of Time Series

2007-16 Tim Bollerslev, Michael Gibson and Hao Zhou: Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities

2007-17 Tim Bollerslev and Hao Zhou: Expected Stock Returns and Variance Risk Premia

2007-18 Torben G. Andersen, Tim Bollerslev and Francis X. Diebold: Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility

2007-19 Tim Bollerslev, Tzuo Hann Law and George Tauchen:Risk, Jumps, and Diversification

2007-20 Torben G. Andersen, Tim Bollerslev, Francis X. Diebold and Clara Vega: Real-Time Price Discovery in Global Stock, Bond and Foreign Exchange Markets

2007-21 Torben G. Andersen, Tim Bollerslev, Per Houmann Frederiksen and Morten Ørregaard Nielsen: Continuous-Time Models, Realized Volatilities, and Testable Distributional Implications for Daily Stock Returns

2007-22 Tim Bollerslev, Uta Kretschmer, Christian Pigorsch and George Tauchen: A Discrete-Time Model for Daily S&P500 Returns and Realized Variations: Jumps and Leverage Effects

2007-23 Olaf Posch: Structural estimation of jump-diffusion processes in macroeconomics

2007-24 Torben G. Andersen and Oleg Bondarenko: Construction and Interpretation of Model-Free Implied Volatility

2007-25 Torben G. Andersen and Luca Benzoni: Do Bonds Span Volatility Risk in the U.S. Treasury Market? A Specification Test for Affine Term Structure Models

2007-26: Mark Podolskij and Daniel Ziggel: A Range-Based Test for the Parametric Form of the Volatility in Diffusion Models

2007-27: Mark Podolskij and Mathias Vetter: Estimation of Volatility Functionals in the Simultaneous Presence of Microstructure Noise and Jump

2007-28: Julie Lyng Forman and Michael Sørensen: The Pearson diffusions: A class of statistically tractable diffusion processes