

NONPARAMETRIC PRICING OF INTEREST RATE DERIVATIVE SECURITIES

BY YACINE AÏT-SHALIA¹

We propose a nonparametric estimation procedure for continuous-time stochastic models. Because prices of derivative securities depend crucially on the form of the instantaneous volatility of the underlying process, we leave the volatility function unrestricted and estimate it nonparametrically. Only discrete data are used but the estimation procedure still does not rely on replacing the continuous-time model by some discrete approximation. Instead the drift and volatility functions are forced to match the densities of the process. We estimate the stochastic differential equation followed by the short-term interest rate and compute nonparametric prices for bonds and bond options.

KEYWORDS: Estimation of stochastic differential equations, nonparametric kernel estimation, discrete-time sampling, option pricing, term structure of interest rates.

1. INTRODUCTION

DERIVATIVE SECURITIES ARE CONTINGENT CLAIMS whose payoffs depend upon another asset's payoff, or nontraded factor, such as an interest rate. Stock options, futures, swaps, caps, floors, bonds, callable bonds, convertibles, bond options, all fall into this category. Derivative securities are widely traded both over the counter and on exchanges, and, more often than not, the volume traded is much larger for the derivative securities than for the underlying assets that sustain them.

The theory of derivative security pricing relies essentially on continuous-time arbitrage arguments since the pioneering Black and Scholes (1973) paper. As further demonstrated by the work of Merton (e.g., Merton (1973, 1990)), pricing derivatives in the theoretical finance literature is generally much more tractable—as well as elegant—in a continuous-time framework than through binomial or other discrete approximations. The empirical option pricing literature, however, has not followed suit. It is typical there to abandon the continuous-time model altogether when estimating derivative pricing models. This paper develops tools to estimate the actual model used in the theoretical construction (which is in continuous-time) by using only the data available (which are in discrete-time). We will estimate the short-term interest rate process and subsequently price bonds and bond options.

¹ This paper is part of my Ph.D. dissertation at the MIT Department of Economics. I am very grateful to my advisors Jerry Hausman, Andrew Lo, and Whitney Newey, as well as John Cox, Lars Hansen, José Scheinkman, and Daniel Stroock for suggestions. The comments of a co-editor and three referees led to considerable improvement of the paper. I thank seminar participants at Berkeley, British Columbia, Carnegie-Mellon, Chicago, Columbia, Harvard, LSE, Michigan, MIT, Northwestern, Princeton, Stanford, Toulouse, ULB, Washington University, Wharton, and Yale for comments. Financial support from Ecole Polytechnique and MIT Fellowships is also gratefully acknowledged. All errors are mine.

The underlying process of interest $\{r_t, t \geq 0\}$ is a diffusion represented by the Itô stochastic differential equation

$$(1.1) \quad dr_t = \mu(r_t) dt + \sigma(r_t) dW_t$$

where $\{W_t, t \geq 0\}$ is a standard Brownian motion. The functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ are respectively the drift (or instantaneous mean) and the diffusion (or instantaneous variance) functions of the process. It has long been recognized in the finance literature that one of the most important features of (1.1) for derivative security pricing is the specification of the function $\sigma^2(\cdot)$.

As a consequence every model has tried to specify $\sigma^2(\cdot)$ correctly. To price interest rate derivatives, Vasicek (1977) specifies that the instantaneous volatility of the spot rate process is constant. The Cox-Ingersoll-Ross (1985b) (CIR hereafter) model of the term structure assumes that the instantaneous variance is a linear function of the level of the spot rate r , so the standard deviation is a square-root. There are many other models in the literature for the instantaneous variance of the short-term interest rate, all being in general mutually exclusive (see Table I). In the absence of any theoretical rationale for adopting

TABLE I
ALTERNATIVE SPECIFICATIONS OF THE SPOT INTEREST RATE PROCESS
 $dr_t = \mu(r_t) dt + \sigma(r_t) dW_t$

$\mu(r)$	$\sigma(r)$	Stationary	Reference
$\beta(\alpha - r)$	σ	Yes	Vasicek (1977)
$\beta(\alpha - r)$	$\sigma r^{1/2}$	Yes	Cox-Ingersoll-Ross (1985b) Brown-Dybvig (1986) Gibbons-Ramaswamy (1993)
$\beta(\alpha - r)$	σr	Yes	Courtadon (1982)
$\beta(\alpha - r)$	σr^λ	Yes	Chan et al. (1992)
$\beta(\alpha - r)$	$\sqrt{\sigma + \gamma r}$	Yes	Duffie-Kan (1993)
$\beta r(\alpha - \ln(r))$	σr	Yes	Brennan-Schwartz (1979) [one-factor]
$\alpha r^{-(1-\delta)} + \beta r$	$\sigma r^{\delta/2}$	Yes	Marsh-Rosenfeld (1983)
$\alpha + \beta r + \gamma r^2$	$\sigma + \gamma r$	Yes	Constantinides (1992)
β	σ	No	Merton (1973)
0	σr	No	Dothan (1978)
0	$\sigma r^{3/2}$	No	Cox (1975) Cox-Ingersoll-Ross (1980)

one particular specification of the diffusion over another, the question must ultimately be decided by taking the models to the data.

The statistical literature contains methods appropriate for a continuous-time record of observations (e.g., Basawa and Prakasa-Rao (1980), Florens-Zmirou (1993)). Nelson (1990) studies the limiting behavior of discrete approximations as the sampling interval goes to zero. Banon (1978) examined the estimation of the drift when the diffusion is either a constant or a known function. With discrete data, an approach based on maximum-likelihood estimation of the parameters is due to Lo (1988). This method, however, requires that a partial differential equation be solved numerically for each maximum-likelihood iteration—except in the few cases where it is known explicitly: see Pearson and Sun (1994) for an application to the CIR model. Duffie and Singleton (1993) propose a method of moments where sample paths are simulated for given parameter values and the moments computed (see also Gouriéroux, Monfort, and Renault (1993)). Parameter estimates make the simulated moments close to the sample moments. This requires that new sample paths be simulated every time the parameters are adjusted. For a particular parameterization of $\mu(\cdot)$ and $\sigma^2(\cdot)$, this method could be applied with sample paths simulated from (1.1). Hansen and Scheinkman (1995) derive theoretical moment conditions characterizing the infinitesimal generators of Markov processes.

A commonly used method to estimate (1.1) consists in first parameterizing $\mu(\cdot)$ and $\sigma^2(\cdot)$, then discretizing the model in order to estimate the parameters using, for example, Hansen (1982)'s General Methods of Moments (e.g., Chan et al. (1992)). Discretization-based methods implicitly assume that more data means more frequent data on a fixed period of observation. This hardly matches the way new data are added to the sample: we typically add today's interest rate at the end of the sample, not the 3 PM interest rate on January 15, 1992 in between two already existing observations in the sample. Even if such data were available, it is likely that market micro-structure problems, such as the bid-ask spread, the discreteness of the prices observed, and the irregularity of the intra-day sampling interval, would complicate considerably the analysis of high frequency data compared to daily or weekly data. Furthermore, identification of the drift is generally impossible on a fixed sampling period, no matter how small the sampling interval. In contrast to this approach, we will not require that the sampling interval shrink to obtain asymptotic properties for our estimators.

Our setup has three characteristic features. First, the precise form of the diffusion function of the interest rate process is crucial to price options, as demonstrated by the extensive literature trying to capture a correct specification. Second, it is hard to form an a priori idea of the functional form of the diffusion function, as the instantaneous volatility of financial series is not observed. Third, long time-series of daily data on spot interest rates are available. These three elements together, importance of the specification, lack of a priori information regarding that specification, and availability of the data, constitute the perfect setup to try to estimate the instantaneous volatility function nonparametrically. This is the objective of this paper. It is achieved by

density-matching: the instantaneous drift and diffusion functions of the short rate process are derived to be consistent with the observed distribution of the discrete data.

The paper is organized as follows. The first part examines the identification and estimation of the drift and diffusion functions of a continuous-time process. Section 2 identifies nonparametrically the diffusion given a restriction on the drift function. Section 3 then constructs the actual estimator, and describes its properties. The second part of the paper uses the results of the first to derive nonparametric option prices for interest rate derivative securities (Section 4). Section 5 computes these prices successively for discount bonds and options on discount bonds. Section 6 concludes. Technical assumptions and details are in Appendix 1–4, while proofs are in Appendix 5.

2. NONPARAMETRIC IDENTIFICATION OF THE DIFFUSION FUNCTION

2.1. Identification via Density-Matching

When estimating the CAPM, it is known empirically (see Merton (1980)) that the estimates of expected returns tend to have low precision when the observation period is finite, even though the diffusion can be estimated very precisely when the sampling interval is small. For example, suppose that a stock price were to follow $dX_t/X_t = \mu dt + \sigma dW_t$ with μ and σ constant. The maximum-likelihood estimate of μ from observations between dates 1 and n at interval Δ is the average of the log-returns, $\hat{\mu} = (1/n)\sum_{t=1}^n \ln(X_t)$. But $\hat{\mu} = (\ln(X_n) - \ln(X_1))/n$ is independent of the sampling interval Δ , and cannot be consistent as $\Delta \rightarrow 0$ for fixed n . The only hope to identify the drift μ of the continuous-time process from the data consists in letting the sample period n increase. We will therefore rely on increasing n to derive identification and estimation results.

Identifying without restrictions both the drift and diffusion functions from discretely sampled data is impossible in general. In particular, without further constraints, a pair of functions (μ, σ^2) cannot be distinguished from $(a\mu, a\sigma^2)$ for any constant a , with data discretely sampled at a fixed interval. To fully identify the process, we therefore impose a restriction on the form of the drift function so that we can leave the diffusion function unrestricted. Since nothing so far is parameterized, the drift function cannot be restricted simply by fixing some parameter values. Instead, the restriction operates at one extra level of generality. It takes an otherwise completely unrestricted drift function and makes it belong to a smaller class of functions, namely a parametric class.

We will then use an essential property of stochastic differential equations. Consider a discrete normal random variable. Its distribution is obviously entirely characterized by its first two moments, mean and variance. The continuous-time process r in (1.1) is not normally distributed in general. However, because the Brownian increments are Gaussian, it turns out that, under regularity conditions, an analogous property will hold for stochastic differential equations: the distributions of the process (marginal and transitional densities) are entirely

characterized by the first two moments of the process, which are here the drift and diffusion functions. By the Markov property, it suffices to consider only one set of transitions of the process; longer transitions can always be derived by iterating the shorter transitions.

In particular, the joint parameterizations of (μ, σ^2) adopted in the literature imply specific forms for the marginal and transitional densities of the process. For example, an Ornstein-Uhlenbeck process $dr_t = \beta(\alpha - r_t)dt + \gamma dW_t$ generates Gaussian transitional and marginal densities. The stochastic differential equation $dr_t = \beta(\alpha - r_t)dt + \gamma r_t^{1/2} dW_t$ yields a noncentral chi-squared transitional distribution, while the marginal density is a Gamma distribution (Feller (1951)). The former parameterization for the spot interest rate has been used by Vasicek (1977), while the latter has been derived in a general equilibrium framework by Cox-Ingersoll-Ross (1985a, 1985b). The functional forms of the transitional densities corresponding to specifications essentially different from these two are not known explicitly (Wong (1964) summarizes the only specifications where they are: linear μ and quadratic σ^2).

Our estimation approach relies on the equivalence between (μ, σ^2) and densities in the other direction: from the densities back to (μ, σ^2) . To price derivatives, the drift and diffusion of the short-term interest rate process need to be estimated. But they are not observed, nor can they be estimated directly. However, the densities of the process can be straightforwardly estimated from the data on the short-term rate. Therefore, instead of specifying parametrically both the drift and diffusion functions, and then accepting whatever marginal and transitional densities are implied by these choices (as the various parametric specifications do), we start with nonparametric estimates of the densities. Then we reconstruct the drift and diffusion of the continuous-time process by matching these densities, which we can, in turn, use for derivative pricing.

Our construction of the pair (μ, σ^2) that will match the densities is the following. Assume that the spot interest rate process $\{r_t, t \geq 0\}$ follows (1.1) under Assumptions A1–A2 in Appendix 1. These assumptions guarantee the existence and uniqueness of a strong stationary solution to (1.1), on $D = (0, \infty)$, since nominal interest rates are positive.² In particular, they rule out the possibility that starting from any interest rate level in $(0, \infty)$ the barriers $\underline{r} = 0$ and $\bar{r} = \infty$ could be attained in finite expected time. The drift $\mu(\cdot, \theta)$ depends on an unknown parameter vector θ , while the diffusion $\sigma^2(\cdot)$ is an unknown function. The available data consist of realizations of the process sampled at equally spaced discrete dates $1, \dots, n$. The continuous record of observations between each sampling date is unobservable. The asymptotic properties of the estimators as $n \rightarrow \infty$ are derived for an expanding sampling period, i.e., the interest rate is observed over a longer period of time, not by sampling more frequently. The sampling interval is fixed at $\Delta = 1$. Let $\pi(\cdot)$ be the marginal

² The Vasicek model has the undesirable feature of being distributed on $(-\infty, +\infty)$ as opposed to $(0, \infty)$.

density of the spot rate, and $p(\Delta, r_{t+\Delta}|r_t)$ the transition density function between two successive observations.

Consider the Kolmogorov forward equation (e.g., Karlin and Taylor (1981, p. 219)

$$(2.1) \quad \frac{\partial p(\Delta, r_{t+\Delta}|r_t)}{\partial \Delta} = - \frac{\partial}{\partial r_{t+\Delta}} (\mu(r_{t+\Delta}, \theta) p(\Delta, r_{t+\Delta}|r_t)) \\ + \frac{1}{2} \frac{\partial^2}{\partial r_{t+\Delta}^2} (\sigma^2(r_{t+\Delta}) p(\Delta, r_{t+\Delta}|r_t)).$$

In order to construct an estimator of σ^2 from the densities of the process, we use (2.1) to characterize the diffusion function. Note that by stationarity $\int_0^{+\infty} p(\Delta, r_{t+\Delta}|r_t) \pi(r_t) dr_t = \pi(r_{t+\Delta})$ has a partial derivative with respect to time equal to zero. Multiply therefore the forward equation (2.1) by the marginal density $\pi(r_t)$ and integrate through the equation with respect to the conditioning variable r_t to obtain the ordinary differential equation

$$(2.2) \quad \frac{d^2}{dr^2} (\sigma^2(r) \pi(r)) = 2 \frac{d}{dr} (\mu(r, \theta) \pi(r))$$

where $r = r_{t+\Delta}$. This equation must be satisfied at any point r in $(0, \infty)$, and at the true parameter value θ . Integrating (2.2) twice with the boundary condition $\pi(0) = 0$ now yields

$$(2.3) \quad \sigma^2(r) = \frac{2}{\pi(r)} \int_0^r \mu(u, \theta) \pi(u) du.$$

This equation shows that once the drift parameter vector θ has been identified, the diffusion function can be identified from the marginal distribution $\pi(\cdot)$. The identification of θ will be based on the transitional distribution. An identifying restriction on the drift is the linear mean-reverting specification $\mu(r_t, \theta) \equiv \beta(\alpha - r_t)$, $\theta \equiv (\alpha, \beta)$. Heuristically, the interest rate is elastically attracted to its equilibrium value α at a speed βdt . This is consistent with the parameterization of the drift used in most spot rate models in the literature and will make comparisons with these models possible. The first step consists of deriving

$$(2.4) \quad E[r_{t+\Delta}|r_t] = \alpha + e^{-\beta\Delta}(r_t - \alpha).$$

To see why this must hold, let $\vartheta(r, t)$ be the solution of the backward Kolmogorov equation $\partial \vartheta(r, t) / \partial t = A \vartheta(r, t)$ with initial condition $\vartheta(r, 0) = r$, where A is the backward Kolmogorov operator: $A \vartheta(r, t) \equiv \mu(r, \theta) \partial \vartheta(r, t) / \partial r + (1/2) \sigma^2(r) \partial^2 \vartheta(r, t) / \partial r^2$. Under Assumptions A1–A2, this partial differential equation has the unique solution $\vartheta(r, t) = E[r_t | r_0 = r]$, by Dynkin's formula (e.g., Karlin and Taylor (1981, p. 310)). Now it is easy to verify directly that the

function $\zeta(r, t) \equiv \alpha + e^{-\beta \Delta}(r - \alpha)$ also satisfies the equation with the same initial equation. Thus $\zeta = \vartheta$. Evaluating the equality at $t = \Delta$, and invoking stationarity, gives (2.4).

This procedure can be extended to cover the case of nonlinear restrictions on the drift. In that case, the only part of the identification and estimation procedure that has to be modified is (2.4): see Appendix 2. The drift parameters, once identified, can be plugged into (2.3), which remains valid for a generic drift function. We now use (2.4) to identify the parameter vector θ . Ordinary least squares (OLS) clearly identifies the parameters γ and δ in $E[r_{t+\Delta} - r_t | r_t] = \gamma + \delta r_t$, and therefore indirectly α and β . A discussion of efficiency is provided in the next section. Finally the diffusion function can be completely identified from the joint and marginal density functions $p(\cdot, \cdot | \cdot)$ and $\pi(\cdot)$ through (2.3)–(2.4). The drift is identified from the conditional mean and, given the drift, the marginal density yields the diffusion function.

THEOREM 1: *Under Assumptions A1–A2, the diffusion function of the underlying spot interest rate process can be identified from its joint and marginal densities by (2.3), where the unknown parameter vector θ is identified by (2.4).*

2.2. Diffusion for Large and Small Values of the Spot Rate

The behavior for $r \rightarrow +\infty$ of the diffusion function given by Theorem 1 is not immediate since the numerator and the denominator of the ratio (2.3) tend both to zero. The resulting effect can nevertheless be characterized by L'Hôpital's Rule as

$$(2.5) \quad \sigma^2(r) \cong_{r \rightarrow +\infty} 2\beta r \pi(r) / |\pi^{(1)}(r)|$$

where $\phi(r) \cong_{r \rightarrow +\infty} \varphi(r)$ means $\lim_{r \rightarrow +\infty} \phi(r) / \varphi(r) = 1$ and $\pi^{(1)}$ is the first derivative of π .

The asymptotic trend of the diffusion function is determined by the speed of decrease of the marginal density $\pi(\cdot)$ to zero: the faster the density goes to zero, the smaller the corresponding diffusion function. Suppose that today's spot rate is high, say 20%. The stationary data on the distribution of the process essentially say that it is very unlikely for the spot rate to be so high. Consider now the spot rate level tomorrow. The increment between today and tomorrow is given by (1.1). If a large level is reached today ($r_t = 20\%$), the mean-reverting drift term will be substantially negative ($\beta > 0$) and drive the process back toward lower levels (say $\alpha = 9\%$), which goes in the right direction. If, however, $\sigma(20\%)$ happens to be very large, then for given realizations of the Brownian increment the stochastic term might compensate the deterministic mean-reverting drift with substantial probability. This would generate a high level of the spot rate which would be incompatible with $\pi(\cdot)$ very close to zero above 20%.

The link between density and shape of the diffusion function can be made evident by applying successively (2.5) to a parametric family of densities which

exhibit an increasingly rapid rate of decrease to zero ($\nu > 1$ with $\omega > 0$):

$$(2.6a) \quad \pi(r) \equiv \xi r^{\nu-1} e^{-\omega r}$$

generates $\sigma^2(r) \cong_{r \rightarrow +\infty} (2\beta/\omega)r$ (the CIR case),

$$(2.6b) \quad \pi(r) \equiv \xi r^{\nu-1} e^{-\omega r^2}$$

generates $\sigma^2(r) \cong_{r \rightarrow +\infty} (\beta/\omega)$ (the Gaussian case),

$$(2.6c) \quad \pi(r) \equiv \xi r^{\nu-1} e^{-\omega r^3}$$

generates $\sigma^2(r) \cong_{r \rightarrow +\infty} (2\beta/3\omega)(1/r)$ (faster decay).

The behavior near zero of the diffusion function is also given by L'Hôpital's Rule as $\sigma^2(r) \cong_{r \rightarrow 0^+} 2\beta\alpha\pi(r)/|\pi^{(1)}(r)|$. So for all densities (2.6) we have: $\sigma^2(r) \cong_{r \rightarrow 0^+} 2\beta\alpha r/(\nu-1)$. Take densities polynomial in zero ($\pi(r) \cong_{r \rightarrow 0^+} \xi r^{\nu-1}$, $\nu > 1$). This encompasses most potential densities since π must be zero at zero and smooth. The result is then that every density polynomial in zero will yield a diffusion function with the same linear curvature near zero as the square-root process. Thus the CIR parameterization must be a reasonable approximation for small levels of the spot rate. Because the pricing equation for interest rate derivative securities has a strong local character, these securities would tend to be priced correctly by the CIR model when interest rates are low.

3. NONPARAMETRIC ESTIMATION OF THE DIFFUSION

We propose to replace θ and $\pi(\cdot)$ in (2.3) by consistent estimators to obtain an estimate of $\sigma^2(\cdot)$. Starting with $\pi(\cdot)$, we use the interest rate data $\{r_i, i = 1, \dots, n\}$ to form the smooth density estimator $\hat{\pi}(r) \equiv (1/nh_n) \sum_{i=1}^n K((r - r_i)/h_n)$, based on a kernel function $K(\cdot)$ and bandwidth h_n (see Silverman (1986) for an introduction to kernel density estimation and Scott (1992) for more details). Regularity conditions on the time-series dependence in the data (Assumption A3), the kernel (Assumption A4) and bandwidth (Assumption A5) are given in Appendix 1. The smooth estimator $\hat{\pi}(\cdot)$ can be used in the denominator of (2.3) as well as inside the integral, or only in the denominator with the density inside the integral replaced by an empirical density. The two resulting estimators of the diffusion function share the same asymptotic properties.

To estimate θ consistently, one can simply use OLS in (2.4). The efficiency of the estimator of θ will not matter for $\sigma^2(\cdot)$, because $\hat{\theta}$ will converge at speed $n^{1/2}$ while $\hat{\pi}(\cdot)$ in the denominator will force the convergence of $\hat{\sigma}^2(\cdot)$ to be slower. To obtain nevertheless a better drift estimator than OLS, useful later when pricing derivatives, consider the following procedure. We form the first step OLS estimator of $\theta = (\alpha, \beta)'$ by a one-to-one transformation from the OLS estimates of (γ, δ) : $\alpha = -\gamma/\delta$ and $\beta = -\ln(1 + \delta)/\Delta$. The OLS estimator of θ can now be plugged along with the kernel estimator $\hat{\pi}(\cdot)$ into (2.3) to form a nonparametric estimator of $\sigma^2(\cdot)$.

Then the diffusion estimator can be used to correct for heteroskedasticity in the residuals from the regression (2.4). We construct the weighting matrix for the second step feasible generalized least squares (FGLS) estimation of (γ, δ) and thus (α, β) . Note that no serial correlation adjustment is needed in (2.4). Any diffusion process is a Markov process so, in particular, $E[r_{t+\Delta}|r_t, r_{t-\Delta}] = E[r_{t+\Delta}|r_t]$. The residuals $\varepsilon_{t+\Delta} \equiv r_{t+\Delta} - (\gamma + (1 + \delta)r_t)$ from the regression $E[r_{t+\Delta} - r_t|r_t] = \gamma + \delta r_t$ are therefore conditionally uncorrelated; that is, $E[\varepsilon_{t+\Delta}\varepsilon_t|r_t, r_{t-\Delta}] = 0$. If the interest rate data were independent and identically distributed, this procedure would achieve the semiparametric efficiency bound of Chamberlain (1987) (see also Robinson (1987) for a two-stage setup). To our knowledge, there exists no known efficiency bound for dependent data. The following result details the asymptotic properties of the diffusion estimator.

THEOREM 2: *Under Assumptions A1–A5:*

(i) *The estimator $\hat{\sigma}^2(\cdot)$ is pointwise consistent and asymptotically normal:*

$$h_n^{1/2} n^{1/2} \{ \hat{\sigma}^2(r) - \sigma^2(r) \} \xrightarrow{d} N(0, V_{\sigma^2}(r)), \text{ with asymptotic variance}$$

$$(3.1) \quad V_{\sigma^2}(r) = \left\{ \int_{-\infty}^{+\infty} K(u)^2 du \right\} \sigma^4(r) / \pi(r).$$

(ii) *The asymptotic variance $V_{\sigma^2}(r)$ can be consistently estimated by*

$$(3.2) \quad \hat{V}_{\sigma^2}(r) = \left\{ \int_{-\infty}^{+\infty} K(u)^2 du \right\} \hat{\sigma}^4(r) / \hat{\pi}(r).$$

(iii) *At different points r and r' in $(0, \infty)$, $\hat{\sigma}^2(r)$ and $\hat{\sigma}^2(r')$ are asymptotically independent.*

The intuition behind the result is the following. There are three elements plugged-into the right-hand side of (2.3): $\hat{\theta}$, $\hat{\pi}(\cdot)$ inside the integral, and $\hat{\pi}(\cdot)$ in the denominator. The first two are responsible for $n^{1/2}$ terms, while the third generates a term converging at the slower speed $h_n^{1/2} n^{1/2}$. The fast terms can be considered as fixed when computing the asymptotic distribution. Only the slowest term will matter. Part (iii) of the theorem is typical of pointwise kernel estimators (see Robinson (1983)) and is useful to know for inference purposes. The consistent estimator of the pointwise asymptotic variance makes it possible to construct pointwise confidence intervals, or later obtain estimates of the variance of derivative security prices.

4. PRICING INTEREST-RATE-DERIVATIVE SECURITIES

4.1. Nonparametric Plug-In

Given the underlying spot interest rate process $\{r_t, t \geq 0\}$ described by a stochastic differential equation of the type (1.1), standard arbitrage arguments

(as for example in Vasicek (1977, Section 3)) determine the price $U(r, t)$ of any interest-rate-derivative security with time-to-maturity t and maturity date T when r is the spot interest rate. Let $\lambda(r)$ be the risk premium factor or market price of interest rate risk (which must be the same for all securities), $c(r, t)$ the cash flow rate paid by the security per unit of time, and $g(r)$ the payoff of the derivative security at maturity. $U(r, t)$ satisfies the following partial differential equation:

$$(4.1) \quad LU(r, t) = -c(r, t)$$

where L is the parabolic differential operator

$$(4.2) \quad LU \equiv -\partial U / \partial t + \{\sigma^2(r)/2\}(\partial^2 U / \partial r^2) \\ + \{\mu(r, \theta) - \lambda(r)\sigma(r)\}(\partial U / \partial r) - rU.$$

This is a Cauchy problem with initial condition $U(r, 0) = g(r)$. The continuous payment rate $c(\cdot, \cdot)$, the payoff $g(\cdot)$ at maturity as well boundary condition(s) depend on the particular security considered. Assumptions A6–A8 in Appendix 3 impose regularity conditions. Consider first a discount bond with face value equal to \$100 and maturing at date T . Let $U(r, t) \equiv B(r, t, T)$ be its price at date $(T - t)$ when the spot interest rate is r . This case corresponds to

$$(4.3) \quad c(r, t) = 0 \text{ (zero-coupon), } g(r) = 100 \\ \text{for all } r \geq 0 \text{ (initial condition),} \\ \lim_{r \rightarrow \infty} U(r, t) = 0 \text{ for all } t \geq 0 \text{ (boundary condition).}$$

Consider next a call option³ on a discount bond. The call option expires at date T , has an exercise (or strike) price X , and the underlying discount bond matures at date S where $T \leq S$. Let $U(r, t) \equiv C(r, t, T; s, X)$ be the call price at time-to-maturity t when the spot interest rate is r . Then

$$(4.4) \quad c(r, t) = 0, \quad g(r) = \max(0, B(r, S - T, S) - X) \text{ for all } r \geq 0, \\ \lim_{r \rightarrow \infty} U(r, t) = 0 \text{ for all } t \geq 0.$$

Other examples can be treated similarly. A coupon bond would be identical to (4.3), except that c is then the coupon rate received at t . An interest rate swap of r against an exogenous rate r' can be viewed as a contract paying at rate $c = r - r'$. A cap (resp. floor) is a loan at variable rate that is guaranteed to be less (resp. more) than some level \bar{r} (resp. r); it can be seen as a derivative with $c = \min(r, \bar{r})$ (resp. $c = \max(r, \underline{r})$). In all these cases g is the constant face value of the contract. A yield curve call option struck at X has $c = 0$ and $g(r) = \max(0, Y(n, T + n, r) - X)$ where $Y(n, T + n, r)$ is the yield-to-maturity at date T of an n -year bond. A binary yield option has $c = 0$ and $g(r) = 1$ ($Y(n, T + n, r) \geq X$). A power- p yield option has $c = 0$ and $g(r) = Y(n, T + n, r)^p$. A yield

³ It is never optimal to exercise early an American call because the underlying bond pays no coupon, so American and European calls on discount bonds have the same value.

curve slope call option has $c = 0$ and $g(r) = \max(0, (Y(n, T + n, r) - Y(m, T + m, r)) / (n - m) - X)$, for some $n > m$.

When $\sigma(r) = \sigma$ and $\lambda(r, t) = \lambda$ are constant, (4.1)–(4.3) and (4.1)–(4.4) have known solutions (Vasicek (1977) and Jamshidian (1989) respectively). When $\sigma(r) = \sigma r^{1/2}$ and $\lambda(r, t) = \lambda r^{1/2} / \sigma$, (4.1)–(4.3) and (4.1)–(4.4) also have known solutions (Cox-Ingersoll-Ross (1985b)). As was discussed earlier, these closed-form solutions all assume a particular parameterization of the diffusion term of the underlying spot interest rate. Since the assumptions are different, the pricing formulas differ. Hull and White (1990) compared them for a set of arbitrary values of the parameters (not estimated from actual data and time-dependent) and found that the formulas can differ by as much as 15% (see their Table 1).

We estimate derivative prices by solving the pricing partial differential equation with the drift parameters replaced by the estimates $\hat{\alpha}$ and $\hat{\beta}$ and the unknown function $\sigma^2(\cdot)$ by its nonparametric estimator $\hat{\sigma}^2(\cdot)$ described in Section 3. We specify that the market price of interest rate risk is constant⁴ and estimate it by minimizing the squared deviations between a given yield curve and that implied by the model. The only justification for the assumptions leading to $\lambda(r, t) = \lambda_{VAS}$ or $\lambda(r, t) = \lambda_{CIR} r^{1/2} / \sigma_{CIR}$ in the Vasicek and CIR models respectively is that they yield explicit solutions given their respective choices of drift and diffusion. Here any other function for the market price of risk could be used.

4.2. The Asymptotic Distribution of Nonparametric Prices

We show next that derivative prices \hat{U} still converge at the usual parametric speed $n^{1/2}$. This holds even though the σ^2 function is estimated nonparametrically and converges slower than $n^{1/2}$ (see Theorem 2). The intuition for the result is as follows. Consider the ordinary differential equation: $dU(r)/dr = \sigma^2(r)U(r)$ with initial condition $U(0) = U_0$. Its solution is simply: $U(r) = U_0 \exp(\int_0^r \sigma^2(x) dx)$. Suppose now that the coefficient function $\sigma^2(\cdot)$ of the differential equation was unknown and estimated nonparametrically by $\hat{\sigma}^2(\cdot)$ at speed $n^{1/2} h_n^{1/2}$. Estimate the solution $U(\cdot)$ by $\hat{U}(r) = U_0 \exp(\int_0^r \hat{\sigma}^2(x) dx)$. Now the integral $\int_0^r \hat{\sigma}^2(x) dx$ will converge at speed $n^{1/2}$, and so will the estimator $\hat{U}(r)$. A similar example would have the derivative price U depend upon a nonparametric estimate of the cumulative distribution function (converging at speed $n^{1/2}$) as opposed to the density (converging at speed $n^{1/2} h_n^{1/2}$).

The partial differential equation (4.1) satisfied by derivative prices is obviously substantially more complex than the simple ordinary differential equation above. However, it shares the two main insights of this example. First, the solution is also characterized as an integral over the coefficient functions. Second, the

⁴ Learning more about the specification of the market price of risk would require data on cross-sections of derivatives. The market price of risk could potentially be estimated nonparametrically as well by fitting a price curve exactly, along the lines of the “arbitrage-free” literature on interest rate derivatives.

coefficient functions are similarly estimated at speed $n^{1/2}h_n^{1/2}$ (the diffusion function σ^2) or faster ($n^{1/2}$ for all the parameters α , β , and λ). Given this intuition, the first step of the proof is to obtain a representation of derivative prices as an integral over the diffusion estimator. We consider the fundamental solution $\Gamma(r, t; x)$ of (4.1).⁵ This is the solution for x fixed in $(0, \infty)$ of the Cauchy problem $L\Gamma(r, t; x) = 0$ with initial condition $\Gamma(r, 0; x) = \delta_{(x)}(r)$ (a mass point at x). The following lemma will provide the integral characterization needed to obtain the asymptotic distribution of the derivative prices.

LEMMA 1: *Under Assumptions A6–A8, there exists a unique solution on $(0, \infty) \times (0, T]$ to the pricing partial differential equation, and it has the form*

$$(4.5) \quad U(r, t) = \int_0^{+\infty} \Gamma(r, 0; x) g(x) dx + \int_0^t \int_0^{+\infty} \Gamma(r, \tau; x) c(x, \tau) dx d\tau.$$

The fundamental solution $\Gamma(r, t; x)$ is twice-differentiable in r and x , once in t , and is a twice-differentiable function of $\sigma^2(x)$. It does not depend on $\sigma^2(r)$.

The dependence of the fundamental solution $\Gamma(r, t; x)$ on $\sigma^2(x)$ is detailed in the proof of Lemma 1. As a result $U(r, t)$ will depend on σ^2 only through integrals of functions of $\sigma^2(x) dx$, which converge at speed $n^{1/2}$. Therefore the speed of convergence of the derivative prices will not be slowed down by the use of a nonparametric estimator of σ^2 . The benefits from using a nonparametric estimator of σ^2 can be substantial in terms of avoiding misspecification of the derivative prices. This gain in robustness is achieved at no cost in terms of convergence speed for the prices.

THEOREM 3: *Under Assumptions A1–A9 the nonparametric price of a derivative security satisfies*

$$(4.6) \quad n^{1/2}\{\hat{U}(r, t) - U(r, t)\} \xrightarrow{d} N(0, V(r, t))$$

*with asymptotic variance $V(r, t) = \text{var}(u_{(r,t)}(x_t)) + 2\sum_{k=1}^{\infty} \text{cov}(u_{(r,t)}(x_t), u_{(r,t)}(x_{t+k}))$, where $x_t \equiv (r_t, r_{t+\Delta})$ and $u_{(r,t)}$ is a cadlag function.*⁶

To implement Theorem 3 in practice one needs the addition of a consistent estimator of the asymptotic variance $V(r, t)$. Since the expression for $u_{(r,t)}$ is complicated and $V(r, t)$ depends upon the entire serial correlation structure of the interest rate data, the bootstrap technique can be used in practice. The

⁵ For example, the transition density of the process over finite intervals of length Δ , $p(\Delta, r_{t+\Delta}|r_t)$, is the fundamental solution Γ corresponding to the backward Kolmogorov operator $-\partial U/\partial t + AU$ (i.e., the same as L in (4.2) but with $\lambda \equiv 0$). If μ and σ were parameterized in (1.1), then the parameters could be estimated by maximum-likelihood: solve numerically for Γ this partial differential equation to obtain the transition density of the model. By the Markov property, this is sufficient to characterize the likelihood function.

⁶ Cadlag = Right-continuous-left-limit.

validity of the bootstrap estimator of $V(r, t)$ in this context is proved in Ait-Sahalia (1992). The estimator consists of two simple steps:⁷

(1) Redraw from the original spot interest rate data. The resampling procedure redraws from blocks of contiguous observations to preserve the serial correlation existing in the original data, an idea introduced by Künsch (1989) and Liu and Singh (1992).

(2) Next estimate the drift and diffusion function associated with this new dataset as in Section 3 and then estimate the market price of risk and compute the resulting bond prices $\hat{U}^*(r, t)$. The bootstrap estimator $\hat{V}^*(r, t)$ of $V(r, t)$ is the sample variance of the difference of bond prices $\hat{U}^*(r, t) - \hat{U}(r, t)$.

Finally, instead of solving (4.1), derivative prices could have been computed by Monte-Carlo simulations of the sample paths of the risk-neutral process $dr_t = \{\mu(r_t, \theta) - \lambda(r_t)\sigma(r_t)\} dt + \sigma(r_t) dW_t$. The sample paths, all starting at r at date $T-t$ and finishing at date T , would be simulated with the risk-neutral drift and diffusion replaced by their estimates. Rewriting (4.5) as the Feynman-Kac conditional expectation under the risk-neutral dynamics gives the prices

$$(4.7) \quad U(r, t) = E \left[g(r_T) \exp \left\{ - \int_{T-t}^T r_u du \right\} + \int_{T-t}^T \exp \left\{ - \int_{T-t}^{\tau} r_u du \right\} c(r_\tau, \tau) d\tau \mid r_{T-t} = r \right].$$

The price $U(r, t)$ can then be estimated by averaging the argument of the conditional expectation over the simulated sample paths.

5. DISCOUNT BOND AND BOND OPTION PRICING

5.1. The Data

The spot rate used is the seven-day Eurodollar deposit rate, bid-ask midpoint, from Bank of America. The data are daily from June 1, 1973 to February 25, 1995. Interest rates paid on other short-term financial assets, such as commercial paper or T-Bills, typically move closely with the Eurodollar rate. A time-series plot of the data is provided in Figure 1. Choosing a seven-day rate—such as the seven-day Eurodollar—as the underlying factor for pricing derivatives is a necessary compromise between: (i) literally taking an “instantaneous” rate and (ii) avoiding some of the spurious microstructure effects associated with overnight rates. For example, the second-Wednesday settlement effect in the Federal Funds market creates a spike in the raw Federal Funds data that would have to be smoothed.

The rates quoted are originally bond-equivalent yields. They were transformed to continuously-compounded yield-to-maturity, deriving the current price

⁷ In terms of computer programming, the computation of the bootstrap standard errors requires only the addition of a loop for the resampling scheme.

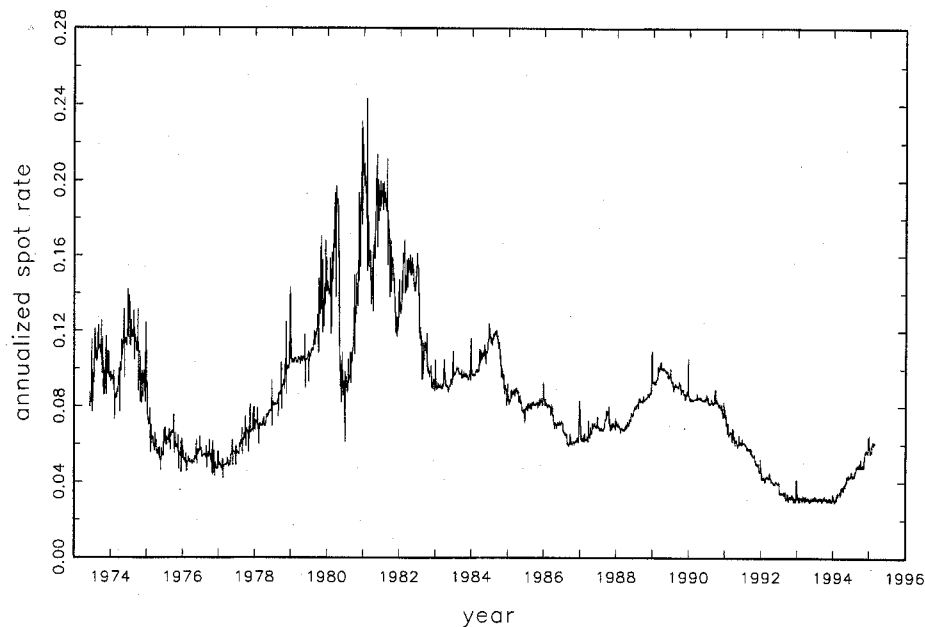


FIGURE 1—Spot interest rate.

B of the instrument from $r_{BEY} = ((100 - B)/B)(365/(T - t))$ and then computing the continuously compounded yield-to-maturity from $B = 100 \exp(-r_{YTM}(T - t))$ with time now measured in years. Monday is taken as the first day after Friday. Whereas weekend effects have been documented extensively for stock prices, there does not seem to be a conclusive weekend effect in money market instruments. Descriptive statistics are provided in Table II. Since the autocorrelations of the interest rate levels decay slowly, Table II also reports the results of an augmented Dickey-Fuller nonstationarity test. The nonstationarity hypothesis is rejected at the 90% level. The test is known to have low power, so even a slight rejection means that stationarity of the series is likely.

Four other daily interest rate series have also been used to check the robustness of the results: first, the overnight Federal Funds rate (source: Telerate); second, the one-month London Eurodollar deposit rate (average LIBOR bid-ask, source: Bank of America and Reuters); third, the one and three-month continuously-compounded bond-equivalent yields on Treasury bills (source: Bank of America). In general, the coefficient β of mean-reversion is lower when estimated on a longer-maturity proxy for the spot rate (e.g., estimated on three-month T-Bills vs. seven-day Eurodollar vs. Federal Funds). Also, longer maturity proxies are less volatile. Overall, all five series produce qualitatively similar shapes for σ^2 .

TABLE II
DESCRIPTIVE STATISTICS

Source: Bank of America 7-Day Eurodollar (Deposit Rate Midpoint Bid-Ask)		
Frequency: Daily		
Sample Period: 06/01/1973-2/25/1995		
Sample Size: 5505 observations		
Type: Continuously-Compounded Yield-to-Maturity (annualized rate)		
	Spot Interest Rate	First Difference of Spot Interest Rate
Mean	0.08362	-0.0000035
Standard Deviation	0.03591	0.004063
Monthly ρ_1	0.9389	0.02136
ρ_2	0.8785	-0.00689
ρ_3	0.8300	-0.01658
ρ_4	0.8014	0.00242
ρ_5	0.7783	0.00858
ρ_6	0.7715	0.01573
ρ_7	0.7361	0.00056
Augmented Daily Dickey-Fuller	-2.60	
H ₀ : Nonstationary	Reject at 90% (critical value = -2.57)	

Notes: (i) The augmented Dickey-Fuller test statistic is computed as $\hat{\tau}_\mu = \hat{\phi}/\text{ase}(\hat{\phi})$ in the model: $\Delta r_t = \mu + \phi r_{t-1} + \sum_{j=1}^p \phi_j \Delta r_{t-j} + u_t$, with $p = 30$ lags. See, e.g., Harvey (1993, Section 5.4).

(ii) The justification for using the Dickey-Fuller table when the residuals are heteroscedastic and possibly serially dependent is provided by Said and Dickey (1984) and Phillips (1987).

5.2. Nonparametric Diffusion Estimation

5.2.1. Drift Estimation

The first step consists of estimating the drift coefficients α and β by OLS. Drift estimates are reported in Table III for daily sampling of the annualized spot rate. These estimates will be used to construct the nonparametric diffusion estimator. Given the diffusion estimator, the second-step semiparametric FGLS estimates of the drift are computed. For both stages of drift estimates, the speed of mean reversion β is slow.

For comparison purposes, generalized method of moments (GMM) estimates of the Vasicek and CIR models are in Table III. The GMM estimates of α , β , and σ^2 for the Vasicek and CIR models are obtained from the following four moment conditions ($\Delta = 1$ day):

$$(5.1) \quad f_t(\theta)' \equiv [\varepsilon_{t+\Delta}, \varepsilon_{t+\Delta} r_t, \varepsilon_{t+\Delta}^2 - E[\varepsilon_{t+\Delta}^2 | r_t], (\varepsilon_{t+\Delta}^2 - E[\varepsilon_{t+\Delta}^2 | r_t]) r_t]$$

where $\varepsilon_{t+\Delta} \equiv (r_{t+\Delta} - r_t) - E[(r_{t+\Delta} - r_t) | r_t]$ with $E[r_{t+\Delta} - r_t | r_t] = (1 - e^{-\beta\Delta})(\alpha - r_t)$ for both models. The exact conditional variance of interest rate changes over

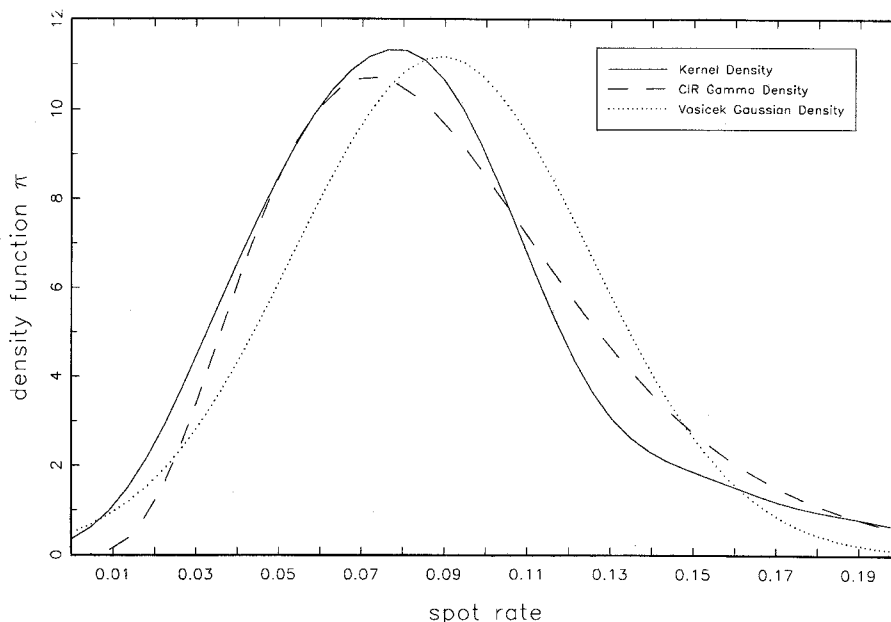


FIGURE 2—Nonparametric kernel density, CIR, and Vasicek densities.

happen to the spot rate. Alternatively, a model allowing for multiple interest rate regimes might be considered appropriate.

The density also exhibits a very fast decay to zero above 20%, due to the fact that in the sample very few interest rate levels are recorded outside this range and that the extreme observations are clustered next to the absolute smallest and largest ones (as opposed to being slowly spread on both ends of the density support). Compared to the Vasicek and CIR densities, the nonparametric density exhibits more variation and a combination of a longer tail and a faster rate of decay for large values of the interest rate. These differences will naturally be reflected in the diffusion function. A 99% pointwise confidence band for the nonparametric density is plotted in Figure 3, using the classical asymptotic distribution of the kernel density estimator. The large sample size makes it possible to estimate the density very precisely.

5.2.3. Diffusion Estimation

Given the drift parameters and nonparametric density estimator already constructed, the method of Section 3 yields the nonparametric estimator of the diffusion function, reported in Figure 4. The first noteworthy aspect of the graph is that σ^2 is globally an increasing function of the level of the interest rate between 0 and 14%. This gives credence to the widely expected result that high

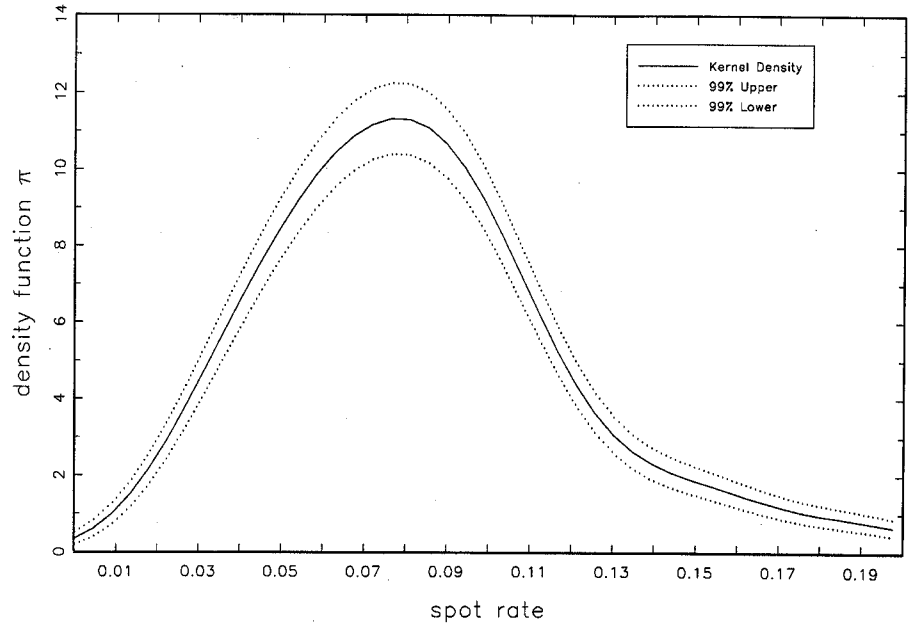


FIGURE 3—Nonparametric kernel density with 99% confidence band.

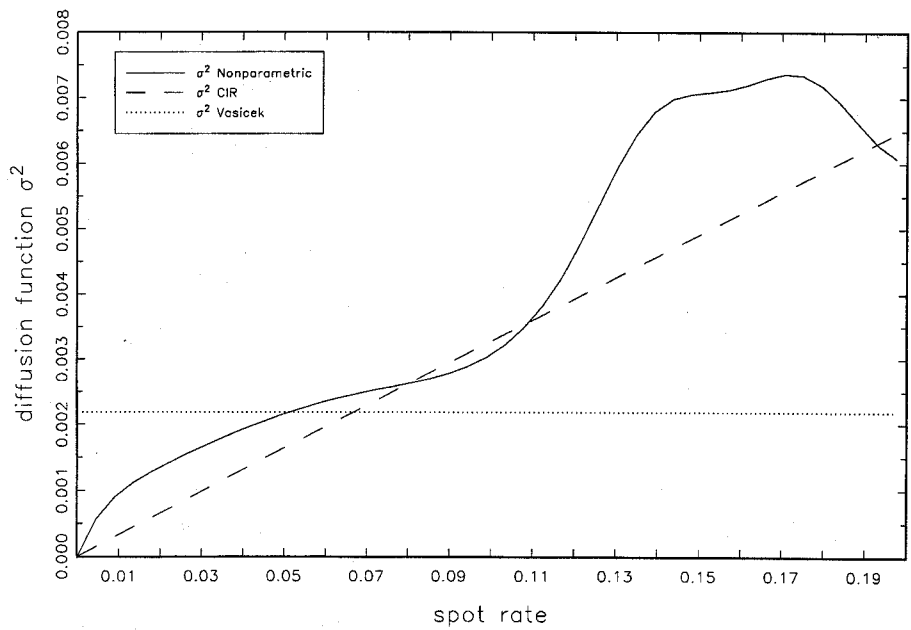


FIGURE 4—Nonparametric diffusion estimator compared to CIR and Vasicek.

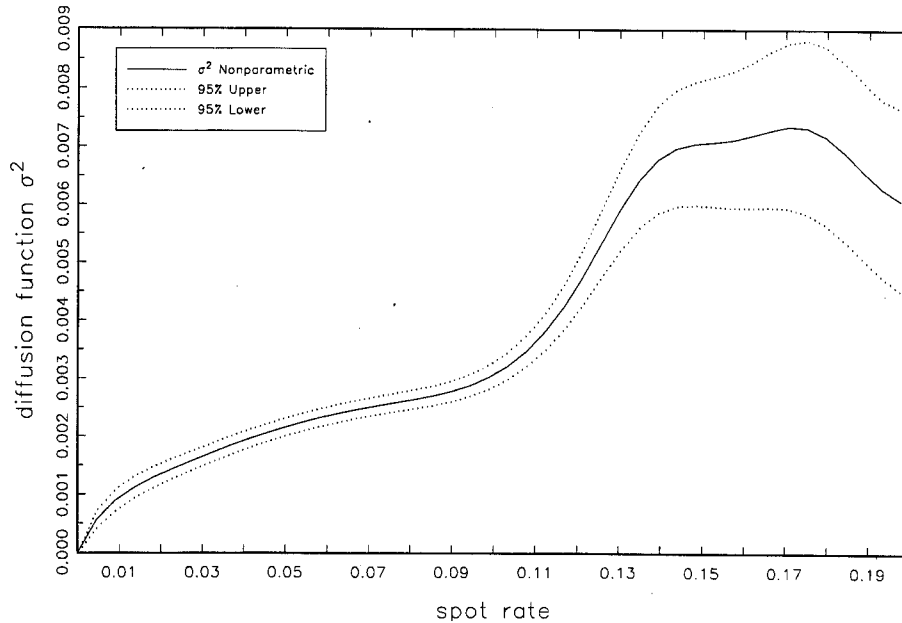


FIGURE 5—Nonparametric diffusion estimator with 95% confidence band.

interest rates should vary more than low interest rates. Above 14%, however, the diffusion function flattens and then decreases.

At high interest rates the drift in the process (2.1) pulls the process back toward the level of $\alpha = 9\%$. The fact that the density translates into a diffusion function with this shape depends upon the specific strength of the mean reversion effect in the drift. For example, the specification of a stronger mean-reverting drift, such as any power of the current specification, e.g., $\mu(r, \theta) \equiv \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3/r$, would make the process “more able” to pull back the spot rate toward 9% from a high level even in the presence of a larger variance (see Ait-Sahalia (1996)). A stronger drift (taking large negative values at high levels of r) would compensate for that higher potential variation and tilt the transitional density toward the left as does a medium-strength drift (e.g., linear) combined with a low diffusion. It is easy to extend the discussion in Section 2.2 to see the balancing effects of a stronger (resp. weaker) mean-reverting drift with a higher (resp. lower) diffusion.

At any rate, if a model insists on specifying a linear drift $\mu(r, \theta) \equiv \beta(r - \alpha)$ as do the Vasicek, CIR, and most other models, then the diffusion function compatible with the data is given by Figure 4 and looks neither linear nor flat. A formal test of misspecification of these parameterizations can be found in Ait-Sahalia (1996). A linear specification for the diffusion (CIR) is a relatively good approximation only on 0 to 9%. This is not too surprising since it was proved in Section 2.2 that any density polynomial in zero would give rise to a

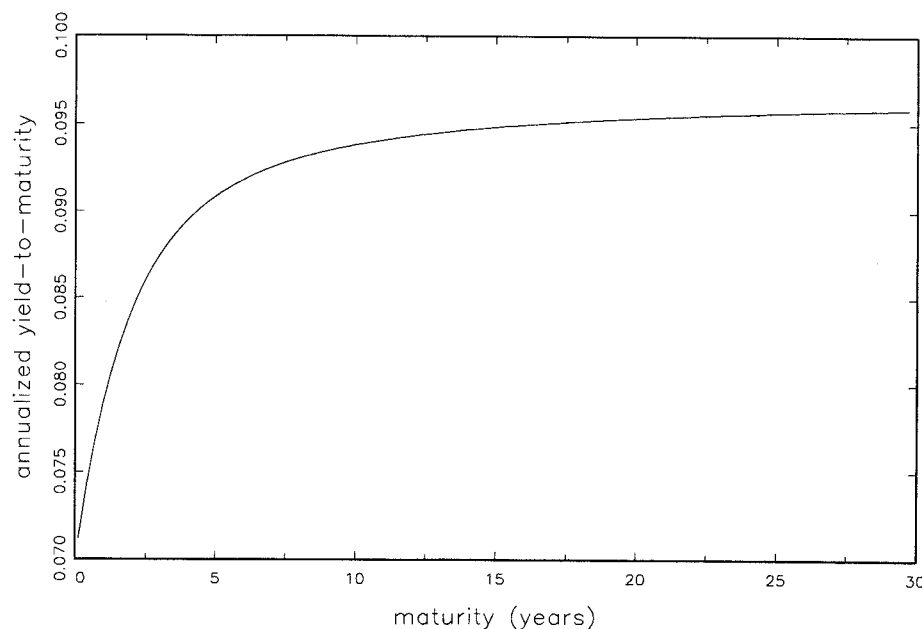


FIGURE 6—Target yield curve.

diffusion function linear for small values of the interest rate. The nonparametrically estimated σ^2 function increases faster than r for medium values of the interest rate (9 to 14%), flattens (14 to 17%), then decreases (above 17%). The CIR diffusion instead increases steadily as a linear function of r .

Figure 5 reports pointwise confidence intervals for the nonparametric diffusion estimator. The asymptotic distribution given by Theorem 2 has the feature that the diffusion function is estimated more precisely in highly populated interest rate regions. Conversely, the confidence band tends to increase above 14%, reflecting the relative scarcity of the data. The estimates of σ^2 are stable across subperiods of the sample, with the exception that subperiods which do not include the high interest rate years 1980–82 produce estimates of π with a truncated right tail, which in turn shift the graph of σ^2 to the left, and have a lower overall level of volatility.

5.3. *The Term Structure of Interest Rates*

The market price of risk is estimated for each model by minimizing the sum of squared deviations across maturities between a given target yield curve and the yields produced by the respective model. To obtain a realistic target yield curve, we take the average yield curve over the sample period obtained from the end-of-month Reuters data. The yield curve for coupon bonds is transformed to

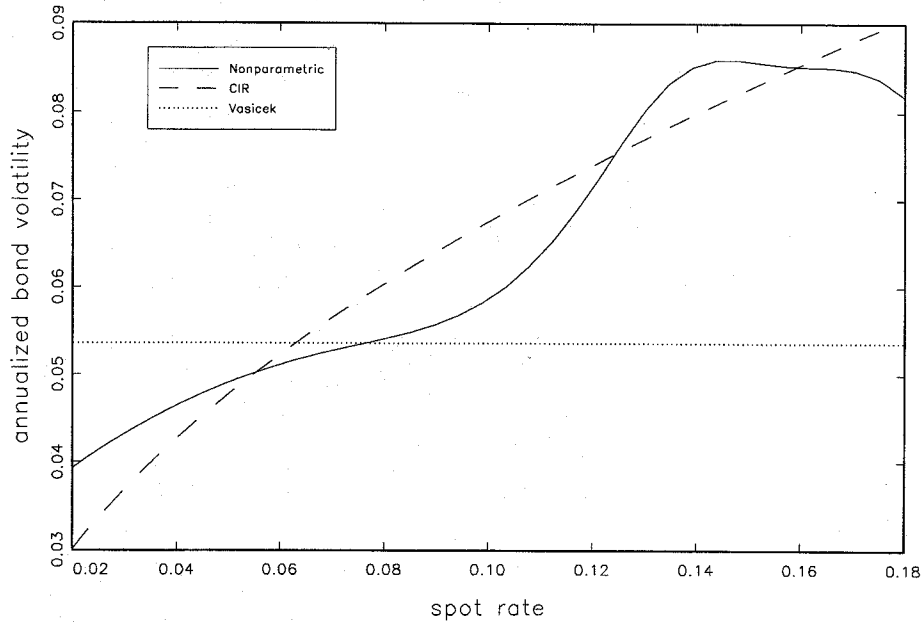


FIGURE 7—5-year bond price volatilities.

a discount yield curve by using the standard technique of fitting third-order polynomial splines. From now on, we treat this yield curve as fixed.

The target yield curve is plotted in Figure 6. Market price of risk estimates are in Table IV. Table V details the bond prices computed under the CIR, Vasicek, and nonparametric models. Standard errors for the nonparametric prices, derived as indicated in Section 4, are also reported. As a result of fitting a common yield curve, the CIR and Vasicek prices are generally within one standard deviation of the nonparametric prices. By eliminating differences in the prices of the underlying bonds, differences in option prices can be attributed to differences in the volatility of the underlying bonds, which is the determining factor.

TABLE IV
MARKET PRICE OF INTEREST RATE RISK

	Nonparametric	CIR	Vasicek
λ	$-2.4451 \cdot 10^{-1}$ (4.52)	$-7.8905 \cdot 10^{-2}$ (3.94)	$-1.6988 \cdot 10^{-1}$ (6.79)

Notes: (i) The market price of risk was estimated by minimizing the squared deviations between the respective model's bond yields and those in Figure 6.
(ii) The bootstrap t statistic of the estimate is in parentheses.

TABLE V
NONPARAMETRIC UNDERLYING BOND PRICES

Maturity (Years)	Annualized Spot Rate						
	0.02	0.04	0.06	0.08	0.10	0.12	0.14
0.5	98.2408	97.4936	96.7065	95.9064	95.1389	94.3724	93.6150
	(0.1078)	(0.0940)	(0.0875)	(0.0709)	(0.0861)	(0.0893)	(0.0960)
	98.3089	97.5054	96.7084	95.9179	95.1339	94.3563	93.5851
	98.2859	97.4900	96.7005	95.9175	95.1408	94.3703	93.6061
1	95.5091	94.2250	92.9944	91.7708	90.5475	89.3639	88.2045
	(0.1146)	(0.0923)	(0.0839)	(0.0655)	(0.0740)	(0.0832)	(0.0992)
	95.6082	94.3132	93.0357	91.7755	90.5323	89.3060	88.0964
	95.5461	94.2721	93.0150	91.7747	90.5509	89.3435	88.1521
5	66.8409	65.4272	64.0544	62.7142	61.4002	60.1130	58.8665
	(0.1125)	(0.1088)	(0.0984)	(0.0664)	(0.0753)	(0.0888)	(0.0991)
	67.3719	65.7962	64.2574	62.7546	61.2869	59.8535	58.4537
	67.2425	65.7148	64.2218	62.7627	61.3368	59.9433	58.5814
10	41.3704	40.4903	39.5658	38.6818	37.8343	37.0836	36.3885
	(0.0920)	(0.0806)	(0.0724)	(0.0678)	(0.0622)	(0.0706)	(0.0945)
	41.5477	40.5622	39.6000	38.6606	37.7436	36.8483	35.9742
	41.4768	40.5218	39.5887	38.6771	37.7865	36.9164	36.0664
30	5.9825	5.8550	5.7211	5.5802	5.4617	5.3256	5.2232
	(0.0873)	(0.0752)	(0.0738)	(0.0705)	(0.0814)	(0.0878)	(0.0800)
	5.9788	5.8369	5.6985	5.5633	5.4313	5.3024	5.1766
	5.9759	5.8383	5.7038	5.5724	5.4441	5.3187	5.1962

Notes: (i) All prices correspond to a face value of the bond equal to \$100. The four elements of each cell are from top to bottom: The nonparametric price, its bootstrap standard error (in parentheses), the CIR price, and the Vasicek price.

(ii) The nonparametric prices were computed using the 2-step FGLS estimates of α and β (second column of Table III), the nonparametric diffusion estimator (Figure 4), and the market price of risk from Table IV. The CIR and Vasicek prices were computed using the GMM estimates of α , β , and σ^2 reported in Table III (third and fourth column respectively), and the market prices of interest rate risk reported in Table IV.

(iii) The bootstrap standard errors are produced by $nb = 100$ replications with blocks containing $nk = 200$ consecutive spot observations.

5.4. Bond Option Pricing

Table VI reports call option prices on a 5-year discount bond and the nonparametric prices standard errors. Option prices reflect mostly differences in the second moment of the spot rate, where differences across models are more pronounced. By Itô's Lemma, the risk-neutral bond price follows $dB/B = rdt + \{\sigma(r)(\partial B/\partial r)/B\}dZ$ when the risk-neutral spot rate has dynamics: $dr = \{\mu(r) - \lambda(r)\sigma(r)\}dt + \sigma(r)dZ$. Figure 7 reports the volatilities $\sigma(r)(\partial B/\partial r)/B$ of a 5-year bond price as a function of the spot rate level r . How large the bond volatility is under each model provides some intuition for the relative magnitude of the three prices in Table VI.

When the short rate is 2%, we see that the bond price volatility from the Vasicek model is larger than that from the CIR and nonparametric models, which are close to each other. As a result, the Vasicek option prices are larger

TABLE VI
NONPARAMETRIC CALL OPTION PRICES ON A 5-YEAR BOND

Annualized Spot Rate (r)	Option Expiration (t : years)	Exercise Price				
		0.96	0.98	1.00	1.02	1.04
0.02	0.25	3.1451	1.8624	0.7926	0.1696	0.0054
		(0.0092)	(0.0097)	(0.0035)	(0.0030)	(0.0026)
		3.1432	1.8422	0.7179	0.0966	0.0002
		3.1513	1.9061	0.9049	0.3071	0.0694
0.02	0.5	3.8179	2.5660	1.4512	0.6141	0.1543
		(0.0094)	(0.0110)	(0.0094)	(0.0089)	(0.0072)
		3.0871	2.5379	1.3910	0.5231	0.0865
		3.8196	2.5855	1.5232	0.7478	0.2950
0.02	1	5.5679	4.3117	3.1139	2.0454	1.1620
		(0.0109)	(0.0134)	(0.0155)	(0.0128)	(0.0092)
		5.5574	4.3061	3.1056	2.0077	1.0911
		5.5718	4.3133	3.1159	2.0514	1.2014
0.14	0.25	4.2449	3.1665	2.1716	1.3281	0.6962
		(0.0125)	(0.0156)	(0.0127)	(0.0148)	(0.0124)
		4.2400	3.1636	2.1670	1.3163	0.6779
		4.2211	3.0941	2.0059	1.0653	0.4243
0.14	0.5	5.9629	4.9094	3.8868	2.9515	2.0940
		(0.0175)	(0.0176)	(0.0152)	(0.0156)	(0.0139)
		5.9595	4.8964	3.8679	2.9010	2.0322
		5.9392	4.8437	3.7558	2.7009	1.7418
0.14	1	9.0293	8.0119	6.9534	5.9949	5.0125
		(0.0194)	(0.0187)	(0.0176)	(0.0175)	(0.0163)
		9.0241	8.0013	6.9857	5.9828	5.0019
		9.0063	7.9735	6.9408	5.9089	4.8809

Notes: (i) All call option prices correspond to a face value of the discount bond equal to \$100. The exercise price is expressed as a proportion of the corresponding bond price for each model.

(ii) The four elements of each cell are from top to bottom: the nonparametric price, its bootstrap standard error in parentheses, the closed-form CIR price, and the closed-form Vasicek price. The standard errors of the nonparametric prices were computed under the same conditions as Table V.

than the other two, and significantly different from the nonparametric prices. Not surprisingly this effect is more visible for deep-out-of-the-money options (the right column of Table VI), and shorter maturity options (the upper rows in both parts of Table VI). Intuitively, the shorter the option, the more its price depends on the values of the parameters evaluated close to the current value of the spot rate r_t . Similarly when the short rate is 14%, the bond price volatility from the nonparametric spot process is higher than CIR, which, in turn, is higher than Vasicek. The ordering of the option prices is modified accordingly. Both the Vasicek and CIR prices then fall outside one or more standard deviations of the nonparametric prices.

6. CONCLUSIONS

This paper has proposed a procedure to estimate stochastic models while preserving their basic continuous-time structure. Properties of stochastic differential equations were exploited to both identify and estimate the drift and diffusion function of the process of interest. In particular, it was shown that the entire diffusion function can be estimated nonparametrically once a parameterization of the drift is adopted. The density-matching procedure starts from estimates of the densities of the process (which could as well be parametric) and then recovers the drift and the diffusion functions that are consistent with the density estimates. Asymptotic distributions are computed for the diffusion function of the underlying asset process and for the prices of the derivative securities. Even though the diffusion is estimated nonparametrically, and therefore converges slower than root- n , the speed of convergence of the derivative prices is not penalized.

The general density-matching approach could potentially be applied to a wide variety of processes in economics and finance, when it is important to identify precisely the instantaneous variance, using discrete data without discretizing the continuous-time process. The pricing of interest-rate derivative securities was precisely such a problem, since option prices depend on the particular shape of the diffusion function of the underlying spot interest rate. With a linear mean-reverting drift, consistent with most of the literature, the diffusion function was found to be increasing linearly, then exponentially, before flattening and decreasing. This was shown to have implications for the pricing of derivative securities. Other relevant applications can be considered in the future. The method can easily be extended to cover multi-dimensional stochastic differential equations and, therefore, multi-factor models of the term structure. Hedging strategies in the option market can be assessed with the help of the nonparametric prices. A similar estimation approach for the drift and diffusion could be implemented to estimate the process followed by stock returns, or exchange rates, to price foreign currency options, or identify the consequences of target zone bands.

Graduate School of Business, The University of Chicago, 1101 East 58th Street, Chicago, IL 60637-1561, U.S.A.,

and

National Bureau of Economic Research.

Manuscript received June, 1993; final revision received April, 1995.

APPENDIX 1

REGULARITY CONDITIONS FOR THE NONPARAMETRIC ESTIMATION OF THE DIFFUSION

- ASSUMPTION A1: (i) *The drift and diffusion functions have $s \geq 2$ continuous derivatives on $(0, \infty)$.*
(ii) $\sigma^2 > 0$ on $(0, \infty)$.
(iii) *The integral of $m(v) \equiv (1/\sigma^2(v)) \exp\{-\int_0^v [2\mu(u, \theta)/\sigma^2(u)] du\}$, the speed measure, converges at both boundaries of $(0, \infty)$.*

(iv) The integral of $s(v) \equiv \exp\{\int_v^{\bar{x}} [2\mu(u, \theta)/\sigma^2(u)] du\}$, the scale measure, diverges at both boundaries of $(0, \infty)$.

In A1(iii)–(iv), \bar{x} is fixed in $(0, \infty)$ and its particular choice is irrelevant. Under A1(i)–(ii), the stochastic differential equation (1.1) admits a unique strong solution. Note that by the mean-value theorem A1(i) implies the local Lipschitz and growth conditions:

(LLG) For each parameter vector θ and each compact subset of $D = (0, \infty)$ of the form $K \equiv [1/R, R]$, $R > 0$, there exist constants C_1 and C_2 such that for all x and y in K :

$$\begin{cases} |\mu(x, \theta) - \mu(y, \theta)| + |\sigma(x) - \sigma(y)| \leq C_1|x - y|, \\ |\mu(x, \theta)| + |\sigma(x)| \leq C_2(1 + |x|). \end{cases}$$

Global Lipschitz and growth conditions are usually imposed on the drift and diffusion functions⁸ to guarantee existence and uniqueness of a strong solution to the stochastic differential equation (1.1). These global conditions fail to be satisfied for many models of interest here. It is known that the local Lipschitz condition is sufficient for pathwise uniqueness of the solution (Itô's Theorem; see, e.g., Karatzas and Shreve (1991, Theorem 5.2.5, p. 287)). Now A1(i)–(ii) guarantees the existence of a weak solution (see, e.g., Karatzas and Shreve (1991, Theorem 5.5.15, p. 341)), up possibly to an explosion time. Then the existence of a weak solution combined with pathwise uniqueness imply strong existence (a corollary to the Yamada and Watanabe Theorem; e.g., Karatzas and Shreve (1991, Proposition 5.3.20, p. 309–310)). Therefore A1(i)–(ii) imply that (1.1) admits a unique strong solution up possibly to an explosion time.

The role of A1(iii)–(iv) is to rule out explosions in finite expected time and guarantee stationarity of the process. The solution is a Markov process with time-homogeneous transition densities (i.e., such that transition probability density $p(s, y; t, x)$ from x at time t to y at time s depends only on $t - s$, not on t and s separately). A1(iv) further guarantees that starting from any point in the interior of the state space $D = (0, \infty)$, the boundaries $x = 0$ and $\bar{x} = +\infty$ cannot be attained in finite expected time. Define the scale measure $S(\varepsilon, \bar{x}) \equiv \int_\varepsilon^{\bar{x}} s(v) dv$. Zero cannot be attained in finite time by the spot rate if $\lim_{\varepsilon \rightarrow 0} S(\varepsilon, \bar{x}) = +\infty$ (see Karlin and Taylor (1981, Lemma 15.6.3 and Table 15.6.2)). Infinity cannot be attained in finite time if: $\lim_{R \rightarrow +\infty} S(\bar{x}, R) = +\infty$.

An alternative condition can be formulated in terms of Lyapunov functions (see Hasminskii (1980)): Assume that there exists a twice continuously differentiable nonnegative function ϑ on D such that $\mu(x, \theta) \partial\vartheta(x)/\partial x + (1/2)\sigma^2(x) \partial^2\vartheta(x)/\partial x^2 \leq -1$ for all x in $D = (0, +\infty)$ outside a compact interval of the form $K \equiv [1/R, R]$, $R > 0$. Also assume that $\lim_{x \rightarrow 0} \vartheta(x) = \lim_{x \rightarrow \infty} \vartheta(x) = \infty$. Since $\vartheta \geq 0$, this condition implies the weaker inequality: $\mu(x, \theta) \partial\vartheta(x)/\partial x + (1/2)\sigma^2(x) \partial^2\vartheta(x)/\partial x^2 \leq c\vartheta(x)$ for some (in fact any) constant $c \geq 0$. Under this weaker inequality, we can only be sure to obtain a Markov solution with time-homogeneous transition densities (Hasminskii (1980, Theorem III.4.1 and Remark III.4.1)). This is necessary but not sufficient for the process to be stationary: for example a Brownian motion has time-homogeneous transition densities (Gaussian) but is obviously not a stationary process. For our estimation strategy to work, we need the process solution to (1.1) to be stationary and therefore impose in A1(iii) the stronger requirement that $\mu(x, \theta) \partial\vartheta(x)/\partial x + (1/2)\sigma^2(x) \partial^2\vartheta(x)/\partial x^2 \leq -1$ (use Theorem III.5.1 and footnote 9 in Hasminskii (1980), or see directly that this condition implies the existence of an integrable solution π to (2.2) for given drift and diffusion functions). It then follows that the stationary marginal density π exists (and integrates to one by construction) and is continuously differentiable on D up to order $s \geq 2$.

⁸ Global means here that the same constants C_1 and C_2 must be valid on the entire domain D of the diffusion. For a statement of Itô's classical result with global conditions, see e.g., Øksendal (1992, Theorem 5.5), or Karatzas and Shreve (1991, Theorem 5.2.5, page 287). Even the weaker Yamada-Watanabe condition (see Karatzas and Shreve (1991, 5.2.13, page 291)), fails to be satisfied for interesting economic models. Both would imply existence and uniqueness of a strong solution to (1.1).

ASSUMPTION A2: *The stationary density π is strictly positive on $(0, \infty)$, and the initial random variable r_0 is distributed as π .*

A2 makes it possible to characterize the diffusion function by (2.3) and subsequently construct the nonparametric estimator based on that expression. Under A1, the candidate stationary distribution $\pi(\cdot)$ exists. Stationarity of the process follows from initializing it with the stationary distribution itself. We now make assumptions required to guarantee that the Central Limit Theorem to be used in the proofs of Theorems 2 and 3 holds.

ASSUMPTION A3: *The observed data sequence $\{r_i, i = 1, \dots, n\}$ is a strictly stationary β -mixing sequence satisfying: $k \beta_k \xrightarrow{k \rightarrow +\infty} 0$ for some fixed $\delta > 1$.*

A3 restricts the amount of dependence allowed in the sequence. Without this assumption, the Central Limit Theorem can fail. As long as $\beta_k \xrightarrow{k \rightarrow +\infty} 0$, the sequence is said to be absolutely regular. In order for the discretely observed process $\{r_i, i \geq 1\}$ to satisfy A3, the stochastic process $\{r_t, t \geq 0\}$ has to verify a stronger continuous-time mixing condition (see Prakasa Rao (1987)). A condition on the drift and diffusion sufficient to insure that the discrete observations will satisfy A3 is the following:

ASSUMPTION A3': *$\lim_{r \rightarrow 0 \text{ or } r \rightarrow \infty} \sigma(r)\pi(r) = 0$ and $\lim_{r \rightarrow 0 \text{ or } r \rightarrow \infty} |\sigma(r)/\{2\mu(r, \theta) - \sigma(r)\sigma^{(1)}(r)\}| < \infty$.*

Geometric ergodicity follows from A3'. Indeed, under A3', Hansen and Scheinkman (1995) prove that the integral operator T (such that the infinitesimal generator A is the derivative of T) is a strong contraction. This can be used here. It follows from the strong contraction property of T that there exists λ , $0 < \lambda < 1$, such that for each measurable f and g : $|E[f(r_i)g(r_{i+k})]| \leq \|f\| \|g\| \lambda^k$. Apply this to the indicator $f = 1_A$ and $g = 1_B$ to obtain the mixing bounds for the discrete data: $2\alpha_k \leq \beta_k \leq \phi_k \leq \psi_k/2 \leq \lambda^k/2$ (where α , β , ϕ , and ψ are the various classical mixing coefficients). This is clearly a stronger mixing property of the discrete data than what is strictly required by A3. Only A3 is necessary. The next assumption is the standard regularity condition imposed on a kernel function:

ASSUMPTION A4: (i) *The kernel $K(\cdot)$ is an even function, continuously differentiable up to order r on R with $2 \leq r \leq s$, belongs to $L^2(R)$, and $\int_{-\infty}^{+\infty} K(x) dx = 1$.*

(ii) *$K(\cdot)$ is of order r : $\int_{-\infty}^{+\infty} x^i K(x) dx = 0$, $i = 1, \dots, r-1$, and: $\int_{-\infty}^{+\infty} x^r K(x) dx \neq 0$, $\int_{-\infty}^{+\infty} |x|^r |K(x)| dx < \infty$.*

Note that, in order to use effectively a kernel of order r for bias reduction, the density π must have at least that many derivatives. Hence the requirement in A1(i) that the drift and diffusion functions have $s \geq 2$ derivatives, and then the choice in A5 of $2 \leq r \leq s$. If a kernel of order $r = 2$ is used, then only $s = 2$ derivatives can be required in A1(i). More generally, when s derivatives of π exist, the optimal rate of convergence of the diffusion estimator is $n^{-s/(2s+1)}$. It is achieved by a kernel of order $r = s$ (even) with bandwidth set to $h_n = O(n^{-1/(2s+1)})$. The optimal rate of convergence for π and σ^2 can only be attained with a nonvanishing bias term (meaning that the asymptotic distribution of Theorem 2 would not be centered at 0). To avoid this undesirable feature, we require A5(i)–(ii).

ASSUMPTION A5 (Admissible Bandwidth Choices): *As $n \rightarrow \infty$, $h_n \rightarrow 0$ and:*

- (i) *to estimate π : $n^{1/2} h_n^{(2r+1)/2} \rightarrow 0$, and $n^{1/2} h_n^{1/2} \rightarrow \infty$;*
- (ii) *to estimate σ^2 : $n^{1/2} h_n^{(2r+1)/2} \rightarrow 0$, and $n^{1/2} h_n^{3/2} \rightarrow \infty$;*
- (iii) *to estimate derivative prices U : $n^{1/2} h_n^r \rightarrow 0$.*

A5 gives the admissible range of bandwidth choices depending upon the object to be estimated. The speed of decrease of the bandwidth h_n to 0 as n becomes large is restricted. Optimality within the respective admissible ranges is discussed later. In A5(i) the first condition insures that the bias in

the normalized density estimator is asymptotically negligible. The second condition insures that the variance of the estimator which is of order $n^{-1}h_n^{-1}$ goes to zero. In A5(ii) the first condition serves the same purpose. Given the first, the second condition insures that the asymptotic distribution of the diffusion estimator is indeed given by the linear term in its Taylor expansion (i.e., the remainder term in the expansion is also asymptotically negligible). As a consequence of $n^{1/2}h_n^2 \rightarrow \infty$, it follows that $n^{1/2}h_n^{1/2} \rightarrow \infty$ and therefore the variance of the estimator which is also of order $n^{-1}h_n^{-1}$ goes to zero. In A5(iii), the condition makes the bias negligible and is sufficient to make the remainder term in the Taylor expansion go to zero.

The empirical results for the density and diffusion estimators (Figures 2–3 and 4–5) are obtained with a Gaussian kernel $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$ of order $r=2$, so we only need to assume the existence of $s=2$ derivatives of π . The optimal bandwidth choice (in the integrated mean squared error sense) within the range specified by either A5(i) or A5(ii) requires getting as close as possible to $n^{-1/(2r+1)}$. Choosing the bandwidth at this order would produce a bias term which does not vanish asymptotically. We therefore set the bandwidth as $h_n = c_n n^{-1/(2r+1)} = c_n n^{-1/5}$, where $c_n = c/\ln(n)$ with the constant c chosen to minimize the integrated mean square error of the diffusion estimator. The diffusion estimator can be calculated explicitly. For the actual sample size $n = 5505$, the bandwidth was $h_n = 1.6033 \cdot 10^{-2}$.

The higher order kernel $K(u) = (3/\sqrt{8\pi})(1-u^2/3)\exp(-u^2/2)$ (order $r=4$) with optimal bandwidth $h_n = c_n n^{-1/(1+2r)} = c_n n^{-1/9}$, as well as exponential, sinusoidal, and Epanechnikov kernels, produce similar results. The results are qualitatively robust to the choice of the kernel and small changes in the bandwidth parameter around the optimal value. Selection of the bandwidth by other various forms of cross-validation also produces similar estimates. The tail behavior of the kernel does not modify substantially the tail behavior of the diffusion estimator in the interval considered (the large values of the spot rate are clustered together in the sample so the tail behavior of the density is estimated fairly accurately up to 22%).

The empirical results for the derivative prices (Tables V–VI and Figure 7) are obtained with the same Gaussian kernel, but the optimal choice in the integrated mean square error sense is now $h_n = cn^{-1/3}$ where c is a scaling factor. The scaling factor is still set to minimize the integrated mean square error of the estimator. The precise choice of the bandwidth in this case does not modify the rate of convergence $n^{1/2}$ of derivative prices (it only affects higher order terms in the integrated mean squared error). Hence, as long as A5(iii) is satisfied, it is substantially less important to be close to the optimal bandwidth than for pointwise estimation of σ^2 . Finally note that the optimal bandwidth choice when estimating σ^2 pointwise is larger (less smoothing) than is admissible when estimating derivative prices. This is due to the respective size of the bias terms in the two cases. In particular the estimate of derivative prices would be biased if they were computed from plugging in the estimate of σ^2 computed at its pointwise bandwidth choice (A5'(i)). We summarize the actual bandwidth choice in Assumption A5':

ASSUMPTION A5' (Actual Bandwidth Choice): (i) To estimate π and σ^2 : $h_n = c_n n^{-1/(2r+1)} = c_n n^{-1/5}$ with $c_n = c/\ln(n)$.

(ii) To estimate derivative prices U : $h_n = cn^{-1/3}$.

(iii) In both cases, the scaling factor c is set to minimize the asymptotic integrated mean squared error of the estimator.

Monte-Carlo evidence shows that the asymptotic distribution given by Theorem 2 is a good approximation for the actual distribution of the diffusion estimator given the same sample size as the data set. The coverage probabilities are consistently above 0.85 when simulating data with various specifications (CIR, various power functions for the diffusion: $\sigma^2(r) = r^\gamma$) and computing the diffusion estimator.

APPENDIX 2: IDENTIFICATION WITH A GENERIC PARAMETERIZATION OF THE DRIFT

Let $\mu(r, \theta)$ be a parameterization of the drift. The identification restriction states that there exists a unique parameter vector θ such that the true drift is $\mu(r, \theta)$. This Appendix shows how to

modify (2.4) when μ is not necessarily linear. Let $\eta(r) \equiv \int_0^r \mu(u, \theta) \pi(u) du$. Then the equation $\partial \vartheta(r, t) / \partial t = A \vartheta(r, t)$ with initial condition $\vartheta(r, 0) = r$ implies

$$\begin{aligned} \pi(r) \partial \vartheta / \partial t &= \mu(r, \theta) \pi(r) \partial \vartheta / \partial x + (1/2) \sigma^2(r) \pi(r) \partial^2 \vartheta / \partial r^2 \\ &= \{\partial \eta / \partial r\} \{\partial \vartheta / \partial r\} + \eta(r) \{\partial^2 \vartheta / \partial r^2\}. \end{aligned}$$

Therefore ϑ solves $\pi(r) \partial \vartheta / \partial t = \{\partial / \partial r\} [\eta(r) \partial \vartheta / \partial r]$. By Dynkin's formula again this partial differential equation has a unique solution given by $\vartheta(r, t) = E[r_t | r_0 = r]$. Solving this equation yields ϑ . Then θ can be estimated consistently by nonlinear least squares with this generalization of (2.4).

APPENDIX 3: REGULARITY CONDITIONS FOR THE PRICING EQUATION

The following assumptions are introduced to obtain nonparametric derivative prices and their asymptotic distribution. They are not needed to estimate the diffusion function itself.

ASSUMPTION A6: *The drift, diffusion, and density functions of the risk-neutral spot rate process, $dr_t = \{\mu(r_t, \theta) - \lambda(r_t)\sigma(r_t)\} dt + \sigma(r_t) dW_t$, satisfy the conditions in A1–A2.*

ASSUMPTION A7: *The derivative continuous payment rate c and final payoff g are right-continuous-left-limit functions and satisfy for (r, t) in $(0, \infty) \times (0, T]$: $|c(r, t)| \leq N \exp(\nu r^2)$ and $|g(r)| \leq N \exp(\nu r^2)$ for some $\nu, N > 0$.*

ASSUMPTION A8: *The derivative price $U(r, t)$ satisfies the exponential boundedness condition: $\int_0^T \int_0^\infty e^{-kr^2} |U(r, t)| dr dt < \infty$ for some $k > 0$.*

The solution to (4.1) is only unique in the class of functions satisfying A8. The bound is sharp as the solution fails to be unique in general if A8 is replaced by $\int_0^T \int_0^\infty e^{-\varepsilon kr^{2+\varepsilon}} |U(r, t)| dr dt < \infty$ for some $\varepsilon > 0$ (see Friedman (1975)).

ASSUMPTION A9: *The market price of interest rate risk depends on a parameter λ and is estimated by minimizing the sum of squared deviations between a fixed set of derivative prices and the corresponding nonparametric prices.*

APPENDIX 4: SOLVING THE PRICING PARTIAL DIFFERENTIAL EQUATION

The partial differential equation (4.1) with (4.3) and then (4.4) will be solved numerically using finite-differencing with a combination of an implicit scheme for the advective term $\partial U / \partial r$ and the diffusion term $\partial^2 U / \partial r^2$. Since the numerical solution is guaranteed to converge only on a compact set, we make the change of variable $s \equiv cr / (1 + cr)$ where the constant c is large enough to over-represent interest rate levels between 0 and 20%. This increases the accuracy of the numerical procedure. Let $V(s, t) \equiv U(r, t)$, so we can replace in (4.1): $\partial U / \partial t = (\partial V / \partial t)$, $\partial U / \partial r = (\partial V / \partial s)(ds/dr)$, and $\partial^2 U / \partial r^2 = (\partial^2 V / \partial s^2)(ds/dr)^2 + (\partial V / \partial s)(d^2s/dr^2)$. The discretized counterpart to (4.1) is

$$\begin{aligned} \frac{v_j^m - v_j^{m-1}}{\Delta t} + p_j \left(\frac{(v_{j+1}^m - 2v_j^m + v_{j-1}^m) + (v_{j+1}^{m-1} - 2v_j^{m-1} + v_{j-1}^{m-1})}{2(\Delta s)^2} \right) \\ + q_j \left(\frac{(v_{j+1}^m - v_{j-1}^m) + (v_{j+1}^{m-1} - v_{j-1}^{m-1})}{4 \Delta s} \right) - r(s_j) \frac{v_j^m + v_j^{m-1}}{2 \Delta s} + c(r(s_j), t^m) = 0 \end{aligned}$$

where $v_j^m \equiv V(s_j, t^m)$, and t^m and s_j are the discretized values of the time and spot interest rate transformed variable, $j = 2, \dots, J - 1$ and $m = 1, \dots, M$. $r(s_j)$ is the inverted change of variable. The coefficients p and q can easily be obtained from (4.1) and the differentials above. For $j = 1$, corresponding to 0, the diffusion function is zero and thus the equation becomes

$$\frac{v_1^m - v_1^{m-1}}{\Delta t} + q_1 \left(\frac{(v_2^m - v_1^m) + (v_2^{m-1} - v_1^{m-1})}{4 \Delta s} \right) + c(0, t^m) = 0.$$

For $j = J$, v_j^m is obtained explicitly from v_j^{m-1} , v_{j-1}^{m-1} , and v_{j-2}^{m-1} .

Obviously, the equation could be solved more easily with a fully explicit representation. However, explicit finite-differencing of parabolic equations is known to yield estimates that do not necessarily converge to the true solution of the equation (see Ames (1977) for a von Neumann stability analysis). We have adopted the method recommended by Ames (1977) consisting of an implicit discretization of the second order partial derivative and an explicit discretization of the first order derivatives. This method is known as the Crank-Nicolson scheme. Convergence to the solution as $\Delta r \rightarrow 0$ and $\Delta t \rightarrow 0$ is guaranteed. For given $\{u_j^m / j = 1, \dots, J\}$, (7.1) gives $\{u_j^{m-1} / j = 1, \dots, J\}$ as the solution of an easily solvable triangular system. The first set $\{u_j^M / j = 1, \dots, J\}$ is derived directly from the final condition (4.3) for bond pricing, and later (4.4) for bond option pricing. Then the equation is used recursively backwards to derive successively the sets $\{u_j^m / j = 1, \dots, J\}$ for $m = M - 1, M - 2$, etc., up to $m = 1$.

When actually pricing bonds and bond options under the CIR model and comparing the numerical solutions to the closed-form formulas, this procedure was found to be very accurate and the approximation error to shrink as much as wanted when decreasing the step sizes.

APPENDIX 5: PROOFS

PROOF OF THEOREM 2: The proof uses the functional delta results of Ait-Sahalia (1992). The diffusion function (2.3) is viewed as a functional of the joint cumulative distribution function F of observations sampled $\Delta = 1$ day apart, $(r_t, r_{t+\Delta})$. Compute a first-order expansion of the diffusion away from F . From (2.3),

$$\sigma^2[F + H](x) = \frac{2}{\pi[F + H](x)} \int_0^x \mu(u, \theta[F + H]) \pi[F + H](u) du.$$

The drift parameters are estimated at the parametric speed, i.e., $\theta(\cdot)$ is a regular functional. That is, there exists a linear functional $\theta^{(1)}(\cdot, \cdot) \in C^{-1}$ such that

$$\theta[F + H] = \theta[F] + \int_0^{+\infty} \int_0^{+\infty} \theta^{(1)}(u, v) h(u, v) du dv + O(\|H\|_{L(\infty, 1)}^2).$$

C^{-1} is the space of cadlag functions. The norm $L(\infty, 1)$ is the sum of the supremum norms for H and its derivative h . The fact that $\theta(\cdot)$ is evaluated at the empirical distribution instead of a kernel plug-in does not change the asymptotic distribution.

Let $\delta_{(x)}$ denote a Dirac mass centered at x . The expansion for the marginal density is

$$\pi[F + H](x) = \pi[F](x) + \int_0^{+\infty} \int_0^{+\infty} \delta_{(x)}(u) h(u, v) du dv + O(\|H\|_{L(\infty, 1)}^2).$$

Collecting terms,

$$\begin{aligned} \sigma^2[F + H](x) = & \sigma^2[F](x) + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left\{ \frac{1}{\pi(x)} (-\sigma^2(x) \delta_{(x)}(u, v) + B[F](x)) \right\} \\ & \times h(u, v) du dv + O(\|H\|_{L(\infty, 1)}^2). \end{aligned}$$

$B[F](\cdot)$ is a functional in C^{-1} which does not influence the asymptotic distribution, and therefore need not be identified. It follows that $\sigma^2[\cdot](x)$ is $L(\infty, 1)$ -differentiable with respect to F . That is, there exists a distribution $\varphi[F](\cdot, \cdot)$ such that the functional $\sigma^2(\cdot)$ satisfies

$$\sigma^2[F + H](x) = \sigma^2[F](x) + \int_0^{+\infty} \int_0^{+\infty} \varphi[F](u, v) h(u, v) \, du \, dv + O(\|H\|_{L(\infty, 1)}^2).$$

The derivative of the functional $\sigma^2(\cdot)$ is

$$\varphi[F](u, v) = \left\{ \frac{-\sigma^2[F](x)}{\pi[F](x)} \right\} \delta_{(x)}(u, v) + B(x) \in C^{-2} \setminus C^{-1}$$

where C^{-2} is the space of Dirac delta functions. Its asymptotic distribution follows directly from the functional delta method.

PROOF OF LEMMA 1: Consider the compact set $K_\varepsilon = [\varepsilon, 1/\varepsilon]$ for a small fixed $\varepsilon > 0$, and fix x in K_ε . Consider the equation $L_\varepsilon \Gamma = 0$, where $L_\varepsilon U \equiv LU + \varepsilon(\partial^2 U / \partial r^2)$. It is necessary to perturbate L slightly because $\sigma^2(r) > 0$ for all $r > 0$ but $\sigma^2(0) = 0$; thus L is parabolic on $(0, \infty) \times (0, T]$ instead of the entire domain $(0, \infty) \times (0, T]$. Under A6, L_ε is uniformly parabolic on $K_\varepsilon \times (0, T]$ and its coefficient functions satisfy the hypotheses of Friedman (1964, Section I.4). We can therefore construct the fundamental solution Γ_ε corresponding to the operator L_ε by using the parametrix method. Alternatively, Γ_ε could be obtained as the limit of the Green functions associated with L_ε on $K_\varepsilon \times (0, T]$ as in Friedman (1976, Lemma 15.1.1); see also Freidlin (1985).

Form the parametrix solution for any (r, t) in $K_\varepsilon \times (0, T]$,

$$\Gamma_\varepsilon(r, t; x) = Z(r, t; x) + \int_0^t \int_\varepsilon^{1/\varepsilon} Z(r, t-s; y) \Phi_\varepsilon(y, s; x) \, dy \, ds,$$

where we have introduced $Z(r, t; x) \equiv (1/\sqrt{2\pi t} \sigma(x)) \exp\{- (r-x)^2 / 2t\sigma^2(x)\}$ and $\Phi_\varepsilon(y, s; x) \equiv \sum_{p=1}^\infty (L_\varepsilon Z)_p(y, s; x)$. The differential operator $(L_\varepsilon \cdot)_p$ is defined recursively by $(L_\varepsilon Z)_1 = L_\varepsilon Z$ and

$$(L_\varepsilon Z)_{p+1}(y, s; x) = \int_0^s \int_\varepsilon^{1/\varepsilon} [L_\varepsilon Z(y, s-\tau; z)] [(L_\varepsilon Z)_p(z, \tau; x)] \, dz \, d\tau.$$

By Theorem I.8 in Friedman (1964) the infinite series in Φ_ε converges and the parametrix solution is well-defined.

Next it is clear that the parametrix expression is a fundamental solution of (4.1) on $K_\varepsilon \times (0, T]$. Write

$$\begin{aligned} L_\varepsilon \Gamma_\varepsilon(r, t; x) &= L_\varepsilon Z(r, t; x) + L_\varepsilon \left\{ \int_0^t \int_\varepsilon^{1/\varepsilon} Z(r, t-s; y) \Phi_\varepsilon(y, s; x) \, dy \, ds \right\} \\ &= L_\varepsilon Z(r, t; x) - \Phi_\varepsilon(r, t; x) + \int_0^t \int_\varepsilon^{1/\varepsilon} L_\varepsilon Z(r, t-s; y) \Phi_\varepsilon(y, s; x) \, dy \, ds. \end{aligned}$$

Thus for Γ_ε to be a fundamental solution Φ_ε must satisfy the Volterra integral equation

$$\Phi_\varepsilon(r, t; x) = L_\varepsilon Z(r, t; x) + \int_0^t \int_\varepsilon^{1/\varepsilon} L_\varepsilon Z(r, t-s; y) \Phi_\varepsilon(y, s; x) \, dy \, ds.$$

By construction the Volterra equation is satisfied by $\sum_{p=1}^\infty (L_\varepsilon Z)_p(r, t; x)$.

Then as in Theorem 15.1.2 in Friedman (1976) we obtain that as ε goes to zero $\Gamma_\varepsilon(r, t; x)$ converges to a limit $\Gamma(r, t; x)$, together with the first two r -derivatives, x -derivatives, and the first t -derivative uniformly for all r and x in $(0, \infty)$ and t in $(0, T]$. Furthermore the limit $\Gamma(r, t; x)$ satisfies (4.1) on $(0, \infty) \times (0, T]$ as in Corollary 15.1.3 in Friedman (1976). The fundamental solution has the form $\Gamma(r, t; x) = \gamma(r, t, x, \sigma^2(x), \mu(x))$ with γ continuously differentiable in each of its arguments (see the parametrix form). The only difference here with the results in Chapter 15 in

Friedman (1976) is that the parabolic operator L contains the term $(-rU)$. The proof is otherwise identical. Finally the fundamental solution is unique in the class of nonexplosive solutions (Assumption A8). By Theorem I.16 in Friedman (1964), for each $\varepsilon > 0$ K_ε is unique on $K_\varepsilon \times (0, T]$. Any two solutions on $(0, \infty) \times (0, T]$ would have to coincide on $K_\varepsilon \times (0, T]$.

PROOF OF THEOREM 3: We first consider the case where the market price of risk is known, or, without lack of generality, $\lambda \equiv 0$. Then the coefficients of the differential operator L are functionals of the joint cumulative distribution function F of the data $(r_t, r_{t+\Delta})$ so the differential operator is $L[F]$. Then consider the equations followed by the derivative prices associated to F and $F + H$ respectively: $L[F]U[F](r, t) = -c(r, t)$ with initial condition $U[F](r, 0) = g[F](r)$, and $L[F + H]U[F + H](r, t) = -c(r, t)$ with initial condition: $U[F + h](r, 0) = g[F + H](r)$. The final payoff $g(\cdot)$ depends on F because it could be the price of a derivative previously computed: for example, the price of the underlying bond when valuing a bond option.

Lemma 1 characterizes the fundamental solution Γ in the form $\Gamma(r, t; x) = \gamma(r, t, x, \sigma^2(x), \mu(x))$ with γ continuously differentiable in each of its arguments. From this characterization, and the (linear in H) expansions of the drift and diffusion functionals in Theorem 2, it follows that $\Delta U[F, H] \equiv U[F + H] - U[F]$ admits the expansion ($L(\infty, 0)$ is the standard supremum norm for H)

$$\begin{aligned} \Delta U[F, H](r, t) &= \int_0^{+\infty} \Gamma[F + H](r, 0; x) g[F + H](x) dx \\ &\quad - \int_0^{+\infty} \Gamma[F](r, 0; x) g[F](x) dx \\ &\quad + \int_0^t \int_0^{+\infty} \{\Gamma[F + H] - \Gamma[F]\}(r, \tau; x) c(x, \tau) dx d\tau \\ &= \int_0^{+\infty} \int_0^{+\infty} u[F](r, t, x, y) h(x, y) dx dy + O(\|H\|_{L(\infty, 0)}^2). \end{aligned}$$

Suppose that for the derivative under consideration the final payoff $g(\cdot)$ is independent of F (e.g., a bond for which $g = 1$, or an interest rate option: $g = \max(0, r - k)$). Then the functional $F \mapsto U[F](r, t)$ for fixed (r, t) in $(0, \infty) \times (0, T]$ is $L(\infty, 1)$ -differentiable with respect to F with a continuous functional derivative. Its functional derivative $u(\cdot)$ is continuous as can be seen from collecting terms from the linearizations of $\mu(\cdot)$ and $\sigma^2(\cdot)$ (which were computed in Theorem 2) and the partial derivatives of $\gamma(r, t, x, \sigma^2(x), \mu(x))$ with respect to σ^2 and μ . If now the derivative's payoff g depends on F , typically through the previously computed price of a more primitive derivative, then the expansion of $g(\cdot)$ also matters. An example of this situation is a bond call option where $g(F)$ is the maximum of zero and the price of the underlying bond (dependent on F) net of the strike price. By what precedes, the bond price admits a first-order expansion in F with derivative in C^{-1} (since its payoff is independent of F). Under Assumption A7 the final payoff for the derivative also satisfies this requirement. The functional derivative $u(\cdot)$ then has additional terms which reflect this additional dependence on F .

We next show that the introduction of the market price of risk does not modify the rate of convergence (only the asymptotic variance). This holds provided that λ is estimated at a speed⁹ that is at least as fast as that of σ^2 , in particular when the estimator of λ converges at speed $n^{1/2}$. This will hold as a further special case when the market price of risk is estimated by matching the prices a cross-section of derivatives $\{\{U_i\}, i = 1, \dots, J\}$. Indeed under Assumption A9 set $\hat{\lambda} \equiv \arg \min_{\lambda \in L} \sum_{j=1}^J \{\hat{U}_j(r, t_j) - U_j\}^2$ by minimizing the squared deviations between the derivative prices $\{\{U_i\}\}$ and those predicted by the model $\{\hat{U}_j(r, t_j)\}$ evaluated at λ .

⁹ If however λ were estimated at a slower speed than that of σ^2 (e.g., its derivative were in $C^{-3} \setminus C^{-2}$, the space of first derivatives of Dirac masses), then $U[\cdot, \cdot]$ would no longer converge at speed $n^{1/2}$.

Note that under Assumption A9 we assume that we are matching a fixed set of derivative prices and therefore do not have to account for the variability of prices being matched.¹⁰ Recall that the solution U_λ has μ replaced by $\mu - \lambda\sigma$. Furthermore the characterization of the fundamental solution has $\Gamma(r, t; x) = \gamma(r, t, x, \sigma^2(x), \mu(x))$ with γ continuously differentiable in each of its arguments. Therefore $\hat{U}_\lambda(r, t_i)$ is continuously differentiable in λ and $\lambda(\cdot)$ is $L(\infty, 1)$ -differentiable as a functional of F with derivative in C^{-1} by the implicit functional theorem: view λ as the implicit solution of the first order condition

$$\sum_{j=1}^J \{\partial \hat{U}_\lambda(r, t_j) / \partial \lambda\} \{\hat{U}_\lambda(r, t_j) - U_j\} = 0.$$

As a consequence, this particular $\hat{\lambda}$ converges at speed $n^{1/2}$ and \hat{U} also converges at speed $n^{1/2}$. Collect all the linear terms in the first-order expansion of U and let $u_{(r,t)}$ be the derivative (in C^{-1} by what precedes) of U . A complete expression for $u_{(r,t)}$ can be computed explicitly but is not needed since the asymptotic variance of the prices will be obtained by the bootstrap method.

REFERENCES

- AÏT-SAHALIA, Y. (1992): "The Delta and Bootstrap Methods for Nonparametric Kernel Functionals," MIT mimeo.
- (1996): "Testing Continuous-Time Models of the Spot Interest Rate," forthcoming in the *Review of Financial Studies*, 9.
- AMES, W. F. (1977): *Numerical Methods for Partial Differential Equations*. New York: Academic Press.
- BANON, G. (1978): "Nonparametric Identification for Diffusion Processes," *S.I.A.M. Journal of Control and Optimization*, 16, 380–395.
- BASAWA, I. V., AND B. L. S. PRAKASA-RAO (1980): *Statistical Inference for Stochastic Processes*. New York: Academic Press.
- BLACK, F., AND M. SCHOLES (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 3, 133–155.
- BRENNAN, M. J., AND E. S. SCHWARTZ (1979): "A Continuous-Time Approach to the Pricing of Bonds," *Journal of Banking and Finance*, 3, 133–155.
- BROWN, S. J., AND P. H. DYBVIG (1986): "The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates," *Journal of Finance*, 41, 617–630.
- CHAMBERLAIN, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.
- CHAN, K. C., G. A. KAROLYI, F. A. LONGSTAFF, AND A. B. SANDERS (1992): "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 47, 1209–1227.
- CONSTANTINIDES, G. M. (1992): "A Theory of the Nominal Term Structure of Interest Rates," *Review of Financial Studies*, 5, 531–552.
- COURTADON, G. (1982): "The Pricing of Options on Default-Free Bonds," *Journal of Financial and Quantitative Analysis*, 17, 75–100.
- COX, J. C. (1975): "Notes on Option Pricing I: Constant Elasticity of Diffusions," Stanford University Mimeo.
- COX, J. C., J. E. INGERSOLL, AND S. A. ROSS (1980): "An Analysis of Variable Rate Loan Contracts," *Journal of Finance*, 35, 389–403.
- (1985a): "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363–384.

¹⁰ If we were matching a random set of derivative prices, such as the term structure on a particular day, then derivatives prices would depend not only on F but also on G , the joint cdf of the prices being matched. For example G would be the joint cdf of discount bond prices at the fixed maturities t_1, \dots, t_j . A formal proof in that case goes beyond the scope of this paper.

- (1985b): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407.
- DOTHAN, L. U. (1978): "On the Term Structure of Interest Rates," *Journal of Financial Economics*, 6, 59–69.
- DUFFIE, D., AND R. KAN (1993): "A Yield Factor Model of Interest Rates," Stanford University Mimeo.
- DUFFIE, D., AND K. SINGLETON (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.
- FELLER, W. (1951): "Two Singular Diffusion Problems," *Annals of Mathematics*, 54, 173–182.
- FLORENS-ZMIROU, D. (1993): "On Estimating the Diffusion Coefficient from Discrete Observations," *Journal of Applied Probability*, 30, 790–804.
- FREIDLIN, M. (1985): *Functional Integration and Partial Differential Equations*, Annals of Mathematics Studies, Number 109. Princeton, N.J.: Princeton University Press.
- FRIEDMAN, A. (1964): *Partial Differential Equations of Parabolic Type*. Englewood Cliff, N.J.: Prentice-Hall.
- (1975): *Stochastic Differential Equations and Applications*, Vol. 1. New York: Academic Press.
- (1976): *Stochastic Differential Equations and Applications*, Vol. 2. New York: Academic Press.
- GIBBONS, M. R., AND K. RAMASWAMY (1993): "A Test of the Cox, Ingersoll and Ross Model of the Term Structure," *Review of Financial Studies*, 6, 619–658.
- GOURIÉROUX, C., A. MONFORT, AND E. RENAULT (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.
- HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.
- HANSEN, L. P., AND J. A. SCHEINKMAN (1995): "Back to the Future: Generating Moment Implications for Continuous Time Markov Processes," *Econometrica*, 63, 767–804.
- HARVEY, A. C. (1993): *Time Series Models*, Second Edition. Cambridge, MA: The MIT Press.
- HASMINSKIĬ, R. Z. (1980): *Stochastic Stability of Differential Equations*. Alphen aan den Rijn, The Netherlands: Sijthoff and Noordhoff Publishers.
- HULL, J., AND A. WHITE (1990): "Pricing Interest-Rate-Derivative Securities," *Review of Financial Studies*, 3, 573–592.
- JAMSHIDIAN, F. (1989): "An Exact Bond Option Formula," *Journal of Finance*, 44, 205–290.
- KARATZAS, I., AND S. E. SHREVE (1991): *Brownian Motion and Stochastic Calculus*, Second Edition. New York: Springer-Verlag.
- KARLIN, S., AND H. M. TAYLOR (1981): *A Second Course in Stochastic Processes*. New York: Academic Press.
- KÜNSCH, H. R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.
- LIU, R. Y., AND K. SINGH (1992): "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of Bootstrap*, ed. by R. LePage and Y. Billard. New York: Wiley.
- LO, A. W. (1988): "Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data," *Econometric Theory*, 4, 231–247.
- MARSH, T. A., AND E. R. ROSENFELD (1983): "Stochastic Processes for Interest Rates and Equilibrium Bond Prices," *Journal of Finance*, 38, 635–646.
- MERTON, R. C. (1973): "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4, 141–183.
- (1980): "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.
- (1990): *Continuous-Time Finance*. Cambridge, MA: Blackwell.
- NELSON, D. B. (1990): "ARCH Models as Diffusion Approximations," *Journal of Econometrics*, 45, 7–38.
- ØKSENDAL, B. (1992): *Stochastic Differential Equations*. Third Edition. New York: Springer-Verlag.
- PEARSON, N. D., AND T.-S. SUN (1994): "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model," *Journal of Finance*, 49, 1279–1304.

- PHILLIPS, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–310.
- PRAKASA-RAO, B. L. S. (1987): "On Mixing for Flows of σ -Algebras," Technical Report No. 95, University of California, Davis.
- ROBINSON, P. M. (1983): "Nonparametric Estimators for Time Series," *Journal of Time Series Analysis*, 4, 185–207.
- (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55, 875–891.
- SAID, E. S., AND D. A. DICKEY (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599–607.
- SCOTT, D. W. (1992): *Multivariate Density Estimation: Theory, Practice and Visualization*. New York: Wiley.
- SILVERMAN, B. W. (1986): *Density Estimation for Statistical and Data Analysis*. London: Chapman-Hall.
- VASICEK, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.
- 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, Vol. XVI, ed. by R. Bellman. Providence, RI: American Mathematical Society, 264–276.