Econometrica, Vol. 70, No. 1 (January, 2002), 223–262

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

BY YACINE AÏT-SAHALIA[1]

When a continuous-time diffusion is observed only at discrete dates, in most cases the transition distribution and hence the likelihood function of the observations is not explicitly computable. Using Hermite polynomials, I construct an explicit sequence of closed-form functions and show that it converges to the true (but unknown) likelihood function. I document that the approximation is very accurate and prove that maximizing the sequence results in an estimator that converges to the true maximum likelihood estimator and shares its asymptotic properties. Monte Carlo evidence reveals that this method outperforms other approximation schemes in situations relevant for financial models.

KEYWORDS: Maximum-likelihood estimation, continuous-time diffusion, discrete sampling, transition density, Hermite expansion.

1. INTRODUCTION

CONSIDER A CONTINUOUS-TIME parametric diffusion

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

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

While the model is written in continuous time, the available data are always sampled discretely in time. Ignoring the difference can result in inconsistent estimators (see, e.g., Merton (1980) and Melino (1994)). A number of econometric methods have been recently developed to estimate the parameters of (1.1), without requiring that a continuous record of observations be available. Some of these methods are based on simulations (Gouriéroux, Monfort, and Renault (1993), Gallant and Tauchen (1996)), others on the generalized method of moments (Hansen and Scheinkman (1995), Duffie and Glynn (1997), Kessler and Sorensen

[1] I am grateful to David Bates, René Carmona, Freddy Delbaen, Ron Gallant, Lars Hansen, Bjarke Jensen, Per Mykland, Peter C. B. Phillips, Rolf Poulsen, Peter Robinson, Chris Rogers, Angel Serrat, Chris Sims, George Tauchen, and in particular a co-editor and three anonymous referees for very helpful comments and suggestions. Robert Kimmel and Ernst Schaumburg provided excellent research assistance. This research was supported by an Alfred P. Sloan Research Fellowship and by the NSF under Grant SBR-9996023. *Mathematica* code to calculate the closed-form density sequence can be found at http://www.princeton.edu/~yacine.

(1999)), nonparametric density-matching (Aït-Sahalia (1996a, 1996b)), nonparametric regression for approximate moments (Stanton (1997)), or are Bayesian (Eraker (1997) and Jones (1997)).

As in most contexts, provided one trusts the parametric specification (1.1), maximum-likelihood is the method of choice. The major caveat in the present context is that the likelihood function for discrete observations generated by (1.1) cannot be determined explicitly for most models. Let $p_X(\Delta, x|x_0; \theta)$ 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. Assume that we observe the process at dates $\{t = i\Delta | i = 0, \ldots, n\}$, where $\Delta > 0$ is fixed.[2] Bayes' rule combined with the Markovian nature of (1.1), which the discrete data inherit, imply that the log-likelihood function has the simple form

$$\ell_n(\theta) \equiv \sum_{i=1}^{n} \ln\{p_X(\Delta, X_{i\Delta} \mid X_{(i-1)\Delta}; \theta)\}. \tag{1.2}$$

For some of the rare exceptions where p_X is available in closed-form, see Wong (1964); in finance, the models of Black and Scholes (1973), Vasicek (1977), Cox, Ingersoll, and Ross (1985), and Cox (1975) all rely on the known closed-form expressions.

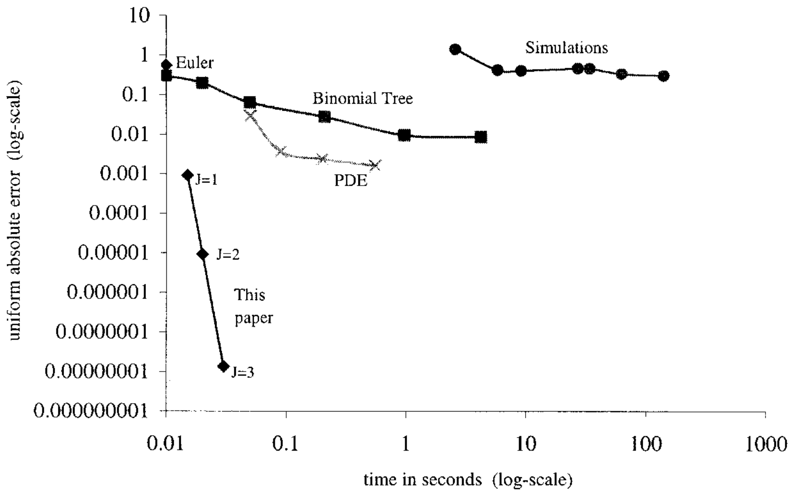
If sampling of the process were continuous, the situation would be simpler. First, the likelihood function for a continuous record can be obtained by means of a classical absolutely continuous change of measure (see, e.g., Basawa and Prakasa Rao (1980)).[3] Second, when the sampling interval goes to zero, expansions for the transition function "in small time" are available in the statistical literature (see, e.g., Azencott (1981)). Dacunha-Castelle and Florens-Zmirou (1986) calculate expressions for the transition function in terms of functionals of a Brownian Bridge. With discrete-time sampling, the available methods to compute the likelihood function involve either solving numerically the Fokker-Planck-Kolmogorov partial differential equation (see, e.g., Lo (1988)) or simulating a large number of sample paths along which the process is sampled very finely (see Pedersen (1995) and Santa-Clara (1995)). 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 a sum over the outcome of the simulations.

By contrast, I construct a *closed-form* sequence $p_X^{(J)}$ of approximations to the transition density, hence from (1.2) a sequence $\ell_n^{(J)}$ of approximations to the log-likelihood function ℓ_n. I also provide empirical evidence that $J = 2$ or 3 is amply adequate for models that are relevant in finance.[4] Since the expression

[2] See Section 3.1 for extensions to the cases where the sampling interval Δ is time-varying and even possibly random.

[3] Note that the continuous-observation likelihood is only defined if the diffusion function σ is known.

[4] In addition, Jensen and Poulsen (1999) have recently completed a comparison of the method of this paper against four alternatives: a discrete Euler approximation of the continuous-time model



Notes: This figure reports the average uniform absolute error of various density approximation techniques applied to the Vasicek, Cox-Ingersoll-Ross and Black-Scholes models. "Euler" refers to the discrete-time, continuous-state, first-order Gaussian approximation scheme for the transition density given in equation (5.4); "Binomial Tree" refers to the discrete-time, discrete-state (two) approximation; "Simulations" refers to an implementation of Pedersen (1995)'s simulated-likelihood method; "PDE" refers to the numerical solution of the Fokker-Planck-Kolmogorov partial differential equation satisfied by the transition density, using the Crank-Nicolson algorithm. For implementation details on the different methods considered, see Jensen and Poulsen (1999).

FIGURE 1.—Accuracy and speed of different approximation methods for p_X.

$\ell_n^{(J)}$ to be maximized is explicit, the effort involved is minimal, identical to a standard maximum-likelihood problem with a known likelihood function. Examples are contained in a companion paper (Aït-Sahalia (1999)), which provides, for different models, the corresponding expression of $p_X^{(J)}$. Besides making maximum-likelihood estimation feasible, these closed-form approximations have other applications in financial econometrics. For instance, they could be used for derivative pricing, for indirect inference (see Gouriéroux, Monfort, and Renault (1993)), which in its simplest version uses an Euler approximation as instrumental model, or for Bayesian inference—basically whenever an expression for the transition density is required.

The paper is organized as follows. Section 2 describes the sequence of density approximations and proves its convergence. Section 3 studies the properties of the resulting maximum-likelihood estimator. In Section 4, I show how to calculate in closed-form the coefficients of the approximation and readers primarily interested in applying these results to a specific model can go there directly.

(1.1), a binomial tree approximation, the numerical solution of the PDE, and simulation-based methods, all in the context of various specifications and parameter values that are relevant for interest rate and stock return models. To give an idea of the relative accuracy and speed of these approximations, Figure 1 summarizes their main results. As is clear from the figure, the approximation of the transition function derived here provides a degree of accuracy and speed that is unmatched by any of the other methods.

Section 5 gives the results of Monte Carlo simulations. Section 6 concludes. All proofs are in the Appendix.

2. A SEQUENCE OF EXPANSIONS OF THE TRANSITION FUNCTION

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

But in what sense does such an expansion converge? In the CLT case, the convergence is understood to mean that the series with a *fixed* number of corrective terms (i.e., fixed J) converges when the sample size n goes to infinity. In fact, for a fixed n, 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. I will make this statement precise later, using the criterion of Cramér (1925): the density $p(z)$ to be expanded around a $N(0, 1)$ must have tails sufficiently thin for $\exp(z^2/2)p'(z)^2$ to be integrable.

The point however is that the density p_X cannot in general be approximated for fixed Δ around a Normal density, because the distribution of the diffusion 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, that tail is of order $x^{-1}\exp(-\ln^2(x))$ as x tends to $+\infty$. Similarly, the expansion of any $N(0, v)$ density around a $N(0, 1)$ diverges if $v > 2$, and hence the class of transition densities p_X for which straight Hermite expansions converge in the sense of adding more terms (J increases with Δ fixed) is quite limited.

To obtain nevertheless an expansion that converges as more correction terms are added while Δ remains fixed, I will show that the transformation of the diffusion process X into Z in fact guarantees (unlike the CLT situation) that the resulting variable Z has a density p_Z that belongs to the class of densities for which the Hermite series converges as more polynomial terms are added. I will then construct a convergent Hermite series for p_Z. Since Z is a known transformation of X, I will be able to revert the transformation from X to Z and by the Jacobian formula obtain an expansion for the density of X. As a result of transforming Z back into X, which in general is a nonlinear transformation (unless $\sigma(x;\theta)$ is independent of the state variable x), the leading term of the expansion for the density p_X will be a *deformed*, or stretched, Normal density rather than the $N(0,1)$ leading term of the expansion for p_Z. The rest of this section makes this basic intuition rigorous. In particular, Theorem 1 will prove that such an expansion converges uniformly to the unknown p_X.

2.1. *Assumptions and First Transformation $X \to Y$*

I start by making fairly general assumptions on the functions μ and σ. In particular, I do not assume that μ and σ satisfy the typical growth conditions at infinity, nor do I restrict attention to stationary diffusions only. Let $D_X = (\underline{x}, \bar{x})$ denote the domain of the diffusion X. I will consider the two cases where $D_X = (-\infty, +\infty)$ and $D_X = (0, +\infty)$. The latter case is often 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 $\lim_{x\to 0^+} \sigma(x;\theta) = 0$ and μ and/or σ violate the linear growth conditions near the boundaries. For these reasons, I will devise a set of assumptions where growth conditions (without constraint on the sign of the drift function near the boundaries) are replaced by assumptions on the sign of the drift near the boundaries (without restriction on the growth of the coefficients). The assumptions are:

ASSUMPTION 1 (Smoothness of the Coefficients): *The functions $\mu(x;\theta)$ and $\sigma(x;\theta)$ are infinitely differentiable in x, and three times continuously differentiable in θ, for all $x \in D_X$ and $\theta \in \Theta$.*

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 $D_X = (0, +\infty)$, $\lim_{x\to 0^+} \sigma(x;\theta) = 0$ is possible, but then there exist constants $\xi_0 > 0$, $\omega > 0$, $\rho \geq 0$ such that $\sigma(x;\theta) \geq \omega x^\rho$ for all $0 < x \leq \xi_0$ and $\theta \in \Theta$. Whether or not $\lim_{x\to 0^+} \sigma(x;\theta) = 0$, σ is a nondegenerate on $(0, +\infty)$, 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 I employ 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[5]

$$(2.1)\qquad Y \equiv \gamma(X;\theta) = \int^X du/\sigma(u;\theta)$$

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 for all $\theta \in \Theta$. 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 $Y = (1-\rho)X^{1-\rho}$ if $0 < \rho < 1$ (so $D_Y = (0, +\infty)$), $Y = \ln(X)$ if $\rho = 1$ (so $D_Y = (-\infty, +\infty)$) and $Y = -(\rho-1)X^{-(\rho-1)}$ if $\rho > 1$ (so $D_Y = (-\infty, 0)$). For a given model under consideration, assume that the parameter space Θ is restricted in such a way that D_Y is 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, that is

$$(2.2)\qquad dY_t = \mu_Y(Y_t;\theta)dt + dW_t, \qquad \text{where}$$
$$\mu_Y(y;\theta) = \frac{\mu(\gamma^{-1}(y;\theta);\theta)}{\sigma(\gamma^{-1}(y;\theta);\theta)} - \frac{1}{2}\frac{\partial\sigma}{\partial x}(\gamma^{-1}(y;\theta);\theta).$$

ASSUMPTION 3 (Boundary Behavior): *For all* $\theta \in \Theta$, $\mu_Y(y;\theta)$ *and its derivatives with respect to* y *and* θ *have at most polynomial growth*[6] *near the boundaries and* $\lim_{y\to\underline{y}^+ \text{ or } \bar{y}^-} \lambda_Y(y;\theta) < +\infty$ *where* $-\lambda_Y$ *is the potential, i.e.,* $\lambda_Y(y;\theta) \equiv -\{\mu_Y^2(y;\theta) + \partial\mu_Y(y;\theta)/\partial y\}/2$.

1. *Left Boundary*: *If* $\underline{y} = 0$, *there exist constants* $\varepsilon_0, \kappa, \alpha$ *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* $\kappa \geq 1$. *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*: *If* $\bar{y} = +\infty$, *there exist constants* $E_0 > 0$ *and* $K > 0$ *such that for all* $y \geq E_0$ *and* $\theta \in \Theta$, $\mu_Y(y;\theta) \leq Ky$. *If* $\bar{y} = 0$, *there exist constants* $\varepsilon_0, \kappa, \alpha$ *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* $\kappa \geq 1/2$.

Note that λ_Y is not restricted from going to $-\infty$ near the boundaries. Assumption 3 is formulated in terms of the function μ_Y for reasons of convenience, but the restriction it imposes on the original functions μ and σ follows from (2.1). Assumption 3 only restricts how large μ_Y can grow if it has the "wrong" sign, meaning that μ_Y is positive near $\bar{y}$ and negative near $\underline{y}$: then linear growth is the maximum possible rate. But if μ_Y has the "right" sign, the process is being pulled

[5] The same transformation, sometimes referred to as the Lamperti transform, has been used, for instance, by Florens (1999).

[6] Define an infinitely differentiable function f as having 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 $p = 1$, f is said to have at most linear growth, and if $p = 2$ at most quadratic growth. Near 0, polynomial growth means that $|y|^{+p}|f(y)|$ is bounded.

back away from the boundaries and I do not restrict how fast mean-reversion occurs (up to an arbitrary large polynomial rate for technical reasons). The constraints on the behavior of the function μ_Y are essentially the best possible for the following reasons. If μ_Y has the wrong sign near an infinity boundary, and grows faster than linearly, then Y explodes (i.e., can reach the infinity boundary) in finite time. Near a zero boundary, say $\underline{y}=0$, if there exist $\kappa>0$ and $\alpha<1$ such that $\mu_Y(y;\theta)\leq ky^{-\alpha}$ in a neighborhood of 0^+, then 0 becomes attainable. The behavior of the diffusion Y implied by the assumptions made is fully characterized by the following proposition, where $T_Y\equiv\inf\{t\geq 0 \mid Y_t\notin D_Y=(\underline{y},\bar{y})\}$ denotes the exit time from D_Y:

PROPOSITION 1: *Under Assumptions* 1–3, (2.2) *admits a weak solution* $\{Y_t \mid t\geq 0\}$, *unique in probability law, for every distribution of its initial value* Y_0.[7] *The boundaries of* D_Y *are unattainable, in the sense that* $\text{Prob}(T_Y=\infty)=1$. *Finally, if* $+\infty$ *is a right 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 left 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 right or left*), *then it is entrance*.[8]

Note also that 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 common unconditional (marginal) density for all Y_t

$$\pi_Y(y;\theta)\equiv\exp\left\{2\int^{y}\mu_Y(u;\theta)\,du\right\}\Big/\int_{\underline{y}}^{\bar{y}}\exp\left\{2\int^{v}\mu_Y(u;\theta)\,du\right\}dv, \tag{2.3}$$

provided that the initial random variable Y_0 is itself distributed with density (2.3) (see, e.g., Karlin and Taylor (1981)). When at least one of the boundaries is natural, stationarity is neither precluded nor implied in that the (only) possible candidate for stationary density, π_Y, may or may not be integrable near

[7] A weak solution to (2.2) in the interval D_Y is a pair (Y,W), a probability space and a filtration, such that (Y,W) satisfies the stochastic integral equation that underlies the stochastic differential equation (2.2). For a formal definition, see, e.g., Karatzas and Shreve (1991, Definition 5.5.20). Uniqueness in law means that two solutions would have identical finite-dimensional distributions, i.e., in particular the same observable implications for *any* discrete-time data. From the perspective of statistical inference from discrete observations, this is therefore the appropriate concept of uniqueness.

[8] Natural boundaries can neither be reached in finite time, nor can the diffusion be started or escape 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 considerations require the boundaries to be unattainable; however, they say 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.

the boundaries.[9] Next, I show that the diffusion Y admits a smooth transition density:

PROPOSITION 2: *Under Assumptions* 1–3, *Y admits a transition density $p_Y(\Delta, y \mid y_0; \theta)$ that is continuously differentiable in $\Delta > 0$, infinitely differentiable in $y \in D_Y$ and $y_0 \in D_Y$, and three times continuously differentiable in $\theta \in \Theta$. Furthermore, there exists $\overline{\Delta} > 0$ such that for every $\Delta \in (0, \overline{\Delta})$, there exist positive constants C_i, $i = 0, \ldots, 4$, and D_0 such that for every $\theta \in \Theta$ and $(y, y_0) \in D_Y^2$*:

$$(2.4)\qquad 0 < p_Y(\Delta, y \mid y_0; \theta) \leq C_0 \Delta^{-1/2} e^{-3(y-y_0)^2/8\Delta} e^{C_1|y-y_0||y_0| + C_2|y-y_0| + C_3|y_0| + C_4 y_0^2},$$

$$(2.5)\qquad \left|\partial p_Y(\Delta, y \mid y_0; \theta)/\partial y\right| \leq D_0 \Delta^{-1/2} e^{-3(y-y_0)^2/8\Delta} P(|y|, |y_0|) e^{C_1|y-y_0||y_0| + C_2|y-y_0| + C_3|y_0| + C_4 y_0^2},$$

where P is a polynomial of finite order in $(|y|, |y_0|)$, with coefficients uniformly bounded in $\theta \in \Theta$. Finally, if $\mu_Y \leq 0$ near the right boundary $+\infty$ and $\mu_Y \geq 0$ near the left boundary (*either* 0 *or* $-\infty$), *then $\overline{\Delta} = +\infty$.*

The next result shows that these properties essentially extend to the diffusion X of original interest.

COROLLARY 1: *Under Assumptions* 1–3, (1.1) *admits a weak solution $\{X_t \mid t \geq 0\}$, unique in probability law, for every distribution of its initial value X_0. The boundaries of D_X are unattainable, in the sense that* $\text{Prob}(T_X = \infty) = 1$ *where $T_X \equiv \inf\{t \geq 0 \mid X_t \notin D_X\}$. In addition, X admits a transition density $p_X(\Delta, x|x_0; \theta)$ which is continuously differentiable in $\Delta > 0$, infinitely differentiable in $x \in D_X$ and $x_0 \in D_X$, and three times continuously differentiable in $\theta \in \Theta$.*

2.2. *Second Transformation $Y \to Z$*

The bound (2.4) implies that the tails of p_Y have a Gaussian-like upper bound. In light of the discussion at the beginning of Section 2 about the requirements for convergence of a Hermite series, this is a big step forward. However, while Y, thanks to its unit diffusion ($\sigma_Y = 1$), is "closer" to a Normal variable than X is, it is not practical to expand p_Y. This is due to the fact that p_Y gets peaked around the conditional value y_0 when Δ gets small. And a Dirac mass is not a particularly appealing leading term for an expansion. For that reason, I perform a further transformation. For given $\Delta > 0$, $\theta \in \Theta$, and $y_0 \in R$, define the "pseudo-normalized" increment of Y as

$$(2.6)\qquad Z \equiv \Delta^{-1/2}(Y - y_0).$$

[9] 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 (null recurrent) is not.

Of course, since I do not require that $\Delta \to 0$, I make no claim regarding the degree of accuracy of this standardization device, hence the term "pseudo." However, I will show that for fixed Δ, Z defined in (2.6) happens to be close enough to a $N(0,1)$ variable to make it possible to create a convergent series of expansions for its density p_Z around a $N(0,1)$. In other words, Z turns out to be the appropriate transformation of X if we are going to start an expansion with a $N(0,1)$ term. Expansions starting with a different leading term could be considered (with matching orthogonal polynomials) but, should Δ in fact be small, they would have the drawback of starting with an inadequate leading term and therefore requiring additional correction.[10]

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

$$p_Z(\Delta, z|y_0;\theta) \equiv \Delta^{1/2} p_Y(\Delta, \Delta^{1/2}z + y_0|y_0;\theta). \tag{2.7}$$

Once I have obtained a sequence of approximations to the function $(z, y_0) \mapsto p_Z(\Delta, z|y_0;\theta)$, I will backtrack and infer a sequence of approximations to the function $(y, y_0) \mapsto p_Y(\Delta, y|y_0;\theta)$ by inverting (2.7):

$$p_Y(\Delta, y|y_0;\theta) \equiv \Delta^{-1/2} p_Z(\Delta, \Delta^{-1/2}(y-y_0)|y_0;\theta), \tag{2.8}$$

and then back to the object of interest $(x, x_0) \mapsto p_X(\Delta, x|x_0;\theta)$, by applying again the Jacobian formula for the change of density:

$$p_X(\Delta, x|x_0;\theta) = \sigma(x;\theta)^{-1} \times p_Y(\Delta, \gamma(x;\theta)|\gamma(x_0;\theta);\theta). \tag{2.9}$$

2.3. *Approximation of the Transition Function of the Transformed Data*

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

$$H_j(z) \equiv e^{z^2/2} \frac{d^j}{dz^j}\left[e^{-z^2/2}\right], \quad j \geq 0, \tag{2.10}$$

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

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

as the Hermite expansion of the density function $z \mapsto p_Z(\Delta, z|y_0;\theta)$ (for fixed Δ, y_0, and θ).[11] By orthonormality of the Hermite polynomials, divided by $\sqrt{j!}$,

[10] This is because the limiting form of the density for a diffusion, which is driven by a Brownian motion, is Gaussian. However a different leading term would be appropriate for processes of a different kind (for example driven by a non-Brownian Lévy process).

[11] Hence the boundary behavior of the transition density approximation is designed to match that of the true density as the forward variable (not the backward variable) nears the boundaries of the support: under the assumptions made, $p_Z \to 0$ near the boundaries.

with respect to the $L^2(\phi)$ scalar product weighted by the Normal density, the coefficients $\eta_Z^{(j)}$ are given by

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

Section 4 will indicate how to approximate these coefficients in closed-form, yielding a fully explicit sequence of approximations to p_Z.

By analogy with (2.8), I can then form the sequence of approximations to p_Y as

$$(2.13) \qquad p_Y^{(J)}(\Delta, y|y_0; \theta) \equiv \Delta^{-1/2} p_Z^{(J)}(\Delta, \Delta^{-1/2}(y - y_0)|y_0; \theta)$$

and finally approximate p_X by mimicking (2.9), i.e.,

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

The following theorem proves that the expansion (2.14) converges uniformly as more terms are added, and that the limit is indeed the true (but unknown) density function p_X.

THEOREM 1: *Under Assumptions* 1–3, *there exists* $\overline{\Delta} > 0$ (*given in Proposition* 1) *such that for every* $\Delta \in (0, \overline{\Delta})$, $\theta \in \Theta$ *and* $(x, x_0) \in D_X^2$:

$$(2.15) \qquad p_X^{(J)}(\Delta, x|x_0; \theta) \underset{J\to\infty}{\longrightarrow} p_X(\Delta, x|x_0; \theta).$$

In addition, the convergence is uniform in θ *over* Θ *and in* x_0 *over compact subsets of* D_X. *If* $\sigma(x; \theta) > c > 0$ *on* D_X, *then the convergence is further uniform in* x *over the entire domain* D_X. *If* $D_X = (0, +\infty)$ *and* $\lim_{x\to 0^+} \sigma(x; \theta) = 0$, *then the convergence is uniform in* x *in each interval of the form* $[\varepsilon, +\infty)$, $\varepsilon > 0$.

3. A SEQUENCE OF APPROXIMATIONS TO THE MAXIMUM-LIKELIHOOD ESTIMATOR

I now study the properties of the sequence of maximum-likelihood estimators $\hat{\theta}_n^{(J)}$ derived from maximizing over θ in Θ the approximate likelihood function computed from $p_X^{(J)}$, i.e., (1.2) with p_X replaced by $p_X^{(J)}$.[12] I will show that $\hat{\theta}_n^{(J)}$ converges as $J \to \infty$ to the true (but uncomputable in practice) maximum-likelihood estimator $\hat{\theta}_n$. I further prove that when the sample size gets larger $(n \to \infty)$, one can find $J_n \to \infty$ such that $\hat{\theta}_n^{(J_n)}$ converges to the true parameter value θ_0.[13]

[12] The roots of the Hermite polynomials are such that $p_X^{(J)} > 0$ on an interval $[-c_J; c_J]$ with $c_J \to \infty$ as $J \to \infty$. Let a_J be a positive sequence converging to 0 as $J \to \infty$. Define ω_J as a (smooth) version of the trimming index taking value 1 if $p_X^{(J)} > a_J$ and a_J otherwise. Before taking the logarithm, replace $p_X^{(J)}$ by $\omega_J p_X^{(J)}$. It is shown in the Appendix that such trimming is asymptotically irrelevant.

[13] This setup is different from either the psuedo-maximum likelihood one (see White (1982) and Gouriéroux, Monfort, and Trognon (1984)), or the semi-nonparametric case (Gallant and Nychka

3.1. *Likelihood Function: Initial Observation and Random Sampling Extension*

When defining the log-likelihood function in (1.2), I ignored the unconditional density of the first observation, $\ln(\pi_0(X_0;\theta))$, because it is dominated by the sum of the conditional density terms $\ln\{p_X(\Delta, X_{i\Delta} \mid X_{(i-1)\Delta};\theta)\}$ as $n \to \infty$. The sample contains only one observation on the unconditional density π and n on the transition function, so that the information on π_0 contained in the sample does not increase with n. All the distributional properties below will be asymptotic, so the definition (1.2) is appropriate for the log-likelihood function (see Billingsley (1961)). In any case, re-introducing the term $\ln(\pi_0(X_0;\theta))$ back into the log-likelihood poses no difficulty.

Note also that I have assumed for convenience that Δ is identical across pairs of successive observations. If instead Δ varies deterministically, say Δ_i is the time interval between the $(i-1)$th and ith observations, then it is clear from (1.2) that it suffices to replace Δ by its actual value Δ_i when evaluating the transition density for the ith pair of observations. If the sampling interval is random, then one can write down the joint likelihood function of the pair of observations and Δ_i and utilize Bayes' Rule to express it as the product of the conditional density of the ith observation $\widetilde{X}_i$ given the $(i-1)$th *and* Δ_i, times the marginal density q of Δ_i: that is $p_X(\Delta_i, \widetilde{X}_i \mid \widetilde{X}_{i-1};\theta) \times q(\Delta_j;\kappa)$ where κ is a parameter vector parameterizing the sampling density.[14] If the sampling process is independent of X and θ, then the marginal density is irrelevant for the likelihood maximization and the conditional density is the same function p_X as before, evaluated at the realization Δ_i. Hence for the purpose of estimating θ, the criterion function (1.2) is unchanged and as in the deterministic case it suffices to replace Δ by the realization Δ_i corresponding to the time interval between the $(i-1)$th and ith observations. By contrast, when the sampling interval is random and informative about the parameters of the underlying process (for example, if more rapid arrivals of trade signal an increase of price volatility), then the joint density cannot be integrated out as simply. I now return to the base case of fixed sampling at interval Δ.

3.2. *Properties of the Maximum-Likelihood Estimator*

To analyze the properties of the estimators $\hat{\theta}_n$ and $\hat{\theta}_n^{(J)}$, I introduce the following notation. Define the $K \times K$ identity matrix as Id and $L_i(\theta) \equiv$

(1987)). We are in a somewhat atypical situation in the sense that the psuedo-likelihood does approximate the true likelihood function, and we wish to exploit this fact. In particular, the choice of J is independent of n and J can always be chosen sufficiently large to make the resulting estimator arbitrarily close to the true MLE. This paper is not concerned with the potential misspecification of the true likelihood function, i.e., it accepts (1.1) as the data generating process, but then does not require that the densities belong to specific classes such as the linear exponential family.

[14] To insure that Theorem 1 remains applicable when Δ is not constant, assume that the distribution of Δ has a support contained in an interval of the form $(\underline{\delta}, \bar{\delta})$ where $0 < \underline{\delta} < \bar{\delta} < \overline{\Delta}$. In this case, the convergence in Theorem 1 is uniform in Δ.

$\ln(p_X(\Delta, X_{i\Delta} \mid X_{(i-1)\Delta}; \theta))$. $\dot{L}_i(\theta)$ (and additional dots) denotes differentiation with respect to θ, and ^{T} denotes transposition. The score vector $V_n(\theta) \equiv \sum_{i=1}^n \dot{L}_i(\theta)$ is a martingale. Corollary 1 proved that p_X admits three continuous derivatives with respect to θ in Θ; the same holds for $p_X^{(J)}$ by direct inspection of its expression given in Section 2.3. Next define

$$i_n(\theta) \equiv \sum_{i=1}^n E_\theta[\dot{L}_i(\theta)\dot{L}_i(\theta)^T], \quad H_n(\theta) \equiv -\sum_{i=1}^n \ddot{L}_i(\theta),$$
$$I_n(\theta) \equiv \mathrm{diag}\{i_n(\theta)\}, \quad T_n(\theta) \equiv -\sum_{i=1}^n \dddot{L}_i(\theta). \tag{3.1}$$

The finiteness of $i_n(\theta)$ for every n is proved as part of Proposition 3 below. Note that if the process is not stationary. $E_\theta[\dot{L}_i(\theta)\dot{L}_i(\theta)^T]$ varies with the time index i because it depends on the joint distribution of $(X_{i\Delta}, X_{(i-1)\Delta})$. The square root of the diagonal element in $i_n(\theta)$ will determine the appropriate speeds of convergence for the corresponding component of $\hat{\theta}_n - \theta_0$, and I define the local $I_n^{1/2}(\theta)$-neighborhoods of the true parameter as $N_n^\varepsilon(\theta) \equiv \{\tilde{\theta} \in \Theta / \|I_n^{1/2}(\theta)(\tilde{\theta} - \theta)\| \leq \varepsilon\}$, where $\|\,\|$ denotes the Euclidean norm on $\mathbb{R}^K$. Recall that $E_\theta[H_n(\theta)] = i_n(\theta)$.[15] To identify the parameters, we make the following assumption.

ASSUMPTION 4 (Identification): *The true parameter vector θ_0 belongs to Θ, $I_n(\theta)$ is invertible,*

$$I_n^{-1}(\theta) \xrightarrow{\text{a.s}} 0 \text{ as } n \to \infty, \text{ uniformly in } \theta \in \Theta, \tag{3.2}$$

and $R_n(\theta, \tilde{\theta}) \equiv I_n^{-1/2}(\theta) T_n(\tilde{\theta}) I_n^{-1/2}(\theta)$ is uniformly bounded in probability for all $\tilde{\theta}$ in an $I_n^{1/2}(\theta)$-neighborhood of θ.

If X is a stationary diffusion, a sufficient condition that guarantees (3.2) is that for all $k = 1, \ldots, K$, $\theta \in \Theta$, and $x_0 \in D_X$,

$$0 < [I(\theta)]_{kk} = \int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} \{\partial \ln(p_X(\Delta, x \mid x_0; \theta))/\partial\theta_k\}^2 p_X(\Delta, x, x_0; \theta)\, dx\, dx_0 < +\infty \tag{3.3}$$

uniformly in θ (where $p_X(\Delta, x, x_0; \theta) = p_X(\Delta, x \mid x_0; \theta)\pi(x_0; \theta)$ denotes the joint density of observations sampled Δ units of time apart) since in that case $I_n^{-1}(\theta) = n^{-1} I^{-1}(\theta) \xrightarrow{\text{a.s.}} 0$. For the upper bound, it is sufficient that $|\partial \ln(p_X(\Delta, x \mid x_0; \theta))/\partial\theta_k|$ remain bounded as x varies in D_X, but not necessary. For the lower bound, it is sufficient that $\partial p_X(\Delta, x \mid x_0; \theta)/\partial\theta_k$ not be zero in a

[15] 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 due to the smoothness of the log-likelihood resulting from Corollary 1.

region (x, x_0) where the joint density has positive mass, i.e., the transition function p_X must not be uniformly flat in the direction of any one of the parameters θ_k. Otherwise $\partial p_X(\Delta, x \mid x_0; \theta)/\partial\theta_k \equiv 0$ for all (x, x_0) and the parameter vector cannot be identified. Furthermore, a sufficient condition for (3.3) is that $\mu(x;\theta) = \mu(x;\theta_0)$ and $\sigma(x;\theta) = \sigma(x;\theta_0)$ for π-almost all x imply $\theta = \theta_0$. I show in the proof of Proposition 3 that the boundedness condition on $R_n(\theta, \tilde{\theta})$ in Assumption 4 is automatically satisfied in the stationary case. A nonstationary example is provided in Section 5.

The strategy I employ to study the asymptotic properties of $\hat{\theta}_n^{(J)}$ is to first determine those of $\hat{\theta}_n$ (see Proposition 3) and then show that $\hat{\theta}_n^{(J)}$ and $\hat{\theta}_n$ share the same asymptotic properties provided one lets J go to infinity with n (Theorem 2). In Proposition 3, I show that general results pertaining to time series asymptotics (see, e.g., Basawa and Scott (1983) and Jeganathan (1995)) can be applied to the present context. These properties follow from first establishing that the likelihood ratio has the locally asymptotically quadratic (LAQ) structure, i.e.,

$$\ell_n(\theta + I_n^{-1/2}(\theta)h_n) - \ell_n(\theta) = h_n S_n(\theta) - h_n^T G_n(\theta) h_n/2 + o_p(1) \tag{3.4}$$

for every bounded sequence h_n such that $\theta + I_n^{-1/2}(\theta)h_n \in \Theta$, where $S_n(\theta) \equiv I_n^{-1/2}(\theta)V_n(\theta)$ and $G_n(\theta) \equiv I_n^{-1/2}(\theta)H_n(\theta)I_n^{-1/2}(\theta)$. Then, depending upon the joint distribution of (S_n, G_n), different cases arise:

PROPOSITION 3: *Under Assumptions* 1–4, *and for* $\Delta \in (0, \overline{\Delta})$, *the likelihood ratio satisfies the LAQ structure* (3.4), *the MLE* $\hat{\theta}_n$ *is consistent and has the following properties*:

i. (*Locally Asymptotically Mixed Normal Structure*): *If*

$$(S_n(\theta_0), G_n(\theta_0)) \xrightarrow{d} (G^{1/2}(\theta_0) \times Z, G(\theta_0)) \tag{3.5}$$

where Z *is a* $N(0, Id)$ *variable independent of the possibly random but almost surely finite and positive definite matrix* $G(\theta)$, *then*

$$I_n^{1/2}(\theta_0)(\hat{\theta}_n - \theta_0) \xrightarrow{d} G^{-1/2}(\theta_0) \times N(0, Id). \tag{3.6}$$

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

$$I_n^{1/2}(\theta)(\tilde{\theta}_n - \theta - I_n^{-1/2}(\theta)h) \xrightarrow{d} F(\theta) \text{ under } P_{\theta + I_n^{1/2}(\theta)h} \tag{3.7}$$

where $F(\theta)$ *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* $\varepsilon > 0$

$$\lim_{n\to\infty} \text{Prob}_{\theta_0}\left(I_n^{1/2}(\theta_0)(\hat{\theta}_n - \theta_0) \in C_\varepsilon\right) \geq \lim_{n\to\infty} \text{Prob}_{\theta_0}\left(I_n^{1/2}(\theta_0)(\tilde{\theta}_n - \theta) \in C_\varepsilon\right) \tag{3.8}$$

where $C_\varepsilon \equiv [-\varepsilon, +\varepsilon]^K$. *Further, if* $\tilde{\theta}_n$ *has the distribution* $I_n^{1/2}(\theta_0)(\tilde{\theta}_n - \theta_0) \xrightarrow{d} G^{-1/2}(\theta_0) \times N(0, \widetilde{V}_0)$ *under* P_{θ_0}, *then* $\widetilde{V}_0 - Id$ *is non-negative definite*.

ii. (*Locally Asymptotically Normal Structure*): *If X is a stationary diffusion, then a special case of the LAMN structure arises where* (3.3) *is a sufficient condition for Assumption* 4, $i(\theta) \equiv E_\theta[\dot{L}_1(\theta)\dot{L}_1(\theta)^T]$ *is Fisher's information matrix,* $i_n(\theta) = ni(\theta), I(\theta) \equiv \mathrm{diag}\{i(\theta)\}, I_n(\theta) = nI(\theta), G(\theta) = I^{-1/2}(\theta)i(\theta)I^{-1/2}(\theta)$ *is a nonrandom matrix and* (3.6) *reduces to*

$$n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, i(\theta_0)^{-1}). \tag{3.9}$$

The efficiency result simplifies to the Fisher-Rao form: $i(\theta_0)^{-1}$ *is the smallest possible asymptotic variance among that of all consistent and asymptotically Normal estimators of* θ_0.

iii. (*Locally Asymptotically Brownian Functional structure*): *If*

$$(S_n(\theta_0), G_n(\theta_0)) \xrightarrow{d} \left(\int_0^1 M_\tau \, dW_\tau, \int_0^1 M_\tau M_\tau^T \, d\tau \right) \tag{3.10}$$

where (M_τ, W_τ) *is a Gaussian process such that* W_τ *is a standard Brownian motion, then*

$$I_n^{1/2}(\theta_0)(\hat{\theta}_n - \theta_0) \xrightarrow{d} \left(\int_0^1 M_\tau M_\tau^T \, d\tau \right)^{-1} \times \int_0^1 M_\tau \, dW_\tau. \tag{3.11}$$

If M_τ *and* W_τ *are independent, then LABF is a special case of LAMN, but not otherwise.*

If one had normed the difference $(\hat{\theta}_n - \theta_0)$ by the stochastic factor $\mathrm{diag}\{H_n(\theta_0)\}^{1/2}$ rather than by the deterministic factor $I_n^{1/2}(\theta_0)$, then the asymptotic distribution of the estimator would have been $N(0, Id)$ rather than $G^{-1/2}(\theta_0) \times N(0, Id)$ (see the example in Section 5). In other words, the stochastic norming, while intrinsically more complicated, may be useful if the distribution of $G(\theta_0)$ is intractable, since 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 the stochastically normed difference would simply be $N(0, Id)$. None of these difficulties are present in the stationary case, where $G(\theta)$ is nonrandom.[16]

Sufficient conditions can be given that insure that the LAMN structure holds: for instance, if $G_n(\theta) \xrightarrow{p} G(\theta)$ uniformly in θ, over compact subsets of Θ, then (3.5) necessarily holds by applying Theorem 1 in Basawa and Scott (1983, page 34). Note also that 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 stationary case (see again the example in Section 5). Proposition 3 is not an end in itself since in our context $\hat{\theta}_n$ cannot be computed explicitly. It becomes useful however when one proves that the approximate maximum-likelihood estimator $\hat{\theta}_n^{(J)}$ is a good substitute for $\hat{\theta}_n$, in the sense

[16] In the terminology of Basawa and Scott (1983), when $G(\theta)$ is deterministic (resp. random), the model is called ergodic (resp. nonergodic). But the LAMN situation where $G(\theta)$ is random is only one particularly tractable form of nonergodicity.

that the asymptotic properties of $\hat{\theta}_n$ identified in Proposition 3 carry over to $\hat{\theta}_n^{(J)}$. For technical reasons, a minor refinement of Assumption 2 is needed:

ASSUMPTION 5 (Strengthening of Assumption 2 in the limiting case where $\alpha = 1$ 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). *If* $\alpha = 1$, *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*.

The following theorem shows that $\hat{\theta}_n^{(J)}$ inherit the asymptotic properties of the (uncomputable) true MLE $\hat{\theta}_n$:

THEOREM 2: *Under Assumptions* 1–5, *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* $n \to \infty$, *a sequence* $J_n \to \infty$ *can be chosen sufficiently large to deliver any rate of convergence of* $\hat{\theta}_n^{(J_n)}$ *to* $\hat{\theta}_n$. *In particular, there exists a sequence* $J_n \to \infty$ *such that* $\hat{\theta}_n^{(J_n)} - \hat{\theta}_n = o_p(I_n^{-1/2}(\theta_0))$ *under* P_{θ_0} *which then makes* $\hat{\theta}_n^{(J_n)}$ *and* $\hat{\theta}_n$ *share the same asymptotic distribution described in Proposition* 3.

4. EXPLICIT EXPRESSIONS FOR THE DENSITY EXPANSION

I now turn to the explicit computation of the terms in the density expansion. Theorem 1 showed that

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

Recall that $p_Z^{(J)}(\Delta, z|y_0; \theta)$ denotes the partial sum in (4.1) up to $j = J$. From (2.12), we have

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

so that the coefficients $\eta_Z^{(j)}$ are specific conditional moments of the process Y. As such, they can be computed in a number of ways, including for instance Monte Carlo integration. A particularly attractive alternative is to calculate explicitly a Taylor series expansion in Δ for the coefficients $\eta_Z^{(j)}$. Let $f(y, y_0)$ be a polynomial.

Taylor's Theorem applied to the function $s \mapsto E[f(Y_{t+s}, y_0)|Y_t = y_0]$ yields

$$E\big[f(Y_{t+\Delta}, y_0)\big|Y_t = y_0\big] = \sum_{k=0}^{K} A^k(\theta)\cdot f(y_0, y_0)\frac{\Delta^k}{k!} + E\big[A^{K+1}(\theta)\cdot f(Y_{t+\delta}, y_0)\big|Y_t = y_0\big]\frac{\Delta^{K+1}}{(K+1)!} \tag{4.3}$$

where $A(\theta)$ is the infinitesimal generator of the diffusion Y, defined as the operator $A(\theta)$: $f \mapsto \mu_Y(\cdot;\theta)\partial f/\partial y + (1/2)\partial^2 f/\partial y^2$. The following proposition provides sufficient conditions under which the series (4.3) is convergent:

PROPOSITION 4: *Under Assumptions* 1–3, *suppose that for the relevant boundaries of* $D_Y = (\underline{y}, \bar{y})$, *near* $\bar{y} = +\infty$: $\mu_Y(y;\theta) \leq -Ky^{\beta}$ *for some* $\beta > 1$; *near* $\underline{y} = -\infty$: $\mu_Y(y;\theta) \geq K|y|^{\beta}$ *for some* $\beta > 1$; *near* $\underline{y} = 0$: $\mu_Y(y;\theta) \geq \kappa y^{-\alpha}$ *for some* $\alpha > 1$ *and* $\kappa > 0$; *and near* $\bar{y} = 0$ $\mu_Y(y;\theta) \leq -\kappa|y|^{-\alpha}$ *for some* $\alpha > 1$ *and* $\kappa > 0$. *Then the diffusion Y is stationary with unconditional density* π_Y *and the series* (4.3) *converges in* $L^2(\pi_Y)$ *for fixed* $\Delta > 0$.

Now let $p_Z^{(J,K)}$ denote the Taylor series up to order K in Δ of $p_Z^{(J)}$. The series for the first seven Hermite coefficients ($j = 0, \ldots, 6$) are given by $\eta_Z^{(0)} = 1$, and to order $K = 3$ by:

$$\begin{aligned}\eta_Z^{(1,3)} = {} & -\mu_Y\Delta^{1/2} - \big(2\mu_Y\mu_Y^{[1]} + \mu_Y^{[2]}\big)\Delta^{3/2}/4 \\ & - \big(4\mu_Y\mu_Y^{[1]2} + 4\mu_Y^2\mu_Y^{[2]} + 6\mu_Y^{[1]}\mu_Y^{[2]} + 4\mu_Y\mu_Y^{[3]} + \mu_Y^{[4]}\big)\Delta^{5/2}/24,\end{aligned} \tag{4.4}$$

$$\begin{aligned}\eta_Z^{(2,3)} = {} & \big(\mu_Y^2 + \mu_Y^{[1]}\big)\Delta/2 + \big(6\mu_Y^2\mu_Y^{[1]} + 4\mu_Y^{[1]2} + 7\mu_Y\mu_Y^{[2]} + 2\mu_Y^{[3]}\big)\Delta^2/12 \\ & + \big(28\mu_Y^2\mu_Y^{[1]2} + 28\mu_Y^2\mu_Y^{[3]} + 16\mu_Y^{[1]3} + 16\mu_Y^3\mu_Y^{[2]} + 88\mu_Y\mu_Y^{[1]}\mu_Y^{[2]} \\ & + 21\mu_Y^{[2]2} + 32\mu_Y^{[1]}\mu_Y^{[3]} + 16\mu_Y\mu_Y^{[4]} + 3\mu_Y^{[5]}\big)\Delta^3/96,\end{aligned} \tag{4.5}$$

$$\begin{aligned}\eta_Z^{(3,3)} = {} & -\big(\mu_Y^3 + 3\mu_Y\mu_Y^{[1]} + \mu_Y^{[2]}\big)\Delta^{3/2}/6 - \big(12\mu_Y^3\mu_Y^{[1]} + 28\mu_Y\mu_Y^{[1]2} \\ & + 22\mu_Y^2\mu_Y^{[2]} + 24\mu_Y^{[1]}\mu_Y^{[2]} + 14\mu_Y\mu_Y^{[3]} + 3\mu_Y^{[4]}\big)\Delta^{5/2}/48,\end{aligned} \tag{4.6}$$

$$\begin{aligned}\eta_Z^{(4,3)} = {} & \big(\mu_Y^4 + 6\mu_Y^2\mu_Y^{[1]} + 3\mu_Y^{[1]2} + 4\mu_Y\mu_Y^{[2]} + \mu_Y^{[3]}\big)\Delta^2/24 \\ & + \big(20\mu_Y^4\mu_Y^{[1]} + 50\mu_Y^3\mu_Y^{[2]} + 100\mu_Y^2\mu_Y^{[1]2} + 50\mu_Y^2\mu_Y^{[3]} + 23\mu_Y\mu_Y^{[4]} \\ & + 180\mu_Y\mu_Y^{[1]}\mu_Y^{[2]} + 40\mu_Y^{[1]3} + 34\mu_Y^{[2]2} + 52\mu_Y^{[1]}\mu_Y^{[3]} + 4\mu_Y^{[5]}\big)\Delta^3/240,\end{aligned} \tag{4.7}$$

$$\begin{aligned}\eta_Z^{(5,3)} = {} & -\big(\mu_Y^5 + 10\mu_Y^3\mu_Y^{[1]} + 15\mu_Y\mu_Y^{[1]2} + 10\mu_Y^2\mu_Y^{[2]} \\ & + 10\mu_Y^{[1]}\mu_Y^{[2]} + 5\mu_Y\mu_Y^{[3]} + \mu_Y^{[4]}\big)\Delta^{5/2}/120,\end{aligned} \tag{4.8}$$

$$\begin{aligned}\eta_Z^{(6,3)} = {} & \big(\mu_Y^6 + 15\mu_Y^4\mu_Y^{[1]} + 15\mu_Y^{[1]3} + 20\mu_Y^3\mu_Y^{[2]} + 15\mu_Y^{[1]}\mu_Y^{[3]} + 45\mu_Y^2\mu_Y^{[1]2} \\ & + 10\mu_Y^{[2]2} + 15\mu_Y^2\mu_Y^{[3]} + 60\mu_Y\mu_Y^{[1]}\mu_Y^{[2]} + 6\mu_Y\mu_Y^{[4]} + \mu_Y^{[5]}\big)\Delta^3/720,\end{aligned} \tag{4.9}$$

where I have used the more compact notation $\mu_Y^{[k]m}$ for $(\partial^k \mu_Y(y_0;\theta)/\partial y_0^k)^m$.

Different ways of gathering the terms are available (as in the CLT, where for example both the Edgeworth and Gram-Charlier expansions are based on a Hermite expansion). Here, if we gather all the terms according to increasing powers of Δ instead of increasing order of the Hermite polynomials, and let $\tilde{p}_Z^{(K)} \equiv p_Z^{(\infty,K)}$ (and similarly for Y, so that $\tilde{p}_Y^{(K)}(\Delta, y\,|\,y_0;\theta) = \Delta^{-1/2}\tilde{p}_Z^{(K)}(\Delta, \Delta^{-1/2}(y-y_0)|y_0;\theta)$, and then for X), we obtain an explicit representation of $\tilde{p}_Y^{(K)}$, given by

$$\tilde{p}_Y^{(K)}(\Delta,y|y_0;\theta)=\Delta^{-1/2}\phi\left(\frac{y-y_0}{\Delta^{1/2}}\right)\exp\left(\int_{y_0}^{y}\mu_Y(w;\theta)dw\right)\sum_{k=0}^{K}c_k(y|y_0;\theta)\frac{\Delta^k}{k!} \tag{4.10}$$

where $c_0(y|y_0;\theta)=1$ and for all $j \geq 1$:

$$\begin{aligned} c_j(y|y_0;\theta) &= j(y-y_0)^{-j}\int_{y_0}^{y}(w-y_0)^{j-1} \\ &\quad\times\left\{\lambda_Y(w;\theta)c_{j-1}(w|y_0;\theta)+\left(\partial^2 c_{j-1}(w|y_0;\theta)/\partial w^2\right)\Big/2\right\}dw. \end{aligned} \tag{4.11}$$

Finally, note that in general, conditional moments of the process Y need not be analytic in time,[17] in which case (4.3) and (4.10) must be interpreted strictly as Taylor series. Even when that is the case, their relevance for empirical work lies in the fact that including a small number of terms (one, two, or three) makes the approximation very accurate for the values these variables typically take in financial econometrics, as we shall now see.[18]

5. ACCURACY OF THE APPROXIMATIONS AND MONTE CARLO EVIDENCE

While Figure 1 shows that the approximation of p_X was extremely accurate as a function of the state variables, it does not necessarily imply that the resulting parameter estimates would in practice necessarily be close to the true MLE, as was proved theoretically in Theorem 2. To answer that question, I perform Monte Carlo experiments. Consider first the Ornstein-Uhlenbeck specification, $dX_t = -\beta X_t dt + \sigma\, dW_t$, where $\theta \equiv (\beta, \sigma^2)$ and $D_X = (-\infty, +\infty)$. The process X has a Gaussian transition density with mean $x_0 e^{-\beta\Delta}$ and variance $(1-e^{-2\beta\Delta})\sigma^2/2\beta$. In this case, $Y = \gamma(X;\theta) = \sigma^{-1}X$, $\mu_Y(y;\theta) = -\beta y$, and the additional terms in

[17] Note however that as a result of Theorem 1 the transition function is analytic in the forward state variable. The expansion is designed to deliver an approximation of the density function $y \mapsto p_Y(\Delta, y\,|\,y_0;\theta)$ for a fixed value of the backward (conditioning) variable y_0. Therefore, except in the limit where Δ becomes infinitely small, it is not designed to reproduce the limiting behavior of p_Y in the limit where y_0 tends to the boundaries. The expansion delivers the correct behavior for y tending to the boundaries, except in the limiting situation of $\mu_Y(y;\theta) \sim \kappa y^{-1}$ in Assumption 3.1 where it is only appropriate if Δ becomes infinitesimally small.

[18] See also the companion paper (Aït-Sahalia (1999)) for examples and an application to the estimation of interest rate models.

the approximation $p_Z^{(J)}$ need only correct for the inadequacy of the conditional moments in the leading term $p_Z^{(0)}$, not for any non-Gaussianity. In other words, in the transformation from X to Y being linear, there is no deformation or stretching of the Gaussian leading term when going from the approximation of p_Y to that of p_X. By specializing Proposition 3 to this model, one obtains the following asymptotic distributions for the MLE:[19]

COROLLARY 2: (Asymptotic Distribution of the MLE for the Ornstein-Uhlenbeck Model):

i. *If* $\beta > 0$ (*LAN*, *stationary case*):

$$\sqrt{n}\left(\begin{pmatrix}\hat{\beta}_n \\ \hat{\sigma}_n^2\end{pmatrix} - \begin{pmatrix}\beta \\ \sigma^2\end{pmatrix}\right) \xrightarrow{d} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} \frac{e^{2\beta\Delta}-1}{\Delta^2} & \frac{\sigma^2(e^{2\beta\Delta}-1-2\beta\Delta)}{\beta\Delta^2} \\ \frac{\sigma^2(e^{2\beta\Delta}-1-2\beta\Delta)}{\beta\Delta^2} & \frac{\sigma^4\left((e^{2\beta\Delta}-1)^2+2\beta^2\Delta^2(e^{2\beta\Delta}+1)+4\beta\Delta(e^{2\beta\Delta}-1)\right)}{\beta^2\Delta^2(e^{2\beta\Delta}-1)} \end{pmatrix}\right). \tag{5.1}$$

ii. *If* $\beta < 0$ (*LAMN*, *explosive case*), *assume* $X_0 = 0$; *then*

$$\frac{e^{-(n+1)\beta\Delta}\Delta}{e^{-2\beta\Delta}-1}(\hat{\beta}_n - \beta) \xrightarrow{d} G^{-1/2} \times N(0,1) \quad \textit{and} \quad \sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4) \tag{5.2}$$

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

iii. *If* $\beta = 0$ (*LAQ*, *unit root case*), *assume* $X_0 = 0$; *then*

$$n\hat{\beta}_n \xrightarrow{d} (1 - W_1^2)\Big/\left(2\Delta\int_0^1 W_t^2\,dt\right) \quad \textit{and} \quad \sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4) \tag{5.3}$$

where W_t *denotes a standard Brownian motion*.

In Monte Carlo experiments, I study the behavior of the true MLE of θ (which is computable in this example since the transition function is known in closed-form), the Euler estimator, and the estimators of this paper corresponding to one and two orders in Δ respectively. The Euler approximation corresponds to a simple discretization of the continuous-time stochastic differential equation, where the differential equation (1.1) is replaced by the difference equation $X_{t+\Delta} - X_t = \mu(X_t, ; \theta)\Delta + \sigma(X_t; \theta)\sqrt{\Delta}\varepsilon_{t+\Delta}$ with $\varepsilon_{t+\Delta} \sim N(0,1)$, so that

$$p_X^{\text{Euler}}(\Delta, x|x_0; \theta) = (2\pi\Delta\sigma^2(x_0;\theta))^{-1/2} \times \exp\left\{-\left(x - x_0 - \mu(x_0;\theta)\Delta\right)^2 / 2\Delta\sigma^2(x_0;\theta)\right\}. \tag{5.4}$$

[19] Since there is no confusion possible in what follows, the subscript 0 is omitted when denoting the true parameter values.

I set the true value of σ^2 at 1.0, and examine the behavior of the various estimators of β and σ^2 for the different cases of Corollary 2, by setting $\beta = 10$, 5, and 1 (stationary root LAN), $\beta = 0$ (unit root LABF), and $\beta = -1$ (explosive root LAMN). For each value of the parameters, I perform $M = 5{,}000$ Monte Carlo simulations of the sample paths generated by the model, each containing $n = 1{,}000$ observations.

These Monte Carlo experiments answer four separate questions. Firstly, how accurate are the various asymptotic distributions in Corollary 2? This question is answered in Figures 2 and 3, where I plot the finite sample distributions of the estimators (histograms) and the corresponding asymptotic distribution (solid line). The asymptotic distribution of $\hat{\theta}_n - \theta$ reported in Panels A–C of Figure 2 is from (5.1). Not surprisingly, as the drift parameter β makes the process closer and closer to a unit root (β decreasing from 10 to 1), the quality of the asymptotic approximation (5.1) deteriorates and the small sample distribution starts to resemble (5.3), which is strongly skewed. This only affects the drift parameter; the estimator of σ^2 behaves in small samples as predicted by the asymptotic distribution—which is compatible with the fact that the distribution for estimating σ^2 is continuous when going through the $\beta = 0$ boundary. Panel A of Figure 3 reports results for the unit root case, with the asymptotic distribution given in (5.3). In the explosive case ($\beta < 0$), Panel B of Figure 3 is based on the Cauchy distribution (5.2), while Panel C exploits the possibility of random norming to obtain a Gaussian asymptotic distribution of the drift coefficient (see (A.69) in the Appendix). The diffusion estimator is identical in both Panels B and C, and is therefore not repeated in Panel C. Since the rate of convergence in nonstationary cases varies, both Panels B and C report the distribution of the drift estimator scaled by the relevant rate of convergence, rather than the raw distribution of $\hat{\beta}_n - \beta$ as in all other panels. The simulations show that in both nonstationary cases, and in the stationary case when sufficiently far away from a unit root, the asymptotic distribution of the drift estimator is an accurate guide to its small sample distribution.

The second question these experiments address is: what is the dispersion of the MLE around the true value? Tables I and II report the first four moments of the finite sample and asymptotic distributions. For each of the parameter values and the M samples, I also report in these tables the first two moments of the differences between the true MLE estimators of β and σ^2, their Euler versions and the estimators from using the method of this paper with one and two terms. This makes it possible, thirdly, to compare the MLE dispersion, or sampling noise, to the distance between the MLE and the various approximations under consideration. In particular, when selecting the order of approximation, it is unnecessary to select a value larger than what is required to make the distance between $\hat{\theta}_n^{(J)}$ and $\hat{\theta}_n$ an order of magnitude smaller than the distance between $\hat{\theta}_n$ and the true value θ (as measured by the exact MLE sampling distribution). These simulations show that the parameter estimates obtained with one and even more so two terms are several orders of magnitude closer to the exact MLE than

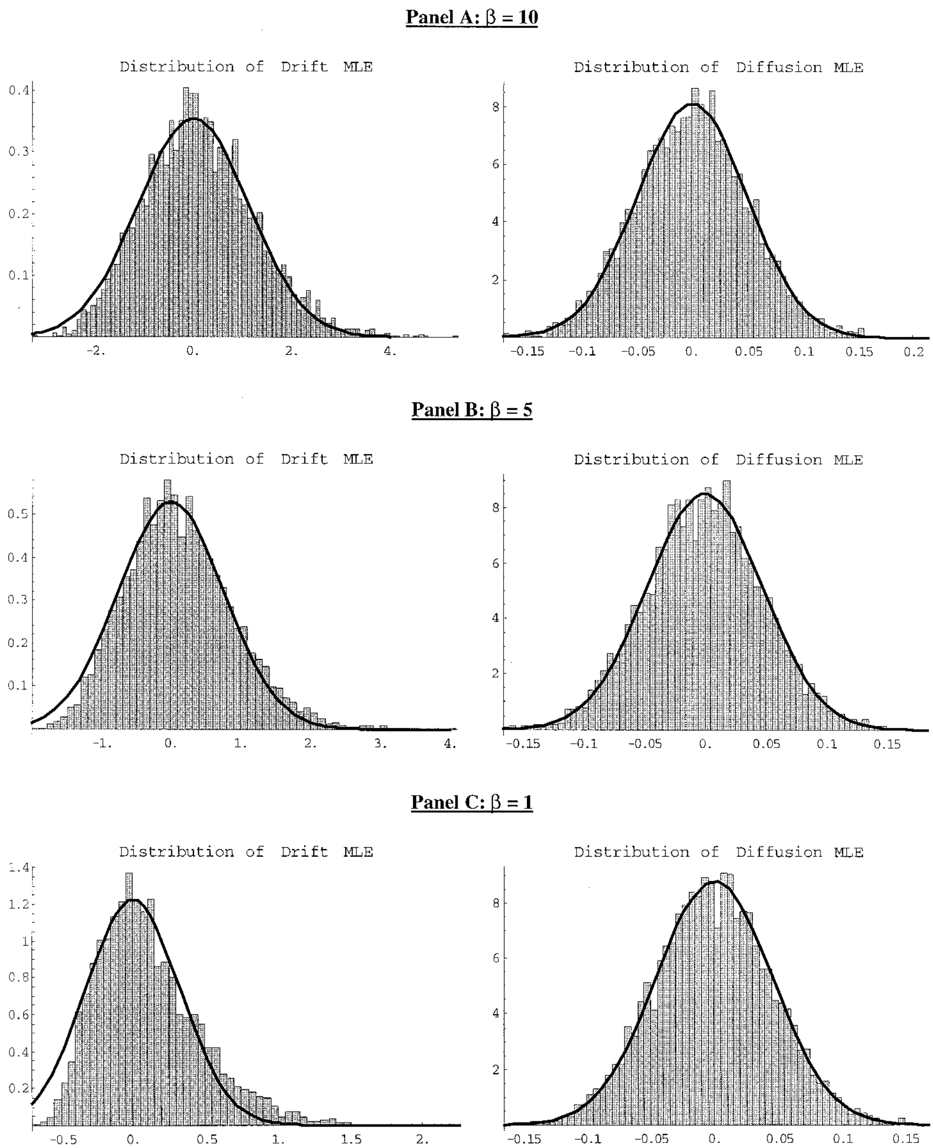


FIGURE 2.— Small sample and asymptotic distributions of the MLE for the Ornstein-Uhlenbeck process: stationary processes.

the MLE is to the parameter, so that the approximate estimates can be used in place of the exact MLE in practice.

Finally, these Monte Carlo experiments make it possible to compare the relative accuracy of the three estimators based on the Euler discretization approximation, and those of this paper. The results of the bottom part of Tables I and II,

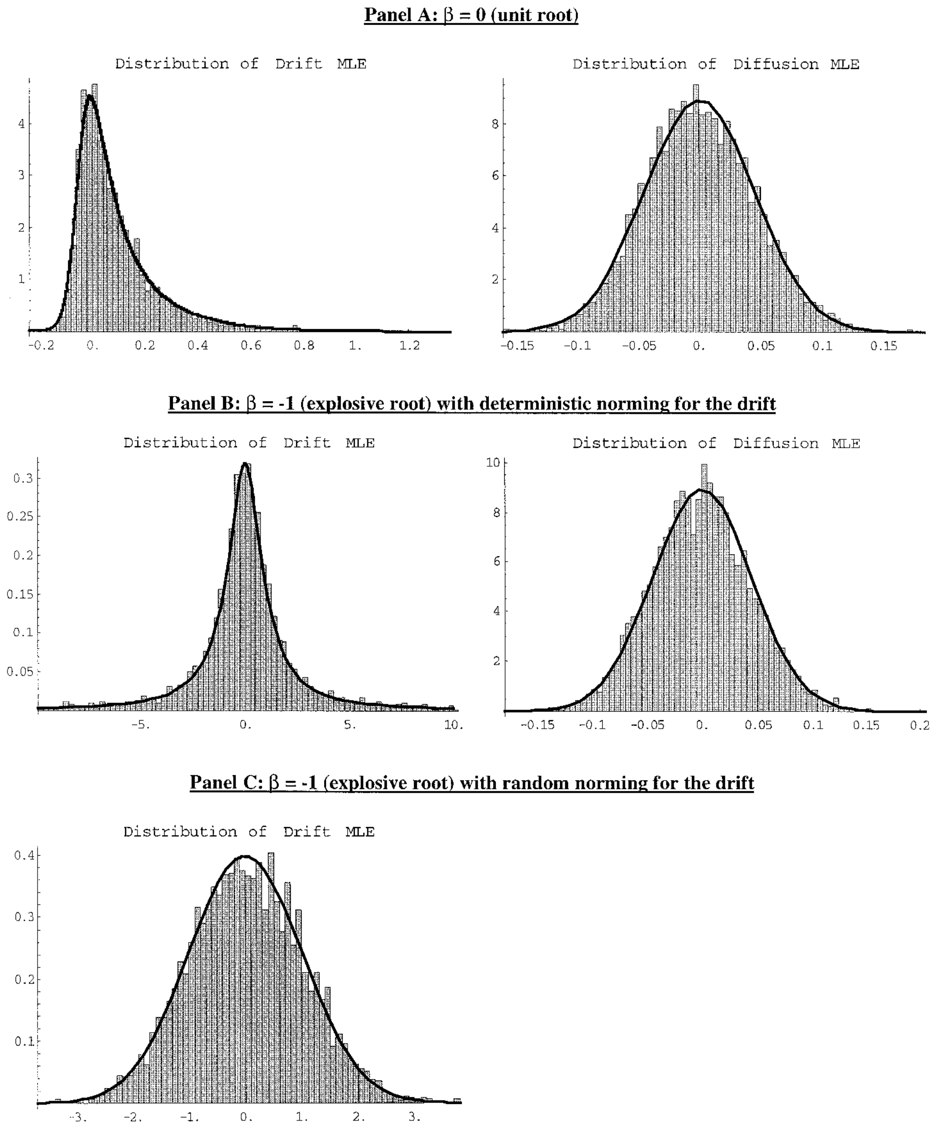


FIGURE 3.— Small sample and asymptotic distributions of the MLE for the Ornstein-Uhlenbeck process: nonstationary processes.

comparing the differences between the approximate and exact estimators, show that the estimators with one and even more so with two terms are substantially more accurate than the Euler estimator, even though the latter is in an ideal situation in this example. Indeed, since the true transition function is Gaussian, the only approximation involved in the Euler estimation consists in using first order Taylor series expansions of the true conditional mean and variances rather

TABLE I

COMPARISON OF APPROXIMATE ESTIMATORS FOR THE ORNSTEIN-UHLENBECK PROCESS

			Stationary Processes		
			Panel A: $\beta = 10$	Panel B: $\beta = 5$	Panel C: $\beta = 1$
$\hat{\beta}^{(\mathrm{MLE})} - \beta^{(\mathrm{TRUE})}$	Mean	Asymptotic	0	0	0
		Sample	0.11826	0.11522	0.10965
	Stand. Dev.	Asymptotic	1.12619	0.75209	0.32561
		Sample	1.12894	0.77012	0.36598
	Skewness	Asymptotic	0	0	0
		Sample	0.362	0.457	0.959
	Kurtosis	Asymptotic	3	3	3
		Sample	3.295	3.432	4.484
$\hat{\beta}^{(\mathrm{EUL})} - \hat{\beta}^{(\mathrm{MLE})}$	Mean	Sample	−0.933615	−0.248697	−0.0130134
	Stand. Dev.	Sample	0.203568	0.074819	0.0091453
$\hat{\beta}^{(1)} - \hat{\beta}^{(\mathrm{MLE})}$	Mean	Sample	−0.279554	−0.070803	−0.0028979
	Stand. Dev.	Sample	0.162931	0.049566	0.0046431
$\hat{\beta}^{(2)} - \hat{\beta}^{(\mathrm{MLE})}$	Mean	Sample	0.013228	0.001209	−0.0000005
	Stand. Dev.	Sample	0.050978	0.006701	0.0000163
$\hat{v}^{(\mathrm{MLE})} - v^{(\mathrm{TRUE})}$	Mean	Asymptotic	0	0	0
		Sample	−0.0002404	−0.0003080	−0.0004357
	Stand. Dev.	Asymptotic	0.0491017	0.0468901	0.0451521
		Sample	0.0504349	0.0480553	0.0462890
	Skewness	Asymptotic	0	0	0
		Sample	0.067	0.054	0.061
	Kurtosis	Asymptotic	3	3	3
		Sample	3.066	3.046	3.041
$\hat{v}^{(\mathrm{EUL})} - \hat{v}^{(\mathrm{MLE})}$	Mean	Sample	−0.1716920	−0.092254	−0.021040
	Stand. Dev.	Sample	0.0218769	0.014914	0.007035
$\hat{v}^{(1)} - \hat{v}^{(\mathrm{MLE})}$	Mean	Sample	−0.0000395	0.000732	0.000992
	Stand. Dev.	Sample	0.0014829	0.000439	0.000059
$\hat{v}^{(2)} - \hat{v}^{(\mathrm{MLE})}$	Mean	Sample	0.0000182	0.000010	0.000010
	Stand. Dev.	Sample	0.0000433	0.000075	0.000046
$\mathrm{Cov}(\hat{\beta}^{(\mathrm{MLE})}, \hat{v}^{(\mathrm{MLE})})$		Asymptotic	0.022831	0.010673	0.002026
		Sample	0.023627	0.010958	0.002240
$\mathrm{Cov}(\hat{\beta}^{(\mathrm{EUL})}, \hat{v}^{(\mathrm{EUL})})$		Sample	0.000529	−0.000027	−0.000298
$\mathrm{Cov}(\hat{\beta}^{(1)}, \hat{v}^{(1)})$		Sample	0.022308	0.010632	0.002220
$\mathrm{Cov}(\hat{\beta}^{(2)}, \hat{v}^{(2)})$		Sample	0.023785	0.010978	0.002242

Notes: The model is $dX_t = -\beta X_t\,dt + \sigma\,dW_t$. In the table, v designates the diffusion parameter σ^2, whose true value is 1.0. The superscripts (MLE), (EUL), (1) and (2) refer to the exact estimator, the estimator based on the Euler approximation, and the estimators based on the methods of this paper with one and two terms respectively (see (4.10)). Panels A, B, and C in this table correspond to the same panels in Figure 2. The asymptotic values correspond to the asymptotic distribution given in Corollary 2. The sample moments are averages over 5,000 Monte Carlo simulations.

TABLE II

COMPARISON OF APPROXIMATE ESTIMATORS FOR THE ORNSTEIN-UHLENBECK PROCESS

			Non-Stationary Processes		
			Unit Root Panel A: $\beta = 0$	Explosive Root Panel B: $\beta = -1$	Explosive Root Panel C: $\beta = -1$
$\hat{\beta}^{(MLE)} - \beta^{(TRUE)}$	Mean	Asymptotic	0.09226	0	0
		Sample	0.09274	0.6026	0.0153
	Stand. Dev.	Asymptotic	0.16701	$+\infty$	1
		Sample	0.16494	70.0266	1.0239
	Skewness	Asymptotic	2.265	undefined	0
		Sample	2.154	27.685	0.0428
	Kurtosis	Asymptotic	11.582	undefined	3
		Sample	9.527	1588.61	3.002
$\hat{\beta}^{(EUL)} - \hat{\beta}^{(MLE)}$	Mean	Sample	−0.00034314		
	Stand. Dev.	Sample	0.001004		
$\hat{\beta}^{(1)} - \hat{\beta}^{(MLE)}$	Mean	Sample	0.0000179		
	Stand. Dev.	Sample	0.0003434		
$\hat{\beta}^{(2)} - \hat{\beta}^{(MLE)}$	Mean	Sample	−0.000000039		
	Stand. Dev.	Sample	0.000000313		
$\hat{v}^{(MLE)} - v^{(TRUE)}$	Mean	Asymptotic	0	0	0
		Sample	0.0002837	−0.001027	−0.001027
	Stand. Dev.	Asymptotic	0.0447214	0.044730	0.044730
		Sample	0.0443634	0.044962	0.044962
	Skewness	Asymptotic	0	0	0
		Sample	0.125	0.097	0.097
	Kurtosis	Asymptotic	3	3	3
		Sample	2.982	3.079	3.079
$\hat{v}^{(EUL)} - \hat{v}^{(MLE)}$	Mean	Sample	−0.001782		
	Stand. Dev.	Sample	0.003171		
$\hat{v}^{(1)} - \hat{v}^{(MLE)}$	Mean	Sample	0.000103		
	Stand. Dev.	Sample	0.000045		
$\hat{v}^{(2)} - \hat{v}^{(MLE)}$	Mean	Sample	0.000011		
	Stand. Dev.	Sample	0.000044		
$\mathrm{Cov}(\hat{\beta}^{(MLE)}, \hat{v}^{(MLE)})$		Asymptotic	0	0	0
		Sample	0.0003960	$-1.09\,10^{-10}$	$-1.09\,10^{-10}$
$\mathrm{Cov}(\hat{\beta}^{(EUL)}, \hat{v}^{(EUL)})$		Sample	−0.0001255		
$\mathrm{Cov}(\hat{\beta}^{(1)}, \hat{v}^{(1)})$		Sample	0.0003955		
$\mathrm{Cov}(\hat{\beta}^{(2)}, \hat{v}^{(2)})$		Sample	0.0003953		

Notes: The same notes as in Table I apply. In the explosive case, the dispersion of the simulated data around the mean of the process (zero) makes it impractical to simulate the approximate estimators. The panels match those of Figure 3. The diffusion estimators in Panels B and C are identical.

TABLE III

COMPARISON OF APPROXIMATE ESTIMATORS FOR THE VASICEK, COX-INGERSOLL-ROSS, AND BLACK-SCHOLES MODELS

		Vasicek $dX_t = \beta(\alpha - X_t)\,dt + \sigma\,dW_t$	Cox-Ingersoll-Ross $dX_t = \beta(\alpha - X_t)\,dt + \sigma X_t^{.5}\,dW_t$	Black-Scholes $dX_t = \beta X_t\,dt + \sigma X_t\,dW_t$
$\hat{\beta}^{(\mathrm{MLE})} - \beta^{(\mathrm{TRUE})}$	Mean	0.099674	0.09711	−0.0002561
	Stand. Dev.	0.178366	0.18772	0.0468815
$\hat{\beta}^{(\mathrm{EUL})} - \hat{\beta}^{(\mathrm{MLE})}$	Mean	−0.015993	−0.00164	0.0017667
	Stand. Dev.	0.009873	0.03250	0.0008121
$\hat{\beta}^{(1)} - \hat{\beta}^{(\mathrm{MLE})}$	Mean	−0.003675	0.00053	−0.0020946
	Stand. Dev.	0.005003	0.00105	0.0017052
$\hat{\beta}^{(2)} - \hat{\beta}^{(\mathrm{MLE})}$	Mean	−0.000012	−0.00036	0.0000197
	Stand. Dev.	0.000270	0.00494	0.0000294
$\hat{\alpha}^{(\mathrm{MLE})} - \alpha^{(\mathrm{TRUE})}$	Mean	0.000023341	0.0006947	not applicable
	Stand. Dev.	0.009078321	0.0011893	not applicable
$\hat{\alpha}^{(\mathrm{EUL})} - \hat{\alpha}^{(\mathrm{MLE})}$	Mean	−0.000000003	0.0000089	not applicable
	Stand. Dev.	0.000000071	0.0001789	not applicable
$\hat{\alpha}^{(1)} - \hat{\alpha}^{(\mathrm{MLE})}$	Mean	0.000001102	0.0001747	not applicable
	Stand. Dev.	0.000109126	0.0002057	not applicable
$\hat{\alpha}^{(2)} - \hat{\alpha}^{(\mathrm{MLE})}$	Mean	−0.000000017	−0.0000009	not applicable
	Stand. Dev.	0.000003544	0.0001322	not applicable
$\hat{\sigma}^{(\mathrm{MLE})} - \sigma^{(\mathrm{TRUE})}$	Mean	0.00008690	0.000560	−0.00000165
	Stand. Dev.	0.00101568	0.004905	0.00966928
$\hat{\sigma}^{(\mathrm{EUL})} - \hat{\sigma}^{(\mathrm{MLE})}$	Mean	−0.00073620	−0.002768	0.00562312
	Stand. Dev.	0.00022012	0.001492	0.00189766
$\hat{\sigma}^{(1)} - \hat{\sigma}^{(\mathrm{MLE})}$	Mean	−0.00000043	0.000029	0.00005585
	Stand. Dev.	0.00000248	0.000391	0.00006205
$\hat{\sigma}^{(2)} - \hat{\sigma}^{(\mathrm{MLE})}$	Mean	−0.00000002	0.000013	−0.00000058
	Stand. Dev.	0.00000029	0.000387	0.00000117
$\mathrm{Cov}(\hat{\beta}^{(\mathrm{MLE})}, \hat{\sigma}^{(\mathrm{MLE})})$		0.0000308	0.000202	0.00003267
$\mathrm{Cov}(\hat{\beta}^{(\mathrm{EUL})}, \hat{\sigma}^{(\mathrm{EUL})})$		−0.0000077	−0.000017	0.00227254
$\mathrm{Cov}(\hat{\beta}^{(1)}, \hat{\sigma}^{(1)})$		0.0000305	0.000199	0.00003588
$\mathrm{Cov}(\hat{\beta}^{(2)}, \hat{\sigma}^{(2)})$		0.0000308	0.000200	0.00003262
$\mathrm{Cov}(\hat{\beta}^{(\mathrm{MLE})}, \hat{\alpha}^{(\mathrm{MLE})})$		0.0000368	−0.00102	not applicable
$\mathrm{Cov}(\hat{\beta}^{(\mathrm{EUL})}, \hat{\alpha}^{(\mathrm{EUL})})$		0.0000345	−0.00099	not applicable
$\mathrm{Cov}(\hat{\beta}^{(1)}, \hat{\alpha}^{(1)})$		0.0000371	−0.00102	not applicable
$\mathrm{Cov}(\hat{\beta}^{(2)}, \hat{\alpha}^{(2)})$		0.0000367	−0.00102	not applicable
$\mathrm{Cov}(\hat{\alpha}^{(\mathrm{MLE})}, \hat{\sigma}^{(\mathrm{MLE})})$		0.0000003112	0.0000006	not applicable
$\mathrm{Cov}(\hat{\alpha}^{(\mathrm{EUL})}, \hat{\sigma}^{(\mathrm{EUL})})$		0.0000002616	0.0000057	not applicable
$\mathrm{Cov}(\hat{\alpha}^{(1)}, \hat{\sigma}^{(1)})$		0.0000003134	0.0000006	not applicable
$\mathrm{Cov}(\hat{\alpha}^{(2)}, \hat{\sigma}^{(2)})$		0.0000003112	0.0000006	not applicable

Notes: The true values of the parameters, chosen to be realistic for US interest rates (Vasicek and CIR) and stock prices (Black-Scholes) respectively, are: $\beta = 0.5$, $\alpha = 0.06$, $\sigma = 0.03$ (Vasicek), $\beta = 0.5$, $\alpha = 0.06$, $\sigma = 0.15$ (CIR), and $\beta = 0.2$, $\sigma = 0.3$ (Black-Scholes). All moments reported are averages over 5,000 Monte Carlo replications.

than the exact expressions. By contrast, the approximate estimators corresponding to one and two terms mimic the moments of the MLE finite sample distribution extremely closely, often to multiple accurate decimal places. Further Monte Carlo experiments for three standard models in finance (Black-Scholes (1973), Vasicek (1977), Cox-Ingersoll-Ross (1985)) reported in Table III reveal that the estimators proposed here outperform by orders of magnitude the Euler estimator, especially in non-Gaussian situations.

6. CONCLUSIONS

This paper has constructed a series of explicit functions, based on Hermite expansions and converging to the conditional density of the diffusion process, under mild regularity conditions. This method makes maximum-likelihood a practical option for the estimation of parameters in discretely-sampled diffusion models. Beyond maximum-likelihood, the formulae for the expansion of p_X apply to any specification of (μ, σ^2), including nonparametric ones. Different types of evidence have been provided in favor of this method. First, it largely outperforms discrete approximations, binomial trees, PDE methods, and simulation-based methods in a direct comparison of speed and accuracy (Figure 1). Second, Monte Carlo experiments show that maximizing the log-likelihood approximation provides parameter estimates that are very close to the true MLE (Tables I, II, and III) and outperforms by several orders of magnitude the alternative methods—not only in terms of computational speed and ease of implementation but also in terms of accuracy.

Extensions to multi-dimensional diffusions (including unobservable state variables to be integrated out of the likelihood function, such as stochastic volatility) and applications to derivative pricing will be considered in future work. A further appeal of this method lies in its potential to be generalized to yet other types of Markov processes, such as those driven by non-Brownian Lévy processes for instance. As I remarked earlier, this generalization would involve different scaling $X \to Y \to Z$, a non-Gaussian leading term for p_Z (in this case a natural choice is the limiting transition density of the driving process), and orthogonal functions that correspond to this leading term. But the basic principle remains valid: first form an orthogonal series to approximate the density and prove its convergence; then determine its coefficients using repeated iterations of the infinitesimal generator of the Markov process under consideration.

Department of Economics, Princeton University, Princeton, NJ 08544-1021, U.S.A., and NBER; yacine@princeton.edu; http://www.princeton.edu/~yacine

Manuscript received December, 1997; final revision received October, 2000.

APPENDIX: PROOFS

PROOF OF PROPOSITION 1: I treat the case where $D_Y = (0, +\infty)$, the other boundary configurations being dealt with similarly. Let $s_Y(v; \theta) \equiv \exp\{-\int^v 2\mu_Y(u; \theta)\, du\}$ be the scale density of Y and

$S_Y(y;\theta) \equiv \int^y s_Y(v;\theta)\,dv$ its scale function.[20] In each case, the lower limit of integration is a fixed value in D_Y, the choice of which is irrelevant in what follows (i.e., for the purpose of determining whether or not the relevant quantities below are infinite or not). Let $m_Y(v;\theta) \equiv 1/s_Y(v;\theta)$ be the speed density of Y.

Step 1—Existence and unicity in law of a weak solution: This follows from the Engelbert-Schmidt criterion (see, e.g., Theorem 5.5.15 in Karatzas and Shreve (1991), replacing $\mathbb{R}$ by D_Y throughout). To apply this result, note that continuity of μ_Y (and of course $\sigma_Y = 1$) implies the local integrability requirements for $|\mu_Y|/\sigma_Y^2$ and $1/\sigma_Y^2$. Explosions are ruled out in Step 2 of this proof.

Step 2—Unattainability of the boundaries 0 and $+\infty$: Define

$$\text{(A.1)}\qquad \begin{cases} \Sigma_\infty \equiv \int_y^{+\infty}\left\{\int_y^v m_Y(u;\theta)\,du\right\} s_Y(v;\theta)\,dv = \int_y^{+\infty}\left\{\int_u^{+\infty} s_Y(v;\theta)\,dv\right\} m_Y(u;\theta)\,du, \\ \Sigma_0 \equiv \int_0^y\left\{\int_v^y m_Y(u;\theta)\,du\right\} s_Y(v;\theta)\,dv = \int_0^y\left\{\int_0^u s_Y(v;\theta)\,dv\right\} m_Y(u;\theta)\,du. \end{cases}$$

From Feller's test for explosions, $\text{Prob}(T_Y = \infty) = 1$ if and only if $\Sigma_\infty = \infty$ and $\Sigma_0 = \infty$ (see, e.g., Karatzas and Shreve (1991, Theorem 5.5.29) or Karlin and Taylor (1981, Section 15.6)). Near $\bar{y} = +\infty$, Assumption 3.1 gives the upper bound $\mu_Y(y;\theta) \le Ky$ for all $y \ge E$ (without restraining how negative μ_Y can get); thus

$$\text{(A.2)}\qquad \begin{aligned} \Sigma_\infty &= \int_y^{+\infty}\left\{\int_u^{+\infty} s_Y(v;\theta)\,dv\right\} s_Y^{-1}(u;\theta)\,du = \int_y^{+\infty}\int_u^{+\infty} e^{-\int_u^v 2\mu_Y(w;\theta)dw}\,dvdu \\ &\ge \int_y^{+\infty}\int_u^{+\infty} e^{-\int_u^v 2Kw\,dw}\,dvdu = \int_y^{+\infty}\left\{\int_u^{+\infty} e^{-Kv^2}\,dv\right\} e^{Ku^2}\,du. \end{aligned}$$

Now by integration by parts

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

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

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

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

Therefore

$$\text{(A.3)}\qquad \Sigma_\infty \ge \int_y^{+\infty}\left\{\int_u^{+\infty} e^{-Kv^2}\,dv\right\} e^{Ku^2}\,du \ge \int_y^{+\infty} (2Ku + u^{-1})^{-1} e^{-Ku^2} e^{Ku^2}\,du = +\infty.$$

If $y = 0$, there exist constants $\varepsilon_0, \kappa, \alpha$ such that for all $0 < y \le \varepsilon_0$ and $\theta \in \Theta$, $\mu_Y(y;\theta) \ge \kappa y^{-\alpha}$ where either $\alpha > 1$ and $\kappa > 0$, or $\alpha = 1$ and $k \ge 1$. If $\alpha > 1$, we have for $0 < v \le \varepsilon_0$

$$\text{(A.4)}\qquad s_Y(v;\theta) = \exp\left\{\int_v 2\mu_Y(w;\theta)dw\right\} \ge \exp\left\{\int_v 2\kappa w^{-\alpha} dw\right\} = \kappa_0 \exp\left\{2\kappa(\alpha-1)v^{-(\alpha-1)}\right\}$$

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

$$\text{(A.5)}\qquad s_Y(v;\theta) \ge \exp\left\{\int_v 2\kappa w^{-1} dw\right\} = k_0 \exp\{-2\kappa \ln(v)\} = k_0 v^{-2\kappa}$$

[20] The scale function has the following intuitive interpretation: with $\underline{x} < a < x_0 < b < \bar{x}$, the probability that X will reach a before b (resp. b before a) starting from x_0 is $\left(S(b;\theta) - S(x_0;\theta)\right)/\left(S(b;\theta) - S(a;\theta)\right)$ (resp. one minus this number). Taking the limit $b \to \bar{x}^-$ and $a \to \underline{x}^+$ respectively, we see that under Assumption 2.2 the probability that X will reach either boundary of D_X in finite time is zero.

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 $\kappa \geq 1$ when $\alpha = 1$ (in fact, $\kappa \geq 1/2$ would be enough to obtain an entrance boundary, but we have also required that $\kappa \geq 1$ to insure that $\lim_{y\to 0^+} \lambda_Y(y;\theta) < +\infty$ since $\lambda_Y(y;\theta) = \kappa(1-\kappa)y^{-2}$ if $\mu_Y(y;\theta) = \kappa y^{-1}$). 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., $\underline{y} = 0$ too is unattainable.

Step 3—Boundary classification for $\bar{y} = +\infty^*$: The boundary $+\infty$ is a natural boundary when $\Sigma_\infty = N_\infty = \infty$, and an entrance boundary when $\Sigma_\infty = \infty$ and $N_\infty < \infty$ (see, e.g., Karlin and Taylor (1981, Table 6.2)), where

$$N_\infty \equiv \int_y^{+\infty}\left\{\int_y^v s_Y(u;\theta)\,du\right\}m_Y(v;\theta)\,dv = \int_y^{+\infty}\left\{\int_u^{+\infty} m_Y(v;\theta)\,dv\right\}s_Y(u;\theta)\,du. \tag{A.6}$$

Under Assumption 3, consider first the case where there exists $E > 0$ such that $-Ky \leq \mu_Y(y;\theta) \leq Ky$ for all $y \geq E$. We then have

$$\begin{aligned} N_\infty &= \int_y^{+\infty}\left\{\int_u^{+\infty} m_Y(v;\theta)\,dv\right\}m_Y^{-1}(u;\theta)\,du = \int_y^{+\infty}\int_u^{+\infty} e^{\int_u^v 2\mu_Y(w;\theta)dw}\,dv\,du \\ &\geq \int_y^{+\infty}\int_u^{+\infty} e^{-\int_u^v 2Kw dw}\,dv\,du = \int_y^{+\infty}\left\{\int_u^{+\infty} e^{-Kv^2}\,dv\right\}e^{Ku^2}\,du = +\infty \end{aligned} \tag{A.7}$$

as in (A.3). If instead we have $\mu_Y(y;\theta) \leq -Ky^\beta$, $\beta > 1$, for all $y \geq E$, then

$$\begin{aligned} N_\infty &= \int_y^{+\infty}\int_u^{+\infty} e^{\int_u^v 2\mu_Y(w;\theta)dw}\,dv\,du \leq \int_y^{+\infty}\int_u^{+\infty} e^{-\int_u^v 2Kw^\beta dw}\,dv\,du \\ &= \int_y^{+\infty}\left\{\int_u^{+\infty} e^{-\zeta v^{\beta+1}}\,dv\right\}e^{\zeta u^{\beta+1}}\,du \end{aligned} \tag{A.8}$$

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

$$\begin{aligned} \int_u^{+\infty} e^{-\zeta v^{\beta+1}}\,dv &= \int_u^{+\infty} v^{-\beta}v^{\beta}e^{-\zeta v^{\beta+1}}\,dv \\ &= (\zeta(\beta+1))^{-1}u^{-\beta}e^{-\zeta u^{\beta+1}} - \zeta^{-1}(\beta+1)^{-2}\int_u^{+\infty} v^{-\beta-1}e^{-\zeta v^{\beta+1}}\,dv; \end{aligned}$$

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

$$N_\infty \leq \int_y^{+\infty}\left\{\int_u^{+\infty} e^{-\zeta v^{\beta+1}}\,dv\right\}e^{\zeta u^{\beta+1}}\,du < (2K)^{-1}\int_y^{+\infty} u^{-\beta}e^{-\zeta u^{\beta+1}}e^{\zeta u^{\beta+1}}\,du < +\infty. \tag{A.9}$$

Step 4—Boundary classification for $\underline{y} = 0$: Among unattainable boundaries (i.e., given that $\Sigma_0 = \infty$), whether 0 is an entrance or a natural boundary depends upon whether $N_0 < \infty$ or $N_0 = \infty$ respectively, where

$$N_0 \equiv \int_0^y\left\{\int_v^y s_Y(u;\theta)\,du\right\}m_Y(v;\theta)\,dv = \int_0^y\left\{\int_0^u m_Y(v;\theta)\,dv\right\}s_Y(u;\theta)\,du. \tag{A.10}$$

We have in all cases $\mu_Y(w;\theta) \geq \kappa w^{-1}$ for some $\kappa > 0$ (since if $\alpha > 1$, $\mu_Y(w;\theta) \geq \kappa w^{-\alpha} > \kappa w^{-1}$; note that this constant κ is not necessarily $\geq 1/2$). Then we can bound N_0 as follows:

$$\begin{aligned} N_0 &= \int_0^y\int_0^u \exp\left\{\int_u^v 2\mu_Y(w;\theta)dw\right\}dv\,du = \int_0^y\int_0^u \exp\left\{-\int_v^u 2\mu_Y(w;\theta)\,dw\right\}dv\,du \\ &\leq \int_0^y\int_0^u e^{-\int_v^u 2\kappa/w dw}\,dv\,du = \int_0^y\left\{\int_0^u v^{2\kappa}\,dv\right\}u^{-2\kappa}\,du = (2\kappa+1)^{-1}\int_0^y\left\{u^{2\kappa+1}\right\}u^{-2\kappa}\,du \\ &= (2\kappa+1)^{-1}y^2/2 < +\infty. \end{aligned} \tag{A.11}$$

Therefore $\underline{y} = 0$ is an entrance boundary for all $\alpha \geq 1$.

PROOF OF PROPOSITION 2: *Step 1*—Existence of the transition density p_Y: Consider first the case where $D_Y = (-\infty, +\infty)$. The fact that Girsanov's Theorem can be applied to Y follows from Karatzas and Shreve (1991, 5.5.38); note that the explosion time of Y, T_Y, is infinity with probability 1 as was proved in Proposition 1. By Girsanov's formula, for every A in the usual σ-field,

$$\text{(A.12)} \qquad \operatorname{Prob}\big(Y_\Delta \in A \mid Y_0 = y_0; \theta\big) = E[M_\Delta \cdot 1(W_\Delta \in A) | W_0 = y_0]$$

where $1(\cdot)$ denotes the indicator function and the nonnegative supermartingale

$$\text{(A.13)} \qquad M_\Delta \equiv \exp\left\{\int_0^\Delta \mu_Y(W_\tau;\theta)\,dW_\tau - \frac{1}{2}\int_0^\Delta \mu_Y^2(W_\tau;\theta)d\tau\right\}$$

is in fact a martingale for all $\Delta > 0$. Setting $\Phi_Y(\Delta, y|y_0;\theta) \equiv E[M_\Delta | W_\Delta = y, W_0 = y_0]$, (A.12) becomes

$$\text{(A.14)} \qquad \operatorname{Prob}(Y_\Delta \in A \mid Y_0 = y_0; \theta) = \int_{-\infty}^{+\infty} 1(y \in A)\Phi_Y(\Delta, y \mid y_0;\theta)p_{\mathrm{BM}}(\Delta, y \mid y_0)\,dy$$

where $p_{\mathrm{BM}}(\Delta, y \mid y_0) = (2\pi\Delta)^{-1/2}\exp\big\{-(y-y_0)^2/(2\Delta)\big\}$. The existence of the transition density p_Y follows from (A.14), and is given by $p_Y(\Delta, y \mid y_0;\theta) = \Phi_Y(\Delta, y \mid y_0;\theta)p_{\mathrm{BM}}(\Delta, y \mid y_0)$. Integration by parts inside the conditional expectation defining Φ_Y and the scaling property of Brownian motion allows Φ_Y to be further simplified (see Gihman and Skorohod (1972, Chapter 3.13), Dacunha-Castelle and Florens-Zmirou (1986), or Rogers (1985)) so that

$$\text{(A.15)} \qquad p_Y(\Delta, y \mid y_0;\theta) = (2\pi\Delta)^{-1/2}e^{-(y-y_0)^2/2\Delta + \int_{y_0}^{y}\mu_Y(w;\theta)\,dw}E\left[e^{\Delta\int_0^1 \lambda_Y\left((1-u)y_0+uy+\Delta^{1/2}B_u;\theta\right)du}\right]$$

where $\{B_u/u \in [0,1]\}$ designates a Brownian Bridge with $B_0 = B_1 = 0$.

Step 2—Bound for p_Y: The strict positivity of p_Y (lower bound) follows from (A.15). From Assumption 3, we obtain $\int_{y_0}^{y}\mu_Y(w;\theta)dw \leq H + L|y-y_0|(1+|y_0|) + Q(y-y_0)^2$ for all (y, y_0) in D_Y^2, where H, L, and Q 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 $Q = K$. This is an upper bound for the integral itself, not its absolute value. Then by the continuity of $\lambda_Y(w;\theta)$ in w, and its limit behavior near the boundaries under Assumption 3, it follows that there exists $\gamma \geq 0$ such that $\lambda_Y(w;\theta) \leq \gamma$ for all $w > 0$ and $\theta \in \Theta$ (in general, however, λ_Y will not be bounded below). Therefore

$$\text{(A.16)} \qquad E\left[\exp\left\{\Delta\int_0^1 \lambda_Y\big((1-u)y_0 + uy + \Delta^{1/2}B_u;\theta\big)\,du\right\}\right] \leq e^{\gamma\Delta}.$$

Collecting all terms we have that

$$\begin{aligned}\text{(A.17)} \qquad p_Y(\Delta, y|y_0;\theta) &\leq (2\pi\Delta)^{-1/2}e^{-(y-y_0)^2/2\Delta + H + L|y-y_0|(1+|y_0|) + K(y-y_0)^2} \times e^{\gamma\Delta} \\ &\leq C_0\Delta^{-1/2}e^{-3(y-y_0)^2/8\Delta} \times e^{C_1|y-y_0||y_0| + C_2|y-y_0| + C_3|y_0| + C_4y_0^2}\end{aligned}$$

provided that $-1/(2\Delta) + Q \leq -3/(8\Delta)$, i.e., that $0 < \Delta \leq \overline{\Delta} \equiv (8Q)^{-1}$. 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 $\overline{\Delta}$, but this will not be necessary. Further, when $\mu_Y \leq 0$ near $+\infty$ and $\mu_Y \geq 0$ near $-\infty$, Q can be set to 0 in the bound for $\int_{y_0}^{y}\mu_Y(w;\theta)dw$ and hence $\overline{\Delta} = +\infty$ (in which case we could also replace $3/(8\Delta)$ by $1/(2\Delta)$).

Step 3—Differentiability of p_Y: Suppose for now that we are allowed to differentiate under the expectation sign in (A.15). It follows from the assumed smoothness of μ and σ (hence μ_Y and λ_Y)

that

$$\text{(A.18)} \quad \frac{\partial p_Y(\Delta, y \mid y_0; \theta)}{\partial y} = (2\pi\Delta)^{-1/2} e^{\frac{(y-y_0)^2}{2\Delta} + \int_{y_0}^{y} \mu_Y(w;\theta)\,dw}$$
$$\times \left\{ \left\{ -\frac{(y-y_0)}{\Delta} + \mu_Y(y;\theta) \right\} E\left[e^{\Delta \int_0^1 \lambda_Y\left((1-u)y_0 + uy + \Delta^{1/2} B_u;\theta\right) du} \right] \right.$$
$$+ E\left[\Delta \int_0^1 u \lambda'_Y\big((1-u)y_0 + uy + \Delta^{1/2} B_u;\theta\big)\,du \right.$$
$$\left.\left. \times\, e^{\Delta \int_0^1 \lambda_Y\left((1-u)y_0 + uy + \Delta^{1/2} B_u;\theta\right) du} \right] \right\}$$

where $\lambda'_Y(w;\theta) \equiv \partial\lambda_Y(w;\theta)/\partial w$. The functions under the expectations depend continuously on y and I will now show that they are bounded by variables having constant expectation themselves. By uniform convergence, differentiating under the expectation will then have been legitimate and result in a continuous derivative. First, we have $|-(y-y_0)/\Delta + \mu_Y(y;\theta)| \le Q_1(|y|,|y_0|)$ where Q_1 is a polynomial of degree one in $(|y|,|y_0|)$, with coefficients uniformly bounded in $\theta \in \Theta$. Second $\left|E\left[Ae^B\right]\right| \le E\left[|A|e^B\right]$ combined with (A.16) imply

$$\text{(A.19)} \quad \left| E\left[\Delta \int_0^1 u\lambda'_Y\left((1-u)y_0 + uy + \Delta^{1/2}B_u;\theta\right) du \times e^{\Delta\int_0^1 \lambda_Y\left((1-u)y_0+uy+\Delta^{1/2}B_u;\theta\right)du} \right] \right|$$
$$\le \Delta E\left[\int_0^1 u \left| \lambda'_Y\left((1-u)y_0 + uy + \Delta^{1/2}B_u;\theta\right) \right| du \right] e^{\gamma\Delta}.$$

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

$$\text{(A.20)} \quad E\left[\int_0^1 u \left| \lambda'_Y\big((1-u)y_0 + uy + \Delta^{1/2}B_u;\theta\big) \right| du \right]$$
$$\le GE\left[\int_0^1 u e^{|(1-u)y_0 + uy + \Delta^{1/2}B_u|}\,du \right]$$
$$= G\int_0^1 uE\left[e^{|(1-u)y_0+uy+\Delta^{1/2}B_u|}\right] du \le G\int_0^1 ue^{|(1-u)y_0|+|uy|}E\left[e^{\Delta^{1/2}|B_u|}\right] du.$$

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 $2(2\pi)^{-1/2}\sigma^{-1}\exp\{-x^2/2\sigma^2\}$, $x \ge 0$. Therefore for any constant a:

$$\text{(A.21)} \quad E[e^{a|N|}] = 2(2\pi)^{-1/2}\sigma^{-1}\int_0^{+\infty} e^{ax}e^{-x^2/2\sigma^2}\,dx = 2(2\pi)^{-1/2}\sigma^{-1}e^{a^2\sigma^2/2}\int_0^{+\infty} e^{-(x-a\sigma^2)^2/2\sigma^2}\,dx$$
$$= e^{a^2\sigma^2/2}(2\pi)^{-1/2}\sigma^{-1}\int_{-\infty}^{+\infty} e^{-(x-a\sigma^2)^2/2\sigma^2}\,dx = e^{a^2\sigma^2/2}$$

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

$$\text{(A.22)} \quad E\left[\int_0^1 u \left| \lambda'_Y\big((1-u)y_0 + uy + \Delta^{1/2}B_u;\theta\big) \right| du \right] \le G\int_0^1 ue^{(1-u)|y_0|+u|y|+\Delta u(1-u)/2}\,du \le Ge^{|y_0|+|y|}$$

(since u runs from 0 to 1) and we obtain (2.5) for all $0 < \Delta < \overline{\Delta}$, where the constant D_0 is uniform in θ and P is a polynomial of finite degree with coefficients also uniform in θ.

Step 4—Consider finally (briefly) the case where $D_Y = (0, +\infty)$. What is required in the proof of Theorem 1 is to show that the integral $\int e^{w^2/2}\{\partial p_Z(\Delta, w|y_0;\theta)/\partial w\}^2 dw$ converges. That is, after a change of variable $Z \to Y$, we need to show that the integral

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

converges at both boundaries 0^+ and $+\infty$. The boundary 0^+ is either an entrance or a natural boundary for Y, and in both cases $\lim_{y\to 0^+} \partial p_Y(\Delta, y|y_0;\theta)/\partial y = 0$ (see McKean (1956, Remark 4.2, page 541). Hence the integral converges at the left boundary 0. The change of measure in Step 1 above is no longer applicable in its simplest form, because the distribution of Y and that of a Brownian motion 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 still transform Y into a Brownian motion, but the Radon-Nikodym derivative is only a local martingale instead of a martingale. Girsanov's Theorem now gives for $y > 0$, $y_0 > 0$:

$$\text{(A.23)} \qquad p_Y(\Delta, y\,|\,y_0;\theta) = p_{\mathrm{BM}}(\Delta, y\,|\,y_0) e^{\int_{y_0}^{y} \mu_Y(w;\theta)\,dw} E\big[e^{\int_0^\Delta \lambda_Y(W_u;\theta)\,du}\big| W_\Delta = y, W_0 = y_0, \Delta < T_0\big]$$

where inside the expectation W follows the law of a Brownian motion and T_0 indicates the first time W hits 0. From (A.23), the same bounds can be derived.

PROOF OF COROLLARY 1: The existence and unicity in law of a solution of (1.1) follows, as in Proposition 1, from an application of Theorem 5.5.15 in Karatzas and Shreve (1991) replacing $\mathbb{R}$ by D_X throughout. Note that $\sigma(x;\theta) > 0$ for every x in D_X and θ in Θ; hence the nondegeneracy condition of the theorem is fulfilled (the only possible local degeneracy of σ, if any, occurs as $x \to 0^+$, but $0 \notin D_X$). The continuity of μ and σ implies the local integrability requirements for $|\mu|/\sigma^2$ and $1/\sigma^2$. Explosions are ruled out by showing that $\mathrm{Prob}(T_X = \infty) = 1$. This in turn follows from the fact that $Y_t = \gamma(X_t;\theta)$. The fact that $\gamma(x;\theta)$ tends to one of the boundaries of D_Y when x tends to one of the boundaries of D_X means that X would not be able to reach one of its boundaries without Y also doing so. But we already know that Y cannot do it (recall Proposition 1). Hence X cannot explode. Finally, the existence of p_X and its derivatives follows from the Jacobian formula; specifically $p_X(\Delta, x\,|\,x_0;\theta) = \sigma(x;\theta)^{-1} p_Y(\Delta, \gamma(x;\theta)\,|\,\gamma(x_0;\theta);\theta)$ and the differentiability of p_Y proved in Proposition 2 (and of course the differentiability of σ and γ which results from Assumption 2) extend to p_X.

PROOF OF THEOREM 1: *Step 1*—Let $\bar{\Delta} > 0$ be the constant defined in Proposition 2 (possibly $\bar{\Delta} = \infty$). 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 that contains the values of $\gamma(x_0;\theta)$ as x_0 varies in A_X and θ in the closure of Θ (recall that Θ is bounded). Define $\zeta(\Delta, x|x_0;\theta) \equiv \Delta^{-1/2}(\gamma(x;\theta) - \gamma(x_0;\theta))$. We seek to bound:

$$\text{(A.24)} \qquad \begin{aligned} |p_X(\Delta, x|x_0;\theta) - p_X^{(J)}(\Delta, x|x_0;\theta)| &= \sigma(x;\theta)^{-1}\Delta^{-1/2}|p_Z(\Delta, \zeta(\Delta, x|x_0;\theta)|\gamma(x_0;\theta);\theta) \\ &\quad - p_Z^{(J)}(\Delta, \zeta(\Delta, x|x_0;\theta)|\gamma(x_0;\theta);\theta)|. \end{aligned}$$

For that purpose, we will bound the jth coefficient in the approximating function $p_Z^{(J)}$. The $\eta_Z^{(J)}$'s in (2.12) are well-defined since by (2.4), the moments $u_Y(\Delta|y_0;\theta, j) \equiv \int_{-\infty}^{+\infty} |y|^j p_Y(\Delta, y|y_0;\theta)\,dy$ are finite for all $j \geq 0$ as a result of

$$\text{(A.25)} \qquad u_Y(\Delta|y_0;\theta, j) \leq e^{\{C_3|y_0| + C_4 y_0^2\}} \frac{C_0}{\Delta^{1/2}} \int_{-\infty}^{+\infty} |w - y_0|^j e^{\{-3w^2/8\Delta + C_1|y_0|\,|w| + C_2|w|\}}\,dw$$

where the variable y has been changed 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 $-3w^2/8\Delta + C_1|y_0|\,|w| + C_2|w| \leq -5w^2/16\Delta$.

Next, integration by parts with $(j+1)H_j(z) = -dH_{j+1}(z)/dz$ yields

$$\text{(A.26)} \qquad \begin{aligned} \eta_Z^{(j)}(\Delta, y_0;\theta) &= (j!)^{-1} \int_{-\infty}^{+\infty} H_j(w) p_Z(\Delta, w|y_0;\theta)\,dw \\ &= -(j!)^{-1}(j+1)^{-1} \int_{-\infty}^{+\infty} H_{j+1}'(w) p_Z(\Delta, w\,|\,y_0;\theta)\,dw \\ &= -((j+1)!)^{-1} H_{j+1}(w) p_Z(\Delta, w|y_0;\theta)\Big]_{-\infty}^{+\infty} \\ &\quad + ((j+1)!)^{-1} \int_{-\infty}^{+\infty} H_{j+1}(w)\{\partial p_Z(\Delta, w|y_0;\theta)/\partial w\}\,dw. \end{aligned}$$

With $y = y_0 + \Delta^{1/2} w$ and (2.4), we have that

$$0 < p_Z(\Delta, w|y_0;\theta) \leq a_0 \exp\{-3w^2/8\} \exp\{a_1|w||y_0| + a_2|w| + a_3|y_0| + a_4 y_0^2\} \tag{A.27}$$

where the constants $a_i, i = 0, \ldots, 4$, are uniform in $\theta \in \Theta$. By Theorem II in Stone (1928), there exists a constant K such that for all z in R and every integer j, $|H_j(z)| \leq K(j!)^{1/2} j^{-1/4}\{1 + 2^{-5/4}|z^{5/2}|\} e^{z^2/4}$. Therefore

$$\begin{aligned} &\left|((j+1)!)^{-1} H_{j+1}(w) p_Z(\Delta, w \mid y_0;\theta)\right| \\ &\quad \leq ((j+1)!)^{-1/2}(j+1)^{-1/4} K\{1 + |w^{5/2}/2^{5/4}|\} e^{w^2/4} a_0 e^{-3w^2/8} e^{a_1|w||y_0| + a_2|w| + a_3|y_0| + a_4 y_0^2} \end{aligned} \tag{A.28}$$

and hence $((j+1)!)^{-1} H_{j+1}(w) p_Z(\Delta, w|y_0;\theta)\big]_{-\infty}^{+\infty} = 0$.

Step 2—Proof that the expansion $p_Z^{(J)}$ of p_Z converges: Define

$$\nu_j(\Delta, y_0;\theta) \equiv (j!)^{-1} \int_{-\infty}^{+\infty} H_j(w)\{\partial p_Z(\Delta, w|y_0;\theta)/\partial w\} dw. \tag{A.29}$$

We can bound the terms of order $j \geq 1$ in the series for p_Z according to

$$\begin{aligned} |\eta_Z^{(j)}(\Delta, y_0;\theta) H_j(z)| &= ((j+1)!)^{-1} \left| \int_{-\infty}^{+\infty} H_{j+1}(w)\{\partial p_Z(\Delta, w \mid y_0;\theta)/\partial w\}\, dw \right| |H_j(z)| \\ &= |\nu_{j+1}(\Delta, y_0;\theta)||H_j(z)| \\ &\leq K\{1 + |z^{5/2}/2^{5/4}|\} e^{z^2/4} \times \{j^{-1/4}(j!)^{1/2}|\nu_{j+1}(\Delta, y_0;\theta)|\} \\ &\leq K\{1 + |z^{5/2}/2^{5/4}|\} e^{z^2/4} \times \{j^{-1/4}(j+1)^{-1/2}((j+1)!)^{1/2}|\nu_{j+1}(\Delta, y_0;\theta)|\} \\ &\leq K\{1 + |z^{5/2}/2^{5/4}|\} e^{z^2/4}\{j^{-1/2}(j+1)^{-1} + (j+1)!\nu_{j+1}^2(\Delta, y_0;\theta)\}/2 \end{aligned} \tag{A.30}$$

since $|\alpha\beta| \leq (\alpha^2 + \beta^2)/2$. The first series on the right-hand side, $\sum j^{-1/2}(j+1)^{-1}$, is convergent. It remains to prove that the series $\sum_{j=0}^{\infty} j!\nu_j^2(\Delta, y_0;\theta)$ converges. The integral $\int_{-\infty}^{+\infty} e^{w^2/2}\{\partial p_Z(\Delta, w \mid y_0;\theta)/\partial w\}^2 dw$ converges, since from (2.5) one can conclude that:

$$|\partial p_Z(\Delta, w \mid y_0;\theta)/\partial w| \leq b_0 e^{-3w^2/8} R(|w|, |y_0|) e^{b_1|w||y_0| + b_2|w| + b_3|y_0| + b_4 y_0^2} \tag{A.31}$$

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

Then expand the squared term in

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

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.

Hence $p_Z^{(J)}(\Delta, z \mid y_0; \theta) = \phi(z)\sum_{j=0}^{J} \eta_Z^{(j)}(\Delta, y_0; \theta)H_j(z)$ is convergent as $J \to \infty$. Note that the convergence is uniform in z over the entire real line since the two series in (A.30) are independent of z and hence converge 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.

Step 3—Proof that the limit of $p_Z^{(J)}(\Delta, z \mid y_0; \theta)$ (which we now know exists) 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 $\varepsilon_{j+1} \equiv j^{-1/2}(j+1)^{-1} + (j+1)!\nu_{j+1}^2(\Delta, y_0; \theta)$, note that there exists a constant K_0 such that

$$\text{(A.32)} \qquad \phi(z)\left|\eta_Z^{(j)}(\Delta, y_0; \theta)H_j(z)\right| \le K\{1+\left|z^{5/2}/2^{5/4}\right|\}e^{-z^2/4}\varepsilon_{j+1} \le K_0 e^{-3z^2/8}\varepsilon_{j+1}$$

(for z large enough) and hence q_Z satisfies the same bound as p_Z (which itself follows from that of p_Y in Proposition 2). Therefore the integral $(k!)^{-1}\int_{-\infty}^{+\infty} q_Z(\Delta, w \mid y_0; \theta)H_k(w)\, dw$ exists and

$$\text{(A.33)} \qquad (k!)^{-1}\int_{-\infty}^{+\infty} p_Z^{(J)}(\Delta, w \mid y_0; \theta)H_k(w)\, dw$$
$$= (k!)^{-1}\sum_{j=0}^{J} \eta_Z^{(j)}(\Delta, y_0; \theta)\int_{-\infty}^{+\infty}(2\pi)^{-1/2}e^{-w^2/2}H_j(w)H_k(w)\, dw$$
$$= \begin{cases} \eta_Z^{(k)}(\Delta, y_0; \theta) & \text{if } k \le J \\ 0 & \text{if } k > J \end{cases}$$

because $\int_{-\infty}^{+\infty}(2\pi)^{-1/2}e^{-w^2/2}H_j(w)\,H_k(w)\, dw = j!$ if $k = j$, and 0 otherwise (see, e.g., Sansone (1991, page 308)). Hence it follows 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 $k \ge 0$. To finish, consider two continuous functions satisfying the same first bound as in Proposition 2 and sharing the same η_k coefficients for all k: they 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 $k \ge 0$ 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.

Step 4—Back to p_X: I have shown that, for every $\varepsilon > 0$, there exists $J_\varepsilon(A_Y; \Theta)$ such that for all $J \ge J_\varepsilon(A_Y; \Theta)$, the bound $\left|p_Z(\Delta, z|y_0; \theta) - p_Z^{(J)}(\Delta, z|y_0; \theta)\right| \le \varepsilon$ holds for all $z \in R$, $y_0 \in A_Y$, and $\theta \in \Theta$. If σ is globally nondegenerate under Assumption 2(1), $\sigma^{-1}(x; \theta) < c^{-1} < +\infty$ implies that for all $J \ge J_\varepsilon(A_X; \Theta)$, $\left|p_X(\Delta, x|x_0; \theta) - p_X^{(J)}(\Delta, x|x_0; \theta)\right| \le \varepsilon$ for all x in R, $x_0 \in A_X$ and $\theta \in \Theta$. If not, 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 equation (A.24) implies

$$\text{(A.34)} \qquad \left|p_X(\Delta, x \mid x_0; \theta) - p_X^{(J)}(\Delta, x \mid x_0, \theta)\right|$$
$$\le c_\varepsilon^{-1}\Delta^{-1/2}\left|p_Z(\Delta, \zeta(\Delta, x|x_0; \theta)|\gamma(x_0; \theta); \theta) - p_Z^{(J)}(\Delta, \zeta(\Delta, x|x_0; \theta)|\gamma(x_0; \theta); \theta)\right|.$$

PROOF OF PROPOSITION 3: First verify that (3.4) holds. By Taylor's Theorem, we have

$$\text{(A.35)} \qquad \ell_n(\theta + I_n^{-1/2}(\theta)h_n) - \ell_n(\theta) = h_n I_n^{-1/2}(\theta)\dot{\ell}_n(\theta) + h_n^T I_n^{-1/2}(\theta)\ddot{\ell}_n(\tilde{\theta})I_n^{-1/2}(\theta)h_n/2$$

for every bounded sequence h_n such that $\theta + I_n^{-1/2}(\theta)h_n \in \Theta$, with $\tilde{\theta}_n$ between θ and $\theta + I_n^{-1/2}(\theta)h_n$. Now under Assumption 4, we have, again by Taylor's Theorem,

$$\text{(A.36)} \qquad \left\|I_n^{-1/2}(\theta)\ddot{\ell}_n(\tilde{\theta})I_n^{-1/2}(\theta) - G_n(\theta)\right\| \le \left\|I_n^{-1/2}(\theta)\dddot{\ell}_n(\tilde{\tilde{\theta}})I_n^{-1/2}(\theta)\right\| \times \left\|I_n^{-1/2}(\theta)\right\|$$

where the first term on the right-hand side is bounded in probability as the norm of $R_n(\theta, \tilde{\tilde{\theta}})$, while the second term, which arises because both $\tilde{\theta}$ and $\tilde{\tilde{\theta}}$ are in the same $I_n^{1/2}(\theta)$-neighborhood, goes to zero, so

$$\text{(A.37)} \qquad \ell_n(\theta + I_n^{-1/2}(\theta)h_n) - \ell_n(\theta) = h_n S_n(\theta) - h_n^T G_n(\theta)h_n/2 + o_p(1).$$

Therefore under (3.5) we have the LAMN structure (see, e.g., Jeganathan (1995, Definition 3, page 837)), and under (3.10) the LABF structure (see, e.g., Jeganathan (1995, Definition 4, page 850)).

By Taylor's Theorem applied to the score function, $\dot{\ell}_n(\theta) - \ell_n(\hat{\theta}) = -\ddot{\ell}_n(\tilde{\theta})(\hat{\theta}_n - \theta)$, i.e., $S_n(\theta) = I_n^{-1/2}(\theta)V_n(\theta) = I_n^{-1/2}(\theta)H_n(\tilde{\theta})I_n^{-1/2}(\theta)I_n^{1/2}(\theta)(\hat{\theta}_n - \theta)$ so

$$\text{(A.38)} \qquad I_n^{1/2}(\theta)(\hat{\theta}_n - \theta) = \left[I_n^{-1/2}(\theta)H_n(\tilde{\theta})I_n^{-1/2}(\theta)\right]^{-1} S_n(\theta)$$

and hence as in (A.36) we have

$$\text{(A.39)} \qquad I_n^{1/2}(\theta)(\hat{\theta}_n - \theta) - G_n^{-1}(\theta)S_n(\theta) = o_p(1).$$

Now both (3.6) and (3.11) follow from the joint convergence in distribution of (S_n, G_n) under LAMN and LABF respectively, and the Continuous Mapping Theorem (e.g., Hall and Heyde (1980, Theorem A.3, page 276) applied to (A.39). The efficiency statement (3.8) under LAMN follows from applying Theorem 3 in Basawa and Scott (1983, Chapter 2.4, Theorem 3, page 60); the Normal asymptotic variance comparison follows from Chapter 2.3, Corollary 2, page 53.

Under stationarity, the convergence in (3.5) follows from the Central Limit Theorem and the Law of Large Numbers (see, e.g., Hall and Heyde (1980)), and the fact that $E_\theta\left[n^{-1}H_n(\theta)\right] = n^{-1}i_n(\theta) = i(\theta)$, so

$$\text{(A.40)} \qquad \begin{aligned} G_n(\theta) &= I_n^{-1/2}(\theta)H_n(\theta)I_n^{-1/2}(\theta) = I^{-1/2}(\theta)\left[n^{-1}H_n(\theta)\right]I^{-1/2}(\theta) \\ &\xrightarrow{p} I^{-1/2}(\theta)i(\theta)I^{-1/2}(\theta) \equiv G(\theta). \end{aligned}$$

$G(\theta)$ is a nonrandom positive definite matrix provided that $i(\theta)$ is (which is guaranteed by (3.2)), and we obtain the classical result (3.9) (see, e.g., Billingsley (1961)).

I now show that the condition on $R_n(\theta, \tilde{\theta})$ in Assumption 4 is automatically satisfied under stationarity. In (A.15), let

$$\text{(A.41)} \qquad b(\theta) \equiv \int_{y_0}^{y} \mu_Y(w;\theta)\,dw, \quad f_u(\theta) \equiv \Delta\lambda_Y\big((1-u)y_0 + uy + \Delta^{1/2}B_u;\theta\big), \quad c(\theta) \equiv \int_0^1 f_u(\theta)\,du,$$

$$\text{(A.42)} \qquad q(\theta) \equiv \ln\big(p_Y(\Delta, y \mid y_0;\theta)\big) = -\ln(2\pi\Delta)/2 - (y-y_0)^2/2\Delta + b(\theta) + \ln\big(E[e^{c(\theta)}]\big),$$

and recall that μ_Y and λ_Y, hence f_u and c, are three times differentiable in θ under Assumption 2.1. From (A.42), it follows that

$$\text{(A.43)} \qquad \dot{q}(\theta) = \dot{b}(\theta) + \frac{E[\dot{c}(\theta)e^{c(\theta)}]}{E[e^{c(\theta)}]}, \quad \ddot{q}(\theta) = \ddot{b}(\theta) + \frac{E[(\ddot{c}(\theta) + \dot{c}(\theta)^2)e^{c(\theta)}]}{E[e^{c(\theta)}]} - \frac{E[\dot{c}(\theta)e^{c(\theta)}]^2}{E[e^{c(\theta)}]^2},$$

$$\text{(A.44)} \qquad \begin{aligned} \dddot{q}(\theta) &= \dddot{b}(\theta) + \frac{E[(\dddot{c}(\theta) + 3\dot{c}(\theta)\ddot{c}(\theta) + \dot{c}(\theta)^3)e^{c(\theta)}]}{E[e^{c(\theta)}]} \\ &\quad - \frac{3E[(\ddot{c}(\theta) + \dot{c}(\theta)^2)e^{c(\theta)}]E[\dot{c}(\theta)e^{c(\theta)}]}{E[e^{c(\theta)}]^2} + \frac{2E[\dot{c}(\theta)e^{c(\theta)}]^3}{E[e^{c(\theta)}]^3}, \end{aligned}$$

where for simplicity I use the same notation as if the parameter vector were one-dimensional. Let $v(\theta)$ play the role of $\dot{c}(\theta)$, $\ddot{c}(\theta) + \dot{c}(\theta)^2$, and $\dddot{c}(\theta) + 3\dot{c}(\theta)\ddot{c}(\theta) + \dot{c}(\theta)^3$ respectively and apply Hölder's Inequality, with $p = 4/3$ and $r = 4$, to

$$\text{(A.45)} \qquad |E[v(\theta)e^{c(\theta)}]| \le E[|v(\theta)|e^{c(\theta)}] = E[e^{c(\theta)/p}e^{c(\theta)/q}|v(\theta)|] \le E[e^{c(\theta)}]^{1/p}E[e^{c(\theta)}|v(\theta)|^q]^{1/q}.$$

Under Assumption 3, the function $v(\theta)$ has at most polynomial growth in w and as a result it follows from the same calculations as in (A.19)–(A.22), and the fact that $c(\theta) \le \gamma$ that

$$\text{(A.46)} \qquad E[e^{c(\theta)}|v(\theta)|^q]^{1/q} \le Ge^{\gamma + |y_0| + |y|}$$

(where G is a constant) for all the $v(\theta)$ listed above. We therefore have shown that $|E[v(\theta)e^{c(\theta)}]|/E[e^{c(\theta)}] \leq Ge^{\gamma+|y_0|+|y|}E[e^{c(\theta)}]^{1/p-1} \leq Ge^{\gamma+|y_0|+|y|}E[e^{c(\theta)}]^{-1/4}$. Recall next that $\dot{\mu}_Y$, $\ddot{\mu}_Y$, and $\dddot{\mu}_Y$ all have at most polynomial growth. Hence there exist a constant G and a finite order polynomial P such that

$$\text{(A.47)} \qquad \begin{cases} |\dot{q}(\theta)| \leq P(|y|,|y_0|) + Ge^{a(\gamma+|y|+|y_0|)}(E[e^{c(\theta)}]^{-1/4}), \\ |\ddot{q}(\theta)| \leq P(|y|,|y_0|) + Ge^{a(\gamma+|y|+|y_0|)}(E[e^{c(\theta)}]^{-1/4} + E[e^{c(\theta)}]^{-1/2}), \\ |\dddot{q}(\theta)| \leq P(|y|,|y_0|) + Ge^{a(\gamma+|y|+|y_0|)}(E[e^{c(\theta)}]^{-1/4} + E[e^{c(\theta)}]^{-1/2} + E[e^{c(\theta)}]^{-3/4}). \end{cases}$$

From this it follows that

$$\text{(A.48)} \qquad E\big[|\dot{q}(\theta)|\big|Y_0 = y_0\big] \leq \int_{\underline{y}}^{\bar{y}} \Big\{P(|y|,|y_0|) + Ge^{a(\gamma+|y|+|y_0|)}E[e^{c(\theta)}]^{-1/4}\Big\} \times (2\pi\Delta)^{-1/2}e^{-(y-y_0)^2/2\Delta+b(\theta)}E[e^{c(\theta)}]\,dy$$

is finite (the negative powers of $E[e^{c(\theta)}]$ get compensated), and similarly for $E[|\ddot{q}(\theta)|\big|Y_0 = y_0]$ and $E[|\dddot{q}(\theta)|\big|Y_0 = y_0]$. Hence $E[|\dddot{q}(\theta)|]$ is bounded, and by the Law of Large Numbers

$$\text{(A.49)} \qquad R_n(\theta,\tilde{\theta}) = I_n^{-1/2}(\theta)T_n(\tilde{\theta})I_n^{-1/2}(\theta) \xrightarrow{p} I^{-1/2}(\theta)E\big[\dddot{q}(\tilde{\theta})\big]I^{-1/2}(\theta)$$

which is a finite constant, uniformly bounded in $\tilde{\theta}$.

By using the top inequality in (A.47) and squaring it, it also follows as in (A.48) that $E[|\dot{q}(\theta)|^2\big|Y_0 = y_0]$ is bounded (the highest negative power becomes $(E[e^{c(\theta)}]^{-1/4})^2$) and therefore $i_n(\theta)$ is finite.

The fact that the derivatives of $\ln(p_X)$ are bounded follows from the bounds just given for the derivatives of $\ln(p_Y)$, and the differentiation chain rule applied to (2.9). Under Assumption 2, $1/\sigma$ is bounded (except possibly near a 0 boundary) and the function γ defined in (2.1) and its derivatives have at most polynomial growth. The same bounds as in (A.48) (with negative powers $E[e^{c(\theta)}]^{-1/2}$ and $E[e^{c(\theta)}]^{-3/4}$ similarly annihilated by $E[e^{c(\theta)}]$) apply to the second and third derivatives of $\ln(p_Y)$ with respect to y and y_0 rather than θ. Thus

$$\text{(A.50)} \qquad \frac{\partial \ln(p_X)}{\partial\theta} = -\frac{\dot{\sigma}}{\sigma} + \frac{\partial \ln(p_Y)}{\partial\theta} + \dot{\gamma}(x;\theta)\frac{\partial \ln(p_Y)}{\partial y} + \dot{\gamma}(x_0;\theta)\frac{\partial \ln(p_Y)}{\partial y_0}$$

(and the next two derivatives) are bounded similarly to (A.48) using the bounds for the derivatives of $\ln(p_Y)$ in (A.50).

PROOF OF THEOREM 2: *Step 1*—Fix $\varepsilon > 0$ and $x_0 \in R$. Let $r_X^{(J)}(\Delta, x|x_0;\theta) = p_X(\Delta, x|x_0;\theta) - p_X^{(J)}(\Delta, x|x_0;\theta)$,

$$\text{(A.51)} \qquad R_X^{(J)}(\Delta, x|x_0;\Theta) \equiv \sup_{\theta\in\Theta}\big|r_X^{(J)}(\Delta, x|x_0;\theta)\big|$$

and also define the corresponding quantities for Y and Z. 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_0 over bounded subsets A_Y of D_y. Hence there exists $J_\varepsilon(\Delta, A_Y;\theta)$ such that for all $J \geq J_\varepsilon(\Delta, A_Y;\Theta)$: $\sup_{\theta\in\Theta}\sup_{y\in D_Y}\sup_{y_0\in A_Y}|r_Y^{(J)}(\Delta, y|y_0;\theta)| < \varepsilon$. Now recall:

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

and for given x_0 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 2.1), A_Y is bounded. It follows that for all $J \geq J_\varepsilon(\Delta, A_Y;\Theta)$

$$\text{(A.52)} \qquad R_X^{(J)}(\Delta, x|x_0;\Theta) \leq \sup_{\theta\in\Theta}\big\{\sigma(x;\theta)^{-1}\big\} \times \sup_{\theta\in\Theta}\sup_{y\in D_Y}\sup_{y_0\in A_Y}\big|r_Y^{(J)}(\Delta, y|y_0;\theta)\big| \leq \sup_{\theta\in\Theta}\big\{\sigma(x;\theta)^{-1}\big\} \times \varepsilon.$$

Let $\Sigma^{-1}(x) \equiv \sup_{\theta\in\Theta}\{\sigma(x;\theta)^{-1}\}$, 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

$$\text{(A.53)}\qquad \left|E_{\theta_0}\left[\left\{R_X^{(J)}(\Delta, X_{t+\Delta}|X_t;\Theta)\right\}^m \big| X_t = x_0\right]\right| \le \int_{\underline{x}}^{\bar{x}} \left|R_X^{(J)}(\Delta, x|x_0;\Theta)\right|^m p_X(\Delta, x|x_0;\theta_0)\,dx$$
$$\le \varepsilon^m \int_{\underline{x}}^{\bar{x}} \Sigma^{-m}(x) p_X(\Delta, x|x_0;\theta_0)\,dx$$

i.e., $\lim_{J\to\infty} E_{\theta_0}\big[\{R_X^{(J)}(\Delta, X_{t+\Delta}|X_t;\Theta)\}^m \big| X_t = x_0\big] = 0$ 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.

Step 2—Bounding the integral in the right-hand side of (A.53): A difficulty only arises when $D_X = (0,+\infty)$, $\lim_{x\to 0^+}\sigma(x;\theta)=0$ (otherwise $\Sigma^{-m}(x)\le c^{-m}$ and then $\int_{\underline{x}}^{\bar{x}} \Sigma^{-m}(x)p_X(\Delta, x|x_0;\theta_0)dx \le c^{-m}$). Applying the change of variable $X\to Y$, I will prove convergence of the integral $\int_{\underline{y}}^{\bar{y}} \Sigma^{-m}(\gamma^{-1}(y;\theta))\times p_Y(\Delta, y|y_0;\theta)dy$ (Σ^{-1} means $1/\Sigma$ whereas γ^{-1} represents the inverse of the function γ). 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 2.1, we have that $\sigma^{-1}(x;\theta)\le \omega^{-1}x^{-\rho}$ for all $0<x\le\xi_0$ and $\theta\in\Theta$. For $0<x\le \xi_0$, we have $\int_0^x du/\sigma(u;\theta) \le \int_0^x \omega^{-1}u^{-\rho}\,du = \omega^{-1}(1-\rho)^{-1}x^{1-\rho}$ if $0\le\rho<1$, and therefore $\underline{y}=0^+$ by taking the limit as x tends to 0^+. Let $x=\gamma^{-1}(y;\theta)$, and I have just shown that for y near 0^+, $y\le \omega^{-1}(1-\rho)^{-1}x^{1-\rho}$, from which it follows that $\gamma^{-1}(y;\theta)\ge(\omega(1-\rho)y)^{1/(1-\rho)}$ and consequently

$$\text{(A.54)}\qquad \Sigma^{-m}(\gamma^{-1}(y;\theta)) \le \omega^{-m}[\gamma^{-1}(y;\theta)]^{-m\rho} \le \omega^{-m}(\omega(1-\rho)y)^{-m\rho/(1-\rho)}.$$

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 this question, we need to call upon Assumption 3.1. For $0<y\le\varepsilon_0$,

$$\text{(A.55)}\qquad e^{\int_{\varepsilon_0}^{y}\mu_Y(w;\theta)dw} = e^{-\int_y^{\varepsilon_0}\mu_Y(w;\theta)dw} \le e^{-\kappa\int_y^{\varepsilon_0} w^{-\alpha}dw} = \begin{cases}\varepsilon_0^{-\kappa}y^{\kappa} & \text{if } \alpha=1,\\ e^{\kappa(\alpha-1)\varepsilon_0^{-(\alpha-1)}-\kappa(\alpha-1)y^{-(\alpha-1)}} & \text{if } \alpha>1,\end{cases}$$

will provide an upper bound to p_Y for y near 0^+ (see the proof of Proposition 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

$$\text{(A.56)}\qquad \Sigma^{-m}(\gamma^{-1}(y;\theta)) \le \omega^{-m}[\gamma^{-1}(y;\theta)]^{-m\rho} \le \omega^{-m}(\omega(1-\rho)y)^{-m\rho/(1-\rho)}$$

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 $\kappa\ge 2\rho/(1-\rho)$. This is given by Assumption 5. If instead $\rho\ge 1$, then

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

and

$$\text{(A.57)}\qquad \lim_{x\to 0^+}\int_{\xi_0}^{x}\sigma^{-1}(u;\theta)\,du \le \lim_{x\to 0^+}\int_{\xi_0}^{x}\omega^{-1}u^{-\rho}\,du = \lim_{x\to 0^+}\begin{cases}\omega^{-1}\ln(x) & \text{if } \rho=1,\\ -\omega^{-1}(\rho-1)^{-1}x^{-(\rho-1)} & \text{if } \rho>1,\end{cases}$$

which is $-\infty$, 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 $y\le\omega^{-1}\ln(x)$ 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 (A.17), 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 $\gamma^{-1}(y;\theta)\ge(-\omega(\rho-1)y)^{-1/(\rho-1)}$, and thus

$\Sigma^{-m}(\gamma^{-1}(y;\theta)) \leq \omega^{-m}(-\omega(\rho-1)y)^{m\rho/(\rho-1)}$, 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 $\int_{\underline{y}}^{\bar{y}} \Sigma^{-m}(\gamma^{-1}(y;\theta))p_Y(\Delta, y\,|\,y_0;\theta)dy$ converges near $\underline{y}=-\infty$ when $\rho \geq 1$. Therefore from (A.53) we conclude that

$$\text{(A.58)} \qquad \lim_{J\to\infty} E_{\theta_0}\left[\{R_X^{(J)}(\Delta, X_{t+\Delta}|X_t;\Theta)\}^m\,|\,X_t=x_0\right]=0 \quad \text{for} \quad m=1,2.$$

Step 3—The convergence of its first two moments given by (A.58) to zero imply by Chebyshev's Inequality that the sequence $R_X^{(J)}(\Delta, X_{t+\Delta}|X_t;\Theta)$ converges to zero in probability, given $X_t=x_0$, that is:

$$\text{(A.59)} \qquad \lim_{J\to\infty} \text{Prob}\left(\left|R_X^{(J)}(\Delta, X_{t+\Delta}|X_t;\Theta)\right| > \varepsilon|X_t=x_0;\theta_0\right)=0.$$

By Bayes' Rule we have

$$\text{(A.60)} \qquad \text{Prob}\left(\left|R_X^{(J)}(\Delta, X_{t+\Delta}|X_t;\Theta)\right| > \varepsilon;\theta_0\right)$$
$$= \int_{-\infty}^{+\infty} \text{Prob}\left(\left|R_X^{J}(\Delta, X_{t+\Delta}|X_t;\Theta)\right| > \varepsilon\middle|X_t=x_0;\theta_0\right)\pi_t(x_0;\theta_0)dx_0$$

where $\pi_t(x_0;\theta_0)\equiv \partial\,\text{Prob}(X_t\leq x_0;\theta_0)/\partial x_0$ 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 $0\leq \text{Prob}\left(\left|R_X^{(J)}(\Delta, X_{t+\Delta}|X_t;\Theta)\right| > \varepsilon|X_t=x_0;\theta_0\right)\leq 1$ and $\int_{-\infty}^{+\infty}\pi_t(x_0;\theta_0)dx_0=1$ it follows from Lebesgue's Dominated Convergence Theorem (see, e.g., Haaser and Sullivan (1991, Theorem 6.8.6)) that

$$\text{(A.61)} \qquad \lim_{J\to\infty} \text{Prob}\Big(\left|R_X^{(J)}(\Delta, X_{t+\Delta}|X_t;\Theta)\right| > \varepsilon;\theta_0\Big)=0.$$

Step 4—Convergence as $J\to\infty$: I have now established that

$$p_X^{(J)}(\Delta, X_{t+\Delta}|X_t;\theta) \xrightarrow{p} p_X(\Delta, X_{t+\Delta}|X_t;\theta) \quad \text{as} \quad J\to\infty.$$

Before taking the logarithm of $p_X^{(J)}$, we need to trim it to insure that it is positive on the entire support D_X. The roots of the Hermite polynomials are such that $p_X^{(J)}>0$ on an interval $[-c_J;c_J]$ with $c_J\to\infty$ as $J\to\infty$. Let a_J be a positive sequence converging to 0 as $J\to\infty$. Define ω_J as a (smooth) version of the trimming index taking value 1 if $p_X^{(J)}>a_J$ and a_J otherwise. As a consequence of the bound (2.4), p_X is tight, i.e., for every $\varepsilon>0$ there always exists a compact space $K_\varepsilon\subset D_X$ that contains $1-\varepsilon$ of the mass of the density p_X. As a result, trimming by ω_J is asymptotically irrelevant (as $J\to\infty$), that is $\omega_J\xrightarrow{p}1$. It follows that $\omega_J p_X^{(J)}\xrightarrow{p}p_X$ and furthermore $\ln[\omega_J p_X^{(J)}(\Delta, X_{t+\Delta}\,|\,X_t;\theta)]\xrightarrow{p}\ln[p_X(\Delta, X_{t+\Delta}|X_t;\theta)]$ by the continuity of the logarithm. Thus for any given n we have obtained that $\ell_n^{(J)}(\theta)\xrightarrow{p}\ell_n(\theta)$ as $J\to\infty$, uniformly in θ. The convergence of the respective argmax in $\hat{\theta}_n^{(J)}\xrightarrow{p}\hat{\theta}_n$ as $J\to\infty$ is then an application of standard methods since $\ell_n^{(J)}(\theta)$ and $\ell_n(\theta)$ and their derivatives are both continuous in θ for all n and J. This proves part (i) of the Theorem.

Step 5—Convergence as $n\to\infty$. From Step 4, a value of J_n can be chosen for each n to make $|\hat{\theta}_n^{(J_n)}-\hat{\theta}_n|$ arbitrarily small in probability. In particular, one can select $J_n\to\infty$ such that $\hat{\theta}_n^{(J_n)}-\hat{\theta}_n=o_p(I_n^{-1/2}(\theta_0))$ as $n\to\infty$. This proves part (ii) of the theorem.

Proof of Proposition 4:[21] Under the assumptions made on μ_Y, the boundaries are entrance (see Proposition 1). The former insures stationarity of Y (see the discussion following Proposition 1). It is also the case that the scale measure $s_Y(v;\theta)\equiv\exp\{-\int^v 2\mu_Y(u;\theta)\,du\}$ diverges *exponentially* fast at each boundary. The spectrum of $A(\theta)$ is discrete, from Section 4.1 in Hansen, Scheinkman,

[21] I am grateful to Ernst Schaumburg for providing a key element in this proof.

and Touzi (1998).[22] That is, there exists a countable number of eigenvalues λ_p and eigenfunctions ϕ_p, $p = 0, 1, \ldots$, forming an orthonormal basis in $L^2(\pi_Y)$ such that for all functions f in the domain of $A(\theta)$, $A(\theta) \cdot f = \sum_{p=0}^{\infty} \lambda_p \langle f, \phi_p \rangle \phi_p$ where $\langle \cdot, \cdot \rangle$ denotes the natural inner product of $L^2(\pi_Y)$. Now polynomials f and their iterates (by repeated application of the generator) retain their polynomial growth characteristic near the boundaries; so they are all in $L^2(\pi_Y)$ and satisfy $\lim_{y \to \underline{y}} f'(y)/s_Y(y; \theta) = \lim_{y \to \bar{y}} f'(y)/s_Y(y; \theta) = 0$. This follows from the exponential divergence of $s_Y(y; \theta)$ near both boundaries whereas polynomials and their iterates diverge at most polynomially (recall that under Assumption 3 μ_Y and its derivatives have at most polynomial growth; multiplying and adding functions with polynomial growth yields a function still with polynomial growth). Using then the Hansen, Scheinkman, and Touzi (1998, page 10) characterization of the domain of the generator of a scalar diffusion, polynomials and their iterates are in the domain of the generator. Since f is in $L^2(\pi_Y)$, it follows that $f = \sum_{p=0}^{\infty} \langle f, \phi_p \rangle \phi_p$ with $\sum_{p=0}^{\infty} \langle f, \phi_p \rangle^2 < \infty$. Moreover, $A^k(\theta) \cdot f = \sum_{p=0}^{\infty} \lambda_p^k \langle f, \phi_p \rangle \phi_p$. Therefore:

$$\text{(A.62)} \qquad \sum_{k=0}^{K} A^k(\theta) \cdot f k!^{-1} \Delta^k = \sum_{k=0}^{K} \left(\sum_{p=0}^{\infty} \lambda_p^k \langle f, \phi_p \rangle \phi_p \right) k!^{-1} \Delta^k = \sum_{p=0}^{\infty} \left(\sum_{k=0}^{K} \lambda_p^k k!^{-1} \Delta^k \right) \langle f, \phi_p \rangle \phi_p$$

by Fubini's Theorem. Y being a time-reversible diffusion, its eigenvalues λ_p are all real and negative. Therefore $| \sum_{k=0}^{K} \lambda_p^k k!^{-1} \Delta^k | \le 1$, with limit $\exp(\lambda_p \Delta) \le 1$, and it follows that

$$\text{(A.63)} \qquad \sum_{p=0}^{\infty} \left(\sum_{k=0}^{K} \lambda_p^k k!^{-1} \Delta^k \right)^2 \langle f, \phi_p \rangle^2 \le \sum_{p=0}^{\infty} \langle f, \phi_p \rangle^2 < \infty$$

for all J and by the dominated convergence theorem the series $\sum_{k=0}^{K} A^k(\theta) \cdot f k!^{-1} \Delta^k$ converges as $K \to \infty$.

PROOF OF COROLLARY 2: *Step 1*—Part (i) follows directly from (3.9).

Step 2—Part (ii): With the notation $\varepsilon_{i\Delta} \equiv X_{i\Delta} - e^{-\beta\Delta} X_{(i-1)\Delta}$, the matrices $H_n(\theta)$, $I_n(\theta)$, and $G(\theta)$ of Section 3 can be calculated explicitly. It is easy to see that the terms of $H_n(\theta)$ are of the form:

$$\text{(A.64)} \qquad [H_n(\theta)]_{11} = a_1(\theta) n + a_2(\theta) \sum_{i=1}^{n} X_{(i-1)\Delta}^2 + a_3(\theta) \sum_{i=1}^{n} X_{(i-1)\Delta} \varepsilon_{i\Delta} + a_4(\theta) \sum_{i=1}^{n} \varepsilon_{i\Delta}^2,$$

$$\text{(A.65)} \qquad [H_n(\theta)]_{22} = b_1(\theta) n + b_2(\theta) \sum_{i=1}^{n} \varepsilon_{i\Delta}^2,$$

$$\text{(A.66)} \qquad [H_n(\theta)]_{12} = c_2(\theta) \sum_{i=1}^{n} X_{(i-1)\Delta} \varepsilon_{i\Delta} + c_3(\theta) \sum_{i=1}^{n} \varepsilon_{i\Delta}^2,$$

where a_k, b_k, and c_k are functions of the parameters θ. Now $E[\sum_{i=1}^{n} X_{(i-1)\Delta}^2]$ and $E[\sum_{i=1}^{n} \varepsilon_{i\Delta}^2]$ are asymptotically equivalent as $n \to \infty$ to $e^{-2n\beta\Delta} \sigma^2 / 2\beta(e^{-2\beta\Delta} - 1)$ and $n\sigma^2(e^{-2\beta\Delta} - 1)/2\beta$ respectively, and $E[\sum_{i=1}^{n} X_{(i-1)\Delta} \varepsilon_{i\Delta}] = 0$ (see White (1958) and Anderson (1959)). So to calculate an asymptotic equivalent for $I_n(\theta) = \operatorname{diag}\{E_\theta[H_n(\theta)]\}$ we only need $a_2(\theta) = 2\beta\Delta^2 e^{-2\beta\Delta} / \sigma^2(e^{-2\beta\Delta} - 1)$, $b_1(\theta) = -1/2\sigma^4$,

[22] Natural boundaries do not necessarily lead to a discrete spectrum (for example, in some instances a mixed discrete-continuous spectrum results). A stationary Ornstein-Uhlenbeck process (i.e., one with positive mean reversion) is an example of a process with natural boundaries and a discrete spectrum. What Proposition 4 shows is that having a discrete spectrum is a *sufficient* condition for the convergence of the series (4.3).

and $b_2(\theta)=\beta/2\sigma^6(e^{-2\beta\Delta}-1)$ to obtain that $[I_n(\theta)]_{11}$ is equivalent to $e^{-2(n+1)\beta\Delta}\Delta^2/(e^{-2\beta\Delta}-1)^2$ while $[I_n(\theta)]_{22}=n/2\sigma^4$. Finally,

$$G_n(\theta)=\begin{pmatrix}[H_n(\theta)]_{11}/[I_n(\theta)]_{11} & [H_n(\theta)]_{12}/\sqrt{[I_n(\theta)]_{11}[I_n(\theta)]_{22}}\\ [H_n(\theta)]_{12}/\sqrt{[I_n(\theta)]_{11}[I_n(\theta)]_{22}} & [H_n(\theta)]_{22}/[I_n(\theta)]_{22}\end{pmatrix}\xrightarrow{p}G(\theta)$$
$$=\begin{pmatrix}\chi^2[1] & 0\\ 0 & 1\end{pmatrix}, \tag{A.67}$$

because $\{2\beta(1-e^{-2\beta\Delta})/\sigma^2e^{-2n\beta\Delta}\}\sum_{i=1}^n X_{(i-1)\Delta}^2\xrightarrow{p}\chi^2[1]=N(0,1)^2$, while $\sum_{i=1}^n\varepsilon_{i\Delta}^2=O_p(n^{1/2})$ and $\sum_{i=1}^n X_{(i-1)\Delta}\varepsilon_{i\Delta}=O_p(e^{-n\beta\Delta})$. Therefore (5.2) follows, which is a non-Gaussian distribution under deterministic scaling by the asymptotic equivalent of $\sqrt{[I_n(\theta)]_{11}}$. Since

$$\text{Prob}\big([G^{-1/2}]_{11}\times N(0,1)\le z\big)=\text{Prob}\big(N(0,1)\le z[G^{1/2}]_{11}\big)=\int_0^{+\infty}\int_{-\infty}^{z\sqrt{g}}\frac{e^{-n^2/2}}{\sqrt{2\pi}}\frac{e^{-g/2}}{\sqrt{2\pi g}}\,dn\,dg \tag{A.68}$$

yields by differentiation with respect to z the density $1/\pi(1+z^2)$, $[G^{-1/2}]_{11}\times N(0,1)$ is the Cauchy distribution. Alternatively, we obtain a Gaussian distribution under random scaling by the asymptotic equivalent of $\sqrt{[H_n(\theta)]_{11}}$:

$$\sqrt{\frac{2\beta\Delta^2\sum_{i=1}^n X_{(i-1)\Delta}^2}{\sigma^2(e^{2\beta\Delta}-1)}}(\hat\beta_n-\beta)\xrightarrow{d}N(0,1). \tag{A.69}$$

Step 3—Part (iii): $[I_n(\theta_0)]_{11}$ is asymptotically equivalent to $\Delta\sigma^{-2}E\big[\sum_{i=1}^n X_{(i-1)\Delta}^2\big]$, i.e., to $n^2\Delta^2/2$, when $\beta=0$. Further,

$$[S_n(\theta_0)]_1=[I_n(\theta_0)]_{11}^{-1/2}\left\{2^{-1}n\Delta-\sigma^{-2}\sum_{i=1}^n X_{(i-1)\Delta}\varepsilon_{i\Delta}-2^{-1}\sigma^{-2}\sum_{i=1}^n\varepsilon_{i\Delta}^2\right\}, \tag{A.70}$$

$$[G_n(\theta_0)]_{11}=[I_n(\theta_0)]_{11}^{-1}\left\{6^{-1}n\Delta^2+\Delta\sigma^{-2}\sum_{i=1}^n X_{(i-1)\Delta}^2+\Delta\sigma^{-2}\sum_{i=1}^n X_{(i-1)\Delta}\varepsilon_{i\Delta}+\Delta 3^{-1}\sigma^{-2}\sum_{i=1}^n\varepsilon_{i\Delta}^2\right\}. \tag{A.71}$$

Thus $\big([S_n(\theta_0)]_1,[G_n(\theta_0)]_{11}\big)\xrightarrow{d}\big(2^{-1/2}(1-W_1^2),2\int_0^1 W_\tau^2d\tau\big)$ since from White (1958):

$$\left(\frac{\sum_{i=1}^n X_{(i-1)\Delta}\varepsilon_{i\Delta}}{2n\sigma^2\Delta},\frac{\sum_{i=1}^n X_{(i-1)\Delta}^2}{n^2\sigma^2\Delta}\right)\xrightarrow{d}\left(W_1^2-1,\int_0^1 W_\tau^2d\tau\right). \tag{A.72}$$

And (5.3) follows from (3.11) with $M_\tau=W_\tau$ and $2\int_0^1 W_\tau\,dW_\tau=\int_0^1 dW_\tau^2-\int_0^1 d\tau=W_1^2-1$. In this case the convergence $G_n\to G$ occurs in distribution but not in probability (which as discussed in the text would have been sufficient to insure an LAMN likelihood ratio structure). In both nonstationary cases ($\beta\le 0$), Assumption 4, including the boundedness condition on R_n, is verified explicitly from the exact expression of the likelihood function (whose second and third derivatives diverge at the same rate: differentiate once more with respect to θ the expressions (A.64)–(A.66)). Finally, when $\beta\le 0$ but not when $\beta>0$, the asymptotic distribution of the diffusion coefficient σ^2 is unaffected by the estimation of the drift, since the convergence rate of the latter is faster when $\beta\le 0$.

REFERENCES

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

——— (1996b): "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.

——— (1999): "Transition Densities for Interest Rate and Other Nonlinear Diffusions," *Journal of Finance*, 54, 1361–1395.

ANDERSON, T. W. (1959): "On Asymptotic Distributions of Parameters of Stochastic Difference Equations," *Annals of Mathematical Statistics*, 30, 676–687.

AZENCOTT, R. (1981): *Géodésiques et Diffusions en Temps Petit*, ed. by R. Azencott. Paris, France: Société Mathématique de France.

BASAWA, I. V., AND B. L. S. PRAKASA RAO (1980): *Statistical Inference for Stochastic Processes.* London, UK: Academic Press.

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

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

BLACK, F., AND M. SCHOLES (1993): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

COX, J. C. (1975): "The Constant Elasticity of Variance Option Pricing Model," published in a special issue of *The Journal of Portfolio Management,* 1996, 15–17.

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," Mimeo, Stanford University.

ERAKER, B. (1997): "MCMC Anaysis of Diffusion Models with Application to Finance," Mimeo, Norwegian School of Economics, Bergen.

FLORENS, D. (1999): "Estimation of the Diffusion Coefficient from Crossings," *Statistical Inference for Stochastic Processes*, 1, 175–195.

GALLANT, A. R., AND D. W. NYCHKA (1987): "Semi-nonparametric Maximum-Likelihood Estimation," *Econometrica*, 55, 363–390.

GALLANT, A. R., AND G. TAUCHEN (1996): "Which Moments to Match?" *Econometric Theory*, 12, 657–681.

GIHMAN, I. I., AND A. V. SKOROHOD (1972): *Stochastic Differential Equations.* New York, NY: Springer-Verlag.

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, NY: Dover.

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and its Applications.* San Diego, CA: Academic Press.

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 (1998): "Identification of Scalar Diffusions Using Eigenvectors," *Journal of Econometrics,* 86, 1–32.

JEGANATHAN, P. (1995): "Some Aspects of Asymptotic Theory with Applications to Time Series Models," *Econometric Theory*, 11, 818–887.

JENSEN, B., AND R. POULSEN (1999): "A Comparison of Approximation Techniques for Transition Densities of Diffusion Processes," Mimeo, Aarhus University.

JONES, C. S. (1997): "Bayesian Analysis of the Short-Term Interest Rate," Mimeo, The Wharton School, University of Pennsylvania.

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

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

KESSLER, M., AND M. SORENSEN (1999): "Estimating Equations Based on Eigenfunctions for a Discretely Observed Diffusion," *Bernoulli*, 5, 299–314.

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 *Advances in Econometrics, Sixth World Congress, Volume II*, ed. by C. Sims. Cambridge, UK: Cambridge University Press.

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, NY: Dover.

SANTA-CLARA, P. (1995): "Simulated Likelihood Estimation of Diffusion with an Application to the Short Term Interest Rate," Mimeo, UCLA.

STANTON, R. (1997): "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance*, 52, 1973–2002.

STONE, M. H. (1928): "Development in Hermite Polynomials," *Annals of Mathematics*, 29, 1–13.

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.

WHITE J. S. (1958): "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *Annals of Mathematical Statistics*, 29, 1188–1197.

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