

Moment equations and Hermite expansion for nonlinear stochastic differential equations with application to stock price models

Hermann Singer

Published online: 9 August 2006
© Physica-Verlag 2006

Abstract 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, in *Ann Math* 54: 173–182, 1951).

Keywords 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 for the dynamics is given. Examples are mechanical systems (Newton's equations) or stock price movements where no natural trading interval can be identified. 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

H. Singer (✉)
FernUniversität in Hagen, Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung, 58084 Hagen, Germany
e-mail: hermann.singer@fernuni-hagen.de

the measurement times (time series models). In contrast, we consider stochastic differential equations (SDE) for the state $Y(t)$, but 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 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 (Pedersen 1995; 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). Using the first and second conditional moments, Bibby and Sorensen (1995) discuss martingale estimators which emerge from corrections to the discretized likelihood.

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 Sect. 2 the basic model and the maximum likelihood method is defined. Section 3 introduces the Hermite expansion used to approximate the transition density and the moment equations are derived in Sect. 4. In Sect. 5, 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 Model and likelihood

We discuss the nonlinear stochastic differential equation (SDE)

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

where discrete measurements Y_i are taken at times $\{t_0, t_1, \dots, t_T\}$, $t_0 \leq t \leq t_T$, $Y_i := Y(t_i)$. 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 $f : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \rightarrow \mathbb{R}^p$ and $g : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \rightarrow \mathbb{R}^p \times \mathbb{R}^r$ 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 (see appendix). Extensions to nonlinear noisy measurements are given in Gordon et al. (1993), Kitagawa (1987, 1996), Hürzeler and Künsch (1998) and Singer (2003). Using the Markov property of $Y(t)$ the likelihood is given by

$$L_\psi(y) := p(y_T, \dots, y_1|y_0; \psi) = \prod_{i=0}^{T-1} p(y_{i+1}|y_i; \psi). \tag{2}$$

The transition density $p(y_{i+1}|y_i; \psi)$ can be computed analytically only in the linear or some special cases [e.g. constant elasticity of variance (CEV) diffusion process; cf. Feller 1951; Cox and Ross 1976]. In the general case, approximate numerical procedures must be employed (cf. the references in Sect. 1). Conditional Gaussian approximations work well when the sampling intervals $\Delta t_i = t_{i+1} - t_i$ are not too large in comparison 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.

3 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.,

$$\int_{-\infty}^{\infty} H_n(x)H_m(x)w(x)dx = n!\delta_{nm} \tag{3}$$

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

$$\phi^{(n)}(x) := (d/dx)^n \phi(x) = (-1)^n \phi(x)H_n(x). \tag{4}$$

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

$$p(x) = \phi(x) \sum_{n=0}^{\infty} c_n H_n(x). \tag{5}$$

and the Fourier coefficients are given by

$$c_n := (1/n!) \int_{-\infty}^{\infty} H_n(x)p(x)dx = (1/n!)E[H_n(X)] \tag{6}$$

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. $\mu_k = E[X^k]$. Using the standardized variables $Z = (X - \mu)/\sigma$ with $\mu = E[X], \sigma^2 = E[X^2] - \mu^2, E[Z] = 0, E[Z^2] = 1, E[Z^k] := v_k$ one obtains the simplified expressions $c_0 = 1, c_1 = 0, c_2 = 0$,

$$c_3 := (1/3!)E[Z^3] = (1/3!)v_3 \tag{7}$$

$$c_4 := (1/4!)E[Z^4 - 6Z^2 + 3] = (1/24)(v_4 - 3) \tag{8}$$

and the density expansion

$$p_z(z) := \phi(z)[1 + (1/6)v_3H_3(z) + (1/24)(v_4 - 3)H_4(z) + \dots] \tag{9}$$

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 obtains the desired Hermite expansion for $p_x(x)$. The standardized moments $v_k = E[Z^k] = E[(X - \mu)^k]/\sigma^k := m_k/\sigma^k$ can be expressed in terms of *centered moments*

$$m_k := E[M_k] := E[(X - \mu)^k]. \tag{10}$$

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

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., $\int q(x)^2 dx = \int \exp(x^2/2)p^2(x)dx < \infty$ (Courant and Hilbert, 1968, p. 81–82).

4 Scalar moment equations

4.1 Differential equations for the moments

The first conditional moment $E[Y(t)|Y^i]$ fulfils the exact equation

$$\dot{\mu}(t|t_i) = E[f(Y, t)|Y^i]. \tag{11}$$

The time dependence of higher order conditional moments

$$m_k(t|t_i) := E[M_k(t|t_i)|Y^i] := E[(Y(t) - \mu(t|t_i))^k|Y^i]. \tag{12}$$

is (see appendix)

$$\dot{m}_k = kE[f(Y) * (M_{k-1} - m_{k-1})] + \frac{1}{2}k(k-1)E[\Omega(Y) * M_{k-2}] \tag{13}$$

with initial condition $m_k(t_i|t_i) = 0$. These are not differential equations, and Taylor expansion of f and Ω around μ and inserting this into (11, 13) yields

$$\dot{\mu} := \sum_{l=0}^{\infty} f^{(l)}(\mu) \frac{m_l}{l!} \tag{14}$$

$$\begin{aligned} \dot{m}_k &= k \sum_{l=1}^{\infty} \frac{f^{(l)}(\mu)}{l!} (m_{l+k-1} - m_l m_{k-1}) \\ &+ \frac{1}{2}k(k-1) \sum_{l=0}^{\infty} \frac{\Omega^{(l)}(\mu)}{l!} m_{l+k-2} \quad (k \geq 2). \end{aligned} \tag{15}$$

For practical applications, three problems must be solved:

1. One must choose a number K of moments to consider.
2. The expansion of f and Ω must be truncated somewhere ($l = 0, \dots, L$).
3. On the right hand side moments of maximum 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 can be used:
 - (a) Higher order moments are neglected: $m_k = 0; k > K$
 - (b) Higher order moments are factorized by the Gaussian assumption

$$m_k = \begin{cases} (k-1)!!m_2^{k/2}; & k > K \text{ is even} \\ 0 & k > K \text{ is odd} \end{cases} \tag{16}$$

If we use only $K = 2$ moments and truncate at $L = 1$, the time update of the extended Kalman filter (EKF) and the second order nonlinear filter (SNF; $L = 2$) is obtained. In analogy to EKF and SNF, the abbreviation HNF(K, L) (higher order nonlinear filter) will be used.

4.2 Example: square root stock price model

The square root stock price model (cf. Feller 1951; Cox and Ross 1976)

$$dY(t) = rY(t)dt + \sigma Y(t)^{1/2}dW(t). \tag{17}$$

has a linear drift and diffusion term $\Omega(y) = \sigma^2 y$ with derivatives $\Omega'(y) = \sigma^2$, $\Omega^{(l)}(y) = 0, l \geq 2$ and yields a closed system (see, e.g. Bibby and Sorensen 1995). The first and second moments fulfil

$$\begin{aligned} \dot{\mu} &= r\mu & (18) \\ \dot{m}_2 &= 2r m_2 + \sigma^2 \mu & (19) \end{aligned}$$

with explicit solution

$$\begin{aligned} \mu(t|t_i) &= \exp[r(t - t_i)]Y_i & (20) \\ m_2(t|t_i) &= \frac{\sigma^2}{r} [\exp(2r(t - t_i)) - \exp(r(t - t_i))]Y_i. & (21) \end{aligned}$$

The equation for the third moment

$$\dot{m}_3 = 3r m_3 + 3\sigma^2 m_2 \tag{22}$$

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

$$\dot{m}_4 = 4r m_4 + 6\sigma^2(\mu m_2 + m_3) \tag{23}$$

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 $\psi = \{r, \sigma\} = \{0.1, 0.2\}$ the moment equations were solved by an Euler scheme with discretization interval $\delta t = 1/250$ year and $T = 1,000$ time steps corresponding to $T\delta t = 4$ years and initial condition $m(t_i|t_i) = [1, 0, 0, 0]'$. Figure 1 demonstrates the improvement of the EKF density over the Euler approximation. The square root model (17) can be solved exactly using Bessel functions (Feller 1951) and the moments m_k were computed numerically from this exact density. Inspection of the function $q^2(z) = p_z(z)^2/\phi(z)$, where $p_z(z) = p_y(\mu + \sigma z)\sigma$ is the standardized density function, reveals that the convergence condition (footnote 1) is not fulfilled. Expanding up to order $k = 21$, the nonconvergence is shown in Fig. 2. However, low order approximations such as $k = 3, 6$ are nevertheless quite good. Alternatively, one could transform the process such that the diffusion coefficient is constant. It can be shown that the resulting transition density $p_y(y)$ is sufficiently close to a normal density (Ait-Sahalia, loc. cit., prop. 2), so that a convergent Hermite expansion is possible. In the multivariate case, such a transformation is difficult, so the idea is to study low order approximants of the nonconvergent Hermite series.

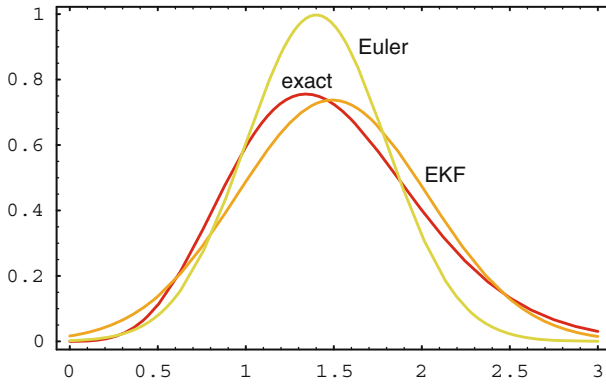


Fig. 1 Square root model: exact density (red), approximate density $p_{2,1}(y)$ with Hermite expansion up to $K = 2$ (EKF, orange) and Euler density (green)

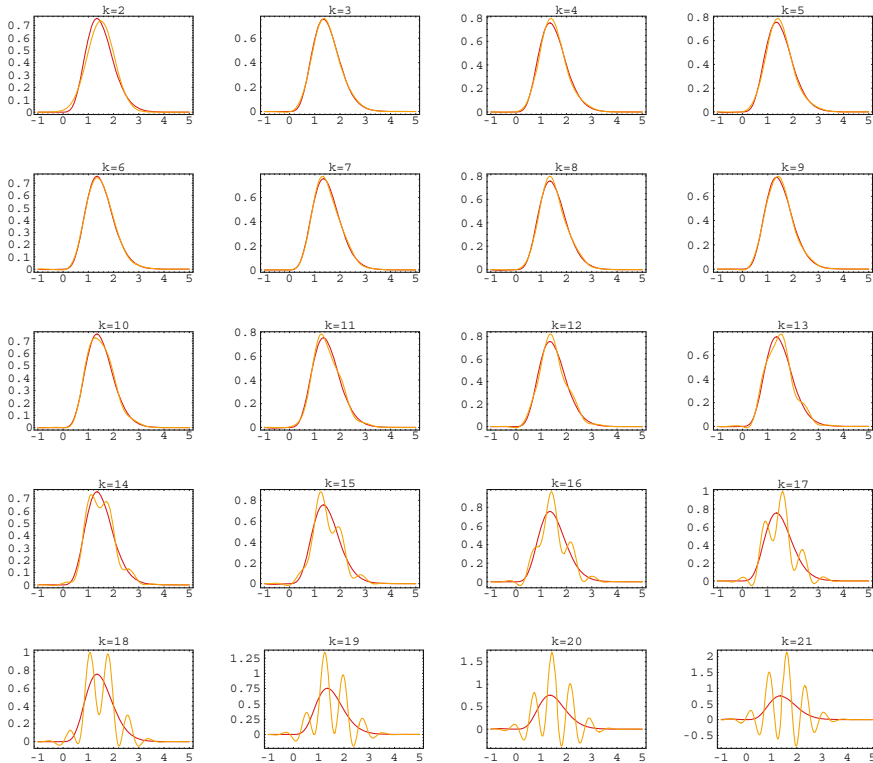


Fig. 2 Square root model: Approximate densities $p_k(y)$ with Hermite expansion up to $k = 21$ (orange) and exact density (red). The moments were computed from the exact density function

5 Simulation studies

The Hermite expansion approach was tested in simulation studies and compared with the Euler approach, the Nowman method² and the exact ML method using the Feller density. Weekly and monthly 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

$$y_{j+1} = y_j + f(y_j, t_j)\delta t + g(y_j, t_j)\delta W_j, \quad (24)$$

$\delta W_j, \sim N(0, \delta t)$ i.i.d., $j = 0, \dots, J$. The data were sampled weekly and monthly at times $j_i = (\Delta t/\delta t)i, i = 0, \dots, T = 500$ with $\Delta t = 7/365, 30/365$ (year). Thus for the sampled data of length $T = 500, J = 500 \times 7 = 3,500$ and $J = 500 \times 30 = 15,000$ daily data were simulated. The parameter values in the CEV model

$$dY(t) = rY(t)dt + \sigma Y(t)^{\alpha/2}dW(t) \quad (25)$$

are $\psi = \{r = 0.1, \alpha = 1, \sigma = 0.2\}$ corresponding to a square root model. Also, scenarios with high drift $r = 0.2$ and high volatility $\sigma = 0.4$ were considered. The data were simulated using these true parameter vectors, but in the estimation procedure no restrictions (such as $\alpha = 1$) were employed.

5.1 Weekly data

Table 1 displays the estimation results for weekly sampling interval $\Delta t = 7/365$ year. Comparing the estimation methods, the exact ML approach is best in terms of root mean square error $\text{RMSE} = \sqrt{\text{Bias}^2 + \text{Std}^2}$, $\text{Bias} := \bar{\hat{\theta}} - \theta$, $\text{Std} = \sqrt{(M-1)^{-1} \sum_m (\hat{\theta}_m - \bar{\hat{\theta}})^2}$, except for r , where the HNF(4,4) is better. The third order approximation HNF(3,3) is only slightly worse than the exact approach and HNF(4,4).

Nowman's method, which approximates the moment equation for m_2 with constant diffusion term, leads to slightly better results than the EKF. Also, the simple Euler estimator performs well.

Generally, for weekly sampling, all methods show small bias and are comparable in terms of RMSE.

5.2 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) and HNF(4,4)] have slight disadvantages in relation to exact ML.

² Here a simplified SDE with approximate diffusion coefficient $g(y) = g(y_i)$ is used (cf. Yu and Phillips 2001).

Table 1 Square root model: means and standard deviations of ML estimates in $M = 500$ replications

Weekly measurements: $\Delta t = 7/365, T = 500, J = 3500$				
True values	Mean	Std	Bias	RMSE
Exact density (Feller)				
$r = 0.1$	0.0871301	0.066783	-0.0128699	0.0680118
$\alpha = 1$	0.98403	0.234105	-0.0159705	0.234649
$\sigma = 0.2$	0.200053	0.0115526	0.0000532826	0.0115527
Euler density				
$r = 0.1$	0.0864449	0.0669437	-0.0135551	0.0683023
$\alpha = 1$	0.989414	0.248798	-0.0105862	0.249023
$\sigma = 0.2$	0.200244	0.0118366	0.000243827	0.0118391
Nowman method				
$r = 0.1$	0.0863305	0.0669343	-0.0136695	0.0683159
$\alpha = 1$	0.989415	0.248799	-0.0105848	0.249024
$\sigma = 0.2$	0.200078	0.0118301	0.0000783523	0.0118304
EKF ($K = 2, L = 1$)				
$r = 0.1$	0.0861456	0.0670187	-0.0138544	0.0684357
$\alpha = 1$	0.987432	0.25246	-0.012568	0.252773
$\sigma = 0.2$	0.200027	0.01183	0.0000268234	0.01183
SNF ($K = 2, L = 2$)				
$r = 0.1$	0.0863455	0.0669369	-0.0136545	0.0683154
$\alpha = 1$	0.98941	0.248831	-0.0105904	0.249057
$\sigma = 0.2$	0.200031	0.0118434	0.0000306548	0.0118434
HNF ($K = 3, L = 3$)				
$r = 0.1$	0.0865024	0.0667861	-0.0134976	0.0681364
$\alpha = 1$	0.996269	0.248761	-0.00373099	0.248789
$\sigma = 0.2$	0.199797	0.0117871	-0.00020323	0.0117888
HNF ($K = 4, L = 4$)				
$r = 0.1$	0.0882339	0.0635362	-0.0117661	0.0646164
$\alpha = 1$	1.0047	0.244924	0.00469891	0.244969
$\sigma = 0.2$	0.199093	0.0121452	-0.000907132	0.012179

Weekly measurements of daily series ($\delta t = 1/365$ year, $\Delta t = 7/365$ year)

Again, the differences are not very pronounced. Surprisingly, Euler is best for parameter r , but with higher RMSE for the other parameters. In this case, all 500 samples converged except for exact ML, where 15 samples run into a step halving with no higher likelihood value (cf. Dennis and Schnabel 1983, for convergence issues).

5.3 Other parameter constellations

The approximation methods were also tested using a scenario with higher drift, i.e. $\psi = \{0.2, 1, 0.2\}$ and higher volatility $\psi = \{0.1, 1, 0.4\}$.

5.3.1 High drift

For high drift $r = 0.2$, the exact ML is best for weekly and monthly sampling interval. The Euler method displays higher RMSE and the other methods perform roughly the same.

Table 2 Square root model: means and standard deviations of ML estimates in $M = 500$ replications

Monthly measurements: $\Delta t = 30/365$, $T = 500$, $J = 15,000$				
True values	Mean	Std	Bias	RMSE
Exact density (Feller)				
$r = 0.1$	0.0920491	0.0402635	-0.0079509	0.0410411
$\alpha = 1$	0.999559	0.0624444	-0.000440569	0.0624459
$\sigma = 0.2$	0.199905	0.0122831	-0.0000954044	0.0122834
Euler density				
$r = 0.1$	0.0928074	0.039442	-0.00719255	0.0400925
$\alpha = 1$	0.998668	0.0625336	-0.00133205	0.0625478
$\sigma = 0.2$	0.201168	0.0124711	0.00116806	0.0125257
Nowman method				
$r = 0.1$	0.0923911	0.0397238	-0.0076089	0.040446
$\alpha = 1$	0.998668	0.0625347	-0.00133228	0.0625489
$\sigma = 0.2$	0.200402	0.0123862	0.000402193	0.0123928
EKF ($K = 2$, $L = 1$)				
$r = 0.1$	0.0924049	0.0397147	-0.00759511	0.0404344
$\alpha = 1$	0.998667	0.0625338	-0.00133316	0.0625481
$\sigma = 0.2$	0.200061	0.0123691	0.0000611214	0.0123692
SNF ($K = 2$, $L = 2$)				
$r = 0.1$	0.0924057	0.0397026	-0.00759426	0.0404224
$\alpha = 1$	0.998661	0.0625818	-0.00133926	0.0625961
$\sigma = 0.2$	0.200061	0.0123727	0.0000607948	0.0123729
HNF ($K = 3$, $L = 3$)				
$r = 0.1$	0.0924237	0.0396978	-0.00757626	0.0404143
$\alpha = 1$	1.00004	0.0620625	0.0000447874	0.0620625
$\sigma = 0.2$	0.199829	0.0122826	-0.000171326	0.0122838
HNF ($K = 4$, $L = 4$)				
$r = 0.1$	0.0924242	0.0394029	-0.0075758	0.0401245
$\alpha = 1$	0.999801	0.0618515	-0.000199132	0.0618518
$\sigma = 0.2$	0.199937	0.0123067	-0.0000625403	0.0123069

Monthly measurements of daily series ($\delta t = 1/365$ year, $\Delta t = 30/365$ year)

5.3.2 High volatility

In the high volatility scenario $\sigma = 0.4$, a considerable number of samples did not converge (neither $||\text{score}||$ nor $|\text{step}|$ were smaller than $\epsilon = 0.01$). Since the percentages (up to 25%) depend on the method, the results are difficult to compare and are not reported.

5.3.3 Small sample properties

The normal scenario ($\psi = \{0.1, 1, 0.2\}$) was also tested for $T = 50$ sampled data (i.e., 50 weeks or months). In this case the estimates for $\alpha = 1$ were strongly biased and the convergence rates were different for each method. No clear picture emerged which method should be preferred.

6 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 K of moments and the order L 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 exact ML method is best, followed by HNF(4,4), HNF(3,3) and the other approximation methods except Euler. 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.

The practical implications are as follows: In cases where the exact density is unknown, use the HNF(4,4), HNF(3,3) or at least the EKF, SNF or Nowman. The simple Euler estimator is fast but has degraded performance.

Further studies will use higher order Hermite approximations and derive equations for the expansion coefficients of a convergent Hermite series using oscillator eigenfunctions. Moreover, generalizations to the vector case will be derived.

Appendix: derivation of the moment equations

The conditional density $p(y|x,s)$ fulfils the Fokker–Planck equation

$$\begin{aligned} \frac{\partial p(y, t|x, s)}{\partial t} &= - \sum_i \frac{\partial}{\partial y_i} [f_i(y, t)p(y, t|x, s)] \\ &\quad + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial y_i \partial y_j} [\Omega_{ij}(y, t)p(y, t|x, s)] \\ &:= F(y, t)p(y, t|x, s) \end{aligned} \tag{26}$$

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

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

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)

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

References

- Abramowitz M, Stegun I (1965) Handbook of mathematical functions. Dover, New York
- Aït-Sahalia Y (2002) Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica* 70(1):223–262
- Andersen T, Lund J (1997) Estimating continuous-time stochastic volatility models of the short-term interest rate. *J Econom* 77:343–377
- Arnold L (1974) Stochastic differential equations. Wiley, New York
- Bergstrom A (1990) Continuous time econometric modelling. Oxford University Press, Oxford
- Bibby M, Sorensen M (1995) Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* 1:1–39
- Courant R, Hilbert D (1968) Methoden der mathematischen physik. Springer, Berlin Heidelberg New York
- Cox J, Ross S (1976) The valuation of options for alternative stochastic processes. *J Financ Econom* 3:145–166
- Dennis Jr. J, Schnabel R (1983) Numerical methods for unconstrained optimization and nonlinear equations. Prentice Hall, Englewood Cliffs
- Elerian O, Chib S, Shephard N (2001) Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* 69(4):959–993
- Feller W (1951) Two singular diffusion problems. *Ann Math* 54:173–182
- Gordon N, Salmond D, Smith A (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Trans Radar Signal Process* 140(2):107–113
- Hürzeler M, Künsch H (1998) Monte carlo approximations for general state-space models. *J Comput Graph Stat* 7(2):175–193
- Jensen B, Poulsen R (2002) Transition densities of diffusion processes: numerical comparison of approximation techniques. *Inst Invest* 18–32
- Kitagawa G (1987) Non-gaussian state space modeling of nonstationary time series. *J Am Stat Assoc* 82:1032–1063
- Kitagawa G (1996) Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J Comput Graph Stat* 5(1):1–25
- Pedersen A (1995) A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand J Stat* 22:55–71
- Shoji I (2003) Approximation of conditional moments. *J Comput Math* 2:163–190
- Shoji I, Ozaki T (1997) Comparative study of estimation methods for continuous time stochastic processes. *J Time Series Anal* 18(5):485–506
- Shoji I, Ozaki T (1998) A statistical method of estimation and simulation for systems of stochastic differential equations. *Biometrika* 85(1):240–243

- Singer H (1995) Analytical score function for irregularly sampled continuous time stochastic processes with control variables and missing values. *Econ Theory* 11:721–735
- Singer H (2002) Parameter estimation of nonlinear stochastic differential equations: simulated maximum likelihood vs. extended Kalman filter and Itô-Taylor expansion. *J Comput Graph Stat* 11(4):972–995
- Singer H (2003) Simulated maximum likelihood in nonlinear continuous-discrete state space models: importance sampling by approximate smoothing. *Comput Stat* 18(1):79–106
- Yu Y, Phillips P (2001) A Gaussian approach for continuous time models of the short term interest rate. *Econ J* 4:210–224