CHAPTER 17

Likelihood Inference for Diffusions: A Survey

Yacine Aït-Sahalia
Department of Economics and Bendheim Center for Finance
Princeton University and NBER
Princeton, NJ 08544
yacine@princeton.edu

This chapter reviews closed form expansions for discretely sampled diffusions, and their use for likelihood inference and testing. The method is applicable to a large class of models, covering both univariate and multivariate processes, and with a state vector that is either fully or partially observed. Examples are included.

17.1. Introduction

This chapter surveys recent results on closed form likelihood expansions for discretely sampled diffusions. The basic model is written in the form of a stochastic differential equation for the state vector X

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \tag{17.1}$$

where W_t is an $m-$dimensional standard Brownian motion. In the parametric case, the functions μ and σ are known, but not the parameter vector θ which is the object of interest. Available data are discrete observations on the process sampled at dates Δ, $2\Delta, ..., N\Delta$, with Δ fixed. The case where Δ is either deterministic and time-varying or random (as long as independent from X) introduces no further difficulties.

One major impediment to both theoretical modeling and empirical work with continuous-time models is the fact that in most cases little can be said about the implications of the instantaneous dynamics (17.1) for X_t for longer time intervals Δ. One cannot in general characterize in closed form an object as simple, yet fundamental for everything from prediction to estimation and derivative pricing, as the conditional density of $X_{t+\Delta}$

given the current value X_t, also known as the transition function of the process. For a list of the rare exceptions, see Wong (1964). In finance, the well-known models of Black and Scholes (1973) (the geometric Brownian motion $dX_t = \beta X_t dt + \sigma X_t dW_t$), Vasicek (1977) (the Ornstein-Uhlenbeck process $dX_t = \beta (\alpha - X_t) dt + \sigma dW_t$) and Cox, Ingersoll, and Ross (1985) (Feller's square root process $dX_t = \beta (\alpha - X_t) dt + \sigma X_t^{1/2} dW_t$) rely on these existing closed-form expressions.

In many cases that are relevant in finance, however, the transition function is unknown: see for example the models used in Courtadon (1982) ($dX_t = \beta (\alpha - X_t) dt + \sigma X_t dW_t$), Marsh and Rosenfeld (1982) ($dX_t = (\alpha X_t^{-(1-\delta)} + \beta) dt + \sigma X_t^{\delta/2} dW_t$), Cox (1975) and the more general version of Chan, Karolyi, Longstaff, and Sanders (1992) ($dX_t = \beta (\alpha - X_t) dt + \sigma X_t^{\gamma} dW_t$), Constantinides (1992) ($dX_t = \left(\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2\right) dt + (\sigma_0 + \sigma_1 X_t) dW_t$), the affine class of models in Duffie and Kan (1996) and Dai and Singleton (2000) ($dX_t = \beta (\alpha - X_t) dt + \sqrt{\sigma_0 + \sigma_1 X_t} dW_t$), the nonlinear mean reversion model in Aït-Sahalia (1996) ($dX_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_{-1}/X_t) dt + (\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}) dW_t$). While it is possible to write down the continuous-time likelihood function for the full sample sample path, ignoring the difference between sampling at a fixed time interval Δ and seeing the full sample path can lead to inconsistent estimators of the parameter vector θ.

In Aït-Sahalia (1999) (examples and application to interest rate data), Aït-Sahalia (2002) (univariate theory) and Aït-Sahalia (2001) (multivariate theory), I developed a method which produces accurate approximations *in closed form* to the unknown transition function $p_X(x|x_0, \Delta; \theta)$, that is, the conditional density that $X_{n\Delta} = x$ given $X_{(n-1)\Delta} = x_0$ in an amount of time Δ implied by the model in equation (17.1).

Bayes' rule combined with the Markovian nature of the process, which the discrete data inherit, imply that the log-likelihood function has the simple form

$$\ell_N(\theta) \equiv \sum_{i=1}^{N} l_X\left(X_{i\Delta} | X_{(i-1)\Delta}, \Delta; \theta\right) \qquad (17.2)$$

where $l_X \equiv \ln p_X$, and the asymptotically irrelevant density of the initial observation, X_0, has been left out. As is clear from (17.2), the availability of tractable formulae for p_X is what makes likelihood inference feasible under these conditions.

The rest of this paper is devoted to reviewing these methods and their applications. I start with the univariate case in Section 17.2, then move on to the multivariate case in Section 17.3. Section 17.4 shows a connection between this method and saddlepoint approximations. I then provide two examples, one of a nonlinear univariate model, and one of a multivariate model, in Sections 17.5 and 17.6 respectively. Section 17.7 discusses inference using this method when the state vector is only partially observed, as in stochastic volatility or term structure models. Section 17.8 outlines the use of this method in specification testing while Section 17.9 sketches derivative pricing applications. Finally, Section 17.10 discusses likelihood inference for continuous time models when the underlying process is nonstationary.

17.2. The Univariate Case

Existing methods to derive MLE for discretely sampled diffusions required solving numerically the Fokker-Planck-Kolmogorov partial differential equation satisfied by p_X (see e.g., Lo (1988)), or simulating a large number of sample paths along which the process is sampled very finely (Pedersen (1995), Brandt and Santa-Clara (2002)). Neither methods produce a closed-form expression, so they both result in a large computational effort since the likelihood must be recomputed for each observed realization of the state vector, and each value of the parameter vector θ along the maximization. Both methods deliver a sequence of approximations to $\ell_N(\theta)$ which become increasingly accurate as some control parameter J tends to infinity.

By contrast, the closed form likelihood expressions that I will describe here make MLE a feasible choice for estimating θ in practical applications. The method involves no simulations and no PDE to solve numerically. Like these two methods, I also construct a sequence $\ell_N^{(J)}$ for $J = 1, 2, ...$ of approximations to ℓ_N, but the essential difference is that $\ell_N^{(J)}$ will be obtained in closed-form. It converges to ℓ_N as $J \to \infty$ and maximizing $\ell_N^{(J)}$ in lieu of the true but incomputable ℓ_N results in an estimator which converges to the true MLE. Since $\ell_N^{(J)}$ is explicit, the effort involved is minimal.

17.2.1. *The density approximation sequence*

To understand the construction of the sequence of approximations to p_X, the following analogy may be helpful. Consider a standardized sum of random variables to which the Central Limit Theorem (CLT) applies. Often, one is willing to approximate the actual sample size n by infinity and use the $N(0,1)$ limiting distribution for the properly standardized transformation of the data. If not, higher order terms of the limiting distribution (for example the classical Edgeworth expansion based on Hermite polynomials) can be calculated to improve the small sample performance of the approximation. Consider now approximating the transition density of a diffusion, and think of the sampling interval Δ as playing the role of the sample size N in the CLT. If we properly standardize the data, then we can find out the limiting distribution of the standardized data as Δ tends to 0 (by analogy with what happens in the CLT when N tends to ∞). Properly standardizing the data in the CLT means summing them and dividing by $N^{1/2}$; here it involves transforming the original diffusion X into another one, called Z below. In both cases, the appropriate standardization makes $N(0,1)$ the leading term of the approximation. This $N(0,1)$ approximation is then refined by "correcting" for the fact that Δ is not 0 (just like in practical applications of the CLT N is not infinity). As in the CLT case, it is natural to consider higher order terms based on Hermite polynomials, which are orthogonal with respect to the leading $N(0,1)$ term.

This is not a standard Edgeworth expansion, however: we want convergence as $J \to\infty$, not $N \to\infty$. Further, in general, p_X cannot be approximated for fixed Δ around a Normal density by standard series because the distribution of X is too far from that of a Normal: for instance, if X follows a geometric Brownian motion, the right tail of p_X is too thick, and the Edgeworth expansion diverges as $J \to\infty$. Therefore there is a need for a transformation of X that standardizes the tails of its distribution.

Since Z is a known transformation of X, one can then revert the transformation from X to Z and obtain an expansion for the density of X. As a result of transforming Z back into X, which in general is a nonlinear transformation (unless $\sigma(x;\theta)$ is independent of the state variable x), the leading term of the expansion for the transition function of X will be a *deformed*, or stretched, normal density rather than the $N(0,1)$ leading term of the expansion for p_Z.

The first step towards constructing the sequence of approximations to p_X consists in standardizing the diffusion function of X, i.e., transforming X into Y defined as

$$Y_t \equiv \gamma\left(X_t;\theta\right) = \int^{X_t} \frac{du}{\sigma\left(u;\theta\right)} \tag{17.3}$$

so that

$$dY_t = \mu_Y\left(Y_t;\theta\right) dt + dW_t \tag{17.4}$$

where

$$\mu_Y\left(y;\theta\right)=\frac{\mu\left(\gamma^{-1}\left(y;\theta\right);\theta\right)}{\sigma\left(\gamma^{-1}\left(y;\theta\right);\theta\right)}-\frac{1}{2}\frac{\partial\sigma}{\partial x}\left(\gamma^{-1}\left(y;\theta\right);\theta\right). \tag{17.5}$$

Let p_Y denote the transition function of Y. The tails of p_Y have a Gaussian-like upper bound, so Y is "closer" to a Normal variable than X is. But it is still not practical to expand p_Y. This is due to the fact that p_Y gets peaked around the conditioning value y_0 when Δ gets small. And a Dirac mass is not a particularly appealing leading term for an expansion. For that reason, a further transformation is performed, defining the "pseudo-normalized" increment of Y as

$$Z_t\equiv\Delta^{-1/2}\left(Y_t-y_0\right). \tag{17.6}$$

Given the density of Y, we can work back to the density of X by applying the Jacobian formula:

$$p_X\left(x|x_0,\Delta;\theta\right)=\frac{p_Y\left(\gamma\left(x;\theta\right)|\gamma\left(x_0;\theta\right),\Delta;\theta\right)}{\sigma\left(\gamma\left(x;\theta\right);\theta\right)} \tag{17.7}$$

where p_Y can itself be deduced from the density p_Z of Z :

$$p_Y\left(y|y_0,\Delta;\theta\right)=\Delta^{-1/2}p_Z\left(\Delta^{-1/2}\left(y-y_0\right)\Big|y_0,\Delta;\theta\right). \tag{17.8}$$

So this leaves us with the need to approximate the density function p_Z. Consider a Hermite series expansion for the conditional density of the variable Z_t, which has been constructed precisely so that it be close enough to a $N(0,1)$ variable for an expansion around a $N(0,1)$ density to converge. The classical Hermite polynomials are

$$H_j\left(z\right)\equiv e^{z^2/2}\frac{d^j}{dz^j}\left[e^{-z^2/2}\right],\quad j\geq 0, \tag{17.9}$$

and let $\phi\left(z\right)\equiv e^{-z^2/2}/\sqrt{2\pi}$ denote the $N(0,1)$ density function. Also, define

$$p_Z^{(J)}\left(z|y_0,\Delta;\theta\right)\equiv\phi\left(z\right)\sum_{j=0}^{J}\eta_j\left(\Delta,y_0;\theta\right)\,H_j\left(z\right) \tag{17.10}$$

as the Hermite expansion of the density function $z\mapsto p_Z\left(z|y_0,\Delta;\theta\right)$ (for fixed Δ, y_0 and θ). The coefficients $\eta_Z^{(j)}$ are given by:

$$
\begin{aligned}
\eta_Z^{(j)}(\Delta, y_0;\theta) &= (1/j!)\int_{-\infty}^{+\infty} H_j(z)\, p_Z(z|y_0,\Delta;\theta)\, dz \\
&= (1/j!)\int_{-\infty}^{+\infty} H_j(z)\, \Delta^{1/2} p_Y\left(\Delta^{1/2}z + y_0 \middle| y_0,\Delta;\theta\right) dz \\
&= (1/j!)\int_{-\infty}^{+\infty} H_j\left(\Delta^{-1/2}(y-y_0)\right) p_Y(y|y_0,\Delta;\theta)\, dy \\
&= (1/j!)\, E\Big[H_j\left(\Delta^{-1/2}(Y_{t+\Delta}-y_0)\right) \Big| Y_t = y_0;\, \theta\Big] \qquad (17.11)
\end{aligned}
$$

so that the coefficients $\eta_Z^{(j)}$ are specific conditional moments of the process Y. As such, they can be computed in a number of ways, including for instance Monte Carlo integration.

A particularly attractive alternative, however, is to calculate explicitly a Taylor series expansion in Δ for the coefficients $\eta_Z^{(j)}$. Let $f(y, y_0)$ be a polynomial. Polynomials and their iterates obtained by repeated application of the generator $\mathcal{A}$ are in $D(\mathcal{A})$ under regularity assumptions on the boundary behavior of the process. $\mathcal{A}$ is the operator which under regularity conditions returns

$$
\mathcal{A}\cdot f = \frac{\partial f}{\partial \delta} + \mu(y)\frac{\partial f}{\partial y} + \frac{1}{2}\sigma^2(y)\frac{\partial^2 f}{\partial y^2} \qquad (17.12)
$$

when applied to functions $f(\delta, y, y_0)$ that are sufficiently differentiable and display an appropriate growth behavior (this includes the Hermite polynomials under mild restrictions on (μ, σ)). For such an f, we have

$$
\begin{aligned}
E[f(\Delta, Y_{t+\Delta}, y_0)\,|\,Y_t = y_0] = \sum_{k=0}^{K} &\mathcal{A}^k(\theta) \bullet f(0, y_0, y_0)\frac{\Delta^k}{k!} \\
&+ O(\Delta^{K+1}), \qquad (17.13)
\end{aligned}
$$

which is then applied to the Taylor-expand (17.11) in powers of Δ. This can be viewed as an expansion in small time, although one that is fully explicit since it merely requires the ability to differentiate repeatedly (μ, σ).

17.2.2. *Explicit expressions for the transition function expansion*

I now apply the method just described. Let $p_Z^{(J,K)}$ denote the Taylor series up to order K in Δ of $p_Z^{(J)}$, formed by using the Taylor series in Δ, up to order K, of the coefficients $\eta_Z^{(j)}$. The series $\eta_Z^{(j,K)}$ of the first seven Hermite coefficients ($j = 0, \ldots, 6$) are given by $\eta_Z^{(0)} = 1$, and, to order $K = 3$, by:

$$
\begin{aligned}
\eta_Z^{(1,3)} = & -\mu_Y \Delta^{1/2} - \left(2\mu_Y \mu_Y^{[1]} + \mu_Y^{[2]}\right) \Delta^{3/2}/4 \\
& \times \left(4\mu_Y \mu_Y^{[1]2} + 4\mu_Y^2 \mu_Y^{[2]} + 6\mu_Y^{[1]} \mu_Y^{[2]} + 4\mu_Y \mu_Y^{[3]} + \mu_Y^{[4]}\right) \Delta^{5/2}/24 \\
\eta_Z^{(2,3)} = & \left(\mu_Y^2 + \mu_Y^{[1]}\right) \Delta/2 + \left(6\mu_Y^2 \mu_Y^{[1]} + 4\mu_Y^{[1]2} + 7\mu_Y \mu_Y^{[2]} + 2\mu_Y^{[3]}\right) \Delta^2/12 \\
& + \left(28\mu_Y^2 \mu_Y^{[1]2} + 28\mu_Y^2 \mu_Y^{[3]} + 16\mu_Y^{[1]3} + 16\mu_Y^3 \mu_Y^{[2]}\right. \\
& \left. + 88\mu_Y \mu_Y^{[1]} \mu_Y^{[2]} + 21\mu_Y^{[2]2} + 32\mu_Y^{[1]} \mu_Y^{[3]} + 16\mu_Y \mu_Y^{[4]} + 3\mu_Y^{[5]}\right) \Delta^3/96 \\
\eta_Z^{(3,3)} = & -\left(\mu_Y^3 + 3\mu_Y \mu_Y^{[1]} + \mu_Y^{[2]}\right) \Delta^{3/2}/6 - \left(12\mu_Y^3 \mu_Y^{[1]} + 28\mu_Y \mu_Y^{[1]2} + 22\mu_Y^2 \mu_Y^{[2]}\right. \\
& \left. + 24\mu_Y^{[1]} \mu_Y^{[2]} + 14\mu_Y \mu_Y^{[3]} + 3\mu_Y^{[4]}\right) \Delta^{5/2}/48 \\
\eta_Z^{(4,3)} = & \left(\mu_Y^4 + 6\mu_Y^2 \mu_Y^{[1]} + 3\mu_Y^{[1]2} + 4\mu_Y \mu_Y^{[2]} + \mu_Y^{[3]}\right) \Delta^2/24 \\
& + \left(20\mu_Y^4 \mu_Y^{[1]} + 50\mu_Y^3 \mu_Y^{[2]} + 100\mu_Y^2 \mu_Y^{[1]2} + 50\mu_Y^2 \mu_Y^{[3]} + 23\mu_Y \mu_Y^{[4]}\right. \\
& \left. + 180\mu_Y \mu_Y^{[1]} \mu_Y^{[2]} + 40\mu_Y^{[1]3} + 34\mu_Y^{[2]2} + 52\mu_Y^{[1]} \mu_Y^{[3]} + 4\mu_Y^{[5]}\right) \Delta^3/240 \\
\eta_Z^{(5,3)} = & -\left(\mu_Y^5 + 10\mu_Y^3 \mu_Y^{[1]} + 15\mu_Y \mu_Y^{[1]2} + 10\mu_Y^2 \mu_Y^{[2]}\right. \\
& \left. + 10\mu_Y^{[1]} \mu_Y^{[2]} + 5\mu_Y \mu_Y^{[3]} + \mu_Y^{[4]}\right) \Delta^{5/2}/120 \\
\eta_Z^{(6,3)} = & \left(\mu_Y^6 + 15\mu_Y^4 \mu_Y^{[1]} + 15\mu_Y^{[1]3} + 20\mu_Y^3 \mu_Y^{[2]} + 15\mu_Y^{[1]} \mu_Y^{[3]} + 45\mu_Y^2 \mu_Y^{[1]2}\right. \\
& \left. + 10\mu_Y^{[2]2} + 15\mu_Y^2 \mu_Y^{[3]} + 60\mu_Y \mu_Y^{[1]} \mu_Y^{[2]} + 6\mu_Y \mu_Y^{[4]} + \mu_Y^{[5]}\right) \Delta^3/720
\end{aligned}
$$

where $\mu_Y^{[k]m} \equiv \left(\partial^k \mu_Y\left(y_0; \theta\right)/\partial y_0^k\right)^m$.

Different ways of gathering the terms are available (as in the CLT, where for example both the Edgeworth and Gram-Charlier expansions are based on a Hermite expansion). Here, if we gather all the terms according to increasing powers of Δ instead of increasing order of the Hermite polynomials, and let $\tilde{p}_Z^{(K)} \equiv p_Z^{(\infty,K)}$–*and similarly for* Y–we obtain an explicit representation of $\tilde{p}_Y^{(K)}$, given by:

$$
\begin{aligned}
\tilde{p}_Y^{(K)}\left(y|y_0, \Delta; \theta\right) = \Delta^{-1/2} \phi\left(\frac{y - y_0}{\Delta^{1/2}}\right) \exp\left(\int_{y_0}^{y} \mu_Y\left(w; \theta\right) dw\right) \\
\times \sum_{k=0}^{K} c_k\left(y|y_0; \theta\right) \frac{\Delta^k}{k!}
\end{aligned}
\tag{17.14}
$$

where $c_0 = 1$ and for all $j \geqslant 1$

$$c_k\left(y|y_0;\theta\right) = k\left(y-y_0\right)^{-k}\int_{y_0}^{y}\left(w-y_0\right)^{k-1}\Big\{\lambda_Y\left(w;\theta\right)c_{k-1}\left(w|y_0;\theta\right) + \left(\partial^2 c_{k-1}\left(w|y_0;\theta\right)/\partial w^2\right)/2\Big\}dw \tag{17.15}$$

with

$$\lambda_Y\left(y;\theta\right) \equiv -\left(\mu_Y^2(y;\theta) + \partial\mu_Y(y;\theta)/\partial y\right)/2. \tag{17.16}$$

This equation allows the determination of the coefficients c_k recursively starting from c_0. These calculations are easily amenable to an implementation using software such as *Mathematica*. This implementation is typically the most convenient and accurate in empirical applications. Of course, that calculation needs only be done once for a particular model; once the formulae are obtained, they can be used in a standard MLE routine.

The first two coefficients are given by

$$c_1\left(y|y_0;\theta\right) = \frac{\int_{y_0}^{y}\lambda_Y(u;\theta)du}{y-y_0} \tag{17.17}$$

$$c_2\left(y|y_0;\theta\right) = \frac{1}{(y-y_0)^2}\int_{y_0}^{y}\frac{dw}{(y_0-w)^2}\Bigg\{2\left(\int_{y_0}^{w}\lambda_Y(u;\theta)du\right) \times\left(\lambda_Y(w;\theta)(y_0-w)^2+1\right) + 2\lambda_Y(w;\theta)(y_0-w) + (y_0-w)^2\lambda_Y'(w;\theta)\Bigg\}. \tag{17.18}$$

These formulae solve the FPK equations up to order Δ^K, both forward and backward:

$$\frac{\partial\tilde{p}_Y^{(K)}}{\partial\Delta} + \frac{\partial}{\partial y}\left\{\mu_Y\left(y;\theta\right)\tilde{p}_Y^{(K)}\right\} - \frac{1}{2}\frac{\partial^2\tilde{p}_Y^{(K)}}{\partial y^2} = O\left(\Delta^K\right) \tag{17.19}$$

$$\frac{\partial\tilde{p}_Y^{(K)}}{\partial\Delta} - \mu_Y\left(y_0;\theta\right)\frac{\partial\tilde{p}_Y^{(K)}}{\partial y_0} - \frac{1}{2}\frac{\partial^2\tilde{p}_Y^{(K)}}{\partial y_0^2} = O\left(\Delta^K\right). \tag{17.20}$$

The boundary behavior of $\tilde{p}_Y^{(K)}$ is similar to that of p_Y : $\lim_{y\to\underline{y}\ or\ \bar{y}} p_Y = 0$. The expansion is designed to deliver an approximation of the density function $y \mapsto p_Y\left(\Delta, y|y_0;\theta\right)$ for a fixed value of conditioning variable y_0. It is not designed to reproduce the limiting behavior of p_Y in the limit where y_0 tends to the boundaries.

By applying (17.7), we obtain the corresponding expression for $\tilde{p}_X^{(K)}$. For instance, at order $K=1$ we get

$$\begin{aligned}
\tilde{p}_X^{(1)}(x|x_0,\Delta;\theta) &= \sigma(\gamma(x;\theta);\theta)^{-1}\Delta^{-1/2}\phi\left(\Delta^{-1/2}\int_{x_0}^{x}\frac{du}{\sigma(u;\theta)}\right)\\
&\quad\times\exp\left(\int_{x_0}^{x}\mu_Y(\gamma(u;\theta);\theta)du/\sigma(u;\theta)\right)\\
&\quad\times\Big(1+c_1(\gamma(x;\theta)|\gamma(x_0;\theta);\theta)\Delta\Big)\\
&=\left(\frac{\sigma(x_0;\theta)}{2\pi\Delta\sigma^3(x;\theta)}\right)^{1/2}\exp\left\{-\frac{1}{2\Delta}\left(\int_{x_0}^{x}\frac{du}{\sigma(u;\theta)}\right)^2\right.\\
&\quad\left.+\int_{x_0}^{x}\frac{\mu(u;\theta)}{\sigma^2(u;\theta)}du\right\}\Big(1+c_1(\gamma(x;\theta)|\gamma(x_0;\theta);\theta)\Delta\Big) \qquad (17.21)
\end{aligned}$$

where

$$c_1(\gamma(x;\theta)|\gamma(x_0;\theta);\theta)=\frac{\int_{x_0}^{x}\lambda_Y(\gamma(u;\theta);\theta)du/\sigma(u;\theta)}{\int_{x_0}^{x}du/\sigma(u;\theta)}. \qquad (17.22)$$

17.2.3. *Convergence of the density sequence*

Aït-Sahalia (2002) shows that the resulting expansion converges as more correction terms are added. Under regularity conditions, there exists $\bar{\Delta}>0$ such that for every $\Delta\in(0,\bar{\Delta})$, $\theta\in\Theta$ and $(x,x_0)\in D_X^2$:

$$p_X^{(J)}(\Delta,x|x_0;\theta)\ \to\ p_X(\Delta,x|x_0;\theta) \quad \text{as } J\to\infty.$$

In addition, the convergence is uniform in θ over Θ, in x over D_X, and in x_0 over compact subsets of D_X.

Finally, maximizing $\ell_N^{(J)}(\theta)$ instead of the true $\ell_N(\theta)$ results in an estimator $\widehat{\theta}_N^{(J)}$ which converges to the true (but incomputable) MLE $\widehat{\theta}_N$ as $J\to\infty$ and inherits all its asymptotic properties. In general, the expansion $\tilde{p}_X^{(K)}$ will converge to p_X as $\Delta\to 0$.

17.2.4. *Extensions and comparison with other methods*

Jensen and Poulsen (2002), Stramer and Yan (2005) and Hurn, Jeisman, and Lindsay (2005) conducted extensive comparisons of different techniques for approximating the transition function and demonstrated that the method described is both the most accurate and the fastest to implement for the types of problems and sampling frequencies one encounters in finance. The method has been extended to time inhomogeneous processes by Egorov, Li, and Xu (2003) and to jump-diffusions by Schaumburg (2001) and Yu (2003). DiPietro (2001) has extended the methodology to make it applicable in a Bayesian setting. Bakshi and Yu (2002) propose an alternative centering to (17.6) in the univariate case. Li (2005) considers the case of "damped diffusion" processes.

17.3. Multivariate Likelihood Expansions

Of course, many models of interest in finance are inherently multivariate. The main difficulty in the multivariate case is that the transformation from X to Y that played a crucial role in the construction of the expansions in the univariate case above is, in general, not possible.

17.3.1. *Reducibilty*

As defined in Aït-Sahalia (2001), a diffusion X is *reducible* if and if only if there exists a one-to-one transformation of the diffusion X into a diffusion Y whose diffusion matrix σ_Y is the identity matrix. That is, there exists an invertible function $\gamma(x;\theta)$ such that $Y_t \equiv \gamma(X_t;\theta)$ satisfies the stochastic differential equation

$$dY_t = \mu_Y(Y_t;\theta)\,dt + dW_t. \tag{17.23}$$

Every univariate diffusion is reducible, through the transformation (17.3). Whether or not a given multivariate diffusion is reducible depends on the specification of its σ matrix. Specifically, Proposition 1 of Aït-Sahalia (2001) provides a necessary and sufficient condition for reducibility: the diffusion X is reducible if and only if the inverse diffusion matrix $\sigma^{-1} = \left[\sigma_{i,j}^{-1}\right]_{i,j=1,\dots,m}$ satisfies on $\mathcal{S}_X \times \Theta$ the condition that

$$\frac{\partial \sigma_{ij}^{-1}(x;\theta)}{\partial x_k} = \frac{\partial \sigma_{ik}^{-1}(x;\theta)}{\partial x_j} \tag{17.24}$$

for each triplet $(i,j,k) = 1,\dots,m$ such that $k > j$, or equivalently

$$\sum_{l=1}^{m} \frac{\partial \sigma_{ik}(x;\theta)}{\partial x_l}\sigma_{lj}(x;\theta) = \sum_{l=1}^{m} \frac{\partial \sigma_{ij}(x;\theta)}{\partial x_l}\sigma_{lk}(x;\theta). \tag{17.25}$$

Whenever a diffusion is reducible, an expansion can be computed for the transition density p_X of X by first computing it for the density p_Y of Y and then transforming Y back into X (see Section 17.3.2). When a diffusion is not reducible, the situation is going to be more involved (see Section 17.3.3), although it still leads to a closed form expression.

17.3.2. *Determination of the coefficients in the reducible case*

The expansion for l_Y is of the form

$$l_Y^{(K)}\left(\Delta, y|y_0;\theta\right) = -\frac{m}{2}\ln\left(2\pi\Delta\right) + \frac{C_Y^{(-1)}\left(y|y_0;\theta\right)}{\Delta} + \sum_{k=0}^{K} C_Y^{(k)}\left(y|y_0;\theta\right)\frac{\Delta^k}{k!}. \tag{17.26}$$

The coefficients of the expansion are given explicitly by:

$$C_Y^{(-1)}\left(y|y_0;\theta\right) = -\frac{1}{2}\sum_{i=1}^{m}\left(y_i - y_{0i}\right)^2 \tag{17.27}$$

$$C_Y^{(0)}\left(y|y_0;\theta\right) = \sum_{i=1}^{m}\left(y_i - y_{0i}\right)\int_0^1 \mu_{Yi}\left(y_0 + u\left(y - y_0\right);\theta\right)du \tag{17.28}$$

and, for $k \geq 1$,

$$C_Y^{(k)}\left(y|y_0;\theta\right) = k\int_0^1 G_Y^{(k)}\left(y_0 + u\left(y - y_0\right)|y_0;\theta\right)u^{k-1}du \tag{17.29}$$

where

$$G_Y^{(1)}\left(y|y_0;\theta\right) = -\sum_{i=1}^{m}\frac{\partial\mu_{Yi}\left(y;\theta\right)}{\partial y_i} - \sum_{i=1}^{m}\mu_{Yi}\left(y;\theta\right)\frac{\partial C_Y^{(0)}\left(y|y_0;\theta\right)}{\partial y_i} + \frac{1}{2}\sum_{i=1}^{m}\left\{\frac{\partial^2 C_Y^{(0)}\left(y|y_0;\theta\right)}{\partial y_i^2} + \left[\frac{\partial C_Y^{(0)}\left(y|y_0;\theta\right)}{\partial y_i}\right]^2\right\} \tag{17.30}$$

and for $k \geq 2$

$$G_Y^{(k)}\left(y|y_0;\theta\right) = -\sum_{i=1}^{m}\mu_{Yi}\left(y;\theta\right)\frac{\partial C_Y^{(k-1)}\left(y|y_0;\theta\right)}{\partial y_i} + \frac{1}{2}\sum_{i=1}^{m}\frac{\partial^2 C_Y^{(k-1)}\left(y|y_0;\theta\right)}{\partial y_i^2} + \frac{1}{2}\sum_{i=1}^{m}\sum_{h=0}^{k-1}\binom{k-1}{h}\frac{\partial C_Y^{(h)}\left(y|y_0;\theta\right)}{\partial y_i}\frac{\partial C_Y^{(k-1-h)}\left(y|y_0;\theta\right)}{\partial y_i}. \tag{17.31}$$

Given an expansion for the density p_Y of Y, an expansion for the density p_X of X can be obtained by a direct application of the Jacobian formula:

$$l_X^{(K)}\left(\Delta, x|x_0;\theta\right) = -\frac{m}{2}\ln\left(2\pi\Delta\right) - D_v\left(x;\theta\right) + \frac{C_Y^{(-1)}\left(\gamma\left(x;\theta\right)|\gamma\left(x_0;\theta\right);\theta\right)}{\Delta} + \sum_{k=0}^{K} C_Y^{(k)}\left(\gamma\left(x;\theta\right)|\gamma\left(x_0;\theta\right);\theta\right)\frac{\Delta^k}{k!} \tag{17.32}$$

from $l_Y^{(K)}$ given in (17.26), using the coefficients $C_Y^{(k)}$, $k = -1, 0, ..., K$ given above, and where

$$v(x;\theta) \equiv \sigma(x;\theta)\sigma(x;\theta)^T \tag{17.33}$$

$$D_v\,(x;\theta) \equiv \frac{1}{2}\ln\left(Det[v(x;\theta)]\right). \tag{17.34}$$

17.3.3. *Determination of the coefficients in the irreducible case*

In the irreducible case, the expansion of the log likelihood is taken in the form

$$\begin{aligned} l_X^{(K)}\,(\Delta, x|x_0;\theta) = &-\frac{m}{2}\ln\left(2\pi\Delta\right) - D_v\,(x;\theta) \\ &+\frac{C_X^{(-1)}\,(x|x_0;\theta)}{\Delta} + \sum_{k=0}^{K} C_X^{(k)}\,(x|x_0;\theta)\,\frac{\Delta^k}{k!}. \end{aligned} \tag{17.35}$$

The approach is now to calculate a Taylor series in $(x-x_0)$ of each coefficient $C_X^{(k)}$, at order j_k in $(x-x_0)$. Such an expansion will be denoted by $C_X^{(j_k,k)}$ at order $j_k = 2(K-k)$, for $k = -1, 0, ..., K$.

The resulting expansion will then be

$$\begin{aligned} \tilde{l}_X^{(K)}\,(\Delta, x|x_0;\theta) = &-\frac{m}{2}\ln\left(2\pi\Delta\right) - D_v\,(x;\theta) \\ &+\frac{C_X^{(j_{-1},-1)}\,(x|x_0;\theta)}{\Delta} + \sum_{k=0}^{K} C_X^{(j_k,k)}\,(x|x_0;\theta)\,\frac{\Delta^k}{k!}. \end{aligned} \tag{17.36}$$

Such a Taylor expansion was unnecessary in the reducible case: the expressions given in Section 17.3.2 provide the explicit expressions of the coefficients $C_Y^{(k)}$ and then in (17.32) we have the corresponding ones for $C_X^{(k)}$. However, even for an irreducible diffusion, it is still possible to compute the coefficients $C_X^{(j_k,k)}$ explicitly.

With $v\,(x;\theta) \equiv \sigma\,(x;\theta)\,\sigma^T\,(x;\theta)$, define the following functions of the coefficients and their derivatives:

$$\begin{aligned}G_X^{(0)}(x|x_0;\theta) = \frac{m}{2} &- \sum_{i=1}^{m} \mu_i(x;\theta)\,\frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_i}\\ &+ \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial v_{ij}(x;\theta)}{\partial x_i}\frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_j}\\ &+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta)\,\frac{\partial^2 C_X^{(-1)}(x|x_0;\theta)}{\partial x_i \partial x_j}\\ &- \sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta)\,\frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_i}\frac{\partial D_v(x;\theta)}{\partial x_j},\end{aligned} \tag{17.37}$$

$$\begin{aligned}G_X^{(1)}(x|x_0;\theta) = &- \sum_{i=1}^{m} \frac{\partial \mu_i(x;\theta)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial^2 v_{ij}(x;\theta)}{\partial x_i \partial x_j}\\ &- \sum_{i=1}^{m} \mu_i(x;\theta)\left(\frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_i} - \frac{\partial D_v(x;\theta)}{\partial x_i}\right)\\ &+ \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial v_{ij}(x;\theta)}{\partial x_i}\left(\frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_j} - \frac{\partial D_v(x;\theta)}{\partial x_j}\right)\\ &+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta)\left\{\frac{\partial^2 C_X^{(0)}(x|x_0;\theta)}{\partial x_i \partial x_j} - \frac{\partial^2 D_v(x;\theta)}{\partial x_i \partial x_j}\right.\\ &+ \left(\frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_i} - \frac{\partial D_v(x;\theta)}{\partial x_i}\right)\\ &\left.\times \left(\frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_j} - \frac{\partial D_v(x;\theta)}{\partial x_j}\right)\right\}\end{aligned} \tag{17.38}$$

and for $k \geq 2$:

$$
\begin{aligned}
G_X^{(k)}(x|x_0;\theta) = &-\sum_{i=1}^{m}\mu_i(x;\theta)\frac{\partial C_X^{(k-1)}(x|x_0;\theta)}{\partial x_i} \\
&+\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial v_{ij}(x;\theta)}{\partial x_i}\frac{\partial C_X^{(k-1)}(x|x_0;\theta)}{\partial x_j} \\
&+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}(x;\theta)\frac{\partial^2 C_X^{(k-1)}(x|x_0;\theta)}{\partial x_i\partial x_j}+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}(x;\theta) \\
&\times\left\{2\left(\frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_i}-\frac{\partial D_v(x;\theta)}{\partial x_i}\right)\frac{\partial C_X^{(k-1)}(x|x_0;\theta)}{\partial x_j}\right. \\
&\left.+\sum_{h=1}^{k-2}\binom{k-2}{h}\frac{\partial C_X^{(h)}(x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(k-1-h)}(x|x_0;\theta)}{\partial x_j}\right\}.
\end{aligned}
\tag{17.39}
$$

For each $k = -1, 0, ..., K$, the coefficient $C_X^{(k)}(x|x_0;\theta)$ in (17.35) solves the equation

$$
f_X^{(k-1)}(x|x_0;\theta) = 0 \tag{17.40}
$$

where

$$
\begin{aligned}
f_X^{(-2)}(x|x_0;\theta) = &-2C_X^{(-1)}(x|x_0;\theta) \\
&-\sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_j} \\
f_X^{(-1)}(x|x_0;\theta) = &-\sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_j} \\
&-G_X^{(0)}(x|x_0;\theta)
\end{aligned}
$$

and for $k \geq 1$

$$
\begin{aligned}
f_X^{(k-1)}(x|x_0;\theta) = C_X^{(k)}(x|x_0;\theta) &- \sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_i} \\
&\times\frac{\partial C_X^{(k)}(x|x_0;\theta)}{\partial x_j}-G_X^{(k)}(x|x_0;\theta),
\end{aligned}
$$

where the functions $G_X^{(k)}$, $k = 0, 1, ..., K$ are given above. $G_X^{(k)}$ involves only the coefficients $C_X^{(h)}$ for $h = -1, ..., k-1$, so this system of equation can be utilized to solve recursively for each coefficient at a time. Specifically, the equation $f_X^{(-2)} = 0$ determines $C_X^{(-1)}$; given $C_X^{(-1)}$, $G_X^{(0)}$ becomes known and the equation $f_X^{(-1)} = 0$ determines $C_X^{(0)}$; given $C_X^{(-1)}$ and $C_X^{(0)}$, $G_X^{(1)}$ becomes known and the equation $f_X^{(0)} = 0$ then determines $C_X^{(1)}$, etc. It turns out that this results in a system of linear equations in the coefficients of the polynomials $C_X^{(j_k,k)}$, so each one of these equations can be solved explicitly in the form of the Taylor expansion $C_X^{(j_k,k)}$ of the coefficient $C_X^{(k)}$, at order j_k in $(x - x_0)$. Convergence results for the expansion are proved in Aït-Sahalia (2001).

As in the univariate case, these calculations are straightforward to implement using software. For actual implementation of this method to practical problems in various contexts and with various datasets, see Aït-Sahalia and Kimmel (2002), Aït-Sahalia and Kimmel (2004), Thompson (2004), Cheridito, Filipović, and Kimmel (2005), Mosburger and Schneider (2005), Takamizawa (2005) and Schneider (2006).

17.4. Connection to Saddlepoint Approximations

Aït-Sahalia and Yu (2005) developed an alternative strategy for constructing closed form approximations to the transition density of a continuous time Markov process. Instead of expanding the transition function in orthogonal polynomials around a leading term, we rely on the saddlepoint method, which originates in the work of Daniels (1954). We show that, in the case of diffusions, it is possible by expanding the cumulant generating function of the process to obtain an alternative closed form expansion of its transition density. We also show there that this approach provides an alternative gathering of the correction terms beyond the leading term that is equivalent at order Δ to the irreducible expansion of the transition density just described.

To understand the connection to the saddlepoint approach, it is useful to contrast it with the Hermite-based method described in Section 17.2. That expansion can be viewed as analogous to a small sample correction to the CLT. As in the CLT case, it is natural to consider higher order terms based on Hermite polynomials, which are orthogonal with respect to the leading $N(0,1)$ term. This is an Edgeworth (or Gram–Charlier, depending upon how the terms are gathered) type of expansion. By contrast, saddlepoint expansions rely on first tilting the original density — *transforming it into another one* — and then applying an Edgeworth-like expansion to the tilted density. If the tilted density is chosen wisely, the resulting approximation can be quite accurate in the tails, and applicable fairly generally. In order to be able to calculate a saddlepoint approximation, one needs to be able to calculate the Laplace transform or characteristic function of the process of interest. This requirement is a restriction on the applicability of the method, but as we will see, one that is possible to satisfy in many cases in our context of Markov processes. But even when such a computation is not possible explicitly, we go one step further by showing how useful approximations can be obtained by replacing the characteristic function by an expansion in small time. Expansions in small time, which involve the infinitesimal generator of the Markov process, are a key element shared with the Hermite-based expansions described above.

The key to our approach is to approximate the Laplace transform of the process, and the resulting saddlepoint, as a Taylor series in Δ around their continuous-time limit. This will result in an approximation (in Δ) to the saddlepoint (which itself is an approximation to the true but unknown transition density of the process). By applying (17.13) to the function $f(\delta, x, x_0) = \exp(ux)$, u treated as a fixed parameter, we can compute the expansion of the Laplace transform $\varphi(\Delta, u|x_0)$ in Δ. At order $n_2 = 1$, the result is

$$\varphi^{(1)}(u|x_0, \Delta; \theta) = e^{ux_0}\left(1 + \left(\mu(x_0;\theta)u + \frac{1}{2}\sigma^2(x_0;\theta)u^2\right)\Delta\right).$$

Then, by taking its log, we see that the expansion at order Δ of the cumulant transform K is simply

$$K^{(1)}(u|x_0, \Delta; \theta) = ux_0 + \left(\mu(x_0;\theta)u + \frac{1}{2}\sigma^2(x_0;\theta)u^2\right)\Delta.$$

The first order saddlepoint $\widehat{u}^{(1)}$ solves $\partial K^{(1)}(u|x_0, \Delta; \theta)/\partial u = x$, that is

$$\widehat{u}^{(1)}(x|x_0, \Delta; \theta) = \frac{x - (x_0 + \mu(x_0;\theta)\Delta)}{\sigma^2(x_0;\theta)\Delta}$$

and, when evaluated at $x = x_0 + z\Delta^{1/2}$, we have

$$\widehat{u}^{(1)}\left(x_0 + z\Delta^{1/2}|x_0, \Delta; \theta\right) = \frac{z}{\sigma^2(x_0;\theta)\Delta^{1/2}} + O(1) \tag{17.41}$$

and

$$\begin{aligned}&K^{(1)}\left(\widehat{u}^{(1)}\left(x_0 + z\Delta^{1/2}|x_0, \Delta; \theta\right)|x_0, \Delta; \theta\right)\\ &- \widehat{u}^{(1)}\left(x_0 + z\Delta^{1/2}|x_0, \Delta; \theta\right)\cdot\left(x_0 + z\Delta^{1/2}\right) = -\frac{z^2}{2\sigma^2(x_0;\theta)} + O(\Delta^{1/2}).\end{aligned}$$

Similarly, a second order expansion in Δ of $K^{(2)}$ is obtained as

$$
\begin{aligned}
&K^{(2)}\left(u|x_0,\Delta;\theta\right) = ux_0 + \Delta\left(\mu(x_0;\theta)u + \frac{1}{2}\sigma^2(x_0;\theta)u^2\right)\\
&+\ \frac{\Delta^2 u}{8}\left\{4\mu(x_0;\theta)\mu'(x_0;\theta) + 2\sigma^2(x_0;\theta)\mu''(x_0;\theta)\right.\\
&+\ u\left(4\sigma^2(x_0;\theta)\mu'(x_0;\theta) + 2\mu(x_0;\theta)(\sigma^2)'(x_0;\theta) + \sigma^2(x_0;\theta)(\sigma^2)''(x_0;\theta)\right)\\
&\left.+\ 2u^2\sigma^2(x_0;\theta)(\sigma^2)'(x_0;\theta)\right\} + O(\Delta^3).
\end{aligned}
$$

The second order saddlepoint $\widehat{u}^{(2)}$ solves $\partial K^{(2)}\left(u|x_0,\Delta;\theta\right)/\partial u = x$, which is a quadratic equation explicitly solvable in u, and we see after some calculations that

$$
\begin{aligned}
&\widehat{u}^{(2)}\left(x_0 + z\Delta^{1/2}|x_0,\Delta;\theta\right)\\
&= \frac{z}{\sigma^2(x_0;\theta)\Delta^{1/2}} - \left\{\frac{\mu(x_0;\theta)}{\sigma^2(x_0;\theta)} + \frac{3(\sigma^2)'(x_0;\theta)}{4\sigma^4(x_0;\theta)}z^2\right\} + O(\Delta^{1/2})
\end{aligned}
\tag{17.42}
$$

and

$$
\begin{aligned}
&K^{(2)}\left(\widehat{u}^{(2)}\left(x_0 + z\Delta^{1/2}|x_0,\Delta;\theta\right)|x_0,\Delta;\theta\right)\\
&- \widehat{u}^{(2)}\left(x_0 + z\Delta^{1/2}|x_0,\Delta;\theta\right)\cdot\left(x_0 + z\Delta^{1/2}\right)\\
&= -\frac{z^2}{2\sigma^2(x_0;\theta)} + \left\{\frac{\mu(x_0;\theta)}{\sigma^2(x_0;\theta)}z + \frac{(\sigma^2)'(x_0;\theta)}{4\sigma^4(x_0;\theta)}z^3\right\}\Delta^{1/2} + O(\Delta).
\end{aligned}
$$

The way the correction terms in $\varphi^{(2)}\left(u|x_0,\Delta;\theta\right)$ are grouped is similar to that of an Edgeworth expansion. Higher order approximate Laplace transforms can be constructed (see Aït-Sahalia and Yu (2005)). Write $p^{(n_1,n_2)}$ to indicate a saddlepoint approximation of order n_1 using a Taylor expansion in Δ of the Laplace transform φ, that is correct at order n_2 in Δ. When the expansions in Δ are analytic at zero, then $p^{(n_1,\infty)} = p^{(n_1)}$. First, the leading term of the saddlepoint approximation at the first order in Δ and with a Gaussian base coincides with the classical Euler approximation of the transition density,

$$
\begin{aligned}
p_X^{(0,1)}\left(x|x_0,\Delta;\theta\right) = &\left(2\pi\Delta\sigma^2(x_0;\theta)\right)^{-1/2}\\
&\times\exp\left(-\frac{\left(x - x_0 - \mu(x_0;\theta)\Delta\right)^2}{\sigma^2(x_0;\theta)\Delta}\right).
\end{aligned}
\tag{17.43}
$$

The first order saddlepoint approximation at the first order in Δ and with a Gaussian base is

$$\begin{aligned} & p_X^{(1,1)}\left(x_0 + z\Delta^{1/2}|x_0,\Delta;\theta\right) \\ &= \frac{\exp\left(-\frac{z^2}{2\sigma^2(x_0;\theta)} + e_{1/2}(z|x_0;\theta)\Delta^{1/2} + e_1(z|x_0;\theta)\Delta\right)}{\sqrt{2\pi}\sigma(x_0;\theta)\Delta^{1/2}\left\{1 + d_{1/2}(z|x_0;\theta)\Delta^{1/2} + d_1(z|x_0;\theta)\Delta\right\}} \\ &\quad \times\{1 + c_1(z|x_0;\theta)\Delta\} \end{aligned} \tag{17.44}$$

where

$$\begin{aligned} e_{1/2}(z|x_0;\theta) &= \frac{z\mu(x_0;\theta)}{\sigma^2(x_0;\theta)} + \frac{z^3(\sigma^2)'(x_0;\theta)}{4\sigma^4(x_0;\theta)} \\ e_1(z|x_0;\theta) &= -\frac{\mu(x_0;\theta)^2}{2\sigma^2(x_0;\theta)} \\ &+ \frac{z^2\left(12\sigma^2(x_0;\theta)\left(4\mu'(x_0;\theta) + (\sigma^2)''(x_0;\theta)\right) - 48\mu(x_0;\theta)(\sigma^2)'(x_0;\theta)\right)}{96\sigma^4(x_0;\theta)} \\ &+ \frac{z^4\left(8\sigma^2(x_0;\theta)(\sigma^2)''(x_0;\theta) - 15(\sigma^2)'(x_0;\theta)^2\right)}{96\sigma^6(x_0;\theta)} \end{aligned} \tag{17.45}$$

$$\begin{aligned} d_{1/2}(z|x_0;\theta) &= \frac{3z\sigma'(x_0;\theta)}{2\sigma(x_0;\theta)} \\ d_1(z|x_0;\theta) &= \frac{\mu'(x_0;\theta)}{2} - \frac{\mu(x_0;\theta)\sigma'(x_0;\theta)}{\sigma(x_0;\theta)} + \frac{\sigma'(x_0;\theta)^2}{4} + \frac{\sigma(x_0;\theta)\sigma''(x_0;\theta)}{4} \\ &\quad + z^2\left(\frac{5\sigma'(x_0;\theta)^2}{8\sigma(x_0;\theta)^2} + \frac{\sigma''(x_0;\theta)}{\sigma(x_0;\theta)}\right) \qquad (17.46) \\ c_1(z|x_0;\theta) &= \frac{1}{4}(\sigma^2)''(x_0;\theta) - \frac{3}{32}\frac{(\sigma^2)'(x_0;\theta)^2}{\sigma^2(x_0;\theta)}. \end{aligned}$$

The expression (17.44) provides an alternative gathering of the correction terms beyond the leading term that is equivalent at order Δ to the irreducible expansion of the transition density resulting from the irreducible method described in Section 17.3.3.

17.5. An Example with Nonlinear Drift and Diffusion Specifications

The likelihood expansions are given for many specific models in Aït-Sahalia (1999), including the Ornstein-Uhlenbeck specification of Vasicek (1977), the Feller square root model of Cox, Ingersoll, and Ross (1985), the linear drift with CEV diffusion model of Cox (1975) and the more general version of Chan, Karolyi, Longstaff, and Sanders (1992),

$$dX_t = \kappa\left(\alpha - X_t\right)dt + \sigma X_t^{\rho} dW_t$$

a double well model

$$dX_t = \left(\alpha_1 X_t - \alpha_3 X_t^3\right)dt + dW_t$$

and a simpler version of the nonlinear model of Aït-Sahalia (1996),

$$dX_t = \left(\alpha_{-1} X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2\right)dt + \sigma X_t^{\rho} dW_t.$$

One example that was not included in full generality in Aït-Sahalia (1999), however, is the general model proposed for the short term interest rate in Aït-Sahalia (1996)

$$\begin{aligned} dX_t = &\left(\alpha_{-1} X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2\right)dt \\ &+ (\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3})dW_t \end{aligned} \tag{17.47}$$

because even in the univariate case the transformation $X \mapsto Y$ does not lead to an explicit integration in (17.3). But, as discussed in Aït-Sahalia (2001), one can use the irreducible method in that case, thereby bypassing that transformation. For instance, at order $K = 1$ in Δ, the irreducible expansion for the generic model $dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$ is given by (17.36) with $m = 1$, namely:

$$\begin{aligned} \tilde{l}_X^{(1)}\left(\Delta, x|x_0;\theta\right) = &-\frac{1}{2}\ln\left(2\pi\Delta\right) - D_v\left(x;\theta\right) + \frac{C_X^{(4,-1)}\left(x|x_0;\theta\right)}{\Delta} \\ &+ C_X^{(2,0)}\left(x|x_0;\theta\right) + C_X^{(0,1)}\left(x|x_0;\theta\right)\Delta \end{aligned}$$

with $D_v\left(x;\theta\right) = \ln(\sigma\left(x;\theta\right))$. The coefficients $C_X^{(j_k,k)}$, $k = -1, 0, 1$ are given by

$$\begin{aligned} C_X^{(4,-1)}\left(x|x_0;\theta\right) = &-\frac{1}{2\sigma(x_0;\theta)^2}(x-x_0)^2 + \frac{\sigma'(x_0;\theta)}{2\sigma(x_0;\theta)^3}(x-x_0)^3 \\ &+ \frac{\left(4\sigma(x_0;\theta)\sigma''(x_0;\theta) - 11\sigma'(x_0;\theta)^2\right)}{24\sigma(x_0;\theta)^4}(x-x_0)^4 \end{aligned}$$

$$\begin{aligned} C_X^{(2,0)}\left(x|x_0;\theta\right) = &\frac{(2\mu(x_0;\theta) - \sigma(x_0;\theta)\sigma'(x_0;\theta))}{2\sigma(x_0;\theta)^2}(x-x_0) \\ &+ \frac{1}{4\sigma(x_0;\theta)^3}\Big\{\left(\sigma'(x_0;\theta)^2 + 2\mu'(x_0;\theta)\right)\sigma(x_0;\theta) \\ &- 4\mu(x_0;\theta)\sigma'(x_0;\theta) - \sigma''(x_0;\theta)\sigma(x_0;\theta)^2\Big\}(x-x_0)^2 \end{aligned}$$

$$C_X^{(0,1)}(x|x_0;\theta) = \frac{1}{8}\left(2\sigma(x_0;\theta)\sigma''(x_0;\theta) - \frac{4\mu(x_0;\theta)^2}{\sigma(x_0;\theta)^2} + \frac{8\sigma'(x_0;\theta)\mu(x_0;\theta)}{\sigma(x_0;\theta)} - \sigma'(x_0;\theta)^2 - 4\mu'(x_0;\theta)\right) \quad (17.48)$$

In the case of model (17.47), this specializes to the following expressions:

$$\begin{aligned}
C_X^{(4,-1)}(x|x_0;\theta) = &- \frac{1}{2\left(\beta_2 {x_0}^{\beta_3} + \beta_1 x_0 + \beta_0\right)^2}(x-x_0)^2 \\
&+ \frac{\left(\beta_2\beta_3 {x_0}^{\beta_3-1} + \beta_1\right)}{2\left(\beta_2 {x_0}^{\beta_3} + \beta_1 x_0 + \beta_0\right)^3}(x-x_0)^3 \\
&+ \frac{1}{24\left(\beta_2 {x_0}^{\beta_3} + \beta_1 x_0 + \beta_0\right)^4}\Big\{4\beta_2(\beta_3-1)\beta_3 {x_0}^{\beta_3-2} \\
&\times \left(\beta_2 {x_0}^{\beta_3} + \beta_1 x_0 + \beta_0\right) - 11\left(\beta_2\beta_3 {x_0}^{\beta_3-1} + \beta_1\right)^2\Big\}(x-x_0)^4
\end{aligned}$$

$$\begin{aligned}
C_X^{(2,0)}(x|x_0;\theta) = &\frac{1}{2\left(\beta_2 {x_0}^{\beta_3} + \beta_1 x_0 + \beta_0\right)^2}\Bigg\{\left(-\beta_2\beta_3 {x_0}^{\beta_3-1} - \beta_1\right) \\
&\times \left(\beta_2 {x_0}^{\beta_3} + \beta_1 x_0 + \beta_0\right) + 2\left(\alpha_0 + x_0(\alpha_1 + \alpha_2 x_0) + \frac{\alpha_{-1}}{x_0}\right)\Bigg\} \\
&\times (x-x_0) + \Bigg\{\frac{1}{4\left(\beta_2 {x_0}^{\beta_3} + \beta_1 x_0 + \beta_0\right)^3}\Bigg\{-\beta_2(\beta_3-1)\beta_3 \\
&\times \left(\beta_2 {x_0}^{\beta_3} + \beta_1 x_0 + \beta_0\right)^2 {x_0}^{\beta_3-2} \\
&-4\left(\beta_2\beta_3 {x_0}^{\beta_3-1} + \beta_1\right)\left(\alpha_0 + x_0(\alpha_1 + \alpha_2 x_0) + \frac{\alpha_{-1}}{x_0}\right)\Bigg\} \\
&+ \frac{\left(\beta_2\beta_3 {x_0}^{\beta_3-1} + \beta_1\right)^2 + 2\left(\alpha_1 + 2\alpha_2 x_0 - \frac{\alpha_{-1}}{{x_0}^2}\right)}{4\left(\beta_2 {x_0}^{\beta_3} + \beta_1 x_0 + \beta_0\right)^2}\Bigg\}(x-x_0)^2
\end{aligned}$$

$$C_X^{(0,1)}(x|x_0;\theta) = \frac{1}{8}\Big\{2\beta_2(\beta_3-1)\beta_3\big(\beta_2 {x_0}^{\beta_3}+\beta_1 x_0+\beta_0\big){x_0}^{\beta_3-2}$$
$$-\big(\beta_2\beta_3 {x_0}^{\beta_3-1}+\beta_1\big)^2-4\Big(\alpha_1+2\alpha_2 x_0-\frac{\alpha_{-1}}{{x_0}^2}\Big)\Big\}$$
$$+\frac{1}{8\big(\beta_2 {x_0}^{\beta_3}+\beta_1 x_0+\beta_0\big)^2}\Big\{8\big(\beta_2\beta_3 {x_0}^{\beta_3-1}+\beta_1\big)\Big(\alpha_0+x_0(\alpha_1+\alpha_2 x_0)+\frac{\alpha_{-1}}{x_0}\Big)$$
$$\times\ \big(\beta_2 {x_0}^{\beta_3}+\beta_1 x_0+\beta_0\big)\ -4\Big(\alpha_0+x_0(\alpha_1+\alpha_2 x_0)+\frac{\alpha_{-1}}{x_0}\Big)^2\Big\}.$$

Bakshi, Ju and Qu-Yang (2006) provide an application to equity volatility dynamics for a variety of models.

17.6. An Example with Stochastic Volatility

Consider as a second example a typical stochastic volatility model

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \end{pmatrix} = \begin{pmatrix} \mu \\ \kappa\,(\alpha - X_{2t}) \end{pmatrix} dt + \begin{pmatrix} \gamma_{11}\exp(X_{2t}) & 0 \\ 0 & \gamma_{22} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \end{pmatrix} \quad (17.49)$$

where X_{1t} plays the role of the log of an asset price and $\exp(X_{2t})$ is the stochastic volatility variable. While the term $\exp(X_{2t})$ violates the linear growth condition, it does not cause explosions due to the mean reverting nature of the stochastic volatility. This model has no closed-form solution.

The diffusion (17.49) is in general not reducible, so I will apply the irreducible method described above to derive the expansion. The expansion at order $K = 3$ is of the form (17.35), with the coefficients $C_X^{(j_k,k)}$, $k = -1, 0, ..., 3$ given explicitly by:

$$C_X^{(8,-1)}(x|x_0;\theta) = -\frac{1}{2}\frac{(x_1-x_{01})^2}{e^{2x_{02}}\gamma_{11}^2} - \frac{1}{2}\frac{(x_2-x_{02})^2}{\gamma_{22}^2} + \frac{(x_1-x_{01})^2\,(x_2-x_{02})}{2e^{2x_{02}}\gamma_{11}^2}$$
$$-\frac{(x_1-x_{01})^2(x_2-x_{02})^2}{6e^{2x_{02}}\gamma_{11}^2} + \frac{(x_1-x_{01})^4\gamma_{22}^2}{24e^{4x_{02}}\gamma_{11}^4} - \frac{(x_1-x_{01})^4\,(x_2-x_{02})\,\gamma_{22}^2}{12e^{4x_{02}}\gamma_{11}^4}$$
$$+\frac{(x_1-x_{01})^2(x_2-x_{02})^4}{90e^{2x_{02}}\gamma_{11}^2} + \frac{(x_1-x_{01})^4(x_2-x_{02})^2\gamma_{22}^2}{15e^{4x_{02}}\gamma_{11}^4} - \frac{(x_1-x_{01})^6\gamma_{22}^4}{180e^{6x_{02}}\gamma_{11}^6}$$
$$-\frac{(x_1-x_{01})^4(x_2-x_{02})^3\gamma_{22}^2}{45e^{4x_{02}}\gamma_{11}^4} + \frac{(x_1-x_{01})^6\,(x_2-x_{02})\,\gamma_{22}^4}{60e^{6x_{02}}\gamma_{11}^6}$$
$$-\frac{(x_1-x_{01})^2(x_2-x_{02})^6}{945e^{2x_{02}}\gamma_{11}^2} - \frac{(x_1-x_{01})^4(x_2-x_{02})^4\gamma_{22}^2}{630e^{4x_{02}}\gamma_{11}^4}$$
$$-\frac{3(x_1-x_{01})^6(x_2-x_{02})^2\gamma_{22}^4}{140e^{6x_{02}}\gamma_{11}^6} + \frac{(x_1-x_{01})^8\gamma_{22}^6}{1120e^{8x_{02}}\gamma_{11}^8}$$

$$
\begin{aligned}
&C_X^{(6,0)}(x|x_0;\theta) = \frac{\mu(x_1-x_{01})}{e^{2x_{02}}\gamma_{11}^2} + (x_2-x_{02})\left(\frac{1}{2}+\frac{\kappa(\alpha-x_{02})}{\gamma_{22}^2}\right)\\
&-\frac{\mu(x_1-x_{01})(x_2-x_{02})}{e^{2x_{02}}\gamma_{11}^2} - \frac{(x_1-x_{01})^2\gamma_{22}^2}{12e^{2x_{02}}\gamma_{11}^2} - \frac{(x_2-x_{02})^2\left(6\kappa+\gamma_{22}^2\right)}{12\gamma_{22}^2}\\
&+\frac{\mu(x_1-x_{01})(x_2-x_{02})^2}{3e^{2x_{02}}\gamma_{11}^2} - \frac{\mu(x_1-x_{01})^3\gamma_{22}^2}{6e^{4x_{02}}\gamma_{11}^4} + \frac{(x_1-x_{01})^2(x_2-x_{02})\gamma_{22}^2}{12e^{2x_{02}}\gamma_{11}^2}\\
&+\frac{(x_2-x_{02})^4}{360} + \frac{\mu(x_1-x_{01})^3(x_2-x_{02})\gamma_{22}^2}{3e^{4x_{02}}\gamma_{11}^4} - \frac{(x_1-x_{01})^2(x_2-x_{02})^2\gamma_{22}^2}{45e^{2x_{02}}\gamma_{11}^2}\\
&+\frac{7(x_1-x_{01})^4\gamma_{22}^4}{720e^{4x_{02}}\gamma_{11}^4} - \frac{\mu(x_1-x_{01})(x_2-x_{02})^4}{45e^{2x_{02}}\gamma_{11}^2} - \frac{4\mu(x_1-x_{01})^3(x_2-x_{02})^2\gamma_{22}^2}{15e^{4x_{02}}\gamma_{11}^4}\\
&-\frac{(x_1-x_{01})^2(x_2-x_{02})^3\gamma_{22}^2}{180e^{2x_{02}}\gamma_{11}^2} + \frac{\mu(x_1-x_{01})^5\gamma_{22}^4}{30e^{6x_{02}}\gamma_{11}^6} - \frac{7(x_1-x_{01})^4(x_2-x_{02})\gamma_{22}^4}{360e^{4x_{02}}\gamma_{11}^4}\\
&-\frac{(x_2-x_{02})^6}{5670} + \frac{4\mu(x_1-x_{01})^3(x_2-x_{02})^3\gamma_{22}^2}{45e^{4x_{02}}\gamma_{11}^4} + \frac{(x_1-x_{01})^2(x_2-x_{02})^4\gamma_{22}^2}{315e^{2x_{02}}\gamma_{11}^2}\\
&-\frac{\mu(x_1-x_{01})^5(x_2-x_{02})\gamma_{22}^4}{10e^{6x_{02}}\gamma_{11}^6} + \frac{223(x_1-x_{01})^4(x_2-x_{02})^2\gamma_{22}^4}{15120e^{4x_{02}}\gamma_{11}^4}\\
&-\frac{71(x_1-x_{01})^6\gamma_{22}^6}{45360e^{6x_{02}}\gamma_{11}^6}
\end{aligned}
$$

$$
\begin{aligned}
&C_X^{(2,2)}(x|x_0;\theta) = \frac{1}{180e^{2x_{02}}\gamma_{11}^2}\Big\{-30e^{2x_{02}}\kappa^2\gamma_{11}^2 - 30e^{2x_{02}}\alpha\kappa^2\gamma_{11}^2\\
&+30e^{2x_{02}}\kappa^2x_{02}\gamma_{11}^2 - 30\mu^2\gamma_{22}^2 + e^{2x_{02}}\gamma_{11}^2\gamma_{22}^4\Big\} + \frac{(x_2-x_{02})\left(e^{2x_{02}}\kappa^2\gamma_{11}^2+2\mu^2\gamma_{22}^2\right)}{12e^{2x_{02}}\gamma_{11}^2}\\
&-\frac{\mu(x_1-x_{01})}{90e^{4x_{02}}\gamma_{11}^4}\Big\{30e^{2x_{02}}\alpha\kappa^2\gamma_{11}^2 - 30e^{2x_{02}}\kappa^2x_{02}\gamma_{11}^2 + 30\mu^2\gamma_{22}^2 + e^{2x_{02}}\gamma_{11}^2\gamma_{22}^4\Big\}\\
&+\frac{\mu(x_1-x_{01})(x_2-x_{02})}{90e^{4x_{02}}\gamma_{11}^4}\Big\{15e^{2x_{02}}\kappa^2\gamma_{11}^2 + 30e^{2x_{02}}\alpha\kappa^2\gamma_{11}^2 - 30e^{2x_{02}}\kappa^2x_{02}\gamma_{11}^2\\
&+60\mu^2\gamma_{22}^2 + e^{2x_{02}}\gamma_{11}^2\gamma_{22}^4\Big\} - \frac{(x_1-x_{01})^2\gamma_{22}^2}{3780e^{4x_{02}}\gamma_{11}^4}\Big\{-105e^{2x_{02}}\kappa^2\gamma_{11}^2 - 21e^{2x_{02}}\alpha\kappa^2\gamma_{11}^2\\
&+21e^{2x_{02}}\kappa^2x_{02}\gamma_{11}^2 - 441\mu^2\gamma_{22}^2 + 4e^{2x_{02}}\gamma_{11}^2\gamma_{22}^4\Big\} - \frac{(x_2-x_{02})^2}{3780e^{2x_{02}}\gamma_{11}^2}\Big\{-21e^{2x_{02}}\kappa^2\gamma_{11}^2\\
&-42e^{2x_{02}}\alpha\kappa^2\gamma_{11}^2 + 42e^{2x_{02}}\kappa^2x_{02}\gamma_{11}^2 + 168\mu^2\gamma_{22}^2 + 4e^{2x_{02}}\gamma_{11}^2\gamma_{22}^4\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&C_X^{(4,1)}(x|x_0;\theta) = -\frac{1}{24e^{2x_{02}}\gamma_{11}^2\gamma_{22}^2}\Big\{12e^{2x_{02}}\alpha^2\kappa^2\gamma_{11}^2 - 24e^{2x_{02}}\alpha\kappa^2 x_{02}\gamma_{11}^2 \\
&+ 12e^{2x_{02}}\kappa^2 x_{02}\gamma_{11}^2 + 12\mu^2\gamma_{22}^2 12e^{2x_{02}}\kappa\gamma_{11}^2\gamma_{22}^2 + e^{2x_{02}}\gamma_{11}^2\gamma_{22}^4\Big\} + \frac{\mu(x_1-x_{01})\gamma_{22}^2}{6e^{2x_{02}}\gamma_{11}^2} \\
&- \frac{(x_2-x_{02})}{2e^{2x_{02}}\gamma_{11}^2\gamma_{22}^2}\Big\{-e^{2x_{02}}\alpha\kappa^2\gamma_{11}^2 + e^{2x_{02}}\kappa^2 x_{02}\gamma_{11}^2 - \mu^2\gamma_{22}^2\Big\} \\
&- \frac{\mu(x_1-x_{01})(x_2-x_{02})\gamma_{22}^2}{6e^{2x_{02}}\gamma_{11}^2} - \frac{(x_1-x_{01})^2}{360e^{4x_{02}}\gamma_{11}^4}\Big\{-30e^{2x_{02}}\alpha\kappa^2\gamma_{11}^2 + 30e^{2x_{02}}\kappa^2 x_{02}\gamma_{11}^2 \\
&- 90\mu^2\gamma_{22}^2 - e^{2x_{02}}\gamma_{11}^2\gamma_{22}^4\Big\} + \frac{(x_2-x_{02})^2}{360e^{2x_{02}}\gamma_{11}^2\gamma_{22}^2}\Big\{-60e^{2x_{02}}\kappa^2\gamma_{11}^2 - 60\mu^2\gamma_{22}^2 \\
&+ e^{2x_{02}}\gamma_{11}^2\gamma_{22}^4\Big\} + \frac{2\mu(x_1-x_{01})(x_2-x_{02})^2\gamma_{22}^2}{45e^{2x_{02}}\gamma_{11}^2} - \frac{7\mu(x_1-x_{01})^3\gamma_{22}^4}{180e^{4x_{02}}\gamma_{11}^4} \\
&- \frac{(x_1-x_{01})^2(x_2-x_{02})}{360e^{4x_{02}}\gamma_{11}^4}\Big\{15e^{2x_{02}}\kappa^2\gamma_{11}^2 + 30e^{2x_{02}}\alpha\kappa^2\gamma_{11}^2 - 30e^{2x_{02}}\kappa^2 x_{02}\gamma_{11}^2 \\
&+ 180\mu^2\gamma_{22}^2 + e^{2x_{02}}\gamma_{11}^2\gamma_{22}^4\Big\} + \frac{\mu(x_1-x_{01})(x_2-x_{02})^3\gamma_{22}^2}{90e^{2x_{02}}\gamma_{11}^2} \\
&+ \frac{7\mu(x_1-x_{01})^3(x_2-x_{02})\gamma_{22}^4}{90e^{4x_{02}}\gamma_{11}^4} - \frac{(x_2-x_{02})^4\left(-42\mu^2 + e^{2x_{02}}\gamma_{11}^2\gamma_{22}^2\right)}{3780e^{2x_{02}}\gamma_{11}^2} \\
&+ \frac{(x_1-x_{01})^2(x_2-x_{02})^2}{2520e^{4x_{02}}\gamma_{11}^4}\Big\{98e^{2x_{02}}\kappa^2\gamma_{11}^2 + 56e^{2x_{02}}\alpha\kappa^2\gamma_{11}^2 - 56e^{2x_{02}}\kappa^2 x_{02}\gamma_{11}^2 \\
&+ 1008\mu^2\gamma_{22}^2 + e^{2x_{02}}\gamma_{11}^2\gamma_{22}^4\Big\} - \frac{(x_1-x_{01})^4\gamma_{22}^2}{10080e^{6x_{02}}\gamma_{11}^6}\Big\{42e^{2x_{02}}\kappa^2\gamma_{11}^2 + 112e^{2x_{02}}\alpha\kappa^2\gamma_{11}^2 \\
&- 112e^{2x_{02}}\kappa^2 x_{02}\gamma_{11}^2 + 840\mu^2\gamma_{22}^2 + 5e^{2x_{02}}\gamma_{11}^2\gamma_{22}^4\Big\}
\end{aligned}
$$

and

$$
\begin{aligned}
&C_X^{(0,3)}(x|x_0;\theta) = \frac{1}{7560e^{4x_{02}}\gamma_{11}^4\gamma_{22}^2}\Big\{1890\mu^4\gamma_{22}^4 \\
&+ 126e^{2x_{02}}\mu^2\gamma_{11}^2\gamma_{22}^2\left(30\kappa^2(\alpha-x_{02})\gamma_{22}^4\right) + e^{4x_{02}}\gamma_{11}^4\Big(1890\kappa^4(x_{02}-\alpha)^2 \\
&- 63\kappa^2(1-2\alpha+2x_{02})\gamma_{22}^4 - 16\gamma_{22}^8\Big)\Big\}.
\end{aligned}
$$

17.7. Inference When the State is Partially Observed

In many cases, the state vector is of the form $X_t = [S_t; V_t]'$, where the $(m-q)$−dimensional vector S_t is observed but the q−dimensional V_t is not. Two typical examples in finance consist of stochastic volatility models, such as the example just discussed, where V_t is the volatility state variable(s), and term structure models, where V_t is a vector of factors or yields. One can conduct likelihood inference in this setting, without resorting to the statistically sound but computationally infeasible integration of the latent variables from the likelihood function. The idea is simple: write down in closed form an expansion for the log-likelihood of the state vector X, including its unobservable components. Then enlarge the observation state by adding variables that are observed and functions of X. For example, in the stochastic volatility case, an option price or an option-implied volatility; in term structure models, as many bonds as there are factors. Then, using the Jacobian formula, write down the likelihood function of the pair consisting of the observed components of X and the additional observed variables, and maximize it.

Identification of the parameter vector must be ensured. In fact, identifying a multivariate continuous-time Markov process from discrete-time data can be problematic when the process is not reversible, as an aliasing problem can be present in the multivariate case (see Philips (1973) and Hansen and Sargent (1983)). As for the distributional properties of the resulting estimator, a fixed interval sample of a time-homogenous continuous-time Markov process is a Markov process in discrete time. Given that the Markov state vector is observed and the unknown parameters are identified, properties of the MLE follow from what is known about ML estimation of discrete-time Markov processes (see Billingsley (1961)).

17.7.1. *Likelihood inference for stochastic volatility models*

In a stochastic volatility model, the asset price process S_t follows

$$dS_t = (r-\delta)\,S_t dt + \sigma_1\,(X_t;\theta)\,dW_t^Q \tag{17.50}$$

where r is the riskfree rate, δ is the dividend yield paid by the asset (both taken to be constant for simplicity only), σ_1 denotes the first row of the matrix σ and Q denotes the equivalent martingale measure (see e.g., Harrison and Kreps (1979)). The volatility state variables V_t then follow a SDE on their own. For example, in the Heston (1993) model, $m=2$ and $q=1$:

$$\begin{aligned} dX_t = d\begin{bmatrix} S_t \\ V_t \end{bmatrix} &= \begin{bmatrix} (r-\delta)\,S_t \\ \kappa\,(\gamma - V_t) \end{bmatrix} dt \\ &\quad + \begin{bmatrix} \sqrt{(1-\rho^2)\,V_t}S_t & \rho\sqrt{V_t}S_t \\ 0 & \sigma\sqrt{V_t} \end{bmatrix} d\begin{bmatrix} W_1^Q\,(t) \\ W_2^Q\,(t) \end{bmatrix}. \end{aligned} \tag{17.51}$$

The model is completed by the specification of a vector of market prices of risk for the different sources of risk (W_1 and W_2 here), such as

$$\Lambda\left(X_t;\theta\right)=\left[\lambda_1\sqrt{\left(1-\rho^2\right)V_t},\ \lambda_2\sqrt{V_t}\right]', \tag{17.52}$$

which characterizes the change of measure from Q back to the physical probability measure P.

Likelihood inference for this and other stochastic volatility models is discussed in Aït-Sahalia and Kimmel (2004). Given a time series of observations of both the asset price, S_t, and a vector of option prices (which, for simplicity, we take to be call options) C_t, the time series of V_t can then be inferred from the observed C_t. If V_t is multidimensional, sufficiently many options are required with varying strike prices and maturities to allow extraction of the current value of V_t from the observed stock and call prices. Otherwise, only a single option is needed. For reasons of statistical efficiency, we seek to determine the joint likelihood function of the observed data, as opposed to, for example, conditional or unconditional moments. We employ the closed-form approximation technique described above, which yields in closed form the joint likelihood function of $[S_t;V_t]'$. From there, the joint likelihood function of the observations on $G_t=[S_t;C_t]'=f\left(X_t;\theta\right)$ is obtained simply by multiplying the likelihood of $X_t=[S_t;V_t]'$ by the Jacobian term J_t:

$$\begin{aligned}\ln p_G\left(g|g_0,\Delta;\theta\right)=&-\ln J_t\left(g|g_0,\Delta;\theta\right)\\&+l_X(f^{-1}\left(g;\theta\right)|f^{-1}\left(g_0;\theta\right);\Delta,\theta)\end{aligned} \tag{17.53}$$

with l_X obtained as described above.

If a proxy for V_t is used directly, this last step is not necessary. Indeed, we can avoid the computation of the function f by first transforming C_t into a proxy for V_t. The simplest one consists in using the Black-Scholes implied volatility of a short-maturity at-the-money option in place of the true instantaneous volatility state variable. The use of this proxy is justified in theory by the fact that the implied volatility of such an option converges to the instantaneous volatility of the logarithmic stock price as the maturity of the option goes to zero. An alternate proxy (which we call the integrated volatility proxy) corrects for the effect of mean reversion in volatility during the life of an option. If V_t is the instantaneous variance of the logarithmic stock price, we can express the integral of variance from time t to T as

$$V\left(t,T\right)=\int_t^T V_u du \tag{17.54}$$

If the volatility process is instantaneously uncorrelated with the logarithmic stock price process, then we can calculate option prices by taking the expected value of the Black-Scholes option price (with $V(t,T)$ as implied variance) over the probability distribution of $V(t,T)$ (see Hull and White (1987)). If the two processes are correlated, then the price of the option is a weighted average of Black-Scholes prices evaluated at different stock prices and volatilities (see Romano and Touzi (1997)).

The proxy we examine is determined by calculating the expected value of $V(t,T)$ first, and substituting this value into the Black-Scholes formula as implied variance. This proxy is model-free, in that it can be calculated whether or not an exact volatility can be computed and results in a straightforward estimation procedure. On the other hand, this procedure is in general approximate, first because the volatility process is unlikely to be instantaneously uncorrelated with the logarithmic stock price process, and second, because the expectation is taken before substituting $V(t,T)$ into the Black-Scholes formula rather than after and we examine in Monte Carlo simulations the respective impact of these approximations, with the objective of determining whether the trade-off involved between simplicity and exactitude is worthwhile.

The idea is to adjust the Black-Scholes implied volatility for the effect of mean reversion in volatility, essentially undoing the averaging that takes place in equation (17.54). Specifically, if the Q-measure drift of Y_t is of the form $a + bY_t$ (as it is in many of the stochastic volatility models in use), then the expected value of $V(t,T)$ is given by:

$$E_t\left[V\left(t,T\right)\right] = \left(\frac{e^{b(T-t)}-1}{b}\right)\left(V_t + \frac{a}{b}\right) - \frac{a}{b}\left(T-t\right). \tag{17.55}$$

A similar expression can be derived in the special case where $b = 0$. By taking the expected value on the left-hand side to be the observed implied variance $V_{\mathrm{imp}}(t,T)$ of a short maturity T at-the-money option, our adjusted proxy is then given by:

$$V_t \approx \frac{bV_{\mathrm{imp}}(t,T) + a\left(T-t\right)}{e^{b(T-t)}-1} - \frac{a}{b}. \tag{17.56}$$

Then we can simply take $[S_t; V_{\mathrm{imp}}(t,T)]'$ as the state vector, write its likelihood from that of $[S_t; V_t]'$ using a Jacobian term for the change of variable (17.56).

It is possible to refine the implied volatility proxy by expressing it in the form of a Taylor series in the "volatility of volatility" parameter σ in the case of the CEV model, where the Q-measure drift of Y_t is of the form

$a+bY_t$, and the Q-measure diffusion of Y_t is of the form σY_t^β (Lewis (2000)). However, unlike (17.56), the relationship between the observed $V_{\text{imp}}(t,T)$ and the latent Y_t is not invertible without numerical computation of the parameter-dependent integral.

17.7.2. *Likelihood inference for term structure models*

Another example of a class of models where the state can be only partially observed consist of term structure models. A multivariate term structure model specifies that the instantaneous riskless rate r_t is a deterministic function of an $m-$dimensional vector of state variables, X_t

$$r_t = r\left(X_t;\theta\right) \tag{17.57}$$

which will typically not be fully observable. Under the equivalent martingale measure Q, the state vector X follows the dynamics given in (17.1). In order to avoid arbitrage opportunities, the price at t of a zero-coupon bond maturing at T is given by the Feynman-Kac representation:

$$P\left(x,t,T;\theta\right) = E^Q\left[\exp\left(-\int_t^T r_u du\right)\middle| X_t = x\right]. \tag{17.58}$$

An affine yield model is any model where the short rate (17.57) is an affine function of the state vector and the risk-neutral dynamics (17.1) are affine:

$$dX_t = \left(\tilde{A} + \widetilde{B}X_t\right)dt + \Sigma\sqrt{S\left(X_t;\alpha,\beta\right)}dW_t^Q \tag{17.59}$$

where $\tilde{A}$ is an m–element column vector, $\widetilde{B}$ and Σ are $m\times m$ matrices, and $S\left(X_t;\alpha,\beta\right)$ is the diagonal matrix with elements $S_{ii} = \alpha_i + X_t'\beta_i$, with each α_i a scalar and each β_i an $m\times 1$ vector, $1\le i\le m$ (see Dai and Singleton (2000)).

It can then be shown that, in affine models, bond prices have the exponential affine form

$$P\left(x,t,T;\theta\right) = \exp\left(-\gamma_0\left(\tau;\theta\right) - \gamma\left(\tau;\theta\right)' x\right) \tag{17.60}$$

where $\tau = T-t$ is the bond's time to maturity. That is, bond yields (non-annualized, and denoted by $g\left(x,t,T;\theta\right) = -\ln\left(P\left(x,t,T;\theta\right)\right)$) are affine functions of the state vector:

$$g(x,t,T;\theta) = \gamma_0\left(\tau;\theta\right) + \gamma\left(\tau;\theta\right)' x. \tag{17.61}$$

Alternatively, one can start with the requirement that the yields be affine, and show that the dynamics of the state vector must be affine (see Duffie and Kan (1996)).

The final condition for the bond price implies that $\gamma_0(0;\theta) = \gamma(0;\theta) = 0$, while

$$r_t = \delta_0 + \delta' x. \tag{17.62}$$

Affine yield models owe much of their popularity to the fact that bond prices can be calculated quickly as solutions to a system of ordinary differential equations. Under non-linear term structure models, bond prices will normally be solutions to a partial differential equation that is far more difficult to solve.

Aït-Sahalia and Kimmel (2002) consider likelihood inference for affine term structure models. They derive the likelihood expansions for the nine canonical models of Dai and Singleton (2000) in dimensions $m = 1, 2$ and 3. For instance, in dimension $m = 3$, the four canonical models are respectively

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \\ dX_{3t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{pmatrix} \begin{pmatrix} -X_{1t} \\ -X_{2t} \\ -X_{3t} \end{pmatrix} dt + \begin{pmatrix} dW_{1t} \\ dW_{2t} \\ dW_{3t} \end{pmatrix},$$

$$\begin{aligned} \begin{pmatrix} dX_{1t} \\ dX_{2t} \\ dX_{3t} \end{pmatrix} &= \begin{pmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{pmatrix} \begin{pmatrix} \theta_1 - X_{1t} \\ -X_{2t} \\ -X_{3t} \end{pmatrix} dt \\ &\quad + \begin{pmatrix} X_{1t}^{1/2} & 0 & 0 \\ 0 & (1+\beta_{21}X_{1t})^{\frac{1}{2}} & 0 \\ 0 & 0 & (1+\beta_{21}X_{1t})^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \\ dW_{3t} \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} \begin{pmatrix} dX_{1t} \\ dX_{2t} \\ dX_{3t} \end{pmatrix} &= \begin{pmatrix} \kappa_{11} & \kappa_{12} & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{pmatrix} \begin{pmatrix} \theta_1 - X_{1t} \\ \theta_2 - X_{2t} \\ -X_{3t} \end{pmatrix} dt \\ &\quad + \begin{pmatrix} X_{1t}^{1/2} & 0 & 0 \\ 0 & X_{2t}^{1/2} & 0 \\ 0 & 0 & (1+\beta_{31}X_{1t}+\beta_{32}X_{2t})^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \\ dW_{3t} \end{pmatrix}, \end{aligned}$$

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \\ dX_{3t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{pmatrix} \begin{pmatrix} \theta_1 - X_{1t} \\ \theta_2 - X_{2t} \\ \theta_3 - X_{3t} \end{pmatrix} dt + \begin{pmatrix} X_{1t}^{1/2} & 0 & 0 \\ 0 & X_{2t}^{1/2} & 0 \\ 0 & 0 & X_{3t}^{1/2} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \\ dW_{3t} \end{pmatrix}.$$

MLE in this case requires evaluation of the likelihood of an observed panel of yield data for each parameter vector considered during a search procedure. The procedure for evaluating the likelihood of the observed yields at a particular value of the parameter vector consists of four steps. First, we extract the value of the state vector X_t (which is not directly observed) from those yields that are treated as observed without error. Second, we evaluate the joint likelihood of the series of implied observations of the state vector X_t, using the closed-form approximations to the likelihood function described above. Third, we multiply this joint likelihood by a Jacobian term, to find the likelihood of the panel of observations of the yields observed without error. Finally, we calculate the likelihood of the observation errors for those yields observed with error, and multiply this likelihood by the likelihood found in the previous step, to find the joint likelihood of the panel of all yields.

The first task is therefore to infer the state vector X_t at date t from the cross-section of bond yields at date t with different maturities. Affine yield models, as their name implies, make yields of zero coupon bonds affine functions of the state vector. Given this simple relationship between yields and the state vector, the likelihood function of bond yields is a simple transformation of the likelihood function of the state vector.

If the number of observed yields at that point in time is smaller than the number N of state variables in the model, then the state is not completely observed, and the vector of observed yields does not follow a Markov process, even if the (unobserved) state vector does, enormously complicating maximum likelihood estimation. On the other hand, if the number of observed yields is larger than the number of state variables, then some of the yields can be expressed as deterministic functions of other observed yields, without error. Even tiny deviations from the predicted values have a likelihood of zero. This problem can be avoided by using a number of yields exactly equal to the number of state variables in the underlying model, but, in general, the market price of risk parameters will not all be identified. Specifically, there are affine yield models that generate identical dynamics for yields with a given set of maturities, but different dynamics for yields with other maturities. A common practice (see, for example, Duffee (2002)) is to use more yields than state variables, and to assume that certain benchmark yields are observed precisely, whereas the other yields are observed with measurement error. The measurement errors are generally held to be i.i.d., and also independent of the state variable processes.

We take this latter approach, and use $N + H$ observed yields, $H \geq 0$, in the postulated model, and include observation errors for H of those yields. At each date t, the state vector X_t is then exactly identified by the yields observed without error, and these N yields jointly follow a Markov process. Denoting the times to maturity of the yields observed without error as $\tau_1, ..., \tau_N$, the observed values of these yields, on the left-hand side, are equated with the predicted values (from (17.61)) given the model parameters and the current values of the state variables, X_t:

$$g_t = \Gamma_0(\theta) + \Gamma(\theta)' X_t. \tag{17.63}$$

The current value of the state vector X_t is obtained by inverting this equation:

$$X_t = \left[\Gamma(\theta)'\right]^{-1}\left[g_t - \Gamma_0(\theta)\right]. \tag{17.64}$$

While the only parameters entering the transformation from observed yields to the state variables are the parameters of the risk-neutral (or Q-measure) dynamics of the state variables, once we have constructed our time series of values of X_t sampled at dates $\tau_0, \tau_1, ..., \tau_n$, the dynamics of the state variable that we will be able to infer from this time series are the dynamics under the physical measure (denoted by P). The first step in the estimation procedure is the only place where we rely on the tractability of the affine bond pricing model. In particular, we can now specify freely (that is, without regard for considerations of analytical tractability) the market prices of risk of the different Brownian motions

$$\begin{aligned} dX_t &= \mu^P(X_t;\theta)\,dt + \sigma(X_t;\theta)\,dW_t^P \\ &= \left\{\mu^Q(X_t;\theta) + \sigma(X_t;\theta)\,\Lambda(X_t;\theta)\right\}dt + \sigma(X_t;\theta)\,dW_t^P. \end{aligned} \tag{17.65}$$

We adopt the simple specification for the market price of risk

$$\Lambda(X_t;\theta) = \sigma(X_t;\theta)'\lambda \tag{17.66}$$

with λ an $m \times 1$ vector of constant parameters, so that under P, the instantaneous drift of each state variables is its drift under the risk-neutral measure, plus a constant times its volatility squared. Under this specification, the drift of the state vector is then affine under both the physical and risk-neutral measures, since

$$\mu^P(X_t;\theta) = \left(\tilde{A} + \widetilde{B}X_t\right) + \Sigma S(X_t;\beta)'\,\Sigma'\lambda \equiv A + BX_t. \tag{17.67}$$

An affine μ^P is not required for our likelihood expansions. Since we can derive likelihood expansions for arbitrary diffusions, μ^P may contain terms that are non-affine, such as the square root of linear functions of the state vector, as in Duarte (2004) for instance. Duffee (2002) and Cheridito, Filipović, and Kimmel (2005) also allow for a more general market price of risk specifications than Dai and Singleton (2000), but retain the affinity of μ^Q and μ^P (and also of the diffusion matrix). However, we do rely on the affine character of the dynamics under Q because those allow us to go from state to yields in the tractable manner given by (17.64).

These closed form likelihood expansions are used in various contexts by Thompson (2004), Takamizawa (2005) and Schneider (2006) for interest rate and term structure models of affine or more general type.

17.8. Application to Specification Testing

Aït-Sahalia, Fan, and Peng (2005) develop a specification test for the transition density of the process, based on a direct comparison of the nonparametric estimate of the transition function to the parametric transition function $p_X(x|x_0, \Delta; \theta)$ implied by the model in order to test

$$\begin{aligned} &H_0 : p_X(x|x_0, \Delta) = p_X(x|x_0, \Delta; \theta) \\ \text{vs.} \quad &H_1 : p_X(x|x_0, \Delta) \neq p_X(x|x_0, \Delta; \theta). \end{aligned} \tag{17.68}$$

As in the parametric situation of (17.2), note that the logarithm of the likelihood function of the observed data $\{X_1, \cdots, X_{n+\Delta}\}$ is

$$\ell(p_X) = \sum_{i=1}^{N} \ln p_X(X_{i\Delta}|X_{(i-1)\Delta}, \Delta),$$

after ignoring the stationary density of X_0. A natural test statistic is then to compare the likelihood ratio under the null and alternative hypotheses. This leads to the test statistic

$$\begin{aligned} T_0 = \sum_{i=1}^{N} \ln \left(\widehat{p}_X(X_{i\Delta}|X_{(i-1)\Delta}, \Delta) / p_X(X_{i\Delta}|X_{(i-1)\Delta}, \Delta; \widehat{\theta}) \right) \\ \times w(X_{(i-1)\Delta}, X_{i\Delta}) \end{aligned} \tag{17.69}$$

where w is a weight function, $\widehat{p}_X$ a nonparametric estimator of the transition function based on locally linear polynomials (see Fan,Yao and Tong (1996)) and $p_X(\cdot, \widehat{\theta})$ a parametric estimator based on the closed form expressions described above. Aït-Sahalia, Fan, and Peng (2005) consider other distance measures and tests, and derive their asymptotic properties.

A complementary approach to this is the one proposed by Hong and Li (2005), who use the fact that under the null hypothesis, the random variables $\{P(X_{i\Delta}|X_{(i-1)\Delta}, \Delta, \theta)\}$ are a sequence of i.i.d. uniform random variables; see also Thompson (2004), Chen and Gao (2004) and Corradi and Swanson (2005). That approach will only work in the univariate case, however, unlike one based on (17.69).

17.9. Derivative Pricing Applications

Consider a generic derivative security with payoff function $\Psi\left(X_\Delta\right)$ at time Δ. If the derivative is written on a traded underlying asset, with price process X and risk-neutral dynamics

$$dX_t/X_t = \{r-q\}\,dt + \sigma\left(X_t;\theta\right)dW_t \tag{17.70}$$

where r is the risk-free rate, q the dividend yield paid by that asset, both viewed as constant, then with complete markets, absence of arbitrage opportunities implies that the price at time 0 of the derivative is

$$\begin{aligned} P_0 &= e^{-r\Delta}E\left[\Psi\left(X_\Delta\right)\middle|\,X_0=x_0\right] \\ &= e^{-r\Delta}\int_0^{+\infty}\Psi\left(x\right)p_X\left(\Delta,x|x_0;\theta\right)dx. \end{aligned} \tag{17.71}$$

In general, the transition function p_X corresponding to the dynamics (17.70) is unknown and one will either solve numerically the PDE solved by P or perform Monte Carlo integration of (17.71).

But we can instead, as long as Δ is not too large, use the $p_X^{(J)}$ corresponding to the SDE (17.70), and get a closed form approximation of the derivative price in the form

$$P_0^{(J)} = e^{-r\Delta}\int_0^{+\infty}\Psi\left(x\right)p_X^{(J)}\left(\Delta,x|x_0;\theta\right)dx \tag{17.72}$$

which is of a different nature than the ad hoc "corrections" to the Black-Scholes Merton formula (as in for example Jarrow and Rudd (1982)), which break the link between the derivative price and the dynamic model for the underlying asset price by assuming directly a functional form for p_X. By contrast, (17.72) is the option pricing formula (of order J in Δ) which matches the dynamics of the underlying asset. Being in closed form, comparative statics, etc. are possible. Being an expansion in small time, accuracy will be limited to relatively small values of Δ (of the order of 3 months in practical applications.)

17.10. Likelihood Inference for Diffusions under Nonstationarity

There is an extensive literature applicable to discrete-time stationary Markov processes starting with the work of Billingsley (1961). The asymptotic covariance matrix for the ML estimator is the inverse of the score covariance or information matrix where the score at date t is $\partial \ln p(X_{t+\Delta}|X_t, \Delta, \theta)/\partial\theta$ where $\ln p(\cdot|x, \Delta, \theta)$ is the logarithm of the conditional density over an interval of time Δ and a parameter value θ. In the stationary case, the MLE will under standard regularity conditions converge at speed $n^{1/2}$ to a normal distribution whose variance is given by the inverse of Fisher's information matrix.

When the underlying Markov process is nonstationary, the score process inherits this nonstationarity. The rate of convergence and the limiting distribution of the maximum likelihood estimator depends upon growth properties of the score process (e.g. see Hall and Heyde (1980) Chapter 6.2). A nondegenerate limiting distribution can be obtained when the score process behaves in a sufficiently regular fashion. The limiting distribution can be deduced by showing that general results pertaining to time series asymptotics (see e.g., Jeganathan (1995)) can be applied to the present context. One first establishes that the likelihood ratio has the locally asymptotically quadratic (LAQ) structure, then within that class separates between the locally asymptotically mixed Normal (LAMN), locally asymptotically Normal (LAN) and locally asymptotically Brownian functional (LABF) structures. As we have seen, when the data generating process is stationary and ergodic, the estimation is typically in the LAN class. The LAMN class can be used to justify many of the standard inference methods given the ability to estimate the covariance matrix pertinent for the conditional normal approximating distribution. Rules for inference are special for the LABF case. These structures are familiar from the linear time series literature on unit roots and co-integration. Details for the case of a nonlinear Markov process can be found in Aït-Sahalia (2002).

As an example of the types of results that can be derived, consider the Ornstein-Uhlenbeck specification, $dX_t = -\kappa X_t dt + \sigma dW_t$, where $\theta = (\kappa, \sigma^2)$. The sampled process is a first-order scalar autoregression, which has received extensive attention in the literature on time series. The discrete-time process obtained by sampling at a fixed interval Δ is a Gaussian first-order autoregressive process with autoregressive parameter $\exp(-\kappa\Delta)$ and innovation variance $\sigma^2 \left(1 - e^{-2\kappa\Delta}\right)/(2\kappa)$. White (1958) and Anderson

(1959) originally characterized the limiting distribution for the discrete-time autoregressive parameter when the Markov process is not stationary. Alternatively, by specializing the general theory of the limiting behavior of the ML estimation to this model, one obtains the following asymptotic distribution for the MLE of the continuous-time parameterization (see Corollary 2 in Aït-Sahalia (2002)):

- If $\kappa > 0$ (LAN, stationary case):

$$\sqrt{N}\left(\begin{pmatrix}\widehat{\kappa}_N\\ \widehat{\sigma}_N^2\end{pmatrix}-\begin{pmatrix}\kappa\\ \sigma^2\end{pmatrix}\right)\longrightarrow N\left(\begin{pmatrix}0\\0\end{pmatrix},\right.$$
$$\left.\begin{pmatrix}\frac{e^{2\kappa\Delta}-1}{\Delta^2} & \frac{\sigma^2\left(e^{2\kappa\Delta}-1-2\kappa\Delta\right)}{\kappa\Delta^2}\\ \frac{\sigma^2\left(e^{2\kappa\Delta}-1-2\kappa\Delta\right)}{\kappa\Delta^2} & \frac{\sigma^4\left(\left(e^{2\kappa\Delta}-1\right)^2+2\kappa^2\Delta^2\left(e^{2\kappa\Delta}+1\right)+4\kappa\Delta\left(e^{2\kappa\Delta}-1\right)\right)}{\kappa^2\Delta^2\left(e^{2\kappa\Delta}-1\right)}\end{pmatrix}\right) \tag{17.73}$$

- If $\kappa < 0$ (LAMN, explosive case), assume $X_0 = 0$, then:

$$\frac{e^{-(N+1)\kappa\Delta}\Delta}{e^{-2\kappa\Delta}-1}\left(\widehat{\kappa}_N-\kappa\right)\to G^{-1/2}\times N\left(0,1\right)$$
$$\sqrt{N}\left(\widehat{\sigma}_N^2-\sigma^2\right)\to N\left(0,2\sigma^4\right) \tag{17.74}$$

  where G has a $\chi^2[1]$ distribution independent of the $N(0,1)$. $G^{-1/2}\times N\left(0,1\right)$ is a Cauchy distribution.

- If $\kappa = 0$ (LABF, unit root case), assume $X_0 = 0$, then:

$$N\,\widehat{\kappa}_N\to\left(1-W_1^2\right)\left(2\Delta\int_0^1 W_t^2dt\right)^{-1}$$
$$\sqrt{N}\left(\widehat{\sigma}_N^2-\sigma^2\right)\to N\left(0,2\sigma^4\right) \tag{17.75}$$

  where N is the sample size and $\{W_t : t \geq 0\}$ is a standard Brownian motion.

Acknowledgments

Financial support from the NSF under grants SES-0350772 and DMS-0532370 is gratefully acknowledged.

References

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies* **9**, 385–426.

Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance* **54**, 1361–1395.

Aït-Sahalia, Y. (2001). Closed-form likelihood expansions for multivariate diffusions. Tech. rep., Princeton University.

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

Aït-Sahalia, Y., Fan, J., Peng, H. (2005). Nonparametric transition-based tests for diffusions. Tech. rep., Princeton University.

Aït-Sahalia, Y., Kimmel, R. (2002). Estimating affine multifactor term structure models using closed-form likelihood expansions. Tech. rep., Princeton University.

Aït-Sahalia, Y., Kimmel, R. (2004). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics* forthcoming.

Aït-Sahalia, Y., Yu, J. (2005). Saddlepoint approximations for continuous-time markov processes. *Journal of Econometrics* forthcoming.

Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics*, **30** 676–687.

Bakshi, G. S., Ju, N. and Qu-Yang, H. (2006). Estimation of continuous-time models with an application to equity volatility dynamics, *Journal of Financial Economics*, forthcoming.

Bakshi, G. S., Yu, N. (2005). A refinement to Aït-Sahalia's (2002) "Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach". *Journal of Business* **78**, 2037–2052.

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

Black, F., Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* **81**, 637–654.

Brandt, M., Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics* **63**, 161–210.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* **48**, 1209–1227.

Chen, S. X., Gao, J. (2004). An adaptive empirical likelihood test for time series models. Tech. rep., Iowa State University.

Cheridito, P., Filipović, D., Kimmel, R. (2005). Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics* forthcoming.

Constantinides, G. M. (1992). A theory of the nominal term structure of interest rates. *Review of Financial Studies* **5**, 531–552.

Corradi, V., Swanson, N. R. (2005). A bootstrap specification test for diffusion processes. *Journal of Econometrics*, forthcoming.

Courtadon, G. (1982). The pricing of options on default free bonds. *Journal of Financial and Quantitative Analysis* **17**, 75–100.

Cox, J. C. (1975). The constant elasticity of variance option pricing model. The Journal of Portfolio Management, Special Issue published in 1996 , 15–

17.
Cox, J. C., Ingersoll, J. E., Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica* **53**, 385–408.
Dai, Q., Singleton, K. J. (2000). Specification analysis of affine term structure models. *Journal of Finance* **55**, 1943–1978.
Daniels, H. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics* **25**, 631–650.
DiPietro, M. (2001). Bayesian inference for discretely sampled diffusion processes with financial applications. Ph.D. thesis, Department of Statistics, Carnegie-Mellon University.
Duarte, J. (2004). Evaluating an alternative risk preference in affine term structure models. *Review of Financial Studies* **17**, 379–404.
Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance* **57**, 405–443.
Duffie, D., Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance* **6**, 379–406.
Egorov, A. V., Li, H., Xu, Y. (2003). Maximum likelihood estimation of time inhomogeneous diffusions. *Journal of Econometrics* **114**, 107–139.
Fan, J., Yao, Q., Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* **83**, 189–206.
Hall, P., Heyde, C. C. (1980). Martingale Limit Theory and Its Application. Academic Press, Boston.
Hansen, L. P., Sargent, T. J. (1983). The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* **51**, 377–387.
Harrison, M., Kreps, D., 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* **20**, 381–408.
Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bonds and currency options. *Review of Financial Studies* **6**, 327–343.
Hong, Y., Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies* **18**, 37–84.
Hull, J., White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance* **42**, 281–300.
Hurn, A. S., Jeisman, J., Lindsay, K. (2005). Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. Tech. rep., School of Economics and Finance, Queensland University of Technology.
Jarrow, R., Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics* **10**, 347–369.
Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* **11**, 818–887.
Jensen, B., Poulsen, R. (2002). Transition densities of diffusion processes: Numerical comparison of approximation techniques. *Journal of Derivatives* **9**, 1–15.
Lewis, A. L. (2000). Option Valuation under Stochastic Volatility. Finance Press,

Newport Beach, CA.

Li, M. (2005). A damped diffusion framework for financial modeling and closed-form maximum likelihood estimation. Tech. rep. Georgia Institute of Technology.

Lo, A. W. (1988). Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory* **4**, 231–247.

Marsh, T., Rosenfeld, E. (1982). Stochastic processes for interest rates and equilibrium bond prices. *Journal of Finance* **38**, 635–646.

Mosburger, G., Schneider, P., 2005. Modelling international bond markets with affine term structure models. Tech. rep., Vienna University of Economics and Business Administration.

Pedersen, A. R. (1995). A new approach to maximum-likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian journal of Statistics* **22**, 55–71.

Philips, P. C. B. (1973). The problem of identification in finite parameter continuous time models. *Journal of Econometrics* **1**, 351–362.

Romano, M., Touzi, N. (1997). Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance* **7**, 399–412.

Schaumburg, E. (2001). Maximum likelihood estimation of jump processes with applications to finance. Ph.D. thesis, Princeton University.

Schneider, P. (2006). Approximations of transition densities for nonlinear multivariate diffusions with an application to dynamic term structure models. Tech. rep. Vienna University of Economics and Business Administration.

Stramer, O., Yan, J. (2005). On simulated likelihood of discretely observed diffusion processes and comparison to closed-from approximation. Tech. rep., University of Iowa.

Takamizawa, H. (2005). Is nonlinear drift implied by the short-end of the term structure? Tech. rep., Hitotsubashi Uninversity.

Thompson, S. (2004). Identifying term structure volatility from the LIBOR-swap curve. Tech. rep., Harvard University.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* **5**, 177–188.

White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* **29**, 1188–1197.

Wong, E. (1964). The construction of a class of stationary Markoff processes. In: Bellman, R. (Ed.), Sixteenth Symposium in Applied Mathematics - Stochastic Processes in Mathematical Physics and Engineering. American Mathematical Society, Providence, RI, 264 – 276.

Yu, J. (2003). Closed-form likelihood estimation of jump-diffusions with an application to the realignment risk premium of the Chinese yuan. Ph.D. thesis, Princeton University.