MOMENT EQUATIONS AND HERMITE EXPANSION FOR NONLINEAR STOCHASTIC DIFFERENTIAL EQUATIONS WITH APPLICATION TO STOCK PRICE MODELS

Hermann Singer
FernUniversität in Hagen *

Exact moment equations for nonlinear Itô processes are derived. Taylor expansion of the drift and diffusion coefficients around the first conditional moment gives a hierarchy of coupled moment equations which can be closed by truncation or a Gaussian assumption. The state transition density is expanded into a Hermite orthogonal series with leading Gaussian term and the Fourier coefficients are expressed in terms of the moments. The resulting approximate likelihood is maximized by using a quasi Newton algorithm with BFGS secant updates. A simulation study for the CEV stock price model compares the several approximate likelihood estimators with the Euler approximation and the exact ML estimator (Feller, 1951).

Key Words: Stochastic differential equations; Nonlinear systems; Discrete measurements; Maximum likelihood estimation; Moment equations; Extended Kalman Filter; Hermite Expansion

1 INTRODUCTION

Continuous time stochastic processes are appropriate models for phenomena where no natural time interval in the dynamics is given. Examples are mechanical systems (Newton's equations) or stock price movements where no natural trading interval can be identified. Different from that, measurements of this continuous time process $Y(t)$ are frequently obtained only at discrete time points t_i (daily, weekly, quarterly, etc.), so that dynamical models in econometrics are mostly formulated for the measurement times (time series models). In contrast, we consider stochastic differential equations (SDE) for the state $Y(t)$, but

*Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung, D-58084 Hagen, Germany, Hermann.Singer@FernUni-Hagen.de

assume that only a sampled trajectory $Y_i := Y(t_i)$ can be measured (cf. e.g. Bergstrom, 1990, Singer, 1995). Therefore, maximum likelihood estimation for sampled continuous time models must be based on the transition probabilities in the observation interval Δt. Unfortunately, this key quantity is not analytically available in most cases and must be computed by approximate schemes. The most simple is based on the Euler approximation of the SDE. The resulting discrete time scheme leads to conditionally Gaussian transition densities. A related approach is based on the moment equations for the first and second moment (for a survey, cf. Singer, 2002). Again, a conditionally Gaussian scheme is obtained. Alternatively, the drift coefficient can be expanded around the measurements to obtain a locally linear SDE leading again to a conditionally Gaussian scheme (Shoji and Ozaki, 1997, 1998). Quasi likelihood methods using conditional moments are also discussed in Shoji (2002). Still another Gaussian approach using stopping times is discussed by Yu and Phillips (2001).
Whereas these approximations are extremely useful for small sampling intervals where the transition density only slightly deviates from normality, for larger intervals corrections are necessary which take account of skewness and kurtosis (and higher order characteristics) of the true density. Among these approaches are Monte Carlo simulations (Andersen and Lund, 1997, Elerian et al., 2001, Singer, 2002, 2003), approximate analytical approaches (Aït-Sahalia, 2002) and finite difference methods for the Fokker-Planck equation (cf. Jensen and Poulsen, 2002). In this paper we consider a Hermite expansion with leading Gaussian term, but in contrast to Aït-Sahalia (2002) the expansion coefficients are expressed in terms of conditional moments and computed by solving deterministic moment equations.
The article is outlined as follows: In section 2 the basic model is stated and the equation for the transition density is formulated. Section 3 briefly introduces the maximum likelihood method. Section 4 introduces the Hermite expansion used to approximate the transition density and the moment equations are derived in section 5. In section 6, a simulation study is performed using an SDE with nonlinear diffusion coefficient (Constant Elasticity of Variance – CEV), and the performance of the several density approximations are compared with the exact solution. Finally, in an appendix, the moment equations are derived.

2 NONLINEAR CONTINUOUS/DISCRETE STATE SPACE MODELS

We discuss the nonlinear *stochastic differential equation* (SDE)

$$dY(t) = f(Y(t), t, \psi)dt + g(Y(t), t, \psi)dW(t) \tag{1}$$

where discrete measurements Y_i are taken at times $\{t_0, t_1, \ldots, t_T\}$ and $t_0 \leq t \leq t_T$ according to

$$Y_i = Y(t_i) \tag{2}$$

In the state equation (1), $W(t)$ denotes a r-dimensional Wiener process and the state is described by the p-dimensional state vector $Y(t)$. It fulfils a system of stochastic differential equations in the sense of Itô (cf. Arnold, 1974) with initial condition $Y(t_0)$. The functions

$f : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \to \mathbb{R}^p$ and $g : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \to \mathbb{R}^p \times \mathbb{R}^r$ are called drift and diffusion coefficients, respectively.

Parametric estimation is based on the u-dimensional parameter vector ψ. The key quantity for the computation of the likelihood function is the transition probability $p(y, t|x, s)$ which is a solution of the Fokker-Planck equation

$$\begin{aligned} \frac{\partial p(y,t|x,s)}{\partial t} &= -\sum_i \frac{\partial}{\partial y_i}[f_i(y,t,\psi)p(y,t|x,s)] \\ &+\tfrac{1}{2}\sum_{ij} \frac{\partial^2}{\partial y_i \partial y_j}[\Omega_{ij}(y,t,\psi)p(y,t|x,s)] \end{aligned} \tag{3}$$

subject to the initial condition $p(y, s|x, s) = \delta(y - x)$ (Dirac delta function). The diffusion matrix is given by $\Omega = gg' : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \to \mathbb{R}^p \times \mathbb{R}^p$. Under certain technical conditions the solution of (3) is the conditional density of $Y(t)$ given $Y(s) = x$ (see, e.g. Wong and Hajek, 1985, ch. 4).

Extensions to nonlinear noisy measurements are given in Gordon et al. (1993), Kitagawa (1987, 1996), Hürzeler and Künsch (1998) and Singer (2003).

In order to model exogenous influences, f and g are assumed to depend on deterministic regressor variables $x(t) : \mathbb{R} \to \mathbb{R}^q$, i.e. $f(.) = f(y, t, x(t), \psi)$ etc. For notational simplicity, the dependence on the $x(t)$ will be suppressed.

3 COMPUTATION OF THE LIKELIHOOD FUNCTION

In order to compute the likelihood function of system (1, 2), we can express the probability distribution of states $Y(t_0), \ldots, Y(t_T)$ in terms of solutions of Fokker-Planck equation (3). Using the Markov property of $Y(t)$ we obtain

$$p(y_T, \ldots, y_1|y_0; \psi) = \prod_{i=0}^{T-1} p(y_{i+1}|y_i; \psi), \tag{4}$$

where $p(y_T, \ldots, y_1|y_0; \psi)$ is the joint distribution of measurements conditional on $Y(t_0) = y_0$ (cf. Lo, 1988) and $p(y_{i+1}|y_i; \psi) := p(y_{i+1}, t_{i+1}|y_i, t_i; \psi)$ is the transition probability density. Defining the likelihood function as $L_\psi(y) := p(y_T, \ldots, y_1|y_0; \psi)$ and the ML estimator as $\hat{\psi} := \arg\max_\psi L_\psi(Y)$ we must solve Fokker-Planck equation (3) repeatedly in a nonlinear optimization algorithm. Only in the case of a linear vector field f and state independent diffusion coefficient g we obtain a Gaussian transition density but otherwise complicated functions arise. In some special cases analytical solutions have been derived. For example, in the case of linear $f(y) = \mu y$ and $g(y) = \sigma y^{\alpha/2}$, which is the well known constant elasticity of variance (CEV) diffusion process used in option pricing (cf. Feller, 1951, Cox and Ross, 1976), an analytical solution has been derived by Feller involving Bessel functions. In the general multivariate case, we cannot hope to obtain analytical solutions and must resort to approximations and numerical procedures for (3) (matrix continued-fractions, finite differences, Monte Carlo methods etc.; cf. Risken, 1989, Press et al., 1992, Ames, 1992, Kloeden and Platen, 1992). Alternatively, the prediction error identification method (Ljung and Söderström, 1983), the extended Kalman filter EKF or other linearization methods (e.g. Shoji and Ozaki, 1997) lead to approximations of the conditional density in terms of conditional Gauss distributions (for a survey, see Jensen and Poulsen, 2002).

Conditional Gaussian approximations work well when the sampling intervals $\Delta t_i = t_{i+1} - t_i$ are not too large in comparision with the dynamics as specified in f and g. On the other hand, time series and panel data often involve large sampling intervals which are fixed by the design of the study. Therefore, corrections must be made to the Gaussian transition probability. Here we use a Hermite expansion with leading Gaussian term and corrections involving higher order moments.

4 HERMITE EXPANSION

The transition density $p(y_{i+1}|y_i;\psi)$ can be expanded into a Fourier series (cf. Courant and Hilbert, 1968, ch. II, 9, Abramowitz and Stegun, 1965, ch. 22) by using the complete set of Hermite polynomials which are orthogonal with respect to the weight function $w(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ (standard Gaussian density), i.e.

$$\int_{-\infty}^{\infty} H_n(x)H_m(x)w(x)dx = \delta_{nm}n! \qquad (5)$$

The Hermite polynomials $H_n(x)$ are defined by

$$\phi^{(n)}(x) := (d/dx)^n\phi(x) = (-1)^n\phi(x)H_n(x). \qquad (6)$$

and are given explicitly by $H_0 = 1, H_1 = x, H_2 = x^2 - 1, H_3 = x^3 - 3x, H_4 = x^4 - 6x^2 + 3$ etc. Therefore, the density function $p(x)$ can be expanded as [1]

$$p(x) = \phi(x)\sum_{n=0}^{\infty} c_n H_n(x). \qquad (7)$$

and the Fourier coefficients are given by

$$c_n := (1/n!)\int_{-\infty}^{\infty} H_n(x)p(x)dx = (1/n!)E[H_n(X)] \qquad (8)$$

where X is a random variable with density $p(x)$. Since the Hermite polynomials contain powers of x, the expansion coefficients can be expressed in terms of moments of X, i.e. $\mu_k = E[X^k]$. Explicitly the first terms are given by

$$c_0 := 1 \qquad (9)$$

$$c_1 := E[H_1(X)] = E[X] = \mu_1 := \mu \qquad (10)$$

[1] Actually, the expansion is in terms of the orthogonal system $\psi_n(x) = \phi(x)^{1/2}H_n(x)$ (oscillator eigenfunctions, i.e. $q(x) := p(x)/\phi(x)^{1/2} = \sum_{n=0}^{\infty} c_n\psi_n(x)$, so the expansion of $q = p/\phi^{1/2}$ must converge. The function to be expanded must be square integrable in the interval $(-\infty, +\infty)$, i.e. $\int q(x)^2dx = \int \exp(x^2/2)p^2(x)dx < \infty$ (Courant and Hilbert, 1968, p. 81–82).

$$
\begin{aligned}
c_2 &:= (1/2!)E[H_2(X)] = (1/2)E[X^2 - 1] = (1/2)(\mu_2 - 1) && (11)\\
c_3 &:= (1/3!)E[H_3(X)] = (1/6)E[X^3 - 3X] = (1/6)(\mu_3 - 3\mu) && (12)\\
c_4 &:= (1/4!)E[H_4(X)] = (1/24)E[X^4 - 6X^2 + 3] = (1/24)(\mu_4 - 6\mu_2 + 3) && (13)
\end{aligned}
$$

etc. Using the standardized variables $Z = (X - \mu)/\sigma$ with $\mu = E[X], \sigma^2 = E[X^2] - \mu^2$, $E[Z] = 0, E[Z^2] = 1, E[Z^k] := \nu_k$ one obtains the simplified expressions

$$
\begin{aligned}
c_0 &:= 1 && (14)\\
c_1 &:= 0 && (15)\\
c_2 &:= 0 && (16)\\
c_3 &:= (1/3!)E[Z^3] = (1/3!)\nu_3 && (17)\\
c_4 &:= (1/4!)E[Z^4 - 6Z^2 + 3] = (1/24)(\nu_4 - 3) && (18)
\end{aligned}
$$

and the density expansion

$$
p_z(z) \ := \ \phi(z)[1 + (1/6)\nu_3 H_3(z) + (1/24)(\nu_4 - 3)H_4(z) + ...] \qquad (19)
$$

which shows that the leading Gaussian term is corrected by higher order contributions containing skewness and kurtosis excess. For a Gaussian random variable, $p_z(z) = \phi(z)$, so the coefficients $c_k, k \geq 3$ all vanish. For example, the kurtosis of Z is $E[Z^4] = 3$, so $c_4 = 0$. Using the expansion for the standardized variable and the change of variables formula $p_x(x) = (1/\sigma)p_z(z); z = (x - \mu)/\sigma$ one derives the convenient formula

$$
p_x(x) \ := \ (1/\sigma)\phi(z)[1 + (1/6)\nu_3 H_3(z) + (1/24)(\nu_4 - 3)H_4(z) + ...] \qquad (20)
$$

and the standardized moments $\nu_k = E[Z^k] = E[(X - \mu)^k]/\sigma^k := m_k/\sigma^k$ can be expressed in terms of *centered moments*

$$
m_k := E[M_k] := E[(X - \mu)^k]. \qquad (21)
$$

For these moments differential equations will be derived in the following.

5 SCALAR MOMENT EQUATIONS

5.1 Conditionally Gaussian model

Denoting the conditional mean and variance as $\mu(t|t_i) = E[Y(t)|Y^i]$ and $m_2(t|t_i) = \mathrm{Var}[Y(t)|Y^i]$, where $Y^i = \{Y_i, \ldots, Y_0\}$ are the measurements up to time t_i, these fulfil the exact equations ($t_i \leq t \leq t_{i+1}$)

$$
\begin{aligned}
\dot{\mu}(t|t_i) &= E[f(Y(t), t)|Y^i] && (22)\\
\dot{m}_2(t|t_i) &= 2E[f(Y(t), t) * (Y(t) - \mu(t|t_i))|Y^i] + E[\Omega(Y(t), t)|Y^i] && (23)
\end{aligned}
$$

between measurements with initial condition $\mu(t_i|t_i) = Y_i; m_2(t_i|t_i) = 0$. These equations can be derived from Fokker-Planck equation (3) using partial integration (see appendix). They are not differential equations, however, because the expectation values require knowledge of the conditional density $p(y,t|Y^i)$.

Expanding the drift f and the diffusion term Ω around the mean $\mu(t|t_i)$ up to first order, and inserting this into (22-23) leads to the coupled closed system of approximate moment equations (Jazwinski, 1970, ch. 9)

$$\dot{\mu}(t|t_i) = f(\mu(t|t_i),t) \tag{24}$$

$$\dot{m}_2(t|t_i) = 2f'(\mu(t|t_i),t)m_2(t|t_i) + \Omega(\mu(t|t_i),t) \tag{25}$$

where the first derivative of the drift is defined by

$$f'(y,t) := \frac{\partial f(y,t)}{\partial y}. \tag{26}$$

These are the time update equations of the extended Kalman filter (EKF). If the diffusion coefficient is frozen at time t_i with measurement Y_i, i.e. $\Omega(\mu(t|t_i),t) = \Omega(Y_i,t_i)$, Nowman's method is obtained (Nowman, 1997, Yu and Phillips, 2001).

Expanding the drift f and the diffusion matrix Ω up to second order around the estimate $\mu(t|t_i)$ and inserting this into (22-23) leads to second order equations which are used in the so called second order nonlinear filter (SNF; Jazwinski, 1970, ch. 9; cf. eqn. 44).

A related method is the so called local linearization method (LL) of Shoji and Ozaki (1997, 1998). They use Itô's lemma and expand the drift into

$$f(Y(t),t) = f(Y_i,t_i) + \int_{t_i}^t f_y(Y,s)dY(s) + \int_{t_i}^t [f_s(Y,s) + \tfrac{1}{2}f_{yy}(Y,s)\Omega(Y,s)]ds \tag{27}$$

Freezing the coefficients at (Y_i,t_i) and approximation of the integrals yields

$$\begin{aligned} f(Y(t),t) \approx\ & f(Y_i,t_i) + f_y(Y_i,t_i)(Y(t)-Y_i) + \\ & + [f_s(Y_i,t_i) + \tfrac{1}{2}f_{yy}(Y_i,t_i)\Omega(Y_i,t_i)](t-t_i) \end{aligned} \tag{28}$$

Therefore the drift is approximately linear and one obtains the linear SDE

$$\begin{aligned} dY(t) \approx\ & f_y(Y_i,t_i)Y(t)dt + \\ & + [f(Y_i,t_i) - f_y(Y_i,t_i)Y_i + (f_s(Y_i,t_i) + \tfrac{1}{2}f_{yy}(Y_i,t_i)\Omega(Y_i,t_i))(t-t_i)]dt + \\ & + g(Y_i,t_i)dW(t) \end{aligned} \tag{29}$$

From this one obtains moment equations similar to the SNF (for a thorough discussion see Singer, 2002, sect. 3.3.–3.4).

In all cases the approximate likelihood function is computed recursively using the prediction error decomposition (Schweppe, 1965)

$$L_\psi(z) = \prod_{i=0}^{T-1} |2\pi\Gamma_{i+1|i}|^{-1/2} \exp\left\{-\tfrac{1}{2}\nu_{i+1}^2/\Gamma_{i+1|i}\right\} \tag{30}$$

$$\nu_{i+1} = Y_{i+1} - \mu(t_{i+1}|t_i) \tag{31}$$

$$\Gamma_{i+1|i} = m_2(t_{i+1}|t_i) \tag{32}$$

with prediction error ν_{i+1} and conditional variance $m_2(t_{i+1}|t_i)$. Since higher order moments were neglected, the (quasi) likelihood is a product of conditional Gaussian densities.

5.2 Higher order moments

The higher order conditional moments (for simplicity the condition and the time argument is suppressed)

$$m_k := E[M_k] := E[(Y-\mu)^k]. \tag{33}$$

fulfil the equations (see appendix)

$$\dot{m}_k = kE[f(Y)*(M_{k-1}-m_{k-1})] + \tfrac{1}{2}k(k-1)E[\Omega(Y)*M_{k-2}] \tag{34}$$

with initial condition $m_k(t_i|t_i) = 0$. Again, these are not differential equations, and Taylor expansion of f and Ω around μ yields

$$f(y) \ := \ \sum_{l=0}^{\infty} f^{(l)}(\mu)\frac{(y-\mu)^l}{l!} \tag{35}$$

Inserting this into (22, 34) yields

$$\dot{\mu} \ := \ \sum_{l=0}^{\infty} f^{(l)}(\mu)\frac{m_l}{l!} \tag{36}$$

$$= \ f(\mu) + \tfrac{1}{2}f''(\mu)m_2 + \tfrac{1}{6}f'''(\mu)m_3 + ... \tag{37}$$

and $(k \geq 2)$

$$\dot{m}_k \ = \ k\sum_{l=1}^{\infty}\frac{f^{(l)}(\mu)}{l!}(m_{l+k-1} - m_l m_{k-1}) + \tfrac{1}{2}k(k-1)\sum_{l=0}^{\infty}\frac{\Omega^{(l)}(\mu)}{l!}m_{l+k-2}. \tag{38}$$

In analogy to EKF and SNF, the abbreviation HNF(K, L) (higher order nonlinear filter) will be used.
For practical applications, three problems must be solved:

1. One must chose a number K of moments to consider.
2. The expansion of f and Ω must be truncated somewhere $(l = 0, ..., L)$.
3. On the right hand side moments of maximal order $L + K - 1$ occur, so that only in the special case $L = 1$ (locally linear approximation of f and Ω) a closed system of equations results. In other cases, two methods are frequently used:
   (a) Higher order moments are neglected: $m_k = 0; k > K$
   (b) Higher order moments are factorized by the Gaussian assumption

$$m_k = \begin{cases} (k-1)!!m_2^{k/2}; k > K & \text{is even} \\ 0; \qquad\qquad\quad k > K & \text{is odd} \end{cases} \tag{39}$$

5.2.1 Example: expansion up to 4th order (truncation)

Expanding f and Ω up to 4th order and using 4 moments with truncation, one obtains, for quick reference, the explicit system

$$\dot{\mu} = f(\mu) + (1/2)f''(\mu)m_2 + (1/6)f'''(\mu)m_3 + (1/24)f''''(\mu)m_4 \quad (40)$$

$$\dot{m}_2 = 2f'(\mu)m_2 + f''(\mu)m_3 + (1/3)f'''(\mu)m_4 + \Omega(\mu) + (1/2)\Omega''(\mu)m_2 + (1/6)\Omega'''(\mu)m_3 + (1/24)\Omega''''(\mu)m_4 \quad (41)$$

$$\dot{m}_3 = 3[f'(\mu)m_3 + (1/2)f''(\mu)(m_4 - m_2^2) - (1/6)f'''(\mu)m_2m_3 - (1/24)f''''(\mu)m_2m_4] + 3[\Omega'(\mu)m_2 + (1/2)\Omega''(\mu)m_3 + (1/6)\Omega'''(\mu)m_4)] \quad (42)$$

$$\dot{m}_4 = 4[f'(\mu)m_4 - (1/2)f''(\mu)m_2m_3 - (1/6)f'''(\mu)m_3^2 - (1/24)f''''(\mu)m_3m_4] + 6[\Omega(\mu)m_2 + \Omega'(\mu)m_3 + (1/2)\Omega''(\mu)m_4] \quad (43)$$

5.2.2 Example: second order nonlinear filter (SNF)

For example, setting $(K = 2, L = 2)$, one again obtains the second order nonlinear filter (SNF)[2]

$$\dot{\mu} = f(\mu) + \tfrac{1}{2}f''(\mu)m_2 \quad (44)$$

$$\dot{m}_2 = 2f'(\mu)m_2 + f''(\mu)m_3 + \Omega(\mu) + \tfrac{1}{2}\Omega''(\mu)m_2 \quad (45)$$

Neglecting m_3 (truncation or Gaussian asumption) a closed system occurs. Setting $(K = 2, L = 1)$ reproduces (24-25), the extended Kalman filter EKF.

5.2.3 Example: locally linear approximation $L = 1$

If we chose $L = 1$ (locally linear approximation), the moment equations yield the closed system

$$\dot{\mu} = f(\mu) \quad (46)$$

and $(k \geq 2)$

$$\dot{m}_k = kf'(\mu)m_k + \tfrac{1}{2}k(k-1)[\Omega(\mu)m_{k-2} + \Omega'(\mu)m_{k-1}] \quad (47)$$

For the second and third moment we obtain

$$\dot{m}_2 = 2f'(\mu)m_2 + \Omega(\mu) \quad (48)$$

$$\dot{m}_3 = 3f'(\mu)m_3 + 3\Omega'(\mu)m_2 \quad (49)$$

Thus, if $m_3(t_i|t_i) = 0$, which is the case at the times of measurement, and for a state independent diffusion coefficient $(\Omega(y,t) = \Omega(t))$ the solution $m_3(t|t_i)$ will remain zero and

$$\dot{m}_4 = 4f'(\mu)m_4 + 6\Omega(\mu)m_2 \quad (50)$$

[2]with trivial measurement model $z = y$

is solved by $m_4 = 3m_2^2$, as required by a Gaussian solution. This is not surprising since linear systems (1) generate Gaussian stochastic processes. Including higher derivatives f'', Ω' etc. yields deviations from Gaussianity and the skewness and kurtosis excess $m_4 - 3m_2^2$ will not remain null for large time intervals Δt_i.

For example, the square root stock price model (cf. Feller, 1951, Cox and Ross, 1976)

$$dY(t) = rY(t)dt + \sigma Y(t)^{1/2} dW(t). \tag{51}$$

has a linear drift and diffusion term $\Omega(y) = \sigma^2 y$ with derivatives $\Omega'(y) = \sigma^2$, $\Omega^{(l)}(y) = 0, l \geq 2$. In this case, the exact equations for the first and second moments

$$\dot{\mu} = r\mu \tag{52}$$

$$\dot{m}_2 = 2rm_2 + \sigma^2\mu \tag{53}$$

yield a closed linear system (see, e.g. Bibby and Sorensen, 1995). It can be solved explicitly by

$$\mu(t|t_i) = \exp[r(t - t_i)]Y_i \tag{54}$$

$$m_2(t|t_i) = \frac{\sigma^2}{r}[\exp(2r(t - t_i)) - \exp(r(t - t_i))]Y_i. \tag{55}$$

Freezing the diffusion term $\sigma^2\mu(t|t_i) = \sigma^2 Y_i$ yields the Nowman approximation method with solution

$$\mu(t|t_i) = \exp[r(t - t_i)]Y_i \tag{56}$$

$$m_2(t|t_i) = \frac{\sigma^2}{2r}[\exp(2r(t - t_i)) - 1]Y_i. \tag{57}$$

Both expressions coincide for small $r \to 0$. Since the drift is linear, the Shoji-Ozaki method (29) yields the same equations as the Nowman approximation.

The equation for the third moment

$$\dot{m}_3 = 3rm_3 + 3\sigma^2 m_2 \tag{58}$$

contains an inhomogenous term yielding a skewed density after some time. Moreover,

$$\dot{m}_4 = 4rm_4 + 6\sigma^2(\mu m_2 + m_3) \tag{59}$$

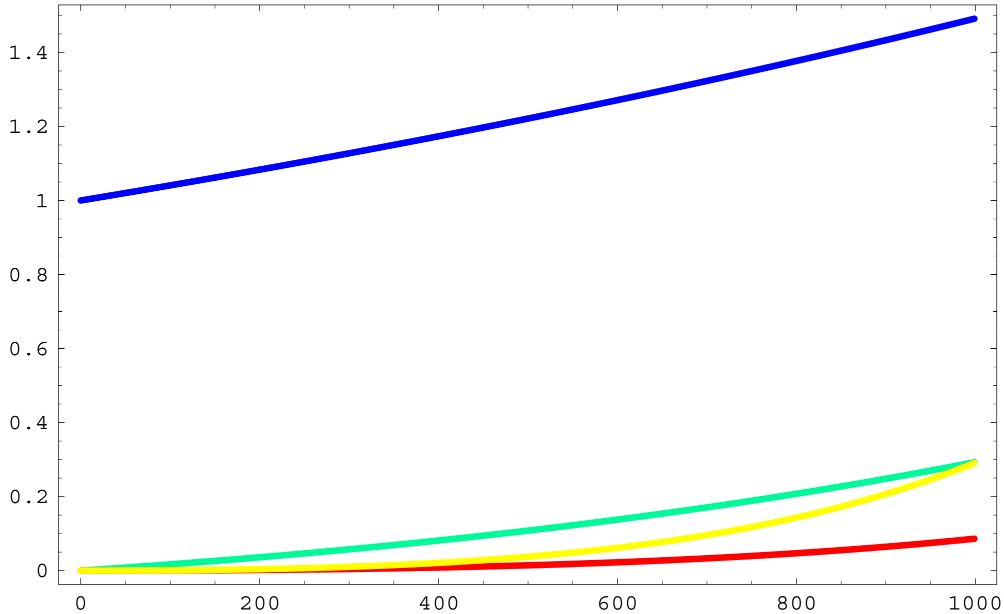
is not solved anymore by the Gaussian factorization $m_4 = 3m_2^2$ due to the skewness term $\sigma^2 m_3$.

For the parameter vector $\psi = \{r, \sigma\} = \{0.1, 0.2\}$ we obtain the equations $(K = 4, L = 1)$

$$d/dt \begin{bmatrix} \mu \\ m_2 \\ m_3 \\ m_4 \end{bmatrix} = \begin{bmatrix} 0.1\mu \\ 0.04\mu + 0.2m_2 \\ 0.12m_2 + 0.3m_3 \\ 6(0.04\mu m_2 + 0.04m_3) + 0.4m_4 \end{bmatrix} \tag{60}$$

The equations were solved by an Euler scheme with discretization interval $\delta t = 1/250$ year and $T = 1000$ time steps corresponding to $T\delta t = 4$ years and initial condition $m(t_i|t_i) =$

Figure 1: Square root model: evolution of 4 conditional moments in the time interval [0,4] using the Euler method with $\delta t = 1/250$ and $T = 1000$ time steps (μ=red, m_2=green, m_3=red, m_4=yellow).



$[1, 0, 0, 0]'$. The evolution of the 4 moments is shown in fig. (1). It can be seen that the skewness does not remain zero. The corresponding approximate densities $p_{k,L=1}(y)$, $k = 2, ..., 7$. (cf. eqn. 20) are plotted in fig. (2) together with the exact solution (Feller, 1951).
Unfortunately, the expansion does not converge, although low order approximations such as $k = 3$ are quite good (see figs. 6, 7 and next section).

5.3 Transformed equations and Jacobi terms

The Hermite expansion (7) is actually based on the Fourier series (footnote 1)

$$p(x)/\phi(x)^{1/2} = \sum_{n=0}^{\infty} c_n \psi_n(x) \tag{61}$$

$$\psi_n(x) = \phi(x)^{1/2} H_n(x) \tag{62}$$

in terms of oscillator eigenfunctions. This means, that the expansion for $q = p/\phi^{1/2}$ should exist ($\int p^2/\phi \, dx < \infty$) and p must be close to a normal distribution.

5.3.1 Log-normal density

For example, the transition density for the geometric brownian motion

$$dX = rXdt + \sigma XdW \tag{63}$$

Figure 2: Square root model: Approximate densities $p_{k,1}(y)$ with Hermite expansion up to $K = 7$ (orange) and exact density (red).

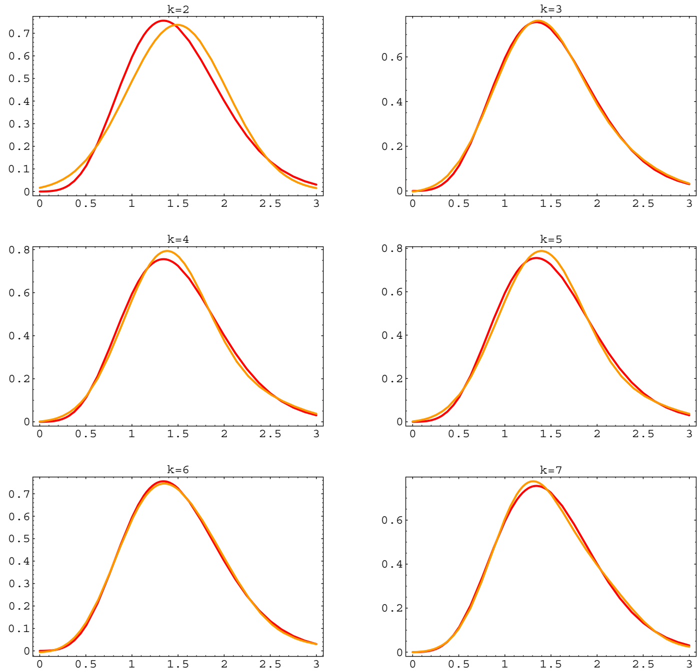
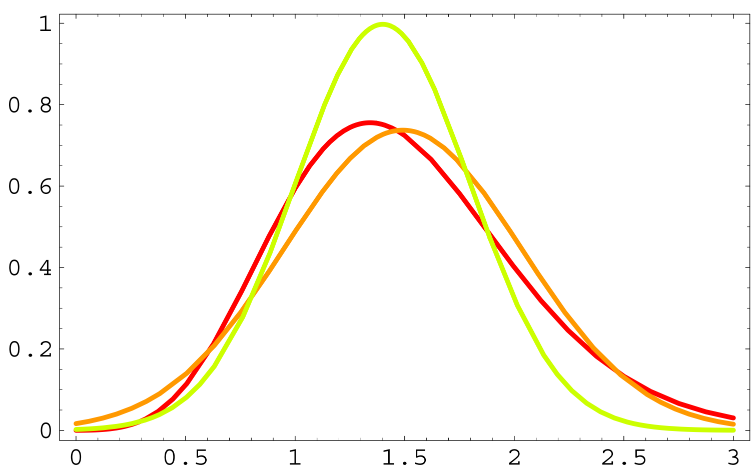


Figure 3: Square root model: exact density (red), approximate density $p_{2,1}(y)$ with Hermite expansion up to $K = 2$ (EKF, orange) and Euler density (green).



is lognormal, so that

$$\begin{aligned} p(x,t|x_0,t_0) &= 1/(x\sqrt{2\pi\gamma^2})\exp[-(\log(x/x_0)-\nu)^2/(2\gamma^2)] && (64)\\ \nu &= (r-\sigma^2/2)(t-t_0) && (65)\\ \gamma^2 &= \sigma^2(t-t_0), && (66)\end{aligned}$$

$p/\phi^{1/2} \to \infty$, and the Hermite series does not converge. As an example the log-normal variable $X = \exp(Y)$ with parameters $E[Y] = \nu = 1$ and $\mathrm{Var}[Y] = \gamma^2 = 1$ is considered. Thus we obtain $E[X] = \exp(\nu + \gamma^2/2) = 4.48169$. The normalized series expansion for $p(x)$ is shown in fig. (4). On the other hand, the direct expansion of $p(x)$

$$\begin{aligned} p(x) &= \sum_{n=0}^{\infty} b_n \psi_n(x) && (67)\\ \psi_n(x) &= \phi(x)^{1/2} H_n(x) && (68)\end{aligned}$$

in terms of oscillator eigenfunctions does converge (cf. footnote 1 and fig. 5), but the expansion coefficients

$$b_n := (1/n!)\int_{-\infty}^{\infty}\phi(x)^{1/2}H_n(x)p(x)dx = (1/n!)E[H_n(X)\phi(X)^{1/2}] \quad (69)$$

cannot be easily expressed in terms of moments of X. But these are the quantities we can compute from the moment equations (34) or by other approximation procedures.

5.3.2 Square root model

The square root model of section (5.2.3) can be solved exactly using Bessel functions (Feller, 1951) and the moments m_k were computed numerically from this exact density. A plot of the function $q^2(z) = p_z(z)^2/\phi(z)$ (fig. 6), where $p_z(z) = p_y(\mu + \sigma z)\sigma$ is the standardized density function, reveals that the convergence condition is not fulfilled. Expanding up to order $K = 19$, the nonconvergence is shown in fig. 7. As mentioned earlier, low order approximations such as $k = 3, 6$ are nevertheless quite good. Again, a direct expansion in terms of oscillator eigenfunctions yields a convergent series (67).

5.3.3 Transformation

Following an idea of Aït-Sahalia (2002), the Itô process $X(t)$ of interest is first transformed into $Y = \tau(X)$ using Itô's lemma, such that the diffusion coefficient is constant. It can be shown that the resulting transition density $p_y(y)$ is suffiently close to a normal density (Aït-Sahalia, loc. cit., prop. 2), so that a convergent Hermite expansion is possible. The original density can be computed by the change of variable formula

$$\begin{aligned} p_x(x)dx &= p_y(\tau(x))\tau'(x)dx && (70)\\ y &= \tau(x) && (71)\end{aligned}$$

Figure 4: Log-Normal density LN($\nu = 1, \gamma^2 = 1$) (red) and nonconvergent Hermite expansion of $p(x)$ with order 2 (orange), 3 (yellow) and 4 (green).

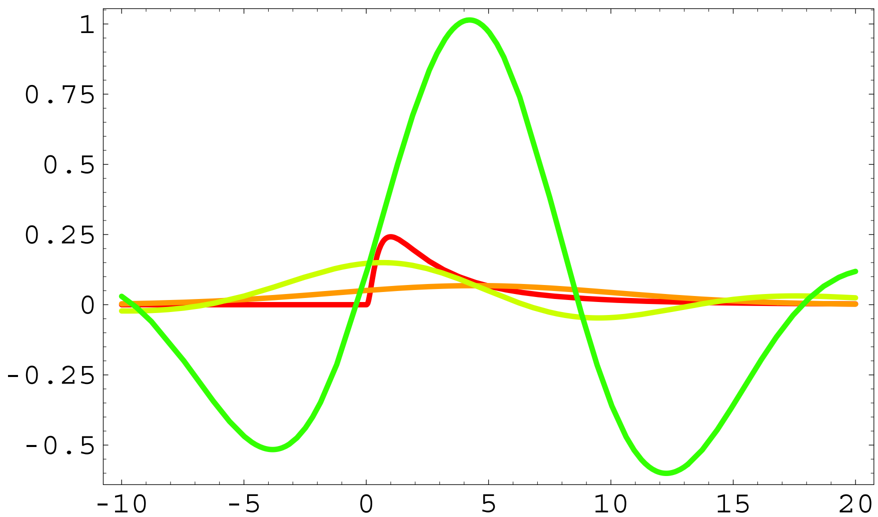


Figure 5: Log-Normal density LN($\nu = 1, \gamma^2 = 1$) (red) and convergent direct Hermite expansion of $p(x)$ in terms of oscillator eigenfunctions with order 20 (orange), 40 (yellow) and 60 (green).

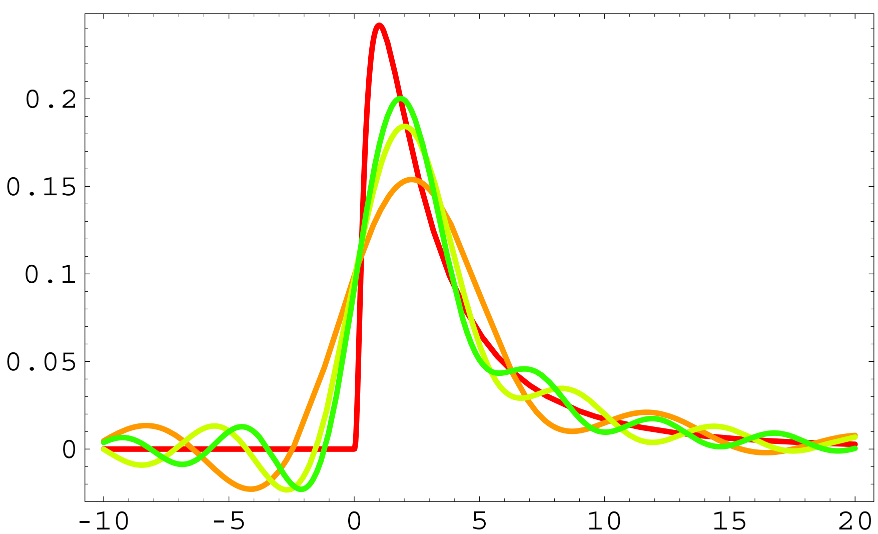
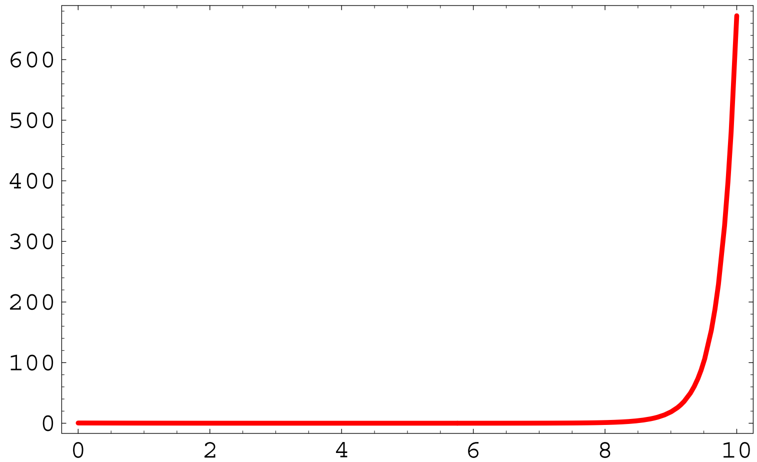


Figure 6: Square root model: plot of convergence condition $q^2(z) = p_z(z)^2/\phi(z)$ (q must be square integrable).



For the expansion of $p_y(y)$ we use the standardized expression

$$p_y(y) \quad := \quad (1/\sigma)\phi(z)[1 + (1/6)\nu_3 H_3(z) + (1/24)(\nu_4 - 3)H_4(z) + ...], \tag{72}$$

$Z = (Y - \mu)/\sigma; \mu = E[Y], \sigma^2 = \mathrm{Var}(Y)$ (cf. eqn. 20). Thus, for small time spans Δt, the conditional variance $\sigma^2 \approx \Delta t$, so Z corresponds to Aït-Sahalia's "pseudo-normalized" increment $(Y - \mu)/\sqrt{\Delta t}$. It remains to determine the transformation function. Itô's lemma yields

$$dY \quad = \quad \tau_x(X,t)dX + \tau_t(X,t)dt + (1/2)\tau_{xx}(X,t)dX^2 \tag{73}$$

$$\quad = \quad [\tau_x(X,t)f + \tau_t(X,t) + (1/2)\tau_{xx}(X,t)g^2(X,t)]dt + \tau_x(X,t)g(X,t)dW \tag{74}$$

$$dX \quad = \quad f(X,t)dt + g(X,t)dW \tag{75}$$

and thus

$$\tau_x(x,t)g(x,t) \quad = \quad 1 \tag{76}$$

$$\tau(x,t) \quad = \quad \int^x dx'/g(x',t). \tag{77}$$

For example, the diffusion term of geometric brownian motion is $g(x) = \sigma x$ and we obtain $y = \tau(x) = \int^x dx'/(\sigma x') = (1/\sigma)\log(x)$.
The transformation approach is simple in the scalar case, but for vector processes the system

$$\sum_{j=1}^{p} \frac{\partial \tau_i(x,t)}{\partial\, x_j} g_{jk}(x,t) \quad = \quad \delta_{ik};\ i = 1,...,p, k = 1,...,r \tag{78}$$

must be solved. Therefore, in the sequel we study the scalar case without transformation in order to apply the method in the multivariate case.

Figure 7: Square root model: Approximate densities $p_k(y)$ with Hermite expansion up to $K = 19$ (orange) and exact density (red). The moments were computed from the exact density function.

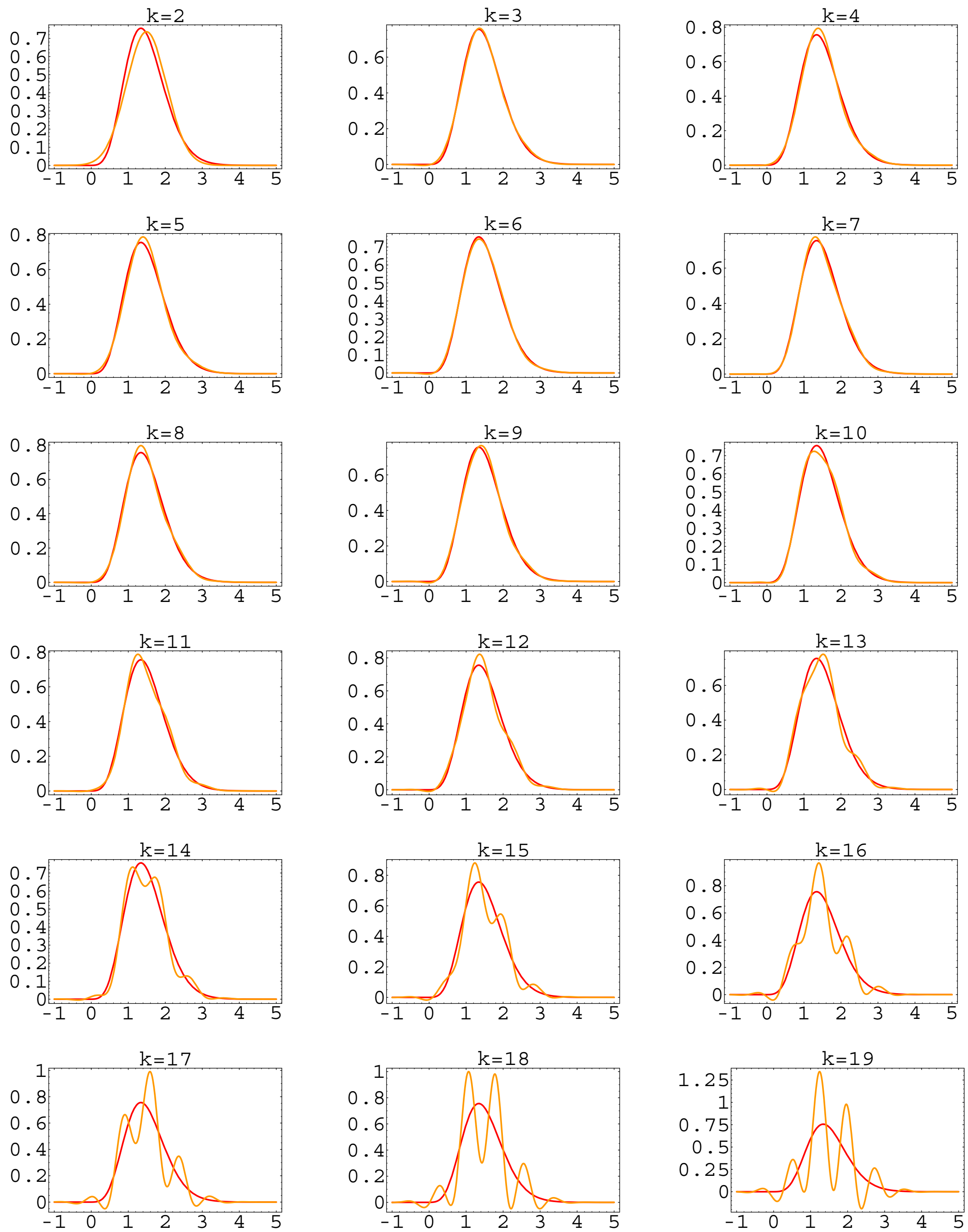
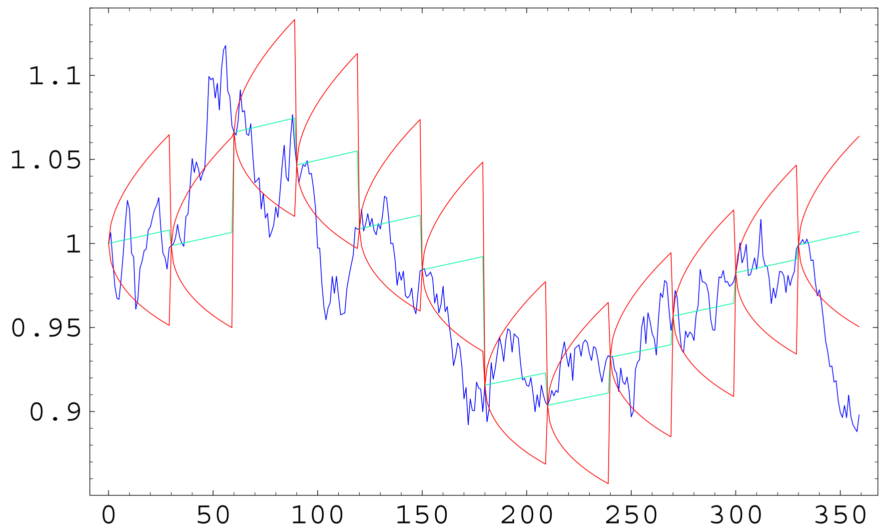


Figure 8: Square root model: simulated trajectory and approximate 67% prediction intervals $\mu(t|t_i) \pm \sqrt{m_2(t|t_i)}$ ($K = 3$).



6 SIMULATION STUDIES

The Hermite expansion approach was tested in simulation studies and compared with the Euler approach, the Nowman method (a simplified EKF; see 5.1) and the exact ML method using the Feller density. Weekly, monthly and quarterly observations of the square root model were generated on a daily basis, i.e. we chose a discretization interval of $\delta t = 1/365$ (year) and simulated daily series using the Euler-Maruyama scheme

$$y_{j+1} \quad = \quad y_j + f(y_j, t_j)dt + g(y_j, t_j)\delta W_j, \tag{79}$$

$\delta W_j, \sim N(0, \delta t)$ i.i.d., $j = 0, ...J$, $J = 3000$. The data were sampled weekly and monthly at times $j_i = (\Delta t/\delta t)i, i = 0, ...T$ with $\Delta t = 7/365, 30/365$ (year) and $j_i \leq J$. Thus the sampled series have length $T = \text{floor}(3000/7) = 428$ and $T = 3000/30 = 100$. In order to obtain a comparable sample size in the case of monthly measurements, also daily series of length $J = 12000$ with sampled length $T = 12000/30 = 400$ were simulated. The parameter values in the CEV model

$$dY(t) \quad = \quad rY(t)dt + \sigma Y(t)^{\alpha/2}dW(t). \tag{80}$$

are $\psi = \{r = .1, \alpha = 1, \sigma = .2\}$ corresponding to a square root model. The data were simulated using this true parameter vector, but in the estimation procedure no restrictions (such as $\alpha = 1$) were employed. Fig. 8 shows a simulated trajectory and approximate 67% prediction intervals $\mu(t|t_i) \pm \sqrt{m_2(t|t_i)}$ for 30 day measurements ($K = 3$). Actually, the transition density is skewed (fig. 7) and the Gaussian prediction interval is only approximate.

Table 1: Square root model: Means and standard deviations of ML estimates in $M = 100$ replications. Weekly measurements of daily series ($\delta t = 1/365$ year, $\Delta t = 7/365$ year).

weekly measurements: $\Delta t = 7/365$, $J = 3000$				
true values	mean	std	bias	RMSE
EKF $(K = 2, L = 1)$				
$r = 0.1$	0.0790961	0.0657432	-0.0209039	0.0689866
$\alpha = 1$	0.972559	0.333523	-0.0274412	0.334649
$\sigma = 0.2$	0.20078	0.0141277	0.00078025	0.0141492
SNF $(K = 2, L = 2)$				
$r = 0.1$	0.0790959	0.0657432	-0.0209041	0.0689866
$\alpha = 1$	0.972554	0.333539	-0.0274455	0.334666
$\sigma = 0.2$	0.200779	0.0141299	0.00077942	0.0141514
HNF $(K = 3, L = 3)$				
$r = 0.1$	0.0793155	0.0655923	-0.0206845	0.0687764
$\alpha = 1$	0.983154	0.343083	-0.0168457	0.343496
$\sigma = 0.2$	0.200648	0.0142035	0.000647821	0.0142183
Euler density				
$r = 0.1$	0.0791831	0.065808	-0.0208169	0.069022
$\alpha = 1$	0.972563	0.333517	-0.0274366	0.334644
$\sigma = 0.2$	0.200976	0.0141384	0.00097614	0.0141721
Nowman method				
$r = 0.1$	0.0790813	0.0657323	-0.0209187	0.0689806
$\alpha = 1$	0.972553	0.333525	-0.0274474	0.334652
$\sigma = 0.2$	0.200824	0.0141201	0.000823618	0.0141441
Exact density (Feller)				
$r = 0.1$	0.0801646	0.0659163	-0.0198354	0.0688361
$\alpha = 1$	0.960829	0.305208	-0.0391709	0.307711
$\sigma = 0.2$	0.201155	0.0138222	0.00115478	0.0138704

6.1 Weekly data

Table (1) displays the estimation results for weekly sampling interval $\Delta t = 7/365$ year. Comparing the estimation methods, the exact approch is best in terms of root mean square error RMSE $= \sqrt{\text{Bias}^2 + \text{Std}^2}$, Bias $:= \bar{\hat{\theta}} - \theta$, Std $= \sqrt{(M-1)^{-1}\sum_m(\hat{\theta}_m - \bar{\hat{\theta}})^2}$. Nowman's method (section 5.1), which approximates the moment equation for m_2, leads to slightly worse results than the EKF. The third order approximation HNF(3,3) shows small bias but somewhat larger standard errors. Also, the simple Euler estimator performs well. Generally, all methods show small bias and are comparable in terms of RMSE.

6.2 Monthly data

The bias of the approximation methods (deviations from conditional normal distribution) should show up for larger sampling interval. Indeed, using monthly data, table (2) shows, that the Euler method and other approximations (except HNF(3,3) have slight disadvantages in relation to the exact ML, in what regards bias. Again, the differences are not pronounced. Since we have only $T = 3000/30 = 100$ sampled observations, the simulation study was repeated using $T = 12000/30 = 400$ sampled observations, which is comparable to $3000/7 \approx 428$ in table (1). The results are shown in table 3. Again, exact ML is best in terms of RMSE (except for σ, where the HNF(3,3) and Nowman's method are better). The Euler method shows the worst results reflecting the large sampling interval. The HNF(3,3) dominates the Nowman method for all three parameters. It is surprising that EKF and SNF perform worse than Nowman's method. However, since α is not restricted to 1 in the estimation procedure, the EKF (SNF) variance equation is not exact, but given as

$$\dot{m}_2 \quad = \quad 2rm_2 + \sigma^2\mu^\alpha + \tfrac{1}{2}\sigma^2\alpha(\alpha-1)\mu^{\alpha-2}. \tag{81}$$

By contrast, Nowman's approximation is (cf. 45)

$$\dot{m}_2 \quad = \quad 2rm_2 + \sigma^2 Y_i^\alpha. \tag{82}$$

7 CONCLUSION

The transition density of a diffusion process was approximated as Hermite series and the expansion coefficients were expressed in terms of conditional moments. Taylor expansion of the drift and diffusion functions leads to a hierarchy of approximations indexed by the number of moments and the order of the Taylor series. The square root model, which is an important model for stock prices, was estimated using a CEV specification. Using weekly and monthly sampling intervals, the different approximation methods were comparable in performance to the exact ML method, but for large sampling intervals the simple Euler approximation has degraded performance in relation to the EKF type Gaussian likelihood and higher order skewed densities. For the chosen parameter values which are typical for stock prices, the differences are not very pronounced, however. Further studies will use higher order Hermite approximations and derive equations for the expansion coefficients of the direct Hermite series (67). Moreover, generalizations to the vector case will be derived.

Table 2: Square root model: Means and standard deviations of ML estimates in $M = 100$ replications. Monthly measurements of daily series ($\delta t = 1/365$ year, $\Delta t = 30/365$ year).

monthly measurements: $\Delta t = 30/365$, $J = 3000$				
true values	mean	std	bias	RMSE
EKF $(K = 2, L = 1)$				
$r = 0.1$	0.0777495	0.0679749	-0.0222505	0.0715239
$\alpha = 1$	0.926859	0.741427	-0.0731408	0.745026
$\sigma = 0.2$	0.198181	0.0277385	-0.00181901	0.0277981
SNF $(K = 2, L = 2)$				
$r = 0.1$	0.0777488	0.0679748	-0.0222512	0.071524
$\alpha = 1$	0.927102	0.740485	-0.0728976	0.744065
$\sigma = 0.2$	0.198153	0.0277338	-0.00184724	0.0277952
HNF $(K = 3, L = 3)$				
$r = 0.1$	0.0786614	0.0674743	-0.0213386	0.070768
$\alpha = 1$	0.974696	0.760442	-0.0253043	0.760863
$\sigma = 0.2$	0.197438	0.0292659	-0.00256184	0.0293778
Euler density				
$r = 0.1$	0.0781703	0.0682771	-0.0218297	0.0716819
$\alpha = 1$	0.926857	0.741369	-0.0731432	0.744969
$\sigma = 0.2$	0.199135	0.0277254	-0.000864986	0.0277389
Nowman method				
$r = 0.1$	0.077735	0.0679604	-0.022265	0.0715146
$\alpha = 1$	0.926825	0.741334	-0.0731752	0.744937
$\sigma = 0.2$	0.1985	0.0276465	-0.00150011	0.0276871
Exact density (Feller)				
$r = 0.1$	0.0772481	0.0674889	-0.0227519	0.0712207
$\alpha = 1$	0.951478	0.734323	-0.0485223	0.735924
$\sigma = 0.2$	0.198262	0.0283593	-0.00173828	0.0284125

Table 3: Square root model: Means and standard deviations of ML estimates in $M = 100$ replications. Monthly measurements of daily series ($\delta t = 1/365$ year, $\Delta t = 30/365$ year).

monthly measurements: $\Delta t = 30/365$, $J = 12000$				
true values	mean	std	bias	RMSE
EKF $(K = 2, L = 1)$				
$r = 0.1$	0.0925493	0.0221771	-0.00745071	0.0233952
$\alpha = 1$	0.99828	0.0777926	-0.00172012	0.0778117
$\sigma = 0.2$	0.199594	0.0130801	-0.000406332	0.0130864
SNF $(K = 2, L = 2)$				
$r = 0.1$	0.0925502	0.0221766	-0.00744983	0.0233944
$\alpha = 1$	0.998273	0.0778161	-0.00172736	0.0778353
$\sigma = 0.2$	0.199595	0.0130841	-0.000405306	0.0130904
HNF $(K = 3, L = 3)$				
$r = 0.1$	0.092583	0.0220525	-0.00741699	0.0232664
$\alpha = 1$	1.00076	0.0773048	0.000759309	0.0773085
$\sigma = 0.2$	0.199235	0.0130515	-0.000764711	0.0130739
Euler density				
$r = 0.1$	0.0929103	0.0222714	-0.00708968	0.0233726
$\alpha = 1$	0.99828	0.0777806	-0.0017197	0.0777996
$\sigma = 0.2$	0.200696	0.0131158	0.000695609	0.0131342
Nowman method				
$r = 0.1$	0.0925383	0.022167	-0.00746173	0.0233892
$\alpha = 1$	0.998279	0.0777901	-0.00172131	0.0778092
$\sigma = 0.2$	0.199934	0.0130753	-0.0000664348	0.0130754
Exact density (Feller)				
$r = 0.1$	0.0925661	0.0220893	-0.00743387	0.0233066
$\alpha = 1$	1.00056	0.0771624	0.000562986	0.0771644
$\sigma = 0.2$	0.199218	0.0130542	-0.000782308	0.0130777

References

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1965.

[2] Y. Aït-Sahalia. Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica*, 70,1:223–262, 2002.

[3] W.F. Ames. *Numerical Methods for Partial Differential Equations*. Academic Press, Boston, third edition, 1992.

[4] T.G. Andersen and J. Lund. Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, 77:343–377, 1997.

[5] L. Arnold. *Stochastic Differential Equations*. John Wiley, New York, 1974.

[6] A.R. Bergstrom. *Continuous Time Econometric Modelling*. Oxford University Press, Oxford, 1990.

[7] M. Bibby and M. Sorensen. Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1:1–39, 1995.

[8] R. Courant and D. Hilbert. *Methoden der Mathematischen Physik*. Springer, Berlin, Heidelberg, New York, 1968.

[9] J.C. Cox and S.A. Ross. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3:145–166, 1976.

[10] O. Elerian, S. Chib, and N. Shephard. Likelihood Inference for Discretely Observed Nonlinear Diffusions. *Econometrica*, 69, 4:959–993, 2001.

[11] W. Feller. Two singular diffusion problems. *Annals of Mathematics*, 54:173–182, 1951.

[12] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Transactions on Radar and Signal Procesing*, 140, 2:107–113, 1993.

[13] M. Hürzeler and H. Künsch. Monte Carlo Approximations for General State-Space Models. *Journal of Computational and Graphical Statistics*, 7,2:175–193, 1998.

[14] B. Jensen and R. Poulsen. Transition Densities of Diffusion Processes: Numerical Comparision of Approximation Techniques. *Institutional Investor*, Summer 2002:18–32, 2002.

[15] G. Kitagawa. Non-Gaussian state space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82:1032 – 1063, 1987.

[16] G. Kitagawa. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics*, 5,1:1–25, 1996.

[17] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.

[18] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. MIT press, Cambridge, Mass., 1983.

[19] A.W. Lo. Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data. *Econometric Theory*, 4:231–247, 1988.

[20] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, second edition, 1992.

[21] H. Risken. *The Fokker-Planck Equation*. Springer, Berlin, Heidelberg, New York, second edition, 1989.

[22] F. Schweppe. Evaluation of likelihood functions for gaussian signals. *IEEE Transactions on Information Theory*, 11:61–70, 1965.

[23] I. Shoji. Approximation of conditional moments. *Journal of Computational Mathematics*, 2:163–190, 2003.

[24] I. Shoji and T. Ozaki. Comparative Study of Estimation Methods for Continuous Time Stochastic Processes. *Journal of Time Series Analysis*, 18, 5:485–506, 1997.

[25] I. Shoji and T. Ozaki. A statistical method of estimation and simulation for systems of stochastic differential equations. *Biometrika*, 85, 1:240–243, 1998.

[26] H. Singer. Analytical score function for irregularly sampled continuous time stochastic processes with control variables and missing values. *Econometric Theory*, 11:721–735, 1995.

[27] H. Singer. Parameter Estimation of Nonlinear Stochastic Differential Equations: Simulated Maximum Likelihood vs. Extended Kalman Filter and Itô-Taylor Expansion. *Journal of Computational and Graphical Statistics*, 11,4:972–995, 2002.

[28] H. Singer. Simulated Maximum Likelihood in Nonlinear Continuous-Discrete State Space Models: Importance Sampling by Approximate Smoothing. *Computational Statistics*, 18,1:79–106, 2003.

[29] E. Wong and B. Hajek. *Stochastic Processes in Engineering Systems*. Springer, New York, 1985.

[30] Y. Yu and P.C.B. Phillips. A Gaussian approach for continuous time models of the short term interest rate. *Econometrics Journal*, 4:210–224, 2001.

Appendix: Derivation of the moment equations

The conditional density $p(yt|xs)$ fulfils the Fokker-Planck equation

$$\begin{aligned} \frac{\partial p(y,t|x,s)}{\partial t} &= -\sum_i \frac{\partial}{\partial y_i}[f_i(y,t)p(y,t|x,s)] \\ &\quad +\tfrac{1}{2}\sum_{ij} \frac{\partial^2}{\partial y_i \partial y_j}[\Omega_{ij}(y,t)p(y,t|x,s)] \\ &:= F(y,t)p(y,t|x,s) \end{aligned}$$

where F is the Fokker-Planck operator. Thus the first conditional moment $\mu(t|t_i) = E[Y(t)|Y^i]$ fulfils

$$\begin{aligned} \dot{\mu}(t|t_i) &= (\partial/\partial t)\int y p(y,t|y_i,t_i)dy \\ &= \int yFp(y,t|y_i,t_i)dy \\ &= \int (Ly)p(y,t|y_i,t_i)dy = E[(Ly)(Y(t))|Y^i] \end{aligned}$$

where $L = \sum_j f_j(y,t)\frac{\partial}{\partial y_j} + \frac{1}{2}\sum_{jk}\Omega_{jk}(y,t)\frac{\partial^2}{\partial y_j \partial y_k}$ is the backward operator. Thus we obtain

$$\dot{\mu}(t|t_i) = \int f(y,t)p(y,t|y_i,t_i)dy = E[f(Y,t)|Y^i].$$

Higher order moments

$$m_k := E[M_k] := E[(Y-\mu)^k].$$

fulfil the equations (scalar notation, condition suppressed)

$$\begin{aligned} \dot{m}_k &= (\partial/\partial t)\int (y-\mu)^k p(y,t)dy \\ &= -\int k(y-\mu)^{k-1}\dot{\mu}p(y,t)dy + \int (L(y-\mu)^k)p(y,t)dy \\ &= -kE[(Y-\mu)^{k-1}]E[f(Y)] + kE[f(Y)*(Y-\mu)^{k-1}] \\ &\quad +\tfrac{1}{2}k(k-1)E[\Omega(Y)(Y-\mu)^{k-2}] \\ &= kE[f(Y)*(M_{k-1}-m_{k-1})] + \tfrac{1}{2}k(k-1)E[\Omega(Y)*M_{k-2}] \end{aligned}$$