Macroeconomic Dynamics, **1**, 1997, 279–311. Printed in the United States of America.

BOOTSTRAPPING THE LONG RUN

TIMOTHY G. CONLEY
Northwestern University

LARS PETER HANSEN
University of Chicago
and
National Bureau of Economic Research

WEN-FANG LIU
University of Chicago

We develop and apply bootstrap methods for diffusion models when fitted to the long run as characterized by the stationary distribution of the data. To obtain bootstrap refinements to statistical inference, we simulate candidate diffusion processes. We use these bootstrap methods to assess measurements of local mean reversion or "pull" to the center of the distribution for short-term interest rates. We also use them to evaluate the fit of the model to the empirical density.

Keywords: Bootstrap Refinements, Statistical Inference, Markov Diffusions, Nonparametric Time-Series Estimation, Short-Term Interest Rates.

1. INTRODUCTION

In this paper we propose and implement bootstrapping methods for stationary diffusion models. We apply these methods to the test-function estimators of Conley et al. (in press) and to the goodness-of-fit density test of Ait-Sahalia (1996). An attractive feature of the test-function estimators is that they are easy to implement in practice, which then makes bootstrap methods feasible. The methods we explore entail simulating diffusion processes.

The econometric estimators we study are based, in large part, on fitting the stationary density of the diffusion. This is by design. Calibrating models to long-run implications is common in a variety of fields of economics. For instance, calibrating dynamic general equilibrium models often is based in part on selecting parameters to fit steady-state relations. Christiano and Eichenbaum (1992) demonstrate how to turn this into a formal econometric exercise. Similarly, Bertola and Caballero (1992) suggest fitting exchange models to stationary distributions. In our case, we expect the diffusion model to be misspecified at high frequencies, and we do not

We benefited from comments by Joel Horowitz, Bruce Lehmann, and George Tauchen. Hansen's research was funded in part by the National Science Foundation and the Guggenheim Foundation. Address correspondence to: Lars Peter Hansen, Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, USA; E-mail address: l_hansen@uchicago.edu.

 1365-1005/97 $9.00 + .10

want our estimation methods to be too sensitive to this misspecification. Fitting low-frequency time-series movements makes it more difficult to obtain reliable methods of statistical inference, however.

We show how to use simulation methods to improve the quality of the inferences based on long-run statistics. Central-limit approximations require an estimate of a long-run covariance matrix. Accurate estimation of this covariance matrix is known to be difficult, especially when the data are highly serially correlated.[1] Our application is to short-term interest rates, which are very persistent. Following on the suggestions of Hall (1992) and others, we use bootstrapping methods to obtain higher-order approximations to central-limit approximations. We bootstrap directly on the simulated diffusion and deliberately avoid using blocking methods, which recently have been advocated in the statistics literature.[2] Although the blocking methods have added robustness at a theoretical level, they seem poorly suited to problems such as ours in which the temporal dependence is substantial.

We also study local methods of estimation and inference. Following Conley et al. (in press), we use local methods to estimate the drift; and following Ait-Sahalia (1996), we use local methods to assess the fit of the model to the empirical stationary density of the data. For weakly dependent data, localization is known to (eventually) eliminate serial correlation, and the resulting central-limit approximations do not contain serial correlation corrections. In other words, the limiting distributions of the estimators and test statistics are the same as if the data were independent and identically distributed (i.i.d.). Because we mistrust these approximations when applied to short-term interest-rate data, we investigate the bootstrap corrections.

2. MODEL AND MEASUREMENT TARGET

Suppose that $\{x_t\}$ is the stationary solution to the stochastic differential equation

$$dx_t = \mu(x_t)\,dt + \sigma(x_t)\,dW_t,$$

where μ is the drift coefficient, σ^2 is a diffusion coefficient, and $\{W_t\}$ is a scalar Brownian motion. We assume that the resulting diffusion process is confined to $(0, \infty)$ with both boundaries nonattracting. Our primary aim is to measure the local pull of the diffusion toward the center of its distribution. Measures of *mean reversion* based solely on the drift ignore the fact that probabilities of moving to the right or the left from a given position depend also on the diffusion coefficient. Following Conley et al. (in press), we use local pull measures obtained as the coefficient in an expansion of the hitting-time probabilities. Starting at a state x, the probability $\rho(\varepsilon \mid x)$ of reaching $x + \varepsilon$ prior to $x - \varepsilon$ is approximately

$$\frac{1}{2} + \frac{\mu(x)}{2\sigma^2(x)}\varepsilon,$$

where the approximation error is of order ε^2. Thus, we take $\mu/2\sigma^2$ as the primary target of measurement in our investigation. In Section 3, we describe our methods

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

for estimating $\mu/2\sigma^2$. These methods are very easy to implement in practice, making it feasible to use bootstrap methods to assess the accuracy of our measurement.

As emphasized by Conley et al. (in press), even if the local hitting-time measure converges to zero for large interest rates, the diffusion still may be stationary. High-volatility elasticities (large responses of the diffusion coefficient to changes in the level of the process) may suffice to induce stationarity. Stationarity is volatility induced when

$$\int_1^{+\infty} \frac{\mu(y)}{\sigma^2(y)}\, dy > -\infty.$$

The estimation methods that we consider permit stationarity to be volatility induced.

3. MEASUREMENT METHOD

We use the test-function method of Conley et al. (in press) to construct asymptotically efficient estimators of the drift of a scalar diffusion process, given the diffusion coefficient and the marginal empirical distribution of the data. The idea is to infer properties of the drift of the diffusion from long-run information.

3.1. Drift Coefficient

We follow Cobb et al. (1983) and Conley et al. (in press) by considering polynomial specifications of the drift coefficient

$$\mu(x_t) = z_t^* \alpha,$$

where the asterisk denotes transposition (leaving a prime to denote differentiation)

$$z_t^* = \begin{bmatrix} (x_t)^{-1} & 1 & x_t & (x_t)^2 \end{bmatrix}.$$

The aim is to nest a linear drift specification and to provide some flexibility in approximating the stationary distribution of the diffusion. Cobb et al. (1983) give a justification for the flexibility of such specifications.[3]

3.2. Diffusion Coefficient

Our specification of the diffusion coefficient follows that of Chan et al. (1992) and assumes that the volatility elasticity is constant

$$\sigma^2(x) = \kappa(x)^\gamma,$$

where γ is the variance elasticity (and $\gamma/2$ is the volatility elasticity). We abstract from estimating the elasticity, but we show how measurements are sensitive to the elasticity specification. Given our parameterization of the drift, it is easy to verify that stationarity is volatility induced if γ exceeds three.

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

For flexible models of the drift, the diffusion coefficient and hence the variance elasticity are not identified from the stationary distribution.[4] Methods for estimating γ are described by Conley et al. (in press), but these estimates tend to be imprecise for short-term interest-rate data. We follow Conley et al. (in press) and display results for a coarse grid of variance elasticities.

3.3. Test-Function Estimator

We initially consider the estimation of α given κ and the variance elasticity γ. We estimate α using a test-function estimator of the form suggested by Hansen and Scheinkman (1995) and Conley et al. (in press). Let $\mathbf{\Psi}$ be a vector of four test functions. Because the diffusion is stationary, the vector Ito process $\{\Psi(x_t)\}$ inherits this stationarity. The time t vector of local means for this process is given by $\Psi'(x_t)\mu(x_t) + \frac{1}{2}\Psi''(x_t)\sigma^2(x_t)$. Under regularity conditions, the vector of local means for the process should have unconditional expectation zero. [See Hansen et al. (in press) for a statement of the regularity conditions.] This implies that the following moment conditions hold:

$$E[\Psi'(x_t)\mu(x_t)] = -\tfrac{1}{2}E\left[\Psi''(x_t)\sigma^2(x_t)\right]. \tag{1}$$

Thus, the parameter α satisfies

$$\alpha = -\left\{E\left[\Psi'(x_t)z_t^*\right]\right\}^{-1}E\left[\tfrac{1}{2}\Psi''(x_t)\sigma^2(x_t)\right].$$

The resulting discrete sample estimator of α is

$$a_T = -\left\{\sum_{t=1}^{T}\Psi'(x_t)z_t^*\right\}^{-1}\frac{1}{2}\sum_{t=1}^{T}\Psi''(x_t)\sigma^2(x_t).$$

For reasons of statistical efficiency, we construct our test-function vector $\mathbf{\Psi}$ to be proportional to the score vector of the log-likelihood for the stationary density function [see Conley et al. (in press)]. Because moment condition (1) depends only on derivatives of $\Psi(\cdot)$, we need only specify the score vector's derivatives, not its level. It is easy to verify that the test-function derivative (with respect to the state)

$$\Psi'(x) = \begin{bmatrix} x^{-\gamma-1} \\ x^{-\gamma} \\ x^{-\gamma+1} \\ x^{-\gamma+2} \end{bmatrix}$$

is proportional to the derivative of the score vector. Statistical efficiency is obtained regardless of the constant of proportionality.

Finally, because our measurement target is $\mu/2\sigma^2$, it suffices to identify μ and σ^2 up to scale. Therefore, in estimation using moment condition (1), we normalize κ to equal one, replacing κx^γ with x^γ.

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

3.4. Asymptotic Distribution

The limiting distribution for the parameter estimator $\{a_T\}$ has the usual form for a generalized method of moments estimator:

$$\sqrt{T}(a_T - \alpha) \approx -\left(E\left[\Psi'(x_t)z_t^*\right]\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\Psi'(x_t)z_t^*\alpha + \tfrac{1}{2}\left[\Psi''(x_t)(x_t)^\gamma\right]\right].$$

Use of this limiting distribution requires an estimate of the covariance matrix Δ in the central-limit approximation

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\Psi'(x_t)z_t^*\alpha + \tfrac{1}{2}E\left[\Psi''(x_t)(x_t)^\gamma\right]\right] \xrightarrow{d} N(0, \Delta).$$

Because the data are temporally dependent, one way to estimate Δ is to use a spectral density estimator at frequency zero. Such estimators are known to have slow rates of convergence, and their inaccuracy might undermine the quality of statistical inference.

For comparison, we also consider the limiting distribution for an (infeasible) estimator that uses a continuous record. The limiting distribution for the continuous-record estimator is

$$\sqrt{T}(a_T - \alpha) \approx -\left\{E\left[\Psi'(x_t)z_t^*\right]\right\}^{-1}\frac{1}{\sqrt{T}}\int_0^T\left[\Psi'(x_t)z_t^*\alpha + \tfrac{1}{2}\left[\Psi''(x_t)(x_t)^\gamma\right]\right]dt.$$

It follows from the continuous-time martingale central-limit theorem for diffusions that

$$\frac{1}{\sqrt{T}}\int_0^T\left[\Psi'(x_t)z_t^*\alpha + \tfrac{1}{2}\left[\Psi''(x_t)(x_t)^\gamma\right]\right]dt \xrightarrow{d} N(0, \Lambda),$$

where

$$\Lambda = \frac{1}{\kappa^2}E\left[\sigma^2(x_t)\Psi'(x_t)\Psi'(x_t)^*\right] = \frac{1}{\kappa}E\left[(x_t)^\gamma\Psi'(x_t)\Psi'(x_t)^*\right]. \qquad (2)$$

[For example, see Bhatacharia (1982) and Hansen and Scheinkman (1995).] When a discrete sample is used, the covariance matrix Λ should provide a lower bound on the matrix Δ. The expression $E[(x_t)^\gamma\Psi'(x_t)\Psi'(x_t)^*]$ used in constructing Λ can be estimated directly by a sample moments estimator. Thus, provided we can estimate κ, we can compute a sample estimator of the lower bound Λ. An estimator of κ is discussed below.

4. DATA

We use a data set of Federal-fund interest rates. It consists of daily observations on "effective" Federal-fund overnight interest rates, measured as annualized percentages, from January 2, 1970, to January 29, 1997.[5] The data set has some missing

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

observations, primarily because of weekends and holidays, but there are still a total of 6826 observations.

5. SCALE ESTIMATION

We explore bootstrap methods that entail simulating diffusion processes. These simulations require an estimate of κ. This parameter pins down the timescale of the process and hence can be interpreted as dictating the sampling frequency of our observations. Larger values of κ make the process less temporally dependent because it corresponds to sampling the (weakly dependent) stationary diffusion less frequently. We obtain an estimate of κ using the sieve estimation method of Chen et al. (1997). They devise a two-step, semiparametric method for estimating the diffusion coefficient. We use a coarse grid for the variance elasticity γ by letting it be 1, 2, 3, and 4. We then apply their method to estimate κ for each of the four choices of γ.[6] This two-step procedure is based on the following identification scheme.

5.1. Identification of κ

Let C denote the set of absolutely continuous functions ψ such that $E[\psi(x_t)^2] = 1$, $E[\psi(x_t)] = 0$, and $E[(x_t)^\gamma \psi'(x_t)^2] < \infty$. We use this set as the constraint set for the following optimization problem:

$$\rho = \min_{C} E\left[(x_t)^\gamma \psi'(x_t)^2\right]. \tag{3}$$

Let ϕ denote the solution to this problem. The solution to this problem is an eigenfunction of the infinitesimal generator of the process and of the conditional-expectation operators over any interval of time. Consequently, the sampled version of the process $\{\phi(x_t)\}$ has a first-order autoregressive representation:

$$\phi(x_{t+1}) = \exp(-\delta)\phi(x_t) + e_{t+1}, \tag{4}$$

where the disturbance term e_{t+1} has expectation zero conditioned on x_t. It can be shown that

$$\kappa = \delta/\rho. \tag{5}$$

5.2. Estimation of ρ

Estimation problem (3) is in the form of a generalized eigenfunction problem, which is the operator counterpart to a generalized eigenvector problem. In practice, we solve a corresponding generalized eigenvector problem using a finite collection of basis functions. We use B-spline wavelets of order three as basis functions. These functions have compact support, are twice continuously differentiable, and are piecewise cubic polynomials. We disperse these functions in an equally spaced fashion throughout the support of the distribution by adding translation terms to the

argument of an initial reference function. Let B denote the resulting n-dimensional vector of functions. Our approximation to ϕ is given by a linear combination of the basis functions

$$\phi_T(x) = \boldsymbol{b}_T^* \boldsymbol{B}(x),$$

where we construct the coefficient vector $\boldsymbol{b}_T$ as described below.

Form two matrices,

$$V_T = \frac{1}{T}\sum_{t=1}^{T}(x_t)^{\gamma}\boldsymbol{B}'(x_t)\boldsymbol{B}'(x_t)^*$$

$$W_T = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{B}(x_t)\boldsymbol{B}(x_t)^*,$$

where an asterisk is used to denote transposition and a prime is used to denote differentiation. A third matrix used to penalize the second derivative of the approximating matrix is built as follows. Construct a matrix K_T such that

$$K_T\boldsymbol{b}_T = \begin{bmatrix} b_{2,T} - b_{1,T} \\ b_{3,T} - 2b_{2,T} + b_{1,T} \\ \cdots \\ b_{n,T} - 2b_{n-1,T} + b_{n-2,T} \\ -b_{n,T} + b_{n-1,T} \end{bmatrix}$$

for $\boldsymbol{b}_T^* = [b_{1,T} \;\; b_{2,T} \;\; \ldots \;\; b_{n-1,T} \;\; b_{n,T}]$. The third matrix is

$$U_T = (K_T)^* D_T K_T,$$

where D_T is a prespecified diagonal matrix with positive diagonal entries.

To approximate ϕ, we solve the problem

$$(\lambda_T U_T + V_T)\boldsymbol{b}_T = \nu_T W_T \boldsymbol{b}_T, \tag{6}$$

where λ_T dictates the magnitude of the penalty on the second differences of the coefficient vector $\boldsymbol{b}_T$. We specify this penalty a priori. The scalar ν_T is a generalized eigenvalue to be computed along with the generalized eigenvector $\boldsymbol{b}_T$. It turns out that the $\nu_T = 0$ and $\boldsymbol{b}_T$ is equal to a coefficient vector with a one in each position that solves (6). Of course, this solution will not have sample mean zero. Instead, we take the (generalized) eigenvector associated with the positive (generalized) eigenvalue that is closest to but distinct from zero. The resulting eigenvector $\boldsymbol{b}_T$ is normalized so that $b_T^* W_T b_T = 1$. Our estimate r_T of ρ is given by

$$r_T = \boldsymbol{b}_T^* V_T \boldsymbol{b}_T.$$

This generalized eigenvector method is easy to implement in practice.

Applying this method, we are led to the estimates of ρ in Table 1.[7]

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

TABLE 1. Estimates of ρ

	γ			
	1	2	3	4
ρ	0.549	4.97	27.4	118

5.3. Estimation of δ

To calibrate our simulations, for each value of γ, we estimate δ in two ways. One method estimates the first-order autoregression (4) of the approximating eigenfunction via least squares. The estimate of δ is the negative of the logarithm of the estimated autoregressive coefficient.

The federal-funds data that we use in our estimation are known to have some high-frequency movements that are not well suited for a (time-invariant) scalar diffusion model. For instance, there is a known two-week institutional cycle that, at the very least, introduces periodicity into the interest-rate difference process [see Hamilton (1996) and others]. Also, the data contain some systematically missing observations because of weekends and holidays. These high-frequency movements may lower the least-squares estimate of δ for uninteresting reasons. As a second approach, we modify the least-squares estimator by omitting high frequencies from a Gaussian approximation to the log likelihood. Thus we use a discrete version of Whittle's pseudo-log-likelihood function, which is constructed as follows. Form the periodogram

$$I(\theta_j) = \frac{1}{T}\left|\sum_{t=1}^{T} x_t \exp(-it\theta_j)\right|^2$$

at frequency θ_j, where

$$\theta_j = \frac{2\pi j}{T}, \qquad j = 1, 2, \ldots, T.$$

We can only construct the spectral density function implied by (4) up to scale from δ via

$$f(\theta_j; \delta) = \frac{1}{1 + \exp(-2\delta) - 2\exp(-\delta)\cos(\theta_j)}.$$

At the same time that we estimate δ, we are compelled to estimate the variance of the implied innovation. Let J define an index set for the included (low) frequencies. Concentrating out the variance estimate and omitting high frequencies results in the pseudo-log-likelihood (up to a constant)

$$Q_T(d) = -\frac{1}{\#J}\sum_{j\in J}\frac{I(\theta_j)}{f(\theta_j; d)} - \frac{1}{\#J}\sum_{j\in J}\log\left[\frac{I(\theta_j)}{f(\theta_j; d)}\right],$$

where $\#J$ denotes the number of elements of J. We exclude from J frequencies with periods less than or equal to 10 time periods. Our choice of 10 time periods as

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

TABLE 2. Estimates of δ

	γ			
Method	1	2	3	4
Least squares	1.09e-2	1.05e-2	2.13e-2	3.12e-2
Whittle	3.51e-3	2.00e-3	3.00e-3	4.01e-3

TABLE 3. Estimates of κ[a]

	γ			
Process	1	2	3	4
Low persistence	1.98e-2	2.11e-3	7.75e-4	2.64e-4
High persistence	6.39e-3	4.02e-4	1.10e-4	3.40e-5

[a]The low-persistence estimates are constructed from the least-squares estimates of δ, and the high-persistence estimates are constructed from the Whittle estimates of δ.

the cutoff point is that it corresponds to the two-week institutional cycle described by Hamilton (1996). In other words, we omit frequencies in the range

$$\pi/5 \leq \theta_j \leq 9\pi/5.$$

By maximizing $Q_T(d)$, we obtain what we refer to as Whittle estimates of δ.

Applying these two different methods we are led to the estimates of δ shown in Table 2. Notice that the Whittle estimates are much smaller than the least-squares estimates. Thus, our calibrated diffusions based on the Whittle estimates are considerably more persistent. We construct κ estimates based on both the least-squares and the Whittle estimates of ρ and assess inference methods using the resulting specifications.

Via equation (5) the above estimates of δ and ρ yield the scale values shown in Table 3.[8] The difference between low-persistence and high-persistence processes is quite substantial. When γ is equal to one, the increase in persistence is roughly equivalent to reducing the sample size to less than one-third of its original size. On the other hand, for the higher volatility elasticity, the effective change in sample size is much more modest (about three-fourths of its original size).

Armed with these calibrations of κ and our previously described estimates of the drift coefficient (up to scale) we have two fully specified scalar diffusion processes for each variance elasticity γ. The resulting processes are simulated in our subsequent bootstrapping and Monte Carlo analysis.

6. GOODNESS-OF-FIT TEST

We use the stationary density specification test of Bickel and Rosenblatt (1973) and Ait-Sahalia (1996) as a goodness-of-fit test. Ait-Sahalia (1996) extends the

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

Bickel and Rosenblatt (1973) analysis to accommodate weakly dependent data and parameter estimation. The basic idea is to test a parametric model of a diffusion process by comparing the model's implied stationary density, in this case $\pi(x; \alpha, \gamma)$, to a nonparametric estimator of the true stationary density: $\pi(x)$. The metric for distance between a candidate parametric density and the truth, $\pi(x)$, is

$$M = \int_0^\infty [\pi(x; \alpha, \gamma) - \pi(x)]^2 \pi(x)\, dx. \tag{7}$$

We estimate the general alternative specification for π using a kernel density estimator

$$\hat{\pi}(x; h_T) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{h_T} K\left(\frac{x - x_t}{h_T}\right),$$

where h_T is a bandwidth sequence that converges to zero at a rate specified below and $K(\cdot)$ is a Gaussian kernel. For each variance elasticity γ, we use the previously described a_T to estimate α. Thus, our normalized estimator of M is

$$\hat{M} = T h_T \int_0^\infty [\pi(x; \boldsymbol{a}_T, \gamma) - \hat{\pi}(x; h_T)]^2 \hat{\pi}(x; h_T)\, dx,$$

where the integral in (7) is approximated numerically.

It follows from Ait-Sahalia (1996) that, under regularity conditions, if the parametric specification is correct and the bandwidth sequence h_T satisfies $\lim_{T\to\infty} T h_T = \infty$ and $\lim_{T\to\infty} T h_T^{4.5} = 0$, then

$$h_T^{-1/2}(\hat{M} - E_M) \xrightarrow{d} N(0, V_M),$$

where the asymptotic mean E_M and variance V_M are given by

$$E_M = \left[\int_{-\infty}^{\infty} K(u)^2\, du\right] \int \pi(x)^2\, dx$$

$$V_M = 2\left\{\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} K(u) K(u+s)\, du\right]^2 ds\right\} \int \pi(x)^4\, dx.$$

This limiting distribution coincides with that obtained by Bickel and Rosenblatt (1973) even though they abstracted from serial correlation and parameter estimation. Because of the difference in rates of convergence for the parametric and nonparametric parts of $\hat{M}$, the parametric estimation does not impact its limiting distribution. The limiting distribution of $\hat{M}$ is the same with weakly dependent data as with independent data because of the local nature of the nonparametric density estimator.

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

We estimate E_M by $\hat{E}_M = [\int_{-\infty}^{\infty} K(u)^2\,du] \int \hat{\pi}(x; h_T)^2\,dx$ and estimate V_M by an analogously defined $\hat{V}_M$, where again the integrals are numerically evaluated. We focus on the standardized version of this test statistic

$$\tilde{M} \equiv \frac{\hat{M} - \hat{E}_M}{\sqrt{h_T \hat{V}_M}},$$

whose limiting distribution is standard normal.

7. SIMULATION

To form bootstrap adjustments to confidence intervals and test statistic distributions, we simulate the estimated diffusion process. In so doing, we use the methods described by Kloeden and Platen (1995).[9] Unfortunately, the regularity conditions imposed by Kloeden and Platen (1995) to justify simulation methods are too strong for our applications. As they emphasize, their approximation theorems only provide sufficient conditions. Consequently, there is scope for simulation methods to be more generally applicable. In this section, we describe the simulation methods we use in our bootstrapping, and we discuss the properties of the error term in the discretization.

7.1. Euler Approximation

The Euler approximation is based on the (first-order) Ito formula,

$$x_{t+\eta} = x_t + \mu(x_t)\eta + \sigma(x_t)(W_{t+\eta} - W_t) + r_{t,\eta},$$

where $r_{t,\eta}$ is a remainder term and satisfies

$$r_{t,\eta} = \int_0^{\eta} [\mu(x_{t+s}) - \mu(x_t)]\,ds + \int_0^{\eta} [\sigma(x_{t+s}) - \sigma(x_t)]\,dW_{t+s},$$

where we interpret the second integral as a stochastic integral. Let $\|\|$ denote the mean-square norm. Then, by the Triangle Inequality and the construction of the stochastic integral,

$$\lim_{\eta \searrow 0} \frac{1}{\eta} \|r_{t,\eta}\| = 0$$

provided that

$$\lim_{\eta \searrow 0} \|\mu(x_{t+\eta}) - \mu(x_t)\| = 0$$

$$\lim_{\eta \searrow 0} \|\sigma(x_{t+\eta}) - \sigma(x_t)\| = 0.$$

These two requirements will be satisfied if

$$E\left[\mu^2(x_t)\right] < \infty$$

$$E\left[\sigma^2(x_t)\right] < \infty.$$

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

These moment restrictions are only problematic for $\gamma = 3$ and $\gamma = 4$. For $\gamma = 3$, the moment restrictions will be satisfied provided that the coefficient on x^2 in the formula for the drift is sufficiently negative, as is the case in our calibrations. Neither of these moment requirements is met for our $\gamma = 4$ calibrations, in part because stationarity is volatility induced. Nevertheless, the diffusions remain stationary and geometrically ergodic.

7.2. Simulating in Logarithms

One way to overcome the potential breakdown of the Euler approximation for high-volatility elasticities is to simulate the process in logarithms. Construct

$$y_t = \log(x_t).$$

The counterparts to the finite-moment conditions for levels are

$$E\left[\mu(x_t)^2 \frac{1}{(x_t)^2}\right] < \infty$$

$$E\left[\sigma^2(x_t) \frac{1}{(x_t)^2}\right] < \infty,$$

both of which are satisfied for the high-volatility elasticity models ($\gamma = 4$). Thus, for the high-volatility elasticity models, we use a first-order simulation scheme applied to logarithms.

The drift for log process is given by

$$\begin{aligned}\mu_y(y_t) &= \frac{1}{x_t}\mu(x_t) - \frac{1}{2(x_t)^2}\sigma^2(x_t) \\ &= \frac{1}{\exp(y_t)}\mu[\exp(y_t)] - \frac{\kappa}{2}\exp[(\gamma - 2)y_t],\end{aligned}$$

and the diffusion coefficient by

$$\sigma_y^2(y_t) = \kappa \exp[(\gamma - 2)y_t].$$

7.3. Second-Order Scheme

A second-order scheme aims to improve the accuracy of the first-order, Euler approximation. The idea is to approximate the evolution of the ingredients μ and σ using an Euler scheme. Thus, we are led to study the local mean and local standard deviation of μ and σ. The local mean of μ is given by $\mu\mu' + \frac{1}{2}\sigma^2\mu''$ and

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

its local standard deviation is $\sigma\mu'$. The local mean of σ is $\mu\sigma' + \frac{1}{2}\sigma^2\sigma''$ with a local standard deviation $\sigma\sigma'$. The approximation takes the form

$$\begin{aligned} x_{t+\eta} = x_t &+ \mu(x_t)\eta + \sigma(x_t)(W_{t+\eta} - W_t) + \left(\mu\mu' + \tfrac{1}{2}\sigma^2\mu''\right)\eta^2 \\ &+ \sigma(x_t)\mu'(x_t)\int_t^{t+\eta}\int_t^u dW_s\,du \\ &+ \left[\mu(x_t)\sigma'(x_t) + \tfrac{1}{2}\sigma^2(x_t)\sigma''(x_t)\right]\int_t^{t+\eta}\int_t^u ds\,dW_u \\ &+ \sigma(x_t)\sigma'(x_t)\int_t^{t+\eta}\int_t^u dW_s\,dW_u + q_{t,\eta}, \end{aligned}$$

where $q_{t,\eta}$ is the approximation error. Additional restrictions are needed to ensure that the resulting stochastic integrals are well behaved. For instance, the two local means and two local volatilities should have finite second moments under the stationary distribution.

This refinement is problematic for the log specification with high-volatility elasticities. In particular,

$$E\left[\sigma_y^2(y_t)\sigma_y'(y_t)^2\right] = \kappa^2(\gamma - 2)^2 E\left[(x_t)^{2\gamma-4}\right] = \infty$$

when $\gamma = 4$.

7.4. Implementation

To implement the simulation methods, we used the second-order scheme for variance elasticities: $\gamma = 1, 2, 3$ with η set to 0.1. For variance elasticity $\gamma = 4$, we simulated the process in logarithms using a first-order scheme with $\eta = 0.05$.

8. BOOTSTRAPPING THE CONFIDENCE INTERVALS

We use simulation methods to refine the confidence interval construction for estimates of $[\mu(y)]/[2\sigma^2(y)]$ for a collection of states y and variance elasticities γ. To construct the bootstrap confidence intervals, we simulate from the values of μ estimated using federal funds data for the prespecified variance elasticities. In conducting the simulations, we use the two different choices of the persistence parameter κ given in Table 3. We simulate 500 samples of length 7000 for each variance elasticity and persistence level.

Hall (1992) shows that bootstrapping methods deliver asymptotic refinements for studentized statistics in i.i.d. environments. More recently, Datta and McCormick (1995) show that bootstrapping an estimated Markov chain can sometimes provide second-order accuracy. Similarly, Gotze and Kunsch (1996) show

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

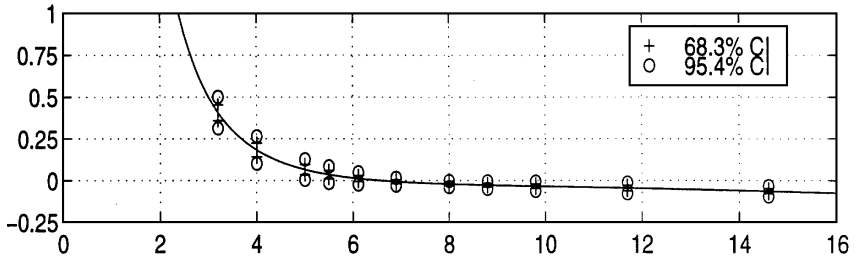
that block bootstrapping methods deliver second-order refinements for some stationary processes. Unfortunately, the results of Datta and McCormick and Gotze and Kunsch (1996) are not directly applicable to our application. While our strategy of simulating fitted diffusions imitates the parametric bootstrapping method for a Markov chain, our standard error estimators in both the actual and bootstrap samples are constructed from Bartlett spectral density estimators. These spectral density estimators are known to have slow rates of convergence. While this complication is addressed by Gotze and Kunsch (1996), we are not using their blocking method because we expect that simulating the diffusion will provide a better approximation of the true dependence structure. Hall (1992) also describes the advantage to constructing symmetric confidence intervals. Unfortunately, the second-order bootstrap refinements shown by Datta and McCormick (1995) and Gotze and Kunsch (1996) are not sufficient to provide the same degree of accuracy as in i.i.d. environments. Nevertheless, we use symmetric confidence intervals in our analysis. Thus, in forming these intervals, we compute an estimate s_T^k of $[\mu(y)]/[2\sigma^2(y)]$ for simulated data set k, subtract the corresponding estimate s_T from the actual data set, and divide by standard error estimated from the simulated sample.[10] Let z_T^j denote the resulting t-statistic. In forming this statistic, we initially use the same standard-error estimation method as Conley et al. (in press). The standard errors are constructed from Bartlett spectral density estimates with a cutoff at 60 lags. Subsequently, we study the robustness of our results to changes in this cutoff point. We obtain critical values for the bootstrap t-statistic distribution by locating the α% right tail of the empirical distribution of $\{|z_T^j|\colon j = 1, 2, \ldots, N\}$. The product of the sample estimate of the standard error and this critical value is added to s_T and subtracted from s_T to construct the symmetric α% confidence interval. This is repeated for different states and different variance elasticities.

8.1. Results

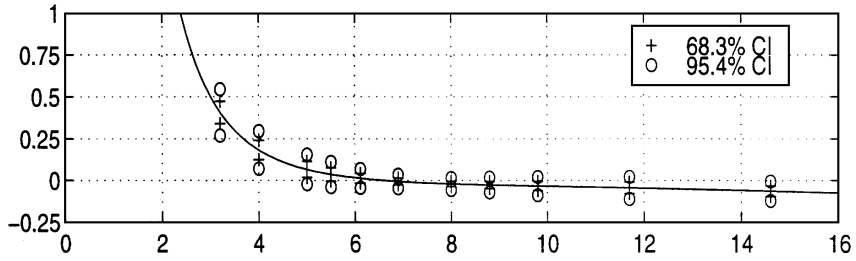
In Figures 1–4, we present point estimates [reported by Conley et al. (in press)] of the pull $(\mu/2\sigma^2)$ with confidence intervals calculated three ways. The confidence intervals in Panel A are one and two standard-error bands computed via the delta method. Panels B and C contain bootstrap confidence intervals for the two different calibrations of κ, which we label as low persistence and high persistence.

Even for the low-persistence specification, the bootstrap method magnifies the confidence intervals quite substantially. This magnification is enhanced when we simulate from a more persistent process. For the less-persistent specification, the confidence intervals at the 68.3% level (one standard deviation for a normal distribution) and 95.4% level (two standard deviations) from the bootstrap distributions are between 1.5 and 2 times wider than the ones from the asymptotic distributions for variance elasticities (γ's) equal to 1, 2, and 3. For variance elasticity 4, they are between 1.1 to 1.5 times larger. The bootstrap confidence intervals for the more persistent specification are even wider. They are about 2 to 3 times larger for γ equal to 1, 2, and 3, and 1.5 to 2 times larger for γ equal to 4.

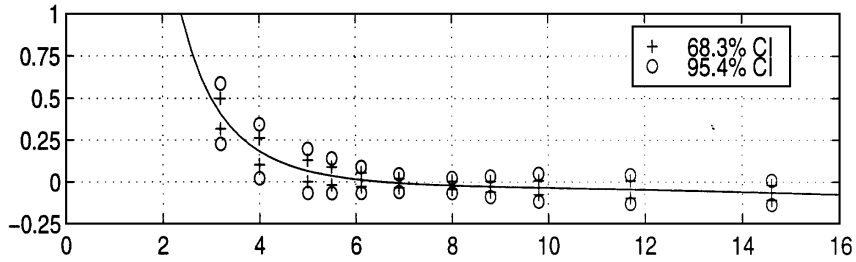
https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press



A) Pull Estimates for $\gamma = 1$



B) Low-Persistence Specification



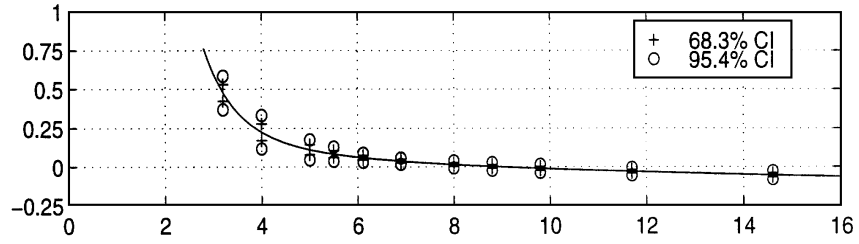
C) High-Persistence Specification

FIGURE 1. Pull $(\mu/2\sigma^2)$ estimates and confidence intervals corresponding to $\gamma = 1$. Point estimates in all panels are obtained using federal funds data. Confidence intervals in Panel A are one and two standard-error bands computed via the delta method and a Bartlett spectral density estimator with 60 lags. Panels B and C contain bootstrap confidence intervals for low- and high-persistence data sets, respectively.

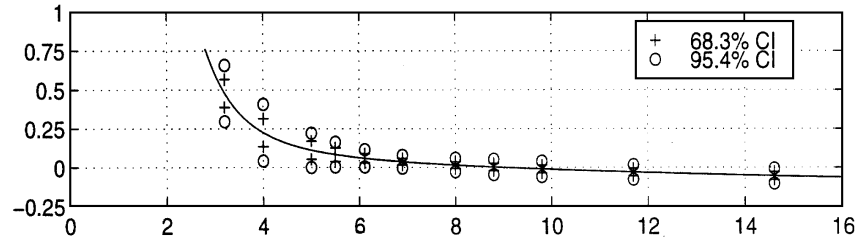
8.2. Alternative Covariance Matrix Estimators

We now explore the sensitivity of our results to changes in the way we estimate the asymptotic covariance matrix for the drift estimator. First we look at how the bootstrap confidence intervals change when we increase the number of lags used in our Bartlett estimator of the long-run covariance matrix Δ. Our previous results imitated those of Conley et al. (in press) by using an estimate of the long-run covariance matrix based on a time-domain Bartlett window with a cutoff of 60 lags. When we expanded the cutoff point to 100 lags, our standard-error

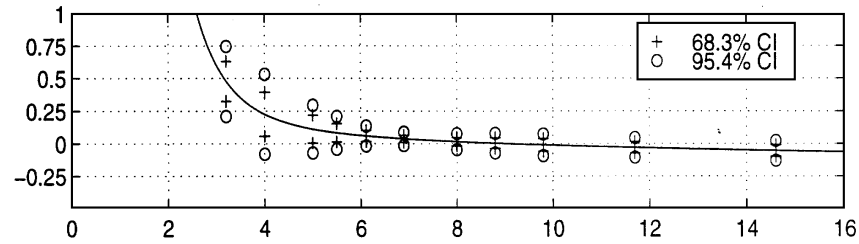
https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press



A) Pull Estimates for $\gamma = 2$



B) Low-Persistence Specification



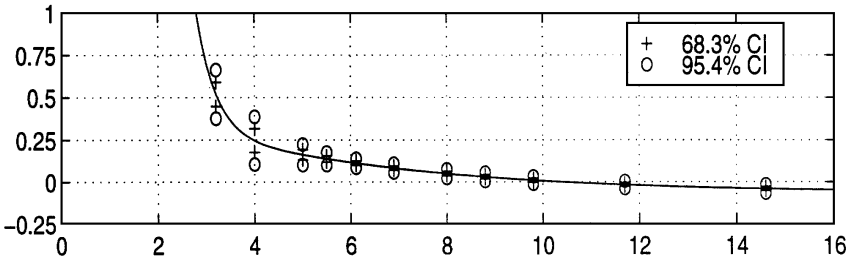
C) High-Persistence Specification

FIGURE 2. Pull $(\mu/2\sigma^2)$ estimates and confidence intervals corresponding to $\gamma = 2$. Point estimates in all panels are obtained using federal funds data. Confidence intervals in Panel A are one and two standard-error bands computed via the delta method and a Bartlett spectral density estimator with 60 lags. Panels B and C contain bootstrap confidence intervals for low- and high-persistence data sets, respectively.

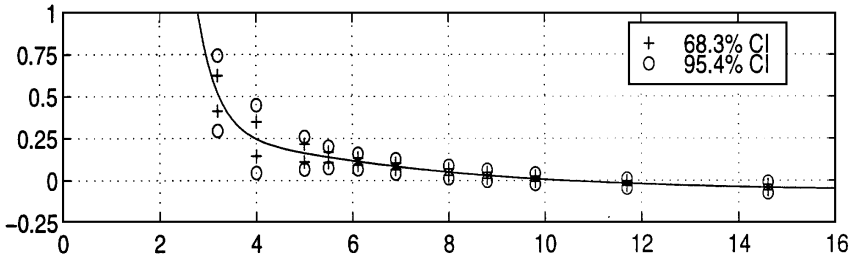
estimates increased and our bootstrap distribution for the t-statistic became more concentrated. These two effects are not fully offsetting, however. For the 68.3% confidence interval, the net effect is that an increase in the length of the Bartlett window from 60 to 100 enlarges the confidence intervals by 4 to 12% for the high-persistence specification, and by 5 to 18% for the low-persistent specification.

Recall that Λ is constructed under the premise that the drift estimator uses a continuous record of data over a finite interval instead of a discrete-time sample. Consequently, from the asymptotic theory, we expect Λ to understate the sampling variability in the drift estimator. However, a Bartlett estimator of Δ is expected to be less reliable than an estimator of Λ because the rate of convergence of the

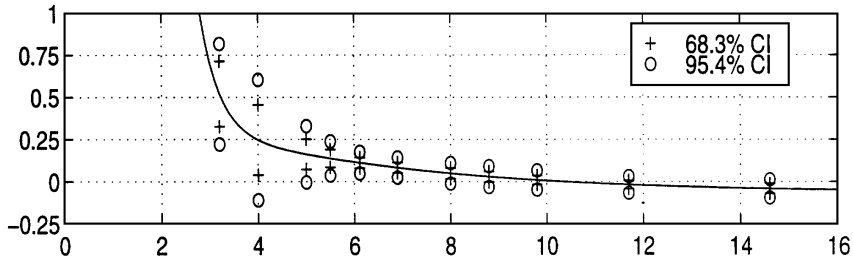
https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press



A) Pull Estimates for $\gamma = 3$



B) Low-Persistence Specification



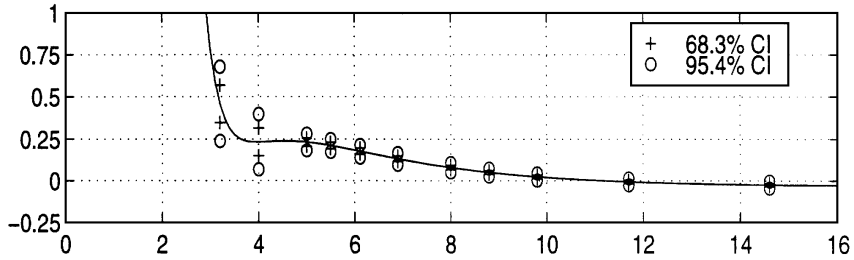
C) High-Persistence Specification

FIGURE 3. Pull $(\mu/2\sigma^2)$ estimates and confidence intervals corresponding to $\gamma = 3$. Point estimates in all panels are obtained using federal-funds data. Confidence intervals in Panel A are one and two standard-error bands computed via the delta method and a Bartlett spectral density estimator with 60 lags. Panels B and C contain bootstrap confidence intervals for low- and high-persistence data sets, respectively.

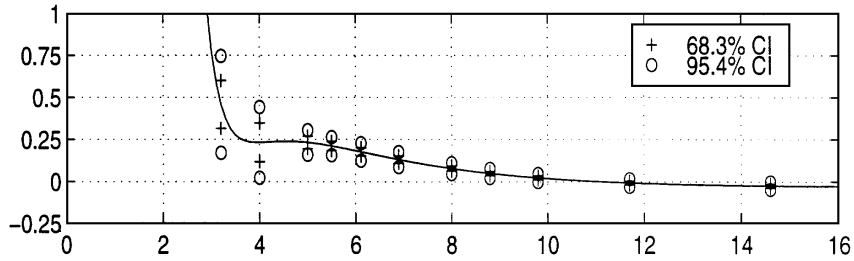
latter estimator is faster. We now compare our estimates of Δ to those of Λ. If our estimates of Λ exceed those of Δ, this suggests that the understatement of the sampling variability can be attributed in part to a downward distortion in our Δ estimates.

Formally, we compare standard errors for our pull measure based on Δ estimates to standard errors based on Λ estimates. We also extract the standard error counterparts from our bootstrap confidence intervals. This is done by using the upper and lower endpoints of the symmetric bootstrap 68.3% confidence intervals.

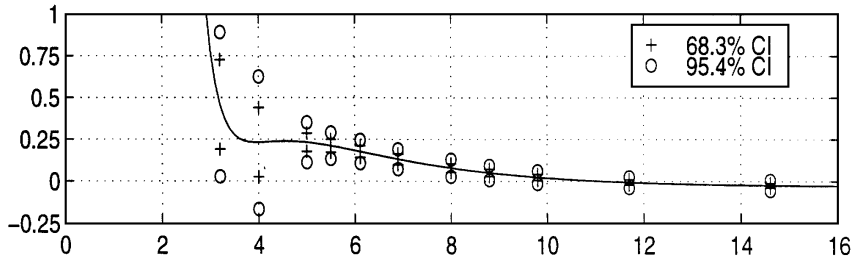
https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press



A) Pull Estimates for $\gamma = 4$



B) Low-Persistence Specification



C) High-Persistence Specification

FIGURE 4. Pull $(\mu/2\sigma^2)$ estimates and confidence intervals corresponding to $\gamma = 4$. Point estimates in all panels are obtained using federal-funds data. Confidence intervals in Panel A are one and two standard-error bands computed via the delta method and a Bartlett spectral density estimator with 60 lags. Panels B and C contain bootstrap confidence intervals for low- and high-persistence data sets, respectively.

(Recall that 0.683 is the probability that a normally distributed random variable is within one standard deviation of its mean.) These three measures of sampling variability are reported in Tables 4 and 5. The standard errors constructed from our Λ estimates are larger than those based on our Δ estimates except when the persistence is low and $\gamma = 4$. Thus, the enhancement of the bootstrap confidence intervals could have been anticipated by the fact that the sample ordering between Λ and Δ is the reverse of the population ordering.

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

TABLE 4. Standard error for federal-funds pull estimates for $\gamma = 1$ and 2 using Δ and Λ

Decile[a] (%)	SE[b]	Low persistence		High persistence	
		Cts. SE[c]	Bootstrap SE[d]	Cts. SE[c]	Bootstrap SE[d]
		(A) Variance Elasticity = 1			
10	0.041	0.040	0.057	0.071	0.080
20	0.031	0.032	0.049	0.056	0.065
30	0.025	0.028	0.040	0.049	0.053
40	0.018	0.024	0.030	0.042	0.042
50	0.011	0.020	0.021	0.036	0.027
60	0.009	0.020	0.017	0.035	0.024
70	0.011	0.021	0.021	0.036	0.032
80	0.014	0.022	0.028	0.039	0.042
90	0.016	0.024	0.035	0.042	0.052
		(B) Variance Elasticity = 2			
10	0.054	0.068	0.091	0.155	0.169
20	0.033	0.045	0.059	0.103	0.106
30	0.023	0.036	0.044	0.082	0.071
40	0.015	0.028	0.030	0.065	0.043
50	0.010	0.024	0.021	0.054	0.028
60	0.011	0.022	0.022	0.051	0.033
70	0.013	0.022	0.024	0.050	0.042
80	0.013	0.021	0.026	0.048	0.048
90	0.013	0.020	0.025	0.045	0.044

[a] Interest rate deciles are 4.0, 5.0, 5.5, 6.1, 6.9, 8.0, 8.8, 9.8, and 11.7.
[b] Standard errors are computed via a Bartlett estimator of Δ.
[c] Standard errors are computed via an estimator of Λ.
[d] Standard errors are based on Δ with the bootstrap adjustment.

As might be expected, for the low-persistence (high values of κ) simulations, the standard-error estimates based on Λ are typically smaller than their bootstrap confidence-interval counterparts. In contrast, for the high-persistence simulations, this ordering is often reversed and magnitude of the difference is sometimes sizable. The high-persistence simulations are the ones in which we expect the asymptotic approximations (including possibly the bootstrap refinements) to be less reliable. The bootstrap is known to fail in the presence of a unit root [see Basawa et al. (1991)] and thus one might be concerned that our bootstrap corrections are also less reliable when the data are more highly persistent.

9. BOOTSTRAPPING THE GOODNESS-OF-FIT TEST

Ait-Sahalia (1994) reports a rejection of the model specifications for Euro-Dollar data using a goodness-of-fit test described in Section 6. However, Pritsker (1996) documents size distortion in Ait-Sahalia's goodness-of-fit test when applied to

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

TABLE 5. Standard errors for federal-funds pull estimates for $\gamma = 3$ and 4 using Δ and Λ

Decile[a] (%)	SE[b]	Low persistence		High persistence	
		Cts. SE[c]	Bootstrap SE[d]	Cts. SE[c]	Bootstrap SE[d]
		(A) Variance Elasticity = 3			
10	0.070	0.062	0.102	0.166	0.20
20	0.030	0.033	0.053	0.088	0.089
30	0.019	0.024	0.032	0.065	0.052
40	0.013	0.019	0.021	0.050	0.031
50	0.013	0.016	0.019	0.042	0.031
60	0.013	0.013	0.019	0.036	0.032
70	0.013	0.012	0.018	0.032	0.031
80	0.012	0.010	0.016	0.028	0.027
90	0.011	0.009	0.014	0.024	0.025
		(B) Variance Elasticity = 4			
10	0.082	0.058	0.115	0.162	0.207
20	0.025	0.025	0.037	0.069	0.054
30	0.019	0.018	0.028	0.049	0.039
40	0.019	0.014	0.025	0.038	0.035
50	0.018	0.011	0.022	0.030	0.032
60	0.014	0.008	0.016	0.022	0.026
70	0.012	0.007	0.013	0.018	0.021
80	0.011	0.005	0.012	0.015	0.017
90	0.010	0.005	0.011	0.013	0.016

[a] Interest rate deciles are 4.0, 5.0, 5.5, 6.1, 6.9, 8.0, 8.8, 9.8, and 11.7.
[b] Standard errors are computed via a Bartlett estimator of Δ.
[c] Standard errors are computed via an estimator of Λ.
[d] Standard errors are based on Δ with the bootstrap adjustment.

Ornstein–Uhlenbeck processes. Motivated by these findings, we compute bootstrap corrections for the goodness-of-fit test applied to federal-funds data.[11]

To implement this test, we must choose a bandwidth for the kernel density estimator. An optimal bandwidth choice for estimating the density itself requires that the bandwidth converge to zero sufficiently slowly for a bias term to enter the limiting normal distribution. However, as noted by Ait-Sahalia (1996), in constructing the density-fit test statistic there should be a little bit of "undersmoothing" to eliminate the contribution of the bias to the limiting distribution.

Even an (asymptotically) optimal bandwidth choice is problematic for our application to short-term interest rates. For instance, a common method of bandwidth choice is cross validation, and there are many results on cross-validation procedures for i.i.d. data. However, there is good reason to believe that these cross-validation procedures will not work well for highly dependent data. Cross-validation techniques systematically omit observations to measure the goodness

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

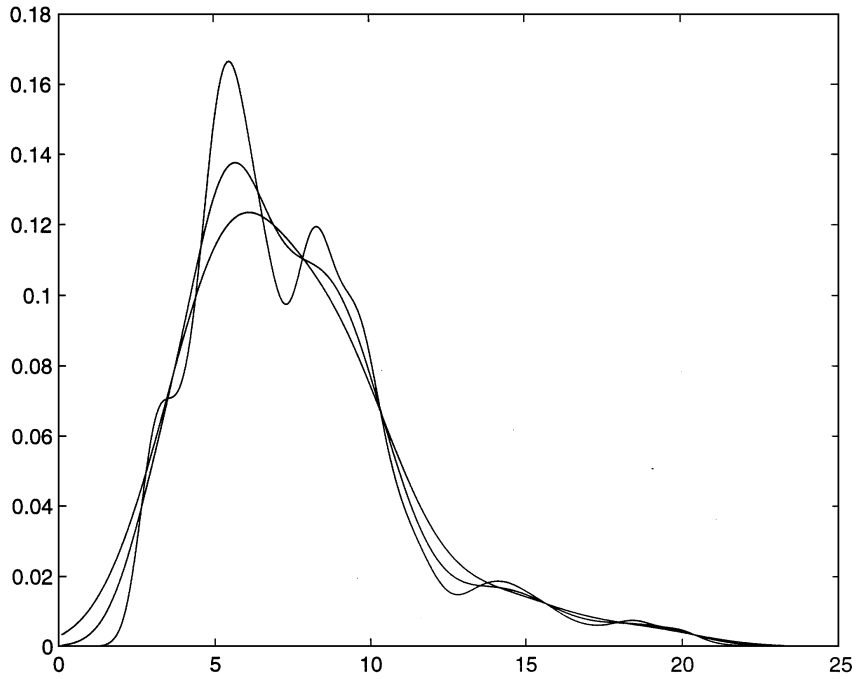


FIGURE 5. Federal-funds stationary density estimates obtained using a Gaussian kernel with bandwidths 0.5, 1, and 1.5.

of fit of the model for a given choice of the bandwidth. The amount of information contained in these omitted observations is much less when the data are positively serially dependent than when the data are i.i.d. Pritsker (1996) discusses this issue in depth for the case of Ornstein–Uhlenbeck processes and documents that the optimal bandwidth choice (minimizing mean integrated squared error) is very sensitive to the level of persistence.[12]

In view of these difficulties in bandwidth selection, we consider a range of bandwidths and calculate the test statistic for each. First, we use the test graph method of Silverman (1978) and obtain a bandwidth of 0.5.[13] We use this bandwidth as a starting point and also use bandwidths of 1 and 1.5 to portray a range of degrees of smoothing. The bandwidth of 1 is the first bandwidth larger than 0.5 (in increments of 0.1) that results in a unimodal density estimate. Finally, motivated by the discussion below, we look at an even larger bandwidth of 1.5 as well. The density estimates for federal-funds data using bandwidths of 0.5, 1, and 1.5 are depicted in Figure 5.

As a check that these bandwidths are in a sensible range, we consider choosing a bandwidth to minimize the average integrated squared error of the density estimate, with the average taken across bootstrap samples. For each of our bootstrap samples, we calculate the integrated squared difference between our kernel density estimate,

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

TABLE 6. Bandwidth choices that minimize average integrated squared errors across bootstrap samples

	γ			
	1	2	3	4
High persistence	1.4	2.1	1.7	1.5
Low persistence	0.8	1.0	0.7	0.4

TABLE 7. Goodness-of-fit test statistics ($\tilde{M}$) for federal-funds data using bandwidths of 0.5, 1, and 1.5

	γ			
Bandwidth	1	2	3	4
0.5	64.13	63.78	69.54	101.59
1	56.15	46.50	39.23	64.41
1.5	123.72	103.97	80.97	90.51

$\hat{\pi}(u; h)$, and the true density $\pi(u)$. Letting $\Upsilon_k(h)$ denote this integrated squared error for sample k,

$$\Upsilon_k(h) = \int [\hat{\pi}(u; h_T) - \pi(u)]^2 \, du.$$

Then, we average $\Upsilon_k(h)$ across our 500 bootstrap samples to get the average integrated squared error for bandwidth h. The values of h that minimize (over a coarse grid of bandwidths) this average integrated squared error are given by Table 6. The impact of data persistence on an optimal bandwidth choice using this criterion is readily apparent. Bandwidth choices for all variance elasticities are smaller for the lower-persistence data than for the high-persistence specification. The bandwidths in Table 6 give us reason to believe that our bandwidth choices from 0.5 to 1.5 give us degrees of smoothing in the right range. For the high-persistence simulations, bandwidth choices of 0.5 and 1 both apparently lead to undersmoothing as required by the asymptotic distribution theory. For the low-persistence simulations, a bandwidth choice of 1 no longer appears to result in undersmoothing.

Results from the goodness-of-fit test applied to federal-funds data are presented in Table 7 for each bandwidth choice. As described previously, the test statistic $\tilde{M}$ has a limiting distribution that is standard normal. The clear inference to be drawn using this limiting distribution is that the model is woefully misspecified. We now investigate the magnitude of bootstrap corrections for the distribution of test statistic $\tilde{M}$ to see if these corrections can explain the apparently poor fit of the model.

Tables 8 and 9 present the results of our bootstrap evaluation of the test's finite sample distribution. Five hundred bootstrap samples of length 7000 were

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

TABLE 8. Percentiles of bootstrap distributions of $\tilde{M}$ for high-persistence data (500 simulations are used for each specification)

	γ			
Percentile	1	2	3	4
	(A) Bandwidth = 0.5			
1	2.33	5.91	5.94	2.06
5	6.71	13.02	10.27	4.38
25	16.64	35.47	24.48	13.88
50	31.28	64.29	47.05	33.96
75	55.66	111.87	86.07	82.23
95	123.59	225.82	174.63	345.05
99	170.92	433.02	292.11	1,386.39
	(B) Bandwidth = 1			
1	7.36	10.60	7.63	6.42
5	11.75	19.48	17.14	11.80
25	29.92	53.96	40.15	25.27
50	50.88	102.67	74.46	64.56
75	86.08	198.27	180.02	168.01
95	178.14	481.09	501.29	1,019.76
99	271.55	1,013.94	1,202.72	2,833.75
	(C) Bandwidth = 1.5			
1	16.57	20.94	31.27	32.86
5	29.72	44.30	41.40	45.84
25	71.39	112.95	84.63	85.51
50	127.74	230.18	172.24	139.99
75	226.46	466.16	352.56	324.54
95	447.79	1,202.61	1,025.95	1,523.52
99	696.58	1,867.39	2,416.21	4,764.15

generated for each variance elasticity, and for low- and high-persistence data sets. Parametric and nonparametric densities were estimated as described above, and then normalized goodness-of-fit statistics $\tilde{M}$ were calculated for each bootstrap sample. Tables 8 and 9 present percentiles of the bootstrap distribution for these test statistics. These bootstrap distributions are very different from the limiting standard normal distribution.[14]

If we conduct inference using these bootstrap distributions in place of the limiting standard normal distribution, we are led to very different conclusions. For the high-persistence data, the values of the observed federal-funds data statistics in Table 6 are never above the 95th percentile of the corresponding bootstrap distribution. For the bandwidth of 1, the test statistics for federal-funds data are below the median of the high-persistence bootstrap distribution for three of the variance elasticities and just above it for the fourth. For the largest bandwidth of 1.5, the federal-funds test statistics are below the 75th percentile for two elasticities and below the median for the other two. For the low-persistence specification, inference

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

TABLE 9. Percentiles of bootstrap distributions of $\tilde{M}$ for low-persistence data (500 simulations are used for each specification)

	γ			
Percentile	1	2	3	4
	(A) Bandwidth = 0.5			
1	1.41	1.72	0.42	−0.47
5	2.48	4.69	1.30	0.73
25	6.57	9.83	4.96	4.15
50	12.04	15.94	9.00	8.64
75	20.88	29.88	16.46	16.07
95	40.02	58.92	32.99	34.91
99	64.31	86.83	45.19	59.64
	(B) Bandwidth = 1			
1	6.27	7.16	6.44	7.70
5	9.18	9.73	8.79	10.06
25	20.66	21.72	16.77	17.20
50	31.09	37.57	26.98	28.59
75	50.02	62.30	45.59	54.20
95	87.93	118.66	86.58	117.24
99	124.52	191.29	130.92	178.97
	(C) Bandwidth = 1.5			
1	24.44	24.09	27.64	42.53
5	32.90	34.94	36.96	47.99
25	72.68	64.08	61.56	57.53
50	108.23	106.94	87.82	74.89
75	155.97	173.15	131.23	111.12
95	246.87	298.18	224.21	219.25
99	333.82	477.35	320.81	306.39

differs by bandwidth choice. For the smallest bandwidth of 0.5, the observed values of the test statistics for federal-funds data are larger than the 95th percentile of the bootstrap distributions in all cases—the model looks incorrect. However, for both larger bandwidths, the federal-funds statistics in Table 6 are consistent with a correct model specification, lying near or below the 75th percentile of the bootstrap distribution for all elasticities.

Recall that for a bandwidth of 0.5, the kernel density estimates reveal two peaks. In contrast, only one peak is present with additional smoothing and in the density implied by our parametric estimates. Thus the evidence against the density fit of our models comes from the existence of the two peaks in the stationary density. With either a unimodal kernel density estimate or the high-persistence data, the model looks to be correctly specified. Regardless of which of these views one finds most compelling, it is clear that different conclusions will be drawn using bootstrap techniques than using first-order asymptotics alone to conduct inference.

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

10. ACCURACY OF LOCAL LINEAR DRIFT ESTIMATOR

In this section we study the accuracy of the pull measure $\mu/2\sigma^2$ when a local estimation method is used for the drift coefficient. As in local linear regression, the idea is to localize a linear parameterization in the neighborhood of a state y through the use of a kernel density. In other words, we envision a linear model of the drift:

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

where we estimate different coefficients α_0 and α_1 for each choice of y. We construct the kernel using the quartic function

$$K(x) \propto \begin{matrix} (x^2 - 1)^2 & \text{for} \quad |x| < 1, \\ 0 & \text{otherwise} \end{matrix}$$

and forming $K[(y - x)/h]$, where h is a bandwidth parameter. In effect, h dictates the domain for which the linear approximation is supposed to hold and determines the smoothness of the resulting estimator in practice.

We implement this localization again using a test-function approach, but we now localize the test function. Formally, we take the efficient test-function vector for a linear drift model and multiply its derivative by $K[(y - x)/h]$:

$$\Psi'(x; h) = \begin{bmatrix} x^{-\gamma} \\ x^{-\gamma-1}\left(\dfrac{x - y}{h}\right) \end{bmatrix} \frac{1}{h} K\left(\frac{y - x}{h}\right).$$

Because the parameterization is linear, we construct $z_t^* = [1 \;\; x_t - y]$, and our test-function estimator of the pull at y is

$$p_T(y; h) = -\frac{1}{2y^\gamma}[1 \quad 0]\left\{2\sum_{t=1}^{T} \Psi'(x_t; h) z_t^*\right\}^{-1} \sum_{t=1}^{T} (x_t)^\gamma \Psi''(x_t; h),$$

which depends on the bandwidth parameter h and the variance elasticity γ. This local estimator is formally justified and applied by Conley et al. (1997).

Our purpose in this section is to study the accuracy of the large-sample approximation of these methods. It is well known that estimators based on localization have the property that there is no "correction" for serial correlation in the first-order asymptotics when the data are weakly dependent [e.g., Robinson (1983)]. Thus, one way to construct confidence intervals is to simplify our previous method by estimating a sample covariance matrix instead of the long-run counterpart. Because ignoring serial correlation seems treacherous for short-term interest-rate data, we also compute standard errors constructed from a Bartlett spectral density estimated with a cutoff point of 60 lags.

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

TABLE 10. Standard errors for local linear pull estimates from federal-funds data

Decile[a] (%)	Bandwidth = 3		Bandwidth = 6	
	Bartlett SE[b]	i.i.d. SE[c]	Bartlett SE[b]	i.i.d. SE[c]
	(A) Variance Elasticity = 1			
10	0.0501	0.0069	0.0291	0.0040
20	0.0396	0.0054	0.0156	0.0021
30	0.0354	0.0049	0.0130	0.0017
40	0.0403	0.0056	0.0117	0.0015
50	0.0402	0.0056	0.0113	0.0015
60	0.0363	0.0051	0.0115	0.0016
70	0.0398	0.0058	0.0120	0.0017
80	0.0515	0.0082	0.0139	0.0020
90	0.0797	0.0142	0.0240	0.0035
	(B) Variance Elasticity = 4			
10	0.0411	0.0056	0.0281	0.0037
20	0.0265	0.0037	0.0215	0.0029
30	0.0256	0.0036	0.0191	0.0025
40	0.0307	0.0043	0.0162	0.0022
50	0.0284	0.0040	0.0131	0.0018
60	0.0229	0.0033	0.0104	0.0014
70	0.0247	0.0037	0.0103	0.0014
80	0.0319	0.0052	0.0108	0.0016
90	0.0468	0.0085	0.0202	0.0030

[a] Interest-rate deciles are 4.0, 5.0, 5.5, 6.1, 6.9, 8.0, 8.8, 9.8, and 11.7.
[b] Standard errors are calculated using a Bartlett spectral density estimator with 60 lags.
[c] Standard errors treat serial correlation as negligible.

Not surprisingly, the two methods of standard-error estimation yield very different results, depicted in Table 10 for variance elasticities $\gamma = 1$ and $\gamma = 4$. The Bartlett standard errors that allow for serial dependence are much larger than those that abstract from serial correlation.

Using the same simulated diffusions as before, we study the reliability of the two alternative estimators of the asymptotic covariance matrix for the local estimator of the drift. Formally, we make this comparison by following in large part the approach described in Section 8. For the generated data sets, we compute standard-error estimates for our pull measure based on the two different methods for estimating the asymptotic covariance matrix. We then form the candidate t-statistics, which by construction should have a limiting normal distribution. The pull-estimate t-statistics are centered around s_T (the series pull estimate from federal-funds data). Finally, we calculate the 68.3% critical values (± 1 standard deviation from the normal distribution) from the bootstrap distribution of absolute values of t-statistics. If initial standard-error estimates are reliable, we expect the computed critical value to be close to one.

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

TABLE 11. Bootstrap t-statistic[a] critical values at 0.683 level for local linear estimators

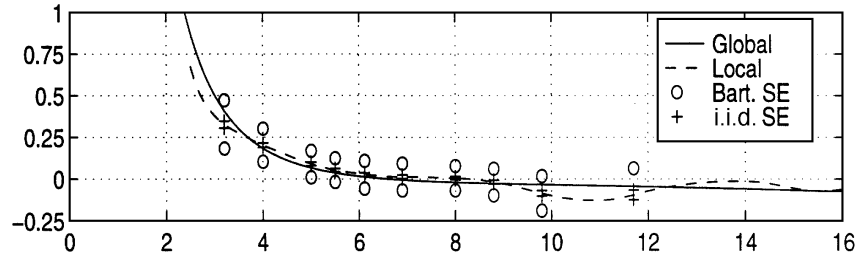
Decile[b] (%)	Low persistence				High persistence			
	Bartlett SE[c]		i.i.d. SE[d]		Bartlett SE[c]		i.i.d. SE[d]	
	$h = 3$	$h = 6$	$h = 3$	$h = 6$	$h = 3$	$h = 6$	$h = 3$	$h = 6$
	(A) Variance Elasticity = 1							
10	1.80	1.97	8.64	10.48	1.96	2.23	12.05	14.47
20	1.95	3.40	10.91	20.78	2.32	3.56	15.56	24.80
30	1.82	3.51	9.93	22.82	2.39	3.89	16.02	27.79
40	1.78	3.14	9.45	20.86	2.41	3.86	15.83	27.98
50	1.56	2.44	8.04	15.92	2.37	3.55	15.33	25.72
60	1.59	2.07	7.82	13.20	2.16	3.36	13.44	23.63
70	1.64	2.04	7.83	12.80	2.04	3.24	12.32	22.60
80	1.57	2.11	7.25	12.97	2.01	3.02	11.81	20.44
90	1.39	1.91	6.15	11.03	2.01	2.85	10.23	18.47
	(B) Variance Elasticity = 4							
10	4.40	7.57	22.70	44.94	4.05	5.49	25.62	38.41
20	1.61	1.64	6.72	8.71	2.66	3.34	16.89	23.50
30	1.47	2.25	5.53	11.19	2.39	3.14	14.47	21.50
40	1.52	3.51	4.59	16.62	2.09	3.37	11.82	22.45
50	1.40	3.58	3.93	15.45	1.81	3.23	9.44	21.01
60	1.19	1.85	2.76	6.83	1.46	2.18	6.82	13.21
70	1.20	1.06	2.38	3.47	1.37	1.63	5.75	9.46
80	1.17	1.26	1.95	3.44	1.29	1.50	4.60	8.19
90	1.05	1.21	1.41	2.56	1.19	1.31	3.39	6.00

[a] Under the asymptotic distribution, this t-statistic critical value is equal to one.
[b] Interest rate deciles are 4.0, 5.0, 5.5, 6.1, 6.9, 8.0, 8.8, 9.8, and 11.7.
[c] Critical values are bootstrap t-statistic estimates that utilize Bartlett spectral density estimators with 60 lags.
[d] Critical values are bootstrap t-statistic estimates that treat serial correlation as negligible.

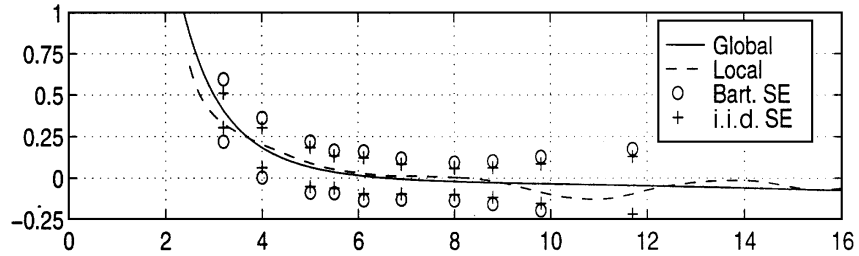
We report results in Table 11 for local linear pull estimates. We examine standard errors for variance elasticities $\gamma = 1$ and $\gamma = 4$. As before, we simulate both a (relatively) high-persistence process and a low-persistence process for each variance elasticity. The critical values resulting from using a Bartlett covariance matrix estimator are labeled Bartlett SE, and ones that abstract from serial correlation are labeled i.i.d. SE. For each covariance estimator, we report two bandwidths so that the impact of different bandwidth choices can be investigated.

The bootstrap t-statistic distributions are spread out much more than the standard normal asymptotic distribution. This is especially pronounced for those using i.i.d. covariance matrix estimators. For a variance elasticity of one, the critical value is between 6 and 22 for the low-persistence case and between 10 and 28 for the high-persistence case, depending on bandwidth and decile of interest. These critical values become larger as the data's persistence increases and as the bandwidth increases from 3 to 6. However, even when we attempt to correct for serial

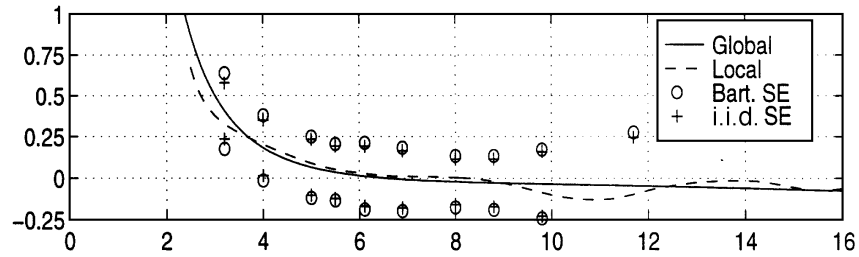
https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press



A) Pull Estimates for $\gamma = 1$



B) Low-Persistence Specification



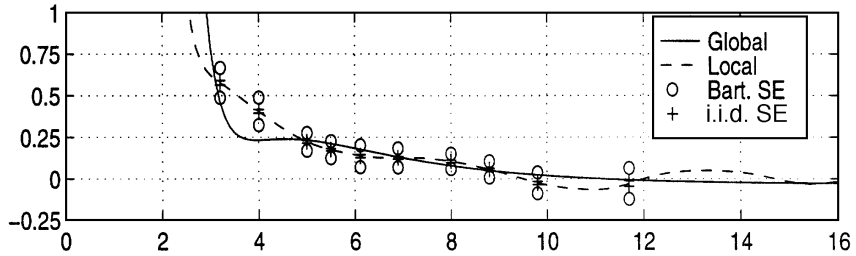
C) High-Persistence Specification

FIGURE 6. Local linear pull estimates and confidence bands for $\gamma = 1$. All panels contain local linear and parametric point estimates for federal-funds data. Panel A contains 95.4% confidence bands estimated from federal-funds data. Panels B and C present bootstrap 95.4% confidence bands for low- and high-persistence specifications, respectively. In all panels, standard error bands labeled Bart. SE are obtained using a Bartlett spectral density estimator with 60 lags and those labeled i.i.d. SE impose zero serial correlation.

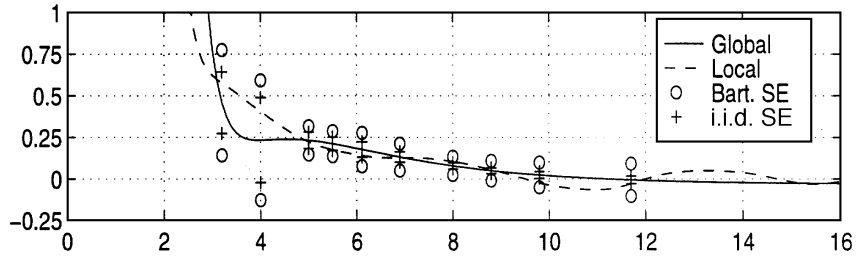
correlation via a Bartlett estimator of the long-run covariance matrix, the critical values still exceed one.

The fact that these bootstrap t-statistic distribution critical values are so large implies that bootstrap confidence intervals will be much wider than those computed from the limiting standard normal distribution. In Figures 6 and 7, we compare the confidence intervals reported in Conley et al. (in press), reproduced in Panels A, to the confidence intervals obtained using bootstrapping. In these figures, we report the 95.4% confidence intervals (corresponding to ± 2 standard deviations of the

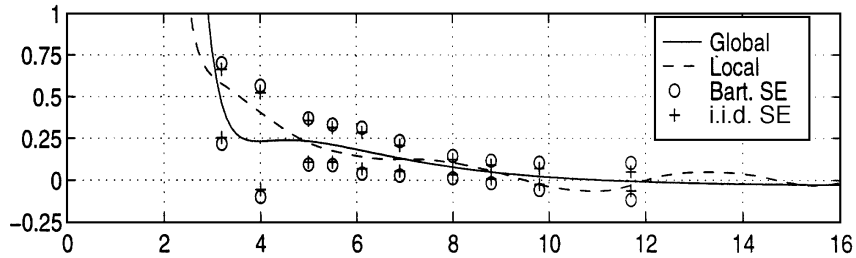
https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press



A) Pull Estimates for $\gamma = 4$



B) Low-Persistence Specification



C) High-Persistence Specification

FIGURE 7. Local linear pull estimates and confidence bands for $\gamma = 4$. All panels contain local linear and parametric point estimates for federal-funds data. Panel A contains 95.4% confidence bands estimated from federal-funds data. Panels B and C present bootstrap 95.4% confidence bands for low- and high-persistence specifications, respectively. In all panels, standard error bands labeled Bart. SE are obtained using a Bartlett spectral density estimator with 60 lags and those labeled i.i.d. SE impose zero serial correlation.

normal distribution). In Panels B and C, we report the bootstrap results from the low- and high-persistence calibrations, respectively. In contrast to the top panel, the confidence intervals are centered around our original (global) estimator of the diffusion pull. This centering is used because it corresponds to processes that were simulated.

As expected, bootstrap corrections to the confidence intervals are substantial, especially for the corrections based on the i.i.d. standard errors. Notice that there is a large discrepancy between the time series and i.i.d. standard errors as

https://doi.org/10.1017/S1365100597003015 Published online by Cambridge University Press

depicted in Panels A. The discrepancy in the corresponding bootstrap confidence intervals is considerably smaller in Panels B and C. This is to be expected because the simulations contain an alternative mechanism for making serial correlation corrections.[15]

Panel A of Figure 7 suggests that the pull estimates using our global model of the drift may understate the pull when $\gamma = 4$ and interest rates are between 3 and 5. However, once we make the bootstrap adjustments, the confidence intervals centered around our global estimates of the pull contain the local estimates. Moreover, in our simulations of the global model, the t-statistic empirical distribution is translated to the right of zero in these cases. Thus, we observed a tendency of the local linear pull estimator to exceed the pull of the process being simulated. Given the bootstrap corrections and the observed distortions in the local estimator, the global pull estimates for variance elasticity 4 in fact may not understate the pull. In summary, the observed differences between the local and global methods for estimating the pull of the diffusion apparently can be attributed to the inaccuracy of the measurements.

11. CONCLUSION

We have explored bootstrap refinements to the estimation and inference methods of Ait-Sahalia (1996) and Conley et al. (in press). We find that the first-order asymptotics can be very misleading and that bootstrap refinements can be quite important. We support this finding using two different methods to calibrate the persistence (say κ) of the candidate diffusion processes. We also show that the persistence calibration can be used to obtain lower-bound estimates for the standard errors with a faster rate of convergence than the standard errors constructed from frequency-zero, spectral density estimates. These lower-bound estimates provide useful hints as to when the bootstrap adjustments to the confidence intervals will be large.

A scalar diffusion is not a good model of federal-funds interest rates. The discrepancy between our two estimates of κ provides informal evidence that the scalar diffusion model is misspecified. Other forms of evidence against a scalar diffusion model with a constant volatility specification for short-term Treasury bill data are given by Gallant and Tauchen (1996) and Anderson and Lund (1997). Although we explore the sensitivity of our results to changes in the persistence of the diffusion, none of our simulated processes are designed to accommodate model misspecification. Hence, a reader may object to our use of these methods as bootstrapping methods and instead prefer to view this paper as simply a Monte Carlo study. Although this complaint is valid, we are skeptical that altering the simulation environment in realistic ways will result in smaller bootstrap corrections.

Gallant and Tauchen (1996) and Anderson and Lund (1997) emphasize the need for an additional volatility factor. Such a factor is permitted in the subordinated diffusion models of Conley et al. (in press), albeit in a different way. For these

models, the estimation methods described in the paper are applicable. Subordinated diffusion models may be viewed as two-factor stochastic volatility models in which the hidden factor shifts both the local mean and the local variance of the process. Equivalently, a subordinated diffusion may be interpreted as emerging because the scalar diffusion is observed at random points in time. This sampling can be modeled with a stationary increment process that is temporally dependent. Such a sampling scheme could explain differences in our κ estimates, provided that there is a substantial amount of positive persistence in the sampling interval process. Diffusion counterparts to the process suggested by Robinson and Zaffaroni (1997) in which the sampling (and hence stochastic volatility) process has long memory may be promising.

NOTES

1. See Robinson and Velasco (1996) for a discussion of a variety and estimation methods and their properties.

2. Blocking methods have been suggested Kunsch (1989) and Liu and Singh (1992), among others.

3. Cobb et al. (1992) consider polynomial drift specifications. Following Conley et al. (in press), we include a negative power because we consider processes on the half line.

4. A more precise statement of this lack of identification is that one can construct a drift for any prespecified logarithmic derivative of the stationary density and any variance elasticity. Similarly, in the literature on calibrating dynamic, stochastic general equilibrium models, only a subset of the parameters can be inferred from the steady-state relations.

5. We thank the Federal Reserve Bank of Chicago for providing the data.

6. The method of Chen et al. (1997) also can be used to make inferences about the variance elasticity.

7. The reported estimates were obtained by centering the cardinal B-splines at integer values of the interest rates, setting D_T to be the identity matrix and setting λ_T to penalize the squared second differences in the coefficients at about 5% of that of the unrestricted ($\lambda_T = 0$) estimates. Strictly speaking, the identity matrix for D_T is not covered in the analysis of Chen et al. (1997). Instead they establish approximation results based on second-difference penalties that increase with the level of the state. We also computed estimates of eigenfunctions and of ρ based on second-difference weighting schemes that are formally covered by the approximation results of Chen et al. (1997) and by λ_T targets designed to impose a common second-difference constraint for the alternative specifications of the variance elasticity γ. The resulting estimates of ρ were mostly insensitive to these changes except for the $\gamma = 4$ case. In this high volatility elasticity case, our estimates of ρ ranged from 117 to 128.

8. For variance elasticity of three, strictly speaking the entries of Table 3 are not quite equal to the ratio of the corresponding entries in Tables 2 and 1. This discrepancy is due to rounding error. The entries in Table 3 are the exact numbers used in our simulations. As mentioned in note 7, we also used alternative second-difference weighting schemes. Using the least squares estimates for δ, the κ estimates ranged from 1.98e-2 to 1.99e-2 for $\gamma = 1$, from 2.11e-3 to 2.20 e-3 for $\gamma = 2$, from 7.07e-4 to 8.32e-4 for $\gamma = 3$, and from 2.26e-4 to 2.71e-4. Thus we found more sensitivity in our κ estimates for higher variance elasticities.

9. The simulation methods have been used fruitfully in the development of an alternative class of simulation-based estimators of diffusion processes. See Gourieroux and Monfort (1996) and Tauchen (1997) for surveys.

10. In estimating the parameter α, we imposed the sign restrictions necessary for a stationary parameterization. Occasionally, these sign restrictions were binding and the numbers of these occurrences for each sample are given by the following table:

	No. of draws			
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
High persistence	32	21	24	4
Low persistence	19	21	0	0

11. The specification test of Ait-Sahalia (1996) and Pritsker (1996) is slightly different from the one we use. First, Ait-Sahalia (1996) estimates $\hat{M}$ using a sample average instead of numerical integration to approximate the integral in (7). We use numerical integration to reduce computation time. Second, Ait-Sahalia uses root-T consistent parameter estimates that minimize $\hat{M}$ whereas ours uses root-T consistent test-function estimators. Our test, although asymptotically equivalent, may have different finite sample properties.

12. Another problem with cross-validation applied to our data set is the presence of small-scale "lumps" in federal-funds interest-rates data—rates are very often whole numbers and fractions in eighths. Cross-validation applied to data with small-scale effects is known to produce estimates of optimal bandwidths that are implausibly small. See Silverman (1986) for a discussion.

13. Silverman (1978) shows that (under regularity conditions on the kernel and true density) if the bandwidth is chosen to ensure that $\sup|\hat{\pi} - \pi|$ converges to zero as fast as possible, then

$$\frac{\sup|\hat{\pi}'' - E\hat{\pi}''|}{\sup|E\hat{\pi}''|}$$

converges to a constant depending on kernel choice (approximately 0.4 for a Gaussian kernel). The test-graph method consists of choosing a bandwidth such that the maximum noise in $\hat{\pi}''$ (corresponding to the numerator of this fraction) is about 0.4 of the maximum trend in the curve $\hat{\pi}''$ (the denominator). See also Silverman (1986).

14. Because we reestimate the parametric model for each Monte Carlo sample, the smoothness of the implied density is allowed to adapt to each sample, albeit in a constrained way. On the other hand, we are holding fixed the bandwidths across Monte Carlo samples. This asymmetry may inflate our bootstrap corrections.

15. There are sometimes quantitatively important differences in the bootstrap confidence intervals depending upon whether the i.i.d. or Bartlett standard errors are used. The standard-errors estimators used in the i.i.d. confidence intervals have a faster rate of convergence than the Bartlett counterparts. However, they are likely to suffer from a greater downward bias. This leads to question of which bootstrap intervals we expect to be more reliable in practice. Although we do not have any Monte Carlo evidence to answer this question, we did find the following. When we used estimates of the long-run covariance matrix that are analogous to Λ, the computed confidence intervals turned out to be closer to the bootstrap i.i.d. confidence intervals than the Bartlett counterparts.

REFERENCES

Ait-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9, 385–426.

Anderson, T.G. & J. Lund (1997) Estimating continuous time stochastic volatility models for short term interest rate data. *Journal of Econometrics* 77, 343–378.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves & R.L. Taylor (1991) Bootstrapping unstable first-order autoregressive processes. *Annals of Statistics* 2, 1098–1101.

Bertola, G. & R. Caballero (1992) Target zones and realignments. *American Economic Review* 82, 520–536.

Bhatacharia, R.N. (1982) On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete* 60, 185–201.

Bickel, P.J. & M. Rosenblatt (1973) On some global measures of the deviations of density function estimates. *Annals of Statistics* 1, 1071–1095.

Chan, K.C., G.A. Karolyi, F.A. Longstaff & A.B. Sanders (1992) An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* 48, 1209–1227.

Chen, X., L.P. Hansen & J.A. Scheinkman (1997) Estimating Eigenfunctions for Sampled Diffusion Processes. Manuscript. University of Chicago.

Christiano, L. & M. Eichenbaum (1992) Current real business cycle theory and aggregate labor market fluctuation. *American Economic Review* 82, 430–450.

Cobb, L., P. Koppstein & N.Y. Chen (1983) Estimation of moment recursion relations for multimodal distributions of the exponential family. *Journal of the American Statistical Association* 78, 124–130.

Conley, T.G., L.P. Hansen, E.G.J. Luttmer & J.A. Scheinkman (in press) Short-term interest rates as subordinated diffusions. *Review of Financial Studies.*

Datta, S. & W. P. McCormick (1995) Some continuous Edgeworth expansions for Markov chains with applications to bootstrap. *Journal of Multivariate Analysis* 52, 83–106.

Gallant, A.R. & G. Tauchen (1996) Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions. Working paper, University of North Carolina at Chapel Hill.

Gotze, F. & H. R. Kunsch (1996) Second-order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics* 24, 1914–1933.

Gourieroux, C. & A. Monfort (1996) Simulation-based econometric methods. In *Core Lectures*, London: Oxford University Press.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion.* New York: Springer–Verlag.

Hamilton, J.D. (1996) The daily market for federal funds. *Journal of Political Economy* 104, 26–56.

Hansen, L.P. & 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 & N. Touzi (in press) Spectral methods for identifying scalar diffusions. *Journal of Econometrics.*

Kloeden, P.E. & E. Platen (1995) *Numerical Solution of Stochastic Differential Equations.* New York: Springer–Verlag.

Kunsch, H.R. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.

Liu, R.Y. & K. Singh (1992) Moving block jackknife and bootstrap capture weak dependence. In R. Lepage and L. Billard (eds.), *Exploring the Limits of the Bootstrap*, pp. 225–258. New York: Wiley.

Pritsker, M. (1996) Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models. Working paper, Federal Reserve Board of Governors.

Robinson, P.M. (1983) Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–207.

Robinson, P.M. & C. Velasco (1996) Autocorrelation-Robust Inference. Discussion paper EM/96/316, London School of Economics.

Robinson, P.M. & P. Zaffaroni (1997) Nonlinear Time Series with Long Memory: A Model for Stochastic Volatility. Discussion paper EM/97/320, London School of Economics.

Silverman, B.W. (1978) Choosing the window width when estimating a density. *Biometrika* 65, 1–11.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Tauchen, G. (1997) New minimum chi-squared methods in empirical finance. In D. Kreps and K. Wallace (eds.), *Advances in Econometrics: Seventh World Congress.* Cambridge, UK: Cambridge University Press.