Short-Term Interest Rates as Subordinated Diffusions

Timothy G. Conley
Northwestern University

Lars Peter Hansen
University of Chicago and NBER

Erzo G. J. Luttmer
Northwestern University

José A. Scheinkman
University of Chicago

In this article we characterize and estimate the process for short-term interest rates using federal funds interest rate data. We presume that we are observing a discrete-time sample of a stationary scalar diffusion. We concentrate on a class of models in which the local volatility elasticity is constant and the drift has a flexible specification. To accommodate missing observations and to break the link between "economic time" and calendar time, we model the sampling scheme as an increasing process that is not directly observed. We propose and implement two new methods for estimation. We find evidence for a volatility elasticity between one and one-half and two. When interest rates are high, local mean reversion is small and the mechanism for inducing stationarity is the increased volatility of the diffusion process.

We thank Yacine Aït-Sahalia, John Cochrane, Ravi Jagannathan, Tom Kurtz, Dan Nelson, Guy Saidenberg, Andrei Taskin, George Tauchen, Nizar Touzi, and three referees for helpful comments; Renee Adams, Gerald Cubbin, and Alexander Monge for computational assistance; and the National Science Foundation for research support. Address correspondence to Erzo G. J. Luttmer, Department of Finance, J. L. Kellogg Graduate School of Management, Northwestern University, Leverone Hall, 2001 Sheridan Rd., Evanston, IL 60208-2006.

The Review of Financial Studies Fall 1997 Vol. 10, No. 3, pp. 525–577

In this article we provide a characterization of the dynamic evolution of short-term interest rates using data on federal funds interest rates. Our characterization uses Markov diffusion models as data interpretative devices. Markov diffusion models have played a preeminent role in the theoretical literature on the term structure of interest rates [e.g., see Cox, Ingersoll, and Ross (1985)]. The factors underlying the term structure are invariably modeled as having linear drifts (local expected changes) and often linear diffusion coefficients (variances of local changes) as in the square-root process of Feller (1951). The short-term interest rate process that is derived from these factors inherits this linear structure. While such specifications are convenient for deriving and estimating explicit models of the term structure of interest rates, from the viewpoint of data description it is important to specify the drift and possibly the diffusion coefficient in more flexible ways.

A Stochastic Process for Interest Rates

Nonlinearities in time-series models are often introduced through the conditional mean. In this article we also emphasize the role of nonlinearities in the conditional variance as a function of the Markov state. As we show, this dependence of the conditional variance on the Markov state has important consequences for both the local and global behavior of the time-series model. To analyze the role of state dependence in the conditional variances it is convenient to use a scalar diffusion model. We presume that an econometrician observes a discrete-time sample of a process $\{x_t\}$ that is a stationary solution to the stochastic differential equation:

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t, \tag{1}$$

where $\{W_t\}$ is a one-dimensional standard Brownian motion.

While we permit flexible models of the drift μ, our parameterization of the local variance σ^2 is, following Chan et al. (1992), a power function of the current state. The magnitude of the power is one of the key parameters of interest. It gives the elasticity of the local variance with respect to the Markov state. Hence, in addition to characterizing the nonlinearities in the drift, we also present measurements of the local variance elasticity that are not premised on strong a priori restrictions on the drift.

In Section 2 we give local and global characterizations of diffusion models with constant variance elasticities. We show that the local mean reversion is conveniently summarized by first-order expansions of *hitting probabilities*, that is, probabilities of reaching a nearby point to the right of the current location prior to an equidistant point to the left. In particular, the first-order term in the expansion is one-half times the ratio of the drift to the diffusion coefficient ($\mu/(2\sigma^2)$). In studying the global behavior we provide sufficient conditions for stationarity and geometric ergodicity. We show that for some parameter configurations the global behavior is dictated primarily by the volatility elasticity, and that nonlinearities in volatility (σ) may guarantee global stationarity even though the corresponding deterministic process (with the same drift μ, but $\sigma = 0$) would be explosive. This is a particular case of what we call *volatility-induced stationarity*. As a consequence, the magnitude of the variance elasticity is a crucial ingredient to both the local and global characterization of the diffusion process for short-term interest rates.

We treat as unobservable the actual amount of economic activity and information flow that elapses between observations. To permit a difference between "economic time" and actual calendar time, we use a device originated by Bochner (1960) and advocated by Clark (1973) for studying certain financial time series. Specifically, there is an increasing process $\{\tau_t\}$ with stationary and ergodic increments that dictates the sampling scheme. Our model of the observed discrete-time process $\{y_j\}$ is given by the "subordination"

$$y_j = x_{\tau_j}$$

sampled at integer points in time j. Thus subordination introduces a temporally dependent but unobserved sampling scheme for the diffusion process. We delineate the properties of the admissible unobserved $\{\tau_j\}$ process in Section 2.

As we will see, this subordination can be interpreted in a variety of ways. One is that it introduces an unobservable hidden factor that shifts both the drift and the diffusion and hence implies a particular form of a stochastic volatility model. Other researchers have concluded that single-factor models of bond prices appear to be too restrictive [see Longstaff and Schwartz (1992) and Pearson and Sun (1994)]. Alternatively, subordination is a device for accommodating systematically missing data points. For instance, in daily federal funds data there are systematically missing data points due to the presence of weekends and holidays. Moreover, as is evident from the empirical work of Hamilton (1996), transition densities for federal funds rates appear to have some strong periodic components due to institutional constraints on participants in the federal funds market.[1]

We impose stationarity on the underlying process $\{x_t\}$ and on the increments of $\{\tau_j\}$. The process $\{y_j\}$ observed by the econometrician

[1] For other characterizations of periodicity in financial time series see Andersen and Bollerslev (1996) and Bollerslev and Ghysels (1996).

in discrete time inherits this stationarity. On the other hand, we deliberately avoid restricting the process $\{x_t\}$ to have a linear drift or a constant diffusion coefficient. In macroeconometrics one often models short-term interest rates as having a unit root and hence being nonstationary. This choice seems dictated in part by the use of multivariate linear time-series models to summarize the data. The nonlinear time-series model we employ here is stationary, but at the same time allows for unit root-like behavior over much of the support of the data.

New Estimation Methods

We analyze nonlinear diffusion processes using some new estimation methods that are justified in this article. Our econometric implementation uses test functions to discriminate among competing models. These test functions are converted into empirically verifiable moment conditions using an operator $\mathcal{A}$ referred to as an *infinitesimal generator.* This operator summarizes conveniently the local evolution of a Markov process. In the case of a scalar Markov diffusion, it can be built directly from the drift and diffusion coefficients via

$$\mathcal{A}\phi = \mu\phi' + \frac{1}{2}\sigma^2\phi''$$

for appropriately smooth functions ϕ. Hence, any parameterization of the drift and diffusion coefficients translates directly into a parametrization of the infinitesimal generator. We use the moment conditions implied by a collection of test functions to ascertain which models within a parameterized family are empirically plausible.[2] We parameterize the drift in a flexible way and use a constant volatility elasticity parameterization of the diffusion coefficient. A feature of the resulting moment implications is that they are applicable independent of how much calendar time has elapsed between observations. As a consequence, they permit the presence of an independent subordinating process.

Our approach to identification is to initially use the marginal distribution of interest rate levels to extract information about the drift of the diffusion for a range of volatility elasticities. This guarantees that our estimated process fits the long-run properties of interest rate data. We then use the distribution of first differences in interest rates to extract information about the volatility elasticity. By looking at first

[2] Hansen and Scheinkman (1995) show that with a sufficiently rich collection of test functions, the drift and diffusion coefficient in principle can be identified nonparametrically up to scale. We will not be that ambitious here for reasons that will be made clear subsequently.

differences, our second step in part imitates other methods of estimating volatility with two important caveats: we adjust for the presence of a nonzero drift and we permit the presence of subordination. In independent work, Aït-Sahalia (1996b) has used similar identifying information via an entirely different econometric implementation.[3] An advantage of both our methods and that of Aït-Sahalia (1996b) over the generalized method of moments approach employed by Chan et al. (1992) is that they avoid taking a coarse difference-equation approximation to the continuous-time evolution equation.

We justify two new methods for implementing the first step of this procedure in Appendixes C and D. In Appendix C we show that the scores of the marginal distribution of the data are (approximately) the most efficient test functions. Thus this method is linked to the method of maximum likelihood except that we use scores as test functions instead of as moment conditions. From this vantage point, our test function approach provides a simple way to adjust the moment conditions from maximum likelihood estimation for the temporal dependence inherent in diffusion models. This method is described in Section 4. In Appendix D we justify a local extension of the efficient test function estimator. Here we adopt a linear parameterization of the drift, but localize this parameterization by our choice of test functions. This method is motivated in part by the extensive literature on local linear regression and is described in Section 5. Both methods allow us to relax the assumption of a linear drift and are easy to implement. We use the second (local) method as a robustness check for the results obtained from our nonlinear parameterization.

We find that nonlinearities in the drift are important for high-variance elasticities (say elasticities greater than four) but not for low ones. By studying hitting time probabilities, we will argue that even with nonlinearities in the drift, interest rates behave locally like a Brownian motion for all but very low interest rates. This is especially true for the higher volatility elasticities. For the high-volatility elasticities, *mean reversion* (pull to the center of the distribution) is dictated primarily by volatility instead of the drift.

Efficient test functions are not readily available for test functions using first differences. Instead we pick test functions to match up with features of the empirical distribution of first differences. Our resulting measurements of the local variance elasticity for federal funds data suggest that the local variance is substantially more sensitive to the

[3] Instead of looking at moment implications for test functions, Aït-Sahalia used the associated forward and backward partial differential equations. This identification scheme is closely related to that in this article when localized versions of test functions are used.

level of these rates than would be the case for the widely used square-root model. This finding is consistent with that of Chan et al. (1992).

1. Constant Volatility Elasticity Diffusion

In our empirical analysis we focus exclusively on the constant volatility elasticity parameterization of the diffusion coefficient for $\{x_t\}$:

$$\sigma^2(x) = \kappa x^\gamma \tag{2}$$

where γ denotes the local variance elasticity. Because of the presence of the subordination, the scale parameter κ is not identified.[4] We normalize it to be one. Our parameterization for the local variance follows, among others, Chan et al. (1992).[5] A variance elasticity of one is consistent with the Feller (1951) square-root process used by Cox, Ingersoll, and Ross (1985) and others. An elasticity of two is consistent with a lognormal specification of the short-term interest process. In contrast to Chan et al. (1992), we do not confine our attention to linear specifications of the drift μ. Instead we start off assuming that μ is a continuous function of the state, and subsequently impose additional restrictions motivated by the implied local and global properties of the diffusion. We argue that *mean reversion* depends on the volatility elasticity as well as on the drift, and we show that high-volatility elasticities may be sufficient to induce stationarity.

While it is most common to construct the scalar diffusion as the solution to a stochastic differential equation, there is an alternative way to build a scalar diffusion from a Brownian motion. This construction leads to a natural measure of mean reversion that takes into account the local variance, in addition to the drift. For convenience, we describe the inverse operation of reducing the scalar diffusion to a Brownian motion via two transformations: a transformation of the scale of the process and a state-dependent transformation of the time scale or speed of the process. By design the scale transformation eliminates the local mean of the process and hence reduces it to a local martingale. The speed transformation eliminates the state dependence in the transformed diffusion coefficient converting the local martingale into a Brownian motion. As we will see, the scale function encodes the amount of "pull" to the right or left of the diffusion. In turn, the speed density is closely related to the stationary density.

---

[4] In the absence of subordination, Hansen and Scheinkman (1995) proposed using an eigenvector for the generator to identify the parameter κ.

[5] Note that Chan et al. (1992) parameterize the elasticity of the standard deviation, whereas we are parameterizing the elasticity of the variance. Hence their estimates must be multiplied by two to be comparable to ours.

1.1 Scale density

Consider first the scale function S. This function is strictly increasing with derivative or density given by

$$s(x) = \exp\left[\int^{x} -\frac{2\mu(y)}{\sigma^2(y)}\, dy\right], \tag{3}$$

where the lower endpoint of the integration is arbitrary. As a consequence the scale density is only defined up to a scale factor. As we indicated previously, the transformed process $\{S(x_t)\}$ has a zero drift and hence is a local martingale. This local martingale has a particularly nice interpretation. Consider two points $0 < r_1 < r_2 < \infty$ and an initialization point y in the interval (r_1, r_2). Suppose the process is stopped whenever either r_1 or r_2 is hit. Then the local martingale gives the stochastic process of conditional probabilities of hitting r_2 prior to hitting r_1. The scale density encodes this information via the following formula. Starting from state y, the probability of hitting r_2 prior to hitting r_1 is given by the ratio: $\int_{r_1}^{y} s(x)\, dx / \int_{r_1}^{r_2} s(x)\, dx$ [e.g., see Karlin and Taylor (1981, p. 195)]. Hence the scale density provides probabilistic measures of the "pull" to the right or the left of the diffusion at alternative points.

1.2 Local hitting time probabilities

Of course, hitting time probabilities are sensitive to the hitting points r_1 and r_2. Instead of specifying these points, in our empirical investigation, we report local hitting time probabilities. Consider an interval centered around an initial condition and compare the probabilities of reaching the right boundary before the left boundary. Recall that for a Brownian motion this probability is equal to one-half. More generally, because of the local Brownian motion character of a diffusion, this probability converges to one-half as the length of the interval shrinks to zero. Using the scale density, the probability of reaching $y + \varepsilon$ before $y - \varepsilon$ starting from y is

$$\pi_y(\varepsilon) \equiv \int_{y-\varepsilon}^{y} s(x)\, dx \Big/ \int_{y-\varepsilon}^{y+\varepsilon} s(x)\, dx.$$

A simple computation shows that

$$\pi_y(\varepsilon) = \frac{1}{2} + \frac{\mu(y)}{2\sigma^2(y)}\varepsilon + o(\varepsilon).$$

Hence $\mu/(2\sigma^2)$ measures, up to first order, the pull to the right. This should be contrasted with the drift μ, which measures the pull per unit time in a corresponding deterministic system obtained by zeroing out the uncertainty. There are two reasons that make the drift μ an unappealing measure of pull in our setting. First, in the presence of subordination, μ is not really a per unit calendar time measure. More problematic is the fact that μ abstracts from the role of uncertainty. As we have just shown, by measuring "pull" using local hitting time probabilities, the contribution of μ is downweighted when σ^2 is large (as a function of the state). This correction occurs because the variance level affects the probability that the system moves toward the center of its stationary distribution.

1.3 Boundary classification

Another use of the scale function is to classify the boundaries 0 and $+\infty$. A boundary is said to be attracting when the scale function (the integral of the scale density) is finite at the boundary point. We impose:

Assumption 1. *$\int_0^x s(y)\,dy = +\infty$ and $\int_x^\infty s(y)\,dy = +\infty$ for any $x \in (0, \infty)$.*

In other words, for our models neither boundary is attracting. For this to be the case, the scale density s must diverge at 0. While the scale density is permitted to converge at the infinite boundary, it cannot converge to zero too quickly. Assumption 1 is sufficient for there to exist a weak solution to the stochastic differential equation given by Equations (1) and (2), and this solution is unique in probability law for any prespecified initial distribution [Theorems 5.7 and 5.13 of Karatzas and Shreve (1991), p. 335].

1.4 Speed density

As we noted previously, by changing the speed of the process we can transform $\{S(x_t)\}$ into a Brownian motion and conversely. This change in speed has the interpretation of subordinating $\{S(x_t)\}$ to an increasing process that depends on the Markov state. More specifically, define the cumulative variance process for $\{S(x_t)\}$;

$$\zeta(t) = \int_0^t \sigma^2(x_s)s^2(x_s)\,ds.$$

Let $\nu = \zeta^{-1}$. We eliminate heteroscedasticity by "inverting" the cumulative variance process and forming $\{S(x_{\nu(t)})\}$. The derivative of ν with respect to time is $1/\{\sigma^2(x_{\nu(t)})s^2(x_{\nu(t)})\}$, which depends only on the state $x_{\nu(t)}$. Measuring this speed transformation in terms of the transformed state gives rise to the *speed density*:

$$\hat{m}(v) = \frac{1}{\sigma^2[S^{-1}(v)]s^2[S^{-1}(v)]}.$$

To guarantee that the diffusion is stationary, we restrict the speed measure to be finite. By the change of variables formula, the corresponding speed density in the original scale is

$$m = \frac{1}{\sigma^2 s}. \tag{4}$$

Assumption 2. $\int_0^\infty \frac{1}{s(y)y^\gamma}\, dy < \infty.$

We initialize the diffusion with a probability density q proportional to the speed density. Given Assumption 2, it is known that with this initialization the diffusion is stationary with stationary density q [e.g., see Karlin and Taylor (1981), p. 221]. In Appendix A we present additional ergodicity restrictions used to justify our large sample approximations in making statistical inferences.

Taking logarithmic derivatives of Equation (4) and using the constant elasticity specification of the diffusion coefficient, it can be verified that the local hitting time correction $\mu/(2\sigma^2)$ satisfies

$$\frac{\mu(y)}{2\sigma^2(y)} = \frac{\gamma}{4y} + \frac{q'(y)}{4q(y)}. \tag{5}$$

Notice that for large y the local measure of pull is approximately equal to $q'/(4q)$. Thus the hitting time correction is approximately zero for large interest rates when the logarithmic derivative of the stationary density approaches zero.

1.5 Volatility-induced stationarity

Some constant volatility elasticity processes are stationary even though the drift is not strictly negative for large interest rates. Assumption 2 does not require the scale density to satisfy:

$$\lim_{y \nearrow +\infty} \inf \; s(y) = +\infty \tag{6}$$

as long as the variance elasticity exceeds one. In particular, the scale density may converge to a finite number. It follows from Equation (3) that this convergence occurs when

$$\int_x^\infty \frac{\mu(y)}{y^\gamma}\, dy > -\infty \tag{7}$$

for some strictly positive x. In cases in which (7) is satisfied and the resulting process is stationary, we say that stationarity is *volatility induced.* Volatility-induced stationarity permits the drift to be positive for large values of the state. In this case the deterministic counterpart obtained by setting $\kappa = 0$ will be explosive.

Volatility-induced stationarity occurs when the drift $\mu \approx cy^{\theta}$ for large values of the Markov state y, and $\theta < \gamma - 1$. Notice that c may be positive in this construction. In these cases the local hitting time correction $\mu/\sigma^2 \approx cy^{\theta-\gamma}$ not only converges to zero but does so faster than y^{-1}. One special case of volatility-induced stationarity is when the drift is constant and positive and the variance elasticity exceeds one. Another is when the drift is linear in y and the variance elasticity exceeds two. These cases are of interest to us because Chan et al. (1992) estimate a high volatility elasticity for short-term interest rates.

2. Subordination

The process observed by the econometrician is constructed by subordinating the scalar diffusion $\{x_t\}$ to a strictly increasing process $\{\tau_j\}$ with stationary increments. What is observed is

$$y_j = x_{\tau_j}.$$

We focus exclusively on identifying and estimating the dynamics governing the process $\{x_t\}$ while permitting the presence of subordination. We do not need to nor do we attempt to model the dynamics of the sampling process $\{\tau_j\}$; it is treated as an unobservable sequence. Thus our fixed interval sample $\{y_j\colon\ j = 1, 2, \ldots, T\}$ is an ordered sequence of observations on the focal process $\{x_t\}$ with unknown gaps of time between observations.

Assumption 3. *The process $\{\tau_j\colon\ j = 0, 1, \ldots\}$ satisfies $\tau_0 = 0$, is increasing, independent of the scalar diffusion $\{x_t\colon\ t \geq 0\}$ and has stationary increments.*

One possibility is that $\{\tau_j\}$ is constructed via the integral:[6]

$$\tau_j = \int_0^j z_s\, ds, \tag{8}$$

where $\{z_t\}$ is stationary. In this case the continuous-time subordinated process is an Ito process with local mean $z_t\mu(y_t)$ and local variance $z_t\sigma^2(y_t)$.[7] Thus when $\{\tau_j\}$ is given by Equation (8), subordination has the effect of being a random proportional shift in the local mean and variance of the original scalar diffusion. When $\{z_t\}$ is also a scalar

[6] Clark (1973) referred to this process as a "directing process" and did not presume that the original process is stationary.

[7] To justify this conclusion, we may use Theorem 2.4 of Revuz and Yor (1991, p. 283) in conjunction with a change of variables (in the time scale) to show that $\{y_t\}$ is an Ito process conditioned on the entire process $\{\tau_j\}$ and hence unconditionally. This conditioning argument works since $\{\tau_j\}$ and $\{x_t\}$ are independent.

diffusion, we obtain a particular form of a stochastic volatility model by introducing a second factor: "economic time."

More generally, the process $\{\tau_j\}$ may have sample paths with discontinuities. Subordination also allows for the presence of periodic patterns in the sampling process (or directing process) $\{\tau_j\}$ due to the presence of weekends, holidays, or day of the week effects. The increment process in our analysis can be deterministic or have deterministic components so long as these components have a systematic pattern that could emerge as a realization from a stationary process.[8]

The independence of the processes $\{x_t\}$ and $\{\tau_j\}$ guarantees that the stationary distribution of the diffusion is also the stationary distribution of the subordinated process $\{y_j\}$. However, the transition densities are not preserved and the subordinated process is not typically Markovian. For instance, the transition densities can have fatter tails for the reasons articulated by Clark (1973).

As we shall see, the diffusion characterization of the underlying process $\{x_t\}$ simplifies the task of constructing econometric estimators, and the independence of the sampling process from the process of interest is critical for justifying these estimators. The large sample approximations we use to make statistical inferences require further restrictions on the subordinating process that are discussed in Appendix A.

3. Identifying the Drift

As argued previously by Banon (1978) and Cobb, Koppstein, and Chen (1983), the drift μ of the scalar diffusion can be identified (nonparametrically) from this distribution for a given volatility elasticity. Following Hansen and Scheinkman (1995), we depict this identification by using moment implications associated with a collection of prespecified test functions. The vehicle for building moment conditions from the test functions is the *infinitesimal generator* for a candidate Markov process. The infinitesimal generator is an operator that gives the local or instantaneous transition probabilities implied by the candidate process. In the case of a scalar diffusion, the generator can be constructed from the drift and diffusion coefficients, or equivalently from the scale and speed densities. In what follows, we present the infinitesimal generator for a scalar diffusion followed by two different collections of moment conditions used to identify the drift μ. One collection is used in combination with a global parameterization

[8] Recall that periodic functions can be viewed as stationary and ergodic processes by randomizing appropriately the initialization.

of μ. This results in an estimation method similar to that of Cobb, Koppstein, and Chen (1983). The second collection is local in nature, and the resulting estimation method is similar to that of Banon (1978).

3.1 Infinitesimal generator

Let $\mathcal{Q}$ denote the stationary distribution induced by x_t. We define $\mathcal{L}^2(\mathcal{Q})$ to be the space of all Borel measurable functions mapping $\mathbb{R} \to \mathbb{R}$ that are square integrable with respect to $\mathcal{Q}$. The infinitesimal generator $\mathcal{A}$ for a Markov process is the time derivative of the conditional expectation operator. For some functions, this time derivative may not exist. In the case of a scalar diffusion, the subspace of functions for which this derivative is well defined is given by

$$\begin{aligned}\mathcal{D} = \{\phi \in \mathcal{L}^2(\mathcal{Q}):\ & \phi' \text{ is absolutely continuous;} \\ & \mu\phi' + \tfrac{1}{2}\sigma^2\phi'' \in \mathcal{L}^2(\mathcal{Q}) \\ & \lim_{y\searrow 0} \tfrac{\phi'(y)}{s(y)} = \lim_{y\nearrow\infty} \tfrac{\phi'(y)}{s(y)} = 0\}\end{aligned}$$

[see Hansen, Scheinkman, and Touzi (1996)]. For functions $\phi \in \mathcal{D}$, an implication of (a generalized version of) Ito's formula is

$$\mathcal{A}\phi = \mu\phi' + \frac{1}{2}\sigma^2\phi'' = \frac{1}{2m}\left(\frac{\phi'}{s}\right)'.$$

3.2 Moment conditions

Consider a test function ϕ in the domain $\mathcal{D}$ of the generator. Since $\{x_t\}$ is stationary, we know that $E[\phi(x_t)]$ is constant over time and hence has a zero derivative. This in turn implies that

$$E[\mathcal{A}\phi(x_t)] = E\left[\mu(x_t)\phi'(x_t) + \frac{1}{2}\sigma^2(x_t)\phi''(x_t)\right] = 0. \tag{9}$$

That is, the local mean of the process $\{\phi(x_t)\}$ has a zero expectation. When stationarity is volatility induced, a linear test function is not in the domain of the generator. This follows because ϕ'/s cannot converge to zero for large values of the Markov state when ϕ is linear. In these cases it is not necessarily true that the drift μ has mean zero under the stationary distribution. In fact, μ could be strictly positive over $(0, \infty)$.[9]

[9] As a consequence, the identification scheme of Aït-Sahalia (1996a) for the diffusion coefficient is not applicable when stationarity is volatility induced.

Since Equation (9) is an implication for the stationary distribution, it follows that it also holds for the subordinated process $\{y_j\}$:

$$E\left[\mu(y_j)\phi'(y_j) + \frac{1}{2}\sigma^2(y_j)\phi''(y_j)\right] = 0. \tag{10}$$

There is great flexibility in the choice of test functions as any function in $\mathcal{D}$ can be used.

3.3 Efficient test functions for a global parameterization

The first collection of test functions we use is motivated by efficiency considerations. For simplicity, suppose we have a continuous record of data between time 0 and time T, and that we directly observe the diffusion (without subordination). We establish in Appendix C that an efficient vector of test functions is given by the score vector of the implied parameterization of the stationary density. That is, let $q(y; a)$ denote the density evaluated at state y and hypothetical parameter vector a. The score vector is defined to be

$$d(y; a) \equiv \frac{\partial}{\partial a} \log q(y; a).$$

In the stylized setting of a continuous record, if each component d is in the domain of the generator for the true parameter vector, then the resulting test function estimator (using $E[\mathcal{A}d] = 0$) will be statistically efficient. In particular, it will be more efficient than a quasi-maximum likelihood that uses moment conditions $E[d] = 0$. In effect, the application of the generator to the score vector adjusts optimally the moment conditions for the presence of temporal dependence.[10]

Although our justification is based on the simplifying assumption of a continuous record, we employ the same test functions for our discrete-time sample of a subordinated diffusion. While we hope the optimality remains approximately true (see Appendix C for a discussion), our discrete-time implementation still results in a consistent and asymptotically normal parameter estimator.

3.4 Localized test functions

When the parameterization of the model is only presumed to work locally, we localize the first derivatives of the previously constructed test functions with respect to y. To localize about the point x, we multiply the first derivative by a smooth kernel K evaluated at

[10] The optimality problem being solved is a second-best problem where we limit the class of estimators to those using only moment conditions obtained from Equation (10).

$(y - x)/\delta$. The kernel K is a differentiable probability density function with compact support. As δ gets small, attention is focused on data points in the vicinity of x, and hence δ dictates the domain of the local parameterization. The corresponding test functions are obtained by integrating up the localized first derivative, although it is unnecessary to perform the integration to implement our estimation method.

To relate this test function approach to the identification scheme of Banon (1978), suppose we take the drift to be locally constant in the vicinity of x. It can be shown in this case that the first derivative of the score of this drift parameter with respect to the state is just $2/\sigma^2(y)$. As δ declines to zero, Equation (10) becomes the first-order differential equation:

$$\frac{\mu(x)}{\sigma^2(x)} - \frac{(\sigma^2)'(x)}{2\sigma^2(x)} - \frac{q'(x)}{2q(x)} = 0. \tag{11}$$

As emphasized by Banon (1978), for a given specification of σ^2 and a given logarithmic derivative of the density, the drift μ can be chosen to satisfy Equation (11). Consequently, armed with just the stationary distribution of the data, we can accommodate any volatility elasticity by selecting the drift appropriately.

4. A Global Model of the Drift

We imitate Cobb, Koppstein, and Chen (1983) by adopting a "flexible" global parameterization of the drift μ in hopes that this parameterization does not "distort" the estimation of γ.

4.1 Parameterization

We parameterize the local mean as

$$\mu(x; \alpha) = \sum_{i=-k}^{\ell} \alpha_i x^i \tag{12}$$

where $\alpha \equiv [\alpha_{-k}, \alpha_{-k+1}, \ldots, \alpha_\ell]$. We assume that $\ell \geq 1$ and $k \geq 0$. When these inequalities are satisfied with equality, the local mean is linear. In our empirical work the most flexible parameterization we consider has $\ell = 2$ and $k = 1$. This still allows for some nonlinearities in the drift. It is straightforward to verify that the following joint restrictions on α_{-k} and α_ℓ suffice for Assumptions 1, 2, and 4 in Appendix A to be satisfied:

- Restrictions on the left boundary

  (i) for $k = 0$ and $0 < \gamma < 1$: no possible choice of α_{-k};[11]

  (ii) for $k = 0$ and $\gamma = 1$: $\alpha_{-k} > 1/2$;

  (iii) for $k = 0$ and $\gamma > 1$: $\alpha_{-k} > 0$;

  (iv) for $k \geq 1$ and $\gamma > 0$: $\alpha_{-k} > 0$;

- Restrictions on the right boundary

  (v) for $\gamma < \ell + 1$: $\alpha_\ell < 0$;

  (vi) for $\gamma = \ell + 1$: $\alpha_\ell < 1/2$;

  (vii) for $\gamma > \ell + 1$: no restriction on α_ℓ.

In case (vii) the integrability of the speed density is coming from the contribution of the diffusion coefficient because the scale density converges to a constant for large x. In other words, we have volatility-induced stationarity. As emphasized by Cobb, Koppstein, and Chen (1983), the family of implied stationary densities not only embeds the gamma density associated with Feller's square-root process but also has considerable flexibility to accommodate multiple modes and approximate a rich family of density functions.

4.2 Test function estimator

To implement our efficient test function estimator, we must compute the first two derivatives (with respect to y) of the parameterized score vector. For our parameterization these derivatives do not depend on the unknown parameter vector α. For this reason, we abuse notation by dropping the dependence of the score vector d on a. The first derivative is

$$d'(y) = \frac{\partial^2}{\partial y \partial a} \log q(y; \alpha) = \frac{\partial}{\partial a} \left[\frac{2\mu(y; \alpha)}{y^\gamma} \right] = \begin{bmatrix} 2y^{-k-\gamma} \\ 2y^{-k-\gamma+1} \\ \cdots \\ 2y^{\ell-\gamma} \end{bmatrix}.$$

Suppose the data available to estimate the parameter vector α are a sample from the subordinated process: $\{y_j\colon\ j = 1, 2, \ldots, T\}$. Our

[11] When $\alpha_0 < 0$, it is possible to build a diffusion model with a reflecting barrier at zero that is stationary and has the prescribed local means and variances. Having a reflecting barrier at zero for short-term interest rates is not an empirically plausible model, however.

discrete-time estimator is obtained by solving the linear equation

$$\frac{1}{T}\sum_{j=1}^{T} d'(y_j) z_j^* a_T + \frac{1}{2T}\sum_{j=1}^{T} d''(y_j)\sigma^2(y_j) = 0 \tag{13}$$

where $*$ denotes the transpose and $z_j = [y_j^{-k} \quad y_j^{-k+1} \quad \cdots \quad y_j^{\ell}]^*$. Therefore

$$a_T = -\left[\frac{1}{T}\sum_{j=1}^{T} d'(y_j) z_j^*\right]^{-1} \frac{1}{2T}\sum_{j=1}^{T} d''(y_j)\sigma^2(y_j). \tag{14}$$

The limiting distribution for this estimator is given by

$$\sqrt{T}(a_T - \alpha) \Rightarrow \mathcal{N}(0, G^{-1}\Lambda G^{-1*})$$

where $\Rightarrow$ denotes convergence in distribution, and where

$$G = E[d'(y_j) z_j^*]$$

and Λ is the spectral density matrix at frequency zero for the multivariate discrete-time process $\{d'(y_j)\mu(y_j) + \frac{1}{2}d''(y_j)\sigma^2(y_j)\colon\ j = 1, 2, \ldots\}$. The spectral density matrix appears because of the central limit approximation,

$$\sqrt{T}\sum_{j=1}^{T}\left[d'(y_j)\mu(y_j) + \frac{1}{2}d''(y_j)\sigma^2(y_j)\right] \Rightarrow \mathcal{N}(0, \Lambda),$$

which is justified in Appendix B. The matrix G can be estimated consistently by a time-series average and the spectral density matrix Λ by standard time-series methods.

Our estimator is similar to that of Cobb, Kobbstein, and Chen (1983) except that they employ polynomial test functions with positive integer powers. For optimality, we let the polynomial powers depend on the variance elasticity. As a result, we do not have to preclude volatility-induced stationarity ($\gamma > \ell + 1$).

4.3 Estimation results

We use a data set of federal funds interest rates. It consists of daily observations on "effective" federal funds overnight interest rates, measured as annualized percentages, from January 2, 1970, to January 29, 1997.[12] The data set has systematically missing observations, primarily due to weekends and holidays. As discussed in Section 2, we take

[12] We thank the Federal Reserve Bank of Chicago for providing the data.

the discrete-time observations $\{y_j\}$ to be generated from an underlying interest rate process $\{x_t\}$ and sampling scheme $\{\tau_j\}$ according to $y_j = x_{\tau_j}$.

The drifts are inferred from Equations (12) and (14) for integer values of the variance elasticity parameter γ between zero and six. The least restricted specification of the drift we consider, with $k = 1$ and $\ell = 2$, has four free parameters [see Equation (12)]. Using this parameterization we consider drifts with more pull for both large and small interest rates than the commonly used linear specification [$k = 0$, $\ell = 1$ in Equation (12)]. Estimates of the drift parameters for our least restricted specification are contained in Appendix G. Here we report estimates of the implied stationary densities (or equivalently speed densities) and our local hitting time measure $\mu/(2\sigma^2)$. The implied density estimates are reported in Figure 1. For comparison, we also include a coarse histogram of the data in the background. Notice that for variance elasticities between zero and three, the implied density estimates track each other remarkably closely. For larger variance elasticities (in excess of four) the implied densities are bimodal. The possibly spurious extra hump becomes more pronounced for larger values of γ. In Figure 2 we report the implied densities when the drift is constrained to be linear. Not surprisingly, the implied densities are much more sensitive to the variance elasticity specification. Nonlinearities seem to be particularly important for very small elasticities and for elasticities in excess of four. On the other hand, when the variance elasticity is one or two, the implied densities for the linear and nonlinear drift models are quite close.

In Figure 3 we report the measure of pull to the right $\mu/(2\sigma^2)$ as a function of the interest rate level. For moderate variance elasticities, there is pull to the center of the distribution for both small and large interest rates. Larger variance elasticities magnify the pull from the left and diminish the pull from the right. In particular, for high variance elasticities, large (but historically observed) interest rate behavior is close to that of a continuous (local) martingale. On the other hand, there is a very substantial pull to the right for small interest rates. Equation (5) indicates that $\mu/(2\sigma^2)$ should increase as the variance elasticity increases. Since our estimate of $q'/(4q)$ varies with the variance elasticity, this monotonicity does not hold over the entire state space. Moreover, since the large interest rate limit of $q'/(4q)$ depends on the parameterization, the corresponding large interest rate hitting time corrections are not constrained to be equal. In Figure 4 we report our estimates of $\mu/(2\sigma^2)$ along with two-standard error bands for variance elasticities one and four. For $\gamma = 1$, the standard error bands widen for large and small interest rate levels. For $\gamma = 4$, the standard error bands widen substantially for small interest rate levels, but not

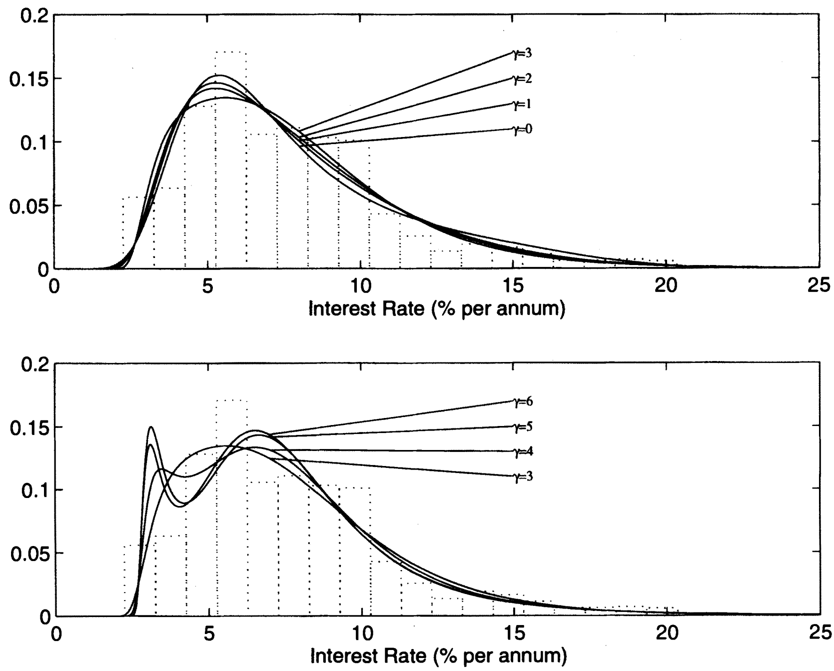

Figure 1
Implied stationary density estimates for federal funds data with the nonlinear drift parameterization and local variance elasticities $\gamma = 0, 1, \ldots, 6$. Background histograms of the data are scaled to integrate to one.

for large interest rates.[13] This latter finding is being driven in part by the fact that when $\gamma = 4$, the parameterization of the drift guarantees that the estimate of $\mu/(2\sigma^2)$ converges to zero by construction as the Markov state gets large.

4.4 Specification testing

To complement our depiction of nonlinearities in the drift coefficient, we now present chi-square tests of the model specification based on the marginal distribution of the data. We are compelled to use different functions for testing than for estimation for the following reason. Suppose the true model has a linear drift and a variance elasticity of one (Feller's square-root process). It turns out that one of the efficient test functions for the nonlinear mean specification we use may fail to be in the domain of the generator and hence may not be a valid

[13] Although not evident from Figure 4, the standard errors increase as the interest rate levels decrease in the left portion of the figure.

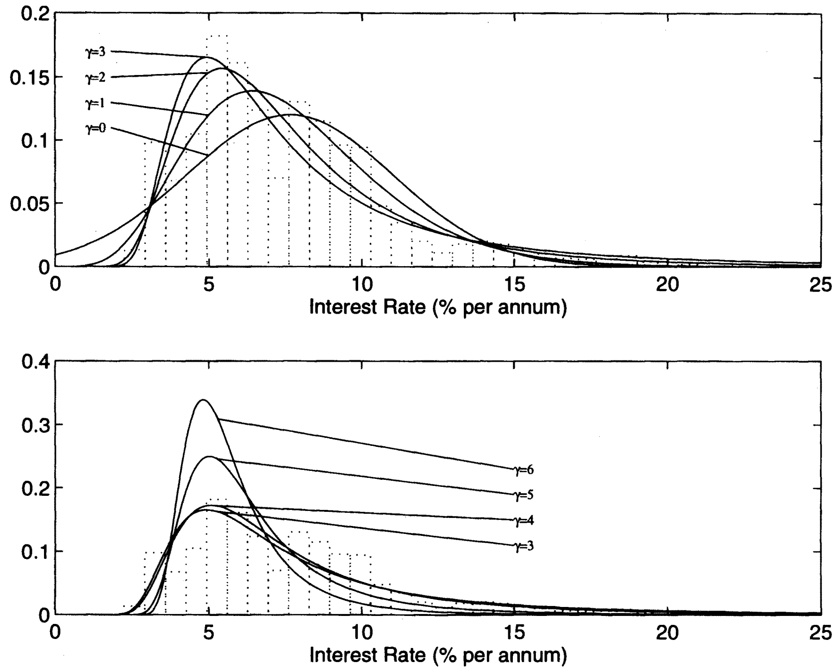

Figure 2
Implied stationary density estimates for federal funds data with the linear drift parameterization and local variance elasticities $\gamma = 0, 1, \ldots, 6$. Background histograms of the data are scaled to integrate to one.

test function in this special case. To circumvent this problem, for the purpose of testing we build test functions that are cumulative normal distributions located at different points of the empirical distribution of the data. We then use standard GMM tests of model specification based on Equation (10) where a chi-square criterion is minimized by choice of the parameters, subject to the sign restrictions discussed after Equation (12). The minimized value of the criterion function has a chi-square distribution with degrees of freedom given by the difference between the number of test functions used and the parameters estimated.[14]

We computed chi-square tests using six and eight cumulative normal test functions for both the linear and the nonlinear drift spec-

[14] The imposition of sign restrictions on the drift coefficients can have a nontrivial implication for the limiting distribution of the coefficients and the GMM tests. Haberman (1989, p. 1645) characterizes the limiting distribution as a possibly nonlinear function of a normally distributed random vector. Wolak (1991) gives bounds on the limiting distribution of the GMM test statistic when nonnegativity constraints are imposed.

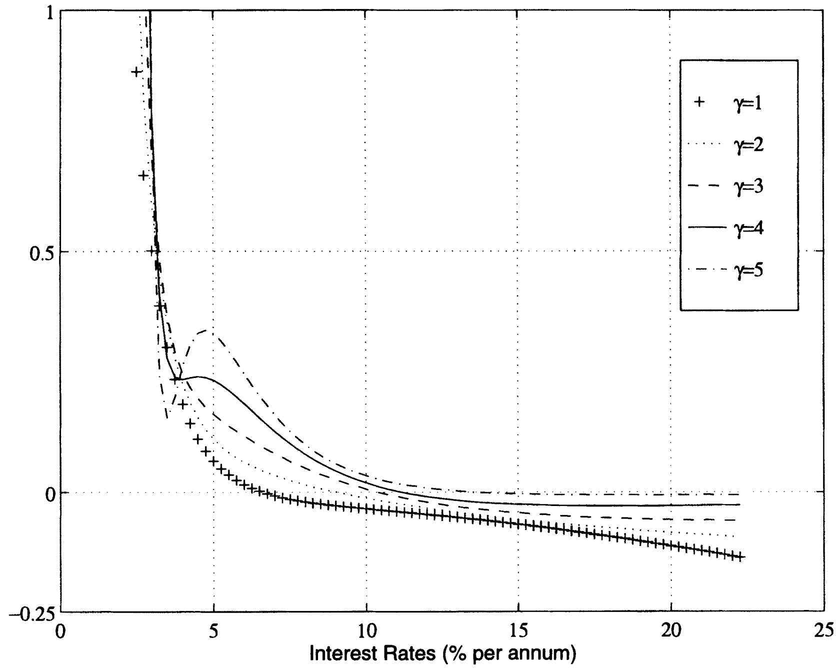

Figure 3
Federal funds pull ($\mu/2\sigma^2$) estimates for the nonlinear drift parameterization and local variance elasticities $\gamma = 1, 2, 3, 4, 5$.

ification. The cumulative normal test functions are adapted to the empirical distribution of $\{y_j\}$ as follows. Let the $(n+1)$-vector ζ contain the quantiles of the empirical distribution associated with the probabilities $.16/n, 1/n, 2/n, \ldots, (n-1)/n, 1-.16/n$. Notice that we deliberately avoid adapting to the smallest and largest interest rates (empirical quantiles zero and one, respectively). Instead we use probabilities $.16/n$ and $1-.16/n$ for the first and last quantiles simply because .16 is the probability that a standard normal random variable exceeds one standard deviation. We take the mean of the ith test function to be $\{\zeta(i+1)+\zeta(i)\}/2$. Its standard deviation is taken to be $\{\zeta(i+1)-\zeta(i)\}/2$. This construction avoids too much overlap between test functions and ensures that some of the test functions are centered close to the upper and lower tails of the empirical distribution, even for small n.

The results are reported in Table 1. In the case of six test functions, there is relatively little change in the chi-square statistics as we change the variance elasticities when the drift is nonlinear. This is not surprising, since we know that with a flexible enough specification

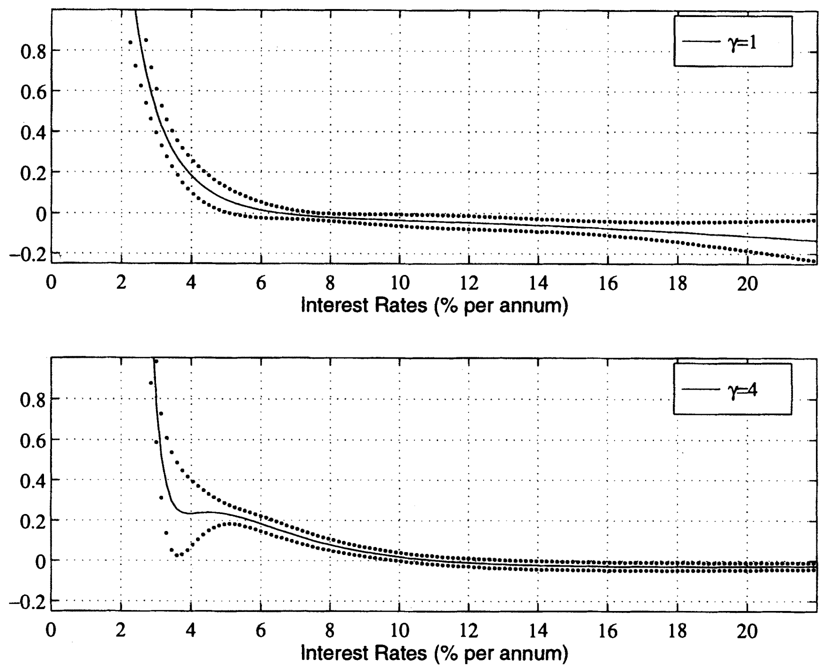

Figure 4
Federal funds pull ($\mu/2\sigma^2$) estimates with two standard error bands for the nonlinear drift parameterization and local variance elasticities $\gamma = 1, 4$. Standard error bands are based on spectral density estimates that were calculated as weighted averages of autocovariances using a Bartlett window that included 60 time lags.

of the drift, we can match an arbitrary stationary density. When the drift is constrained to be linear, there is evidence against the highest variance elasticities of five and six. When we expand the number of test functions to eight, we find that sign restrictions on the parameter estimates are binding for all variance elasticities. Thus the reference to the chi-square distribution in the table heading is particularly misleading in this case. Instead, following Wolak (1991), upper bounds on the probability values may be deduced from chi-square distributions with additional degrees of freedom (corresponding to the additional constraints that are imposed). Taking this into account, there is little evidence against the nonlinear drift model, except possibly for the highest variance elasticities.[15] As before, the linear drift specification is implausible for the two highest volatility elasticities.

[15] While our six test function results were robust to increasing the number of time-domain lags in the Bartlett window, the same was not true for the eight test function results. Using eight test functions and a nonlinear specification of the drift, we found that when more time lags were added to our spectral density estimate, the resulting test statistics declined significantly.

Table 1
Chi-square statistics for specification testing

	6 test functions	6 test functions	8 test functions	8 test functions
γ	nonlin. $\mu : \chi^2(2)$	linear $\mu : \chi^2(4)$	nonlin. $\mu : \chi^2(4)$	linear $\mu : \chi^2(6)$
0	4.12	6.18	8.47 ∘ ∘	11.98
1	4.14	4.82	9.01∘	9.10
2	4.21	5.72	9.33*	10.50
3	4.35	7.18	10.04 * *	14.42
4	4.60	6.13*	11.92 * *	11.98*
5	4.75	18.62*	9.13 * *	26.45*
6	4.63*	38.76*	10.07*	50.76*

GMM criterion function estimates for linear and nonlinear drift specifications using federal funds data. The columns labeled "nonlin. μ" contain the chi-square statistics for the four-parameter nonlinear drift model, and columns labeled "linear μ" contain the chi-square statistics for the two-parameter linear drift model. The numbers in $\chi^2(\cdot)$ give the degrees of freedom for the chi-square statistics. The symbols $*$ and $\circ$ indicate that a parameter constraint is binding on the lowest and highest power terms of the drift, respectively. Multiple symbols indicate that multiple terms are constrained. Weighting matrices were calculated as weighted averages of autocovariances using a Bartlett window that included 60 time lags. We iterated on the weighting matrix four times starting from an identity matrix.

5. A Local Model of the Drift

As an alternative way to obtain a flexible model of the drift, we may localize our parameterization of the drift through our choice of test functions. Suppose the drift is twice continuously differentiable in the Markov state. We exploit this smoothness by using a linear parameterization for the drift locally centered at each hypothetical state x:

$$\mu(y) \approx \alpha_0 + \alpha_1(y - x)$$

Implicit in this notation is that $\alpha_0 = \mu(x)$ and $\alpha_1 = \mu'(x)$. We estimate a different value of α_0 and α_1 for each value of the Markov state x.

For a globally linear parameterization, we previously showed that the efficient test functions have first derivatives given by $2/\sigma^2(y)$ and $2(y - x)/\sigma^2(y)$. As described earlier, we make the linear parameterization of $\mu(y)$ local at x by multiplying two linear combinations of these derivatives by a kernel density K with compact support, evaluated at $(y - x)/\delta$. Specifically, we use a pair of test functions with derivatives given by the vector $\Psi'(y; \delta)$:

$$\Psi'(y; \delta) = \frac{1}{\sigma^2(y)} \begin{bmatrix} 1 \\ \frac{y-x}{\delta} \end{bmatrix} \frac{1}{\delta} K\left(\frac{y - x}{\delta}\right)$$

Thus the density $K((y - x)/\delta)/\delta$ is centered at x and δ is a dispersion parameter. Using Equation (10) and substituting the local approximation for the drift, we have that

$$E\left[\Psi'(y_j;\delta)\left[1 \quad \tfrac{y_j-x}{\delta}\right]D(\delta)\alpha + \frac{1}{2}\sigma^2(y_j)\Psi''(y_j;\delta)\right] \approx 0 \qquad (15)$$

where $\alpha = [\alpha_0 \quad \alpha_1]^*$ and $D(\delta) = \text{diag}\{1, \delta\}$. Relation (15) gives us two moment conditions that are linear in the two unknown parameters α_0 and α_1. The presence of the approximation error leads to a bias term in our local linear estimator of the drift, the magnitude of which is dictated by our choice of δ and, for a kernel K with compact support, the size of the second derivative of μ in the vicinity of x. Our local linear estimator of the drift coefficient vector α is

$$a_T = -\{D(\delta_T)\}^{-1}A_T^{-1}U_T$$

where

$$A_T = \frac{1}{T}\sum_{j=1}^{T}\Psi'(y_j;\delta_T)\left[1 \quad \tfrac{y_j-x}{\delta_T}\right],$$

and

$$U_T = \frac{1}{2T}\sum_{j=1}^{T}\sigma^2(y_j)\Psi''(y_j;\delta_T)$$

for some specification of δ_T.

The kernel we use in our implementation is

$$K(z) \propto \begin{cases}(z^2-1)^2 & \text{for } |z|<1 \\ 0 & \text{otherwise}\end{cases}.$$

This estimation method is most likely related to that of Banon (1978). Banon's method uses a kernel estimator of the stationary density to build a nonparametric estimator of the drift for a given specification of the local variance. His method can be viewed as adopting a "locally constant" estimator of the drift coefficient. Banon's estimator presumed the existence of a continuous record of data. Like Aït-Sahalia (1996a), our method is modified to accommodate discrete-time data.[16] By using a locally linear parameterization we are following the procedures developed by Stone (1977), Fan (1993), Fan and Gijbels (1992), and Fan and Masry (1994) for estimating conditional expectation functions. Although our primary interest is in estimating α_0 for each x, we include the additional slope term in hopes that we will improve the

[16] In contrast to Banon (1978), Aït-Sahalia's (1996a) method uses the density estimator to obtain a nonparametric estimator of the diffusion coefficient given the drift.

performance of the estimator near the edges of the state space, or more generally in places where the data are sparse.[17]

We show in Appendix D that $\sqrt{T\delta_T^3}D(\delta_T)(a_T - \alpha)$ has a normal limiting distribution provided $T\delta_T^3$ diverges and $T\delta_T^7$ converges. The bias in the linear approximation vanishes if $T\delta_T^7$ converges to zero. As we also show in Appendix D, when the probability that $\tau_1 = 0$ is zero, shrinking δ eliminates serial correlation. This is a familiar result for local nonparametric methods applied to weakly dependent data [e.g., Robinson (1983)]. It has the implication that serial correlation can be ignored when estimating the asymptotic covariance matrix of our locally linear estimator. Although theoretically correct, the practice of ignoring serial correlation is not likely to work well for the temporal dependence present in our short-term interest rate data.[18] For this reason, we also calculate the time-series version of M-estimation standard errors as in Hansen (1982).[19] The resulting statistical inferences are based on a normal approximation with a bivariate covariance matrix $\frac{1}{T}D(\delta_T)^{-1}A_T^{-1}V_T(A_T^*)^{-1}D(\delta_T)^{-1}$, where the matrix V_T is equal to a sample spectral density matrix at frequency zero (a "long-run" sample covariance matrix) for the time series $\{Y_{j,T}\}$, where

$$Y_{j,T} = \Psi'(y_j; \delta_T)\begin{bmatrix} 1 & \frac{y_j - x}{\delta_T} \end{bmatrix} D(\delta_T)a_T + \frac{1}{2}\sigma^2(y_j)\Psi''(y_j; \delta_T).$$

5.1 Estimation results

To check the robustness of our previous results, we apply the local linear estimation method to the federal funds data. We initially report results for the case where the dispersion parameter δ equals 3 and for two benchmark specifications of the variance elasticity: $\gamma = 1, 4$. The first specification was chosen because of the popularity of the Feller square-root model of short-term interest rates. The second one is of interest because it is the largest variance elasticity that generates a reasonable looking density under the previous parameterization. For both values of the parameters we compare the local linear estimates of our measure $\mu/(2\sigma^2)$ of pull to our previous estimates.

The $\gamma = 1$ estimates and standard error bands are reported in Figure 5. The solid line depicts our local linear estimate and the dashed

---

[17] In Appendix D we show that in contrast to a locally constant parameterization, the bias depends only on μ'' and not on other features of the stationary density. This is analogous to the advantages of locally linear over locally constant parameterizations of a regression function [see Fan (1993)].

[18] More generally, Robinson (1983) argues that theoretical "results of this kind are not to be taken too seriously."

[19] While this approach is analogous to the usual corrections for serial correlation, Hansen (1982) does not provide a formal justification for local parameterizations.

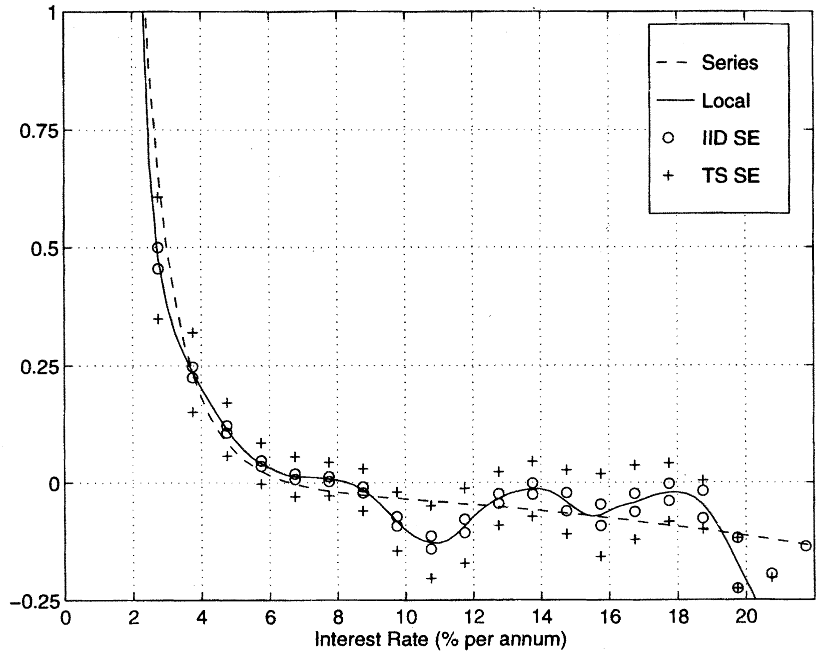

Figure 5
Local linear and series pull ($\mu/2\sigma^2$) estimates using federal funds data for local elasticity power $\gamma = 1$. Circles are two standard error bands imposing the asymptotic result that serial dependence is negligible. Pluses are two standard error bands using spectral density estimates. Spectral density estimates were calculated as weighted averages of autocovariances using a Bartlett window that included 60 time lags.

line our previous four-term series estimates. As expected, the standard errors computed ignoring serial correlation (circles) give the appearance of much more precision than the ones computed with the long-run sample covariance matrix (pluses). Our global series estimate looks like a smoothed version of the locally linear one. Moreover, the smoothed version looks plausible if we consider the standard errors that correct for serial correlation, except possibly at very high interest rates (in excess of 19%). Since there are relatively few interest rate observations of such magnitude, the locally linear estimates may well be even less precise than what is indicated by the standard errors. The series estimates exploit the global functional form, and hence extrapolate from sample information elsewhere in the interest rate distribution.

In Figures 6 and 7 we report local linear estimates for $\gamma = 4$. Recall that δ dictates the size of the "neighborhood" to which the local approximation is meant to apply. Figure 6 depicts the $\delta = 3$ results and Figure 7 the corresponding $\delta = 6$ estimates. Again we include

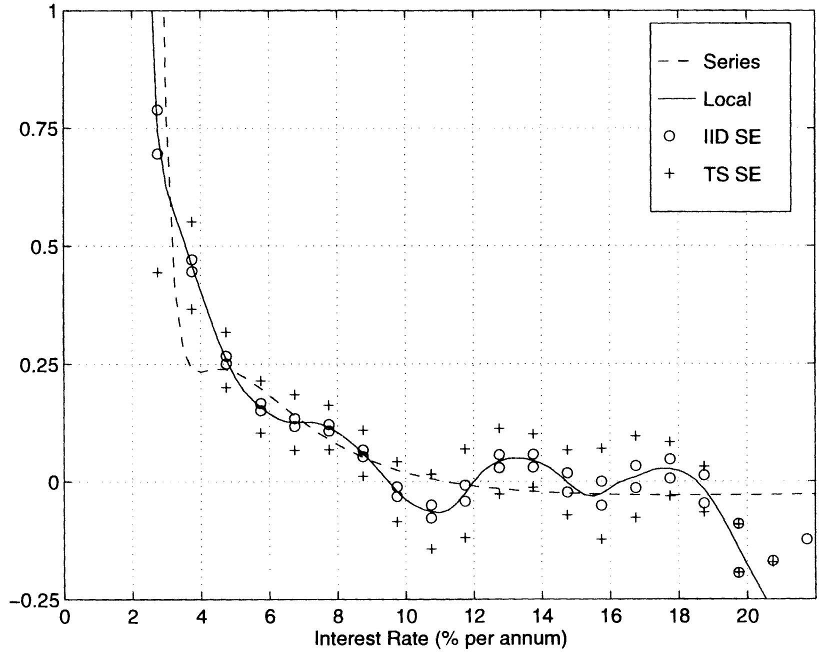

Figure 6
Local linear and series pull ($\mu/2\sigma^2$) estimates using federal funds data for local variance elasticity $\gamma = 4$. Circles are two standard error bands imposing the asymptotic result that serial dependence is negligible. Pluses are two standard error bands using spectral density estimates. Spectral density estimates were calculated as weighted averages of autocovariances using a Bartlett window that included 60 time lags.

the series estimates as a benchmark. In comparing the series and the local linear estimates for $\gamma = 4$, three differences emerge. First, the series estimate has a left (small interest rate) hump that is possibly spurious and not evident in the locally linear estimates. Second, as before, the series and locally linear estimates diverge for very large interest rates. Finally, in Figure 7 ($\delta = 6$) the locally linear estimates detect a hump in the vicinity of 15% interest rates not present in the series estimates. In comparing this figure to Figure 6 ($\delta = 3$), we see that the single hump in Figure 7 results from smoothing over two apparent humps in Figure 6. Even the single hump only appears significant when we abstract from serial correlation. Both the series and the locally linear estimates show Brownian motionlike behavior for interest rates between 10% and 18%. In summary, our four-term series parameterization of the drift does not seem overly restrictive with the proviso that the pull measure beyond interest rates of 19% is sensitive to the extrapolation technique.

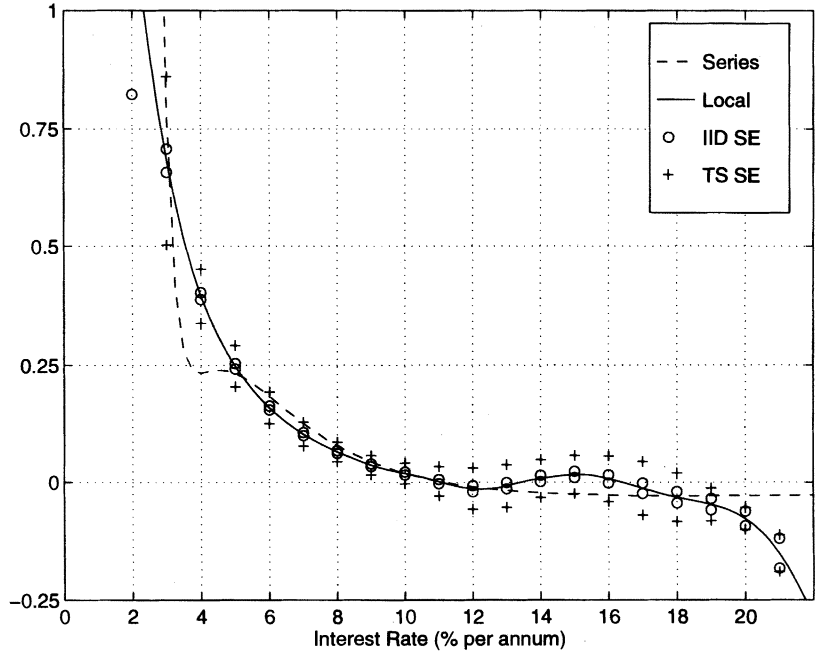

Figure 7
Local linear and series pull ($\mu/2\sigma^2$) estimates using federal funds data for local variance elasticity $\gamma = 4$. Circles are two standard error bands imposing the asymptotic result that serial dependence is negligible. Pluses are two standard error bands using spectral density estimates. Spectral density estimates were calculated as weighted averages of autocovariances using a Bartlett window that included 60 time lags.

6. Identifying the Volatility Elasticity

So far we have demonstrated how to identify and estimate the drift for a prespecified volatility elasticity using the stationary distribution of the data. We now show how to estimate the volatility elasticity using first differences of the data. A feature of constant variance elasticity models is that the variance can be arbitrarily small for small interest rates. For instance, the ratio $\mu/(2\sigma^2)$ used in our hitting time corrections gets arbitrarily large for small interest rates. When measuring volatility, drift corrections appear to be particularly important when the diffusion coefficient is not bounded away from zero.[20] We show how to identify the variance elasticity while allowing for the presence

[20] Some other econometric and statistical methods for measuring volatility, such as those suggested by Florens-Zmirou (1992) and Genot-Catalot and Jacod (1993), abstract from drift corrections when looking at the squared first differences in the time-series data and hence are not well suited for our estimation problem.

of subordination and a flexible model of the drift using test functions that depend on first differences.[21]

Consider a smooth and bounded test function ψ with the property that $\phi_v(u) \equiv \psi(u - v)$ and $\phi_u(v) \equiv \psi(u - v)$ are both in the domain of the generator, for all v and u. Stationary scalar diffusions with barriers that are not attracting are reversible. Following Hansen and Scheinkman (1995) we exploit the idea that under reversibility

$$E[\psi(x_t - x_s)] = E[\psi(x_s - x_t)].$$

In other words, the distribution of $\{x_t - x_s\}$ has to be symmetric around zero. In Appendix E we use an argument of Tom Kurtz to provide conditions under which the right derivatives with respect to t $(t > s)$ of both sides of this equation exist and can be calculated by applying $\mathcal{A}$ to $\phi_v(x_t)$ and to $\phi_u(x_t)$, replacing v and u by x_s, and taking expectations:[22]

$$\begin{aligned} &E[\mu(x_t)\psi'(x_t - x_s) + \tfrac{1}{2}\sigma^2(x_t)\psi''(x_t - x_s)] \\ &\quad = E[-\mu(x_t)\psi'(x_s - x_t) + \tfrac{1}{2}\sigma^2(x_t)\psi''(x_s - x_t)]. \end{aligned} \tag{16}$$

(See Appendix E for a more complete justification.) Since the increment process $\{z_t\}$ is independent of $\{x_t\}$ and since the moment condition in Equation (16) is applicable for all $t > s$, it also applies to the discrete-time subordinated process $\{y_j\}$:

$$\begin{aligned} &E[\mu(y_{j+1})\{\psi'(y_{j+1} - y_j) + \psi'(y_j - y_{j+1})\} \\ &+ \tfrac{1}{2}\sigma^2(y_{j+1})\{\psi''(y_{j+1} - y_j) - \psi''(y_j - y_{j+1}\}] = 0. \end{aligned} \tag{17}$$

Although subordination can alter the distribution of first differences, including fattening the tails as in Clark (1973), this distribution is still implied to be symmetric. Moreover, as reflected by Equation (17), test functions of first differences can reveal information about the volatility elasticity.

Since we have not characterized the efficient test functions of first differences, we take our test functions ψ to be cumulative normal distribution functions with different centerings and dispersions. Thus

---

[21] Foster and Nelson (1996) propose and analyze methods for estimating volatility as a function of calendar time. When the directing process $\tau_j = \int_0^j z_t\, dt$, their methods are aimed at measuring the local variance given by $\kappa z_j(y_j)^\gamma$ in our setup from nearby (in calendar time) squared first differences. Again, drift corrections would seem to be vital for small interest rates.

[22] The argument can be extended to test functions ψ: $\mathbb{R}^2 \to \mathbb{R}$ by using the fact that $E[\psi(x_t, x_s)] = E[\psi(x_s, x_t)]$ under stationarity and reversibility. Hansen and Scheinkman (1995) established the analogous result to moment condition (16) for functions: $\psi(u, v) = \phi(u)\phi^*(v)$.

Equation (17) becomes

$$E\Big[\mu(y_{j+1})\tfrac{1}{\delta}\Big\{\rho'\Big(\tfrac{y_{j+1}-y_j-\eta}{\delta}\Big)+\rho'\Big(\tfrac{y_j-y_{j+1}-\eta}{\delta}\Big)\Big\}$$
$$-\tfrac{1}{2}y_{j+1}^2\tfrac{1}{\delta^2}\Big\{\Big(\tfrac{y_{j+1}-y_j-\eta}{\delta}\Big)\rho'\Big(\tfrac{y_{j+1}-y_j-\eta}{\delta}\Big)-\Big(\tfrac{y_j-y_{j+1}-\eta}{\delta}\Big)\rho'\Big(\tfrac{y_j-y_{j+1}-\eta}{\delta}\Big)\Big\}\Big]$$
$$=0. \tag{18}$$

Notice that the moment condition looks simultaneously at first differences in the data in the vicinity of both η and $-\eta$. Thus, in selecting centerings for test functions of the first differences of the data, it suffices to concentrate on nonnegative choices of η.

6.1 Test functions

We use cumulative normal test functions for moment condition (17). We construct two such functions by imitating the strategy used for levels test functions. Specifically, we form the .5 quantile $\zeta(1)$ and the .92 quantile $\zeta(2)$ from the empirical distribution of $\{|y_{j+1} - y_j|\}$. The derivative of the first test function is centered at 0 and has a standard deviation given by $\zeta(1)$. The derivative of the second test function is then centered at $\{\zeta(2) + \zeta(1)\}/2$ with a standard deviation of $[\zeta(2) - \zeta(1)]/2$. These test function derivatives are presented for federal funds data in Figure 8. Notice that the empirical distributions of first differences have a great deal of mass near zero and relatively fat tails. Thus, the two test functions are aimed at these two types of first differences: one has most of its mass near zero and the other is more spread out, covering the larger first differences. Even with the more disperse test function, the large first-difference outliers in the federal funds data are downweighted substantially.

6.2 Two-step GMM estimation

One way to estimate the variance elasticity is to minimize a GMM criterion function based on the combined set of moment conditions constructed from level and difference test functions, whereby the variance elasticity is treated as an unknown parameter to be estimated along with the drift parameters. To facilitate the interpretation of the GMM test statistics, we use a two-step procedure in which we use our previously described estimators of the drift from Section 4 as a function of the variance elasticity γ and plug them into the moment conditions (17) formed from test functions of first differences to estimate γ. This two-step method adjusts for the initial estimation of the drift. We implement this approach using our global four-parameter model of the drift.

Let c be used for a hypothetical value of the variance elasticity and, as before, use a as a hypothetical parameter vector for the drift. Let f_j

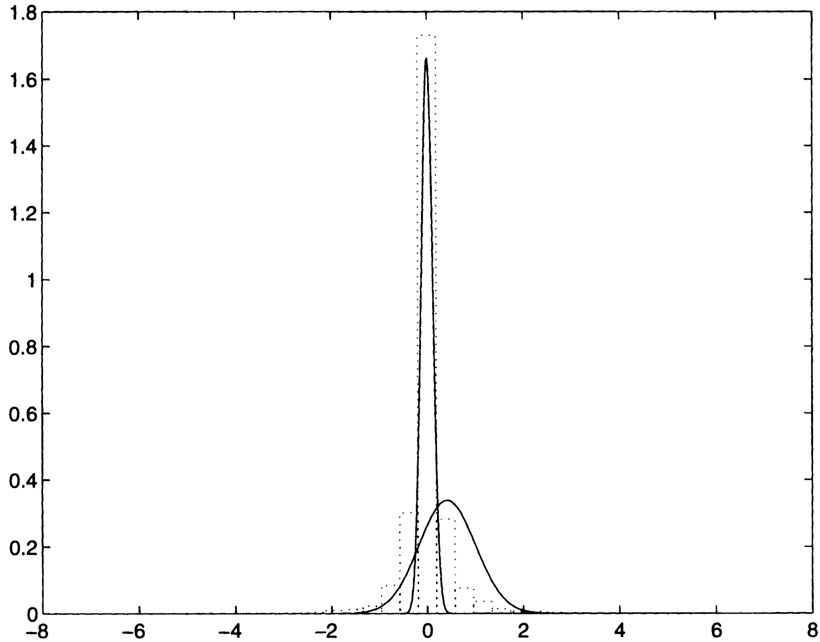

Figure 8
Derivatives of test function constructed from first differences of interest rates and background histograms of first differences. The derivatives of test functions integrate to 1/2 and the histograms integrate to 1.

denote the random function being averaged in Equation (13), viewed as a function of both a and c. In other words, these are the random functions constructed from test functions of interest rate levels. Let g_j denote the vector random function based on Equation (18) using test functions of first differences in interest rates. These moment conditions are included in hopes of extracting sample information about the variance elasticity. The random function g_j also depends on both a and c and the dependence on a is linear.

Write the first-stage estimator as

$$a_T(c) \equiv \arg\min_a \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{T} f_j(a, c) \right)^* [V_T(c)]^{-1} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{T} f_j(a, c) \right),$$

where we have now made explicit the dependence of the drift estimate on the specified variance elasticity c. Let V_0 be the asymptotic covariance matrix for $\{(1/\sqrt{T}) \sum_{j=1}^{T} f_j(\alpha, \gamma)\}$. Consider now estimating the variance elasticity parameter γ using the moment conditions built from test functions of first differences of the data. As is standard in two-step estimation problems, we have to account for the impact of the first-stage estimation on the second set of moment conditions [e.g., see Durbin (1970)]. The random functions f_j and g_j are affine in a:

$$\begin{aligned} f_j(a, c) &= \hat{f}_j(c)a + \tilde{f}_j(c) \\ g_j(a, c) &= \hat{g}_j a + \tilde{g}_j(c). \end{aligned}$$

It follows from Hansen (1982) that

$$\frac{1}{\sqrt{T}} \sum_{j=1}^{T} g_j(\hat{a}_T(\gamma), \gamma) \approx \frac{1}{\sqrt{T}} \sum_{j=1}^{T} [g_j(\alpha, \gamma) - A_0 f_j(\alpha, \gamma)],$$

where

$$A_0 \equiv E[\hat{g}_j](E[\hat{f}_j(\gamma)^*] V_0^{-1} E[\hat{f}_j(\gamma)])^{-1} E[\hat{f}_j(\gamma)^*] V_0^{-1}.$$

Notice that this approximation includes the usual correction term for first-stage estimation of α. Let W_0 be the asymptotic covariance matrix for $\{(1/\sqrt{T}) \sum_{j=1}^{T} [g_j(\alpha, \gamma) - A_0 f_j(\alpha, \gamma)]\}$ and W_T be a consistent estimator of W_0. Then the two-step GMM estimator of γ is given by

$$\hat{c}_T = \arg\min_c \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{T} g_j(a_T(c), c) \right)^* W_T^{-1} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{T} g_j(a_T(c), c) \right).$$

6.3 Updating the weighting matrix as a function of the variance parameter

The GMM estimators just described require a consistent estimator of W_0 for asymptotic efficiency. One way to accomplish this is to use an initial consistent estimator of γ, say $\tilde{c}_T$, and then to use an estimator of the spectral density at frequency zero constructed from the time series $\{g_j(a_T(\tilde{c}_T), \tilde{c}_T) - A_T f_j(a_T(\tilde{c}_T), \tilde{c}_T)\colon\ j = 1, 2, \ldots, T\}$ as in Newey and West (1987) and Andrews (1991). The estimate A_T of A_0 is given by

$$\begin{aligned} A_T = {} & \left(\frac{1}{T} \sum_{j=1}^{T} \hat{g}_j \right) \left[\left(\frac{1}{T} \sum_{j=1}^{T} \hat{f}_j(\tilde{c}_T) \right)^* V_T^{-1} \left(\frac{1}{T} \sum_{j=1}^{T} \hat{f}_j(\tilde{c}_T) \right) \right]^{-1} \\ & \times \left(\frac{1}{T} \sum_{j=1}^{T} \hat{f'}_j \right) V_T^{-1}, \end{aligned}$$

where V_T is a sample spectral density estimator from the time series $\{f_j(a_T(\tilde{c}_T), \tilde{c}_T)\colon\ j = 1, 2, \ldots, T\}$. Recall that the parameter estimate of the local mean is easy to compute for a given value of the variance parameter. For each hypothetical variance parameter c we obtain the estimator $a_T(c)$. We then form $W_T(c)$ by estimating the spectral density at frequency zero from the time series $\{g_j(a_T(c), c) - A_T(c)f_j(a_T(c), c)\colon\ j = 1, 2, \ldots, T\}$. The sample estimator A_T now depends on c because the time series $\{f_j(a_T(c), c)\colon\ j = 1, 2, \ldots, T\}$ is used to build the sample spectral density estimator V_T. This is repeated for each admissible value of c and the resulting criterion is plotted. With this procedure, we use different weighting matrices for each choice of c. The minimized value of the criterion still has an asymptotic chi-square distribution with degrees of freedom given by the number of overidentifying restrictions. Confidence sets for c can be inferred from this function by comparing the criterion function at hypothetical values of c minus the minimized value to critical values inferred from a chi-square one distribution. An alternative method is to compare the criterion directly to the chi-square two distribution without subtracting off the minimum value. Stock and Wright (1995) recommend this second approach for poorly identified models.

Our choice of updating the weighting matrix as a function of the variance parameter c is motivated in part by the Monte Carlo evidence reported in Hansen, Heaton, and Yaron (1996). In a different estimation environment, this latter article documents advantages to continuously updating the weighting matrix, especially when criterion-function-based inferences are made about the parameter of interest.

6.4 Volatility elasticity estimation

Next we consider two-step estimation results. The GMM criterion function for variance elasticity estimation using federal funds data is depicted by the solid line of Figure 9. For each value of $c = \gamma$, the limiting distribution of the criterion is chi-square with two degrees of freedom. A 90% confidence set is given by the region below a horizon line drawn at 4.6, and 95% is given by the corresponding region below 6. Thus, variance elasticities around three to four are much more plausible than small ones and very large ones. The dashed and dotted lines in Figure 9 give the GMM criterion functions using each of the two test functions separately so that we might gauge the relative contributions of each. These criterion functions will have a limiting chi-square distribution with only one degree of freedom. We see that the narrow test function adds little relative to the wide test function.

7. Concluding Remarks

In this article we used the constant volatility elasticity parameterization of the diffusion coefficient to interpret federal funds interest rate data. We allowed for the possibility that our data are generated through an unobservable sampling process that is independent of the underlying

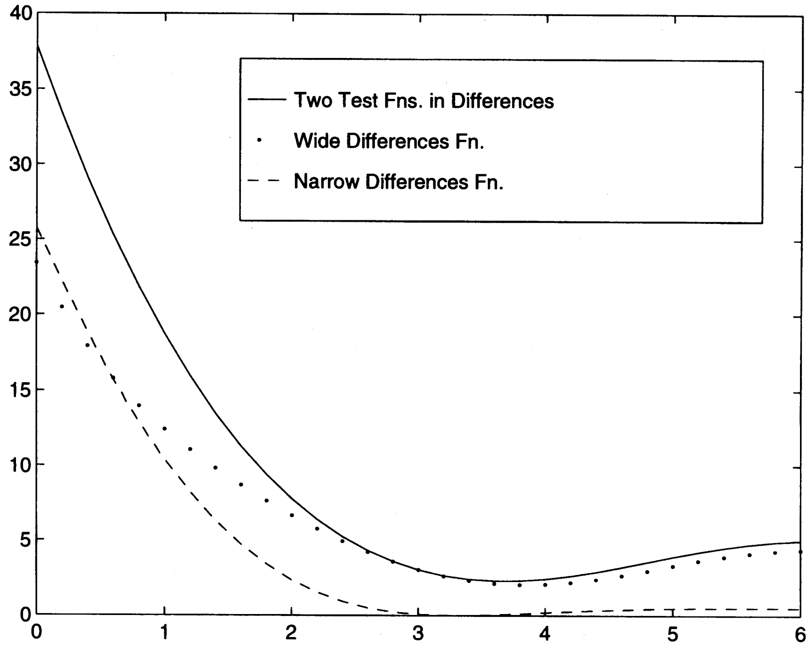

Figure 9
Federal funds GMM criterion functions for two-step estimation of the nonlinear drift specification. The curves are the GMM criteria as a function of the local variance elasticity γ after concentrating out drift estimates. Weighting matrices were calculated using Bartlett spectral density estimators that included 60 time lags.

diffusion, and we permitted the drift to be nonlinear in the state. We proposed and implemented two estimation methods for the drift that are applicable for constant variance elasticity models, and a two-step method for estimating volatility elasticities in the presence of subordination.[23]

Using federal funds data, we found that volatility elasticities between one and one-half and two (variance elasticities between three and four) give reasonable looking stationary densities and at the same

[23] We treated the stationary scalar diffusion process as the target process of the empirical exercise and make no attempt to characterize the "sampling process." Of course the "sampling" or "directing process" is of considerable interest in its own right, but our empirical investigation provides no direct information about it. To extract this information requires other methods and/or data. For examples of empirical implementations of models with time deformation that do characterize the sample process, see Tauchen and Pitts (1983), Harris (1987), Stock (1988), and Ghysels and Jasiak (1995).

time are consistent with the behavior of first differences of short-term interest rates.[24] Although we modeled the interest rate process as stationary, we found very little pull to the center of the distribution except when rates are small. Thus, for high interest rates, we find that the stationarity is primarily volatility induced.

While we find high volatility elasticity models hard to dismiss empirically, perhaps there are other perverse implications of these models that our estimation method is missing. A feature of high volatility elasticity models is that extremely large interest rate movements are likely to occur when interest rates are high. In fact, it is precisely this mechanism that induces stationarity for our fitted models of short-term interest rates. Among other things, the potential for fat-tailed distributions of interest rate differences may undermine the quality of statistical inferences based on large sample approximations. Assessing the finite-sample performance of estimation methods such as ours is an important topic for future research.[25]

Appendix A: Geometric Ergodicity

In this appendix we present sufficient conditions for geometric ergodicity. Geometric ergodicity is a convenient sufficient condition for justifying a variety of large sample approximations. It guarantees that autocorrelations of time-invariant functions of the data decay geometrically [see Rosenblatt (1969)].

A.1 Restrictions on the original diffusion

Geometric ergodicity in continuous time requires that there exists a positive λ such that

$$\mathrm{var}(E[\phi(x_t)|x_0]) \leq \exp(-2\lambda t)\mathrm{var}[\phi(x_0)]$$

for any (Borel measurable) function ϕ of the Markov state for which $\phi(x_t)$ has a finite second moment and any $t \geq 0$. We use the follow-

[24] When we applied these same estimation methods to one-month Treasury bills, we were unable to find evidence against smaller volatility elasticities including an elasticity of one-half. Thus these Treasury-bill results are not in conflict with one of the main findings of Andersen and Lund's (1997) study. They used a tightly parameterized, multivariate diffusion model of the short-term interest rate and estimated a volatility elasticity close to one-half.

[25] This topic is important for another reason. There are discrepancies between some of our results and those reported in concurrent work by Aït-Sahalia (1996b). Aït-Sahalia used alternative econometric methods that exploit similar identifying information. He argued that a more complicated specification of the diffusion coefficient is needed to fit the implied stationary density. In contrast, we claimed that our drift specification if flexible enough to adapt to a wide range of volatility elasticities and still match the stationary density. This tension in results may be due in part to data differences. However, it is an open question as to which, if either, large sample inference method is reliable in practice.

ing sufficient condition for geometric ergodicity of $\{x_t\}$ deduced by Hansen and Scheinkman (1995).

Assumption 4. *The following two boundary conditions are satisfied:*

(i) $\lim_{x\nearrow+\infty} \frac{\gamma}{2}x^{\frac{\gamma}{2}-1} - \frac{2\mu(x)}{x^{\gamma/2}}$ *exists (possibly infinite) and is not zero;*
(ii) $\lim_{x\searrow 0} \frac{\gamma}{2}x^{\frac{\gamma}{2}-1} - \frac{2\mu(x)}{x^{\gamma/2}}$ *exists (possibly infinite) and is not zero.*

If μ/σ converges to a constant for large values of the Markov state, Assumption 4 (i) is satisfied provided that the volatility elasticity γ exceeds two. In this case the integral in (7) is finite and hence stationarity is volatility induced. This shows one way to build geometrically ergodic processes for which stationarity is volatility induced.

A.2 Restrictions on the directing process

To obtain ergodicity results for the subordinated process $\{y_j\}$, we also must restrict $\{\tau_j\}$:

Assumption 5. $E[\exp(-\tau_j)] \leq b\theta^j$ *for some* $0 < \theta < 1$ *and some* $b > 0$.

One sufficient condition for this assumption is that the support of $\tau_j - \tau_{j-1}$ be bounded away from zero. An alternative sufficient condition for Assumption 5 is as follows. Extend the stationary process $\{\tau_j - \tau_{j-1}\}$ to the negative integers. Let $\mathcal{G}_0$ be the sigma algebra generated by $\tau_j - \tau_{j-1}$ for $j \leq 0$. Assume that given η there exists a sufficiently large k such that

$$|E[\exp(-(\tau_k - \tau_{k-1})) - \theta|\mathcal{G}_0]| \leq \eta \tag{A1}$$

almost surely, where $\theta < 1$ is the mean of the random variable $\exp(-(\tau_j - \tau_{j-1}))$. For instance, this restriction is satisfied when $\{\tau_j - \tau_{j-1}\}$ is uniform (ϕ) mixing and $\tau_j - \tau_{j-1}$ is different from zero with positive probability. In addition, periodic behavior can be accommodated by allowing the ϕ-mixing coefficients to converge to $1 - \frac{1}{p}$ instead of zero, where p is the periodicity of the process [see Bradley (1986)].

It follows from (A1) that we can dominate the conditional expectation $E[\exp(-(\tau_k - \tau_{k-1}))|\mathcal{G}_0]$ by $\theta + \eta$ almost surely. We choose the uniform approximation error η so that the sum $\theta + \eta < 1$. For $j \geq k\ell$

$$\begin{aligned} E[\exp(-\tau_j)|\mathcal{G}_0] &= E\left[\exp\left(-\sum_{m=1}^{j}(\tau_m - \tau_{m-1})\right)|\mathcal{G}_0\right] \\ &\leq E\left[\exp\left(-\sum_{m=1}^{\ell}(\tau_{km} - \tau_{km-1})\right)|\mathcal{G}_0\right] \leq (\theta + \eta)^{\ell}. \end{aligned}$$

A.2.1 Subordination via a diffusion. Finally, suppose that $\{\tau_j\}$ is constructed from $\{z_t\}$ via Equation (8). Suppose that $\{z_t\}$ is a stationary scalar diffusion on $[0, \infty)$, with a finite second moment and 0 not attracting. Further suppose that drift and diffusion coefficients are continuous and the drift is positive at zero. Following Runolfsson (1994), construct the semigroup

$$\mathcal{K}_t\phi(u) = E\left[\exp\left(-\int_0^t z_s\, ds\right)\phi(z_t)|z_0 = u\right],$$

although we take the domain of this semigroup to be the space of functions with a finite second moment under the stationary distribution for $\{z_t\}$. The generator for this semigroup is $\hat{\mathcal{B}}\phi(u) = \mathcal{B}\phi(u) - u\phi(u)$ where $\mathcal{B}$ is the generator for the process $\{z_t\}$. The domain of $\hat{\mathcal{B}}$ includes any function ϕ in the domain of $\mathcal{B}$ for which $E[z_t^2\phi^2(z_t)] < \infty$. Let $\phi(h) = \varsigma\exp(-\theta u)$ for $0 \leq u \leq 1$ and equal to $\varsigma\exp(-\theta\chi)$ for some real number χ for $u \geq 2$. We construct ϕ in $(1, 2)$ so that the resulting function is twice continuously differentiable. Both ς and θ are positive, and the scale factor ς is chosen so that $\phi \geq 1$. By the results in Hansen, Scheinkman, and Touzi (1996), ϕ is in the domain of $\mathcal{B}$ and hence $\hat{\mathcal{B}}$. Moreover, it can be verified that for sufficiently small θ, there exists an $\eta > 0$ such that

$$\hat{\mathcal{B}}\phi(u) \leq -\eta\phi(u).$$

As in Runolfsson (1994),

$$E\left[\exp\left(-\int_0^t z_s\, ds\right)|z_0\right] \leq E[\mathcal{K}_t\phi(z_0)] \leq \exp(-\eta t)E[\phi(z_0)] \quad \text{(A2)}$$

(see the second part of the proof of Theorem 3.1).

Periodic behavior may be accommodated by extending the state space so that the resulting process is a multivariate diffusion. We now sketch how our previous analysis may be extended to cases in which $\{z_t\}$ is the first component of a stationary, multivariate diffusion with time t state vector w_t. The semigroup is now

$$\mathcal{K}_t\phi(v) = E\left[\exp\left(-\int_0^t z_s\, ds\right)\phi(w_t)|w_0 = v\right].$$

Partition $v = [u, v_2]$, and let π be the first component of the drift and ω the one-one component of the covariance matrix. We assume that $\pi(u, \cdot)$ and $\omega(u, \cdot)$ are bounded for each $u \geq 0$ and $\pi(0, \cdot) \geq \epsilon > 0$ for some ϵ. We may now repeat the previous construction of an appropriate ϕ to establish the inequality analogous to (A2).

Appendix B: Central Limit Approximations for the Sampled Process

In this appendix we verify that under subordination, if Assumptions 3, 4, and 5 are satisfied, we can justify central limit approximations for the sampled process.

Since $\{x_t\}$ and $\{\tau_j - \tau_{j-1}\}$ are stationary processes, they can be extended backward in time. Let $\mathcal{F}_t$ be the sigma algebra generated by $\{x_s\colon\ s \leq t\}$, $\mathcal{G}_0$ be the sigma algebra generated by $\{\tau_j - \tau_{j-1}\colon\ j \leq 0\}$, $\mathcal{G}_\infty$ be the sigma algebra generated by the entire increment process $\{\tau_j - \tau_{j-1}\}$ and $\mathcal{H}_j = \mathcal{F}_{\tau_j} \vee \mathcal{G}_\infty$. Construct a subspace $\mathcal{Z}$ of the collection of Borel measurable functions mapping $\phi\colon\ \mathbb{R} \to \mathbb{R}$:

$$\mathcal{Z} \equiv \{\phi\colon\ E[\phi(x_t)^2] < \infty, E[\phi(x_t)] = 0\}.$$

We let $\|h\|$ denote the norm $(E[h^2])^{1/2}$.

We first establish that

$$\sum_{j=0}^{\infty} \|E[\phi(y_j)|\mathcal{H}_0]\| \leq \hat{b}\|\phi(y_0)\|, \tag{B1}$$

where $\hat{b}$ is independent of $\phi \in \mathcal{Z}$. The existence of a common $\hat{b}$ will not be needed for the central limit approximations established in this appendix or be referred to in Section 4. It will, however, prove useful in our large sample justification for the locally linear estimator of the drift. Notice that under Assumptions 3 and 4, it follows that there exists a $\lambda > 0$ such that

$$\begin{aligned} E[E[\phi(y_j)|\mathcal{H}_0]^2|\mathcal{G}_\infty] &\leq \exp(-2\lambda\tau_j)E[\phi(x_0)^2|\mathcal{G}_\infty] \\ &\leq \exp(-2\lambda\tau_j)\|\phi(y_0)\|^2 \end{aligned}$$

and hence that

$$E[E[\phi(y_j)|\mathcal{H}_0]^2] \leq E[\exp(-2\lambda\tau_j)]\|\phi(y_0)\|^2.$$

Therefore,

$$\sum_{j=0}^{\infty} \|E[\phi(y_j)|\mathcal{H}_0]\| \leq \sum_{j=0}^{\infty}\{E[\exp(-2\lambda\tau_j)]\}^{1/2}\|\phi(y_0)\|.$$

By Jensen's inequality we have

$$\{E[\exp(-2\lambda\tau_j)]\}^{1/2} \leq \{E[\exp(-\tau_j)]\}^{\min\{\lambda,1/2\}}.$$

Hence, by Assumption 5, relation (B1) is satisfied for some $\hat{b}$.

B.1 Central limit approximation

We may use this relation to obtain a central limit theorem via a martingale approximation for the sampled process. As in Hall and Heyde (1980, chap. 5), the basic idea is to use the decomposition

$$\phi(y_j) = \sum_{m=j}^{\infty} \{E[\phi(y_m)|\mathcal{H}_j] - E[\phi(y_m)|\mathcal{H}_{j-1}]\} - e_{j+1} + e_j, \qquad \text{(B2)}$$

where

$$e_{j+1} = \sum_{m=j+1}^{\infty} \{E[\phi(y_m)|\mathcal{H}_j]\}.$$

The terms in the infinite series in Equation (B2) are each martingale differences, and so we have a martingale difference approximation of the original process if (B1) holds. In that case, partial sums of the martingale difference scaled by the square root of the sample size, and hence $\{(1/\sqrt{T}) \sum_{j=1}^{T} [\phi(y_j)]\}$, will have a limiting normal distribution with mean zero.

Appendix C: Efficient Estimation

In this appendix we describe the optimal choice of test functions for estimation of a parametric model of the infinitesimal generator. The efficiency problem studied is a constrained one in that the only information used is that contained in the empirical distribution of the data. For simplicity, we begin with the case in which there is a continuous record of data. We then show how to extend this result to accommodate discrete sampling. For square-integrable functions ϕ and ψ write $E[\phi(x_t)\psi(x_t)] = \langle \psi | \phi \rangle$.

C.1 Continuous record

Let $\mathcal{A}_a$ denote the parameterized family of generators, with $\mathcal{A}_0$ being the true generator. From Bhattacharya (1982) and Hansen and Scheinkman (1995), it is known that

$$\frac{1}{\sqrt{T}} \int_0^T \mathcal{A}_0 \phi(x_t)\, dt \Rightarrow \mathcal{N}(0, -2\langle \mathcal{A}_0 \phi | \phi \rangle).$$

Let q_a denote the implied parameterization of the stationary density, and let

$$d_0 = \frac{\partial q_0 / \partial a}{q_0}$$

denote the score for this process.

Assumption 6. $d_0 \in \mathcal{L}^2(Q_0)$.

The domain $\mathcal{D}_a$ of each generator $\mathcal{A}_a$ varies with a. However, for parameterizations such as the one considered in this article, where for each a, the corresponding stochastic process is a time-reversible diffusion with a continuous drift and a nonzero continuous diffusion coefficient, the domains have a "large" common intersection. In fact, if C_K^2 denotes the set of all twice continuously differentiable functions with a compact support in $(0, \infty)$, and $\mathcal{L}_a$ is the operator in C_K^2 given by

$$\mathcal{L}_a\phi = \mu_a\phi' + \frac{1}{2}\sigma^2\phi''$$

then $\mathcal{A}_a$ is the closure of $\mathcal{L}_a$ [e.g., see Hansen, Scheinkman, and Touzi (1996)]. For each $\phi \in C_K^2$:

$$\int (\mathcal{A}_a\phi)(x) q_a(x)\, dx = 0. \tag{C1}$$

Imagine using the sample counterpart of this moment condition to estimate the parameter value. Thus form a_T by solving

$$\int_0^T \mathcal{A}_a\phi(x_t)\, dt = 0$$

for a. We index the resulting estimator by the choice of test function used in constructing it. We now study the statistical efficiency of the resulting class of test function estimators.

First we deduce the limiting distribution for each test function estimator. Differentiating Equation (C1) with respect to a, using the fact that the derivative of the integrand with respect to a is a continuously differentiable function with compact support, and that σ is independent of a, yields

$$\int \left[\frac{\partial}{\partial a}(\mathcal{A}_0\phi)(x) + \mathcal{A}_0\phi(x)\, d_0(x)\right] q_0(x)\, dx = 0,$$

or

$$\int \frac{\partial \mu_a}{\partial a}(x)\phi'(x) q_0(x)\, dx = -\int \mathcal{A}_0\phi(x)\, d_0(x) q_0(x)\, dx.$$

From Hansen (1982) the asymptotic variance of the resulting parameter estimator a_T is

$$\sqrt{T} a_T \Rightarrow \mathcal{N}\left(0, \frac{-2\langle \mathcal{A}_0\phi | \phi\rangle}{\langle \mathcal{A}_0\phi | d_0\rangle^2}\right).$$

We define an efficiency bound for a test function estimator via

$$I \equiv \inf_{\phi \in \mathcal{D}_0} \frac{-2\langle \mathcal{A}_0\phi | \phi \rangle}{\langle \mathcal{A}_0\phi | d_0 \rangle^2} = \inf_{\phi \in C_K^2} \frac{-2\langle \mathcal{A}_0\phi | \phi \rangle}{\langle \mathcal{A}_0\phi | d_0 \rangle^2}.$$

***Assumption** 7.* $d_0 \in \mathcal{D}_0$.

C.2 Comparison with QMLE

When Assumption 7 is satisfied,

$$I = \frac{-2}{\langle \mathcal{A}_0 d_0 | d_0 \rangle}.$$

When the data sample is i.i.d., maximum likelihood is asymptotically efficient, and hence the corresponding bound is

$$I = \frac{1}{\langle d_0 | d_0 \rangle}.$$

The temporal dependence in the data alters the efficiency bound by introducing the generator into the computation. Application of (quasi) maximum likelihood to a continuous record sample from a diffusion has an efficiency given by

$$\frac{-2\langle d_0 | \mathcal{A}_0^{-1} d_0 \rangle}{\langle d_0 | d_0 \rangle^2} \geq \frac{-2}{\langle \mathcal{A}_0 d_0 | d_0 \rangle}.$$

In other words quasi-maximum likelihood is not asymptotically efficient among the class of test function estimators. The efficient moment condition to use in estimation is

$$E[\mathcal{A}_0 d_0] = 0,$$

whereas in the i.i.d. context it is

$$E[d_0] = 0$$

corresponding to the first-order conditions from maximum likelihood.

C.3 Parameter-dependent test functions

Although our efficiency calculation did not include the possibility that test functions can depend on the unknown parameters, the efficiency bound is unaltered if this dependence is introduced. To attain the efficiency bound, it is most convenient to view d_a as a parameter-dependent test function. Hence we construct the following function of the data:

$$f(y; a) = \mu_a(y) d_a'(y) + \frac{1}{2}\sigma^2(y) d_a''(y)$$

for hypothetical parameter value a. For multidimensional parameterizations, we build a vector of such functions, one for each score variable.

For the parameterization used in this article a typical score has the form

$$d_a(y) = \frac{2y^{j-\gamma+1}}{j-\gamma+1} - \int \frac{2x^{j-\gamma+1}}{j-\gamma+1} q_a(x)\, dx,$$

and hence

$$f(y; a) = 2\mu_a(y) y^{j-\gamma} + (j-\gamma) y^{j-1}.$$

It is not necessary to evaluate the integral $\int \frac{2x^{j-\gamma+1}}{j-\gamma+1} q_a(x)\, dx$ because any generator maps a constant function into zero.

For the parameterization used in this article, we have an explicit formula for the stationary density up to scale. More generally, we may allow σ to vary with a and we know that

$$\frac{\partial \log q_a(y)}{\partial y} = \frac{2\mu_a(y) - (\sigma_a^2)'(y)}{\sigma_a^2(y)},$$

As a consequence, we can evaluate

$$d_a'(y) = \frac{\partial}{\partial a}\left[\frac{2\mu_a(y) - (\sigma_a^2)'(y)}{\sigma_a^2(y)}\right]$$

and similarly for d_a''.

C.4 Discrete sampling

Consider next the modifications to accommodate discrete sampling at a fixed interval, say h. It follows from Hansen and Scheinkman (1995) that the bound is given by

$$\bar{I} = \inf_{\phi \in D} \frac{\langle \mathcal{A}_0\phi | (I + T_h)(I - T_h)^{-1} \mathcal{A}_0\phi \rangle}{\langle \mathcal{A}_0\phi | d_0 \rangle^2}$$

where T_h is the one-period conditional expectation operator. When an efficient test function exists, it must satisfy

$$(I + T_h)(I - T_h)^{-1} \mathcal{A}_0 \phi_h \propto d_0,$$

where ϕ_h is an efficient test function and where any nonzero proportionality constant works. Hence an efficient test function is

$$\frac{1}{h}\phi_h = \mathcal{A}_0^{-1}(I + T_h)^{-1}(I - T_h)d_0 = (I + T_h)^{-1}(I - T_h)\mathcal{A}_0^{-1}d_0. \quad \text{(C2)}$$

Implementation of this formula in practice is problematic, which is why we choose to use the continuous record calculation in practice. To see the relation between the two formulas for efficient test functions, take limits of Equation (C2) as the sample interval h declines to zero:

$$\lim_{h \searrow 0} \frac{1}{h}\phi_h = -2d_0.$$

Appendix D: Large Sample Behavior of the Local Linear Estimator

In this appendix, we study the large sample approximation for the local linear estimator of the drift. For our analysis of the local linear estimator of the drift, we will require a strengthened version of Assumption 4:

Assumption 8. *The following two boundary conditions are satisfied:*

(i) $\lim_{x \nearrow +\infty} \frac{\gamma}{2}x^{\frac{\gamma}{2}-1} - \frac{2\mu(x)}{x^{\gamma/2}} = +\infty$
(ii) $\lim_{y \searrow 0} \frac{\gamma}{2}x^{\frac{\gamma}{2}-1} - \frac{2\mu(x)}{x^{\gamma/2}} = -\infty$

We will also impose:

Assumption 9. $\Pr\{\tau_1 = 0\} = 0$.

In other words, with probability one, "economic time" moves during a unit of calendar time. To simplify the notation of Section 6, write $\psi_T'(y) = \psi_T'(y; \delta_T)$ and $D_T = D(\delta_T)$. Then

$$\begin{aligned} D_T(a_T - \alpha) &= -A_T^{-1}(U_T + A_T D_T \alpha) \\ &= -A_T^{-1}\left(\frac{1}{T}\sum_{j=1}^{T} \mathcal{A}\Psi_T(y_j)\right) + A_T^{-1}B_T, \end{aligned} \quad \text{(D1)}$$

where

$$B_T = \frac{1}{T}\sum_{j=1}^{T} \frac{1}{\sigma^2(y_j)} \begin{bmatrix} 1 \\ \frac{y_j - x}{\delta_T} \end{bmatrix} \left(\mu(y_j) - \begin{bmatrix} 1 & y_j - x \end{bmatrix} \alpha\right) \frac{1}{\delta_T} K\left(\frac{y_j - x}{\delta_T}\right).$$

Rescale Equation (D1) as

$$\sqrt{\delta_T^3 T} D_T(a_T - \alpha) = -A_T^{-1}\left[\sqrt{\frac{\delta_T}{T}}\sum_{j=1}^{T}\delta_T A\Psi_T(y_j) + \sqrt{\delta_T^3 T} B_T\right].$$

D.1 A uniform variance bound

To justify a central limit approximation, we will need the following inequality relating the variance of the partial sums to the variance of the component in the presence of serial correlation:

$$\left\|\frac{1}{\sqrt{T}}\sum_{j=1}^{T}[\phi(y_j)]\right\| \leq 4\hat{b}\|\phi(y_0)\| \tag{D2}$$

for $\phi \in \mathcal{Z} \equiv \{\phi\colon E[\phi(x_t)^2] < \infty, E\phi[(x_t)] = 0\}$. Inequality (D2) follows from combining (B2) with (B1) and using the fact that in computing the norm of a sum of stationary martingale differences we may ignore cross-product terms.

D.2 Convergence of A_T

We will establish that

$$A_T \to A_\infty \equiv \frac{q(x)}{\sigma^2(x)}\begin{bmatrix}1 & 0\\ 0 & \int u^2 K(u)\,du\end{bmatrix} \tag{D3}$$

in mean-square. First note that

$$A_T = \frac{1}{T}\sum_{j=1}^{T}\Upsilon_T(y_j)$$

for an appropriate Υ_T, each entry of which has a finite second moment under the stationary distribution for T sufficiently large. It is straightforward to show that $E[\Upsilon_T(y_j)]$ converges to the right-hand side of Equation (D3). Next, a simple calculation shows that

$$\delta_T E[\text{trace}\{\Upsilon_T^2(y_j)\}] \text{ has a finite limit.}$$

Mean-square convergence follows from (D2), provided $T\delta_T$ diverges to infinity.

D.3 Bias

Note that the bias term B_T satisfies

$$|B_T| \leq \delta_T^2\left(\frac{b}{2T}\right)\sum_{j=1}^{T}\frac{1}{\sigma^2(y_j)}\left[\frac{1}{\left|\frac{y_j - x}{\delta_T}\right|}\right]\left(\frac{y_j - x}{\delta_T}\right)^2\frac{1}{\delta_T}K\left(\frac{y_j - x}{\delta_T}\right),$$

where b is a bound on the second derivative of μ. Hence

$$\left|\sqrt{\delta_T^3 T} B_T\right| \leq \sqrt{\delta_T^7 T} \frac{1}{T} \sum_{j=1}^{T} \Upsilon_T(y_j)$$

for an appropriate Υ_T, each entry of which has a finite second moment under the stationary distribution for T sufficiently large. First, it is straightforward to show that $E[\Upsilon_T(y_j)]$ converges to

$$\lim_{T\to\infty} E[\Upsilon_T(y_j)] = \frac{bq(x)}{2\sigma^2(x)} \int \begin{bmatrix} u^2 \\ |u|^3 \end{bmatrix} K(u)\, du.$$

Next, a simple calculation shows that $\delta_T E[|\Upsilon_T(y_j)|^2]$ has a finite limit. Therefore,

$$\frac{1}{T} \sum_{j=1}^{T} \Upsilon_T(y_j) \to \frac{bq(x)}{2\sigma^2(x)} \int \begin{bmatrix} u^2 \\ |u|^3 \end{bmatrix} k(u)\, du$$

in mean square, provided that $T\delta_T$ diverges to infinity. Therefore, the bias will vanish as long as

$$T(\delta_T)^7 \to 0.$$

D.4 Central limit approximation

The central limit approximation is obtained by studying the limiting behavior of

$$\left\{\sqrt{\frac{\delta_T}{T}} \sum_{j=1}^{T} \delta_T \mathcal{A} \Psi_T(y_j)\right\}.$$

We know the population mean is zero. Define $\Phi_T(y) = \sqrt{\delta_T^3} \mathcal{A} \Psi_T(y)$, and write

$$\frac{1}{\sqrt{T}} \sum_{j=1}^{T} \Phi_T(y_j) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} m_{T,j} + \frac{1}{\sqrt{T}} \sum_{j=1}^{T} E[\Phi_T(y_j) | \mathcal{H}_{j-1}], \qquad \text{(D4)}$$

where $m_{T,j} = \Phi_T(y_j) - E[\Phi_T(y_j) | \mathcal{H}_{j-1}]$. Our aim is to apply a triangular array central limit theorem for martingale differences to establish the weak convergence of the first term on the right-hand side of Equation (D4).

D.4.1 Mean-square approximation of the martingale error. We show that

$$\frac{1}{\sqrt{T}} \sum_{j=1}^{T} E[\Phi_T(y_j)|\mathcal{H}_{j-1}] \to 0 \tag{D5}$$

in mean-square. It is easy to verify that the sequences of coordinate functions of $\{\Phi_T\}$ converge weakly to zero in $\mathcal{L}^2(Q)$. Further, Assumption 8 implies that the generator $\mathcal{A}$ has a discrete spectrum [e.g., see Hansen, Scheinkman, and Touzi (1996, Theorem 5.1)]. Hence the one-period conditional expectation operator has spectral representation

$$E[\phi(y_j)|\mathcal{H}_{j-1}] = \sum_{i=0}^{\infty} \exp(-\lambda_i(\tau_j - \tau_{j-1}))E[\phi(y_j)\psi_i(y_j)]\psi_i(y_{j-1}),$$

where the orthonormal sequence $\{\psi_i\}$ consists of the eigenfunctions of the generator and $\{-\lambda_i\}$ is the (unbounded) sequence of eigenvalues of $\mathcal{A}$ ordered by their magnitude. Let ϕ_T be one of the coordinate functions of Φ_T. Then

$$\lim_{T\to\infty} E[\phi_T^2(y_j)] \text{ is finite}$$

$$\lim_{T\to\infty} E[\phi_T(y_j)\psi_i(y_j)] = 0 \text{ for each } i.$$

Given $\varepsilon > 0$, choose an N such that $E[\exp(-2\lambda_N\tau_1)] < \varepsilon$. Such an N must exist, since $\lambda_N \to \infty$ and $\Pr\{\tau_1 = 0\}$ (Assumption 9). Then

$$\begin{aligned} &\|E[\phi_T(y_j)|\mathcal{H}_{j-1}]\|^2 \\ &\quad \le \sum_{i=0}^{N-1} E[\exp(-2\lambda_i(\tau_j - \tau_{j-1}))]\{E[\psi_i(y_j)\phi_T(y_j)]\}^2 + \varepsilon E[\phi_T^2(y_j)]. \end{aligned}$$

Hence

$$\|E[\Phi_T(y_j)|\mathcal{H}_{j-1}]\| \to 0.$$

Limit (D5) then follows from (D2).

D.4.2 Asymptotic approximation of the sample second moment. We show that

$$\frac{1}{T}\sum_{j=1}^{T} \Phi_T(y_j)\Phi_T(y_j)^*$$

converges in mean-square. To construct a candidate limit point, we compute $E[\Phi_T(y_j)\Phi_T(y_j)^*]$ and take limits. To compute the expectations we decompose the random vector $\Phi_T(y)$ into three pieces. The first is the contribution of the drift and the logarithmic derivative of the variance. The second one entails differentiating the constant and linear (multiplicative) components of the test functions. The third one entails differentiating the kernel itself. Only the second two terms contribute to the central limit approximation because

$$\int \left(\frac{\sqrt{\delta_T}}{\sigma^2(y)} \begin{bmatrix} 1 \\ \frac{y-x}{\delta_T} \end{bmatrix} \left[\mu(y) - \frac{1}{2}\sigma^{2\prime}(y) \right] K\left(\frac{y-x}{\delta_T}\right) \right)^2 q(y)\, dy \to 0,$$

whereas

$$\frac{\sqrt{\delta_T}}{2\delta_T} \begin{bmatrix} 0 \\ 1 \end{bmatrix} K\left(\frac{y_j - x}{\delta_T}\right)$$

and

$$\frac{\sqrt{\delta_T}}{2\delta_T} \begin{bmatrix} 1 \\ \frac{y-x}{\delta_T} \end{bmatrix} K'\left(\frac{y_j - x}{\delta_T}\right)$$

have nondegenerate second moments. Therefore

$$\lim_{T\to\infty} E[\Phi_T \Phi_T^*] = \Lambda,$$

where

$$\Lambda = \frac{1}{4} q(x) \begin{bmatrix} \int K'(u)^2\, du & \int [uK'(u)^2 + K'(u)K(u)]\, du \\ \int [uK'(u)^2 + K'(u)K(u)]\, du & \int [K(u) + uK'(u)]^2\, du \end{bmatrix}.$$

By imitating our previous weak law arguments with $\Psi_T = \Phi_T \Phi_T^*$, it follows that

$$\frac{1}{T} \sum_{j=1}^{T} \Phi_T(y_j) \Phi_T(y_j)^* \to \Lambda$$

in mean-square.

D.4.3 Applying a triangular array central limit theorem. Combining our two previous intermediate results, it follows that

$$E\left[\frac{1}{\sqrt{T}} \sum_{j=1}^{T} m_{T,j} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{T} m_{T,j} \right)^* \right] = E[m_{T,j}(m_{T,j})^*] \to \Lambda.$$

Further from Cauchy-Schwartz and stationarity it follows that

$$E\left[\frac{1}{T} \sum_{j=1}^{T} |\Phi_T(y_j)|\, |E[\Phi_T(y_j)|\mathcal{H}_{j-1}]| \right] \leq \|\Phi_T(y_j)\|\, \|E[\Phi_T(y_j)|\mathcal{H}_{j-1}]\|.$$

Hence

$$\frac{1}{T}\sum_{j=1}^{T} m_{T,j}(m_{T,j})^* \to \Lambda$$

in probability. Finally, we show that

$$\begin{aligned} \max_{j=1,2,\dots,T} \sqrt{\tfrac{1}{T}}\, m_{T,j} \quad &\to 0 \text{ in probability} \\ E\left(\max_{j=1,2,\dots,T} \tfrac{1}{T}|m_{T,j}|^2\right) &\text{ is bounded in } T. \end{aligned} \tag{D6}$$

Since the kernel K has compact support, the random vector $\sqrt{\frac{1}{T}}\Phi_T(y_j)$ has a uniform bound in j for T, and this bound converges to zero provided $T\delta_T^3 \to +\infty$. The same is true of $\sqrt{\frac{1}{T}}E[\Phi_T(y_j)|\mathcal{H}_{j-1}]$. Therefore, the relations of Equation (D6) hold. We can now apply the triangular array central limit theorem in Hall and Heyde (1980, p. 58, Theorem 3.2).

D.5 Limit distribution

Combining our limit results, we have shown that

$$\sqrt{\delta_T^3 T}D_T(a_T - \alpha) \Rightarrow \mathcal{N}(0, A_\infty^{-1}\Lambda(A_\infty)^{-1}).$$

In the study of optimal bandwidth choice, it is desirable to have an explicit formula for the bias. Optimality requires that $T\delta_T^7$ converges to a constant h, and then the limiting bias shows up as the mean in the limiting normal distribution. Our previous bias calculation in Section D.3 can be refined to show that the limiting bias of the estimate of $\mu(x)$ is $h\frac{\mu''(x)}{2}\int u^2 K(u)\,du$. Hence the bias depends only on the second derivative of the drift and not on other features of the stationary distribution as it would in the case of a locally constant model of the drift.

Appendix E: Derivation of Moment Conditions for Nonseparable Functions

In this appendix we exposit an argument due to Tom Kurtz that allows us to generalize the moment conditions in Hansen and Scheinkman (1995) to a class of nonseparable functions. Let $\{x_t\}$ be a strictly stationary, continuous-time, n-dimensional, vector Markov process defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P}r)$, and $\mathcal{Q}$ be the probability measure induced on $\mathbb{R}^n$ by x_t (for any t), $\mathcal{L}^2(\mathcal{Q})$ be the space of all Borel measurable functions ϕ: $\mathbb{R}^n \to \mathbb{R}$ such that

$$\int_{\mathbb{R}^n} \phi^2 \, dQ < \infty.$$

Let $\mathcal{A}$ be the $\mathcal{L}^2(Q)$ infinitesimal generator associated with $\{x_t\}$ and $\mathcal{A}^*$ the adjoint of $\mathcal{A}$, that is, the generator associated with the reverse time process $x_t^* = x_{-t}$. Suppose ψ: $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is such that

(i) for each $v \in \mathbb{R}^n$, $\psi(\cdot, v) \in \mathcal{D}(\mathcal{A})$ is bounded and continuous,
(ii) for each $u \in \mathbb{R}^n$, $\psi(u, \cdot) \in \mathcal{D}(\mathcal{A}^*)$ is bounded and continuous,
(iii) $\mathcal{A}\psi(\cdot, v)$ is bounded and continuous,
(iv) $\mathcal{A}^*\psi(u, \cdot)$ is bounded and continuous.

Let $t > s$ and $\varepsilon > 0$. Since for each $v \in \mathbb{R}^n$, $\psi(\cdot, v) \in \mathcal{D}(\mathcal{A})$ we have that

$$E[\psi(x_{t+\varepsilon}, v)|\mathcal{F}_t] = \psi(x_t, v) + E\left[\int_t^{t+\varepsilon} \mathcal{A}\psi(x_{t'}, v)\, dt' \middle| \mathcal{F}_t\right].$$

Since we assumed that both $\psi(\cdot, v)$ and, $\mathcal{A}\psi(\cdot, v)$ are bounded and continuous, and x_s is $\mathcal{F}_t$-measurable:

$$E[\psi(x_{t+\varepsilon}, x_s)|\mathcal{F}_t] = \psi(x_t, x_s) + E\left[\int_t^{t+\varepsilon} \mathcal{A}\psi(x_{t'}, x_s)\, dt' \middle| \mathcal{F}_t\right]. \tag{E1}$$

Similarly

$$E[\psi(x_s^*, x_{t+\varepsilon}^*)|\mathcal{F}_t^*] = \psi(x_s^*, x_t^*) + E\left[\int_t^{t+\varepsilon} \mathcal{A}^*\psi(x_s^*, x_{t'}^*)\, dt' \middle| \mathcal{F}_t^*\right]. \tag{E2}$$

The law of iterated expectations and stationarity implies that the expected value of the left-hand sides of Equations (E1) and (E2) are identical and hence

$$E\left[\int_t^{t+\varepsilon} \mathcal{A}\psi(x_{t'}, x_s)\, dt'\right] = E\left[\int_t^{t+\varepsilon} \mathcal{A}^*\psi(x_s^*, x_{t'}^*)\, dt'\right].$$

Dividing by ε, and letting $\varepsilon \to 0$ we obtain:

$$E[\mathcal{A}\psi(x_t, x_s)] = E[\mathcal{A}^*\psi(x_s^*, x_t^*)]. \tag{E3}$$

For the case of interest in this article, in which x_s is a nonnegative Markov process that solves Equation (1) in the introduction, with inaccessible boundaries, we know that conditions (i) to (iv) above hold if ψ is a twice continuously differentiable, bounded function satisfying:

(a) for each $v \in \mathbb{R}_+$, $\mu(u)\psi_1(u, v) + (1/2)\sigma^2(u)\psi_{11}(u, v)$ is continuous and bounded,

(b) for each $v \in \mathbb{R}_+$, $\sigma(u)\psi_1(u, v) \in \mathcal{L}^2(\mathcal{Q})$,
(c) for each $u \in \mathbb{R}_+$, $\mu(v)\psi_2(u, v)+(1/2)\sigma^2(v)\psi_{22}(u, v)$ is continuous and bounded,
(d) for each $u \in \mathbb{R}_+$, $\sigma(v)\psi_2(u, v) \in \mathcal{L}^2(\mathcal{Q})$,
(e) for each $v \in \mathbb{R}_+$, $\lim_{u\to 0} \dfrac{\psi_1(u, v)}{s(u)} = \lim_{u\to\infty} \dfrac{\psi_1(u, v)}{s(u)} = 0$,
(f) for each $u \in \mathbb{R}_+$, $\lim_{v\to 0} \dfrac{\psi_2(u, v)}{s(v)} = \lim_{v\to\infty} \dfrac{\psi_2(u, v)}{s(v)} = 0$.

Further, in this case, for each $v \in \mathbb{R}_+$

$$\mathcal{A}\psi(u, v) = \mu(u)\psi_1(u, v) + \frac{1}{2}\sigma^2(u)\psi_{11}(u, v)$$

and, for each $u \in \mathbb{R}_+$,

$$\mathcal{A}^*\psi(u, v) = \mu(v)\psi_2(u, v) + \frac{1}{2}\sigma^2(v)\psi_{22}(u, v).$$

Substituting these expressions in Equation (E3) and using the fact that stationarity and reversibility imply that $E[\mathcal{A}^*\psi(x_s^*, x_t^*)] = E[\mathcal{A}^*\psi(x_s, x_t)]$ we obtain Equation (16) in Section 6.

Notice that for functions $\psi(u, v)$ that can be expressed as $\phi(u)\phi^*(v)$, we may replace the boundedness of ψ by $\psi(\cdot, v) \in \mathcal{L}^2(\mathcal{Q})$, for each fixed v and $\psi(u, \cdot) \in \mathcal{L}^2(\mathcal{Q})$, for each fixed u. Further, in Assumptions (b) and (c) above we may also substitute square integrability with respect to $\mathcal{Q}$ for boundedness and continuity. This follows from the fact that, in the separable case, Equation (E3) holds for each $\psi(u, v) = \phi(u)\phi^*(v)$, with $\phi \in \mathcal{D}(\mathcal{A})$ and $\phi^* \in \mathcal{D}(\mathcal{A}^*)$ [see Hansen and Scheinkman (1995)].

The test functions we used in our applications (see Section 6) are infinitely smooth bounded functions with derivatives with exponentially thin tails at ∞. Furthermore, the restrictions introduced in Section 4 imply that $\lim_{u\to 0} s(u) = \infty$, and that $\lim_{u\to\infty} s(u) \neq 0$ (possibly $= \infty$) unless $\gamma = \ell + 1$ and $0 < \alpha_\ell < 1/2$. In this case $s(u) \sim u^{-2\alpha_\ell}$ for large u. Using these facts, one can verify that under the parameter restrictions specified in Section 4, conditions (a) to (f) are satisfied.

Appendix F: Central Limit Approximations for Nonseparable Functions

We may extend the approach taken in Appendix B to functions of two arguments (corresponding to two different time periods). We alter our definition of $\mathcal{Z}$:

$$\mathcal{Z} \equiv \{\psi: \text{ for all } s, E[\psi(x_{t+s}, x_t)] = 0 \text{ and } E[\psi(x_{t+s}, x_t)^2] \leq b \text{ for some } b \text{ independent of } s\}.$$

Application of the generator to the cumulative normal test functions we use in Section 6 results in functions in $\mathcal{Z}$. We now use the decomposition

$$\psi(y_{j+1}, y_j) = \sum_{m=j}^{\infty}\{E[\psi(y_{m+1}, y_m)|\mathcal{H}_{j+1}] - E[\psi(y_{m+1}, y_m)|\mathcal{H}_j]\} - e_{j+1} + e_j$$

where

$$e_{j+1} = \sum_{m=j+1}^{\infty} \{E[(y_{m+1}, y_m)|\mathcal{H}_{j+1}]\}$$

to obtain a martingale difference approximator and hence a central limit result. The formal justification of the martingale difference approximation is essentially the same as in Appendix B.

Appendix G: Estimates of Nonlinear Drift Parameters

Table 2 below presents estimates of α as defined in Equation (12), and as used for the nonlinear drift specifications in Figures 1–7 and Figure 9. The estimator is defined in Equation (14).

Table 2
Input into the construction of the figures

γ	$7^{-\gamma} \times \alpha_{-1}$	$10 \times 7^{-\gamma} \times \alpha_0$	$10^2 \times 7^{-\gamma} \times \alpha_1$	$10^3 \times 7^{-\gamma} \times \alpha_2$
0	8.2349	−26.127	24.150	−7.4531
	(1.3657)	(6.3519)	(8.0167)	(2.8276)
1	3.4534	−9.7524	9.8181	−4.5131
	(1.2514)	(6.5243)	(9.0433)	(3.4612)
2	1.6862	−6.4647	12.646	−8.3724
	(1.2508)	(6.9097)	(10.504)	(4.5348)
3	2.4939	−15.250	32.646	−19.518
	(1.2596)	(7.3974)	(12.573)	(6.3055)
4	4.0905	−26.962	55.490	−31.014
	(1.2706)	(7.9583)	(15.000)	(8.5271)
5	3.9811	−26.350	52.023	−25.609
	(1.3211)	(8.6904)	(17.640)	(10.996)
6	1.1220	−5.4129	2.1667	12.449
	(1.2616)	(9.0138)	(20.330)	(14.357)

Standard errors in parentheses. The standard errors are based on spectral density estimates that were calculated as weighted averages of autocovariances using a Bartlett window that included 60 time lags. The normalized parameters $7^{-\gamma}[\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2]$ imply for every γ a local variance equal to one at interest rate level seven.

References

Aït-Sahalia, Y., 1996a, "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica* 64, 527–560.

Aït-Sahalia, Y., 1996b, "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.

Andersen, T. G., and T. Bollerslev, 1996, "Intraday Seasonality and Volatility Persistence in Financial Markets," forthcoming in *Journal of Empirical Finance.*

Andersen, T. G., and J. Lund, 1997, "Estimating Continuous Time Stochastic Volatility Models of the Short-Term Interest Rate," *Journal of Econometrics*, 77, 343–377.

Andrews, D. W. K., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Banon, G., 1978, "Nonparametric Identification for Diffusion Processes," *SIAM Journal of Control and Optimization*, 16, 380–395.

Bhattacharya, R. N., 1982, "On Functional Central Limit Theorem and the Law of the Iterated Logarithm for Markov Processes," *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 60, 185–201.

Bochner, S., 1960, *Harmonic Analysis and the Theory of Probability.* Berkeley: University of California Press.

Bollerslev, T., and E. Ghysels, 1996, "Periodic Autoregressive Conditional Heteroskedasticity," *Journal of Business and Economic Statistics*, 14, 139–151.

Bradley, R. C., 1986, "Basic Properties of Strong Mixing Conditions," in *Dependence in Probability and Statistics, a Survey of Recent Results.* Birkhauser, Oberwolfach.

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*, 48, 1209–1227.

Clark, P. K., 1973, "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135–155.

Cobb, L., P. Koppstein, and N. Y. Chen, 1983, "Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family," *Journal of the American Statistical Association*, 78, 124–130.

Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–406.

Durbin, J., 1970, "Testing for Serial Correlation in Least Squares Regression when some Regressors are Lagged Dependent Variables," *Econometrica*, 38, 410–421.

Fan, J., 1993, "Local Linear Regression Smoothers and Their Minimax Efficiencies," *Annals of Statistics*, 21, 196–216.

Fan, J., and I. Gijbels, 1992, "Variable Bandwidth and Local Linear Regression Smoothers," *Annals of Statistics*, 20, 2008–2036.

Fan, J., and E. Masry, 1994, "Local Polynomial Estimation of Regression Functions for Mixing Processes," manuscript, University of North Carolina, Chapel Hill.

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.

Foster, D. P., and D. B. Nelson, 1996, "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica* 64, 139–174.

Genot-Catelon, V., and J. Jacod, 1993, "On the Estimation of the Diffusion Coefficient for Multi-Dimensional Diffusion Processes," *Ann. Inst. Henri Poincare* 29, 119–151.

Ghysels, E., and J. Jasiak, 1995, "Stochastic Volatility and Time Deformation: An Application to Trading Volume and Leverage Effects," manuscript.

Haberman, S. J., 1989, "Concavity and Estimation," *Annals of Statistics*, 17, 1631–1661.

Hall, P., and C. C. Heyde, 1980, *Martingale Limit Theory and Its Application*, Academic Press, New York.

Hamilton, J. D., 1996, "The Daily Market for Fed Funds," *Journal of Political Economy* 104, 26–56.

Hansen, L. P., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029–1054.

Hansen, L. P., J. Heaton, and A. Yaron, 1996, "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics* 14, 262–280.

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

Hansen, L. P., J. A. Scheinkman, and N. Touzi, 1996, "Spectral Methods for Identifying Scalar Diffusions," forthcoming in *Journal of Econometrics.*

Harris, L., 1987, "Transaction Data Tests of the Mixture of Distributions Hypothesis," *Journal of Financial and Quantitative Analysis*, 22, 127–141.

Karatzas, I., and S. Shreve, 1991, *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York.

Karlin, S., and H. M. Taylor, 1981, *A Second Course in Stochastic Processes*, Academic Press, New York.

Longstaff, F. A., and E. S. Schwartz, 1992, "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance*, 47, 1259–1282.

Newey, W. K., and K. West, 1987, "A Simple Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 57, 1027–1057.

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.

Revuz, D., and M. Yor, 1991, *Continuous Martingales and Brownian Motion*, Springer-Verlag, New York.

Robinson, P. M., 1983, "Nonparametric Estimators for Time Series," *Journal of Time Series Analysis*, 4, 185–207.

Rosenblatt, M., 1969, "Density Estimates and Markov Sequences," in M. L. Puri (ed.), *Nonparametric Techniques in Statistical Inference*, Cambridge University Press, Cambridge, pp. 199–210.

Runolfsson, T., 1994, "Optimal Control of a Stochastic System with an Exponential-of-Integral Performance Criterion," *Systems & Control Letters*, 22, 451–456.

Stock, J. H., 1988, "Estimating Continuous-Time Processes Subject to Time Deformation, An Application to Postwar U.S. GNP," *Journal of the American Statistical Association*, 83, 77–85.

Stock, J. H., and J. Wright, 1995, "Asymptotics for GMM Estimators with Weak Instruments," manuscript, Harvard University.

Stone, C. J., 1977, "Consistent Nonparametric Regression," *Annals of Statistics*, 5, 595–620.

Tauchen, G. E., and M. Pitts, 1983, "The Price Variability Volume Relationship on Speculative Markets," *Econometrica*, 51, 485–505.

Wolak, F. A., 1991, "The Local Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models," *Econometrica*, 59, 981–995.