TECHNICAL WORKING PAPER SERIES

MAXIMUM LIKELIHOOD ESTIMATION OF DISCRETELY SAMPLED DIFFUSIONS: A CLOSED-FORM APPROACH

Yacine Aït-Sahalia

Technical Working Paper 222
http://www.nber.org/papers/T0222

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1998

The latest version of this paper, including *Mathematica* code to calculate the closed-form density sequence, can be found at http://gsbwww.uchicago.edu/fac/yacine.aitsahalia/research. I am grateful to Lars Hansen, Per Mykland and Angel Serrat, as well as seminar participants at LSE for helpful discussions. Financial support from the Center for Research in Security Prices is gratefully acknowledged. All errors are mine. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Maximum Likelihood Estimation of Discretely
Sampled Diffusions: A Closed-Form Approach
Yacine Aït-Sahalia
NBER Technical Working Paper No. 222
February 1998
JEL Nos. G12, C13, C22

ABSTRACT

When a continuous-time diffusion is observed only at discrete dates, not necessarily close together, the likelihood function of the observations is in most cases not explicitly computable. Researchers have relied on simulations of sample paths in between the observations points, or numerical solutions of partial differential equations, to obtain estimates of the function to be maximized. By contrast, we construct a sequence of fully explicit functions which we show converge under very general conditions, including non-ergodicity, to the true (but unknown) likelihood function of the discretely-sampled diffusion. We document that the rate of convergence of the sequence is extremely fast for a number of examples relevant in finance. We then show that maximizing the sequence instead of the true function results in an estimator which converges to the true maximum-likelihood estimator and shares its asymptotic properties of consistency, asymptotic normality and efficiency. Applications to the valuation of derivative securities are also discussed.

Yacine Aït-Sahalia
Graduate School of Business
University of Chicago
1101 East 58th Street
Chicago, IL 60637-1561
and NBER
yacine.aitsahalia@gsb.uchicago.edu

1. Introduction

Consider a continuous-time parametric diffusion

(1.1) $$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dW_t$$

where X_t is the state variable, W_t a standard Brownian motion, $\mu(.;.)$ and $\sigma(.;.)$ are known functions and θ an unknown parameter vector in an open bounded set $\Theta \subset R^K$. Diffusion processes are widely used in theoretical financial models, for instance to represent the stochastic dynamics of asset prices, interest rates, macroeconomic factors, etc.

Obviously, the available data are always sampled discretely, while the model is written in continuous time. As discussed by Merton (1980), Lo (1988) and Melino (1994), ignoring the difference can result in inconsistent estimators. A number of econometric methods have been recently developed to estimate θ in (1.1), without requiring that a continuous record of observations be available. Some of these methods are based on simulations [Duffie and Singleton (1993), Gouriéroux, Monfort and Renault (1993), Gallant and Tauchen (1997), Pedersen (1995), Santa-Clara (1995) and for applications Honoré (1997) and Andersen and Lund (1996)], others on the generalized method of moments [Hansen and Scheinkman (1995), Bibby and Sørensen (1995), Conley et al. (1997)], nonparametric density-matching [Aït-Sahalia (1996a, 1996b), Stanton (1997)] or random sampling of the process to generate moment conditions [Duffie and Glynn (1997)].

As in most contexts, provided we trust the specification (1.1), maximum-likelihood is the method of choice --with only one caveat here: in general, the likelihood function of discrete observations generated by (1.1) cannot be calculated explicitly! Let $p_X(\Delta, x \mid x_0;\theta)$ denote the conditional density of $X_{t+\Delta}=x$ given $X_t=x_0$ induced by the model (1.1), also called the transition function. We observe the process at dates $\{t = i\Delta \mid i = 0,\ldots,n\}$, where Δ is fixed. The Markovian nature of (1.1) implies that the log-likelihood function has the simple form

(1.2) $$\ell_n(\theta) \equiv n^{-1}\sum_{i=1}^{n} \mathrm{Ln}\left\{p_X\left(\Delta, X_{i\Delta} \mid X_{(i-1)\Delta};\theta\right)\right\}$$

but the true density p_X is in general unknown [for a list of the rare exceptions, see Wong (1964); in finance, the models of Black and Scholes (1973), Vasicek (1978) and Cox, Ingersoll and Ross (1985) rely on some of the existing closed-form expressions].

If sampling of the process were continuous, the situation would be radically simpler. First, the likelihood function is known by means of classical absolutely continuous changes of measures [see e.g., Basawa and Prakasa Rao (1980)]. Second, if we are willing to let the sampling interval go to zero, then expansions for the transition function "in small time" are available in the statistical literature [see e.g., Azencott (1981)].

With fixed sampling, two methods have been proposed in the literature to compute p_X. They either involve solving numerically the Fokker-Planck-Kolmogorov partial differential equation [see Lo (1988)] or "filling-in the blanks" between the observation dates by simulating a large number of sample paths along which the process is sampled very finely [see Pedersen (1995) and Santa-Clara (1995)]. Dacunha-Castelle and Florens-Zmirou (1986) calculate expressions for the transition function which involve functionals of a Brownian Bridge, and can potentially be simulated. Neither method produces a closed-form expression to be maximized over θ: the criterion function takes either the form of an implicit solution to a partial differential equation, or an infinite sum over simulated sample paths. In addition, when looking for the maximum value of (1.2), these numerical steps, whether solving the PDE or recalculating averages over simulated sample paths, have to be repeated a number of times: each time the value of θ changes infinitesimally as part of the likelihood maximization algorithm. As a result, these methods are not trivial to implement in practice.

We can think of both the PDE and simulation methods as delivering a sequence of approximations to the true likelihood function, which become increasingly accurate as some control parameter J tends to infinity: for instance, J indexes the number of points on a grid used in a typical discretization scheme to solve numerically the PDE solved by p_X; or J indexes the sampling frequency used in the fine discretization of the sample path between two successive observations, and the number of such simulated paths.

By contrast, we propose here a method to estimate θ by maximum-likelihood which involves neither the numerical solution of a PDE nor any simulations of sample paths. Like the PDE and simulation-based methods, we also construct a sequence $\ell_n^{(J)}$, J = 1,2,... of approximations to the log-likelihood function ℓ_n, but our sequence is in *closed-form*. We then show that $\ell_n^{(J)}$ converges to ℓ_n as J increases, and prove that maximizing $\ell_n^{(J)}$ in lieu of the unknown ℓ_n results in an estimator $\hat{\theta}_n^{(J)} \equiv \arg\max_{\theta\in\Theta} \ell_n^{(J)}(\theta)$ which converges to the true maximum-likelihood estimator $\hat{\theta}_n \equiv \arg\max_{\theta\in\Theta} \ell_n(\theta)$ as J gets larger. Therefore, in practice, it suffices to take a single value of J --large, although we provide empirical evidence for models that are relevant in finance that J=4 is amply adequate-- and maximize $\ell_n^{(J)}$. Since the expression of $\ell_n^{(J)}$ is explicit, the effort involved is minimal -- identical to a standard maximum-likelihood problem with a known likelihood function.

$\{X_t/t\geq 0\}$, and with payoff function $\Psi(X_\Delta)$ at some future date Δ. For simplicity, assume that the underlying asset is traded, so that its risk-neutral dynamics have the form

$$(1.3)\quad dX_t/X_t = \{r-\delta\}dt + \sigma(X_t;\theta)dW_t$$

where r is the riskfree rate and δ the dividend rate paid by the asset, both constant again for simplicity.

As is well-known, when markets are dynamically complete, the only price of the derivative security that is compatible with the absence of arbitrage opportunities is

$$(1.4)\quad P_0 = e^{-r\Delta}E\left[\Psi(X_\Delta)\middle| X_0 = x_0\right] = e^{-r\Delta}\int_0^{+\infty}\Psi(x)p_X(\Delta, x \mid x_0;\theta)dx$$

where p_X is the transition function (or risk-neutral density, or state-price density) induced by (1.3).

The Black-Scholes-Merton option pricing formula is the prime example of (1.4), when $\sigma(S)=\sigma$ is constant. The corresponding p_X is known in closed-form (as a lognormal density) and so the integral in (1.4) can be evaluated explicitly for specific payoff functions [see also Cox and Ross (1976)]. In general, of course, no known expression for p_X is available and one must rely on numerical methods such as solving numerically the PDE satisfied by the derivative price, or Monte-Carlo integration of (1.3). These methods are exact parallels to the two existing approaches to maximum-likelihood estimation that we described earlier.

Here, given the sequence $\{p_X^{(J)}/J\geq 1\}$ of approximations to p_X, our valuation of the derivative security would be based on the explicit formula

$$(1.5)\quad P_0^{(J)} = e^{-r\Delta}\int_0^{+\infty}\Psi(x)p_X^{(J)}(\Delta, x \mid x_0;\theta)dx.$$

Formulas of the type (1.5) have been proposed in the finance literature [see e.g., Jarrow and Rudd (1982)] and justified as "corrections" to the Black-Scholes-Merton formula. There is however an important difference between what we propose and the existing formulae: the latter are based on calculating the integral in (1.4) with an ad hoc density $\widehat{p}_X$ --typically adding free skewness and kurtosis parameters to the lognormal density, so as to allow for departures from the Black-Scholes-Merton formula. In doing so, these formulas entirely ignore the underlying dynamic model (1.3) for the asset price, whereas our method gives in closed-form *the* option pricing formula (of order of precision J, for each J) which corresponds to the given dynamic model (1.3). For instance, we can explore how changes in the specification of the volatility function $\sigma(x;\theta)$ affect the derivative price, which is obviously impossible when the specification of the density $\tilde{p}_X$ to be used in lieu of p_X is unrelated to (1.3).

The paper is organized as follows. In Section 2, we construct the sequence of density approximations for any given parametric specification (1.1) and show that they converge, in a strong sense, to the true density function. We then prove in Section 3 that maximizing the approximation to the likelihood function produces an estimator which can be made arbitrarily close to the true (but not explicitly computable) maximum-likelihood estimator, and shares its asymptotic properties. In Section 4, we show how to calculate in closed-form the coefficients of the approximations.

The reader primarily interested in applying the result may go directly to Section 5. There we give explicitly the first six terms of the approximating density sequence, $\{ p_X^{(J)} / J=1,\ldots,6\}$, and provide a number of examples which show that as a practical matter stopping after the first three terms is sufficient. Section 6 concludes. All proofs are in the Appendix.

2. A Sequence of Expansions of the Transition Function

To understand the construction of our sequence of approximations to p_X, the following analogy may be helpful. Consider the density of the standardized sum of random variables to which the Central Limit Theorem (CLT) apply, and its classical Edgeworth expansion. The convergence of such an expansion is understood in the sense that the number of corrective terms to the Normal density is fixed while the number of observations goes to infinity. In fact, for a fixed sample size, the Edgeworth expansion will typically diverge as more and more corrective terms are added, unless the density of each of these random variables was "close to" a Normal density to start with. There is no need at this point to make this statement precise. In our context, imposing that Δ remain fixed is equivalent to imposing that the number of observations in the CLT remain fixed. By contrast, if Δ goes to zero then p_X converges to a Normal, just like the distribution of the standardized sum in the CLT converges to a Normal as the sample size goes to infinity.

Therefore, in general, the density p_X cannot be approximated for fixed Δ around a Normal density by standard series such as Hermite expansions, because the distribution of X is in general too far from that of a Normal. For instance, if X follows a geometric Brownian motion, the right tail of the corresponding log-normal density p_X is too large for its Hermite expansion to converge. Indeed, the tail is of order $x^{-1}\exp\left\{-(\mathrm{Ln}(x))^2\right\}$ as x tends to $+\infty$. From the work of Cramér (1925), it is known that Hermite series only converge when the density to be expanded is sufficiently close to a Normal density. For instance, an explicit calculation shows that the expansion of any N(0,v) density diverges if v>2, and hence the class of densities functions to which Hermite expansions can be applied is quite limited.

The idea in this paper is to circumvent this difficulty by making two successive transformations of X into a variable Z whose density p_Z happens to belong to the class of densities for which the Hermite series converges. We next construct the converging sequence of approximations for p_Z. We can then revert the transformation $X \to Z$, and through the process of transforming Z back into X, deform the approximation of p_Z to obtain an expansion for the density p_X around a *deformed* Normal density. In Theorem 1, we will prove that such an expansion converges uniformly to the unknown p_X.

2.1 Assumptions and First Transformation

We start by making standard regularity assumptions on the functions μ and σ. We denote by $D_X = (\underline{x}, \bar{x})$ the domain of the diffusion X. We will consider the two cases where $D_X = (-\infty,+\infty)$ and $D_X = (0,+\infty)$. The latter case is often the most relevant in finance, when considering models for asset prices or nominal interest rates. In addition, the function σ is often specified in financial models in such a way that $\sigma(0;\theta)=0$ and μ and/or σ violate the linear growth conditions near the boundaries. For these reasons, we will devise a set of assumptions where we replace growth conditions (without constrain on the sign of the drift function near the boundaries) with assumptions on the sign of the drift near the boundaries (without restriction on the growth of the coefficients).

Assumption 1 (Smoothness of the Coefficients): The functions $\mu(x;\theta)$ and $\sigma(x;\theta)$ are infinitely differentiable in x on D_X, and twice continuously differentiable in θ in the open and bounded parameter space $\Theta \subset R^K$.

Assumption 2 (Non-Degeneracy of the Diffusion):

1. If $D_X = (-\infty,+\infty)$, there exists a constant c such that $\sigma(x;\theta) > c > 0$ for all $x \in D_X$ and $\theta \in \Theta$.

2. If $D_X = (0,+\infty)$, we allow for the possible local degeneracy of σ at x=0: if $\sigma(0;\theta)=0$, then there exist constants ξ_0, $\omega \geq 0$, ρ such that $\sigma(x;\theta) \geq \omega x^{\rho}$ for all $0 < x \leq \xi_0$ and $\theta \in \Theta$. Away from 0, σ is non-degenerate, that is: for each $\xi > 0$, there exists a constant c_ξ such that $\sigma(x;\theta) \geq c_\xi > 0$ for all $x \in [\xi,+\infty)$ and $\theta \in \Theta$.

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

(2.1) $Y_t \equiv \gamma(X_t;\theta) = \int^{X_t} du/\sigma(u;\theta)$

where any primitive of the function $1/\sigma$ may be selected, i.e., the constant of integration is irrelevant. Because $\sigma>0$ on D_X, the function γ is increasing and invertible. It maps D_X into $D_Y = (\underline{y}, \bar{y})$, the domain of Y, where $\underline{y} \equiv \lim_{x\to\underline{x}} \gamma(x;\theta)$ and $\bar{y} \equiv \lim_{x\to\bar{x}} \gamma(x;\theta)$. For example, if $D_X = (0,+\infty)$ and $\sigma(x;\theta) = x^\rho$, then $Y_t = (1-\rho)X_t^{1-\rho}$ if $0<\rho<1$ [so $D_Y = (0,+\infty)$], $Y_t = \mathrm{Ln}(X_t)$ if $\rho=1$ [so $D_Y = (-\infty,+\infty)$] and $Y_t = -(\rho-1)X_t^{-(\rho-1)}$ if $\rho>1$ [so $D_Y = (-\infty,0)$]. We suppose that the parameter space Θ is such that D_Y in independent of θ in Θ. This restriction on Θ is inessential, but it helps keep the notation simple.

By applying Itô's Lemma, Y has unit diffusion:

(2.2) $dY_t = \mu_Y(Y_t;\theta)dt + dW_t$

where

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

We say that an infinitely differentiable function f has at most polynomial growth if there exists an integer $p\geq 0$ such that $|y|^{-p}|f(y)|$ is bounded above in a neighborhood of infinity. If p=1, we say more specifically at most linear, and if p=2 at most quadratic. If there exists a constant $\lambda\geq 0$ such that $\exp\{-\lambda|y|\}|f(y)|$ is bounded above in a neighborhood of infinity then we say that f has at most exponential growth.

Assumption 3 below restricts the behavior of the function μ_Y and its derivatives near the boundaries of D_Y. It is formulated in terms of the function μ_Y for reasons of convenience, but the equivalent formulation directly in terms of the original functions μ and σ is obvious from (2.3). Let $g(y;\theta) \equiv -(\mu_Y^2(y;\theta) + \partial\mu_Y(y;\theta)/\partial y)/2$.

Assumption 3 (Boundary Behavior)**:** For all $\theta\in\Theta$, $\mu_Y(y;\theta)$, $\partial\mu_Y(y;\theta)/\partial y$ and $\partial^2\mu_Y(y;\theta)/\partial y^2$ have at most exponential growth near the infinity boundaries and $\lim_{y\to\underline{y} \text{ or } \bar{y}} g(y;\theta) < +\infty$.

1. Left Boundary:

i. If $\underline{y} = 0^+$, there exist constants ε_0, κ, α such that for all $0 < y \leq \varepsilon_0$ and $\theta\in\Theta$, $\mu_Y(y;\theta) \geq \kappa y^{-\alpha}$ where either $\alpha>1$ and $\kappa>0$ or $\alpha=1$ and $\kappa\geq 1/2$.

ii. If $\underline{y} = -\infty$, there exist constants $E_0>0$ and $K>0$ such that for all $y \leq -E_0$ and $\theta\in\Theta$, $\mu_Y(y;\theta) \geq Ky$.

2. Right Boundary:

i. If $\bar{y} = +\infty$, there exist constants $E_0>0$ and $K>0$ such that for all $y \geq E_0$ and $\theta \in \Theta$, $\mu_Y(y;\theta) \leq Ky$.

ii. If $\bar{y} = 0^-$, there exist constants ε_0, κ, α such that for all $0 > y \geq -\varepsilon_0$ and $\theta \in \Theta$, $\mu_Y(y;\theta) \leq -\kappa |y|^{-\alpha}$ where either $\alpha>1$ and $\kappa>0$ or $\alpha=1$ and $\kappa \geq 1/2$.

In Section 5.2, we will give examples to illustrate the applicability of Assumption 3 to the typical models considered in finance. At this point however, the following remarks can help demonstrate the generality of Assumption 3:

(1) Note that the upper bound $\lim_{y \to \underline{y} \text{ or } \bar{y}} g(y;\theta) < +\infty$ does not restrict g from going to $-\infty$ near the boundaries.

(2) Similarly, Assumption 3 does not preclude μ_Y from going to $-\infty$ very fast near $\bar{y}$, and similarly, from going to $+\infty$ very fast near $\underline{y}$. Assumption 3 only restricts how large μ_Y can grow if it has the "wrong" sign, i.e., if it is positive near $\bar{y}$ and negative near $\underline{y}$: then linear growth is the maximum possible rate. If μ_Y has the "right" sign then the process is being pulled back away from the boundaries and we do not restrict how fast mean-reversion occurs [up to an exponential rate for technical reasons].

(3) The constraints on the behavior of the function μ_Y are essentially the best possible. For example, if μ_Y has the "wrong" sign near an infinity boundary, and grows faster than linearly, then Y explodes in finite time. Near a zero boundary, say 0^+, if there exists $\kappa>0$ and $\alpha<1$ such that $\mu_Y(y;\theta) \leq ky^{-\alpha}$ in a neighborhood of 0^+ then 0 and negative values become attainable.

(4) Finally, we can fully characterize the boundary behavior of the diffusion Y implied by the assumptions made:

Lemma 1: Under Assumptions 1-3, if $+\infty$ is a boundary then it is natural if, near $+\infty$, $|\mu_Y(y;\theta)| \leq Ky$ and entrance if $\mu_Y(y;\theta) \leq -Ky^{\beta}$ for some $\beta>1$. If $-\infty$ is a boundary then it is natural if, near $-\infty$, $|\mu_Y(y;\theta)| \leq K|y|$ and entrance if $\mu_Y(y;\theta) \geq K|y|^{\beta}$ for some $\beta>1$. If 0 is a boundary (either 0^+ or 0^-), then it is entrance.

Both entrance and natural boundaries are unattainable [see Feller (1952) or Karlin and Taylor (1981, Section 15.6) for the definition of boundaries]. Natural boundaries can neither be reached in finite time, nor can the diffusion be started from there. Entrance boundaries, such as 0^+, cannot be reached starting from an interior point in $D_Y = (0,+\infty)$, but it is possible for Y to begin there. In that case, the process moves quickly away from 0 and never returns there. Typically, economic intuition says little about how the process would behave if it were to start at the boundary, or whether that is even possible, and hence it is sensible to allow both types of boundary behavior.

(5) Assumption 3 neither requires nor implies that the process is stationary. When *both* boundaries of the domain D_Y are entrance boundaries then the process is necessarily stationary with unconditional density

$$(2.4)\quad \pi(y;\theta) \equiv \exp\left\{2\int_{\underline{y}}^{y}\mu_Y(u;\theta)du\right\}\Big/\int_{\underline{y}}^{\bar{y}}\exp\left\{2\int_{0}^{v}\mu_Y(u;\theta)du\right\}dv,$$

provided that the initial random variable Y_0 is itself distributed with density (2.4). When at least one of the boundaries is natural, stationarity is neither precluded nor implied. For instance, both an Ornstein-Uhlenbeck process, where $\mu_Y(y;\theta)=\beta(\alpha-y)$, and a Brownian motion, where $\mu_Y(y;\theta)=0$, satisfy the assumptions made, and both have natural boundaries at $-\infty$ and $+\infty$. Yet the former process is stationary, due to mean-reversion, while the latter is not (null recurrent).

2.2 Further Data Transformations

While Y, thanks to its unit diffusion, is "closer" to a Normal variable than X is, in general it is not close enough to allow us to expand its conditional density around the Normal density function, due to the fact that Δ need not be small. For that reason, we need to perform a further transformation. For given $\Delta>0$, $\theta\in\Theta$ and $y_0\in R$, we define the "pseudo-normalized" increment of Y as

$$(2.5)\quad Z_t = Z(\Delta, Y_t \mid y_0;\theta) \equiv \Delta^{-1/2}\left(Y_t - y_0 - \mu_Y(y_0;\theta)\Delta\right)$$

Of course, since we do not require that $\Delta\to 0$, we make no claim regarding the degree of accuracy of this standardization device, hence the term "pseudo". It will turn out below [see Lemma 2] that for fixed Δ, Z_t defined in (2.5) happens to be close enough to a Normal variable to make it possible to create a series of expansions for its density p_Z around the Normal density function.

Let $p_Y(\Delta, y \mid y_0;\theta)$ denote the conditional density of $Y_{t+\Delta} \mid Y_t$, and define the density function

$$(2.6)\quad p_Z(\Delta, z \mid y_0;\theta) \equiv \Delta^{1/2}p_Y\left(\Delta, \Delta^{1/2}z + y_0 + \mu_Y(y_0;\theta)\Delta \,\middle|\, y_0;\theta\right)$$

Once we have constructed a sequence of approximations to the function $(z, y_0) \mapsto p_Z(\Delta, z \mid y_0;\theta)$, we will backtrack and infer a sequence of approximations to the function $(y, y_0) \mapsto p_Y(\Delta, y \mid y_0;\theta)$ by inverting (2.6):

$$\text{(2.7)} \quad p_Y(\Delta, y \mid y_0;\theta) \equiv \Delta^{-1/2} p_Z\left(\Delta, \Delta^{-1/2}\left(y - y_0 - \mu_Y(y_0;\theta)\Delta\right) \middle| y_0;\theta\right)$$

and then back to the object of interest $(x, x_0) \mapsto p_X(\Delta, x \mid x_0;\theta)$. To go from p_X to p_Y and then back from p_Y to p_X, we can compute that

$$\text{(2.8)} \quad p_X(\Delta, x \mid x_0;\theta) = \sigma(x;\theta)^{-1} \times p_Y(\Delta, \gamma(x;\theta) \mid \gamma(x_0;\theta);\theta)$$

$$\text{(2.9)} \quad p_Y(\Delta, y \mid y_0;\theta) = \sigma(\gamma^{-1}(y;\theta);\theta) \times p_X(\Delta, \gamma^{-1}(y;\theta) \mid \gamma^{-1}(y_0;\theta);\theta)$$

since

$$\begin{aligned} p_Y(\Delta, y \mid y_0;\theta) &= \frac{\partial}{\partial y} \operatorname{Prob}\left(Y_{t+\Delta} \le y \middle| Y_t = y_0;\theta\right) \\ &= \frac{\partial}{\partial y} \operatorname{Prob}\left(X_{t+\Delta} \le \gamma^{-1}(y;\theta) \middle| X_t = \gamma^{-1}(y_0;\theta);\theta\right) \\ &= \frac{\partial}{\partial y}\left[\int_{-\infty}^{\gamma^{-1}(y;\theta)} p_X\left(\Delta, x \mid \gamma^{-1}(y_0;\theta);\theta\right) dx\right] \\ &= \sigma(\gamma^{-1}(y;\theta);\theta) \times p_X(\Delta, \gamma^{-1}(y;\theta) \mid \gamma^{-1}(y_0;\theta);\theta) \end{aligned}$$

which follows from

$$\frac{\partial \gamma^{-1}(y;\theta)}{\partial y} = \frac{1}{[\partial\gamma/\partial x](\gamma^{-1}(y;\theta))} = \sigma(\gamma^{-1}(y;\theta);\theta),$$

and similarly for (2.8).

2.3 Approximation of the Transition Function of the Transformed Data

To approximate the density function p_Z, we will construct a Hermite series expansion. We have constructed the variable Z_t precisely so that it be "close" to a Normal variable, for which expansions around a Normal density can be calculated. The rest of this section makes this basic intuition rigorous. Of crucial importance in Theorem 1 is to prove that the variable Z_t is "close enough" to a Normal variable so that our expansion converges uniformly.

Define the classical Hermite polynomials by

(2.10) $H_j(z) \equiv e^{z^2/2} \frac{d^j}{dz^j}\left[e^{-z^2/2}\right]$,

let $\phi(z) \equiv e^{-z^2/2}/\sqrt{2\pi}$ be the N(0,1) density function and

(2.11) $p_Z^{(J)}(\Delta, z \mid y_0;\theta) \equiv \phi(z)\sum_{j=0}^{J} \eta_j(\Delta, y_0;\theta) H_j(z)$

be the Hermite expansion of the density function $z \mapsto p_Z(\Delta, z \mid y_0;\theta)$ (for fixed Δ, y_0 and θ). The coefficients η_j are defined by:

(2.12) $\eta_j(\Delta, y_0;\theta) \equiv (1/j!)\int_{-\infty}^{+\infty} H_j(z) p_Z^{(J)}(\Delta, z \mid y_0;\theta) dz$.

By analogy with (2.7), we then construct the sequence of approximations to p_Y as:

(2.13) $p_Y^{(J)}(\Delta, y \mid y_0;\theta) \equiv \Delta^{-1/2} p_Z^{(J)}\left(\Delta, \Delta^{-1/2}\left(y - y_0 - \mu_Y(y_0;\theta)\Delta\right) \mid y_0;\theta\right)$

and then approximate p_X by mimicking (2.8):

(2.14) $p_X^{(J)}(\Delta, x \mid x_0;\theta) \equiv \sigma(x;\theta)^{-1} p_Y^{(J)}\left(\Delta, \gamma(x;\theta) \mid \gamma(x_0;\theta);\theta\right)$.

2.4 Pointwise Convergence of the Expansion

The following theorem proves that Z is close enough to a Normal variable for the expansion (2.14) to converge uniformly as more terms are added, and that the limit is the true (but unknown) density function. Note that the sampling interval remains fixed; in particular, we do not require that $\Delta \to 0$ for the sequence $p_X^{(J)}(\Delta, x \mid x_0;\theta)$ to converge to $p_X(\Delta, x \mid x_0;\theta)$. Rather, we let the number of terms J grow:

Theorem 1: Under Assumptions 1-3, there exists $\bar{\Delta} > 0$ such that for every $\Delta \in (0, \bar{\Delta})$, $\theta \in \Theta$ and $(x, x_0) \in D_X^2$:

(2.15) $p_X^{(J)}(\Delta, x \mid x_0;\theta) \xrightarrow[J \to \infty]{} p_X(\Delta, x \mid x_0;\theta)$

In addition, the convergence is uniform in θ over Θ and in x_0 over compact subsets of D_X. If σ is non-degenerate, then the convergence is further uniform in x over the entire domain D_X. If σ is degenerate at zero, then the convergence is uniform in x in each interval of the form $[\varepsilon, +\infty)$, $\varepsilon > 0$.

The proof of Theorem 1 relies on the following lemmas:

Lemma 2: Under Assumptions 1-3, there exists $\bar{\Delta} > 0$ such that for every $\Delta \in (0, \bar{\Delta})$, there exist constants C_i, i=0,...,4 and D_0 such that for every $\theta \in \Theta$ and every $(y, y_0) \in R^2$:

$$(2.16) \quad \begin{aligned} 0 < p_Y(\Delta, y \mid y_0;\theta) \leq\ & C_0 \Delta^{-1/2} \exp\{-3(y-y_0)^2/8\Delta\} \\ & \times \exp\{C_1|y-y_0||y_0| + C_2|y-y_0| + C_3|y_0| + C_4 y_0^2\} \end{aligned}$$

and

$$(2.17) \quad \begin{aligned} |\partial p_Y(\Delta, y \mid y_0;\theta)/\partial y| \leq\ & D_0 \Delta^{-1/2} \exp\{-3(y-y_0)^2/8\Delta\} \times P(|y|,|y_0|) \\ & \times \exp\{C_1|y-y_0||y_0| + C_2|y-y_0| + C_3|y_0| + C_4 y_0^2\} \end{aligned}$$

where P is a polynomial of finite order in $(|y|,|y_0|)$, with coefficients uniformly bounded in $\theta \in \Theta$. Further, if $\mu_Y \leq 0$ near $+\infty$ and $\mu_Y \geq 0$ near $-\infty$, then $\bar{\Delta} = +\infty$.

Lemma 3: Under Assumptions 1-3, for every $\Delta \in (0, \bar{\Delta})$, every $\theta \in \Theta$ and every $y_0 \in R$, the moments

$$(2.18) \quad u_Y(\Delta \mid y_0;\theta, j) \equiv \int_{-\infty}^{+\infty} |y|^j p_Y(\Delta, y \mid y_0;\theta) dy$$

are finite for all $j \geq 0$.

Lemma 4: The polynomials H_j satisfy:

(i) $(j+1)H_j(z) = dH_{j+1}(z)/dz$ for all z in R and every integer j.

(ii) $\int_{-\infty}^{+\infty} (2\pi)^{-1/2} e^{-w^2/2} H_j(w) H_k(w) dw = \begin{cases} j! & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}$

(iii) There exists a constant K such that for all z in R and every integer j:

$$|H_j(z)| \leq K (j!)^{1/2} j^{-1/4} \{1 + |z^{5/2}/2^{5/4}|\} e^{z^2/4}.$$

We now study the properties of the sequence of maximum-likelihood estimators derived from maximizing the approximate likelihood function computed from $p_X^{(J)}$.

3. A Sequence of Approximations to the Maximum-Likelihood Estimator

With the sequence of approximation to the transition function in hand, the point of this section is to show that maximizing

(3.1) $\quad \ell_n^{(J)}(\theta) \equiv n^{-1}\sum_{i=1}^{n} \mathrm{Ln}\left\{p_X^{(J)}\left(\Delta, X_{i\Delta} \mid X_{(i-1)\Delta};\theta\right)\right\}$

(with the convention that $\mathrm{Ln}(\alpha) = -\infty$ if $\alpha \leq 0$) over θ in Θ results in an estimator $\hat{\theta}_n^{(J)}$ which converges to the true (but uncomputable in practice) maximum-likelihood estimator $\hat{\theta}_n$ as $J \rightarrow \infty$. We further prove that when the sample size gets large ($n \rightarrow \infty$) then $\hat{\theta}_n^{(J_n)}$ converges to the true parameter value θ_0 where $J_n \rightarrow \infty$ with n. That this would hold is not surprising in light of the strong nature of the convergence of $p_X^{(J)}$ proved in Theorem 1: uniform in x *and* in θ.

This setup is different from the pseudo-maximum likelihood one [see White (1982) and Gouriéroux, Monfort and Trognon (1984)]. We are in an atypical situation in the sense that the pseudo-likelihood does approximate the true likelihood function, and wish to exploit this fact. We are not concerned with the potential misspecification of the true likelihood function, but then do not require that the densities belong to specific classes such as the linear exponential family. Simulation-based or PDE-based methods also produce approximations to the true likelihood. Of course, what makes the convergence proof possible here under very general conditions [including non-ergodicity of the process] is the explicit nature of the approximation.

When defining the log-likelihood function in (3.1), we ignore the unconditional density term $\mathrm{Ln}(\pi(X_0;\theta))$ because it is dominated by the sum of the conditional density terms $\mathrm{Ln}\left\{p_X\left(\Delta, X_{i\Delta} \mid X_{(i-1)\Delta};\theta\right)\right\}$ as $n \rightarrow \infty$. The sample contains only one observation on the unconditional density π and n on the transition function, so that the information on π contained in the sample does not increase with n. All the distributional properties below will obviously be asymptotic, so the definition (3.1) is appropriate for the log-likelihood function [see Billingsley (1961)].

To analyze the properties of the estimators $\hat{\theta}_n$ and $\hat{\theta}_n^{(J)}$, we introduce the following notation. Define the $K \times K$ identity matrix Id, $L_i(\theta) \equiv \mathrm{Ln}\left(p_X\left(\Delta, X_{i\Delta} \mid X_{(i-1)\Delta};\theta\right)\right)$, the $K \times 1$ vector $\dot{L}_i(\theta) \equiv \partial L_i(\theta)/\partial\theta$ and the $K \times K$ matrix $\ddot{L}_i(\theta) \equiv \partial^2 L_i(\theta)/\partial\theta\partial\theta^T$ where ^{T} denotes transposition. From the direct representation of the p_X used in the proof of Lemma 2, and the differentiability of μ and σ in θ [see Assumption 1], $p_X(\Delta, x \mid x_0, \theta)$ admits two continuous derivatives with respect to θ in Θ. The same holds for its approximations of any order J. Let

(3.2) $\quad I_n(\theta) \equiv \sum_{i=1}^{n} \text{diag}\, E_\theta\left[\dot{L}_i(\theta)\dot{L}_i(\theta)^T \middle| X_{(i-1)\Delta}\right]$

and $i_n(\theta) \equiv E_\theta[I_n(\theta)] = \sum_{i=1}^{n} \text{diag}\, E_\theta\left[\dot{L}_i(\theta)\dot{L}_i(\theta)^T\right]$ denote the unconditional expectation of $I_n(\theta)$. Also define $H_n(\theta) \equiv -\sum_{i=1}^{n} \ddot{L}_i(\theta)$ and recall that $E_\theta[H_n(\theta)] = E_\theta[I_n(\theta)] = i_n(\theta)$. The order of differentiation with respect to θ and integration with respect to the conditional density p_X [i.e., computation of conditional expectations] can be interchanged because the expected value of the derivative is continuous in θ.

If the process is not stationary, $E_\theta\left[\dot{L}_i(\theta)\dot{L}_i(\theta)^T\right] = E_\theta\left[E_\theta\left[\dot{L}_i(\theta)\dot{L}_i(\theta)^T \middle| X_{(i-1)\Delta}\right]\right]$ [by the law of iterated expectations] is not independent of the time index i, but rather depends on the joint distribution of $\left(X_{i\Delta}, X_{(i-1)\Delta}\right)$ which is nonstationary. We make the following assumptions:

Assumption 4 (Identification): The true parameter vector θ_0 belongs to Θ, and

(3.3) $\quad i_n^{-1}(\theta) \xrightarrow{a.s.} 0$ uniformly in $\theta \in \Theta$.

If X is stationary, and for all k=1,...,K, $\theta \in \Theta$, and $x_0 \in D_X$,

(3.4) $\quad 0 < \int_{\underline{x}}^{\bar{x}} \left\{\partial \text{Ln}\left(p_X(\Delta, x \mid x_0; \theta)\right)/\partial\theta_k\right\}^2 p_X(\Delta, x \mid x_0; \theta)\,dx < +\infty$

is sufficient to ensure that $I_n(\theta)$ is well-defined and that $i_n^{-1}(\theta) \xrightarrow{a.s.} 0$. For the upper bound, it is obviously sufficient that $\left|\partial \text{Ln}\left(p_X(\Delta, x \mid x_0; \theta)\right)/\partial\theta_k\right|$ remain bounded as x varies in D_X, but not necessary. For the lower bound, the assumption says that the transition function p_X cannot be flat as a function of one of the parameters θ_k, otherwise $\partial p_X(\Delta, x \mid x_0; \theta)/\partial\theta_k \equiv 0$ and the model cannot be identified.

Assumption 5 (Convergence in the Non-Ergodic Case): There exists a (possibly random) matrix $G(\theta)$, almost surely finite and positive definite, such that

(3.5) $\quad G_n(\theta) \equiv i_n^{-1/2}(\theta)\, H_n(\theta)\, i_n^{-1/2}(\theta) \xrightarrow{p} G(\theta)$

uniformly over compact subsets of Θ.

If X is a stationary diffusion, then $G(\theta)$=Id is constant, and (3.4) in fact follows from the Law of Large Numbers [see Hall and Heyde (1980, Theorem 2.18)] and the fact that $E_\theta[H_n(\theta)] = i_n(\theta)$.

Our strategy to study the asymptotic properties of $\hat{\theta}_n^{(J_n)}$ is to first determine those of $\hat{\theta}_n$ [Lemma 5] and then to show that $\hat{\theta}_n^{(J_n)}$ and $\hat{\theta}_n$ share the same asymptotic properties [Theorem 2]. It is easy to see that the score vector $S_n(\theta) \equiv \sum_{i=1}^{n} \dot{L}_i(\theta)$ is a martingale, and this forms the basis of the asymptotic properties of $\hat{\theta}_n$. $\hat{\theta}_n$ is consistent, asymptotically Normal and efficient:

Lemma 5: Under Assumptions 1-5, and for $\Delta \in (0, \bar{\Delta})$, the maximum-likelihood estimator $\hat{\theta}_n$ exists and satisfies:

i. $\hat{\theta}_n \xrightarrow{p} \theta_0$ and $i_n^{1/2}(\theta_0)(\hat{\theta}_n - \theta_0) \xrightarrow{d} G^{-1/2}(\theta_0) \times N(0, Id)$ under P_{θ_0}.

ii. Suppose that $\tilde{\theta}_n$ is an alternative estimator such that for any $h \in R^K$ and $\theta \in \Theta$,

$$i_n^{1/2}(\theta)\left(\tilde{\theta}_n - \theta - i_n^{-1/2}(\theta)h\right) \xrightarrow{d} T(\theta) \text{ under } P_{\theta + I_n^{1/2}(\theta)h}$$

where $T(\theta)$ is a proper law, not necessarily Normal.

Then $\hat{\theta}_n$ has maximum concentration in that class, i.e., is closer to θ_0 than $\tilde{\theta}_n$ is in the sense that for any $\varepsilon > 0$

$$(3.6) \quad \lim_{n\to\infty} \text{Prob}_{\theta_0}\left(i_n^{1/2}(\theta_0)(\hat{\theta}_n - \theta_0) \in C_\varepsilon\right) \geq \lim_{n\to\infty} \text{Prob}_{\theta_0}\left(i_n^{1/2}(\theta_0)(\tilde{\theta}_n - \theta) \in C_\varepsilon\right)$$

where $C_\varepsilon \equiv [-\varepsilon, +\varepsilon]^K$.

(iii) If $\tilde{\theta}_n$ is also asymptotically Normal, $i_n^{1/2}(\theta_0)(\tilde{\theta}_n - \theta_0) \xrightarrow{d} G^{-1/2}(\theta_0) \times N(0, \tilde{V}_0)$ under P_{θ_0}, then $\tilde{V}_0 - Id$ is non-negative definite.

This lemma follows from specializing to the observed process here general results pertaining to maximum-likelihood estimation for stochastic processes [see Hall and Heyde (1980, Section 6) and Basawa and Scott (1983)]. If the process is stationary, then Lemma 5 can be greatly simplified: we can then set $G(\theta) = Id$ and $i_n^{1/2}(\theta_0) = n^{1/2} i^{1/2}(\theta_0)$ where the unconditional expectation

$$(3.7) \quad i(\theta) \equiv \text{diag } E_\theta\left[\dot{L}_i(\theta)\dot{L}_i(\theta)^T\right],$$

is now independent of the time index i and is Fisher's Information Matrix [see Billingsley (1961)].

However, when the process is not ergodic, $G(\theta)$ may well be a stochastic matrix: consider for example the case of a mean-avoiding Ornstein-Uhlenbeck process,

$$dX_t = (\alpha_0 + \alpha_1 X_t)dt + dZ_t,$$

where $\alpha_1 > 0$ [Sørensen (1991, Example 5.2) studied the likelihood of continuously sampled observations from this process and calculated the stochastic limit of the norming factor; we can show that the same result applies to discrete sampling at interval Δ]. Furthermore, when the parameter vector is multidimensional, the K diagonal terms of $i_n^{1/2}(\theta_0)$ do not necessarily go to infinity at the same rate [unlike the common rate $n^{1/2}$ in the ergodic case].

If we had normed the difference $(\hat{\theta}_n - \theta_0)$ by the stochastic factor $I_n^{1/2}(\theta_0)$ rather than by the deterministic factor $i_n^{1/2}(\theta_0) \equiv E_{\theta_0}[I_n(\theta_0)]^{1/2}$ [see Hall and Heyde (1980, Chapter 6)] then the asymptotic distribution of the estimator would have been N(0,Id) rather than $G^{-1/2}(\theta_0) \times N(0, Id)$. In other words, the stochastic norming, while intrinsically more complicated, may be useful if the distribution of $G(\theta_0)$ is untractable. In that case, the distribution of $i_n^{1/2}(\theta_0)(\tilde{\theta}_n - \theta_0)$ need not be asymptotically Normal [and depends on θ_0] whereas that of $I_n^{1/2}(\theta_0)(\tilde{\theta}_n - \theta_0)$ would simply be N(0,Id). Again, none of these difficulties are present in the stationary case, where $G(\theta)$ = Id.

Naturally, Lemma 5 is not an end in itself since in our context $\hat{\theta}_n$ cannot be computed explicitly. The lemma becomes useful however when we can prove that the approximate maximum-likelihood estimator $\hat{\theta}_n^{(J)}$ is a good substitute for $\hat{\theta}_n$, in the sense that the asymptotic properties of $\hat{\theta}_n$ identified in Lemma 5 carry over to $\hat{\theta}_n^{(J)}$. For technical reasons, we need to limit the speed at t a minor additional condition on the

Assumption 6 (Strengthening of Assumption 2 in the limiting case where α=1 and the diffusion is degenerate at 0): Recall the constant ρ in Assumption 2(2), and the constants α and κ in Assumption 3(1.i). If α=1, then either $\rho \geq 1$ with no restriction on κ, or $\kappa \geq 2\rho/(1-\rho)$ if $0 \leq \rho < 1$. If $\alpha > 1$, no restriction is required.

Finally, we have:

Theorem 2: Under Assumptions 1-6, and for $\Delta \in (0, \bar{\Delta})$:

i. Fix the sample size n. Then as $J \to \infty$, $\hat{\theta}_n^{(J)} \xrightarrow{p} \hat{\theta}_n$ under P_{θ_0}.

ii. As $n \to \infty$, there exist a sequence $\bar{J}_n \to \infty$ such that for any $J_n \geq \bar{J}_n$:

(3.8) $\hat{\theta}_n^{(J_n)} \xrightarrow{p} \theta_0$ and $i_n^{1/2}(\theta_0)(\hat{\theta}_n^{(J_n)} - \theta_0) \xrightarrow{d} G^{-1/2}(\theta_0) \times N(0, Id)$ under P_{θ_0}.

iii. And recall from Lemma 5 that Id is the lowest asymptotic variance achievable by $i_n^{1/2}(\theta_0)$-consistent and asymptotically Normal estimators of θ_0.

4. Explicit Expressions for the Expansion

With these desirable asymptotic properties in hand, we can now turn to the computation of the terms in the expansion of p_X. Theorem 1 shows that

$$(4.1)\quad p_Z(\Delta,z \mid y_0;\theta) = \phi(z)\sum_{j=0}^{\infty}\eta_j(\Delta,y_0;\theta)H_j(z).$$

Recall that $p_Z^{(J)}(\Delta,z \mid y_0;\theta)$ denotes the partial sum in (4.1) up to j=J. To fully characterize the expansion for a given J, we now give the explicit expression of the coefficients η_j in the form of a Taylor series in powers of Δ. From (2.12), we have

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

To calculate explicitly the coefficients of the expansion we therefore need to calculate these conditional moments. For that purpose, we rely on Lemma 6:

Lemma 6: Under Assumptions 1-3, let f be a function such as f and all its derivatives have at most exponential growth. Then for $\Delta \in (0,\bar{\Delta})$, $y_0 \in R$ and $\theta \in \Theta$, there exists δ in $[0,\Delta]$ such that

$$(4.3)\quad E\left[f(Y_{t+\Delta}) \middle| Y_t = y_0\right] = \sum_{j=1}^{J} A^j(\theta)\bullet f(y_0)\frac{\Delta^j}{j!} + E\left[A^{J+1}(\theta)\bullet f(Y_{t+\delta}) \middle| Y_t = y_0\right]\frac{\Delta^{J+1}}{(J+1)!}$$

where $A(\theta)$ is the infinitesimal generator of the diffusion Y, defined as the operator:

$$(4.4)\quad A(\theta): f \mapsto \mu_Y(\cdot;\theta)\frac{\partial f}{\partial y}(\cdot) + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(\cdot)$$

and $A^j(\theta)\bullet f(y_0)$ means $A(\theta)$ applied j times to the function $y \mapsto f(y)$, and evaluated at $y=y_0$.

Further, there exists a constant K_J dependent on J, but independent of f and δ, such that

$$(4.5)\quad \left| E\left[A^{J+1}(\theta)\bullet f(Y_{t+\delta}) \middle| Y_t = y_0\right]\right| \leq K_J$$

Note also that in what follows, we do not require that the remainder of the Taylor series converge to zero as $J \to \infty$ for fixed $\Delta \in (0, \bar{\Delta})$, since the operator $A(\theta)$ is unbounded in general. Rather, the convergence of the Taylor series (4.3) should be interpreted in the sense that for any given J

$$(4.6) \quad \lim_{\Delta \to 0} \Delta^{-(J+1)} \left\{ E\left[f(Y_{t+\Delta}) \middle| Y_t = y_0\right] - \sum_{j=1}^{J} A^j(\theta) \bullet f(y_0) \frac{\Delta^j}{j!} \right\} = \frac{A^{J+1}(\theta) \bullet f(y_0)}{(J+1)!}.$$

i.e., as the proof of Lemma 6 makes clear, (4.3) is a Taylor series expansion of the conditional expectation operator in Δ.

Of course, the convergence of the series $p_Z^{(J)}$ to p_Z in (4.1), which follows from Theorem 1, is independent of (4.2). In other words, we first choose J sufficiently large for the remainder in (4.1) to be small, and for that fixed J we then apply (4.2) to calculate the coefficients η_j, $j=0,\ldots,J$.

Let $\tilde{p}_Z^{(J)}$ denote the approximation to p_Z obtained by retaining in $p_Z^{(J)}$ all the terms in η_j, $j=0,\ldots,J$ of order smaller or equal to $\Delta^{J/2}$. Exact calculations show that η_j is or order $\Delta^{J/2}$, so $\phi(z) \sum_{j=J+1}^{\infty} \eta_j(\Delta, y_0; \theta) H_j(z)$ is of order larger than $\Delta^{J/2}$. Hence the expression $\tilde{p}_Z^{(J)}$ does indeed retain all the terms up to $\Delta^{J/2}$ in p_Z, not only in $p_Z^{(J)}$. In the next section, we give the expression of the first few terms of the expansion $\tilde{p}_Z^{(J)}$.

5. Practical Considerations

5.1 The First Terms in the Expansion

The message from Section 3 is that maximizing the closed-form expansion of the likelihood function of order J results in an estimator which gets closer to the exact (but impossible to compute) maximum-likelihood estimator as J increases. If we collect the terms in powers of Δ as indicated in Section 4, and retain only the terms of order smaller or equal to $\Delta^{J/2}$, we obtain an approximation $\tilde{p}_Z^{(J)}$. We call $\tilde{p}_Z^{(J)}$ the approximation of order J to p_Z.

The first six terms of the sequence $\tilde{p}_Z^{(J)}(\Delta, z \mid y_0; \theta)$ are given by:

$$(5.1) \quad \tilde{p}_Z^{(1)} = \phi$$

$$(5.2) \quad \tilde{p}_Z^{(2)} = \tilde{p}_Z^{(1)} + \phi\left[H_2 \mu_Y^{[1]}/2\right]\Delta$$

$$(5.3) \quad \tilde{p}_Z^{(3)} = \tilde{p}_Z^{(2)} - \phi\left[H_1\left\{\mu_Y \mu_Y^{[1]}/2 + \mu_Y^{[2]}/4\right\} + H_3 \mu_Y^{[2]}/6\right]\Delta^{3/2}$$

$$(5.4)\quad \tilde{p}_Z^{(4)} = \tilde{p}_Z^{(3)} + \phi\left[H_2\left\{\mu_Y^{[1]2}/3 + \mu_Y\mu_Y^{[2]}/3 + \mu_Y^{[3]}/6\right\} + H_4\left\{\mu_Y^{[1]2}/8 + \mu_Y^{[3]}/24\right\}\right]\Delta^2$$

$$(5.5)\quad \begin{aligned}\tilde{p}_Z^{(5)} = \tilde{p}_Z^{(4)} - \phi\Big[&H_1\left\{\mu_Y\mu_Y^{[1]2}/6 + \mu_Y^2\mu_Y^{[2]}/6 + \mu_Y^{[1]}\mu_Y^{[2]}/4 + \mu_Y\mu_Y^{[3]}/6 + \mu_Y^{[4]}/24\right\}\\ &+ H_3\left\{\mu_Y\mu_Y^{[1]2}/4 + \mu_Y^{[1]}\mu_Y^{[2]}/2 + \mu_Y\mu_Y^{[3]}/8 + \mu_Y^{[4]}/16\right\}\\ &+ H_5\left\{\mu_Y^{[1]}\mu_Y^{[2]}/12 + \mu_Y^{[4]}/120\right\}\Big]\Delta^{5/2}\end{aligned}$$

$$(5.6)\quad \begin{aligned}\tilde{p}_Z^{(6)} = \tilde{p}_Z^{(5)} + \phi\Big[&H_2\Big\{\mu_Y^2\mu_Y^{[1]2}/8 + \mu_Y^{[1]3}/6 + 2\mu_Y\mu_Y^{[1]}\mu_Y^{[2]}/3 + 7\mu_Y^{[2]2}/32\\ &\quad +\mu_Y^2\mu_Y^{[3]}/8 + \mu_Y^{[1]}\mu_Y^{[3]}/3 + \mu_Y\mu_Y^{[4]}/8 + \mu_Y^{[5]}/32\Big\}\\ &+ H_4\Big\{\mu_Y^{[1]3}/6 + \mu_Y\mu_Y^{[1]}\mu_Y^{[2]}/4 + 17\mu_Y^{[2]2}/120\\ &\quad +13\mu_Y^{[1]}\mu_Y^{[3]}/60 + \mu_Y\mu_Y^{[4]}/30 + \mu_Y^{[5]}/60\Big\}\\ &+ H_6\left\{\mu_Y^{[1]3}/48 + \mu_Y^{[2]2}/72 + \mu_Y^{[1]}\mu_Y^{[3]}/48 + \mu_Y^{[5]}/720\right\}\Big]\Delta^3\end{aligned}$$

where we have used the more compact notation ϕ for $\phi(z)$, the N(0,1) density, H_j for $H_j(z)$ and $\mu_Y^{[k]m}$ for $\left(\partial^k\mu_Y(y_0;\theta)/\partial y_0^k\right)^m$.

The corresponding expressions for $\tilde{p}_X^{(J)}(\Delta, x \mid x_0;\theta)$ are given by:

$$\tilde{p}_X^{(J)}(\Delta, x \mid x_0;\theta) = \sigma(x;\theta)^{-1}\Delta^{-1/2}\,\tilde{p}_Z^{(J)}\left(\Delta, \Delta^{-1/2}\left(\gamma(x;\theta) - \gamma(x_0;\theta) - \mu_Y\left(\gamma(x_0;\theta);\theta\right)\Delta\right) \mid \gamma(x_0;\theta); \theta\right)$$

and then replacing γ and μ_Y by their expressions (2.1) and (2.3) respectively.

For instance, the first term is the Normal density deformed by the function γ:

$$(5.7)\quad \tilde{p}_X^{(1)}(\Delta, x \mid x_0;\theta) = \frac{1}{\sqrt{2\pi\Delta}\,\sigma(x;\theta)} e^{-\left\{\left(\gamma(x;\theta)-\gamma(x_0;\theta)-\mu_Y\left(\gamma(x_0;\theta);\theta\right)\Delta\right)\right\}^2/2\Delta},$$

the second term is

$$(5.8)\quad \begin{aligned}\tilde{p}_X^{(2)}(\Delta, x \mid x_0;\theta) \;=\; &\left\{\sqrt{2\pi\Delta}\,\sigma(x;\theta)\right\}^{-1} e^{-\left\{\left(\gamma(x;\theta)-\gamma(x_0;\theta)-\mu_Y\left(\gamma(x_0;\theta);\theta\right)\Delta\right)\right\}^2/2\Delta} \times\\ &\left\{1 + H_2\left(\frac{\gamma(x;\theta) - \gamma(x_0;\theta) - \mu_Y\left(\gamma(x_0;\theta);\theta\right)\Delta}{\Delta^{1/2}}\right)\frac{\mu_Y^{[1]}\left(\left(\gamma(x_0;\theta);\theta\right)\right)}{2}\Delta\right\}\end{aligned}$$

and so on.

The first six polynomials H_j are given by:

$$(5.9)\quad \begin{aligned}&H_1(z) \equiv -z,\; H_2(z) \equiv z^2 - 1,\; H_3(z) \equiv -z^3 + 3z,\; H_4(z) \equiv z^4 - 6z^2 + 3\\ &H_5(z) \equiv -z^5 + 10z^3 - 15z,\; H_6(z) \equiv z^6 - 15z^4 + 45z^2 - 15\end{aligned}$$

The corresponding terms for the expansion of the log-likelihood for Z are given by:

(5.10) $\tilde{\ell}^{(1)}_{Z,n}(\theta) = n^{-1}\sum_{i=1}^{n} \mathrm{Ln}[\phi]$

(5.11) $\tilde{\ell}^{(2)}_{Z,n}(\theta) = \tilde{\ell}^{(1)}_{Z,n}(\theta) + n^{-1}\sum_{i=1}^{n}\left[H_2\,\mu_Y^{[1]}/2\right]\Delta$

(5.12) $\tilde{\ell}^{(3)}_{Z,n}(\theta) = \tilde{\ell}^{(2)}_{Z,n}(\theta) - n^{-1}\sum_{i=1}^{n}\left[H_1\left\{\mu_Y\mu_Y^{[1]}/2 + \mu_Y^{[2]}/4\right\} + H_3\,\mu_Y^{[2]}/6\right]\Delta^{3/2}$

etc., and the expansion $\tilde{\ell}^{(J)}_{n}(\theta)$ is obtained from $\tilde{\ell}^{(J)}_{Z,n}(\theta)$ just like $\tilde{p}_X^{(J)}$ from $\tilde{p}_Z^{(J)}$ above. In practice, including just the first three terms in the expansion, i.e., maximizing $\tilde{\ell}^{(3)}_{n}(\theta)$ is more than adequate and there is no need to consider higher order expansions.

Note that the third and higher correction terms to the N(0,1) distribution function in the expansion of p_Z do not become smaller as J increases: the third order term remains of order $\Delta^{3/2}$ no matter how large J gets. So constructing better and better approximations to the conditional mean and standard deviation of $Y_{t+\Delta} \mid Y_t$ for the purpose of centering Y more accurately [i.e., constructing a "better" Z] would not improve the performance of the approximation. This is why we do not construct pseudo-maximum-likelihood estimators of θ by maximizing a Normal density function with converging expansions of the mean and standard deviation as in Huggins (1997) and then let J go to infinity --the best one could do would be an approximation error of order $\Delta^{3/2}$, the first correction term present exclusively in the non-Gaussian case. By contrast, the error here is of order $\Delta^{J/2}$ and can be made arbitrarily small by choosing J large enough.

5.2 Examples and Accuracy of the Expansion

We study the size of the approximation made when replacing p_X by $p_X^{(J)}$, and how fast the error decreases as more terms are added, in three classical examples where p_X is known in closed-form. These examples show that the term of order 2, provides an approximation to p_X which is better by a factor of roughly 10 than the term of order 1, and that each additional order produces additional improvements by a factor of roughly 10.

Example 1 (Vasicek's Model): Consider the Ornstein-Uhlenbeck specification proposed by Vasicek (1977) for the short term interest rate, $dX_t = \beta(\overline{\alpha} - X_t)dt + \sigma dW_t$, distributed on $(-\infty,+\infty)$ and for which the transition density is Gaussian:

$$p_X(\Delta, x \mid x_0;\theta) = \left(\pi\sigma^2\left(1-e^{-2\beta\Delta}\right)/\beta\right)^{-1/2} \exp\left\{-\left(x-\overline{\alpha}-\left(x_0-\overline{\alpha}\right)e^{-\beta\Delta}\right)^2\beta\Big/\left(\sigma^2\left(1-e^{-2\beta\Delta}\right)\right)\right\}$$

with $\theta \equiv (\overline{\alpha},\beta,\sigma)$. Here $Y_t = \gamma(X_t;\theta) = \sigma^{-1}X_t$ and $\mu_Y(y;\theta) = \beta\overline{\alpha}\sigma^{-1} - \beta y$.

We first plot in Figure 1 the density p_X as a function of the interest rate value x for a semi-annual sampling frequency (Δ=1/2), evaluated at x_0=0.10 and for the parameter values $\overline{\alpha}$=0.08, β=0.01 and σ=0.02, which are realistic values from U.S. short term interest rates. Below the density p_X, we plot the approximation error $p_X - p_X^{(J)}$ for J=1,...,6. The striking feature of this figure is the speed of convergence to zero of the approximation error as J gets larger. In effect, we can approximate p_X (which is of order 10^{+1}) within 10^{-6} with J=4 and within 10^{-8} with J=6 (even though we are only sampling the process every six months).

Figure 2 reports the same results for a monthly sampling frequency (Δ=1/12). As one would expect, the approximation error gets smaller even faster for this lower value of Δ. With J=4, the approximation error is of order 10^{-7}, and 10^{-11} with J=6, when p_X is of order 10^{+2}. In other words, small values of J already produce extremely precise approximations to the true density p_X, and the approximation is even more precise if Δ is smaller.

Example 2 (The CIR Model): Consider Feller's square-root specification proposed as a model for the short term interest rate, $dX_t = \beta(\overline{\alpha} - X_t)dt + \sigma\sqrt{X_t}dW_t$, by Cox, Ingersoll and Ross (1985). X is distributed on $D_X = (0,+\infty)$ provided that $q \equiv 2\beta\overline{\alpha}/\sigma^2 - 1 \geq 0$. Its transition density is non-central chi-squared:

$$p_X(\Delta, x \mid x_0;\theta) = ce^{-u-v}(v/u)^{q/2} I_q\left(2(uv)^{1/2}\right)$$

with $\theta \equiv (\overline{\alpha},\beta,\sigma)$, all positive, $c \equiv 2\beta/\left(\sigma^2\{1-e^{-\beta\Delta}\}\right)$, $u \equiv cx_0e^{-\beta\Delta}$, $v \equiv cx$ and I_q is the modified Bessel function of the first kind of order q [see Cox, Ingersoll and Ross (1985, (18))]. Here $Y_t = \gamma(X_t;\theta) = 2\sqrt{X_t}/\sigma$ and $\mu_Y(y;\theta) = (q+1/2)/y \; - \beta y/2$. This process satisfies Assumptions 1-3. With regard to Assumption 6, we have α=1, κ=q+1/2 and ρ=1/2. Hence we need q+1/2 $\geq$ 2ρ/(1-ρ)=2, or q$\geq$3/2. The CIR process is indeed the limiting case where α=1 and 0$\leq\rho$<1.

When comparing the approximation to the true density, we find similar results to those of Example 1, with again an extremely fast convergence even for a semi-annual sampling frequency (Figure 3), and even more so if we sample more often (see Figure 4 for monthly sampling). The parameter values are $\overline{\alpha}$=0.08, β=0.05 and σ=0.05 [so q=2.2] and the density is evaluated at x_0=0.10.

Example 3 (The CEV Model): $dX_t = \beta(\overline{\alpha} - X_t)dt + \sigma X_t^{\rho} dW_t$ with $\theta \equiv (\overline{\alpha},\beta,\sigma,\rho)$ is distributed on (0,+∞) when $\overline{\alpha}$>0, β>0 and ρ>1/2 [if ρ=1/2, see Example 2 for an additional constraint]. This model does not admit a closed-form density [see Cox (1975-

1996), Chan et al. (1992)]. For $1/2<\rho<1$, the transformation from X to Y is given by $Y_t = \gamma(X_t;\theta) = X_t^{1-\rho}/\{\sigma(1-\rho)\}$ and

$$\mu_Y(y;\theta) = \varphi_0 y^{-\rho/(1-\rho)} - \psi y^{-1} - \varphi_1 y$$

where $\varphi_0 \equiv \beta\overline{\alpha}\sigma^{-1/(1-\rho)}(1-\rho)^{-\rho/(1-\rho)}$, $\psi \equiv -\rho/\{2(1-\rho)\}$ and $\varphi_1 \equiv \beta(1-\rho)$. Assumptions 1-3 again apply. Assumption 6 is automatically satisfied since $\alpha>1$ when $1/2<\rho<1$.

Example 4 (Nonlinear Mean Reversion): The following model, estimated by Aït-Sahalia (1996b), Conley et al. (1997) and Tauchen (1997), was designed to produce very little mean-reversion while interest rate values remain in the middle part of their domain, and strong nonlinear mean-reversion at either end of the domain:

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

with $\theta \equiv (\alpha_{-1},\alpha_0,\alpha_1,\alpha_2,\sigma,\rho)$. We can again verify that Assumptions 1-3 are satisfied by this model. Indeed, $D_X = (0,+\infty)$, $Y_t = \gamma(X_t;\theta) = X_t^{1-\rho}/\{\sigma(1-\rho)\}$ (for $\rho\neq1$, $Y_t=\sigma \mathrm{Ln}(X_t)$ if $\rho=1$) and for $\rho\neq1$

$$\mu_Y(y;\theta) = \varphi_{-1} y^{-(1+\rho)/(1-\rho)} + \varphi_0 y^{-\rho/(1-\rho)} + \psi y^{-1} + \varphi_1 y + \varphi_2 y^{(2-\rho)/(1-\rho)}$$

where $\varphi_{-1} \equiv \alpha_{-1}\sigma^{-2/(1-\rho)}(1-\rho)^{-(1+\rho)/(1-\rho)}$, $\varphi_0 \equiv \alpha_0\sigma^{-1/(1-\rho)}(1-\rho)^{-\rho/(1-\rho)}$, $\psi \equiv -\rho/\{2(1-\rho)\}$, $\varphi_1 \equiv \alpha_1(1-\rho)$ and $\varphi_2 \equiv \alpha_2\sigma^{1/(1-\rho)}(1-\rho)^{(2-\rho)/(1-\rho)}$. Assumption 6 is satisfied because $\alpha>1$ when $0\leq\rho<1$ due to the term with coefficient φ_{-1}, and $\alpha=1$ (leading term has coefficient ψ near 0^+) can only occur if $\rho>1$.

Example 5 (The Black-Scholes-Merton Model): Consider a geometric Brownian Motion, $dX_t = \mu X_t dt + \sigma X_t dW_t$, which is distributed on $D_X = (0,+\infty)$. Its transition density is log-normal:

$$p_X(\Delta, x \mid x_0;\theta) = \left(2\pi\Delta\sigma^2 x^2\right)^{-1/2} \exp\left\{-\left(\mathrm{Ln}(x/x_0) - \left(\mu - \sigma^2/2\right)\Delta\right)^2 / \left(2\Delta\sigma^2\right)\right\}$$

with $\theta \equiv (\mu,\sigma)$. In this case, we have $Y_t = \gamma(X_t;\theta) = \sigma\, \mathrm{Ln}(X_t)$ so $D_Y = (-\infty,+\infty)$ and $\mu_Y(y;\theta) = \mu/\sigma - \sigma/2$. This process satisfies Assumptions 1-3 and 6 (irrelevant), and is not ergodic. Nevertheless, because Y is an arithmetic Brownian motion (μ_Y is constant), it is easy to see that our approximation gives an exact result starting at the first order, i.e.: $p_X^{(J)} = p_X$ for all $J\geq1$.

5.3 Estimation of the Asymptotic Variance and Test Statistics

Theorem 2 essentially implies that we can replace $\hat{\theta}_n$ by $\hat{\theta}_n^{(J_n)}$ in any of the usual calculations involving maximum-likelihood estimates, without any adverse consequences asymptotically.

For instance, from the convergence in Theorem 2 and the continuity of the gradient of the log-likelihood, it follows in the ergodic case that

$$(5.14)\quad \hat{i}_0^{(J_n)} \equiv n^{-1}\sum_{i=1}^{n} \text{diag}\left\{\dot{L}_i^{(J_n)}(\theta)\dot{L}_i^{(J_n)}(\theta)^T\right\}\Big|_{\theta=\hat{\theta}_n^{(J_n)}} \xrightarrow{\ p\ } i_0(\theta_0)$$

so $\left[\hat{i}_0^{(J_n)}\right]^{-1}$ is a suitable consistent estimator of the asymptotic variance of the maximum-likelihood estimator.

Test statistics can be derived. Suppose that the model is (1.1) and that we wish to test H_0: $\theta=\theta_0$ against the two-sided alternative H_a: $\theta\neq\theta_0$. As a consequence of Theorem 2, the likelihood ratio test statistic evaluated at $\hat{\theta}_n^{(J_n)}$ behaves identically to that $\hat{\theta}_n$

$$(5.15)\quad 2\left\{\ell_n^{(J_n)}\left(\hat{\theta}_n^{(J_n)}\right) - \ell_n^{(J_n)}(\theta_0)\right\} \xrightarrow{\ d\ } \chi_K^2 \text{ under } H_0\text{, and:}$$

$$(5.16)\quad 2\left\{\ell_n^{(J_n)}\left(\hat{\theta}_n^{(J_n)}\right) - \ell_n^{(J_n)}(\theta_0)\right\} \xrightarrow{\ d\ } \left(Z+G^{1/2}(\theta_0)h\right)^2 \text{ under the sequence of alternatives}$$

H_n: $\theta = \theta_0 + i_n^{-1/2}(\theta_0)h$.

It is interesting to note that under H_0, the distribution of the statistic is chi-square whether $G(\theta)$ is random or Id, i.e., whether the diffusion is ergodic or not. However under H_n, the statistic is non-central chi-square if $G(\theta)$ = Id, and a mixture of non-central chi-square distributions with the random non-centrality parameter acting as a mixer.

Distributional results can be also be obtained for tests of a nested model which only allows for $\overline{K}$ free parameters from the K parameters in θ, and we can also consider Rao's efficient score statistic, which depends only on the restricted estimator $\overline{\theta}_n^{(J_n)}$, and Wald's test statistic, which depends only on the unrestricted estimator $\hat{\theta}_n^{(J_n)}$. Basawa and Scott (1983, Chapter 3) derive the properties of these statistics when true maximum-likelihood estimators are used. In effect, under Theorem 2, we can replace $\hat{\theta}_n$ with $\hat{\theta}_n^{(J_n)}$ and the same distributional and power properties apply.

5.4 How Many Terms to Include

From Theorems 1 and 2, the answer is simple: as many as is practically feasible! However, as the examples above have shown, it is not necessary to go much beyond J=3 in the relevant financial examples to estimate the true density with a high degree of precision. More generally, to select an appropriate J at which to stop adding terms to the expansion, we propose the following approach: take J large enough so that the *approximation error* made in replacing p_X by $\tilde{p}_X^{(J)}$ is smaller than the *sampling error* due to the random character of the data, by a predetermined factor.

That is, in

$$\left\|\hat{\theta}_n^{(J)} - \theta_0\right\| \leq \left\|\hat{\theta}_n^{(J)} - \hat{\theta}_n\right\| + \left\|\hat{\theta}_n - \theta_0\right\|, \tag{5.16}$$

we can estimate the asymptotic standard variance of $\hat{\theta}_n$ about θ_0 by (5.14). By Chebyshev's Inequality, we can then bound the second term on the right-hand-side of (5.16). We can then stop considering higher order approximations at an order J such that the distance between the two successive estimates $\hat{\theta}_n^{(J)}$ and $\hat{\theta}_n^{(J-1)}$ is an order of magnitude smaller than the distance between $\hat{\theta}_n$ and θ_0.

6. Conclusions

This paper has constructed a series of explicit functions converging to the conditional density of a diffusion process, under very mild regularity conditions on the process. This method makes maximum-likelihood a practical option for the estimation of parameters in discretely-sampled diffusion models. Further, the formulae for the expansion of p_X apply to any specification of (μ,σ^2), including nonparametric ones. An extension to multi-dimensional diffusions will be considered in future work. Applications to derivative pricing, consisting in obtaining pricing formulas for any underlying price process, have been outlined and will also be developed in future work.

APPENDIX: PROOFS

Proof of Lemma 1: We treat fully the case where $\underline{y}=0^+$ and $\bar{y}=+\infty$, the other boundary configurations being dealt with similarly. Let $s_Y(v;\theta) \equiv \exp\left\{-\int^v 2\mu_Y(u;\theta)\,du\right\}$ be the scale density of Y and $S_Y(v;\theta) \equiv \int^y s_Y(v;\theta)dv$. In each case, the lower limit of integration is a fixed constant, the choice of which is irrelevant in what follows. Also define its speed density $m_Y(v;\theta) \equiv 1/s_Y(v;\theta)$ and $M_Y(v;\theta) \equiv \int^y m_Y(v;\theta)dv$.

(1) To study the boundary $+\infty$, define:

$$\Sigma_\infty \equiv \int_y^{+\infty}\left\{\int_y^v m_Y(u;\theta)du\right\}s_Y(v;\theta)dv = \int_y^{+\infty}\left\{\int_u^{+\infty} s_Y(v;\theta)dv\right\}m_Y(u;\theta)du$$

$$N_\infty \equiv \int_y^{+\infty}\left\{\int_y^v s_Y(u;\theta)du\right\}m_Y(v;\theta)dv = \int_y^{+\infty}\left\{\int_u^{+\infty} m_Y(v;\theta)dv\right\}s_Y(u;\theta)du$$

where the choice of the lower bound of integration y is again irrelevant. The boundary $+\infty$ is a natural boundary when $\Sigma_\infty = N_\infty = \infty$, and an entrance boundary when $\Sigma_\infty = \infty$ and $N_\infty < \infty$.

(i) Let E>0 be such that $-Ky \le \mu_Y(y;\theta) \le Ky$ for all y≥E. We have

$$N_\infty = \int_y^{+\infty}\left\{\int_u^{+\infty} m_Y(v;\theta)dv\right\}m_Y^{-1}(u;\theta)du = \int_y^{+\infty}\int_u^{+\infty} e^{\int_u^v 2\mu_Y(w;\theta)dw}dvdu$$
$$\ge \int_y^{+\infty}\int_u^{+\infty} e^{-\int_u^v 2Kw\,dw}dvdu = \int_y^{+\infty}\left\{\int_u^{+\infty} e^{-Kv^2}dv\right\}e^{Ku^2}du$$

Now by integration by parts

$$\int_u^{+\infty} e^{-Kv^2}dv = \int_u^{+\infty} v^{-1}ve^{-Kv^2}dv = (2Ku)^{-1}e^{-Ku^2} - (2K)^{-1}\int_u^{+\infty} v^{-2}e^{-Kv^2}dv$$

and, since $\int_u^{+\infty} v^{-2}e^{-Kv^2}dv < u^{-2}\int_u^{+\infty} e^{-Kv^2}dv$, it follows that

$$\left(1+(2K)^{-1}u^{-2}\right)\int_u^{+\infty} e^{-Kv^2}dv > (2Ku)^{-1}e^{-Ku^2}.$$

or $\int_u^{+\infty} e^{-Kv^2}dv > \left(2Ku+u^{-1}\right)^{-1}e^{-Ku^2}$. Therefore

$$N_\infty \ge \int_y^{+\infty}\left\{\int_u^{+\infty} e^{-Kv^2}dv\right\}e^{Ku^2}du \ge \int_y^{+\infty}\left(2Ku+u^{-1}\right)^{-1}e^{-Ku^2}e^{Ku^2}du = +\infty$$

that is, $N_\infty = \infty$ and $+\infty$ is a natural boundary.

(ii) If instead we have $\mu_Y(y;\theta) \le -Ky^\beta$, β>1, for all y≥E, then

$$N_\infty = \int_y^{+\infty}\int_u^{+\infty} e^{\int_u^v 2\mu_Y(w;\theta)dw}dvdu \le \int_y^{+\infty}\int_u^{+\infty} e^{-\int_u^v 2Kw^\beta dw}dvdu = \int_y^{+\infty}\left\{\int_u^{+\infty} e^{-\zeta v^{\beta+1}}dv\right\}e^{\zeta u^{\beta+1}}du$$

where $\zeta \equiv 2(\beta+1)^{-1}K$. By integration by parts

$$\int_u^{+\infty} e^{-\zeta v^{\beta+1}}dv = \int_u^{+\infty} v^{-\beta}v^{\beta}e^{-\zeta v^{\beta+1}}dv = (\zeta(\beta+1))^{-1}u^{-\beta}e^{-\zeta u^{\beta+1}} - \zeta^{-1}(\beta+1)^{-2}\int_u^{+\infty} v^{-\beta-1}e^{-\zeta v^{\beta+1}}dv$$

hence $\int_u^{+\infty} e^{-\zeta v^{\beta+1}}dv < (2K)^{-1}u^{-\beta}e^{-\zeta u^{\beta+1}}$, and therefore

$$N_\infty \le \int_y^{+\infty}\left\{\int_u^{+\infty} e^{-\zeta v^{\beta+1}}dv\right\}e^{\zeta u^{\beta+1}}du < (2K)^{-1}\int_y^{+\infty} u^{-\beta}e^{-\zeta u^{\beta+1}}e^{\zeta u^{\beta+1}}du < +\infty$$

and $+\infty$ is an entrance boundary.

(iii) By the same type of calculation as in (i), we have provided that $\mu_Y(y;\theta) \le Ky$ for all $y \ge E$ (irrespectively of how negative μ_Y gets):

$$\Sigma_\infty = \int_y^{+\infty}\left\{\int_u^{+\infty} s_Y(v;\theta)dv\right\}s_Y^{-1}(u;\theta)du = \int_y^{+\infty}\int_u^{+\infty} e^{-\int_u^v 2\mu_Y(w;\theta)dw}dvdu$$

$$\ge \int_y^{+\infty}\int_u^{+\infty} e^{-\int_u^v 2Kw\,dw}dvdu = \int_y^{+\infty}\left\{\int_u^{+\infty} e^{-Kv^2}dv\right\}e^{Ku^2}du = +\infty$$

that is, $\Sigma_\infty = \infty$ and thus $+\infty$ is unattainable.

(2) Near 0^+, define:

$$\Sigma_0 \equiv \int_0^y\left\{\int_v^y m_Y(u;\theta)du\right\}s_Y(v;\theta)dv = \int_0^y\left\{\int_0^u s_Y(v;\theta)dv\right\}m_Y(u;\theta)du$$

$$N_0 \equiv \int_0^y\left\{\int_v^y s_Y(u;\theta)du\right\}m_Y(v;\theta)dv = \int_0^y\left\{\int_0^u m_Y(v;\theta)dv\right\}s_Y(u;\theta)du$$

where the choice of the lower bound of integration y is again irrelevant. Note that we have only assumed that $\mu_Y(y;\theta) \sim \kappa y^{-\alpha}$, but it is clear from $\mu_Y(y;\theta)/\kappa y^{-\alpha} \to 1$ as $y \to 0^+$ that for the purpose of calculating Σ_0 and N_0 we can do as if $\mu_Y(y;\theta) = \kappa y^{-\alpha}$ over the interval $(0,\varepsilon_0]$. Let $0<u\le\varepsilon_0$.

(i) If $\alpha>1$, we have for $0<v\le\varepsilon_0$

$$s_Y(v;\theta) = \exp\left\{\int_v 2\mu_Y(w;\theta)dw\right\} \ge \exp\left\{\int_v 2\kappa w^{-\alpha}dw\right\} = k_0\exp\left\{2\kappa(\alpha-1)v^{-(\alpha-1)}\right\}$$

and hence $\int_0^u s_Y(v;\theta)dv = +\infty$. If $\alpha=1$,

$$s_Y(v;\theta) \ge \exp\left\{\int_v 2\kappa w^{-1}dw\right\} = k_0\exp\{-2\kappa \mathrm{Ln}(v)\} = k_0 v^{-2\kappa}$$

and $\int_0^u s_Y(v;\theta)dv \geq \int_0^u k_0 v^{-2\kappa}dv = +\infty$ again since we have assumed that κ≥1/2 when α=1. In all these inequalities, k_0 denotes a different positive and finite constant. It follows from $\int_0^u s_Y(v;\theta)dv = +\infty$ and the finiteness of the measure m_Y in the second equality defining Σ_0 that $\Sigma_0 = \infty$, i.e., 0 is unattainable.

(ii) Among unattainable boundaries, whether 0 is an entrance or a natural boundary depends upon whether $N_0 < \infty$ or $N_0 = \infty$ respectively. We have in all cases $\mu_Y(w;\theta) \geq \kappa w^{-1}$ for some κ>0 [since if α>1, $\mu_Y(w;\theta) \geq \kappa w^{-\alpha} > \kappa w^{-1}$; note that this constant κ is not necessarily ≥ 1/2]. Then we can bound N_0 above as follows

$$
\begin{aligned}
N_0 &= \int_0^y\int_0^u \exp\left\{\int_u^v 2\mu_Y(w;\theta)dw\right\}dvdu = \int_0^y\int_0^u \exp\left\{-\int_v^u 2\mu_Y(w;\theta)dw\right\}dvdu \\
&\leq \int_0^y\int_0^u e^{-\int_v^u 2\kappa/w\,dw}dv\,du = \int_0^y\left\{\int_0^u v^{2\kappa}dv\right\}u^{-2\kappa}du \\
&= (2\kappa+1)^{-1}\int_0^y\left\{u^{2\kappa+1}\right\}u^{-2\kappa}du \;=\; (2\kappa+1)^{-1}y^2/2 \;<\; +\infty
\end{aligned}
$$

Therefore 0 is an entrance boundary for all α≥1. QED.

Proof of Theorem 1: Let $\bar{\Delta} > 0$ be the constant defined in Lemma 2 below. Let A_X be a compact set contained in D_X, and consider x_0 in A_X. Let A_Y be the compact set which contains the values of $\gamma(x_0;\theta)$ as x_0 varies in A_X and θ in Θ (recall that Θ is bounded). Define

$$
\zeta(\Delta, x \mid x_0;\theta) \;\equiv\; \Delta^{-1/2}\left(\gamma(x;\theta) - \gamma(x_0;\theta) - \mu_Y(\gamma(x_0;\theta);\theta)\Delta\right).
$$

We seek to bound:

$$
\begin{aligned}
&\left|p_X(\Delta, x \mid x_0;\theta) - p_X^{(J)}(\Delta, x \mid x_0;\theta)\right| \\
&\quad = \sigma(x;\theta)^{-1}\left|p_Y(\Delta, \gamma(x;\theta) \mid \gamma(x_0;\theta);\theta) - p_Y^{(J)}(\Delta, \gamma(x;\theta) \mid \gamma(x_0;\theta);\theta)\right| \\
&\quad = \sigma(x;\theta)^{-1}\Delta^{-1/2}\left|p_Z(\Delta, \zeta(\Delta, x \mid x_0;\theta) \mid \gamma(x_0;\theta);\theta) - p_Z^{(J)}(\Delta, \zeta(\Delta, x \mid x_0;\theta) \mid \gamma(x_0;\theta);\theta)\right|
\end{aligned}
$$

Consider the j-th coefficient of the approximating function $p_Z^{(J)}$:

$$
\begin{aligned}
\eta_j(\Delta, y_0;\theta) &= (j!)^{-1}\int_{-\infty}^{+\infty} H_j(w)p_Z(\Delta, w \mid y_0;\theta)dw \\
&= (j!)^{-1}(j+1)^{-1}\int_{-\infty}^{+\infty} H_{j+1}{}'(w)p_Z(\Delta, w \mid y_0;\theta)dw
\end{aligned}
$$

$$= \left((j+1)!\right)^{-1} H_{j+1}(w) p_Z(\Delta, w \mid y_0;\theta)\Big]_{-\infty}^{+\infty}$$
$$- \left((j+1)!\right)^{-1} \int_{-\infty}^{+\infty} H_{j+1}(w) \left\{ \partial p_Z(\Delta, w \mid y_0;\theta)/\partial w \right\} dw$$

From $y = y_0 + \mu_Y(\Delta \mid y_0;\theta)\Delta + \Delta^{1/2} w$, it follows that

$$(y - y_0)^2/\Delta = w^2 + 2w\Delta^{1/2}\mu_Y(\Delta \mid y_0;\theta) + \Delta\mu_Y^2(\Delta \mid y_0;\theta)$$

Therefore, from Lemma 2, the continuity of $\theta \mapsto \mu_Y(\Delta \mid y_0;\theta)$ and the boundedness of Θ, it follows that

$$0 < p_Z(\Delta, w \mid y_0;\theta) \ \le \ a_0 \exp\{-3w^2/8\} \exp\{a_1 |w||y_0| + a_2 |w| + a_3 |y_0| + a_4 y_0^2\}$$

where the constants a_i, i=0,...,4 are uniform in $\theta \in \Theta$. Combine this bound with Lemma 3 (iii) to obtain

$$\left| \left((j+1)!\right)^{-1} H_{j+1}(w) p_Z(\Delta, w \mid y_0;\theta) \right| \ \le \ \left((j+1)!\right)^{-1/2} (j+1)^{-1/4} K \left\{1 + \left| w^{5/2}/2^{5/4} \right| \right\} e^{w^2/4}$$
$$\times\, a_0\, e^{-3w^2/8}\, e^{a_1 |w||y_0| + a_2 |w| + a_3 |y_0| + a_4 y_0^2}$$

and therefore

$$\left((j+1)!\right)^{-1} H_{j+1}(w) p_Z(\Delta, w \mid y_0;\theta)\Big]_{-\infty}^{+\infty} = 0 .$$

We now prove that the series $\sum_{j=0}^{\infty} j!\, \nu_j^2(\Delta, y_0;\theta)$ converges, where

$$\nu_j(\Delta, y_0;\theta) \equiv (j!)^{-1} \int_{-\infty}^{+\infty} H_j(w) \left\{ \partial p_Z(\Delta, w \mid y_0;\theta)/\partial w \right\} dw .$$

First, note that the integral $\int_{-\infty}^{+\infty} e^{w^2/2} \left\{ \partial p_Z(\Delta, w \mid y_0;\theta)/\partial w \right\}^2 dw$ converges, since from the second bound in Lemma 2, we obtain that:

$$\left| \partial p_Z(\Delta, w \mid y_0;\theta)/\partial y \right| \ \le \ b_0 \exp\{-3w^2/8\} \times R(|w|, |y_0|)$$
$$\times \ \exp\{b_1 |w||y_0| + b_2 |w| + b_3 |y_0| + b_4 y_0^2\}$$

where R is a polynomial of finite order in $(|w|,|y_0|)$ with coefficients uniform in $\theta \in \Theta$, and where the constants b_i, i=0,...,4 are uniform in $\theta \in \Theta$.

Second, expand the squared term in

$$0 \ \le \int_{-\infty}^{+\infty} e^{w^2/2} \left\{ \partial p_Z(\Delta, w \mid y_0;\theta)/\partial w - \phi(w) \sum_{j=0}^{J} \nu_j(\Delta, y_0;\theta) H_j(w) \right\}^2 dw$$

$$
\begin{aligned}
&= \int_{-\infty}^{+\infty} e^{w^2/2}\{\partial p_Z(\Delta,w\mid y_0;\theta)/\partial w\}^2 dw \\
&\quad - 2(2\pi)^{-1/2}\sum_{j=0}^{J} v_j(\Delta,y_0;\theta)\int_{-\infty}^{+\infty}\{\partial p_Z(\Delta,w\mid y_0;\theta)/\partial w\}H_j(w)dw \\
&\quad + (2\pi)^{-1}\sum_{j=0}^{J}\sum_{k=0}^{J} v_j(\Delta,y_0;\theta)v_k(\Delta,y_0;\theta)\int_{-\infty}^{+\infty} e^{-w^2/2}H_j(w)H_k(w)dw \\
&= \int_{-\infty}^{+\infty} e^{w^2/2}\{\partial p_Z(\Delta,w\mid y_0;\theta)/\partial w\}^2 dw \ - \ (2\pi)^{-1/2}\sum_{j=0}^{J} j!\, v_j^2(\Delta,y_0;\theta)
\end{aligned}
$$

and the (dominated) convergence of the series on the right-hand-side follows. Further, the series converges uniformly with respect to θ in Θ and to y_0 in the compact set A_Y.

Next, we prove that the expansion $p_Z^{(J)}$ of p_Z converges. We can bound the terms of order $j\geq 1$ in the series according to

$$
\begin{aligned}
\left|\eta_j(\Delta,y_0;\theta)H_j(z)\right| &= ((j+1)!)^{-1}\left|\int_{-\infty}^{+\infty} H_{j+1}(w)\{\partial p_Z(\Delta,w\mid y_0;\theta)/\partial w\}dw\right|\left|H_j(z)\right| \\
&= \left|v_{j+1}(\Delta,y_0;\theta)\right|\left|H_j(z)\right| \\
&\leq K\left\{1+\left|z^{5/2}/2^{5/4}\right|\right\}e^{z^2/4}\times\left\{j^{-1/4}(j!)^{1/2}\left|v_{j+1}(\Delta,y_0;\theta)\right|\right\} \\
&\leq K\left\{1+\left|z^{5/2}/2^{5/4}\right|\right\}e^{z^2/4}\times\left\{j^{-1/4}(j+1)^{-1/2}((j+1)!)^{1/2}\left|v_{j+1}(\Delta,y_0;\theta)\right|\right\} \\
&\leq K\left\{1+\left|z^{5/2}/2^{5/4}\right|\right\}e^{z^2/4}\left\{ j^{-1/2}(j+1)^{-1}+(j+1)!\,v_{j+1}^2(\Delta,y_0;\theta)\right\}/2
\end{aligned}
$$

since $|\alpha\beta|\leq(\alpha^2+\beta^2)/2$. But both series

$$
\sum j^{-1/2}(j+1)^{-1} \text{ and } \sum (j+1)!\, v_{j+1}^2(\Delta,y_0;\theta)
$$

are convergent. Hence $p_Z^{(J)}(\Delta,z\mid y_0;\theta) = \phi(z)\sum_{j=0}^{J}\eta_j(\Delta,y_0;\theta)H_j(z)$ is convergent as $J\to\infty$. Note that the convergence is uniform in z over the entire real line since the two series just above converged uniformly with respect to z. The convergence is also uniform with respect to θ in Θ and y_0 in the compact set A_Y.

The last point that remains to be proved is that the limit of $p_Z^{(J)}(\Delta,z\mid y_0;\theta)$ is indeed $p_Z(\Delta,z\mid y_0;\theta)$. Let $q_Z(\Delta,z\mid y_0;\theta)\equiv\lim_{J\to\infty}p_Z^{(J)}(\Delta,z\mid y_0;\theta)$. q_Z is continuous in z as the uniform limit of a series of continuous functions. Further, with

$$
\varepsilon_{j+1}\equiv j^{-1/2}(j+1)^{-1}+(j+1)!\,v_{j+1}^2(\Delta,y_0;\theta)
$$

note that there exists a constant K_0 such that

$$
\phi(z)\left|\eta_j(\Delta,y_0;\theta)H_j(z)\right| \ \leq \ K\left\{1+\left|z^{5/2}/2^{5/4}\right|\right\}e^{-z^2/4}\,\varepsilon_{j+1} \ \leq \ K_0\,e^{-3z^2/8}\,\varepsilon_{j+1}
$$

(for z large enough) and hence q_Z satisfies the same bound as p_Z in Lemma 2. Therefore the integral $(k!)^{-1}\int_{-\infty}^{+\infty} q_Z(\Delta,w\,|\,y_0;\theta)H_k(w)\,dw$ exists and since

$$\begin{aligned}
&(k!)^{-1}\int_{-\infty}^{+\infty} p_Z^{(J)}(\Delta,w\,|\,y_0;\theta)H_k(w)\,dw \\
&\quad = (k!)^{-1}\sum_{j=0}^{J}\eta_j(\Delta,y_0;\theta)\int_{-\infty}^{+\infty}(2\pi)^{-1/2}e^{-w^2/2}H_j(w)H_k(w)\,dw \\
&\quad = \begin{cases}\eta_k(\Delta,y_0;\theta) & \text{if } k\le J\\ 0 & \text{if } k>J\end{cases}
\end{aligned}$$

we have that $(k!)^{-1}\int_{-\infty}^{+\infty} q_Z(\Delta,w\,|\,y_0;\theta)H_k(w)\,dw=\eta_k(\Delta,y_0;\theta)$, and so p_Z and q_Z have the same η_k coefficients for all $k\geq 0$.

To conclude, it is easy to see that two continuous functions satisfying the same first bound as in Lemma 2 and sharing the same η_k coefficients for all k must be equal. Indeed, define the difference $r_Z(\Delta,w\,|\,y_0;\theta)\equiv q_Z(\Delta,w\,|\,y_0;\theta)-p_Z(\Delta,w\,|\,y_0;\theta)$. The integral of r_Z against polynomials w^k of all orders $k\geq 0$ is equal to zero (since any polynomial of order k is a linear combination of the first k polynomials H_k) and therefore by Weierstrass's approximation theorem the function r_Z is identically zero.

Let us now go back to p_X. We have shown that, for every $\varepsilon>0$, there exists $J_\varepsilon(A_Y;\Theta)$ such that for all $J \geq J_\varepsilon(A_Y;\Theta)$, the bound:

$$\left|p_Z(\Delta,z\,|\,y_0;\theta)-p_Z^{(J)}(\Delta,z\,|\,y_0;\theta)\right| \leq \varepsilon$$

holds for all $z\in R$, $y_0\in A_Y$ and $\theta\in\Theta$.

If σ is globally non-degenerate under Assumption 2(1), $\sigma^{-1}(x;\theta)<c^{-1}<+\infty$ implies that for all $J\geq J_\varepsilon(A_X;\Theta)$:

$$\left|p_X(\Delta,x\,|\,x_0;\theta)-p_X^{(J)}(\Delta,x\,|\,x_0;\theta)\right| \leq \varepsilon$$

for all x in R, $x_0\in A_X$ and $\theta\in\Theta$. Otherwise, under Assumption 2(2), for every $\varepsilon>0$, there exists a constant c_ε such that $\sigma^{-1}(x;\theta)<c_\varepsilon^{-1}<+\infty$ for all $x\in[\varepsilon,+\infty)$ and $\theta\in\Theta$. Therefore the uniform convergence of $p_Z^{(J)}$ to p_Z for z in R implies the uniform convergence of $p_X^{(J)}$ to p_X for x in $[\varepsilon,+\infty)$ since for such x's:

$$\begin{aligned}
&\left|p_X(\Delta,x\,|\,x_0;\theta)-p_X^{(J)}(\Delta,x\,|\,x_0;\theta)\right| \\
&\quad=\sigma(x;\theta)^{-1}\Delta^{-1/2}\left|p_Z(\Delta,\zeta(\Delta,x\,|\,x_0;\theta)\,|\,\gamma(x_0;\theta);\theta)-p_Z^{(J)}(\Delta,\zeta(\Delta,x\,|\,x_0;\theta)\,|\,\gamma(x_0;\theta);\theta)\right| \\
&\quad\leq c_\varepsilon^{-1}\Delta^{-1/2}\left|p_Z(\Delta,\zeta(\Delta,x\,|\,x_0;\theta)\,|\,\gamma(x_0;\theta);\theta)-p_Z^{(J)}(\Delta,\zeta(\Delta,x\,|\,x_0;\theta)\,|\,\gamma(x_0;\theta);\theta)\right|.
\end{aligned}$$

QED.

Proof of Lemma 2: (i) Consider first the case where $D_Y = (-\infty,+\infty)$. Since Y has unit diffusion, an application of Girsanov's Theorem yields

$$p_Y(\Delta, y \mid y_0;\theta) = (2\pi\Delta)^{-1/2} \exp\left\{-(y-y_0)^2/2\Delta + \int_{y_0}^{y} \mu_Y(w;\theta)dw\right\}$$
$$\times\ E\left[\exp\left\{\Delta\int_0^1 g\left((1-u)y_0 + uy + \Delta^{1/2}B_u;\theta\right)du\right\}\right]$$

where $g(w;\theta) \equiv -\left(\mu_Y^2(w;\theta) + \partial\mu_Y(w;\theta)/\partial w\right)/2$ and $\{B_u / u \in [0,1]\}$ designates a Brownian Bridge with $B_0=B_1=0$ [see Rogers (1984) and Dacunha-Castelle and Florens-Zmirou (1986, Lemma 1)]. The strict positivity of p_Y follows from that expression, and the bound from bounding each of the terms.

Assumption 3 gives the bound

$$\int_{y_0}^{y} \mu_Y(w;\theta)dw \ \leq\ H + L\left| y - y_0 \right|\left(1 + \left| y_0 \right|\right) + M\left(y - y_0\right)^2$$

for all y in D_Y, where H and M are positive constants [if $y \geq 0$, decompose the integral from y_0 to E_0, where μ_Y is bounded as a continuous function on a compact interval, and then from E_0 to y, where μ_Y is bounded by Ky; a similar argument holds for $y \leq 0$]. Hence in general M=K. Note that this is an upper bound for the integral itself, not its absolute value. It is useful to note that if $\mu_Y \leq 0$ near $+\infty$ and $\mu_Y \geq 0$ near $-\infty$ then M can be set to 0 in the expression above.

Then by the continuity of $g(w;\theta)$ in w, and its limit behavior near the boundaries under Assumption 3, it follows that there exists $\gamma \geq 0$ such that $g(w;\theta) \leq \gamma$ for all w>0 and $\theta \in \Theta$ [in general, of course, g will not be bounded below]. Therefore

$$E\left[\exp\left\{\Delta\int_0^1 g\left((1-u)y_0 + uy + \Delta^{1/2}B_u;\theta\right)du\right\}\right] \ \leq\ e^{\gamma\Delta}$$

Collecting all terms we have that

$$p_Y(\Delta, y \mid y_0;\theta) \ \leq\ (2\pi\Delta)^{-1/2} e^{-(y-y_0)^2/2\Delta + H + L|y-y_0|(1+|y_0|) + K(y-y_0)^2} \times e^{\gamma\Delta}$$
$$\leq\ C_0\Delta^{-1/2}\, e^{-3(y-y_0)^2/8\Delta} \times e^{C_1|y-y_0||y_0| + C_2|y-y_0| + C_3|y_0| + C_4 y_0^2}$$

provided that $-1/(2\Delta) + M \leq -3/(8\Delta)$, i.e., that $0 < \Delta \leq \bar{\Delta} \equiv (8M)^{-1}$. It is clear from the argument that we could replace $3/(8\Delta)$ in the bound for p_Y by any number less than but arbitrarily close to $1/(2\Delta)$, at the cost of reducing $\bar{\Delta}$, but this will not be necessary.

Further, when $\mu_Y \leq 0$ near $+\infty$ and $\mu_Y \geq 0$ near $-\infty$, M=0 and hence $\overline{\Delta} = +\infty$ and we can replace $3/(8\Delta)$ by $1/(2\Delta)$.

(ii) For the second part of the lemma, we calculate

$$\partial p_Y(\Delta, y \mid y_0;\theta)/\partial y = (2\pi\Delta)^{-1/2} \exp\left\{-(y-y_0)^2/2\Delta + \int_{y_0}^{y} \mu_Y(w;\theta)dw\right\}$$
$$\times\left\{\left\{-(y-y_0)/\Delta + \mu_Y(y;\theta)\right\} E\left[e^{\Delta\int_0^1 g\left((1-u)y_0+uy+\Delta^{1/2}B_u;\theta\right)du}\right]\right.$$
$$\left. + E\left[\Delta\int_0^1 ug'\left((1-u)y_0+uy+\Delta^{1/2}B_u;\theta\right)du \times e^{\Delta\int_0^1 g\left((1-u)y_0+uy+\Delta^{1/2}B_u;\theta\right)du}\right]\right\}$$

where $g'(w;\theta) \equiv \partial g(w;\theta)/\partial w = -\left(2\mu_Y(w;\theta)\partial\mu_Y(w;\theta)/\partial w + \partial^2\mu_Y(w;\theta)/\partial w^2\right)/2$. First, we have $\left|\left\{-(y-y_0)/\Delta + \mu_Y(y;\theta)\right\}\right| \leq Q_1(|y|,|y_0|)$ where Q_1 is a polynomial of degree one in $(|y|,|y_0|)$, with coefficients uniformly bounded in $\theta \in \Theta$.

Second $\left|E\left[Ae^{B}\right]\right| \leq E\left[|A|e^{|B|}\right]$ and $E\left[e^{\Delta\int_0^1 g\left((1-u)y_0+uy+\Delta^{1/2}B_u;\theta\right)du}\right] \leq e^{\gamma\Delta}$, so:

$$\left|E\left[\Delta\int_0^1 ug'\left((1-u)y_0+uy+\Delta^{1/2}B_u;\theta\right)du \times e^{\Delta\int_0^1 g\left((1-u)y_0+uy+\Delta^{1/2}B_u;\theta\right)du}\right]\right|$$
$$\leq \Delta\, E\left[\int_0^1 u\left|g'\left((1-u)y_0+uy+\Delta^{1/2}B_u;\theta\right)\right|du\right] \times e^{\gamma\Delta}$$

To bound the expected value on the right-hand-side, recall that $g'(w;\theta)$ has at most exponential growth. Hence there exists $\lambda>0$ and $G>0$ such that $|g'(w;\theta)| \leq Ge^{\lambda|w|}$ and thus

$$E\left[\int_0^1 u\left|g'\left((1-u)y_0+uy+\Delta^{1/2}B_u;\theta\right)\right|du\right] \leq GE\left[\int_0^1 u e^{\left|(1-u)y_0+uy+\Delta^{1/2}B_u\right|}du\right]$$
$$= G\int_0^1 u E\left[e^{\left|(1-u)y_0+uy+\Delta^{1/2}B_u\right|}\right]du \leq G\int_0^1 u e^{|(1-u)y_0|+|uy|} E\left[e^{\Delta^{1/2}|B_u|}\right]du$$

B_u is distributed as $N(0,u(1-u))$. If N is distributed as $N(0,\sigma^2)$, the density of $|N|$ is given by $2(2\pi)^{-1/2}\sigma^{-1}\exp\left\{-x^2/2\sigma^2\right\}$, $x \geq 0$. Therefore for any constant a:

$$E\left[e^{a|N|}\right] = 2(2\pi)^{-1/2}\sigma^{-1}\int_0^{+\infty} e^{ax}e^{-x^2/2\sigma^2}dx = 2(2\pi)^{-1/2}\sigma^{-1}e^{a^2\sigma^2/2}\int_0^{+\infty} e^{-\left(x-a\sigma^2\right)^2/2\sigma^2}dx$$
$$= e^{a^2\sigma^2/2}(2\pi)^{-1/2}\sigma^{-1}\int_{-\infty}^{+\infty} e^{-\left(x-a\sigma^2\right)^2/2\sigma^2}dx = e^{a^2\sigma^2/2}$$

and it follows that $E\left[e^{\Delta^{1/2}|B_u|}\right] = e^{\Delta u(1-u)/2}$. Hence

$$E\left[\int_0^1 u\left|g'\left((1-u)y_0+uy+\Delta^{1/2}B_u;\theta\right)\right|du\right] \leq G\int_0^1 ue^{(1-u)|y_0|+u|y|+\Delta u(1-u)/2}du \leq Ge^{|y_0|+|y|}$$

(since u runs from 0 to 1) and we can conclude that

$$\left|\partial p_Y(\Delta, y \mid y_0;\theta)/\partial y\right| \leq D_0 \Delta^{-1/2} \exp\{-3(y-y_0)^2/8\Delta\} \times P(|y|,|y_0|) \times \exp\{C_1|y-y_0||y_0| + C_2|y-y_0| + C_3|y_0| + C_4 y_0^2\}$$

for all $0<\Delta<\bar{\Delta}$, where the constant D_0 is uniform in θ and P is a polynomial of finite degree with coefficients also uniform in θ.

(iii) Consider now the case where $D_Y = (0,+\infty)$. From the proof of Theorem 1, we need to show that the integral $\int e^{w^2/2}\{\partial p_Z(\Delta, w \mid y_0;\theta)/\partial w\}^2 dw$ converges. That is, after a change of variable $Z \rightarrow Y$, we need to show that the integral

$$\int_0^{+\infty} \Delta^{1/2} e^{(y-y_0-\mu_Y(y_0;\theta)\Delta)^2/2\Delta} \{\partial p_Y(\Delta, y \mid y_0;\theta)/\partial y\}^2 dy$$

converges at both boundaries 0^+ and $+\infty$. The boundary 0^+ is either an entrance or a natural boundary for Y, and in both cases $\lim_{y\rightarrow 0^+} \partial p_Y(\Delta, y \mid y_0;\theta)/\partial y = 0$ [see McKean (1956), Remark 4.2 page 541]. Hence the integral converges at the boundary 0^+.

The change of measure that we used in (1) above is no longer applicable, because we cannot transform Y into a Brownian motion: the two distributions are no longer absolutely continuous with respect to one another since Y is now distributed on a subset of the real line whereas a Brownian motion is distributed on the entire real line. However, we can transform Y into a Brownian motion on $[0,+\infty)$, reflected at 0. Its transition density is known to be

$$p_{RBM}(\Delta, w \mid w_0) = (2\pi\Delta)^{-1/2}\left\{\exp\left(-(w-w_0)^2/2\Delta\right) + \exp\left(-(w+w_0)^2/2\Delta\right)\right\}$$

for $w \geq 0$, $w_0 \geq 0$.

Therefore, by Girsanov's Theorem, we have for $y>0$, $y_0>0$:

$$p_Y(\Delta, y \mid y_0;\theta) = p_{RBM}(\Delta, y \mid y_0) \times \exp\left\{\int_{y_0}^{y} \mu_Y(w;\theta)dw\right\} \times E^{RBM}\left[\exp\left\{\int_0^{\Delta} g(Y_u;\theta)du\right\} \middle| Y_t = y, Y_0 = y_0\right]$$

where E^{RBM} indicates that inside the expectation Y follows the law of a Brownian motion reflected at 0. We have that

$$p_{RBM}(\Delta, y \mid y_0) < 2(2\pi\Delta)^{-1/2} \exp\left(-(y-y_0)^2/2\Delta\right)$$

since $\exp\left(-(y+y_0)^2/2\Delta\right) < \exp\left(-(y-y_0)^2/2\Delta\right)$ for y>0, y_0>0. Therefore the same bound as in cases (1) and (2) apply. QED

Proof of Lemma 3: Recall the bound for p_Y derived in Lemma 2:

$$u_Y(\Delta \mid y_0;\theta,j) \leq \exp\left\{C_3 \mid y_0 \mid + C_4 y_0^2\right\} \frac{C_0}{\Delta^{1/2}}$$
$$\times \int_{-\infty}^{+\infty} \mid w - y_0 \mid^j \exp\left\{-3w^2/8\Delta + C_1 |y_0| |w| + C_2 |w|\right\} dy$$

where we have changed the variable y to w=y-y_0. For each Δ and y_0 there exists a value $\bar{y}(\Delta, y_0) \geq 0$ such that for all w, $|w| \geq \bar{y}(\Delta, y_0)$ implies that

$$-3w^2/8\Delta + C_1 |y_0| |w| + C_2 |w| \leq -5w^2/16\Delta$$

and the convergence of the integral follows. QED.

Proof of Lemma 4: (i): See e.g., Sansone (1991, page 304); (ii): see e.g., Sansone (1991, page 308); (iii): see Stone (1927, Theorem II).

Proof of Lemma 5: The proof is based on verifying the conditions (B.1)-(B.3) of Basawa and Scott (1983, page 33). First, it is clear that for every $\theta \in \Theta$, $\ddot{L}_i(\theta)$ exists. Indeed, from the expression that we already used in the proof of Lemma 2,

$$\mathrm{Ln}\left(p_Y(\Delta, y \mid y_0;\theta)\right) = -\mathrm{Ln}(2\pi\Delta)/2 - (y-y_0)^2/2\Delta + \int_{y_0}^{y} \mu_Y(w;\theta)dw$$
$$+ \mathrm{Ln}\left(E\left[\exp\left\{\Delta \int_0^1 g\left((1-u)y_0 + uy + \Delta^{1/2} B_u;\theta\right)du\right\}\right]\right)$$

and recall that μ_Y and g are twice differentiable in θ, and that p_X is given by (2.8); apply the differentiation chain rule to conclude.

Next, for every θ and k=1,..,K, the k×k entry in the matrix $E\left[\dot{L}_i(\theta)\dot{L}_i(\theta)^T \middle| X_{(i-1)\Delta}\right]$ is finite, which follows from the finiteness of (3.4). Hence $I_n(\theta)$ is well-defined, and (B.1) is satisfied.

(B.2), i.e., $i_n^{-1}(\theta) \xrightarrow{a.s.} 0$ and $G_n(\theta) \equiv i_n^{-1/2}(\theta)\, H_n(\theta)\, i_n^{-1/2}(\theta) \xrightarrow{P} G(\theta)$ is given by Assumptions 4 and 5.

(B.3) is a continuity requirement: for all $\varepsilon>0$, let $N_n^{\varepsilon}(\theta) \equiv \left\{\tilde{\theta} \in \Theta \middle| \left\| i_n^{1/2}(\theta)\left(\tilde{\theta}-\theta\right)\right\| \leq \varepsilon\right\}$ where $\| \ \|$ denotes the Euclidean norm on R^K, we need that

$$\sup_{\tilde{\theta}\in N_n^{\varepsilon}(\theta)}\left\| i_n\left(\tilde{\theta}\right) i_n^{-1}(\theta) - \mathrm{Id}\right\| \xrightarrow{\ p\ } 0$$

and

$$\sup_{\tilde{\theta}\in N_n^{\varepsilon}(\theta)}\left\| i_n^{-1/2}(\theta)\left\{H_n\left(\tilde{\theta}\right) - H_n(\theta)\right\} i_n^{-1/2}(\theta)\right\| \xrightarrow{\ p\ } 0,$$

both uniformly in θ. A sufficient condition under (3.4) is that the functions

$$\theta \mapsto \int_{\underline{x}}^{\bar{x}} \left\{\partial \mathrm{Ln}\left(p_X(\Delta, x \mid x_0;\theta)\right)/\partial\theta_k\right\}^2 p_X(\Delta, x \mid x_0;\theta)\,dx$$

and $\theta \mapsto \partial^2 \mathrm{Ln}\left(p_X(\Delta, x \mid x_0;\theta)\right)/\partial\theta\partial\theta^T$ be continuous in $\theta\in\Theta$, for all $x_0\in D_X$. This follows from the continuity of $\theta \mapsto p_X(\Delta, x \mid x_0;\theta)$ and its first two derivatives. The continuity is uniform in θ because Θ is bounded.

These conditions are essentially equivalent to multivariate extensions of Assumptions 1 and 2 in Hall and Heyde (1980, Proposition 6.1, page 160). It follows from verifying these conditions that

$$\left\{i_n^{-1/2}(\theta)S_n(\theta), G_n(\theta)\right\} \xrightarrow{\ d\ } \left\{G^{1/2}(\theta)N(0,\mathrm{Id}), G(\theta)\right\}$$

which is a version of the Central Limit Theorem for Martingales in Chapter 1.4, Theorem 1 page 34 [this reference and the following are from Basawa and Scott (1983).

Now the MLE exists and its distribution follows from Chapter 2.4, Theorem 2 page 58, which is a classical Taylor expansion of the score function:

$$i_n^{1/2}(\theta_0)\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\ d\ } G^{-1/2}(\theta_0)\times N(0,\mathrm{Id}) \text{ under } P_{\theta_0}$$

The efficiency part of Lemma 5 is an adaptation of Chapter 2.4, Theorem 3 page 60; the Normal asymptotic variance comparison follows from Chapter 2.3, Corollary 2, page 53. QED.

Proof of Theorem 2: (i) Fix $\varepsilon>0$ and $x_0\in R$. Let $r_X^{(J)}(\Delta, x \mid x_0;\theta) \equiv p_X(\Delta, x \mid x_0;\theta)$ - $p_X^{(J)}(\Delta, x \mid x_0;\theta)$ and

$$R_X^{(J)}(\Delta, x \mid x_0;\Theta) \equiv \sup_{\theta\in\Theta}\left| r_X^{(J)}(\Delta, x \mid x_0;\theta)\right|$$

and also define the corresponding quantities for Y and Z. Recall that

$$\left|p_X(\Delta,x\,|\,x_0;\theta)-p_X^{(J)}(\Delta,x\,|\,x_0;\theta)\right|$$
$$=\sigma(x;\theta)^{-1}\left|p_Y(\Delta,\gamma(x;\theta)\,|\,\gamma(x_0;\theta);\theta)-p_Y^{(J)}(\Delta,\gamma(x;\theta)\,|\,\gamma(x_0;\theta);\theta)\right|$$

and hence $R_X^{(J)}(\Delta,x\,|\,x_0;\Theta) \;=\; \sup_{\theta\in\Theta}\left|\sigma(x;\theta)^{-1}\right|\times R_Y^{(J)}(\Delta,\gamma(x;\theta)\,|\,\gamma(x_0;\theta);\Theta)$

By Theorem 1, the convergence of $p_Y^{(J)}(\Delta,y\,|\,y_0;\theta)$ to $p_Y(\Delta,y\,|\,y_0;\theta)$ is uniform in y over D_Y and in θ over Θ, and in y over bounded subsets A_Y of D_Y. Hence there exists $J_\varepsilon(\Delta,A_Y;\theta)$ such that for all $J \geq J_\varepsilon(\Delta,A_Y;\Theta)$ we have

$$\sup_{\theta\in\Theta}\sup_{y\in D_Y}\sup_{y_0\in A_Y}\left|r_Y^{(J)}(\Delta,y\,|\,y_0;\theta)\right| < \varepsilon.$$

Now recall that

$$\left|p_X(\Delta,x\,|\,x_0;\theta)-p_X^{(J)}(\Delta,x\,|\,x_0;\theta)\right|$$
$$=\sigma(x;\theta)^{-1}\left|p_Y(\Delta,\gamma(x;\theta)\,|\,\gamma(x_0;\theta);\theta)-p_Y^{(J)}(\Delta,\gamma(x;\theta)\,|\,\gamma(x_0;\theta);\theta)\right|$$

x_0 is fixed. Let A_Y be the set of y_0 described by $\gamma(x_0;\theta)$ as θ varies in Θ. Since Θ is bounded and γ is continuous in θ [σ is by Assumption 1] A_Y is bounded, and therefore

$$\begin{aligned}R_X^{(J)}(\Delta,x\,|\,x_0;\Theta) &\leq \sup_{\theta\in\Theta}\left\{\sigma(x;\theta)^{-1}\right\}\times\sup_{\theta\in\Theta}\sup_{y\in D_Y}\sup_{y_0\in A_Y}\left|r_Y^{(J)}(\Delta,y\,|\,y_0;\theta)\right|\\ &\leq \sup_{\theta\in\Theta}\left\{\sigma(x;\theta)^{-1}\right\}\times\varepsilon\end{aligned}$$

Let $\Sigma^{-1}(x)\equiv\sup_{\theta\in\Theta}\left\{\sigma(x;\theta)^{-1}\right\}$, which is finite by the boundedness of Θ and the continuity of σ^{-1} in θ.

Then for m=1 and m=2, we have that

$$\begin{aligned}\left|E_{\theta_0}\left[\left\{R_X^{(J)}(\Delta,X_{t+\Delta}\,|\,X_t;\Theta)\right\}^m\middle|X_t=x_0\right]\right| &\leq \int_{\underline{x}}^{\bar{x}}\left|R_X^{(J)}(\Delta,x\,|\,x_0;\Theta)\right|^m p_X(\Delta,x\,|\,x_0;\theta_0)\,dx\\ &\leq \varepsilon^m\int_{\underline{x}}^{\bar{x}}\Sigma^{-m}(x)p_X(\Delta,x\,|\,x_0;\theta_0)\,dx\end{aligned}$$

i.e., $\lim_{J\to\infty}E_{\theta_0}\left[\left\{R_X^{(J)}(\Delta,X_{t+\Delta}\,|\,X_t;\Theta)\right\}^m\middle|X_t=x_0\right]=0$ for m=1,2, provided that we prove that the two integrals $\int_{\underline{x}}^{\bar{x}}\Sigma^{-m}(x)p_X(\Delta,x\,|\,x_0;\theta_0)\,dx$, m=1,2, converge.

(ii) A difficulty only arises when σ is degenerate [otherwise $\Sigma^{-m}(x)\leq c^{-m}$ under Assumption 2(i), and then $\int_{\underline{x}}^{\bar{x}}\Sigma^{-m}(x)p_X(\Delta,x\,|\,x_0;\theta_0)\,dx\;\leq c^{-m}$]. Under Assumption 2(ii), degeneracy may happen when D = (0,+∞) and σ(0;θ)=0. Applying the change of variable X→Y, we wish to prove convergence of the integral

$$\int_{\underline{y}}^{\bar{y}} \Sigma^{-m}\left(\gamma^{-1}(y;\theta)\right) \times p_Y\left(\Delta, y \mid y_0;\theta\right) dy.$$

[Σ^{-1} means $1/\Sigma$ whereas γ^{-1} represents the inverse of the function γ].

Since the only degeneracy of σ is near the left boundary $\underline{x} = 0^+$, we need to consider the two cases where $\underline{y} \equiv \lim_{x \to 0^+} \gamma(x;\theta)$ is either 0^+ or $-\infty$. Under Assumption 2(ii), we have that $\sigma^{-1}(x;\theta) \leq \omega^{-1} x^{-\rho}$ for all $0<x\leq\xi_0$ and $\theta \in \Theta$. For $0<x\leq\xi_0$, we have

$$\int_0^x du/\sigma(u;\theta) \;\leq\; \int_0^x \omega^{-1} u^{-\rho} du = \omega^{-1}(1-\rho)^{-1} x^{1-\rho}$$

if $0\leq\rho<1$, and therefore $\underline{y} = 0^+$ by taking the limit as x tends to 0^+. Let $x=\gamma^{-1}(y;\theta)$, and we have just shown that for y near 0^+, $y \leq \omega^{-1}(1-\rho)^{-1} x^{1-\rho}$, from which it follows that $\gamma^{-1}(y;\theta) \;\geq\; \left(\omega(1-\rho)y\right)^{1/(1-\rho)}$ and consequently

$$\Sigma^{-m}\left(\gamma^{-1}(y;\theta)\right) \leq \omega^{-m}\left[\gamma^{-1}(y;\theta)\right]^{-m\rho} \leq \omega^{-m}\left(\omega(1-\rho)y\right)^{-m\rho/(1-\rho)}.$$

So naturally the upper bound tends to $+\infty$ as y tends to 0^+. The issue is whether this upper bound increases faster than p_Y decreases as y tends to 0^+. To answer that question, we need to call upon Assumption 3(1.i). For $0<y\leq\varepsilon_0$,

$$\exp\left(\int_{\varepsilon_0}^{y} \mu_Y(w;\theta)dw\right) \;=\; \exp\left(-\int_y^{\varepsilon_0} \mu_Y(w;\theta)dw\right)$$

$$\leq\; \exp\left(-\kappa\int_y^{\varepsilon_0} w^{-\alpha}dw\right) \;=\; \begin{cases} \varepsilon_0^{-\kappa} y^{\kappa} & \text{if } \alpha = 1 \\ e^{\kappa(\alpha-1)\varepsilon_0^{-(\alpha-1)} - \kappa(\alpha-1)y^{-(\alpha-1)}} & \text{if } \alpha > 1 \end{cases}$$

will provide an upper bound to p_Y for y near 0^+ [see the proof of Lemma 2; the other terms are bounded near 0^+]. It is clear that if $\alpha>1$ the left tail of p_Y decays exponentially fast, while the upper bound for $\Sigma^{-m}\left(\gamma^{-1}(y;\theta)\right) \leq \omega^{-m}\left[\gamma^{-1}(y;\theta)\right]^{-m\rho} \leq \omega^{-m}\left(\omega(1-\rho)y\right)^{-m\rho/(1-\rho)}$ increases only geometrically, so the integral will converge. If $\alpha=1$, then the tail of p_Y is bounded above by y^{κ} and therefore the integral will converge if $\kappa \geq 2\rho/(1-\rho)$. This is insured by Assumption 6.

If instead $\rho\geq1$, then

$$\underline{y} = \lim_{x \to 0^+} \int_{+\infty}^{x} \sigma^{-1}(u;\theta)du \;= \int_{+\infty}^{\xi_0} \sigma^{-1}(u;\theta)du + \lim_{x \to 0^+} \int_{\xi_0}^{x} \sigma^{-1}(u;\theta)du$$

where $\int_{+\infty}^{\xi_0} \sigma^{-1}(u;\theta)du \leq 0$ and

$$\lim_{x\to 0^+} \int_{\xi_0}^{x} \sigma^{-1}(u;\theta)du \;\le\; \lim_{x\to 0^+} \int_{\xi_0}^{x} \omega^{-1}u^{-\rho}du$$
$$= \lim_{x\to 0^+} \begin{cases} \omega^{-1}\mathrm{Ln}(x) & \text{if } \rho = 1 \\ -\omega^{-1}(\rho-1)^{-1}x^{-(\rho-1)} & \text{if } \rho > 1 \end{cases}$$
$$= -\infty$$

so $\underline{y} = -\infty$ when $\rho \ge 1$.

In that case, we have for y near $\underline{y}$: $\Sigma^{-m}(\gamma^{-1}(y;\theta)) \le \omega^{-m}[\gamma^{-1}(y;\theta)]^{-m\rho}$. Let $x=\gamma^{-1}(y;\theta)$. From the same calculation as above, we have $y \le \omega^{-1}\mathrm{Ln}(x)$ if $\rho=1$. Thus $\gamma^{-1}(y;\theta) \ge e^{\omega y}$ and therefore $\Sigma^{-m}(\gamma^{-1}(y;\theta)) \le \omega^{-m}e^{-m\rho\omega y}$. Now from Lemma 2, we know that p_Y is bounded above by a term of the form $e^{-3y^2/8\Delta}$, so the integral of $e^{-m\rho\omega y}e^{-3y^2/8\Delta}$ converges for y near $-\infty$. If $\rho>1$, $y \le -\omega^{-1}(\rho-1)^{-1}x^{-(\rho-1)}$ and therefore

$$\gamma^{-1}(y;\theta) \;\ge\; (-\omega(\rho-1)y)^{-1/(\rho-1)} \Rightarrow \Sigma^{-m}(\gamma^{-1}(y;\theta)) \le \omega^{-m}(-\omega(\rho-1)y)^{m\rho/(\rho-1)}$$

which again tends to $+\infty$ as y tends to $-\infty$, but not fast enough to overcome the decay $e^{-3y^2/8\Delta}$ of p_Y. Hence the integral $\int_{\underline{y}}^{\bar{y}} \Sigma^{-m}(\gamma^{-1}(y;\theta)) \times p_Y(\Delta, y \mid y_0;\theta)dy$ converges near $\underline{y}$ = $-\infty$ when $\rho \ge 1$.

We can therefore conclude that $\lim_{J\to\infty} E_{\theta_0}\left[\left\{R_X^{(J)}(\Delta, X_{t+\Delta} \mid X_t;\Theta)\right\}^m \middle| X_t = x_0\right] = 0$ for m=1,2.

(iii) The convergence of its first two moments to zero imply by Chebyshev's Inequality that the sequence $R_X^{(J)}(\Delta, X_{t+\Delta} \mid X_t;\Theta)$ converges to zero in probability, given $X_t=x_0$, that is:

$$\lim_{J\to\infty} \mathrm{Prob}\left(\left|R_X^{(J)}(\Delta, X_{t+\Delta} \mid X_t;\Theta)\right| > \varepsilon \,\middle|\, X_t = x_0;\theta_0\right) = 0.$$

Then, by Bayes' Rule we have

$$\mathrm{Prob}\left(\left|R_X^{(J)}(\Delta, X_{t+\Delta} \mid X_t;\Theta)\right| > \varepsilon;\theta_0\right)$$
$$= \int_{-\infty}^{+\infty} \mathrm{Prob}\left(\left|R_X^{(J)}(\Delta, X_{t+\Delta} \mid X_t;\Theta)\right| > \varepsilon \,\middle|\, X_t = x_0;\theta_0\right) \pi_t(x_0;\theta_0)dx_0$$

where $\pi_t(x_0;\theta_0) \equiv \partial \mathrm{Prob}(X_t \le x_0;\theta_0)/\partial x_0$ denotes the unconditional (or marginal) density of X_t at the true parameter value. Note that since we are not assuming that the process is strictly stationary, that density depends on t.

Now since

$$0 \le \text{Prob}\left(\left|R_X^{(J)}(\Delta, X_{t+\Delta} \mid X_t;\Theta)\right| > \varepsilon \middle| X_t = x_0;\theta_0\right) \le 1 \text{ and } \int_{-\infty}^{+\infty} \pi_t(x_0;\theta_0)dx_0 = 1$$

it follows from Lebesgue Dominated Convergence Theorem [see e.g., Haaser and Sullivan (1991, Theorem 6.8.6)] that

$$\lim_{J\to\infty} \text{Prob}\left(\left|R_X^{(J)}(\Delta, X_{t+\Delta} \mid X_t;\Theta)\right| > \varepsilon;\theta_0\right) = 0.$$

From $p_X^{(J)}(\Delta, X_{t+\Delta} \mid X_t;\theta) \xrightarrow{p} p_X(\Delta, X_{t+\Delta} \mid X_t;\theta)$ and the continuity of the logarithm, it follows that

$$\text{Ln}\left[p_X^{(J)}(\Delta, X_{t+\Delta} \mid X_t;\theta)\right] \xrightarrow{p} \text{Ln}\left[p_X(\Delta, X_{t+\Delta} \mid X_t;\theta)\right]$$

and therefore for fixed n: $\ell_n^{(J)}(\theta) \xrightarrow{p} \ell_n(\theta)$ as $J\to\infty$, uniformly in θ. Once we have reached this stage, the convergence of the respective argmax in $\hat{\theta}_n^{(J)} \xrightarrow{p} \hat{\theta}_n$ is an application of standard methods since $\ell_n^{(J)}(\theta)$ and $\ell_n(\theta)$ are both continuous in θ for all n and J.

(iv) Fix $\varepsilon>0$ and $\delta>0$. From part (ii), for each n, there exists $J_n>0$ such that for all $J\ge J_n$: $\text{Prob}\left(\left|\hat{\theta}_n^{(J)} - \hat{\theta}_n\right| > \varepsilon/2;\theta_0\right) \le e^{-n}$. Then there exists N>0, $e^{-N} < \delta/2$, such that for all $n\ge N$: $\text{Prob}\left(\left|\hat{\theta}_n - \theta_0\right| > \varepsilon/2;\theta_0\right) \le \delta/2$ since $\hat{\theta}_n \xrightarrow{p} \theta_0$. The conclusion follows from:

$$\text{Prob}\left(\left|\hat{\theta}_n^{(J_n)} - \theta_0\right| > \varepsilon;\theta_0\right) \le \text{Prob}\left(\left|\hat{\theta}_n^{(J_n)} - \hat{\theta}_n\right| > \varepsilon/2;\theta_0\right) + \text{Prob}\left(\left|\hat{\theta}_n - \theta_0\right| > \varepsilon/2;\theta_0\right).$$

and similarly for the convergence in distribution since

$$i_n^{1/2}(\theta_0)\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} G^{-1/2}(\theta_0) \times N(0, \text{Id})$$

by Lemma 5. In addition, it is clear that we can replace J_n by any $J_n' \ge J_n$ in the statements above with no modification. QED.

Proof of Lemma 6: Define the function $\varphi(s, y_0;\theta) \equiv E\left[f(Y_{t+s}) \middle| Y_t = y_0\right]$ for $s \in (0, \overline{\Delta})$. φ is differentiable in s (since the function $s \mapsto p_Y(s, y \mid y_0;\theta)$ is), and we have:

$$\begin{aligned}
\frac{\partial\varphi(s, y_0;\theta)}{\partial s} &= \int_{-\infty}^{+\infty} f(y) \frac{\partial p_Y(s, y \mid y_0;\theta)}{\partial s} dy \\
&= \int_{-\infty}^{+\infty} f(y) \left\{ \frac{\partial}{\partial y}\left(-\mu_Y(y;\theta) p_Y(s, y \mid y_0;\theta)\right) + \frac{1}{2}\frac{\partial^2 p_Y(s, y \mid y_0;\theta)}{\partial y^2} \right\} dy \\
&= \int_{-\infty}^{+\infty} \left\{ \mu_Y(y;\theta)\frac{df(y)}{dy} + \frac{1}{2}\frac{d^2 f(y)}{dy^2} \right\} p_Y(s, y \mid y_0;\theta) dy
\end{aligned}$$

where the integral converges due to the upper bound on p_Y in Lemma 2, the growth condition on μ_Y and the fact that f and its derivatives have at most exponential growth. The second equality follows from the Kolmogorov forward equation satisfied by p_Y and the third from integration by parts. When integrating by parts, we have

$$f(y)\mu_Y(y;\theta)p(s,y \mid y_0;\theta)\Big]_{-\infty}^{+\infty}=0$$

(and similarly for the other boundary terms) in light of the same bounds as above. Therefore $\partial\varphi(0,y_0;\theta)/\partial s = A(\theta)\bullet f(y_0)$ and similarly $\partial^j\varphi(0,y_0;\theta)/\partial s^j = A^j(\theta)\bullet f(y_0)$ for all j, by iterating the same steps, starting from:

$$\frac{\partial^2\varphi(s,y_0;\theta)}{\partial s^2}=\int_{-\infty}^{+\infty}\{A(\theta)\bullet f(y)\}\frac{\partial p_Y(s,y \mid y_0;\theta)}{\partial s}dy.$$

The result then follows from applying Taylor's Theorem at s=0, to the function φ evaluated at s=Δ. QED.

REFERENCES

AÏT-SAHALIA, Y., 1996a, Nonparametric Pricing of Interest Rate Derivative Securities, *Econometrica,* 64, 527-560.

AÏT-SAHALIA, Y., 1996b, Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies,* 9, 385-426.

AÏT-SAHALIA, Y., 1997, Do Interest Rates Really Follow Continuous-Time Markov Diffusions?, working paper, University of Chicago.

AMES, W.F., 1992, *Numerical Methods for Partial Differential Equations,* Third Edition, Academic Press, San Diego.

ANDERSEN, T.G. and J. LUND, 1996, The Short Rate Diffusion Revisited: An Investigation Guided by the Efficient Method of Moments, working paper, Northwestern University.

AZENCOTT, R., 1981, *Géodésiques et Diffusions en Temps Petit,* edited by R. Azencott, Société Mathématique de France, Paris, France.

BASAWA, I.V. and B.L.S. PRAKASA RAO, 1980, *Statistical Inference for Stochastic Processes,* Academic Press, London.

BASAWA, I.V. and D.J. SCOTT, 1983, *Asymptotic Optimal Inference for Non-ergodic Models,* Lecture Notes in Statistics 17, Springer Verlag, New York.

BIBBY, B.M. and M. SØRENSEN, 1995, Martingale Estimation Functions for Discretely Observed Diffusion Processes, *Bernoulli,* 1, 17-39.

BILLINGSLEY, P., 1961, *Statistical Inference for Markov Processes,* The University of Chicago Press, Chicago.

CONLEY, T.G., L.P. HANSEN, E.G.J. LUTTMER and J.A. SCHEINKMAN, 1997, Short-Term Interest Rates as Subordinated Diffusions, *Review of Financial Studies,* 10, 525-578.

COX, J.C., 1975-1996, The Constant Elasticity of Variance Option Pricing Model, *The Journal of Portfolio Management*, Special Issue, 1996.

COX, J.C. and S.A. ROSS, 1976, The Valuation of Options for Alternative Stochastic Processes, *Journal of Financial Economics*, 3, 145-166.

COX, J.C., J.E. INGERSOLL and S.A. ROSS, 1985, A Theory of the Term Structure of Interest Rates, *Econometrica*, 53, 385-407.

CRAMÉR, H., 1925, On Some Classes of Series Used in Mathematical Statistics, *Proceedings of the Sixth Scandinavian Mathematical Congress*, 399-425.

DACUNHA-CASTELLE, D. and D. FLORENS-ZMIROU, 1986, Estimation of the Coefficients of a Diffusion from Discrete Observations, *Stochastics*, 19, 263-284.

DUFFIE, D. and P. GLYNN, 1997, Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals, working paper, Stanford University.

DUFFIE, D. and K. SINGLETON, 1993, Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica,* 61, 929-952.

FELLER, W., 1952, The Parabolic Differential Equations and the Associated Semi-Groups of Transformations, *Annals of Mathematics,* 55, 468-519.

FLORENS, J.-P., E. RENAULT and N. TOUZI, 1995, Testing for Embeddability by Stationary Scalar Diffusions, forthcoming in the *Econometric Theory.*

GALLANT, A.R. and G. TAUCHEN, 1997, Estimation of Continuous Time Models for Stock Returns and Interest Rates, *Macroeconomic Dynamics,* 1, 135--168.

GALLANT, A.R. and J.R. LONG, 1997, Estimating Stochastic Differential Equations Efficiently by Minimum Chi-squared, *Biometrika,* 84, 125-141.

GOURIÉROUX, C., A. MONFORT and E. RENAULT, 1993, Indirect Inference, *Journal of Applied Econometrics,* 8, S85-S118.

GOURIÉROUX, C., A. MONFORT and A. TROGNON, 1984, Pseudo Maximum Likelihood Methods: Theory, *Econometrica,* 52, 681-700.

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

HALL, P. and C.C. HEYDE, 1980, *Martingale Limit Theory and its Applications,* Academic Press, San Diego.

HANSEN, L.P. and J.A. SCHEINKMAN, 1995, Back to the Future: Generating Moment Implications for Continuous Time Markov Processes, *Econometrica,* 63, 767-804.

HANSEN, L.P., J.A. SCHEINKMAN and N. TOUZI, 1995, Identification of Scalar Diffusions Using Eigenvectors, forthcoming in the *Journal of Econometrics.*

HONORÉ, P., 1997, Maximum-Likelihood Estimation of Non-Linear Continuous-Time Term Structure Models, working paper, Aarhus University.

HUGGINS, D.J., 1997, Estimation of a Diffusion Process for the U.S. Short Interest Rate Using a Semigroup Pseudo-Likelihood, Ph.D. Dissertation, University of Chicago.

JARROW, R. and A. RUDD, 1982, Approximate Option Valuation for Arbitrary Stochastic Processes, *Journal of Financial Economics,* 10, 347-349.

KARLIN, S. and H.M. TAYLOR, 1981, *A Second Course in Stochastic Processes*, New York: Academic Press.

LO, A.W., 1988, Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data, *Econometric Theory,* 4, 231-247.

MCKEAN, H.P. JR., 1956, Elementary Solutions for Certain Parabolic Partial Differential Equations, *Transactions of the American Mathematical Society,* 82, 519-548.

MELINO, A., 1994, Estimation of Continuous-Time Models in Finance, in C. Sims (ed.), *Advances in Econometrics, Sixth World Congress, Volume II,* Cambridge University Press, Cambridge, England.

MERTON, R.C., 1980, On Estimating the Expected Return on the Market: An Exploratory Investigation, *Journal of Financial Economics,* 8, 323-361.

PEDERSEN, A.R., 1995, A New Approach to Maximum-Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations, *Scandinavian Journal of Statistics*, 22, 55-71.

ROGERS, L.C.G., 1985, Smooth Transition Densities for One-Dimensional Diffusions, *Bulletin of the London Mathematical Society*, 17, 157-161.

SANSONE, G., 1991, *Orthogonal Functions*, New York: Dover.

SANTA-CLARA, P., 1995, Simulated Likelihood Estimation of Diffusion with an Application to the Short Term Interest Rate, working paper, UCLA.

SØRENSEN, M., 1991, Likelihood Methods for Diffusions with Jumps, in Prabhu, N.U. and I.V. Basawa (eds), *Statistical Inference in Stochastic Processes,* Marcel Dekker, New York, 67-105.

STANTON, R., 1997, A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk, forthcoming in the *Journal of Finance.*

STONE, M.H., 1928, Development in Hermite Polynomials, *Annals of Mathematics*, 29, 1-13.

TAUCHEN, G., 1997, New Minimum Chi-Square Methods in Empirical Finance, in D. Kreps and K. Wallis (eds.), *Advances in Econometrics: Seventh World Congress,* Cambridge University Press, Cambridge, UK, 279-317.

TAUCHEN, G., 1998, The Objective Function of Simulation Estimators Near the Boundary of the Unstable Region of the Parameter Space, forthcoming in the *Review of Economics and Statistics*.

VASICEK, O., 1977, An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, 5, 177-188.

WHITE, H., 1982, Maximum Likelihood Estimation of Misspecified Models, *Econometrica,* 50, 1-25.

WONG, E., 1964, The Construction of a Class of Stationary Markov Processes, in *Stochastic Processes in Mathematical Physics and Engineering,* Proceedings of Symposia in Applied Mathematics, 16, ed. by R. Bellman, American Mathematical Society, Providence, RI, 264-276.

Figure 1

Conditional Density Approximations for the Vasicek Model
Semi-Annual Sampling Frequency

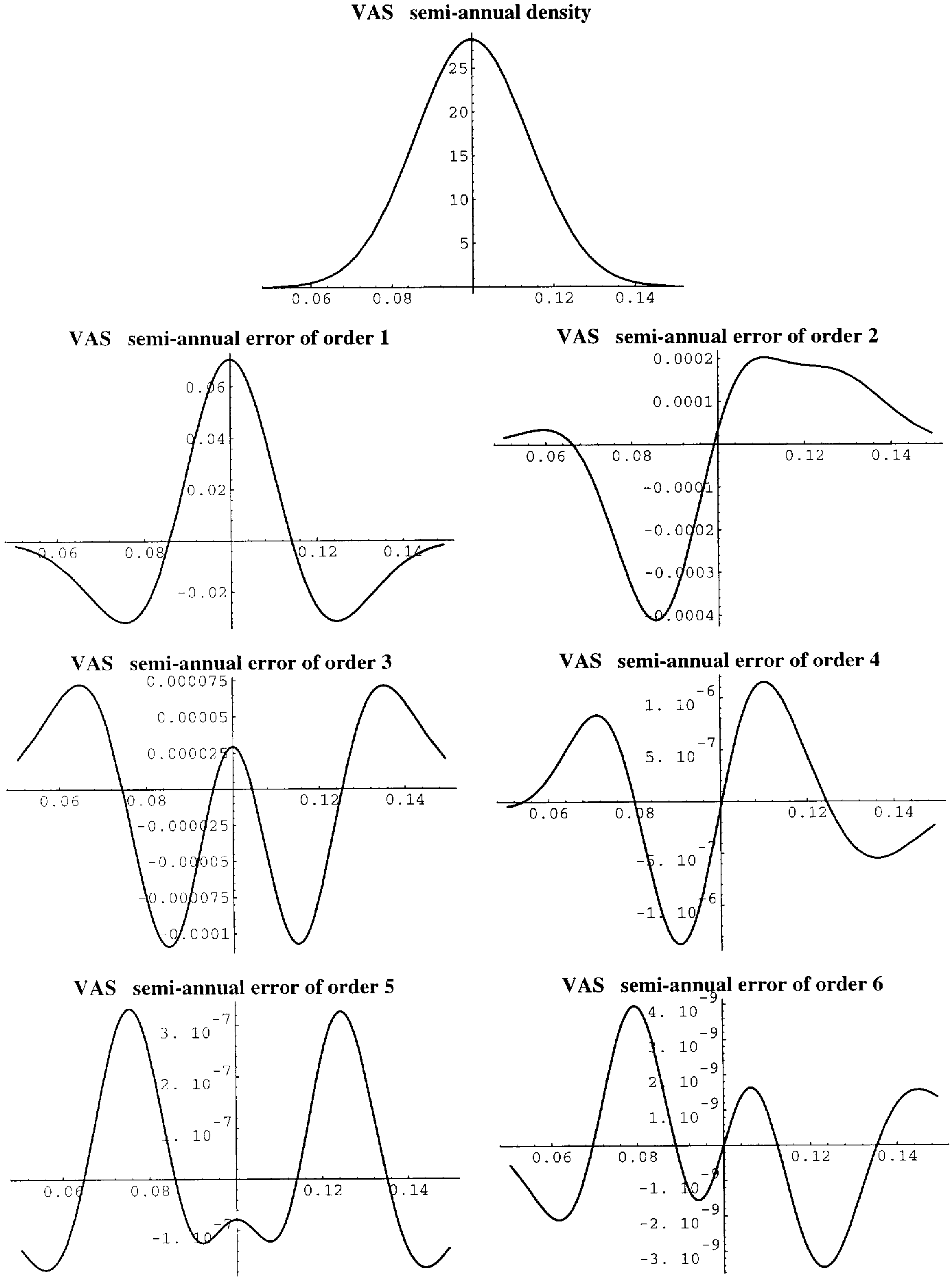

Figure 2

Conditional Density Approximations for the Vasicek Model
Monthly Sampling Frequency

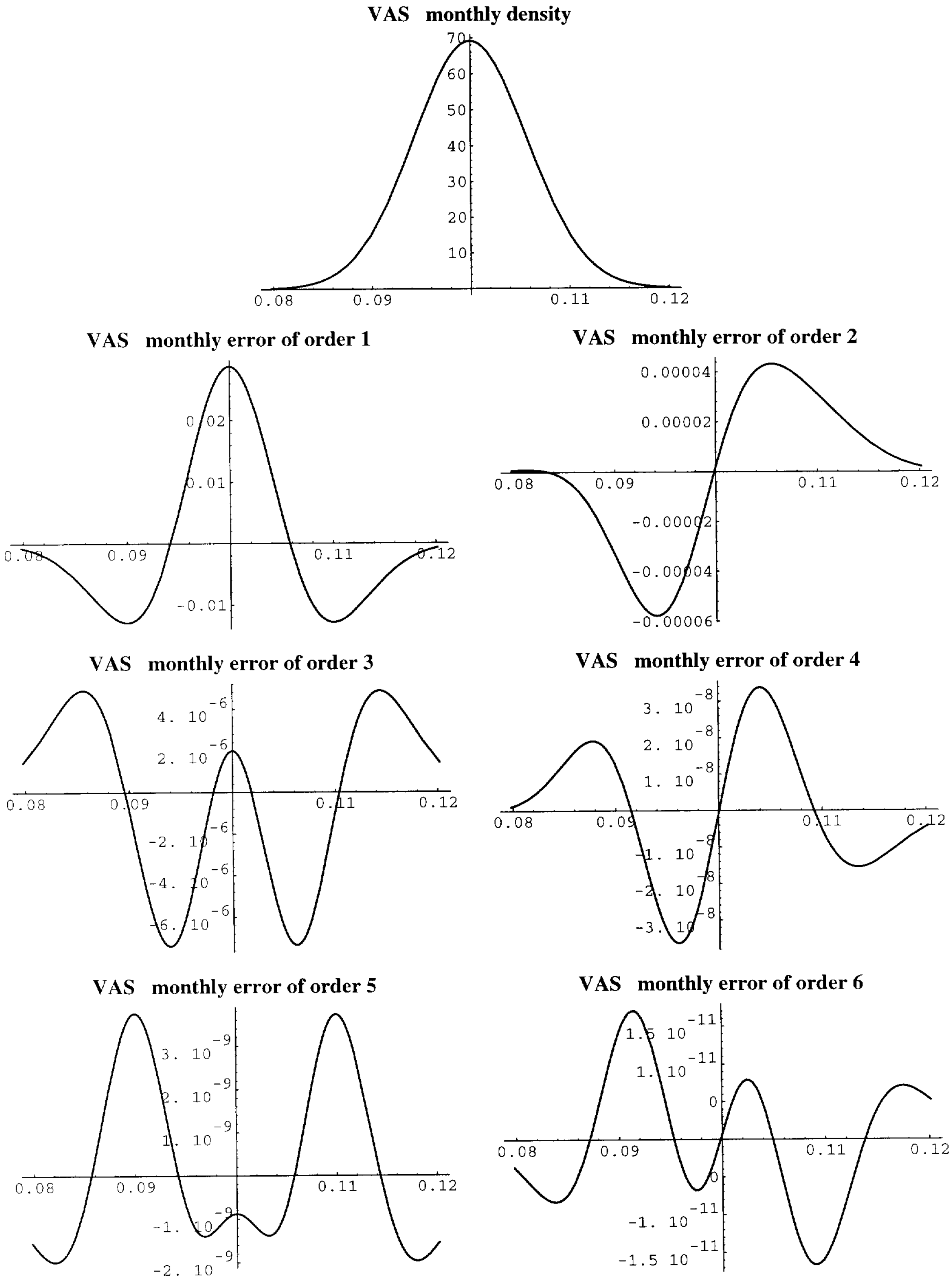

Figure 3

Conditional Density Approximations for the Cox-Ingersoll-Ross Model
Semi-Annual Sampling Frequency

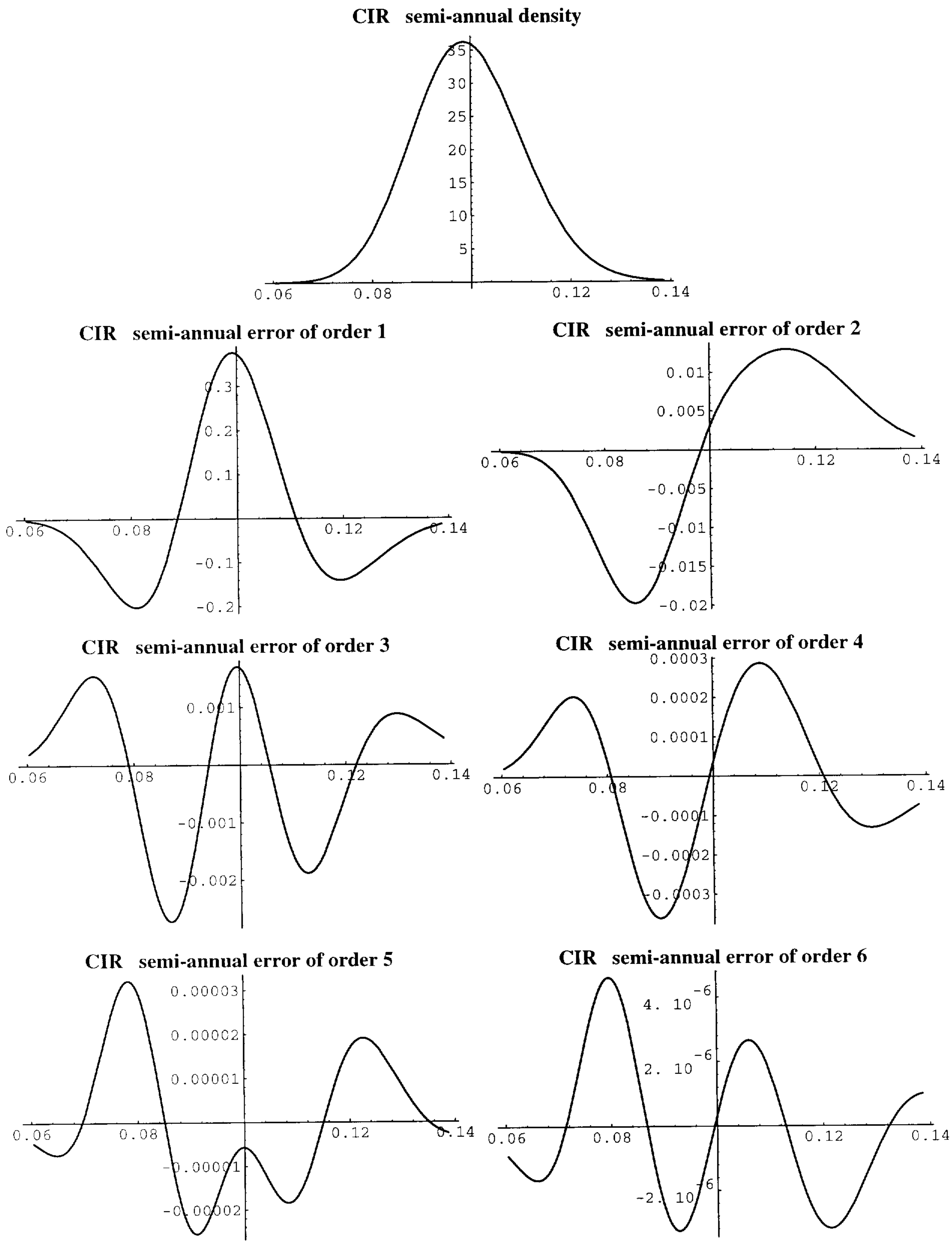

Figure 4

Conditional Density Approximations for the Cox-Ingersoll-Ross Model Monthly Sampling Frequency