CLOSED-FORM LIKELIHOOD EXPANSIONS FOR MULTIVARIATE DIFFUSIONS

Yacine Ait-Sahalia

Closed-Form Likelihood Expansions for Multivariate Diffusions
Yacine Ait-Sahalia
NBER Working Paper No. 8956
May 2002
JEL No. C32, G12

## **ABSTRACT**

This paper provides closed-form expansions for the transition density and likelihood function of arbitrary multivariate diffusions. The expansions are based on a Hermite series, whose coefficients are calculated explicitly by exploiting the special structure afforded by the diffusion hypothesis. Because the transition function for most diffusion models is not known explicitly, the expansions of this paper can help make maximum-likelihood a practical estimation method for discretely sampled multivariate diffusions. Examples of interest in financial econometrics are included.

Yacine Ait-Sahalia
Department of Economics
Princeton University
Princeton, NJ 08544-1021
and NBER
(609) 258-4015
yacine@princeton.edu

# 1    Introduction

Diffusions and more generally continuous-time Markov processes are generally specified in economics and finance by their evolution over infinitesimal instants, that is, by writing down the stochastic differential equation followed by the state vector. However, for most estimation techniques relying on discrete data, we need to be able to infer the implications of the infinitesimal time evolution of the process for longer time intervals, for instance the time interval at which the process is actually sampled, say daily or weekly. The transition function plays a key role in that context. The transition function of a Markov process is the conditional density for the values of the state variable at a fixed future date, given the current level of the state vector. It effectively gives a precise answer to the time aggregation problem inherent in the dichotomy between the time scale of the model (continuous) and that of the observed data (discrete): if the process evolves at each instant according to a given infinitesimal continuous-time equation, what is the distribution of the values of the process after a finite amount of time has elapsed?

Continuous-time models in finance have long been predominantly univariate, whether the variable in an asset price as in the Black-Scholes and Merton models, or an interest rate as in the Cox, Ingersoll and Ross or Vasicek models. In recent years, however, the literature has naturally evolved towards the inclusion of multiple variables in continuous-time diffusion models. Typical examples include asset pricing models with multiple explanatory factors, term structure models with multiple yields or factors, and stochastic volatility or stochastic mean reversion models (see Sundaresan (2000) for a recent survey).

In response to this trend towards multivariate models, this paper describes the construction of closed-form approximations to the transition density of arbitrary multivariate diffusions, thereby extending to the multivariate setting the results of Aït-Sahalia (2002). The form of the likelihood expansions derived here is based on Hermite polynomials. While writing down a Hermite series can be done for any model, the key idea is to exploit the specificity afforded by the diffusion hypothesis in order to obtain the expressions for the coefficients of the series fully explicitly, as functions of the state vectors at the present and future dates, the time interval that separates them and the parameters of the assumed stochastic differential equation. Other methods can be used to approximate the transition function, which involve solving numerically the Fokker-Planck-Kolmogorov equation, simulating the process to Monte Carlo integrate the transition density or approximating the process with binomial trees (see Aït-Sahalia (2002) for a review of the literature, and Jensen and Poulsen (2002) for a comparison of the different methods). None however produces a closed form approximation.

The extension from the univariate to the multivariate setting presents many challenges. Through judicious use of Itô's Lemma, every univariate diffusion can be transformed into one with unit diffusion, whose density can then be approximated around a standard Normal. This is no longer the case for multivariate diffusions. I therefore introduce the concept of reducibility for multivariate diffusions, which essentially characterizes diffusions for which such a transformation exists. For reducible multivariate diffusions, the ideas introduced in the univariate setting can be extended, leading to an expansion for the log-likelihood function in the form of a Taylor series in the time variable, which is a particularly convenient way of gathering the Hermite terms. For irreducible diffusions, however, one must proceed differently. The situation is more involved, yet still amenable

to a closed-form result, but this time in the form of a double Taylor expansion in the time variable and the state vector. Extensions of the results of Aït-Sahalia (2002) in two different univariate directions have also recently been developed, for time-inhomogenous diffusions (Egorov, Li, and Xu (2001)) and for models driven by Lévy processes other than Brownian motion (Schaumburg (2001)).

Once the expansion is computed for the diffusion model at hand, it can be immediately applied to the estimation of parameters of the discretely sampled diffusion by maximum-likelihood, or to a variety of other estimation methods which require an expression for the transition density of the state variables, such as Bayesian methods where one wishes to obtain a posterior distribution for the parameters of a stochastic differential equation. The method can also be applied to generate simulated data at the desired frequency from the continuous-time model, or to serve as the instrumental or auxiliary model in indirect inference and simulated or efficient moments methods. The point is that the explicit nature of the expansion as a function of all the relevant variables makes these computations, whether maximization of the classical likelihood or computation of posterior distributions, straightforward and computationally very efficient.

The paper is organized as follows. Section 2 sets up the model, notation and assumptions. In Section 3, I introduce the concept of reducibility of a diffusion and provide a necessary and sufficient condition for the reducibility of a multivariate diffusion. When diffusions are reducible, the coefficients of the expansion are obtained by a change of variable, which I show in Section 4. When the diffusion is not reducible, the expressions for the coefficients are given in Section 5. Section 6 contains examples of multivariate diffusions relevant for financial econometrics and gives their corresponding likelihood expansions. Finally, Section 7 concludes. All proofs are in the Appendix.

## 2   Setup and Assumptions

Consider the multivariate diffusion

$$dX_t = \mu\left(X_t; \theta\right) dt + \sigma\left(X_t; \theta\right) dW_t \tag{2.1}$$

where $X_t$ and $\mu\left(X_t; \theta\right)$ are $m \times 1$ vectors, $\sigma\left(X_t; \theta\right)$ is an $m \times m$ matrix, $\theta$ is a $p$-dimensional parameter and $W_t$ is an $m \times 1$ vector of independent Brownian motions. Independence of the components is without loss of generality as arbitrary correlation structures between the shocks to the different equations can be modelled through the inclusion of off-diagonal terms in the $\sigma$ matrix. Note that $\sigma$ need not be symmetric, and if convenient attention can be restricted to triangular matrices by appropriate rotation of the $m-$dimensional Brownian motion.

The objective of this paper is to derive closed-form approximations to the transition function $p_X\left(\Delta, x | x_0; \theta\right)$ of the process $X$, that is the conditional density of $X_{t+\Delta} = x$ given $X_t = x_0$ induced by the model (2.1). Assume that we observe the process at dates $\{t = i\Delta \,|\, i = 0, \ldots, n\}$, where $\Delta > 0$ is fixed. Bayes' rule combined with the Markovian nature of (2.1), which the discrete data inherit, imply that the log-likelihood

function has the simple form

$$\ell_n(\theta) \equiv n^{-1} \sum_{i=1}^{n} l_X\left(\Delta, X_{i\Delta} | X_{(i-1)\Delta}; \theta\right) \tag{2.2}$$

where $l_X \equiv \ln p_X$. In practice, the issue is that for most models of interest, the function $p_X$, hence $l_X$, is not available in closed-form.

If the sampling interval $\Delta$ is time-varying deterministically, say $\Delta_i$ is the actual time interval between the $(i-1)^{th}$ and $i^{th}$ observations $\tilde{X}_{i-1}$ and $\tilde{X}_i$, then it suffices to replace $\Delta$ in (2.2) by its actual value $\Delta_i$ when evaluating the transition density for the $i^{th}$ pair of observations. If the sampling interval is random and either drawn independently of the $X$ process or conditionally on $\tilde{X}_{i-1}$, then one can write down the joint likelihood function of the pair of observations and $\Delta_i$ and utilize Bayes' rule to express it as the product of the conditional density of $\tilde{X}_i$ given $\tilde{X}_{i-1}$ *and* $\Delta_i$, times the marginal density $d$ of $\Delta_i$ given $\tilde{X}_{i-1}$, that is $p_X\left(\Delta_i, \tilde{X}_i | \tilde{X}_{i-1}; \theta\right) \times d\left(\Delta_i | \tilde{X}_{i-1}; \kappa\right)$ where $\kappa$ is a parameter vector parametrizing the sampling density $d$. Aït-Sahalia and Mykland (2000) study the different effects resulting in the likelihood framework from randomly and discretely spaced observations. In all cases, an expression is needed for $l_X$, which is what this paper delivers.

I will use the following notation. Let $\mathcal{S}_X$, a subset of $\mathbb{R}^m$, denote the domain of the diffusion $X$ and $\Theta \subset \mathbb{R}^p$ the open parameter space. $\mathcal{S}_X$ can often be taken to be of the form of a product of $m$ intervals with limits $\underline{x}_i$ and $\bar{x}_i$, where possibly $\underline{x}_i = -\infty$ and/or $\bar{x}_i = +\infty$. The intervals are closed at finite limits and open at infinite limits. For simplicity, I will assume that $\Theta$ is such that $\mathcal{S}_X$ is identical for each value of the parameter vector $\theta$ in $\Theta$. I will use $^T$ to denote transposition and, for a function $\eta(x;\theta) = (\eta_1(x;\theta), ..., \eta_d(x;\theta))^T$, differentiable in $x$, I will write $\nabla\eta(x;\theta)$ for the Jacobian matrix of $\eta$, i.e., the matrix $\nabla\eta(x;\theta) = [\partial\eta_i(x;\theta)/\partial x_j]_{i=1,...,d;j=1,...,m}$. For $x \in \mathbb{R}^m$, $\|x\|$ denotes the usual Euclidean norm. The binomial coefficients will be denoted

$$\binom{k}{j} \equiv \frac{k!}{j!(k-j)!}. \tag{2.3}$$

If $a = [a_{ij}]_{i,j=1,...,m}$ is a $m \times m$ invertible matrix then I write $a^{-1} = [a_{ij}^{-1}]_{i,j=1,...,m}$ for the matrix inverse, rather than using the tensor notation (note that $a_{ij}^{-1}$ denotes the element $(i,j)$ of the inverse matrix, not the inverse of the element $(i,j)$ of the original matrix). $Det[a]$ and $tr[a]$ denote the determinant of $a$ and its trace, respectively. If $a = [a_i]_{i=1,...,m}$ is a vector, $tr[a]$ denotes the sum of the elements of $a$. $a = diag[a_i]_{i=1,...,m}$ denotes the $m \times m$ diagonal matrix with diagonal elements $a_i$. When a function $\eta(x;\theta)$ is invertible in $x$, I write $\eta^{\text{inv}}(y;\theta)$ for its inverse, i.e., the solution in $x$ of the equation $y = \eta(x;\theta)$ is $x = \eta^{\text{inv}}(y;\theta)$. By the Inverse Function Theorem (see e.g., Theorem 8.7.8 in Haaser and Sullivan (1991)), $\eta(x;\theta)$ is invertible in $x$ at $x = x_0$ if $\nabla\eta(x;\theta)$ has a bounded matrix inverse at $x = x_0$; the inverse function $\eta^{\text{inv}}$ then inherits the smoothness properties of $\eta$.

Let $A_X$ denote the infinitesimal generator of the process $X$, which is characterized by its action on functions $f(\Delta, x, x_0; \theta)$ in its domain:

$$A_X \cdot f(\Delta, x, x_0; \theta) = \frac{\partial f(\Delta, x, x_0; \theta)}{\partial \Delta} + \sum_{i=1}^{m} \mu_i(x;\theta) \frac{\partial f(\Delta, x, x_0; \theta)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x;\theta) \frac{\partial^2 f(\Delta, x, x_0; \theta)}{\partial x_i \partial x_j}. \tag{2.4}$$

3

The domain of $A_X$ includes at least functions that, for each $(x_0; \theta) \in \mathcal{S}_X \times \Theta$, are once continuously differentiable in $\Delta$ in $\mathbb{R}_+$, twice continuously differentiable in $x$ in $\mathcal{S}_X$ and have compact support.

In some instances, it may be more natural to parametrize directly the infinitesimal variance-covariance matrix of the process

$$v\left(x;\theta\right) \equiv \sigma\left(x;\theta\right)\sigma^T\left(x;\theta\right), \tag{2.5}$$

that $\sigma(x;\theta)$ itself. In that case, $\sigma\left(x;\theta\right)$ is defined indirectly as the positive definite square root of $v\left(x;\theta\right)$,

$$\sigma\left(x;\theta\right) = v\left(x;\theta\right)^{1/2}.$$

$\sigma$ can be obtained by the Cholesky decomposition of the matrix $v\left(x;\theta\right)$. While it is traditional to parametrize the process by $(\mu, \sigma)$, every characterization of the process, such as its transition probability, depends in fact on $(\mu, v)$. In particular, it can be shown that, should there exist a continuum of solutions in $\sigma$ to the equation (2.5), the transition probability of the process is identical for each one of these $\sigma$ (see Remark 5.17 and Section 5.3 in Stroock and Varadhan (1979)). This is also quite clear from the definition (2.4) of the infinitesimal generator of the process, which is an equivalent characterization of the process, and depends on $v$ rather than $\sigma$. As this will pay a role in the likelihood expansions, define

$$D_v\left(x;\theta\right) \equiv \frac{1}{2}\ln\left(Det[v(x;\theta)]\right). \tag{2.6}$$

To avoid the issues associated with the multiple $\sigma$ scenario, I assume from now on that $\sigma$ is uniquely determined, either directly as part of the assumed specification of the model (2.1) or indirectly as the unique solution of (2.5), in which case the form of $v$ is such that it yields a unique square root matrix. I will assume that this matrix $\sigma$ satisfies the following regularity condition:

**Assumption 1.** *The matrix $\sigma\left(x;\theta\right)$ is positive definite for all $x$ in the interior of $\mathcal{S}_X$ and $\theta \in \Theta$.*

Further assumptions are required to insure the existence and unicity of a solution to (2.1), and to make the computation of likelihood expansions possible. I will assume the following:

**Assumption 2.** *For each $\theta \in \Theta$, $\mu\left(x;\theta\right)$ and $\sigma\left(x;\theta\right)$ are infinitely differentiable in $x$ on $\mathcal{S}_X$.*

Assumption 2 insures the unicity of solutions to (2.1). Indeed, Assumption 2 implies in particular that the coefficients of the stochastic differential equation are locally Lipschitz under their assumed (once) differentiability, by applying the mean value theorem. That is, for each $C > 0$, there exists a constant $K > 0$ such that for every $x$ and $x'$ in $\mathcal{S}_X$, $\|x\| \leq C$ and $\|x'\| \leq C$, we have

$$\left|\mu_i\left(x;\theta\right) - \mu_i\left(x';\theta\right)\right| \leq K\left\|x - x'\right\| \tag{2.7}$$

$$\left|\sigma_{ij}\left(x;\theta\right) - \sigma_{ij}\left(x';\theta\right)\right| \leq K\left\|x - x'\right\| \tag{2.8}$$

for $i, j = 1, ..., m$. This insures that a solution, if it exists, will be unique (see e.g., Theorem 5.2.5 in Karatzas and Shreve (1991)). The infinite differentiability assumption in $x$ is unnecessary for that purpose, but it allows

the computation of expansions of the transition density, which as we will see involve repeated differentiation of the coefficient functions $\mu$ and $\sigma$.

There exist models of interest in finance, such as Feller's square-root diffusion used in the Cox, Ingersoll and Ross model of the term structure, that fail to satisfy (2.8) since they violate the differentiabilty requirement of Assumption 2 at a boundary of $\mathcal{S}_X$ : for instance, $\sigma(x;\theta) = \sigma_0 x^{1/2}$ is not differentiable at the left boundary 0 of $\mathcal{S}_X$. Fortunately, it is possible to weaken Assumption 2 to cover such cases:

**Assumption 3.** *(Yamada-Watanabe Conditions) Assumption 2 can be replaced by:*

*1. For each $\theta \in \Theta$, $\mu(x;\theta)$ and $\sigma(x;\theta)$ are infinitely differentiable in $x$ on the interior of $\mathcal{S}_X$.*

*2. There exist real-valued, continuous, positive and increasing functions $\rho(u)$ and $\kappa(u)$ defined on $[0,C)$ for some $C > 0$ such that $\rho(0) = \kappa(0) = 0$, $\rho^2(u)u^{-1}$ and $\kappa(u)$ are concave and satisfy*

$$\lim_{\varepsilon \to 0^+} \int_\varepsilon^C \frac{u}{\rho^2(u)} du = +\infty \tag{2.9}$$

$$\lim_{\varepsilon \to 0^+} \int_\varepsilon^C \frac{1}{\kappa(u)} du = +\infty. \tag{2.10}$$

*Then*

$$|\mu_i(x;\theta) - \mu_i(x';\theta)| \leq \kappa(\|x - x'\|) \tag{2.11}$$

$$|\sigma_{ij}(x;\theta) - \sigma_{ij}(x';\theta)| \leq \rho(\|x - x'\|) \tag{2.12}$$

*for all $(x,y) \in \mathcal{S}_X^2$ such that $\|x - x'\| < C$ and all $i,j = 1,...,m$.*

*3. If $\sigma(x;\theta)$ is of the form $\sigma(x;\theta) = diag\left[\sigma_i(x_i;\theta)\right]_{i=1,...,m}$ (this is always the case if $m = 1$), condition (2.9) can be weakened to*

$$\lim_{\varepsilon \to 0^+} \int_\varepsilon^C \frac{1}{\rho^2(u)} du = +\infty \tag{2.13}$$

*with no concavity requirement.*

*4. If $m = 2$ and $\sigma(x;\theta)$ is of the isotropic form $\sigma(x;\theta) = diag\left[s(x;\theta)\right]_{i=1,2}$ then condition (2.9) can be weakened to*

$$\lim_{\varepsilon \to 0^+} \int_\varepsilon^C \frac{u\ln(1/u)}{\rho^2(u)} du = +\infty \tag{2.14}$$

*provided that $G(u) = u^3 \exp(2/u)\rho^2(\exp(-1/u))$ is concave.*

As in the case of Assumption 2, Assumption 3.1 is there for the purpose of computing likelihood expansions. The fact that Assumption 3.2 insures unicity of the solution follows from Theorem 4 in Watanabe and Yamada (1971); Assumption 3.3 from Theorem 1 in Yamada and Watanabe (1971); Assumption 3.4 from Theorem 3 in Watanabe and Yamada (1971). Examples of functions $\rho$ that satisfy (2.9) are: $\rho(u) = u^\alpha$ with $\alpha \geq 1$, $\rho(u) = u(\ln(1/u))^{1/2}$. The functions $\rho(u) = u^\alpha$ with $\alpha \geq 1/2$ satisfy (2.13). The functions $\rho(u) = u^\alpha$ with $\alpha \geq 1/2$ and $\rho(u) = u\ln(1/u)$ satisfy (2.14). A function $\sigma_{ij}$ satisfying condition (2.12) with $\rho(u) = u^\alpha$ is said to be Hölder-continuous of order $\alpha$.

Assumption 3.3 with $\rho(u) = u^{1/2}$ allows us in particular to consider mutivariate Cox, Ingersoll and Ross models where $\sigma(x; \theta) = diag\left[\eta_i x_i^{1/2}\right]_{i=1,...,m}$ (see the term structure examples in Section 6.3). The issue with these affine models (linear $\mu$ and $v = \sigma\sigma^T$) lies in the non-Lipschitz behavior of the $\sigma$ function rather than that of the $\mu$ function. In that case, Assumption 3 for $\mu$ with $\kappa(u) = k.u$ reduces to the Lipschitz condition (2.7) for the drift $\mu$.

These conditions are essentially the best possible, in that examples where multiple solutions to the stochastic differential equation (2.1) arise when they are violated. If $m \geq 3$, take any subadditive $\rho(u)$ (i.e., $\rho(u + v) \leq \rho(u) + \rho(v)$) such that

$$\lim_{\varepsilon \to 0^+} \int_\varepsilon^C \frac{u}{\rho^2(u)} du < +\infty,$$

for instance $\rho(u) = u^{1/2}$, then the stochastic differential equation $dX_t = \sigma(X_t)dW_t$, $X_0 = 0$, with isotropic $\sigma$ matrix $\sigma(x) = diag\left[\rho\left(\|x\|\right)\right]_{i=1,...,m}$, has, apart from the solution $X_t = 0$, other non-zero solutions. Thus condition (2.9) in Assumption 3.2 is sharp. In dimension $m = 1$, the famous example of Girsanov, $dX_t = |X_t|^\alpha dW_t$, has a unique solution if $\alpha \geq 1/2$, namely $X_t = 0$, but that solution is no longer unique if $0 < \alpha < 1/2$; hence condition (2.13) in Assumption 3.3 is also sharp. In Assumption 3.4 concerning the dimension $m = 2$, the restriction that the matrix $\sigma(x; \theta)$ be of the isotropic form cannot be relaxed: a counterexample was provided recently in Swart (2001). The condition (2.14) is also seen to be sharp, by forming a counterexample with a subadditive $\rho$ as in dimension $m \geq 3$.

The next assumption restricts the growth behavior of the coefficients near the boundaries of the domain:

**Assumption 4.** *The drift and diffusion functions satisfy linear growth conditions, that is, for each $\theta \in \Theta$ there exists a constant $K$ such that for all $x \in \mathcal{S}_X$, and $i, j = 1, ..., m$ :*

$$|\mu_i(x; \theta)| \leq K(1 + \|x\|) \tag{2.15}$$
$$|\sigma_{ij}(x; \theta)| \leq K(1 + \|x\|). \tag{2.16}$$

The role of Assumption 4 is to insure existence of a solution to the stochastic differential equation (2.1) by preventing explosions of the process in finite expected time. While it can be relaxed in specific examples, it is not possible to do so in full generality as shown by the following counterexamples, illustrating the need for restricting the growth of both $\mu$ and $\sigma$. The one-dimensional equation $dX_t = (1 + X_t^2)dt$, $X_0 = 0$, has the exploding solution $X_t = \tan(t)$. The three-dimensional equation $dX_t = (1 + \|X_t\|^2)dW_t$ explodes in finite time. In dimension one, however, finer results are available (see the Engelbert-Schmidt criterion in Theorem 5.5.15 in Karatzas and Shreve (1991)) allowing linear growth to be imposed only when the drift coefficient pulls the process towards an infinity boundary (see Proposition 1 of Aït-Sahalia (2002)). Even in higher dimensions, the condition can sometimes be refined in specific examples (see Section 6.2 below). In all dimensions, the linear growth condition in Assumption 4 is only an issue near the boundaries of $\mathcal{S}_X$. On any compact set, the growth condition (boundedness, in fact) follows from differentiability of the functions and the mean value theorem.

While nothing in this paper hinges upon the stationarity of the process $X$, it is useful to have a sufficient

6

condition that would guarantee it, if need be. From Hasminskii (1980), for given $\theta \in \Theta$, there exists a unique stationary distribution for the process $X$ if there exists $C > 0$ and some positive definite matrix $V$ such that

$$\mu(x;\theta)Vx + \frac{1}{2}tr\left[v\left(x;\theta\right)V\right] < -1 \tag{2.17}$$

for all $x$ in $\mathcal{S}_X$ such that $\|x\| > C$. Then the stationary density of $X$ is the solution $\pi\left(x_0;\theta\right)$ of the equation

$$\sum_{i=1}^{m} \frac{\partial}{\partial x_{0i}} \left[\mu_i\left(x_0;\theta\right)\pi\left(x_0;\theta\right)\right] - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial^2}{\partial x_{0i}\partial x_{0j}} \left[v_{ij}\left(x_0;\theta\right)\pi\left(x_0;\theta\right)\right] = 0 \tag{2.18}$$

that integrates to one. The process $X$ will be stationary provided that the initial random variable $X_0$ is distributed with density $\pi\left(x_0;\theta\right)$. Of course, the process may be stationary for some values of $\theta$ in $\Theta$ and not others. For example, in an Ornstein-Uhlenbeck process stationarity depends upon the positivity of the real parts of the eigenvalues of the mean reversion matrix.

If the approximation to the function $l_X$ is to be used for maximum-likelihood estimation of the parameters $\theta$, then care must be taken to insure that all the parameters are identified. The MLE is well-defined and identification is achieved if we assume:

**Assumption 5.** *For each $x \in \mathcal{S}_X$, $\mu\left(x;\theta\right)$ and $\sigma\left(x;\theta\right)$ are three times continuously differentiable in $\theta$ on $\Theta$, and, if there exist $(\theta,\theta') \in \Theta^2$ such that $p_X\left(\Delta, x|x_0;\theta\right) = p_X\left(\Delta, x|x_0;\theta'\right)$ on a set of values of $(x, x_0) \in \mathcal{S}_X^2$ of non-zero measure, then $\theta = \theta'$.*

More primitive conditions, not involving the function $p_X$, can be given in specific examples, see Section 6 below. This paper deals only with the construction of an approximation to $l_X$, which can then be used for purposes other than maximum likelihood estimation. In that case, there is no reason to assume Assumption 5.

One last remark. The diffusion process $X$ is fully defined by the specification of the functions $\mu$ and $\sigma$ *and* its behavior at the boundaries of $\mathcal{S}_X$. In many examples, the specification of $\mu$ and $\sigma$ predetermines the boundary behavior of the process, but this will not be the case for models that represent limiting situations. For instance, in Cox, Ingersoll and Ross processes with affine $\mu$ and $v$, the behavior at the 0 boundary depends upon the values of the parameters $\theta$ in $(\mu, \sigma)$. When this situation occurs for a particular model, the behavior of the likelihood expansion near such a boundary will be specified exogneously to match that of the assumed model.

## 3  Reducible Diffusions

Whenever possible, I will first transform the diffusion $X$ into one that is more amenable to the derivation of an expansion for its transition density. For that purpose, I introduce the following definition:

**Definition 1.** *(Reducibility) The diffusion $X$ is said to be reducible to unit diffusion (or reducible, in short) if and if only if there exists a one-to-one transformation of the diffusion $X$ into a diffusion $Y$ whose diffusion matrix $\sigma_Y$ is the identity matrix. That is, there exists an invertible function $\gamma\left(x;\theta\right)$, infinitely differentiable in*

7

$X$ on $\mathcal{S}_X$ and three times continuously differentiable in $\theta$ on $\Theta$ such that $Y_t \equiv \gamma(X_t; \theta)$ satisfies the stochastic differential equation

$$dY_t = \mu_Y(Y_t; \theta)\, dt + dW_t \tag{3.1}$$

on the domain $\mathcal{S}_Y$.

To avoid needless complications, I will assume that the domain of the transformed process, $\mathcal{S}_Y$, is independent of the parameter value $\theta$. As discussed for $\mathcal{S}_X$ already, in typical examples, $\mathcal{S}_X$ and $\mathcal{S}_Y$ are both products of intervals with lower limits $\underline{x}_i$ and $\underline{y}_i$ that are either $-\infty$ or 0, and upper limits $\bar{x}_i$ and $\bar{y}_i$ that are either 0 or $+\infty$.

By Itô's Lemma, when the diffusion is reducible, the change of variable $\gamma$ satisfies

$$\nabla\gamma(X_t; \theta) = \sigma^{-1}(x; \theta). \tag{3.2}$$

Every scalar (i.e., one-dimensional) diffusion is reducible, by means of the transformation

$$Y_t \equiv \gamma(X_t; \theta) = \int^{X_t} \frac{du}{\sigma(u; \theta)} \tag{3.3}$$

and we have by Itô's Lemma:

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

This transformation played a critical role in the derivation of closed-form Hermite approximations to the transition density of univariate diffusions in Aït-Sahalia (2002). However, not every multivariate diffusion is reducible. Whether or not a given multivariate diffusion is reducible depends on the specification of its $\sigma$ matrix, namely:

**Proposition 1.** (Necessary and Sufficient Condition for Reducibility) The diffusion $X$ is reducible if and only if the inverse diffusion matrix $\sigma^{-1} = \left[\sigma_{i,j}^{-1}\right]_{i,j=1,\ldots,m}$ satisfies on $\mathcal{S}_X \times \Theta$ the condition that

$$\frac{\partial\sigma_{ij}^{-1}(x; \theta)}{\partial x_k} = \frac{\partial\sigma_{ik}^{-1}(x; \theta)}{\partial x_j} \tag{3.4}$$

for each triplet $(i, j, k) = 1, \ldots, m$ such that $k > j$.

In the bivariate case $m = 2$, the state vector is $X_t = (X_{1t}, X_{2t})^T$ and the components of the $\mu$ vector and $\sigma$ matrix are

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \end{pmatrix} = \begin{pmatrix} \mu_1(X_t; \theta) \\ \mu_2(X_t; \theta) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11}(X_t; \theta) & \sigma_{12}(X_t; \theta) \\ \sigma_{21}(X_t; \theta) & \sigma_{22}(X_t; \theta) \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \end{pmatrix} \tag{3.5}$$

and condition (3.4) reduces to

$$\frac{\partial\sigma_{11}^{-1}(x; \theta)}{\partial x_2} - \frac{\partial\sigma_{12}^{-1}(x; \theta)}{\partial x_1} = \frac{\partial\sigma_{21}^{-1}(x; \theta)}{\partial x_2} - \frac{\partial\sigma_{22}^{-1}(x; \theta)}{\partial x_1} = 0. \tag{3.6}$$

8

**Example 1.** *Diagonal Systems: If $\sigma_{12} = \sigma_{21} = 0$, then the reducibility condition becomes $\partial\sigma_{11}^{-1}/\partial x_2 = \partial\sigma_{22}^{-1}/\partial x_1 = 0$. Since $\sigma_{ii}^{-1} = 1/\sigma_{ii}$ in the diagonal case, reducibility is equivalent to the fact that $\sigma_{ii}$ depends only on $x_i$ (and $\theta$) for each $i = 1, 2$. This is true more generally in dimension $m$. Note that this is not the case if off-diagonal elements are present.*

Another example is provided by the class of stochastic volatility models:

**Example 2.** *Stochastic Volatility: If*

$$\sigma(x;\theta) = \begin{pmatrix} \sigma_{11}(x_2;\theta) & 0 \\ 0 & \sigma_{22}(x_2;\theta) \end{pmatrix}$$

*then the process is not reducible in light of the previous example, as this is a diagonal system where $\sigma_{11}$ depends on $x_2$. However, if*

$$\sigma(x;\theta) = \begin{pmatrix} a(x_1;\theta) & a(x_1;\theta)b(x_2;\theta) \\ 0 & c(x_2;\theta) \end{pmatrix}$$

*then the process is reducible as can be seen by applying (3.6).*

The situation now is as follows. 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 the reduced process $Y$ and then transforming $Y$ back into $X$, proceeding essentially by extending the univariate method: see Section 4. When a diffusion is not reducible, I explain below how to nevertheless derive a closed-form expansion directly for the transition density $p_X$ : this is done in Section 5.

# 4    Closed-Form Expansion for the Transition Density of a Reducible Diffusion

## 4.1    Form of the Hermite Series in the Univariate Case

As discussed above, every univariate diffusion is reducible. To motivate the approach in the multivariate case, let me first recall how one proceeds in the univariate case, summarizing briefly the results of Aït-Sahalia (2002). To understand the construction of the sequence of approximations to the transition function $p_X$, the following analogy may be helpful. Consider a standardized sum of random variables to which the Central Limit Theorem (CLT) apply. 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. The basic idea is to create an analogy between this situation and that of approximating the transition density of a diffusion. Think of the sampling interval $\Delta$ 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 $\Delta$ tends to 0 (by analogy with what

happens in the CLT when $n$ tends to $\infty$). Properly standardizing the data in the CLT means summing them and dividing by $n^{1/2}$; here it will involve transforming the original diffusion $X$ into another one, called $Z$ below. In both cases, the appropriate standardization makes $N(0,1)$ the leading term. I will then refine this $N(0,1)$ approximation by "correcting" for the fact that $\Delta$ is not 0 (just like in practical applications of the CLT $n$ is not infinity), i.e., by computing the higher order terms. 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.

So let $p_Y$ denote the transition function of the process $Y$, whose dynamics are given by (3.1). As shown in Aït-Sahalia (2002), the tails of $p_Y$ have a Gaussian-like upper bound; but while $Y$ is "closer" to a Normal variable than $X$ is, it is not practical to expand $p_Y$. This is due to the fact that $p_Y$ gets peaked around the conditional value $y_0$ when $\Delta$ 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

$$Z_\Delta \equiv \Delta^{-1/2} \left( Y_\Delta - y_0 \right).$$

I then expand the density of $Z$ around a $N(0,1)$, leading to an expansion for $p_Y$ of the form:

$$\breve{p}_Y^{(J)} \left( \Delta, y | y_0; \theta \right) = \frac{1}{(2\pi\Delta)^{1/2}} \exp \left( -\frac{(y-y_0)^2}{2\Delta} \right) \sum_{j=0}^{J} \eta^{(j)} \left( \Delta, y_0; \theta \right) H_j(\Delta^{-1/2}(y-y_0)) \qquad (4.1)$$

where the Hermite coefficients $\eta^{(j)} \left( \Delta, y_0; \theta \right)$ are given by

$$
\begin{aligned}
\eta^{(j)} \left( \Delta, y_0; \theta \right) &= (1/j!) \int_{-\infty}^{+\infty} H_j \left( z \right) p_Z \left( z | y_0, \Delta; \theta \right) dz \\
&= (1/j!) \int_{-\infty}^{+\infty} H_j \left( z \right) \Delta^{1/2} p_Y \left( \Delta^{1/2} z + y_0 \middle| y_0, \Delta; \theta \right) dz \\
&= (1/j!) \int_{-\infty}^{+\infty} H_j \left( \Delta^{-1/2} \left( y - y_0 \right) \right) p_Y \left( y | y_0, \Delta; \theta \right) dy \\
&= (1/j!) E \left[ H_j \left( \Delta^{-1/2} \left( Y_\Delta - y_0 \right) \right) \middle| Y_0 = y_0; \theta \right].
\end{aligned}
\qquad (4.2)
$$

To evaluate the conditional expectation (4.2), I use the Taylor expansion

$$E_{Y_1} \left[ f(\Delta, Y_\Delta, Y_0; \theta) | Y_0 = y_0 \right] = \sum_{k=0}^{K} \frac{\Delta^k}{k!} A_Y^k \cdot f(\delta, y, y_0; \theta)_{| y=y_0, \delta=0} + O \left( \Delta^{K+1} \right) \qquad (4.3)$$

where $A_Y$ is the infinitesimal generator of the process, i.e., the operator whose action is defined by

$$A_Y \cdot f \left( \Delta, y, y_0; \theta \right) = \frac{\partial f \left( \Delta, y, y_0; \theta \right)}{\partial \Delta} + \mu_Y(y, \theta_0) \frac{\partial f \left( \Delta, y, y_0; \theta \right)}{\partial y} + \frac{1}{2} \frac{\partial^2 f \left( \Delta, y, y_0; \theta \right)}{\partial y^2}. \qquad (4.4)$$

In all cases, this expression is a proper Taylor series; whether the series is analytic at $\Delta = 0$ is not guaranteed, although sufficient conditions can be given (see Proposition 4 in Aït-Sahalia (2002), who also discusses the class of functions $f$, such as polynomials, for which this representation is admissible).

Applying (4.3) to $f(\Delta, Y_\Delta, Y_0; \theta) = H_j \left( \Delta^{-1/2} \left( Y_\Delta - Y_0 \right) \right)$ up to order $K$ for the purpose of evaluating $\eta_j$

in $\check{p}_Y^{(J)}$ yields the expansion $\check{p}_Y^{(J,K)}$. Different ways of gathering the terms are available (as in the Central Limit Theorem, where for example both the Edgeworth and Gram-Charlier expansions are based on a Hermite expansion). One particularly convenient way of gathering the terms of the expansion consists in grouping them in powers of $\Delta$. This is then in the same spirit as the "small-time" expansions of Azencott (1984), except that the expansions obtained here are fully explicit instead of relying of moments of functionals of Brownian Bridges. Indeed, if we gather all the terms according to increasing powers of $\Delta$ instead of increasing order of the Hermite polynomials, and let $p_Y^{(K)} \equiv \check{p}_Y^{(\infty,K)}$, we obtain an explicit representation of $p_Y^{(K)}$, given by:

$$p_Y^{(K)}(\Delta, y|y_0; \theta) = \Delta^{-1/2} \phi\left(\frac{y - y_0}{\Delta^{1/2}}\right) \exp\left(\int_{y_0}^{y} \mu_Y(w; \theta)\, dw\right) \sum_{k=0}^{K} c_Y^{(k)}(y|y_0; \theta) \frac{\Delta^k}{k!} \tag{4.5}$$

where $\phi(w) = \exp(-w^2/2)/(2\pi)$ is the $N(0,1)$ density function, $c_Y^{(0)} = 1$ and for all $k > 1$:

$$\begin{aligned}
c_Y^{(k)}(y|y_0; \theta) &= k\ (y - y_0)^{-k} \int_{y_0}^{y} (w - y_0)^{k-1} \Big\{ \lambda_Y(w; \theta)\, c_Y^{(k-1)}(w|y_0; \theta) \\
&\quad + \left(\partial^2 c_Y^{(k-1)}(w|y_0; \theta)/\partial w^2\right)/2 \Big\}\, dw
\end{aligned} \tag{4.6}$$

with

$$\lambda_Y(y; \theta) = -\frac{1}{2}\left(\mu_Y^2(y; \theta) + \frac{\partial \mu_Y(y; \theta)}{\partial y}\right). \tag{4.7}$$

Equation (4.6) allows the recursive computation of the coefficients, starting from $c_Y^{(0)} = 1$.

When we are interested in computing the logarithm of the transition function, an alternative form of the Taylor series can be more amenable to the computation of the log-likelihood, and guarantee positivity of the density. Indeed, the function $l_Y(\Delta, y|y_0; \theta)$ can also be expressed directly as a series in $\Delta$, namely by Taylor-expanding $\ln(\sum_{j=0}^{J} c_Y^{(j)}(y|y_0; \theta) \frac{\Delta^j}{j!})$ in $\Delta$. This yields the form

$$l_Y^{(K)}(\Delta, y|y_0; \theta) = -\frac{1}{2}\ln(2\pi\Delta) + \frac{C_Y^{(-1)}(y|y_0; \theta)}{\Delta} + \sum_{k=0}^{K} C_Y^{(k)}(y|y_0; \theta) \frac{\Delta^k}{k!} \tag{4.8}$$

and, by application of the Jacobian change of variable formula,

$$l_X^{(K)}(\Delta, x|x_0; \theta) = -\frac{1}{2}\ln(\sigma^2(x; \theta)) + l_Y^{(K)}(\Delta, \gamma(x; \theta)|\gamma(x_0; \theta); \theta). \tag{4.9}$$

The coefficients are given by

$$\begin{aligned}
C_Y^{(-1)}(y|y_0; \theta) &= -(y - y_0)^2/2 \tag{4.10} \\
C_Y^{(0)}(y|y_0; \theta) &= \int_{y_0}^{y} \mu_Y(w; \theta)\, dw \tag{4.11} \\
C_Y^{(1)}(y|y_0; \theta) &= c_Y^{(1)}(y|y_0; \theta) = (y - y_0)^{-1} \int_{y_0}^{y} \lambda_Y(w; \theta)\, dw \tag{4.12}
\end{aligned}$$

Note that

$$C_Y^{(-1)}(y_0|y_0; \theta) = 0, \quad C_Y^{(0)}(y_0|y_0; \theta) = 0, \quad C_Y^{(1)}(y_0|y_0; \theta) = \lambda_Y(y_0; \theta), \tag{4.13}$$

11

the last equation being a consequence of L'Hôpital's Rule.

The other coefficients are obtained recursively. Given $C_Y^{(-1)}$, $C_Y^{(0)}$, ..., $C_Y^{(k-1)}$, the coefficient $C_Y^{(k)}$, $k \geq 2$, is given by:

$$
\begin{aligned}
C_Y^{(k)}(y|y_0;\theta) &= k \ (y-y_0)^{-k} \int_{y_0}^{y} (w-y_0)^{k-1} \left\{ \frac{1}{2} \frac{\partial^2 C_Y^{(k-1)}(w|y_0;\theta)}{\partial w^2} \right. \\
&\left. + \frac{1}{2} \sum_{h=1}^{k-2} \binom{k-1}{h} \frac{\partial C_Y^{(h)}(w|y_0;\theta)}{\partial w} \frac{\partial C_Y^{(k-1-h)}(w|y_0;\theta)}{\partial w} \right\} dw.
\end{aligned} \tag{4.14}
$$

For consistency with the multivariate case to appear below, note for now that these expressions can also be written in the form

$$
\begin{aligned}
C_Y^{(k)}(y|y_0;\theta) &= k \ (y-y_0)^{-k} \int_{y_0}^{y} (w-y_0)^{k-1} G_Y^{(k)}(w|y_0;\theta)\,dw \\
&= k \int_0^1 G_Y^{(k)}(y_0 + u\,(y-y_0)\,|y_0;\theta)\,u^{k-1}\,du
\end{aligned} \tag{4.15}
$$

where

$$
\begin{aligned}
G_Y^{(1)}(y|y_0;\theta) &= -\frac{\partial \mu_Y(y;\theta)}{\partial y} - \mu_Y(y;\theta)\frac{\partial C_Y^{(0)}(y|y_0;\theta)}{\partial y} + \frac{1}{2}\frac{\partial^2 C_Y^{(0)}(y|y_0;\theta)}{\partial y^2} + \frac{1}{2}\left[\frac{\partial C_Y^{(0)}(y|y_0;\theta)}{\partial y}\right]^2 \\
&= \lambda_Y(y;\theta)
\end{aligned} \tag{4.16}
$$

and for $k \geq 2$

$$
\begin{aligned}
G_Y^{(k)}(y|y_0;\theta) &= -\mu_Y(y;\theta)\frac{\partial C_Y^{(k-1)}(y|y_0;\theta)}{\partial y} + \frac{1}{2}\frac{\partial^2 C_Y^{(k-1)}(y|y_0;\theta)}{\partial y^2} \\
&+ \frac{1}{2}\sum_{h=0}^{k-1}\binom{k-1}{h}\frac{\partial C_Y^{(h)}(y|y_0;\theta)}{\partial y}\frac{\partial C_Y^{(k-1-h)}(y|y_0;\theta)}{\partial y} \\
&= \frac{1}{2}\frac{\partial^2 C_Y^{(k-1)}(y|y_0;\theta)}{\partial y^2} + \frac{1}{2}\sum_{h=1}^{k-2}\binom{k-1}{h}\frac{\partial C_Y^{(h)}(y|y_0;\theta)}{\partial y}\frac{\partial C_Y^{(k-1-h)}(y|y_0;\theta)}{\partial y}
\end{aligned} \tag{4.17}
$$

## 4.2 Determination of the Coefficients in the Multivariate Reducible Case

In the case of a multivariate reducible diffusion, I proceed along the same lines. Hermite polynomials are available in the multivariate case (see e.g., Chapter 5 of McCullagh (1987) or Withers (2000)). Let $\phi(x)$ denote the density of the $m-$dimensional multivariate Normal distribution with mean zero and covariance matrix $\kappa = [\kappa_{ij}]_{i,j=1,..,m}$. The inverse of $\kappa$ is $\kappa^{-1} = [\kappa_{ij}^{-1}]_{i,j=1,..,m}$, so that

$$
\phi(x;\kappa) = (2\pi)^{-m/2}Det[\kappa]^{-1/2}\exp(-\sum_{i=1}^{m}\sum_{j=1}^{m}\kappa_{ij}^{-1}x_i x_j).
$$

For each vector $h = (h_1, ..., h_m)^T \in \mathbb{N}^m$, recall that $tr[h] = h_1 + ... + h_m$, I will denote by $H_h(x)$ the associated Hermite polynomials, which are defined by

$$
H_h(x;\kappa) = \frac{(-1)^{tr[h]}}{\phi(x;\kappa)}\frac{\partial^{tr[h]}\phi(x;\kappa)}{dx_1^{h_1}...dx_m^{h_m}}
$$

12

and can be computed explicitly to an arbitrary order $tr[h]$. The dual Hermite polynomials are

$$\tilde{H}_h(x;\kappa) = \frac{(-1)^{tr[h]}}{\phi(x;\kappa)} \frac{\partial^{tr[h]}\phi(z;\kappa)}{dz_1^{h_1}...dz_m^{h_m}}$$

at $z = \kappa^{-1}x$. We have that $\tilde{H}_h(x;\kappa) = H_h(\kappa^{-1}x;\kappa^{-1})$. The polynomials are orthogonal with respect to their duals in the sense that

$$\int_{\mathbb{R}^m} H_h(x;\kappa)\tilde{H}_k(x;\kappa)\phi(x;\kappa)dx = h_1!...h_m!$$

if $h = k$ and 0 otherwise.

The Hermite series approximation of $p_Y$ is in the form

$$\breve{p}_Y^{(J)}(\Delta, y|y_0; \theta) = \Delta^{-m/2}\phi\left(\Delta^{-1/2}(y - y_0); I\right) \sum_{h \in \mathbb{N}^m : tr[h] \leq J} \eta_h(\Delta, y_0; \theta) H_h(\Delta^{-1/2}(y - y_0); I) \quad (4.18)$$

i.e., with $\kappa = I$, and the Hermite coefficients $\eta_h(\Delta, y_0; \theta)$ can be computed as in the univariate case, by relying on their orthogonality. Also as in the univariate case, the Hermite expansions can be written directly for the log-density. The key question addressed in this paper is the computation of the coefficients, and this is where I rely on the structure afforded by the diffusion hypothesis (note of course that I do not assume that the characteristic function of the process is known).

The infinitesimal generator $A_Y$ corresponding to the reduced diffusion $Y$ in (3.1) is

$$A_Y \cdot f(\Delta, y, y_0; \theta) = \frac{\partial f(\Delta, y, y_0; \theta)}{\partial \Delta} + \sum_{i=1}^m \mu_{Yi}(y; \theta) \frac{\partial f(\Delta, y, y_0; \theta)}{\partial y_i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 f(\Delta, y, y_0; \theta)}{\partial y_i \partial y_j}. \quad (4.19)$$

Gathering again the coefficients in an expansion in increasing powers of $\Delta$, the form of the expansion analogous to (4.8) is then

$$l_Y^{(K)}(\Delta, y|y_0; \theta) = -\frac{m}{2}\ln(2\pi\Delta) + \frac{C_Y^{(-1)}(y|y_0; \theta)}{\Delta} + \sum_{k=0}^K C_Y^{(k)}(y|y_0; \theta)\frac{\Delta^k}{k!} \quad (4.20)$$

leaving us with the computation of the coefficients $C_Y^{(k)}$, $k = -1, 0, 1, 2, ..., K$. The following result gives an explicit expression for each one of these coefficients:

**Theorem 1.** *The coefficients of the log-density Taylor expansion $l_Y^{(K)}(\Delta, y|y_0; \theta)$ are given explicitly by:*

$$C_Y^{(-1)}(y|y_0; \theta) = -\frac{1}{2}\sum_{i=1}^m (y_i - y_{0i})^2 \quad (4.21)$$

$$C_Y^{(0)}(y|y_0; \theta) = \sum_{i=1}^m (y_i - y_{0i})\int_0^1 \mu_{Yi}(y_0 + u(y - y_0); \theta)\, du \quad (4.22)$$

*and, for $k \geq 1$,*

$$C_Y^{(k)}(y|y_0; \theta) = k\int_0^1 G_Y^{(k)}(y_0 + u(y - y_0)|y_0; \theta)\, u^{k-1}du \quad (4.23)$$

13

*where*

$$G_Y^{(1)}(y|y_0;\theta) = -\sum_{i=1}^{m} \frac{\partial \mu_{Yi}(y;\theta)}{\partial y_i} - \sum_{i=1}^{m} \mu_{Yi}(y;\theta) \frac{\partial C_Y^{(0)}(y|y_0;\theta)}{\partial y_i}$$
$$+ \frac{1}{2} \sum_{i=1}^{m} \left\{ \frac{\partial^2 C_Y^{(0)}(y|y_0;\theta)}{\partial y_i^2} + \left[ \frac{\partial C_Y^{(0)}(y|y_0;\theta)}{\partial y_i} \right]^2 \right\} \tag{4.24}$$

*and for $k \geq 2$*

$$G_Y^{(k)}(y|y_0;\theta) = -\sum_{i=1}^{m} \mu_{Yi}(y;\theta) \frac{\partial C_Y^{(k-1)}(y|y_0;\theta)}{\partial y_i} + \frac{1}{2} \sum_{i=1}^{m} \frac{\partial^2 C_Y^{(k-1)}(y|y_0;\theta)}{\partial y_i^2}$$
$$+ \frac{1}{2} \sum_{i=1}^{m} \sum_{h=0}^{k-1} \binom{k-1}{h} \frac{\partial C_Y^{(h)}(y|y_0;\theta)}{\partial y_i} \frac{\partial C_Y^{(k-1-h)}(y|y_0;\theta)}{\partial y_i}. \tag{4.25}$$

To obtain an expansion for the density $p_Y$ instead of the log-density $l_Y$, one can either take the exponential of $l_Y^{(K)}$, yielding

$$p_Y^{(K)}(\Delta, y|y_0;\theta) = (2\pi\Delta)^{-m/2} \exp\left( \frac{C_Y^{(-1)}(y|y_0;\theta)}{\Delta} + \sum_{k=0}^{K} C_Y^{(k)}(y|y_0;\theta) \frac{\Delta^k}{k!} \right) \tag{4.26}$$

or alternatively, given the coefficients $C_k$ for the log-density, the coefficients $c_k$ for the density expansion can be obtained by matching the coefficients in the two Taylor expansions $l_Y^{(K)}$ and $p_Y^{(K)}$.

## 4.3   Change of Variable

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. Define the Jacobian matrix $\nabla\gamma(x;\theta)$. Then the transition density of $X$ is related to that of $Y$ by

$$p_X(\Delta, x|x_0;\theta) = Det\left[\nabla\gamma(x;\theta)\right] \, p_Y(\Delta, \gamma(x;\theta) \,|\, \gamma(x_0;\theta);\theta). \tag{4.27}$$

Then from (3.2) and (2.6), we have

$$Det\left[\nabla\gamma(x;\theta)\right] = Det\left[\sigma^{-1}(x;\theta)\right] = Det\left[v(x;\theta)\right]^{-1/2}. \tag{4.28}$$

Then, replacing $p_Y$ on the right-hand-side of (4.27) by $p_Y^{(K)}$ yields an expansion $p_X^{(K)}$ for $p_X$.

In terms of log-densities, we have

$$l_X(\Delta, x|x_0;\theta) = -\frac{1}{2} \ln\left(Det\left[v(x;\theta)\right]\right) + l_Y(\Delta, \gamma(x;\theta) \,|\, \gamma(x_0;\theta);\theta)$$
$$= -D_v(x;\theta) + l_Y(\Delta, \gamma(x;\theta) \,|\, \gamma(x_0;\theta);\theta) \tag{4.29}$$

14

which I mimic at the level of the approximations of order $K$ in $\Delta$, thereby defining $l_X^{(K)}$

$$
\begin{aligned}
l_X^{(K)}(\Delta, x|x_0; \theta) &= -D_v(x; \theta) + l_Y^{(K)}(\Delta, \gamma(x; \theta) | \gamma(x_0; \theta); \theta) \\
&= -\frac{m}{2}\ln(2\pi\Delta) - D_v(x; \theta) \\
&\quad + \frac{C_Y^{(-1)}(\gamma(x; \theta) | \gamma(x_0; \theta); \theta)}{\Delta} + \sum_{k=0}^{K} C_Y^{(k)}(\gamma(x; \theta) | \gamma(x_0; \theta); \theta) \frac{\Delta^k}{k!}
\end{aligned}
\tag{4.30}
$$

from $l_Y^{(K)}$ given in (4.20), using the coefficients $C_Y^{(k)}$, $k = -1, 0, ..., K$ given in Theorem 1. This fully describes the construction of the expansion of $l_X$ for a reducible diffusion.

## 4.4 Independent Variables

An important special case occurs when the $m$ variables in (2.1) are independent. In that case, the multivariate transition density $p_X$ is simply the product of the $m$ univariate transition densities, and the log-likelihood $l_X$ is the sum of the univariate ones. The following proposition shows that the expansion shares this feature:

**Proposition 2.** *Suppose that for each $i = 1, ..., m$, $\mu_i(x; \theta)$ and $\sigma_{ii}(x; \theta)$ depend on $x_i$ only, and that $\sigma_{ij}(x; \theta) = 0$ for $j \neq i$. Then the diffusion is reducible and we have*

$$
l_X^{(K)}(\Delta, x|x_0; \theta) = \sum_{i=1}^{m} l_X^{(K)}(\Delta, x_i|x_{0i}; \theta)
\tag{4.31}
$$

*where $l_X^{(K)}(\Delta, x_i|x_{0i}; \theta)$ is the univariate expansion corresponding to the $i^{th}$ variable, defined in (4.9).*

# 5 Closed-Form Expansion for the Transition Density of an Irreducible Diffusion

I now turn to the irreducible case. Mimicking the form of the Taylor expansion in $\Delta$ obtained in the reducible case, namely (4.30), leads to postulating the following form for an expansion of the log likelihood

$$
l_X^{(K)}(\Delta, x|x_0; \theta) = -\frac{m}{2}\ln(2\pi\Delta) - D_v(x; \theta) + \frac{C_X^{(-1)}(x|x_0; \theta)}{\Delta} + \sum_{k=0}^{K} C_X^{(k)}(x|x_0; \theta)\frac{\Delta^k}{k!}.
\tag{5.1}
$$

The idea now is to derive an explicit Taylor approximation in $(x - x_0)$ of the coefficients $C_X^{(k)}(x|x_0; \theta)$, $k = -1, 0, ..., K$. Specifically, I 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 $C_X^{(j_k, k)}$. A Taylor series in $(x - x_0)$ is the form that arises directly from the representation of the Hermite series $\breve{p}_X^{(J)}$ as in the univariate case (4.1), with the order $J$ of the truncation of the series now representing the order of the polynomial term in $(x - x_0)$ as opposed to the order of the Hermite polynomials (which are polynomials in $(x - x_0)$). In the reducible case, we are able to expand that series in powers of $\Delta$, gather the terms as the coefficient of the term $\Delta^k$ in the series, take the limit of the series as the number of Hermite polynomials increase and obtain an explicit expression for

15

$C_X^{(k)} = C_X^{(\infty,k)}$, so that we obtained the coefficients $C_X^{(k)}$ with no need to Taylor-expand them in $(x - x_0)$. This last step is what's no longer possible when the diffusion is irreducible.

However, it is still possible to compute the Taylor expansions $C_X^{(j_k,k)}$ explicitly. Before describing how to compute such a coefficient, one remaining question to solve is the choice of the order $j_k$ (in $(x - x_0)$) corresponding to a given order $k$ (in $\Delta$). For that purpose, recall that $x - x_0 = O_p\left(\Delta^{1/2}\right)$ so that

$$\left| C_X^{(k)}(x|x_0;\theta)\,\Delta^k - C_X^{(j_k,k)}(x|x_0;\theta)\,\Delta^k \right| = O_p\left((x - x_0)^{j_k}\Delta^k\right) = O_p(\Delta^{j_k/2+k}) \tag{5.2}$$

and setting $j_k/2 + k = K$, i.e.,

$$j_k = 2(K - k) \tag{5.3}$$

for $k = -1, 0, ..., K$, will therefore provide an approximation error due to the Taylor expansion in $(x - x_0)$ of the same order $\Delta^K$ for each one of the terms in the series (5.1).

The resulting expansion will then be

$$\tilde{l}_X^{(K)}(\Delta, x|x_0;\theta) = -\frac{m}{2}\ln(2\pi\Delta) - D_v(x;\theta) + \frac{C_X^{(j_{-1},-1)}(x|x_0;\theta)}{\Delta} + \sum_{k=0}^K C_X^{(j_k,k)}(x|x_0;\theta)\frac{\Delta^k}{k!}. \tag{5.4}$$

This double Taylor expansion (in $\Delta$ and in $(x - x_0)$) can be viewed as a Taylor expansion in $\Delta$ only, in light of (5.2). In general, the function need not be analytic at $\Delta = 0$, hence the expansion is to be interpreted strictly as a Taylor expansion.

What remains to be done is to compute explicitly the Taylor expansion $C_X^{(j_k,k)}$ of each coefficient $C_X^{(k)}$. As I will now show, this involves solving a cascade of differential equations, starting with $C_X^{(j_{-1},-1)}$, then use that solution to determine $C_X^{(j_0,0)}$, etc. Fortunately, each one of these differential problems has a closed-form solution as we will now see in Section 5.2.

## 5.1 The Leading Term: Geometric Interpretation

While the leading term $C_X^{(-1)}$ in the case of a reducible diffusion is simply

$$C_X^{(-1)}(x|x_0;\theta) = C_Y^{(-1)}(\gamma(x;\theta)|\gamma(x_0;\theta);\theta),$$

with $C_Y^{(-1)}(y|y_0;\theta) = -\frac{1}{2}\|y - y_0\|^2$ (see (4.21) and (4.30)), the situation is more involved when the diffusion $X$ is not reducible.

Consider the set $\Omega(x|x_0)$ of $m-$dimensional differentiable paths $\omega(\tau)$, starting at $x_0$ at time 0 and ending at $x$ at time 1. An example of such a path is the straight line $\omega(\tau) = x_0 + \tau(x - x_0)$. Consider now the Riemannian metric derived from the coefficients of the matrix $v(x;\theta)^{-1}$, that is the distance between points $x$ and $x + dx$ defined by

$$ds = \left(\sum_{i,j=1}^m v_{ij}^{-1}(x;\theta)\,dx_i dx_j\right)^{1/2}. \tag{5.5}$$

With this metric, the length of any differentiable path $\omega$ is

$$d(\omega;\theta) = \int_0^1 \left( \sum_{i,j=1}^m v_{ij}\left(\omega(\tau);\theta\right) \frac{d\omega_i(\tau)}{d\tau} \frac{\omega_j(\tau)}{d\tau} \right)^{1/2} d\tau.$$

Varadhan (1967) has shown that

$$\lim_{\Delta \to 0} -2\Delta\, l_X\left(\Delta, x | x_0; \theta\right) \quad = \quad \inf_{\omega \in \Omega(x|x_0)}\, d(\omega;\theta)^2.$$

Since from (5.1)

$$\lim_{\Delta \to 0}\, \Delta\, l_X^{(K)}\left(\Delta, x | x_0; \theta\right) \quad = \quad C_X^{(-1)}\left(x | x_0; \theta\right)$$

the appropriate leading term of the expansion (5.1) ought to be

$$C_X^{(-1)}\left(x | x_0; \theta\right) = -\frac{1}{2} \inf_{\omega \in \Omega(x|x_0)} d(\omega;\theta)^2 \tag{5.6}$$

that is, minus one half the square of the shortest distance from $x$ to $x_0$ in the metric induced in $\mathbb{R}^m$ by the matrix $v(x;\theta)^{-1}$.

An important special case occurs when $\sigma$, hence $\nu$, is the identity matrix. In this case, the distance (5.5) reduces to the usual Euclidean distance, the infimum in (5.6) is achieved by the straight line, and we have

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

which is the result obtained in the reducible case for the reduced diffusion $Y$: see equation (4.21).

But, for any $v(x;\theta)$, the distance (5.6) is invariant under coordinate transformations. This applies in particular to the transformation from $X$ to $Y \equiv \gamma(X;\theta)$ when the diffusion is reducible. In this situation, we have

$$C_X^{(-1)}\left(x | x_0; \theta\right) = -\frac{1}{2} \|\gamma(x;\theta) - \gamma(x_0;\theta)\|^2.$$

In dimension $m = 1$, where every diffusion is reducible, this can be recovered directly. We already know from the univariate case that

$$C_X^{(-1)}\left(x | x_0; \theta\right) = -\frac{1}{2} \left( \int_{x_0}^x \frac{1}{\sigma(w;\theta)} dw \right)^2. \tag{5.7}$$

Now, the only way to move on the real line (including the shortest distance path) is to stay on that straight line. Suppose, without loss of generality, that $x \geq x_0$. With $\omega(\tau) = x_0 + \tau(x - x_0)$, we have

$$\begin{aligned}
d(\omega;\theta) &= \int_0^1 \frac{1}{\sigma(\omega(\tau);\theta)} \left(\frac{d\omega(\tau)}{d\tau}\right) d\tau \\
&= (x - x_0) \int_0^1 \frac{1}{\sigma(x_0 + \tau(x - x_0);\theta)} d\tau \\
&= \int_{x_0}^x \frac{1}{\sigma(w;\theta)} dw
\end{aligned}$$

17

with the last equality resulting from the change of variable $\tau \mapsto w = x_0 + \tau(x - x_0)$. Since $\gamma$ is given by (3.3) when $m = 1$, we indeed recover (5.7) from the general formula (5.6).

## 5.2   Determination of the Coefficients in the Multivariate Irreducible Case

I now turn to the determination of a closed-form expression for the Taylor expansions $C_X^{(j_k,k)}$ of the coefficients $C_X^{(k)}$. Essentially, the coefficients are determined one by one, starting with the leading term $C_X^{(j_{-1},-1)}$. Given $C_X^{(j_{-1},-1)}$, the next term $C_X^{(j_0,0)}$ is calculated explicitly, and so on. The orders of the Taylor expansions $j_{-1}$, $j_0$, etc., are chosen to control the order of the remainder terms, setting each $j_k$ according to (5.3). This means in particular that the highest order term ($k = -1$) is Taylor-expanded to a higher degree of precision than the successive terms. This is to be expected, given that $C_X^{(j_{-1},-1)}$ in a input to the differential equation determining $C_X^{(j_0,0)}$, and so on.

In order to state the main result pertaining to the closed-form solutions $C_X^{(j_k,k)}$, I define the following functions of the coefficients and their derivatives:

$$
\begin{aligned}
G_X^{(0)}(x|x_0;\theta) &= \frac{m}{2} - \sum_{i=1}^m \mu_i(x;\theta) \frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_i} + \sum_{i=1}^m \sum_{j=1}^m \frac{\partial v_{ij}(x;\theta)}{\partial x_i} \frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_j} \\
&\quad + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x;\theta) \frac{\partial^2 C_X^{(-1)}(x|x_0;\theta)}{\partial x_i \partial x_j} \\
&\quad - \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x;\theta) \frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_i} \frac{\partial D_v(x;\theta)}{\partial x_j},
\end{aligned}
\tag{5.8}
$$

$$
\begin{aligned}
G_X^{(1)}(x|x_0;\theta) &= -\sum_{i=1}^m \frac{\partial \mu_i(x;\theta)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 v_{ij}(x;\theta)}{\partial x_i \partial x_j} \\
&\quad - \sum_{i=1}^m \mu_i(x;\theta) \left( \frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_i} - \frac{\partial D_v(x;\theta)}{\partial x_i} \right) \\
&\quad + \sum_{i=1}^m \sum_{j=1}^m \frac{\partial v_{ij}(x;\theta)}{\partial x_i} \left( \frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_j} - \frac{\partial D_v(x;\theta)}{\partial x_j} \right) \\
&\quad + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x;\theta) \left\{ \frac{\partial^2 C_X^{(0)}(x|x_0;\theta)}{\partial x_i \partial x_j} - \frac{\partial^2 D_v(x;\theta)}{\partial x_i \partial x_j} \right. \\
&\quad \left. + \left( \frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_i} - \frac{\partial D_v(x;\theta)}{\partial x_i} \right) \left( \frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_j} - \frac{\partial D_v(x;\theta)}{\partial x_j} \right) \right\}
\end{aligned}
\tag{5.9}
$$

and for $k \geq 2$ :

$$
\begin{aligned}
G_X^{(k)}(x|x_0;\theta) &= -\sum_{i=1}^m \mu_i(x;\theta)\frac{\partial C_X^{(k-1)}(x|x_0;\theta)}{\partial x_i} + \sum_{i=1}^m \sum_{j=1}^m \frac{\partial v_{ij}(x;\theta)}{\partial x_i}\frac{\partial C_X^{(k-1)}(x|x_0;\theta)}{\partial x_j} \\
&\quad +\frac{1}{2}\sum_{i=1}^m \sum_{j=1}^m v_{ij}(x;\theta)\frac{\partial^2 C_X^{(k-1)}(x|x_0;\theta)}{\partial x_i \partial x_j} \\
&\quad +\frac{1}{2}\sum_{i=1}^m \sum_{j=1}^m v_{ij}(x;\theta)\left\{2\left(\frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_i}-\frac{\partial D_v(x;\theta)}{\partial x_i}\right)\frac{\partial C_X^{(k-1)}(x|x_0;\theta)}{\partial x_j}\right. \\
&\quad \left. +\sum_{h=1}^{k-2}\binom{k-2}{h}\frac{\partial C_X^{(h)}(x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(k-1-h)}(x|x_0;\theta)}{\partial x_j}\right\}.
\end{aligned}
\tag{5.10}
$$

Note that the computation of each function $G_X^{(k)}$ requires only the ability to differentiate the previously determined coefficients $C_X^{(-1)}$, ..., $C_X^{(k-1)}$. The same applies to their Taylor expansions. Let $i \equiv (i_1, i_2, ..., i_m)$ denote a vector of integers and

$$
I_k = \{i \equiv (i_1, i_2, ..., i_m) \in \mathbb{N}^m : 0 \leq tr[i] \leq j_k\}
\tag{5.11}
$$

so that the form of $C_X^{(j_k,k)}$ is

$$
C_X^{(j_k,k)}(x|x_0;\theta) = \sum_{i \in I_k} \gamma_i^{(k)}(x_0;\theta)(x_1 - x_{01})^{i_1}(x_2 - x_{02})^{i_2}...(x_m - x_{0m})^{i_m}.
\tag{5.12}
$$

The following theorem can now describe how the coefficients $C_X^{(j_k,k)}$, i.e., the coefficients $\gamma_i^{(k)}$, $i \in I_k$, are determined:

**Theorem 2.** *For each $k = -1, 0, ..., K$, the coefficient $C_X^{(k)}(x|x_0;\theta)$ in (5.1) solves the equation*

$$
f_X^{(k-1)}(x|x_0;\theta) = 0
\tag{5.13}
$$

*where*

$$
f_X^{(-2)}(x|x_0;\theta) = -2C_X^{(-1)}(x|x_0;\theta) - \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_j}
\tag{5.14}
$$

$$
f_X^{(-1)}(x|x_0;\theta) = -\sum_{i=1}^m \sum_{j=1}^m v_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(0)}(x|x_0;\theta)}{\partial x_j} - G_X^{(0)}(x|x_0;\theta).
\tag{5.15}
$$

*and for $k \geq 1$*

$$
\begin{aligned}
f_X^{(k-1)}(x|x_0;\theta) &= C_X^{(k)}(x|x_0;\theta) - \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(k)}(x|x_0;\theta)}{\partial x_j} \\
&\quad -G_X^{(k)}(x|x_0;\theta).
\end{aligned}
\tag{5.16}
$$

*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, meaning that 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.*

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 $(x-x_0)$. The coefficients $\gamma_i^{(k)}(x_0;\theta)$, $i \in I_k$ of $C_X^{(j_k,k)}$ are determined by setting the Taylor expansion $f_X^{(j_k,k-1)}$ of $f_X^{(k-1)}$ to zero. The key feature that makes this problem solvable in closed form is that the coefficients solve a succession of systems of linear equations: first determine $\gamma_i^{(k)}$ for $tr[i] = 0$, then $\gamma_i^{(k)}$ for $tr[i] = 1$, and all the way to $tr[i] = j_k$.

Note in particular, for $k = -1 : \gamma_i^{(-1)} = 0$ for $tr[i] = 0, 1$ (i.e., the polynomial has no constant or linear terms) and the terms corresponding to $tr[i] = 2$ (with of course $j_{-1} \geq 2$) are:

$$\sum_{i \in I_{-1}:tr[i]=2} \gamma_i^{(-1)}(x_0;\theta)(x_1 - x_{01})^{i_1}(x_2 - x_{02})^{i_2}...(x_m - x_{0m})^{i_m} = -\frac{1}{2}(x - x_0)^T v^{-1}(x_0;\theta)(x - x_0).$$

which is the anticipated term given the Gaussian limiting behavior of the transition density when $\Delta$ is small. Thus with $j_{-1} \geq 3$, we only need to determine the terms $\gamma_i^{(-1)}$ corresponding to $tr[i] = 3, ..., j_{-1}$.

For $k = 0 : \gamma_i^{(0)} = 0$ for $tr[i] = 0$, so the polynomial has no constant term. For $k \geq 1$, the polynomials have a constant term (for $k \geq 1$, $\gamma_i^{(k)} \neq 0$ for $tr[i] = 0$ in general).

## 5.3 Applying the Irreducible Method to a Reducible Diffusion

Theorem 2 is more general than Theorem 1 in that it does not require that the diffusion be reducible. In exchange for that generality, the coefficients are available in closed form only in the form of a Taylor series expansion in $(x - x_0)$. The following proposition describes the relationship between these two methods when Theorem 2 is applied to a diffusion that is in fact reducible:

**Proposition 3.** *Suppose that the diffusion $X$ is reducible, and let $l_X^{(K)}$ denote its log-likelihood expansion calculated by applying Theorem 1. Suppose now that we also calculate its log-likelihood expansion, $\tilde{l}_X^{(K)}$, without first transforming $X$ into the unit diffusion $Y$, that is by applying Theorem 2 to $X$ directly. Then each coefficient $C_X^{(j_k,k)}(x|x_0;\theta)$ from $\tilde{l}_X^{(K)}$ is a Taylor expansion in $(x - x_0)$ at order $j_k$ of the coefficient $C_X^{(k)}(x|x_0;\theta) = C_Y^{(k)}(\gamma(x;\theta)|\gamma(x_0;\theta);\theta)$ from $l_X^{(K)}$.*

In other words, applying the irreducible method to a diffusion that is in fact reducible involves replacing (needlessly) the exact expression for $C_X^{(k)}(x|x_0;\theta)$ by its Taylor series in $(x - x_0)$. Of course, there is no reason to do so when the diffusion is reducible.

# 6 Examples

In this section, I apply the results above to three examples of multivariate diffusion processes of interest in financial econometrics.

## 6.1 The Bivariate Ornstein-Uhlenbeck Model

Consider the model

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{11}(\alpha_1 - X_{1t}) + \beta_{12}(\alpha_2 - X_{2t}) \\ \beta_{21}(\alpha_1 - X_{1t}) + \beta_{22}(\alpha_2 - X_{2t}) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \end{pmatrix} \tag{6.1}$$

where the parameter vector is $\theta = (\alpha_1, \alpha_2, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \sigma_{11}, \sigma_{12}, \sigma_{21}, \sigma_{22})^T$. Let

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{22} & \beta_{22} \end{pmatrix}, \quad \sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

so that $dX_t = \beta(\alpha - X_t)dt + \sigma dW_t$, and assume that $\beta$ has full rank (as well as $\sigma$, recall Assumption ??.3). This is the most basic model capturing mean reversion in the state variables.

Consider the matrix equation

$$\beta\lambda + \lambda\beta^T = \sigma\sigma^T \tag{6.2}$$

whose solution in the bivariate case is the $2 \times 2$ symmetric matrix $\lambda$ given by

$$\lambda = \frac{1}{2tr[\beta]Det[\beta]} \left( Det[\beta]\sigma\sigma^T + (\beta - tr[\beta])\sigma\sigma^T (\beta - tr[\beta])^T \right). \tag{6.3}$$

When the process is stationary, i.e., when the eigenvalues of the matrix $\beta$ have positive real parts, $\lambda$ is the stationary variance-covariance matrix of the process. That is, the stationary density of $X$ is the bivariate Normal density with mean $\alpha$ and variance-covariance $\lambda$.

The transition density of $X$ is the bivariate Normal density

$$p_X(\Delta, x|x_0; \theta) = (2\pi)^{-1} Det[\Omega(\Delta; \theta)]^{-1/2} \exp(-(x - m(\Delta, x_0; \theta))^T \Omega^{-1}(\Delta; \theta)(x - m(\Delta, x_0; \theta))) \tag{6.4}$$

where

$$m(\Delta, x_0; \theta) = \alpha + \exp(-\beta\Delta)(x_0 - \alpha) \tag{6.5}$$

$$\Omega(\Delta; \theta) = \lambda - \exp(-\beta\Delta)\lambda\exp(-\beta^T\Delta) \tag{6.6}$$

and exp applied to a matrix denotes the matrix exponential (which does not in general reduce to the exponential of each term of the matrix).

I now discuss the identification of the continuous time parameters from the discrete data. This presence of the matrix exponential $\exp(-\beta\Delta)$ provides a clear insight into the aliasing phenomenon as it applies to this model. From the form of the transition function (6.4) with conditional mean and variance (6.5)-(6.6), discrete data sampled at time interval $\Delta$ may not distinguish between two sets of parameters $\beta$ and $\beta'$ such that $\exp(-\beta\Delta) = \exp(-\beta'\Delta)$. The eigenvalues of $\beta$ are either both real, or both complex conjugates. If they are complex, then for any given $B$, there are countably many solutions in $\beta$ to the equation $\exp(-\beta\Delta) = B$. This phenomenon was noted by Philips (1973). If the eigenvalues of $\beta$ are a pair of distinct complex conjugate

numbers that do not differ by an integer multiple of $2\pi i/\Delta$, let $\beta = T\Lambda T^{-1}$ where $T$ and $\Lambda$ are respectively the matrices of eigenvectors and eigenvalues of $\beta$. Then for any integer $g$, the matrix $\beta'$ defined by

$$\beta' = \beta + \frac{2\pi i}{\Delta} T \cdot diag(g, -g) \cdot T^{-1}$$

satisfies $\exp\left(-\beta'\Delta\right) = \exp\left(-\beta\Delta\right) = B$. The phenomenon does not occur if the eigenvalues of $\beta$ are all real because $\beta'$ would then have complex elements since the eigenvalues of $\beta'$ are $\Lambda + (2\pi i/\Delta)\, diag(g, -g)$ with $T$ and $T^{-1}$ real in that case.

This does not necessarily mean that $\beta$ is not identified, because the conditional variance (6.6) conveys identifying information about $\beta$. Indeed, while the matrix $\exp\left(-\beta\Delta\right)$ and $\exp(-\beta^T\Delta) = \exp(-\beta\Delta)^T$ are identical for $\beta$ and $\beta'$, the $\lambda$ matrix may be different and as a result the conditional variances $\Omega\left(\Delta; \theta\right)$ corresponding to $\beta$ and $\beta'$ may be different. To lose identification, we would need to find a pair $(\beta', \sigma')$ which produce the same $(m, \Omega)$ as $(\beta, \sigma)$. Let $v = \sigma\sigma^T$ and $v' = \sigma'\sigma'^T$. Identical conditional variances under both sets of parameters would require that

$$v' = v + \frac{2\pi i}{\Delta}\left(T \cdot diag(g, -g) \cdot T^{-1} \cdot \lambda + \lambda \cdot T \cdot diag(g, -g) \cdot T^{-1}\right).$$

Such a matrix $v'$ always exist but, as pointed out by Hansen and Sargent (1983), except in degenerate cases, there is at most a finite number of integers $g$ for which $v'$ is positive definite (which is necessary since $v' = \sigma'\sigma'^T$). Hence the identification problem is not as severe as it first seems from looking at the infinite number of solutions to the equation $\exp\left(-\beta\Delta\right) = B$ when $\beta$ has complex eigenvalues.

But in any event, if we wish to identify the parameters in $\theta$ from discrete data sampled at the given time interval $\Delta$, then we must restrict the set of admissible parameter values $\Theta$. For instance, we may restrict $\Theta$ in such a way that that the mapping $\beta \mapsto \exp\left(-\beta\Delta\right)$ is invertible, for instance by restricting the admissible parameter matrices $\beta$ to have real eigenvalues. This will be the case for example if we restrict attention to matrices $\beta$ which are triangular (and of course have real elements). For the rest of this discussion, I will assume that $\Theta$ has been restricted in such a way.

By applying Proposition 1, we see that the process $X$ is reducible, and that $\gamma\left(x; \theta\right) = \sigma^{-1}x$ so

$$
\begin{aligned}
dY_t &= \left(\sigma^{-1}\beta\alpha - \sigma^{-1}\beta\sigma Y_t\right)dt + dW_t \\
&= \sigma^{-1}\beta\sigma\left(\sigma^{-1}\alpha - Y_t\right)dt + dW_t \\
&\equiv \kappa\left(\gamma - Y_t\right)dt + dW_t
\end{aligned}
\tag{6.7}
$$

where

$$\gamma = \sigma^{-1}\alpha = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}, \quad \kappa = \sigma^{-1}\beta\sigma = \begin{pmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{pmatrix}.$$

22

One can therefore apply Theorem 1 which gives:

$$C_Y^{(-1)}(y|y_0;\theta) = -\tfrac{1}{2}(y_1 - y_{01})^2 - \tfrac{1}{2}(y_2 - y_{02})^2$$

$$C_Y^{(0)}(y|y_0;\theta) = -\tfrac{1}{2}(y_1 - y_{01})((y_1 + y_{01} - 2\gamma_1)\kappa_{11} + (y_2 + y_{02} - 2\gamma_2)\kappa_{12})$$
$$- \tfrac{1}{2}(y_2 - y_{02})((y_1 + y_{01} - 2\gamma_1)\kappa_{21} + (y_2 + y_{02} - 2\gamma_2)\kappa_{22})$$

$$C_Y^{(1)}(y|y_0;\theta) = \tfrac{1}{2}\left(\kappa_{11} - ((y_{01} - \gamma_1)\kappa_{11} + (y_{02} - \gamma_2)\kappa_{12})^2 + \kappa_{22} - ((y_{01} - \gamma_1)\kappa_{21} + (y_{02} - \gamma_2)\kappa_{22})^2\right)$$
$$- \tfrac{1}{2}(y_1 - y_{01})\left((y_{01} - \gamma_1)(\kappa_{11}^2 + \kappa_{21}^2) + (y_{02} - \gamma_2)(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22})\right)$$
$$+ \tfrac{1}{24}(y_1 - y_{01})^2\left(-4\kappa_{11}^2 + \kappa_{12}^2 - 2\kappa_{12}\kappa_{21} - 3\kappa_{21}^2\right)$$
$$- \tfrac{1}{2}(y_2 - y_{02})\left((y_{01} - \gamma_1)(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22}) + (y_{02} - \gamma_2)(\kappa_{12}^2 + \kappa_{22}^2)\right)$$
$$+ \tfrac{1}{24}(y_2 - y_{02})^2\left(-4\kappa_{22}^2 + \kappa_{21}^2 - 2\kappa_{12}\kappa_{21} - 3\kappa_{12}^2\right)$$
$$- \tfrac{1}{3}(y_1 - y_{01})(y_2 - y_{02})(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22})$$

$$C_Y^{(2)}(y|y_0;\theta) = -\tfrac{1}{12}\left(2\kappa_{11}^2 + 2\kappa_{22}^2 + (\kappa_{12} + \kappa_{21})^2\right)$$
$$+ \tfrac{1}{6}(y_1 - y_{01})(\kappa_{12} - \kappa_{21})\left((y_{01} - \gamma_1)(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22}) + (y_{02} - \gamma_2)(\kappa_{12}^2 + \kappa_{22}^2)\right)$$
$$+ \tfrac{1}{12}(y_1 - y_{01})^2(\kappa_{12} - \kappa_{21})(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22})$$
$$+ \tfrac{1}{6}(y_2 - y_{02})(\kappa_{21} - \kappa_{12})\left((y_{01} - \gamma_1)(\kappa_{11}^2 + \kappa_{21}^2) + (y_{02} - \gamma_2)(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22})\right)$$
$$+ \tfrac{1}{12}(y_2 - y_{02})^2(\kappa_{21} - \kappa_{12})(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22})$$
$$+ \tfrac{1}{12}(y_1 - y_{01})(y_2 - y_{02})(\kappa_{12} - \kappa_{21})(\kappa_{22}^2 + \kappa_{12}^2 - \kappa_{11}^2 + \kappa_{21}^2)$$

Because this is one of the few multivariate models with a known closed-form density, the Ornstein-Uhlenbeck process can serve as a useful benchmark to examine the accuracy of the expansions. Table 1 reports the results of 1,000 Monte Carlo simulations comparing the distribution of the maximum-likelihood estimator $\hat{\theta}^{(EXACT)}$ based on the exact transition density for this model, around the true value of the parameters $\theta_0$, to the distribution of the difference between the exact MLE $\hat{\theta}^{(EXACT)}$ and the approximate MLE $\hat{\theta}^{(2)}$ based on the expansion with $K = 2$ terms shown above. The results in the table show that the difference $\hat{\theta}^{(EXACT)} - \hat{\theta}^{(2)}$ is several orders of magnitude smaller than the difference $\hat{\theta}^{(EXACT)} - \theta_0$ due to the sampling noise.

## 6.2 A Stochastic Volatility Model

Consider as a second example the prototypical stochastic volatility model

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \end{pmatrix} = \begin{pmatrix} \mu \\ \kappa(\alpha - X_{2t}) \end{pmatrix} dt + \begin{pmatrix} \gamma_{11}\exp(X_{2t}) & 0 \\ 0 & \gamma_{22} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \end{pmatrix} \tag{6.8}$$

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 (6.8) is in general not reducible, so I will apply the method of Theorem 2 to derive the

23

expansion. The expansion at order $K = 3$ is given by (5.1), with the coefficients $C_X^{(j_k, k)}$, $k = -1, 0, ..., 3$ given by:

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

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

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

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

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

Finally, while in many instance financial econometricians are willing to let $X_{2t}$ denote an observable volatility variable (option-implied from the underlying asset's option price, direct observation of volatility derivatives contracts such as the VIX, or other sources), if the variable $X_{2t}$ is not observable (latent) then

the transition density $p_X$ cannot be used directly in (2.2). Instead, the latent variable must be integrated out from the joint likelihood of prices and volatility in order to obtain the likelihood function to be maximized. Alternatively, Bayesian methods can make use of $p_X$.

## 6.3   Multivariate Term Structure Models

Aït-Sahalia and Kimmel (2002) apply the method of this paper to the class of affine yield models for the term structure of interest rates. They derive the likelihood expansions for the nine canonical models of Dai and Singleton (2000). For instance, in dimension $m = 3$, the four canonical models are respectively

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

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

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

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

Likelihood expansions for all these models are given in Aït-Sahalia and Kimmel (2002), as well as a Monte Carlo investigation of the properties of maximum-likelihood estimators of the parameters derived from these expansions. They show that error due to replacing the exact transition density (for the models where it is known) with this paper's approximation is again several orders of magnitude smaller than the uncertainty in the parameter estimates due to the sampling noise, and that maximum-likelihood estimates are substantially more efficient (as expected from standard asymptotic theory and the Cramer-Rao lower bound) than alternative estimates for these models.

# 7   Conclusions

This paper provides a method to derive closed-form expansions to the likelihood function of arbitrary multivariate diffusions. The multivariate diffusion setting presents many challenges, including the fact that not all diffusions are reducible. Nevertheless, the paper provides a method that delivers closed form likelihood

expansions whether the diffusion is reducible or not. I hope that this will contribute to making maximum-likelihood the method of choice for estimating diffusion models with discretely sampled data, as is the case for other time series models.

# References

AÏT-SAHALIA, Y. (2002): "Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223–262.

AÏT-SAHALIA, Y., AND R. KIMMEL (2002): "Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions," Discussion paper, Princeton University.

AÏT-SAHALIA, Y., AND P. A. MYKLAND (2000): "The Effects of Random and Discrete Sampling When Estimating Continuous-Time Diffusions," Discussion paper, Princeton University.

AZENCOTT, R. (1984): "Densité des Diffusions en Temps Petit: Développements Asymptotiques, Partie I," in *Seminar on Probability XVIII, Lecture Notes in Mathematics 1059*, pp. 402–498. Springer, Berlin.

DAI, Q., AND K. J. SINGLETON (2000): "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943–1978.

EDWARDS, C. H. (1973): *Advanced Calculus of Several Variables*. Dover, New York.

EGOROV, A., H. LI, AND Y. XU (2001): "Maximum Likelihood Estimation of Time Inhomogeneous Diffusions," Discussion paper, Cornell University.

HAASER, N. B., AND J. A. SULLIVAN (1991): *Real Analysis*. Dover, New York.

HANSEN, L. P., AND T. J. SARGENT (1983): "The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities," *Econometrica*, 51, 377–387.

HASMINSKII, R. Z. (1980): *Stochastic Stability of Differential Equations*. Sijthoff and Noordhoff, The Netherlands.

JENSEN, B., AND R. POULSEN (2002): "Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques," *Journal of Derivatives*, 9, 1–15.

KARATZAS, I., AND S. E. SHREVE (1991): *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York.

MCCULLAGH, P. (1987): *Tensor Methods in Statistics*. Chapman and Hall, London, U.K.

PHILIPS, P. C. B. (1973): "The Problem of Identification in Finite Parameter Continuous Time Models," *Journal of Econometrics*, 1, 351–362.

SCHAUMBURG, E. (2001): "Maximum Likelihood Estimation of Jump Processes with Applications to Finance," Ph.D. thesis, Princeton University.

STROOCK, D. W., AND S. R. S. VARADHAN (1979): *Multidimensional Diffusion Processes*. Springer-Verlag, New York.

SUNDARESAN, S. M. (2000): "Continuous-Time Finance: A Review and an Assessment," *Journal of Finance*, 55, 1569–1622.

SWART, J. M. (2001): "A 2-Dimensional SDE Whose Solutions Are Not Unique," *Electronic Communications in Probability*, 6, 67–71.

VARADHAN, S. R. S. (1967): "Diffusion Processes in a Small Time Interval," *Communications in Pure and Applied Mathematics*, 20, 659–685.

WATANABE, S., AND T. YAMADA (1971): "On the Uniqueness of Solutions of Stochastic Differential Equations II," *Journal of Mathematics of Kyoto University*, 11, 553–563.

WITHERS, C. S. (2000): "A Simple Expression for the Multivariate Hermite Polynomials," *Statistics and Probability Letters*, 47, 165–169.

YAMADA, T., AND S. WATANABE (1971): "On the Uniqueness of Solutions of Stochastic Differential Equations," *Journal of Mathematics of Kyoto University*, 11, 155–167.

# Appendix: Proofs

## A   Proof of Proposition 1

Suppose that a transformation $\gamma(x;\theta) = (\gamma_1(x;\theta),...,\gamma_m(x;\theta))^T$ exists and define $Y_t \equiv \gamma(X_t;\theta)$. By Itô's Lemma, the diffusion matrix of $Y$ is

$$\sigma_Y(Y_t;\theta) = \nabla\gamma(X_t;\theta)\,\sigma(X_t;\theta).$$

For $\sigma_Y$ to be $Id$, it must therefore be that

$$\nabla\gamma(X_t;\theta) = \sigma^{-1}(x;\theta).$$

Thus

$$\sigma_{ij}^{-1}(x;\theta) = \frac{\partial\gamma_i(x;\theta)}{\partial x_j}, \tag{A.1}$$

hence

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

for all $(i,j,k) = 1,...,m$. Continuity of the second order partial derivatives is required for the order of differentiation to be interchangeable. Here, we have infinite differentiability.

Conversely, suppose that $\sigma^{-1}$ satisfies (3.4). Then, for each $i = 1,...,m$, use row $i$ of the matrix $\sigma^{-1}$, $\sigma_{i\cdot}^{-1} = \left[\sigma_{i,j}^{-1}\right]_{,j=1,...,m}$, to define the differential $1-$form

$$\omega_i = \sum_{j=1}^m \sigma_{ij}^{-1} dx_j$$

and calculate its differential, the differential $2-$form $d\omega_i$. We have that

$$
\begin{aligned}
d\omega_i &= \sum_{j=1}^m d(\sigma_{ij}^{-1}) \wedge dx_j = \sum_{j=1}^m \left\{\sum_{k=1}^m \frac{\partial\sigma_{ij}^{-1}}{\partial x_k} dx_k\right\} \wedge dx_j \\
&= \sum_{j=1}^m \sum_{k=j+1}^m \left\{\frac{\partial\sigma_{ij}^{-1}}{\partial x_k} - \frac{\partial\sigma_{ik}^{-1}}{\partial x_j}\right\} dx_k \wedge dx_j
\end{aligned}
$$

since $dx_j \wedge dx_k = -dx_k \wedge dx_j$ and $dx_j \wedge dx_j = 0$ (for notation and definitions of differential forms, see e.g., Chapter V in Edwards (1973)).

Thus condition (3.4) implies that $d\omega_i = 0$, that is the differential $1-$form $\omega_i$ is closed on $\mathcal{S}_X$. Note also that because of its form, the domain $\mathcal{S}_X$ is open and star-shaped (meaning that there exists a point $w$ in its interior such that for every $x \in \mathcal{S}_X$ the line segment from $x$ to $w$ is contained in $\mathcal{S}_X$). Therefore by Poincaré's Lemma (see e.g., Theorem V.8.1 in Edwards (1973)) the form $\omega_i$ is exact, i.e., there exists a differential $0-$form $\gamma_i$ such that $d\gamma_i = \omega_i$. In other words, for each row $i$ of the matrix $\sigma^{-1}$ there exists a function $\gamma_i$ defined by

$$\gamma_i(x;\theta) = \int^{x_j} \sigma_{ij}^{-1}(x;\theta)\, dx_j$$

(the choice of the index $j$ is irrelevant) which satisfies (A.1), the required differentiability properties and is invertible. The function $\gamma$ is then defined by each of its $d$ components $\gamma_i$, $i = 1,...,m$. By construction, $Y_t \equiv \gamma(X_t;\theta)$ has unit diffusion and therefore $X$ is reducible.

29

# B  Proof of Theorem 1

The expression for the coefficients is obtained by computing a Taylor expansion using the multivariate generator (4.19), in effect extending the approach used in the univariate case by Aït-Sahalia (2002). This process establish the form of the solution. But as is often the case when a differential operator is involved, it is easier to verify that a given functional form (in this case established using the generator) is the right solution. Indeed, to show that (4.20) with the coefficients given in the statement of Theorem 1 represent indeed the Taylor expansion in $\Delta$ of the log-density function $l_Y$, at order $K-1$, it suffices to verify that the difference between the left and right hand sides in the Fokker-Planck-Kolmogorov (FPK) forward and backward partial differential equations is of order $\Delta^K$.

The forward and backward FPK equations for $p_Y$ are respectively:

$$
\frac{\partial p_Y\left(\Delta, y|y_0; \theta\right)}{\partial \Delta} = -\sum_{i=1}^{m} \frac{\partial}{\partial y_i}\left\{\mu_{Yi}\left(y; \theta\right) p_Y\left(\Delta, y|y_0; \theta\right)\right\}
$$
$$
+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial^2}{\partial y_i \partial y_j}\left\{v_{ij}\left(y; \theta\right) p_Y\left(\Delta, y|y_0; \theta\right)\right\} \tag{B.1}
$$

$$
\frac{\partial p_Y\left(\Delta, y|y_0; \theta\right)}{\partial \Delta} = \sum_{i=1}^{m} \mu_{Yi}\left(y_0; \theta\right) \frac{\partial p_Y\left(\Delta, y|y_0; \theta\right)}{\partial y_{0i}} + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial^2 p_Y\left(\Delta, y|y_0; \theta\right)}{\partial y_{0i} \partial y_{0j}} \tag{B.2}
$$

Define $F_Y\left(\Delta, y|y_0; \theta\right)$ (resp. $B_Y\left(\Delta, y|y_0; \theta\right)$) as the the difference left and right hand sides of (B.1) (resp. (B.2)), divided by $p_Y\left(\Delta, y|y_0; \theta\right)$; let $F_Y^{(K)}$ and $B_Y^{(K)}$ denote the analogous quantities when $p_Y$ is replaced by the expansion

$$
p_Y^{(K)}\left(\Delta, y|y_0; \theta\right) = (2\pi\Delta)^{-m/2} \exp\left(\frac{C_Y^{(-1)}\left(y|y_0; \theta\right)}{\Delta} + \sum_{k=0}^{K} C_Y^{(k)}\left(y|y_0; \theta\right)\frac{\Delta^k}{k!}\right) \tag{B.3}
$$

obtained by exponentiation of (4.20).

Starting with the Gaussian leading term (4.21), tedious but otherwise straightforward computations show that:

$$
F_Y^{(K)}\left(\Delta, y|y_0; \theta\right) = \sum_{k=-1}^{K-1} f_Y^{(k)}\left(y|y_0; \theta\right)\frac{\Delta^k}{k!} + O\left(\Delta^K\right)
$$

(with the convention that $(-1)! = 0! = 1$). The first term is

$$
f_Y^{(-1)}\left(y|y_0; \theta\right) = -\sum_{i=1}^{m}\left(y_i - y_{0i}\right)\mu_{Yi}\left(y; \theta\right) + \sum_{i=1}^{m}\left(y_i - y_{0i}\right)\frac{\partial C_Y^{(0)}\left(y|y_0; \theta\right)}{\partial y_i}.
$$

Solving the equation

$$
f_Y^{(-1)}\left(y|y_0; \theta\right) = 0
$$

for $C_Y^{(0)}\left(y|y_0; \theta\right)$ with the boundary condition that $C_Y^{(0)}$ be finite when going through the axes $y_j = y_{j0}$ for all $j = 1, ..., m$ yields the solution (4.22). The boundary condition serves to set the generic integration constants $\alpha_{ij}^{(0)}$ in the full solution

$$
C_Y^{(0)}\left(y|y_0; \theta\right) = \sum_{i=1}^{m}\left(y_i - y_{0i}\right)\int_0^1 \mu_{Yi}\left(y_0 + u\left(y - y_0\right); \theta\right) du + \sum_{i,j=1, \ j\neq i}^{m} \alpha_{ij}^{(0)}\frac{y_i - y_{0i}}{y_j - y_{0j}}
$$

to zero.

The next term is

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

where $G_Y^{(1)}$ is given in (4.24) and depends on the previously determined $C_Y^{(-1)}$ and $C_Y^{(0)}$. Solving the equation

$$
f_Y^{(0)}\left(y|y_0;\theta\right) = 0
$$

for $C_Y^{(1)}$, including generic integration constants $\alpha_{ij}^{(1)}$, the explicit solution is

$$
C_Y^{(1)}\left(y|y_0;\theta\right) = \int_0^1 G_Y^{(1)}\left(y_0 + u\left(y - y_0\right)|y_0;\theta\right)du + \sum_{i,j=1,\ j\neq i}^m \alpha_{ij}^{(1)}\frac{y_i - y_{0i}}{\left(y_j - y_{0j}\right)^2}
$$

which reduces to (4.23) after accounting for the same boundary condition as for $C_Y^{(0)}$.

More generally, the term $f_Y^{(k-1)}$, $k \geq 1$, is given by

$$
f_Y^{(k-1)}\left(y|y_0;\theta\right) = C_Y^{(k)}\left(y|y_0;\theta\right) + \frac{1}{k}\sum_{i=1}^m \left(y_i - y_{0i}\right)\frac{\partial C_Y^{(k)}\left(y|y_0;\theta\right)}{\partial y_i} - G_Y^{(k)}\left(y|y_0;\theta\right)
$$

where $G_Y^{(k)}$ is given in (4.25) and depends on the previously determined $C_Y^{(-1)}, C_Y^{(0)}, ..., C_Y^{(k-1)}$. Solving the equation

$$
f_Y^{(k)}\left(y|y_0;\theta\right) = 0
$$

for $C_Y^{(k)}$ (with the same boundary condition as for $C_Y^{(0)}$ and $C_Y^{(1)}$) yields the explicit solution (4.23). In this case, the full solution including generic integration constants $\alpha_{ij}$ is

$$
C_Y^{(k)}\left(y|y_0;\theta\right) = k \int_0^1 G_Y^{(k)}\left(y_0 + u\left(y - y_0\right)|y_0;\theta\right)u^{k-1}du + \sum_{i,j=1,\ j\neq i}^m \alpha_{ij}^{(k)}\frac{y_i - y_{0i}}{\left(y_j - y_{0j}\right)^{k+1}}.
$$

Thus by construction, the solution $C_Y^{(k)}$, $k = -1, 0, ..., K$ given in the statement of the theorem is such that

$$
F_Y^{(K)}\left(\Delta, y|y_0;\theta\right) = O\left(\Delta^K\right)
$$

which along with the linearity of (B.1) in $p_Y$ insures that (4.20) is a Taylor expansion of order $K - 1$ of $l_Y$. Similar calculations show that

$$
B_Y^{(K)}\left(\Delta, y|y_0;\theta\right) = O\left(\Delta^K\right).
$$

# C    Proof of Proposition 2

To establish that $l_X^{(K)}$ is the sum of the univariate components, it suffices to establish that each multivariate coefficient $C_X^{(k)}$ of the expansion is the sum of the corresponding univariate coefficients. Further, it suffices to establish this for the coefficients $C_Y^k$, since the reducibility transformation $\gamma\left(x;\theta\right)$ involves each component separately:

$$
\gamma\left(x;\theta\right) = \left(\gamma_1\left(x_1;\theta\right), ..., \gamma_m\left(x_i;\theta\right)\right)^T
$$

31

where $\gamma_i(x_i; \theta)$ is given from $\sigma_{ii}(x; \theta)$ by equation (3.3). Therefore, we need to establish that

$$C_Y^{(k)}(y|y_0; \theta) = \sum_{i=1}^m C_Y^{(k)}(y_i|y_{0i}; \theta) \tag{C.1}$$

for $k = -1, 0, ..., K$.

From (4.21), it can be seen that (C.1) is always satisfied for $k = -1$ (whether the variables are independent or not). For $k = 0$, we have from (4.22) that

$$
\begin{aligned}
C_Y^{(0)}(y|y_0; \theta) &= \sum_{i=1}^m (y_i - y_{0i}) \int_0^1 \mu_{Yi}(y_0 + u(y - y_0); \theta)\, du \\
&= \sum_{i=1}^m (y_i - y_{0i}) \int_0^1 \mu_{Yi}(y_{0i} + u(y_i - y_{0i}); \theta)\, du \\
&= \sum_{i=1}^m \int_{y_{0i}}^{y_i} \mu_{Yi}(w; \theta)\, dw \\
&= \sum_{i=1}^m C_Y^{(0)}(y_i|y_{0i}; \theta).
\end{aligned}
$$

For $k = 1$, we have

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

and for $k \geq 2$

$$
\begin{aligned}
G_Y^{(k)}(y|y_0; \theta) &= -\sum_{i=1}^m \mu_{Yi}(y_i; \theta) \frac{\partial C_Y^{(k-1)}(y_i|y_{0i}; \theta)}{\partial y_i} + \frac{1}{2} \sum_{i=1}^m \frac{\partial^2 C_Y^{(k-1)}(y_i|y_{0i}; \theta)}{\partial y_i^2} \\
&\quad + \frac{1}{2} \sum_{i=1}^m \sum_{h=0}^{k-1} \binom{k-1}{h} \frac{\partial C_Y^{(h)}(y_i|y_{0i}; \theta)}{\partial y_i} \frac{\partial C_Y^{(k-1-h)}(y_i|y_{0i}; \theta)}{\partial y_i} \\
&= \sum_{i=1}^m G_Y^{(k)}(y_i|y_{0i}; \theta)
\end{aligned}
$$

Therefore, for $k \geq 1$, we have

$$
\begin{aligned}
C_Y^{(k)}(y|y_0; \theta) &= k \int_0^1 G_Y^{(k)}(y_0 + u(y - y_0)|y_0; \theta) u^{k-1}\, du \\
&= k \sum_{i=1}^m \int_0^1 G_Y^{(k)}(y_{0i} + u(y_i - y_{0i})|y_{0i}; \theta) u^{k-1}\, du \\
&= \sum_{i=1}^m C_Y^{(k)}(y_i|y_{0i}; \theta).
\end{aligned}
$$

# D    Proof of Theorem 2

This proof proceed along the same lines as that of Theorem 1. The forward and backward FPK equations for $p_X$ are respectively:

$$\frac{\partial p_X\left(\Delta, x|x_0;\theta\right)}{\partial \Delta} = -\sum_{i=1}^{m}\frac{\partial}{\partial x_i}\left\{\mu_i\left(x;\theta\right)p_X\left(\Delta, x|x_0;\theta\right)\right\}$$

$$+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial^2}{\partial x_i\partial x_j}\left\{v_{ij}\left(x;\theta\right)p_X\left(\Delta, x|x_0;\theta\right)\right\} \tag{D.1}$$

$$\frac{\partial p_X\left(\Delta, x|x_0;\theta\right)}{\partial \Delta} = \sum_{i=1}^{m}\mu_i\left(x_0;\theta\right)\frac{\partial p_X\left(\Delta, x|x_0;\theta\right)}{\partial x_{0i}} + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}v_{ij}\left(x_0;\theta\right)\frac{\partial^2 p_X\left(\Delta, x|x_0;\theta\right)}{\partial x_{0i}\partial x_{0j}} \tag{D.2}$$

Define $F_X\left(\Delta, x|x_0;\theta\right)$ (resp. $B_X\left(\Delta, x|x_0;\theta\right)$) as the the difference left and right hand sides of (D.1) (resp. (D.2)), divided by $p_X\left(\Delta, x|x_0;\theta\right)$; let $F_X^{(K)}$ and $\tilde{F}_X^{(K)}$ (resp. $B_X^{(K)}$ and $\tilde{B}_X^{(K)}$) and denote the analogous quantities when $p_X$ is replaced by the expansions

$$p_X^{(K)}\left(\Delta, x|x_0;\theta\right) = (2\pi\Delta)^{-m/2}\exp\left(-\ln D_v\left(x;\theta\right) + \frac{C_X^{(-1)}\left(x|x_0;\theta\right)}{\Delta} + \sum_{k=0}^{K}C_X^{(k)}\left(x|x_0;\theta\right)\frac{\Delta^k}{k!}\right) \tag{D.3}$$

and

$$\tilde{p}_X^{(K)}\left(\Delta, x|x_0;\theta\right) = (2\pi\Delta)^{-m/2}\exp\left(-\ln D_v\left(x;\theta\right) + \frac{C_X^{(j_{-1},-1)}\left(x|x_0;\theta\right)}{\Delta} + \sum_{k=0}^{K}C_X^{(j_k,k)}\left(x|x_0;\theta\right)\frac{\Delta^k}{k!}\right) \tag{D.4}$$

respectively, obtained by exponentiation of (5.1) and (5.4) respectively.

We have

$$F_X^{(K)}\left(\Delta, x|x_0;\theta\right) = \sum_{k=-2}^{K-1}f_X^{(k)}\left(x|x_0;\theta\right)\frac{\Delta^k}{k!} + O\left(\Delta^K\right)$$

(with the convention that $(-2)! = 2$ and $(-1)! = 0! = 1$). The highest order term is $f_X^{(-2)}$ given by (5.14) and the coefficient function $C_X^{(-1)}$ is such that it sets $f_X^{(-2)}$ to zero. Then we have successively $C_X^{(0)}$ determined by setting $f_X^{(-1)}$ in (5.15) to zero, and more generally, given $C_X^{(-1)}$, $C_X^{(0)}$, ..., $C_X^{(k-1)}$, the expression (5.16) for $f_X^{(k-1)}$ is defined and can be set to zero to determine the next coefficient $C_X^{(k)}$.

To determine the Taylor expansions in $x - x_0$ for each coefficient $C_X^{(k)}$, $k \geq -1$, replace $C_X^{(k)}$ by $C_X^{(j_k,k)}$ in each equation in turn, starting with (5.14). calculate a Taylor expansion of $\tilde{f}_X^{(-2)}$ in $(x - x_0)$ to order $j_{-1}$. This determines a system of equations in the unknown coefficients $\gamma_i^{(-1)}$, $i \in I_{-1}$ (which appear when $C_X^{(-1)}$ is Taylor expanded as in (5.12)). By construction, there are as many equations as unknowns (both are given by the number of elements in $I_{-1}$). This system of equation can always be solved explicitly because it has the following form.

First, $\gamma_i^{(-1)} = 0$ for $tr[i] = 0, 1$ (i.e., the polynomial has no constant or linear terms) and the terms corresponding to $tr[i] = 2$ (with of course $j_{-1} \geq 2$) are:

$$\sum_{i\in I_{-1}:tr[i]=2}\gamma_i^{(-1)}\left(x_0;\theta\right)\left(x_1 - x_{01}\right)^{i_1}\left(x_2 - x_{02}\right)^{i_2}\ldots\left(x_m - x_{0m}\right)^{i_m} = -(x - x_0)^T v^{-1}(x_0;\theta)(x - x_0).$$

which is the anticipated term given the Gaussian limiting behavior of the transition density when $\Delta$ is small. Thus with $j_{-1} \geq 3$, we only need to determine the terms $\gamma_i^{(-1)}$ corresponding to $tr[i] = 3, ..., j_{-1}$.

Then, the next order coefficients in $(x - x_0)$, i.e., the coefficients corresponding to $tr[i] = 3$, each appear linearly in a separate equations. That is, we have a system

$$M_3^{(-1)}\left(x_0;\theta\right) \cdot \gamma_3^{(-1)}\left(x_0;\theta\right) = b_3^{(-1)}\left(x_0;\theta\right)$$

whose explicit solution is given by $\gamma_3^{(-1)}(x_0; \theta) = Inv[M_3^{(-1)}(x_0; \theta)] \cdot b_3^{(-1)}(x_0; \theta)$, and so on. Given the previously determined coefficients corresponding to $tr[i] = 0, ..., r$, the equations determining the coefficients for $tr[i] = r + 1$ are given by a linear system:

$$M_{r+1}^{(-1)}(x_0; \theta) \cdot \gamma_{r+1}^{(-1)}(x_0; \theta) = b_{r+1}^{(-1)}(x_0; \theta)$$

where the matrix $M_{r+1}^{(-1)}$ and the vector $b_{r+1}^{(-1)}$ are functions of the previously determined coefficients $\gamma_i^{(-1)}$ for $tr[i] = 0, ..., r$, and of course $x_0$ and the parameters $\theta$ of the process.

The same principle applies to all values of $k$. For $k = 0 : \gamma_i^{(0)} = 0$ for $tr[i] = 0$, so the polynomial has no constant term. For $k \geq 1$, the polynomials have a constant term (for $k \geq 1$, $\gamma_i^{(k)} \neq 0$ for $tr[i] = 0$ in general). The same principle applies to each equation in turn: once $C_X^{(j-1,-1)}$ is determined, a Taylor expansion of (5.15) determines the coefficients $\gamma_i^{(0)}$, $i \in I_0$, etc.

Finally, note that the term $D_v(x; \theta)$ which arose in the reducible case from the Jacobian transformation is independent of $\Delta$ and so could be built into the $C_X^{(0)}$ coefficient. Doing so however would subject it to being Taylor-expanded in $x - x_0$, which is unnecessary anyway since $D_v(x; \theta)$ is known. If $D_v(x; \theta)$ were being Taylor-expanded along with $C_X^{(j_0,0)}$ in (D.4), we would lose the property that $\tilde{p}_X^{(K)}$ also solves the backward FPK equation (D.2) to order $K - 1$ in $\Delta$. Hence the form of the log-likelihood I adopted in (5.1) with $D_v$ kept separate from $C_X^{(0)}$ is essential to obtain

$$\tilde{B}_X^{(K)}(\Delta, x|x_0; \theta) = O(\Delta^K)$$

in addition to $\tilde{F}_X^{(K)}(\Delta, x|x_0; \theta) = O(\Delta^K)$.

# E   Proof of Proposition 3

If the diffusion $X$ is reducible, then $C_X^{(k)}(x|x_0; \theta) = C_Y^{(k)}(\gamma(x; \theta)|\gamma(x_0; \theta); \theta)$. By construction (see the proof of Theorem 2), the coefficients $C_X^{(j_k,k)}$ are Taylor expansions of the coefficients $C_X^{(k)}$ (which are the expressions solutions of the equations $f_X^{(k-1)} = 0$).

34

| Parameter | $\theta^{(TRUE)}$ | $\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$ | | $\hat{\theta}^{(MLE)} - \hat{\theta}^{(2)}$ | |
|:---:|:---:|:---:|:---:|:---:|:---:|
| | | Mean | Stnd. Dev. | Mean | Stnd. Dev. |
| $\gamma_1$ | 0 | $-0.0013$ | 0.069 | $-0.0000015$ | 0.000035 |
| $\gamma_2$ | 0 | 0.00070 | 0.033 | 0.00000012 | 0.000016 |
| $\kappa_{11}$ | 5 | 0.52 | 1.17 | 0.012 | 0.0085 |
| $\kappa_{12}$ | 1 | $-0.066$ | 1.74 | 0.0087 | 0.017 |
| $\kappa_{22}$ | 5 | 0.35 | 1.50 | 0.069 | 0.029 |

**Table 1: Monte-Carlo Simulations for the Bivariate Ornstein-Uhlenbeck Model**

This table reports the results of 1,000 Monte Carlo simulations comparing the distribution of the maximum-likelihood estimator $\hat{\theta}^{(MLE)}$ based on the exact transition density for this model, around the true value of the parameters $\theta_0$, to the distribution of the difference between the exact MLE $\hat{\theta}^{(MLE)}$ and the approximate MLE $\hat{\theta}^{(2)}$ based on the expansion with $K = 2$ terms, for the process (6.7). To insure full identification, the off-diagonal term $\kappa_{21}$ is constrained to be zero. As discussed in the text, this guarantees that the eigenvalues of the mean reversion matrix are both real and avoids the aliasing problem altogether. The constraints $\kappa_{11} > 0$ and $\kappa_{22} > 0$ are imposed to insure stationarity of the process. The true values of the parameter vector $\theta = (\gamma_1, \gamma_2, \kappa_{11}, \kappa_{12}, \kappa_{22})$ used to generate the data are $\theta^{(TRUE)} = (0, 0, 5, 1, 5)$. Each of the 1,000 samples is a series of $n = 500$ weekly observations ($\Delta = 1/52$), generated using the exact discretization of the process. The results in the table show that the difference $\hat{\theta}^{(MLE)} - \hat{\theta}^{(2)}$ is several orders of magnitude smaller than the difference $\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$ due to the sampling noise.