Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Financial Economics 74 (2004) 31–66

JOURNAL OF Financial ECONOMICS

www.elsevier.com/locate/econbase

Conditional estimation of diffusion processes[☆]

Minqiang Li, Neil D. Pearson*, Allen M. Poteshman

University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA

Received 4 April 2001; accepted 18 June 2003

Abstract

There are a number of circumstances in finance in which it is useful to estimate diffusion processes conditional on some event. In this paper, we develop the theoretical and numerical tools necessary to perform conditional estimation of diffusion processes within a generalized method of moments framework. We illustrate our method by estimating a univariate diffusion process for a standard time-series of interest rate data conditioned to remain between lower and upper boundaries. A test statistic fails to reject by a wide margin the linearity of the conditionally estimated drift coefficient.

JEL classification: C13; C22; G10; G12

Keywords: Estimation; Diffusion process; Interest rates; Nonlinearity

---

1. Introduction

There are a number of situations in finance in which it is desirable to estimate a data-generating process conditional on the occurrence of some event. One context

---

[☆] We thank Yacine Aït-Sahalia for making available the interest rate data used in this paper, Reza Mahani for performing some of the preliminary computations, and William Goetzmann, Matthew Pritsker, Andrea Roncoroni, Bill Schwert (the editor), and participants in seminars at the University of Texas at Austin, Pennsylvania State University, the 2001 WFA meetings, MathWeek 2001, and the 2003 EFMA meetings for helpful comments. We also especially thank an anonymous referee for comments and suggestions that greatly improved the paper. Some of the computations were performed on computer workstations provided by the Intel Corporation under its Technology for Education 2000 program. Minqiang Li was supported by the Corzine Assistantship of the Office for Futures and Options Research of the University of Illinois at Urbana-Champaign.

*Corresponding author.

E-mail addresses: mli1@uiuc.edu (M. Li), pearson2@uiuc.edu (N.D. Pearson), poteshma@uiuc.edu (A.M. Poteshman).

doi:10.1016/j.jfineco.2004.03.001

where conditional estimation should be performed is situations in which only data that satisfy some criteria are accessible to researchers even though the underlying process is capable of generating data that violate the criteria. An example of such a situation is the estimation of stock return processes when data are only available on stock markets that have survived up to the time that the research is conducted. In this case, the assumed stock return model should be estimated conditional on the stock price not having reached some lower boundary. Brown et al. (1995) analyze this problem under the assumption that the unconditioned stock price process obeys geometric Brownian motion.

A second context where performing conditional estimation is appropriate is situations in which any finite data sample will misrepresent in some specifically identifiable and consequential way the underlying population from which it is drawn. Chapman and Pearson (2000) demonstrate that the estimation of short-rate models from time series of interest rate data is an example of this type of situation. They perform a Monte Carlo study in which artificial time series of interest rate data are generated from a diffusion process with linear drift, and then use these artificially generated time series to estimate diffusion models with flexibly specified drift coefficients. One of their findings is that the estimated drift function is negatively biased near the maximum value of the time series of interest rates from which it is estimated. Intuitively, the source of this bias is that at the maximum value of a time series, the next step in the process is necessarily downward even though the underlying data-generating process is capable of producing a greater or lesser value at the next step. Similarly, the estimated drift coefficient is positively biased near the minimum value. A researcher who is concerned about the bias in the drift coefficient may prefer to estimate short-rate models conditional on the process remaining in between the minimum and maximum observed values to estimating these models unconditionally.

Essentially the same issue arises in other settings. Consider regressions of future stock returns on dividend yields. Dividend yields are mean-reverting and most of their variation is produced by movements of the stock price level in their denominators. At the same time, stock returns are essentially changes in stock price levels. Consequently, when the dividend yield is at its maximum value within a finite sample of data (at which point the stock price level will be low), the stock price level will subsequently tend to rise within the given sample even if dividend yields and future stock returns are actually independent of one another. The resulting spurious positive relationship between future stock returns and dividend yields is documented in bootstrap simulations in Goetzmann and Jorion (1993), which are conducted under the null hypothesis that the dividend yield and future stock returns are unrelated. A similar spurious relationship is likely to arise whenever future stock returns are regressed on conditioning variables that are ratios containing stock price levels. Another example of the same underlying issue arises when traders at financial institutions trade based on the spreads between different interest rates, e.g., the Treasury-Eurodollar spread, the swap spread, and various credit spreads. The potential profitability of these strategies is often evaluated using past data, sometimes under the name of "backtesting." Due to the conditioning discussed above, the historical data will likely suggest stronger mean reversion at high and low values of the spread than is actually true of the underlying data-generating process, which would cause backtesting to suggest greater profitability than one can reasonably expect. The methodology in this paper allows one to address these issues by estimating the underlying data-generating process from a sample conditioned to have the dividend yield or interest rate spread remain within upper and lower bounds.

This paper develops an econometric procedure for estimating diffusions with flexibly specified drift and diffusion coefficients conditional on one of a broad class of events. We proceed by building the theoretical tools necessary to apply a generalized method of moments (GMM) procedure to the estimation of a multivariate diffusion process conditioned to remain confined within an open, connected, and bounded region. We then describe numerical techniques that can be used to implement our approach within a GMM framework to estimate a flexibly specified univariate diffusion conditioned to remain between a lower and an upper boundary. Modifying the numerical techniques for other conditioning events is straightforward.[1] We also derive explicit expressions for the case of a univariate geometric Brownian motion conditioned to remain within a lower and an upper boundary.

We illustrate our method by applying it to a standard time series of interest rate data in order to estimate an unconditional process with flexibly specified drift and diffusion coefficients that is conditioned to remain between a lower and an upper boundary. We find that the conditioning decreases the estimate of the drift coefficient for low values of the interest rate and considerably increases the estimate of the drift coefficient for large values of the interest rate. These changes cause the estimated drift coefficient to be closer to linear than when the process is estimated without conditioning. Indeed, a test statistic fails to reject by a wide margin the linearity of the unconditioned drift coefficient.

The statistical justification for our procedure is provided by the conditional frequentist theory worked out most extensively by Kiefer (1975, 1976, 1977), Brownie and Kiefer (1977), and Brown (1978). The basic idea of the conditional frequentist approach is to partition the sample space (e.g., the space of short-rate sample paths) into disjoint sets and then to proceed in the usual frequentist manner conditional on the set to which the realization of the data belongs (Berger, 1986). In our illustrative estimation, we divide the space of interest rate sample paths into those which do and those which do not remain between the observed minimum and maximum, and we carry out our analysis conditional on the first of these sets of paths. Conditioning in this way differs from imposing reflecting or absorbing barriers at the observed minimum and maximum values, because our conditioning eliminates paths that reach the observed minimum or maximum values while the imposition of reflecting or absorbing barriers retains but alters these paths. While both the conditional frequentist approach that we follow and the standard Bayesian

---

[1] In an earlier version of the paper we work out the modifications to the numerical technique necessary to condition on two other events.

approach to statistics condition on the observed data, our conditioning is less severe. For example, when we apply our approach to the interest rate data, we use features of the observed data (i.e., its minimum and maximum) to condition on a large set of sample paths (i.e., all that remain within the minimum and maximum). The Bayesian approach, on the other hand, conditions on the particular sample path observed (Berger et al., 1997).

Our work is most closely related to that of Abhyankar and Basu (2001) who compute the drifts of the Ornstein-Uhlenbeck and Cox-Ingersoll-Ross (1985) "square root" processes conditional on the event that each process is less than a fixed constant, and also the drift of a Wiener process conditioned to remain between upper and lower boundaries. They find that the conditioned drifts are nonlinear, even though the original processes have linear drifts. Our paper differs from this previous work insofar as it permits flexible specifications for both the drift and the diffusion coefficients of the unconditioned process, and it allows conditioning on staying between a minimum and a maximum value for any of the flexible specifications of the unconditioned process. In addition, the techniques that we develop can be extended to other conditioning events. As noted above, our work is also related to the Brown et al. (1995) paper on survivorship bias, which assumes both that the unconditioned stock process follows geometric Brownian motion and that the analyst or econometrician observes only the process conditioned on survival. They find the relation between the drifts of the conditioned and unconditioned processes, and argue that the conditioning bias can be significant in the interpretation of such diverse phenomena as the equity premium, long-term autocorrelation studies, post-announcement drift following earnings announcements, and stock split studies.

The remainder of the paper is organized as follows. Section 2 presents expressions for the conditioned drift vector and diffusion matrix of a multivariate diffusion in terms of the probabilities that the conditioning event will be satisfied. It also shows that the probabilities of the conditioning event as well as the expected change vector of the conditioned process over a finite interval satisfy parabolic partial differential equations. Section 3 first describes numerical tools to solve the partial differential equations when an unconditioned univariate diffusion has flexibly specified drift and diffusion coefficients and the conditioning event is that the diffusion remains between a lower and an upper boundary. Section 3 also obtains explicit expressions for the case in which the unconditioned process is a geometric Brownian motion. The explicit results for the case of geometric Brownian motion are included because they provide a benchmark against which the numerical solutions can be compared. They also will be of interest to researchers studying equity markets, for which geometric Brownian motion is a reasonable model for the unconditioned process. We next illustrate our technique by applying it to interest rate data. Section 4 carries out a Monte Carlo experiment which demonstrates the spurious nonlinearities that arise when drift coefficients of a univariate diffusion are estimated from a time series of interest rate data using a generalized method of moments procedure that does not address the bias identified by Chapman and Pearson (2000). Section 5 then uses our technique to estimate a diffusion model of a standard time series of interest rate data conditional on the process staying between a lower and an upper boundary. Section 6 briefly concludes.

2. Conditioning a multivariate diffusion on the event that it remains within an open, connected, and bounded region $\mathcal{G}$

This section of the paper develops theoretical tools that can be used to estimate a multivariate diffusion process conditional on the event that it remains in an open, connected, and bounded set $\mathcal{G} \subset \Re^d$. We begin by specifying the unconditional, multivariate, time-homogeneous diffusion process in $\Re^d$ by

$$\mathrm{d}x(t) = \mu(x(t))\,\mathrm{d}t + \sigma(x(t))\,\mathrm{d}B(t), \tag{1}$$

where μ is a continuous d-vector, σ is a positive $d \times d$ matrix with elements $\sigma_{ij} \in C^1(\Re^d)$, and B is a vector of standard Brownian motions. When $d = 1$, this is a univariate diffusion process with drift and diffusion coefficients given by μ and σ^2, respectively.

Letting $\{x(u)\}_{u \in [t_1, t_2]}$ denote the path of the process between times t_1 and $t_2 > t_1$, define the event that the process remains in the region $\mathcal{G}$ from time t_1 to t_2 by

$$G(t_1, t_2) \equiv \{\{x(u)\}_{u \in [t_1, t_2]} : x(u) \in \mathcal{G}\ \forall u \in [t_1, t_2]\}. \tag{2}$$

The event that the process remains in the region $\mathcal{G}$ from t to T is then $G(t, T)$. Next, define $\pi(x, t; G(t_1, t_2))$ to be the probability of the event $G(t_1, t_2)$ given that the process is at the value x at time t:

$$\pi(x, t; G(t_1, t_2)) \equiv \mathrm{P}[G(t_1, t_2) \mid x(t) = x]. \tag{3}$$

The main theorem in Pinsky (1985, p. 366) states that the process conditioned on the event $G(t, T)$ is an inhomogeneous diffusion, with drift vector and diffusion matrix given by

$$\mu(x, t \mid G(t, T)) = \mu(x) + \frac{1}{\pi(x, t; G(t, T))}\sigma(x)\sigma(x)^\top \nabla\pi(x, t; G(t, T)) \tag{4}$$

and

$$\sigma(x, t \mid G(t, T))\sigma(x, t \mid G(t, T))^\top = \sigma(x)\sigma(x)^\top, \tag{5}$$

respectively. Note that the diffusion matrix in Eq. (5) is unaffected by the conditioning. For this reason, our main focus in this paper is on the vector of drift coefficients.

Pinsky's main theorem requires that the following two conditions be satisfied: (i) the probability $\pi(x, t; G(t, T)) \in C^2(\mathcal{G})$ as a function of x; and, (ii) $\pi(x, t; G(t, T)) = C_1\varphi_0 \exp(-\lambda_0 T) + o(\exp(-\lambda_0 T))$ as $T \to \infty$ and $\nabla\pi(x, t; G(t, T)) = C_1 \nabla\varphi_0 \exp(-\lambda_0 T) + o(\exp(-\lambda_0 T))$ as $T \to \infty$, with $o(\exp(-\lambda_0 T))$ uniform on compact subsets of $\mathcal{G}$. Pinsky (1985, p. 366) notes that these two conditions are satisfied if $(\sigma\sigma^\top)^{-1}\mu$ is a gradient function, which is always the case for a univariate diffusion.

In order to implement a GMM estimation procedure using daily data, we need to compute the expected change vector (or an accurate approximation to it) of the conditioned process over intervals of one trade date. Eq. (4) provides the vector of expected rates of change at time t of the conditional process over an infinitesimal length of time $\mathrm{d}t$. This equation could be used directly to compute the one-trade-date expected change vector in the conditioned process if the expected change vector of the conditioned process at time t is approximately linear in time over a period of one trade date. While in some (but certainly not all) contexts it may be reasonable to assume that the expected change vector in the unconditioned process is close to linear in time for a period of one trade date, we will see below that the expected change vector in the conditioned process tends to be highly nonlinear near the boundary of $\mathcal{G}$ even over a one-trade-date interval. Consequently, we now turn to the task of computing the expected change vector of the conditioned process over a finite interval. In order to compute this expected change vector, we first need to obtain the formula for the process' conditioned joint transition density and to compute the probability $\pi(x, t; G(t, T))$.

2.1. The conditioned joint transition density

Let $f(x, t, y, s)$ be the joint transition density function for the unconditioned process (1) to be at a value y at time s if it is at a value x at an earlier time t. Similarly, let $f(x, t, y, s \mid G(t, T))$ be the joint transition density function for the process (1) to be at a value y at time s if it is at a value x at an earlier time t conditional on the occurrence of the event $G(t, T)$ (here $t \leqslant s \leqslant T$). For $\delta t > 0$ and $x, y \in \mathcal{G}$, we follow Karlin and Taylor (1975, p. 358; 1981, Chapter 15) and write

$$f(x, t, y, t+\delta t \mid G(t, T))\,\mathrm{d}y = \mathrm{P}[x(t+\delta t) \in \mathrm{d}y \mid x(t) = x, G(t, T)], \tag{6}$$

where $\mathrm{d}y = \mathrm{d}y_1 \mathrm{d}y_2 \cdots \mathrm{d}y_d$ and the notation $x(t+\delta t) \in \mathrm{d}y$ denotes $y_k \leqslant x_k(t+\delta t) \leqslant y_k + \mathrm{d}y_k$ for all k $(1 \leqslant k \leqslant d)$. An application of Bayes' Rule to the right-hand side of (6) gives

$$\begin{aligned} &f(x, t, y, t+\delta t \mid G(t, T))\,\mathrm{d}y \\ &\quad = \frac{\mathrm{P}[x(t+\delta t) \in \mathrm{d}y \mid x(t) = x] \cdot \mathrm{P}[G(t, T) \mid x(t) = x, x(t+\delta t) = y]}{\mathrm{P}[G(t, T) \mid x(t) = x]}, \end{aligned} \tag{7}$$

or

$$\begin{aligned} &f(x, t, y, t+\delta t \mid G(t, T))\,\mathrm{d}y \\ &\quad = \frac{f(x, t, y, t+\delta t)\mathrm{P}[G(t, T) \mid x(t) = x, x(t+\delta t) = y]}{\mathrm{P}[G(t, T) \mid x(t) = x]}\,\mathrm{d}y. \end{aligned} \tag{8}$$

When $T - t \geqslant \delta t$, we have

$$\begin{aligned}
&\mathrm{P}[G(t,T) \,|\, x(t) = x, x(t+\delta t) = y] \\
&\quad = \mathrm{P}[G(t,t+\delta t) \cap G(t+\delta t, T) \,|\, x(t) = x, x(t+\delta t) = y] \\
&\quad = \mathrm{P}[G(t,t+\delta t) \,|\, x(t) = x, x(t+\delta t) = y] \\
&\qquad \times \mathrm{P}[G(t+\delta t, T) \,|\, x(t) = x, x(t+\delta t) = y, G(t,t+\delta t)] \\
&\quad = \mathrm{P}[G(t,t+\delta t) \,|\, x(t) = x, x(t+\delta t) = y]\mathrm{P}[G(t+\delta t, T) \,|\, x(t+\delta t) = y], \qquad (9)
\end{aligned}$$

where the last step uses the Markov property of diffusion processes.

Now define a new probability $\bar{\pi}$ as

$$\bar{\pi}(x, t, y, t+\delta t; G(t, t+\delta t)) = \mathrm{P}[G(t, t+\delta t) \,|\, x(t) = x, x(t+\delta t) = y]. \qquad (10)$$

The second term of the numerator of (8) becomes

$$\begin{aligned}
&\mathrm{P}[G(t,T)|x(t) = x, x(t+\delta t) = y] \\
&\quad = \bar{\pi}(x, t, y, t+\delta t; G(t, t+\delta t))\pi(y, t+\delta t; G(t+\delta t, T)). \qquad (11)
\end{aligned}$$

Recognizing that the denominator of (8) is just $\pi(x, t; G(t, T))$, and substituting this and (11) into (8) yields

$$\begin{aligned}
&f(x, t, y, t+\delta t \,|\, G(t,T))\,\mathrm{d}y \\
&\quad = \frac{\bar{\pi}(x, t, y, t+\delta t; G(t, t+\delta t))\pi(y, t+\delta t; G(t+\delta t, T))f(x, t, y, t+\delta t)}{\pi(x, t; G(t, T))}\,\mathrm{d}y. \qquad (12)
\end{aligned}$$

2.2. *The probability* $\pi(x, t; G(t, T))$

To use (12) we need one fact about the probability $\pi(x, t; G(t, T))$. The Law of Total Probability can be used to write $\pi(x, t; G(t, T))$ as the expected value of a function of the process (1) conditional on an initial value. In particular, if 1_G is the indicator function which takes the value of one if the event G occurs and zero otherwise, then $\pi(x, t; G(t, T)) = \mathrm{E}[1_{G(t,T)} \,|\, x(t) = x]$. Such expectations satisfy the Kolmogorov backward differential equation (see Section 15.5 of Karlin and Taylor, 1981). That is, in the interval $[0, T)$, the probability $\pi(x, t; G(t, T))$ obeys

$$\begin{aligned}
&\frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d}(\sigma(x)\sigma(x)^{\top})_{jk}\frac{\partial^2\pi(x, t; G(t, T))}{\partial x_j \partial x_k} + \mu(x)\cdot\nabla\pi(x, t; G(t, T)) \\
&\quad + \frac{\partial\pi(x, t; G(t, T))}{\partial t} = 0, \qquad (13)
\end{aligned}$$

along with the boundary conditions

$$\pi(\bar{x}, t; G(t, T)) = 0 \quad \text{for } \forall t \in [0, T] \text{ if } \bar{x} \in \partial\mathcal{G} \qquad (14)$$

and

$$\pi(x, T; G(t, T)) = 1 \quad \text{for } \forall x \in \mathcal{G}. \qquad (15)$$

To see that the probability $\pi(x, t; G(t, T))$ satisfies (13), note that the characterization of the probability as an expected value implies that the process $\{\pi(x, t; G(t, T))\}$ is a martingale. Using Itô's Lemma, the process obeys the stochastic differential equation

$$\begin{aligned} &\mathrm{d}\pi(x(t), t; G(t, T)) \\ &\quad = \left[\frac{1}{2} \sum_{j=1}^{d} \sum_{k=1}^{d} (\sigma(x(t))\sigma(x(t))^{\top})_{jk} \frac{\partial^2 \pi(x, t; G(t, T))}{\partial x_j \partial x_k} \bigg|_{x=x(t)} \right. \\ &\qquad \left. + \mu(x(t)) \cdot \nabla \pi(x, t; G(t, T))|_{x=x(t)} + \frac{\partial \pi(x, t; G(t, T))}{\partial t} \bigg|_{x=x(t)} \right] \mathrm{d}t \\ &\qquad + \sum_{j=1}^{d} \sum_{k=1}^{d} \sigma_{jk}(x(t)) \frac{\partial \pi(x, t; G(t, T))}{\partial x_j} \bigg|_{x=x(t)} \mathrm{d}B_k(t). \end{aligned} \tag{16}$$

If (16) is a martingale then its drift must be zero, which implies that the function π satisfies the partial differential equation (13).

2.3. Conditional expected change vector over a finite time interval

We now use the results above to develop an approach for computing the vector of conditional expected changes over a fixed finite time interval $[t, t + \delta t]$. This quantity can be written in terms of the joint conditional transition density. Let $m(x, t, t + \delta t \mid G(t, T))$ be the expected change vector in the process (1) over a finite time step δt when the process is currently at a level $x(t) = x$ conditional on the event $G(t, T)$. Then, $m(x, t, t + \delta t \mid G(t, T))$ is given by

$$m(x, t, t + \delta t \mid G(t, T)) = \int_{\mathscr{G}} (y - x) f(x, t, y, t + \delta t \mid G(t, T))\, \mathrm{d}y. \tag{17}$$

Substituting Eq. (12) into this expression gives

$$\begin{aligned} &m(x, t, t + \delta t \mid G(t, T)) \\ &\quad = \int_{\mathscr{G}} \frac{\bar{\pi}(x, t, y, t + \delta t; G(t, t + \delta t)) \pi(y, t + \delta t; G(t + \delta t, T)) f(x, t, y, t + \delta t)\, y}{\pi(x, t; G(t, T))} \\ &\qquad \times \mathrm{d}y - x \\ &\quad = \frac{1}{\pi(x, t; G(t, T))} \mathrm{E}[\bar{\pi}(x, t, y, t + \delta t; G(t, t + \delta t)) \\ &\qquad \times \pi(y, t + \delta t; G(t + \delta t, T))\, y | x(t) = x] - x \\ &\quad = \frac{v(x, t, t + \delta t \,| G(t, T))}{\pi(x, t; G(t, T))} - x, \end{aligned} \tag{18}$$

where the expectation E is taken over y and we define the vector v to be

$$\begin{aligned} v(x, s, t + \delta t \mid G(s, T)) \equiv\, &\mathrm{E}[\bar{\pi}(x, s, y, t + \delta t; G(s, t + \delta t)) \\ &\times \pi(y, t + \delta t; G(t + \delta t, T)) y \mid x(s) = x]. \end{aligned} \tag{19}$$

Notice here $t \leqslant s \leqslant t+\delta t \leqslant T$. Standard arguments on the relations between expectations of functions of diffusion processes and partial differential equations (see, e.g., Section 15.5 of Karlin and Taylor (1981) or Section 2.2 above) imply that on the fixed interval $[t, t+\delta t)$, each component v_ℓ of the vector v satisfies the backward equation

$$\left[\frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d}(\sigma(x)\sigma(x)^\top)_{jk}\frac{\partial^2}{\partial x_j \partial x_k} + \mu(x)\cdot\nabla + \frac{\partial}{\partial s}\right] v_\ell(x, s, t+\delta t | G(s,T)) = 0, \tag{20}$$

together with spatial boundary conditions

$$\begin{aligned} &v_\ell(\bar{x}, s, t+\delta t \,|\, G(s,T)) \\ &\quad = \mathrm{E}[\bar{\pi}(\bar{x}, s, y, t+\delta t; G(s, t+\delta t))\pi(y, t+\delta t; G(t+\delta t, T))y_\ell \,|\, x(s) = \bar{x}] \\ &\quad = \mathrm{E}[0 \cdot \pi(y, t+\delta t; G(t+\delta t, T))\, y_\ell \,|\, x(s) = \bar{x}] \\ &\quad = 0 \quad \text{for } \forall s \in [t, t+\delta t] \text{ if } \bar{x} \in \partial\mathcal{G}, \end{aligned} \tag{21}$$

and terminal boundary condition at time $t+\delta t$ given by

$$\begin{aligned} &v_\ell(x, t+\delta t, t+\delta t \,|\, G(t+\delta t, T)) \\ &\quad = \mathrm{E}[\bar{\pi}(x, t+\delta t, y, t+\delta t; G(t+\delta t, t+\delta t)) \\ &\qquad \times \pi(y, t+\delta t; G(t+\delta t, T))y_\ell \,|\, x(t+\delta t) = x] \\ &\quad = \mathrm{E}[\delta_{x,y} \cdot \pi(y, t+\delta t; G(t+\delta t, T))\, y_\ell \,|\, x(t+\delta t) = x] \\ &\quad = \pi(x, t+\delta t; G(t+\delta t, T))x_\ell \quad \text{for } x \in \mathcal{G}, \end{aligned} \tag{22}$$

where the expectation is taken over a Dirac measure since when $t = s$ and $x, y \in \mathcal{G}, f(x,t,y,s) = \delta(x-y)$.

The above suggests that one can compute $m(x, t, t+\delta t \,|\, G(t,T))$ by (perhaps numerically) solving Eq. (20) over the interval $[t, t+\delta t]$ to obtain $v(x, t, t+\delta t \,|\, G(t,T))$, solving (13) over the interval $[t, T]$ to obtain $\pi(x, t; G(t,T))$, and then combining $v(x, t, t+\delta t \,|\, G(t,T))$ and $\pi(x, t; G(t,T))$ using (18). We emphasize that (20) (with appropriate boundary conditions) applies over any interval $[t, t+\delta t]$, which allows for the computation of $m(x, t, t+\delta t \,|\, G(t,T))$ over any such interval. By repeatedly solving (20) and using (18) for every interval of the form $[t_i, t_{i+1}]$ over which one has data, one can compute the conditioned expected change vector $m(x(t_i), t_i, t_{i+1} \,|\, G(0,T))$ that appears in moment conditions used for estimation.

3. Computing the conditioned expected change for univariate diffusions confined within a box

The previous section derives the partial differential equations obeyed by the components of the expected change vector over a finite time interval when a multivariate diffusion process is conditioned to remain in an open, connected, and bounded region. In this section of the paper, we analyze the case of a univariate diffusion conditioned to stay between some lower and upper boundaries a and b

from time 0 to time T; i.e, we condition a univariate diffusion to remain within the box $(a, b) \times [0, T]$. We begin in Section 3.1 by describing numerical techniques to solve the partial differential equations when the unconditioned process is a flexibly specified univariate diffusion, and then in Section 3.2 we derive explicit solutions for the special case in which the unconditioned process follows a one-dimensional geometric Brownian motion. These explicit solutions allow us to assess the accuracy of our numerical method and also are of interest in situations (such as the analysis of stock prices) in which the unconditioned process can reasonably be modelled as a one-dimensional geometric Brownian motion.

3.1. Numerical approach for a general specification of the unconditioned univariate diffusion

The starting point for this computation is Eq. (18), which we repeat for convenience:

$$m(x, t, t + \delta t \mid G(t, T)) = \frac{1}{\pi(x, t; G(t, T))} v(x, t, t + \delta t | G(t, T)) - x. \tag{23}$$

The strategy to calculate this quantity will be to compute $\pi(x, t; G(t, T))$ and $v(x, t, t + \delta t \mid G(t, T))$ as the (numerical) solutions of two different partial differential equations, (13) and (20), with appropriate boundary conditions. These numerical computations will then be combined to obtain the expected change $m(x, t, t + \delta t \mid G(t, T))$.

A tricky feature of the problem is that the partial differential equations for $\pi(x, t; G(t, T))$ and $v(x, t, t + \delta t \mid G(t, T))$ hold in different regions. Eq. (23) above is the expected change over the interval $[t, t + \delta t]$, and the differential equation (20) used to compute $v(x, t, t + \delta t \mid G(t, T))$ holds on the interval $[t, t + \delta t)$ with boundary conditions (21)–(22). When performing GMM estimation, we will need to solve this differential equation separately for each interval over which the expected change is required, that is, for each point in a time series of data. In contrast, the probability $\pi(x, t; G(t, T))$, which appears in the denominator of the right-hand side of (23), is the probability of satisfying the condition on the entire interval $[0, T]$; we need to solve the differential equation (13) only once on the interval $[0, T)$, and then refer to the single solution at the various points (x, t) in order to obtain $\pi(x, t; G(t, T))$.

3.1.1. Computation of the probability $\pi(x, t; G(t, T))$

We show above that the probability $\pi(x, t; G(t, T))$ satisfies the partial differential equation

$$\frac{1}{2}\sigma^2(x)\frac{\partial^2 \pi(x, t; G(t, T))}{\partial x^2} + \mu(x)\frac{\partial \pi(x, t; G(t, T))}{\partial x} + \frac{\partial \pi(x, t; G(t, T))}{\partial t} = 0$$

(Eq. (13)), along with the boundary conditions (14)–(15). This partial differential equation can be solved explicitly only for a few special cases of the drift and diffusion coefficients. Consequently, we approximate the solution numerically using a finite difference scheme based on a discretization of the space $(a, b) \times [0, T]$. Except near the terminal boundary, we use the Crank-Nicholson scheme on a mesh with elements of identical size $\Delta x \times \Delta t$. When working with daily data, it is convenient to set the time step of the Crank-Nicholson scheme Δt to be equal to the time interval δt between trade dates, which we generally do.[2]

We check the accuracy of our numerical scheme by comparing the numerical solution for π obtained when the underlying process follows a geometric Brownian motion with the explicit solution derived for this case in Section 3.2. The geometric Brownian motion is described by

$$\mathrm{d}x(t) = \mu x(t)\,\mathrm{d}t + \sigma x(t)\,\mathrm{d}B(t). \tag{24}$$

We set $\mu = 0.05$ and $\sigma = 0.20$, which are reasonable values for a stock index. The process is confined within $(300, 800) \times [0, 12]$; i.e., $a = 300$, $b = 800$, and $T = 12$ years. We set $\Delta x = 1$ and $\Delta t = 1/250$ (approximately one trade date) in most of the region, but use a smaller time step near the terminal boundary. We find that the solution for π from the Crank-Nicholson scheme and the explicit solution match almost perfectly. In most of the region the difference between the two is on the order of only 10^{-7} or less. Larger errors occur in the neighborhoods of $(300, 12)$ and $(800, 12)$, and stem from the discontinuities at these points where the spatial and terminal boundaries meet. In the neighborhoods of these two points, the function π is highly curved and the errors in the finite difference scheme are as large as 10^{-4}. (Near the terminal boundary but away from $(300, 12)$ and $(800, 12)$, the errors are essentially zero.) Due to the diffusive nature of the solution, the errors are smaller for $t < 12$. The somewhat large (i.e., 10^{-4}) errors near the points where the spatial and terminal boundaries meet will not be important for estimation unless the times-series data come close to one of these two corners of the box.

3.1.2. *Computation of* $v(x, t, t + \delta t \mid G(t, T))$

To compute $m(x, t, t + \delta t \mid G(t, T))$ using Eq. (23), we must also calculate

$$\begin{aligned} &v(x, t, t + \delta t \mid G(t, T)) \\ &\quad = \mathrm{E}[\bar{\pi}(x, t, y, t + \delta t; G(t, t + \delta t))\pi(y, t + \delta t; G(t + \delta t, T))y \mid x(t) = x]. \end{aligned}$$

From Eq. (20), in the time interval $[t, t + \delta t)$ the quantity $v(x, s, t + \delta t \mid G(s, T))$ satisfies the partial differential equation

$$\left[\frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2} + \mu(x)\frac{\partial}{\partial x} + \frac{\partial}{\partial s}\right] v(x, s, t + \delta t \mid G(s, T)) = 0, \tag{25}$$

---

[2] Difficulties arise at the terminal boundary because π is discontinuous at the points (a, T) and (b, T), where the spatial and terminal boundaries meet; i.e., $\pi(a, T; G(T, T)) = \pi(b, T; G(T, T)) = 0$ but $\pi(a + \eta, T; G(T, T)) = \pi(b - \eta, T; G(T, T)) = 1$ for any $\eta > 0$. As a result, the function π is very highly curved near these discontinuities. The discontinuity at the terminal boundary is problematic because the Crank-Nicholson scheme uses the function values on the terminal boundary in its approximation of the spatial derivatives, while the curvature near these discontinuities leads to approximation errors in the finite difference scheme. We address these issues by further subdividing the time interval $[T - \Delta t, T]$ nearest the terminal boundary into 10 subintervals of length $\Delta t/10$, and using a fully implicit scheme in this region.

together with the boundary conditions

$$v(a, s, t+\delta t \mid G(s, T)) = 0 \quad \text{for } s \in [t, t+\delta t], \tag{26}$$

$$v(b, s, t+\delta t \mid G(s, T)) = 0 \quad \text{for } s \in [t, t+\delta t], \tag{27}$$

and

$$v(x, t+\delta t, t+\delta t \mid G(t+\delta t, T)) = \pi(x, t+\delta t; G(t+\delta t, T))x \quad \text{for } x \in (a, b). \tag{28}$$

To compute the conditioned expected change $m(x, t, t+\delta t \mid G(t, T))$ for an interval of the form $[t, t+\delta t]$, we must solve the partial differential equation (25).

For the most part, we solve the partial differential equation using the Crank-Nicholson scheme with just one time step. That is, the time increment Δt in the finite difference scheme is usually chosen to be equal to the length of time δt between data points, so that the interval $[t, t+\delta t]$ over which the partial differential equation holds is traversed using just one finite difference time step. However, finer time steps are employed when the interest rate is near either of the boundaries, because $m(x, t, t+\delta t \mid G(t, T))$ and thus $v(x, t, t+\delta t \mid G(t, T))$ are not as well-behaved in these regions. For example, for the geometric Brownian motion described above, if the current level is 799 and there are four years left in the conditioning event, the expected change is approximately -15. When we compute the expected change using an Euler approximation that assumes the drift is constant during the interval, we obtain a result of approximately -40. However, our use of the finer mesh turns out not to make a significant difference in the estimation results that we report below. Because the use of the finer mesh is important at only a limited number of data points, it has little impact on the moment conditions that we use in estimation, and thus little impact on the estimates.

3.1.3. Computation of $m(x, t, t+\delta t \mid G(t, T))$

We compute $m(x, t, t+\delta t \mid G(t, T))$ by substituting into Eq. (23) the numerical solutions of $\pi(x, t; G(t, T)$ and $v(x, t, t+\delta t \mid G(t, T))$ that are described in the previous two subsections. We check the accuracy of our numerical approach to compute $m(x, t, t+\delta t \mid G(t, T))$ by comparing it to the explicit solution that is available in the case of geometric Brownian motion, using the drift and diffusion coefficients that we use above to check the accuracy of the computation of $\pi(x, t; G(t, T))$. Once again a comparison of the explicit solution with the numerical solution indicates that the Crank-Nicholson finite difference approximation for m is highly accurate. The errors are essentially zero about 20 or more mesh points away from the boundaries. Even within 20 mesh points of a boundary, the relative error is only on the order of 0.01% of the value of x. Since in the estimations we perform below errors in computing m of this magnitude occur at only a few data points, their effect on our moment conditions is not important.

3.2. Explicit calculations for geometric Brownian motion

We now derive an explicit expression for $m(x, t, t+\delta t \mid G(t, T))$ when the underlying process obeys a geometric Brownian motion. This expression is used to check the accuracy of the results from our numerical approach in the previous subsection and should also be of interest to researchers studying equity markets in cases in which geometric Brownian motion is a reasonable model for the unconditioned process.

Assume that the stochastic process $\{x(t)\}$ obeys

$$\mathrm{d}x(t) = \mu x(t)\,\mathrm{d}t + \sigma x(t)\,\mathrm{d}B(t), \tag{29}$$

where $B(t)$ is a standard Brownian motion, μ is a constant drift parameter, and σ is a constant volatility parameter. Letting $f(x, t, y, t+\delta t)$ be the unconditioned transition density for the process to go from $x(t) = x$ to $x(t+\delta t) = y$, we have

$$f(x, t, y, t+\delta t)\,\mathrm{d}y = \frac{1}{\sigma y\sqrt{2\pi\delta t}} \exp\left(-\frac{(\ln(y/x) - (\mu - \frac{1}{2}\sigma^2)\delta t)^2}{2\sigma^2\delta t}\right)\mathrm{d}y. \tag{30}$$

To obtain an expression for $m(x, t, t+\delta t \mid G(t, T))$, we proceed by deriving in the following order expressions for the probabilities $\pi(x, t; G(t, T))$ and $\bar{\pi}(x, t, y, t+\delta t; G(t, t+\delta t))$, and the conditioned transition density $f(x, t, y, t+\delta t \mid G(t, T))$.

3.2.1. Probability of the event $G(t, T)$

We turn first to finding a "closed-form" expression for the probability $\pi(x, t; G(t, T))$ of the process staying within (a, b) during the time interval $[t, t+\delta t]$ given the current level $x(t) = x$.

Kunitomo and Ikeda (1992) consider the problem of pricing options with curved boundaries when the underlying asset follows a geometric Brownian motion. Their Theorem 3.1 is a statement about call option prices conditional on the stock price not hitting an upper or lower boundary. A standard result is that the price of a call option can be expressed as an expectation of final payoffs; thus, it is closely related to $\pi(x, t; G(t, T))$.

If we consider the special case of constant boundaries and set the option strike price to the value of our lower boundary, we can modify Kunitomo and Ikeda equation (3.2) to obtain for $x \in (a, b)$ the following "closed-form" expression for $\pi(x, t; G(t, T))$:

$$\begin{aligned}\pi(x, t; G(t, T)) = \sum_{n=-\infty}^{\infty} &\left\{\left(\frac{b^n}{a^n}\right)^c \cdot [N(\bar{d}_{1n}) - N(\bar{d}_{2n})]\right. \\ &\left. - \left(\frac{a^{n+1}}{b^n x}\right)^c \cdot [N(\bar{d}_{3n}) - N(\bar{d}_{4n})]\right\},\end{aligned} \tag{31}$$

where $c = 2\mu/\sigma^2$, $\tau = T - t$,

$$\bar{d}_{1n} = \frac{\ln(xb^{2n}/a^{2n+1}) + (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \tag{32}$$

$$\bar{d}_{2n} = \frac{\ln(xb^{2n-1}/a^{2n}) + (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \tag{33}$$

$$\bar{d}_{3n} = \frac{\ln(a^{2n+1}/xb^{2n}) + (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \tag{34}$$

$$\bar{d}_{4n} = \frac{\ln(a^{2n+2}/xb^{2n+1}) + (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \tag{35}$$

and $N(\cdot)$ is the standard normal distribution function. An interesting special case occurs when we condition on the geometric Brownian motion process never going below a. In this case, we can obtain the probability of satisfying this condition by letting $b \to \infty$ in (31). This probability might be useful for analyzing survivorship bias in the estimation of diffusion models of stock prices, for example. Combined with Eq. (12) and the other formulas in this section, it allows for exact maximum likelihood estimation when the underlying unconditioned process is a geometric Brownian motion.

We use this expression in Section 3.1 above to check the numerical scheme we developed for computing $\pi(x, t; G(t, T))$ when the underlying process follows a flexibly specified univariate diffusion. In our calculations, the series turns out to converge quickly as claimed by Kunitomo and Ikeda (1992).

3.2.2. The probability $\bar{\pi}(x, t, y, t + \delta t; G(t, t + \delta t))$

By definition, $\bar{\pi}(x, t, y, t + \delta t; G(t, t + \delta t))$ is the probability that the geometric Brownian bridge with fixed values x at starting time t and y at ending time $t + \delta t$ never hits either of the two spatial boundaries; that is,

$$\bar{\pi}(x, t, y, t + \delta t; G(t, t + \delta t)) \\ = \mathrm{P}\left[\max_{u \in [t, t+\delta t]} x(u) < b,\ \min_{u \in [t, t+\delta t]} x(u) > a \,|\, x(t) = x, x(t + \delta t) = y\right]. \tag{36}$$

When calculating this probability, it is easier to work with the following transformed process, which is a Brownian motion with drift:

$$X(t) \equiv \frac{1}{\sigma} \ln x(t) = \lambda t + B(t), \tag{37}$$

where $\lambda \equiv (1/\sigma)(\mu - 1/2\sigma^2)$. It follows from Douady's (1998) equation (4.1) and Girsanov's Theorem that

$$\mathrm{P}\left[X(t + \delta t) \in \mathrm{d}y,\ \max_{u \in [t, t+\delta t]} X(u) - X(t) < h,\ \min_{u \in [t, t+\delta t]} X(u) - X(t) > k | X(t) = x\right] \\ = \mathbf{1}_{\{k < x < h, k < y < h\}} \cdot \exp\left(-\frac{\lambda^2 \delta t}{2} + \lambda(y - x)\right) \sum_{n=-\infty}^{\infty} \{g_{\delta t}((y - x) + 2n(h - k)) \\ - g_{\delta t}(2(h - y) - (y - x) + 2n(h - k))\}\, \mathrm{d}y, \tag{38}$$

where

$$g_{\delta t}(z) = \frac{1}{\sqrt{2\pi\delta t}} \exp\left(-\frac{z^2}{2\delta t}\right) \tag{39}$$

is the normal density function with mean zero and variance δt.

Now using the rule $\mathrm{P}[A \mid B, C] = \mathrm{P}[A \cap B \mid C]/\mathrm{P}[B \mid C]$ and the fact that

$$\mathrm{P}[X(t+\delta t) \in \mathrm{d}y \mid X(t) = x] = \frac{1}{\sqrt{2\pi\delta t}} \exp\left(-\frac{(y - x - \lambda\delta t)^2}{2\delta t}\right) \mathrm{d}y, \tag{40}$$

we have

$$\begin{aligned}
&\mathrm{P}\left[\max_{u\in[t,t+\delta t]} X(u) - X(t) < h,\ \min_{u\in[t,t+\delta t]} X(u) - X(t) > k \mid X(t) = x, X(t+\delta t) = y\right] \\
&\quad = \frac{\mathrm{P}[X(t+\delta t)\in \mathrm{d}y, \max_{u\in[t,t+\delta t]} X(u) - X(t) < h, \min_{u\in[t,t+\delta t]} X(u) - X(t) > k \mid X(t) = x]}{\mathrm{P}[X(t+\delta t)\in \mathrm{d}y \mid X(t) = x]} \\
&\quad = \mathbf{1}_{\{k<x<h, k<y<h\}} \cdot \sqrt{2\pi\delta t} \exp\left(\frac{(y-x)^2}{2\delta t}\right) \sum_{n=-\infty}^{\infty} \{g_{\delta t}((y-x) + 2n(h-k)) \\
&\qquad - g_{\delta t}(2(h-y) - (y-x) + 2n(h-k))\},
\end{aligned} \tag{41}$$

where in the last step we substitute the expressions (38) and (40) for the numerator and denominator, and simplify the result.

Going back to our original process $\{x(t)\}$, we have

$$\begin{aligned}
&\bar{\pi}(x, t, y, t+\delta t; G(t, T)) \\
&\quad = \mathrm{P}\left[\max_{u\in[t,t+\delta t]} x(u) < b,\ \min_{u\in[t,t+\delta t]} x(u) > a \mid x(t) = x, x(t+\delta t) = y\right] \\
&\quad = \mathrm{P}\Bigg[\max_{u\in[t,t+\delta t]} X(u) - X(t) < h \equiv \frac{1}{\sigma}(\ln b - \ln x(t)), \\
&\qquad\quad \min_{u\in[t,t+\delta t]} X(u) - X(t) > k \equiv \frac{1}{\sigma}(\ln a - \ln x(t)) \mid X(t) = \frac{1}{\sigma}\ln x, \\
&\qquad\quad X(t+\delta t) = \frac{1}{\sigma}\ln y\Bigg] \\
&\quad = \mathbf{1}_{\{a<x<b, a<y<b\}} \cdot \sqrt{2\pi\delta t} \exp\left(\frac{(\ln y/x)^2}{2\sigma^2\delta t}\right) \\
&\qquad \times \sum_{n=-\infty}^{\infty} \left\{ g_{\delta t}\left(\frac{1}{\sigma}\ln(y/x) + \frac{2n}{\sigma}\ln(b/a)\right)\right. \\
&\qquad \left. - g_{\delta t}\left(\frac{2}{\sigma}\ln(b/x) - \frac{1}{\sigma}\ln(y/x) + \frac{2n}{\sigma}\ln b/a\right)\right\},
\end{aligned} \tag{42}$$

where in the last step we use Eq. (41) and substitute in the expressions for h and k.

Expression (42) is the key result in this subsection. Although it seems quite complicated, in many situations a good approximation can be obtained by including only a few of the terms in the infinite summation.

3.2.3. The conditioned transition density $f(x,t,y,t+\delta t\,|\,G(t,T))$ and the conditioned expected change $m(x,t,t+\delta t\,|\,G(t,T))$

Recall that $f(x,t,y,t+\delta t\,|\,G(t,T))$ is the transition density for the process $\{x(t)\}$ to go from $x(t)=x$ to $x(t+\delta t)=y$ conditional on $G(t,T)$; from (12) we have

$$\begin{aligned} &f(x,t,y,t+\delta t\,|\,G(t,T))\,\mathrm{d}y \\ &\quad=\frac{\bar{\pi}(x,t,y,t+\delta t;G(t,t+\delta t))\pi(y,t+\delta t;G(t+\delta t,T))f(x,t,y,t+\delta t)}{\pi(x,t;G(t,T))}\,\mathrm{d}y, \end{aligned} \tag{43}$$

where here and in the remainder of this subsection we assume that $x\in(a,b)$ and $y\in(a,b)$. Eqs. (30), (31), and (42) provide expressions for all of the quantities on the right-hand side of (43). Simply substituting these quantities into Eq. (43) provides one method for computing the conditioned transition density $f(x,t,y,t+\delta t\,|\,G(t,T))$. However, because the unconditioned transition density $f(x,t,y,t+\delta t)$ is always concentrated around x when δt is small, we can simplify the calculation by approximating $\bar{\pi}(x,t,y,t+\delta t;G(t,t+\delta t))$ as

$$\begin{aligned} &\bar{\pi}(x,t,y,t+\delta t;G(t,t+\delta t)) \\ &\quad\approx\begin{cases} \mathrm{P}\left[\max_{u\in[t,t+\delta t]}x(u)<b\,|\,x(t)=x,x(t+\delta t)=y\right] & \text{if } x(t) \text{ is near } b, \\ \mathrm{P}\left[\min_{u\in[t,t+\delta t]}x(u)>a\,|\,x(t)=x,x(t+\delta t)=y\right] & \text{if } x(t) \text{ is near } a, \\ 1 & \text{otherwise.} \end{cases} \end{aligned} \tag{44}$$

This approximation is excellent when δt is small. We use this approximation (44) in Section 3.1 above when we check the accuracy of our numerical scheme for computing m.

The first probability in the approximation scheme can be worked out by noting that

$$\begin{aligned} &\mathrm{P}\left[X(t+\delta t)\in\mathrm{d}y,\max_{u\in[t,t+\delta t]}X(u)-X(t)<h|X(t)=x\right] \\ &\quad=\frac{1}{\sqrt{2\pi\delta t}}\left[\exp\left(-\frac{(y-x-\lambda\delta t)^2}{2\delta t}\right)\right. \\ &\qquad\left.-\exp\left(2\lambda h-\frac{(2h+\lambda\delta t-(y-x))^2}{2\delta t}\right)\right]\mathrm{d}y \end{aligned} \tag{45}$$

An argument similar to that used in the previous subsection then gives us:

$$\begin{aligned} &\mathrm{P}\left[\min_{u\in[t,t+\delta t]}x(u)<b\,|\,x(t)=x,x(t+\delta t)=y\right] \\ &\quad=1-\exp\left(-\frac{(2\ln(b/x)-\ln(y/x))^2-\ln^2 y/x}{2\sigma^2\delta t}\right). \end{aligned} \tag{46}$$

The second probability in the approximation scheme can be worked out by using equation (46) and the reflection principle,

$$\mathrm{P}[\lambda t + B(t) > k] = \mathrm{P}[-\lambda t + B(t) < -k]. \tag{47}$$

Using these two probabilities, the approximation is

$$\bar{\pi}(x, t, y, t+\delta t; G(t, t+\delta t)) \approx \begin{cases} 1 - \exp\left(-\dfrac{(2\ln b/x - \ln y/x)^2 - \ln^2 y/x}{2\sigma^2 \delta t}\right) & \text{if } x(t) \text{ is near } b, \\ 1 - \exp\left(-\dfrac{(2\ln a/x - \ln y/x)^2 - \ln^2 y/x}{2\sigma^2 \delta t}\right) & \text{if } x(t) \text{ is near } a, \\ 1 & \text{otherwise.} \end{cases} \tag{48}$$

Alternatively, one can start from equation (42) and take the limit $a \to 0$ when x is close to b and the limit $b \to +\infty$ when x is close to a. When $b \to +\infty$, we need only consider $n = 0$ in the first term of the infinite summation and $n = -1$ in the second term. The situation is similar for $a \to 0$.

Finally, the expected change over a finite time interval can be calculated from

$$\begin{aligned} & m(x, t, t+\delta t \,|\, G(t, T)) \\ &\quad = \int_a^b (y - x) f(x, t, y, t+\delta t \,|\, G(t, T))\, \mathrm{d}y \\ &\quad = \frac{1}{\pi(x, t; G(t, T))} \int_a^b f(x, t, y, t+\delta t) \bar{\pi}(x, t, y, t+\delta t; G(t, t+\delta t)) \\ &\qquad \times \pi(y, t+\delta t; G(t+\delta t, T)) y \,\mathrm{d}y - x. \end{aligned} \tag{49}$$

We use an adaptive recursive Newton-Cotes eight-panel rule to evaluate this integral, and find that the result agrees closely with the expected change computed using the Crank-Nicholson scheme.

4. Conditioning bias when estimating a univariate diffusion

The remainder of the paper illustrates the techniques developed above by applying them to a problem of recent interest, the estimation of a univariate, time-homogeneous diffusion process from a time series of interest rate data. In this section of the paper, we perform a Monte Carlo experiment to demonstrate the bias that results when a diffusion process is estimated from a time series of interest rates without taking into account the realized minimum and maximum values. We also considers the discretization bias that arises when the expected change and variance of the process are replaced by approximations that are of relatively low order in δt. In the next section of the paper, we then perform the estimation, conditioning the process to remain between various lower and upper bounds.

The underlying unconditioned process is

$$\mathrm{d}x(t) = \mu(x(t))\,\mathrm{d}t + \sigma(x(t))\,\mathrm{d}B(t), \tag{50}$$

where $\mu(x)$ and $\sigma^2(x)$ are the drift and diffusion coefficients, respectively, and B is a standard Brownian motion. Arithmetic and geometric Brownian motion, the CEV process and various "one-factor" interest rate models are special cases of this specification. It also encompasses the nonparametric model of Stanton (1997) and the flexible specifications proposed by Aït-Sahalia (1996) (henceforth, AS) and Conley et al. (1997) (henceforth, CHLS). The drift and diffusion coefficients of the CHLS specification are given by

$$\mu(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x \tag{51}$$

and

$$\sigma^2(x) = \beta_1 x^{2\beta_2}, \tag{52}$$

where α_i and $\beta_i > 0$ are parameters to be estimated. AS proposes the same specification of the drift and a slightly more general specification for the diffusion.

In order to illustrate the bias that can arise when a diffusion process is estimated from a time series of interest rate data without conditioning, we perform the following simulation experiment. First, 1,000 artificial interest rate series are generated from the CHLS specification with α_2 and α_3 restricted to be equal to zero so that the drift coefficient is linear. Next, these 1,000 paths are used to estimate the CHLS specification without any restriction on α_2 and α_3. Finally, the estimated drift and diffusion coefficients are compared to the drift and diffusion coefficients that generated the data.

We calibrate the CHLS specification using Hansen's (1982) generalized method of moments procedure and the daily time series of the seven-day Eurodollar spot rates (bid-ask midpoints) used in AS. This time series is 5,505 days long and covers 21.73424 years during the period from 1973 to 1995. Its minimum value is 0.02915 (which occurred in early 1993) and its maximum value is 0.24333 (which occurred in 1981). Fig. 1 shows the realized sample path of the interest rate process.

Assume that the interest rate data $\{x_{t_1}, x_{t_2}, \ldots, x_{t_N}\}$ are generated by the CHLS specification and that observations are equally spaced in time. Define $\delta t \equiv t_{i+1} - t_i$ and $\delta x_i \equiv x_{t_{i+1}} - x_{t_i}$. Then, when α_2 and α_3 are restricted to be equal to zero, the CHLS specification becomes

$$\mathrm{d}x(t) = (\alpha_0 + \alpha_1 x(t))\,\mathrm{d}t + \sqrt{\beta_1}\,x(t)^{\beta_2}\,\mathrm{d}B(t). \tag{53}$$

Dividing both sides by $x(t)^{\beta_2}$ (which amounts to a heteroskedasticity correction) results in

$$x(t)^{-\beta_2}\,\mathrm{d}x(t) = x(t)^{-\beta_2}(\alpha_0 + \alpha_1 x(t))\,\mathrm{d}t + \sqrt{\beta_1}\,\mathrm{d}B(t). \tag{54}$$

Using x_i to denote $x(t_i)$ and defining $\tilde{u}_i \equiv x_i^{-\beta_2}\{\delta x_i - \mathrm{E}[\delta x_i | x_i]\}$, Eq. (54) suggests that the parameter vector $\tilde{\theta} = (\alpha_0, \alpha_1, \beta_1, \beta_2)^\top$ be estimated from the

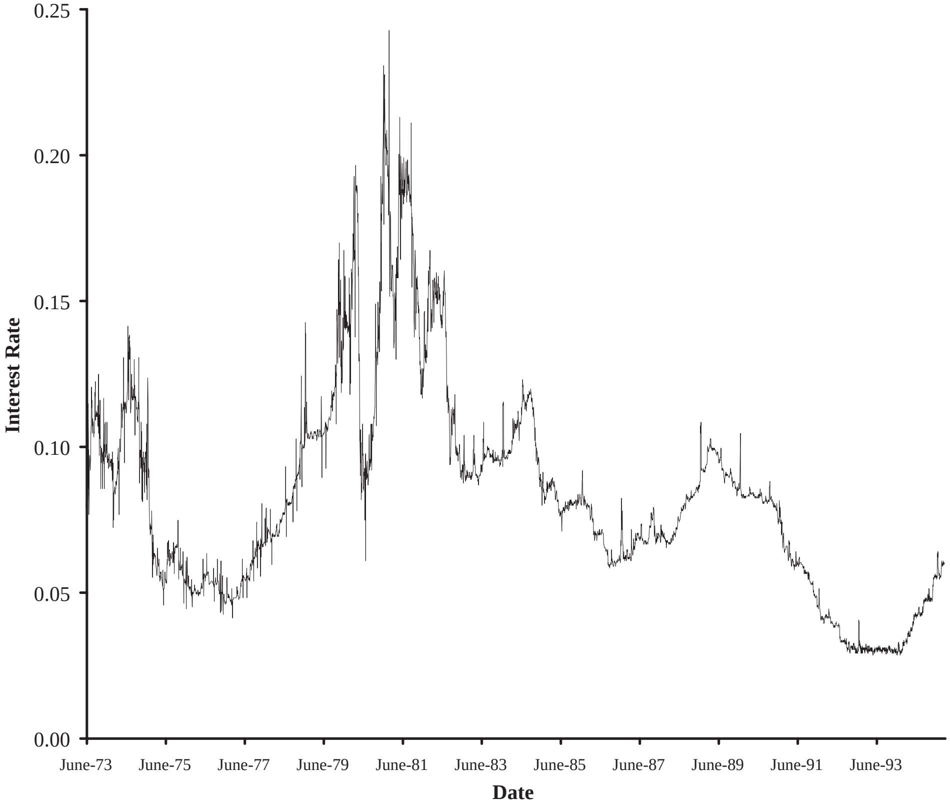

Fig. 1. Time series of 5,505 daily observations of the seven-day Eurodollar deposit spot rate bid–ask midpoints, June 1, 1973 to February 24, 1995. Rates are expressed on an annual basis.

moment conditions

$$\mathrm{E}\begin{bmatrix} \tilde{u}_i \\ x_i\tilde{u}_i \\ \tilde{u}_i^2 - x_i^{-2\beta_2}\mathrm{var}[\delta x_i \mid x_i] \\ x_i\tilde{u}_i^2 - x_i^{1-2\beta_2}\mathrm{var}[\delta x_i \mid x_i] \end{bmatrix} = 0. \tag{55}$$

These moment conditions are those used by Chan et al. (1992) (henceforth, CKLS), except that CKLS do not divide by $x_i^{\beta_2}$ and they use first-order approximations to $\mathrm{E}[\delta x_i|x_i]$ and $\mathrm{var}[\delta x_i|x_i]$.

To implement the moment conditions, we define

$$\tilde{g}_i(\tilde{\theta}) \equiv \begin{bmatrix} \tilde{u}_i \\ x_i\tilde{u}_i \\ \tilde{u}_i^2 - x_i^{-2\beta_2}\,\mathrm{var}[\delta x_i|x_i] \\ x_i\tilde{u}_i^2 - x_i^{1-2\beta_2}\,\mathrm{var}[\delta x_i|x_i] \end{bmatrix} \tag{56}$$

and

$$\tilde{h}(\tilde{\theta}) \equiv \frac{1}{N-1}\sum_{i=1}^{N-1} \tilde{g}_i(\tilde{\theta}), \tag{57}$$

so that $\tilde{h}$ is the sample analogue of the left-hand side of (55). The standard GMM approach is that the estimates are the solution of

$$\widetilde{H}(\theta) = \frac{1}{N-1} \min_{\tilde{\theta}} \tilde{h}(\tilde{\theta})^{\top} \widetilde{W}\tilde{h}(\tilde{\theta}), \tag{58}$$

where $\widetilde{W}$ is a positive definite weighting matrix. Following Hansen (1982), we choose $\widetilde{W} = \widetilde{S}^{-1}(\tilde{\theta})$, where

$$\widetilde{S}(\tilde{\theta}) = \frac{1}{N-2}\sum_{i=1}^{N-1} [(\tilde{g}_i(\tilde{\theta}) - h(\tilde{\theta}))(\tilde{g}_i(\tilde{\theta}) - h(\tilde{\theta}))^{\top}]. \tag{59}$$

One issue is that the data are assumed to be generated by a continuous-time model, yet are observed only at discrete intervals. Aït-Sahalia (1996, 1999, 2002a, 2002b) has argued that failure to account for this fact (e.g., through naive use of the Euler approximation) can result in a discretization bias. As one step in computing an approximation to the transition density, Aït-Sahalia (2002a; Proposition 4 and Eq. (4.3)) provides a series expansion for expectations of polynomial functions (e.g., moments) of a scalar process with a unit diffusion coefficient that converges for fixed time step δt. We can modify this expansion for our more general process and use it to compute approximations to the quantities $\mathrm{E}[\delta x_i|x_i]$ and $\mathrm{E}[(\delta x_i)^2|x_i]$ contained in the terms $\tilde{u}_i$ and $\mathrm{var}[\delta x_i|x_i] = \mathrm{E}[(\delta x_i)^2|x_i] - (\mathrm{E}[\delta x_i|x_i])^2$ that appear in Eq. (56).

The expansion we use is

$$\begin{aligned}\mathrm{E}[f(X_{t+\delta t}, x_0)\,|\,X_t = x_0] = &\sum_{k=0}^{K} A^k(\tilde{\theta})f(X_{t+\delta t}, x_0)\Big|_{X_{t+\delta t}=x_0} \frac{(\delta t)^k}{k!} \\ &+ \mathrm{E}[A^{K+1}(\tilde{\theta})f(X_{t+\delta t}, x_0)|X_t = x_0]\frac{(\delta t)^{K+1}}{(K+1)!},\end{aligned} \tag{60}$$

where the operator $A(\tilde{\theta})$ is defined as

$$A(\tilde{\theta})f(x, x_0) = \mu(x;\tilde{\theta})\frac{\partial f(x, x_0)}{\partial x} + \frac{\sigma^2(x;\tilde{\theta})}{2}\frac{\partial^2 f(x, x_0)}{\partial x^2}. \tag{61}$$

Eq. (60) is Eq. (4.3) in AS (2002a) with some changes in notation. Eq. (4.3) and Proposition 4 in AS (2002a) apply to polynomial functions of a scalar process transformed to have a unit diffusion coefficient. While the functions f we introduce below are polynomial, they are polynomial functions of the original process (50), not the process $\{y(t)\}$ with the unit diffusion coefficient given by the transformation $y = x^{1-\beta_2}/(1-\beta_2)\sqrt{\beta_1}$. We can nonetheless use the expansion in Eq. (4.3) of AS (2002a) because the property of polynomials used in the proof of AS Proposition 4 is that they have polynomial growth. The functions f we introduce below, while not polynomial in y, do have polynomial growth as functions of y provided $\beta_2 \neq 1$.

In order to use Eqs. (60) and (61) to compute $\mathrm{E}[\delta x_i]$, we need to set $f(x, x_0)$ to

$$f(x, x_0) = x - x_0, \tag{62}$$

and in order to use these equations to compute $\mathrm{E}[(\delta x_i)^2|x_i]$, we need to set $f(x, x_0)$ to

$$f(x, x_0) = x^2 - 2xx_0 + x_0^2. \tag{63}$$

Then, first- and second-order approximations to $\mathrm{E}[\delta x_i|x_i]$ are given, respectively, by

$$\mathrm{E}[\delta x_i|x_i] = (\alpha_0 + \alpha_1 x_i)\delta t + O((\delta t)^2) \tag{64}$$

and

$$\mathrm{E}[\delta x_i|x_i] = (\alpha_0 + \alpha_1 x_i)\delta t + \alpha_1(\alpha_0 + \alpha_1 x_i)\frac{(\delta t)^2}{2} + O((\delta t)^3), \tag{65}$$

while the first- and second-order approximations to $\mathrm{E}[(\delta x_i)^2|x_i]$ are given, respectively, by

$$\mathrm{E}[(\delta x_i)^2|x_i] = \beta_1 x_i^{2\beta_2}\delta t + O((\delta t)^2) \tag{66}$$

and

$$\begin{aligned}\mathrm{E}[(\delta x_i)^2|x_i] = {} & \beta_1 x_i^{2\beta_2}\delta t + \left\{\frac{1}{2}\beta_1 x_i^{2\beta_2}[4\alpha_1 + 2\beta_1\beta_2(2\beta_2 - 1)x_i^{-2+2\beta_2}]\right. \\ & \left. + (\alpha_0 + \alpha_1 x_i)[2\beta_1\beta_2 x_i^{-1+2\beta_2} + 2(\alpha_0 + \alpha_1 x_i)]\right\}\frac{(\delta t)^2}{2} + O((\delta t)^3).\end{aligned} \tag{67}$$

We apply the GMM procedure to the AS data using approximations that are second order in δt to compute $\tilde{u}_i$ and $\mathrm{var}[\delta x_i|x_i]$. In particular, we compute $\tilde{u}_i$ using the approximation in Eq. (65). We compute $\mathrm{var}[\delta x_i\,|\,x_i]$ from $\mathrm{var}[\delta x_i\,|\,x_i] = \mathrm{E}[(\delta x_i)^2\,|\,x_i] - (\mathrm{E}[\delta x_i\,|\,x_i])^2$, using Eq. (67) to compute $\mathrm{E}[(\delta x_i)^2\,|\,x_i]$ and Eq. (64) to compute $\mathrm{E}[\delta x_i\,|\,x_i]$. (When the first-order approximation for $\mathrm{E}[\delta x_i\,|\,x_i]$ is squared to produce $(\mathrm{E}[\delta x_i\,|\,x_i])^2$, the resulting quantity is second order in δt, which is the accuracy needed to compute $\mathrm{var}[\delta x_i\,|\,x_i]$ to the second order.) The GMM parameter estimates are $\alpha_0 = 0.063$, $\alpha_1 = -0.74$, $\beta_1 = 2.11$, and $\beta_2 = 1.34$.

We use the parameter values just above (with $\alpha_2 = \alpha_3 = 0$) to generate 1,000 artificial sample paths. The initial value for each sample path is determined by a randomly drawn value from the stationary distribution of the process. Once this initial value is obtained, an Euler scheme is used to simulate the path forward for 550,400 steps, where each step covers an interval of $21.73424/(100 \times 5{,}504)$ years (equal to 1/100 of a trade day). The artificial sample path is then defined as the 1st, 101st, 201st, …, and 550,401st values of the simulated path. The resulting artificial sample paths are 5,505 trade dates long. The subdivision of each trade date into 100 equal parts should ensure that any discretization bias that enters into the generation of the artificial paths is inconsequential.

The CHLS specification is estimated for each of the 1,000 paths using a GMM procedure like the one described above except that α_2 and α_3 are no longer restricted to be equal to zero. Accordingly, the parameter vector to be estimated becomes $\theta = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)^\top$. Similar to before, we define $u_i \equiv x_i^{-\beta_2}(\delta x_i - \mathrm{E}[\delta x_i|x_i])$, where the expectation is now taken with respect to the CHLS process and α_2 and α_3

are not restricted to be equal to zero. The moment conditions are now

$$\mathrm{E}\begin{bmatrix} u_i \\ x_i u_i \\ x_i^2 u_i \\ u_i/x_i \\ u_i^2 - x_i^{-2\beta_2}\,\mathrm{var}[\delta x_i|x_i] \\ x_i u_i^2 - x_i^{1-2\beta_2}\,\mathrm{var}[\delta x_i|x_i] \end{bmatrix} = 0. \tag{68}$$

To implement these moment conditions, we define

$$g_i(\theta) \equiv \begin{bmatrix} u_i \\ x_i u_i \\ x_i^2 u_i \\ u_i/x_i \\ u_i^2 - x_i^{-2\beta_2}\,\mathrm{var}[\delta x_i|x_i] \\ x_i u_i^2 - x_i^{1-2\beta_2}\,\mathrm{var}[\delta x_i|x_i] \end{bmatrix} \tag{69}$$

and

$$h(\theta) \equiv \frac{1}{N-1}\sum_{i=1}^{N-1} g_i(\theta), \tag{70}$$

so that h is the sample analogue of the left-hand side of (68). The estimates are the solution of

$$H = \frac{1}{N-1}\min_{\theta}\, h(\theta)^{\top}\, Wh(\theta), \tag{71}$$

where W is a positive definite weighting matrix. Again following Hansen (1982), we choose $W = S^{-1}(\theta)$, where

$$S(\theta) = \frac{1}{N-2}\sum_{i=1}^{N-1}\,[(g_i(\theta) - h(\theta))(g_i(\theta) - h(\theta))^{\top}]. \tag{72}$$

Table 1 reports the mean value and standard error of the 1,000 estimates of the six parameters when $\tilde{u}_i$ and $\mathrm{var}[\delta x_i\,|\,x_i]$ are computed to various orders in δt. Regardless of the order to which u_i and $\mathrm{var}[\delta x_i\,|\,x_i]$ are computed, the estimates of the drift parameters are quite similar and severely biased. For the case in which both u_i and $\mathrm{var}[\delta x_i\,|\,x_i]$ are computed to first order in δt the mean value of the α_0 estimates is -0.44 even though the data-generating value of α_0 is 0.063; the data-generating value of α_1 is -0.74, but the mean value of the estimates of α_1 is 6.11.

For the α_2 parameter, the data-generating value is zero, but the mean value of the estimates is -28.85. Since the drift coefficient is dominated by the $\alpha_2 x^2$ term for large values of the interest rate, the drift coefficient estimates are downwardly biased when the interest rate is large. This bias occurs because the true data-generating process can increase or decrease at the maximum value of any given sample path, whereas

Table 1
Parameter estimates for a nonlinear diffusion when estimating without conditioning artificial data paths generated from a linear diffusion

One thousand artificial paths are generated from the linear diffusion

$$\mathrm{d}x = (\alpha_0 + \alpha_1 x)\,\mathrm{d}t + \sqrt{\beta_1}x^{\beta_2}\,\mathrm{d}B,$$

with $\alpha_0 = 0.063$, $\alpha_1 = -0.74$, $\beta_1 = 2.11$, and $\beta_2 = 1.34$. The initial value for each of the paths is set to a random draw from the stationary distribution of the linear diffusion. An Euler scheme is then used to simulate each path forward from its initial value for 550,400 steps, where each step covers $1/100$th of a trade day (i.e., each step is $21.73424/(100 \times 5{,}504)$ years long). The artificial sample path is then defined as the 1st, 101st, 201st, …, and 550,401st values of the simulated path. The resulting artificial sample paths then each consist of 5,505 observations spaced at $\delta t =$ one-trade-date intervals. Using a GMM procedure, the following nonlinear diffusion is then estimated for each artificial sample path:

$$\mathrm{d}x = (\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x)\,\mathrm{d}t + \sqrt{\beta_1}x^{\beta_2}\,\mathrm{d}B.$$

The six moment conditions for the GMM procedure are:

$$\mathrm{E}\begin{bmatrix} u_i \\ x_i u_i \\ x_i^2 u_i \\ u_i/x_i \\ u_i^2 - x_i^{-2\beta_2}\,\mathrm{var}[\delta x_i \mid x_i] \\ x_i u_i^2 - x_i^{1-2\beta_2}\,\mathrm{var}[\delta x_i \mid x_i] \end{bmatrix} = 0,$$

where $u_i \equiv x_i^{-\beta_2}\delta x_i - \mathrm{E}[\delta x_i \mid x_i]$ and $\mathrm{var}[\delta x_i \mid x_i] = \mathrm{E}[(\delta x_i)^2 \mid x_i] - (\mathrm{E}[\delta x_i \mid x_i])^2$. The table below reports the mean value and standard error (in parentheses) of the 1,000 estimates of the six parameters when u_i and $\mathrm{var}[\delta x_i]$ are computed to the indicated orders in δt. The reported standard errors are sample standard deviations divided by the square root of 1,000.

Panel A: *Data generating values*

α_0	α_1	α_2	α_3	β_1	β_2
0.063	−0.74	0	0	2.11	1.34

Panel B: *Monte Carlo estimates*

Order of drift	Order of variance	α_0	α_1	α_2	α_3	β_1	β_2
1	1	−0.44	6.11	−28.85	0.0116	2.05	1.3329
		(0.01)	(0.18)	(0.88)	(0.0003)	(0.01)	(0.0008)
2	1	−0.45	6.22	−29.33	0.0118	2.05	1.3329
		(0.01)	(0.18)	(0.90)	(0.0003)	(0.01)	(0.0008)
2	2	−0.45	6.20	−29.25	0.0118	2.11	1.3379
		(0.01)	(0.18)	(0.90)	(0.0003)	(0.01)	(0.0008)

any particular sample path necessarily decreases after its maximum value. As a result, the drift coefficient estimate for a given sample path tends to be negatively biased when the level of the interest rate is near the maximum of the sample path.

The data-generating value of α_3 is zero, but the mean value of the α_3 estimates is 0.0116. Since the drift coefficient is dominated by the α_3/x term for small values of the interest rate, the drift coefficient estimates are upwardly biased when the interest rate is small. The reason for the upward bias for small interest rates is that the true data-generating process can increase or decrease at the minimum value of any given sample path, whereas any particular sample path necessarily increases after its minimum value. As a result, the drift coefficient estimate for a given sample path tends to be positively biased when the level of the interest rate is near the minimum of the sample path. The bias near the minimum value of the process is smaller than near the maximum, because the process is much less volatile near the minimum due to the dependence of the diffusion coefficient on the level of the interest rate. Intuitively, due to this weaker diffusive component of the process, a smaller change in the drift is needed to prevent the process from crossing a fixed boundary.

The bias in the estimation of the diffusion parameters is smaller than that in the estimation of the drift parameters. The mean estimate of the β_1 parameter is 2.05 in comparison with a data-generating value of 2.11. The mean estimate of the β_2 parameter is 1.3329 in comparison with a data-generating value of 1.34.

In order to understand the impact of computing u_i only to order δt, we next re-estimate the 1,000 paths using expansion (60) to compute u_i to order $(\delta t)^2$ and $\mathrm{var}[\delta x_i \mid x_i]$ to order δt. The results are also displayed in Table 1. There is very little change in the estimates of the drift parameters. In all cases, the changes are less than one standard error and are a small fraction of the difference between the mean parameter estimates and the data-generating values. There is no discernable change in the estimates of the diffusion parameters. Finally, in Table 1 we also report the results of estimating the 1,000 paths using expansion (60) to compute both u_i and $\mathrm{var}[\delta x_i|x_i]$ to order $(\delta t)^2$. As explained above, when we compute $\mathrm{var}[\delta x_i|x_i]$ from $\mathrm{E}[(\delta x_i)^2 \mid x_i] - (\mathrm{E}[\delta x_i \mid x_i])^2$, we use Eq. (67) to compute $\mathrm{E}[(\delta x_i)^2|x_i]$ and Eq. (64) to compute $\mathrm{E}[\delta x_i|x_i]$. While computing $\mathrm{var}[\delta x_i|x_i]$ to the second rather than first order in δt improves the estimates of the diffusion parameters, it has almost no effect on the estimation of the drift parameters. Since our main focus in this paper is the estimation of the drift parameters, we will proceed below by computing $\mathrm{E}[\delta x_i|x_i]$ and u_i to a high level of accuracy, but for simplicity will compute $\mathrm{var}[\delta x_i|x_i]$ to only the first order in δt.

5. Estimation results

In this section of the paper we first investigate the potential importance of conditioning the interest rate data to remain between lower and upper boundaries by examining the extent to which the conditioning impacts the expected one-trade-date change in the process. We then estimate the CHLS specification from the data unconditionally and when conditioning on a variety of lower and upper bounds.

5.1. Impact of conditioning on the expected change in rates

Fig. 2 illustrates the importance of conditioning by graphing the conditioned and unconditioned expected changes in the interest rate as functions of the level of the rate. Here, the drift of the unconditioned process is the linear drift $\mu(x) = 0.03400 - 0.02834x$ and the diffusion coefficient is $\sigma^2(x) = 2.0511x^{2(1.3333)}$. We condition on the event that the rate stays between the boundaries of $a = 0.02487$ (10

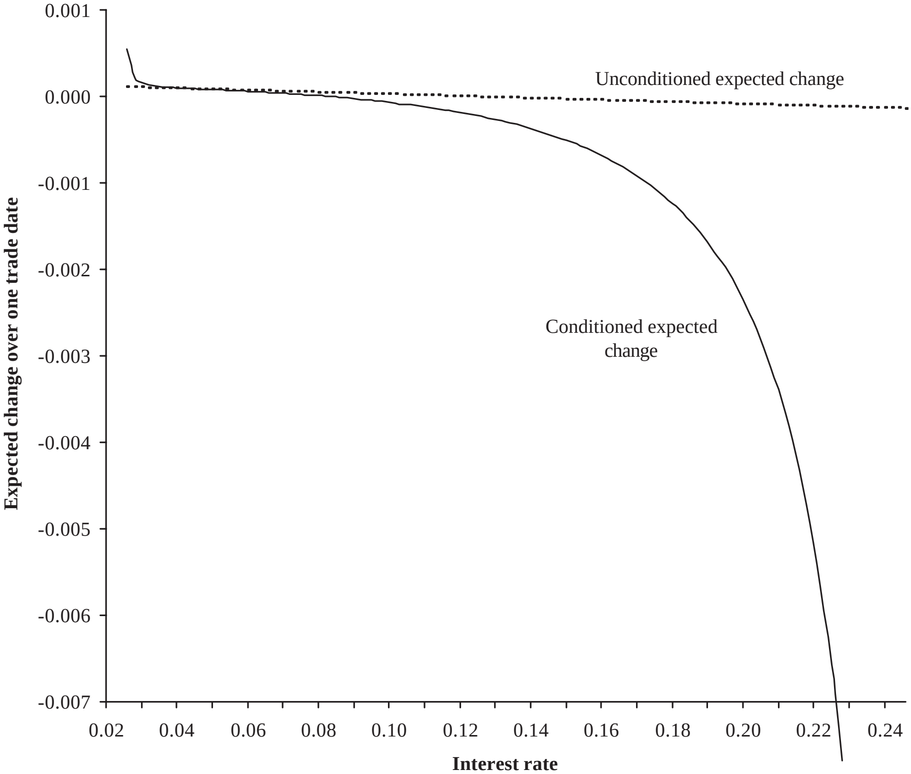

Fig. 2. The one-trade-date expected change for an interest rate process with and without conditioning on the process remaining between $a = 0.02487$ and $b = 0.24761$ for the next twelve years as a function of the current interest rate level. The drift coefficient of the unconditioned process is assumed to be $\mu(x) = 0.03400 - 0.02834x$ and the diffusion coefficient is assumed to be $\sigma^2(x) = 2.0511x^{2(1.3333)}$. The solid line is the conditioned one-trade-date expected change computed using a Crank-Nicholson scheme, and the dashed line is the unconditioned one-trade-date expected change. Rates are expressed on an annual basis.

finite-difference mesh points, or, 42.8 basis points below the realized minimum of 0.02915) and $b = 0.24761$ (10 mesh points, or, 42.8 basis points above the realized maximum of 0.24333) for another 12 years. The choice of 12 years places us roughly in the middle of the 21.73424 years covered by the interest rate sample; the conditioned expected changes at other times are similar to that shown in the figure (and the unconditioned expected changes are identical). This figure shows that the conditioning can have an important effect on the expected change, particularly for large values of the interest rate. Thus, it will not be surprising if conditioning has an important effect on parameter estimation.

Fig. 3 shows that the effect of conditioning illustrated in Fig. 2 is important along the realized sample path of interest rates. In particular, Fig. 3 plots the value of the conditioned expected change $m(x(t), t, t + \delta t | G(t, T))$ along the sample path of interest rates; i.e., it plots the sample path of the conditioned expected change. The

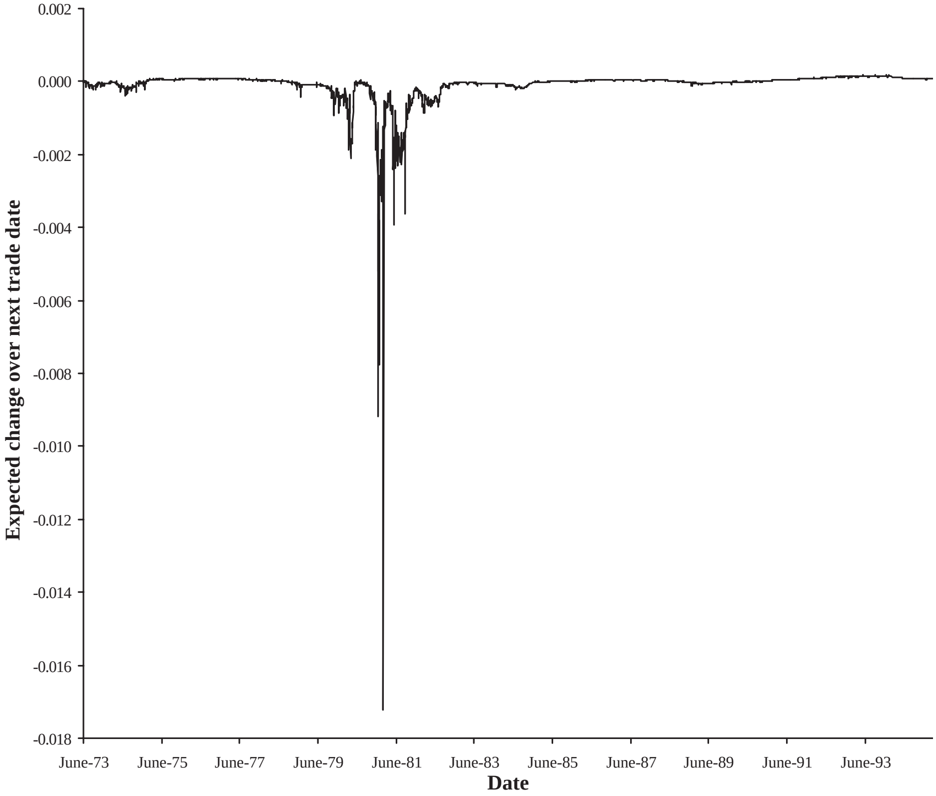

Fig. 3. The expected one-trade-date change in the seven-day Eurodollar rate on each trade date from June 1, 1973 to February 23, 1995, conditional on the rate remaining between $a = 0.02487$ and $b = 0.24761$ until February 24, 1995. The drift coefficient of the unconditioned process is assumed to be $\mu(x) = 0.03400 - 0.02834x$ and the diffusion coefficient is $\sigma^2(x) = 2.0511x^{2(1.3333)}$. The expected one-trade-date change is computed using the Crank-Nicholson scheme on each trade date based on that trade date's observed seven-day Eurodollar rate and the time remaining until February 24, 1995. Rates (and the expected changes in them) are expressed on an annual basis.

sample path of the unconditioned expected change is not shown, but is essentially zero — were the figure to show the unconditioned expected change, it would be difficult visually to distinguish it from zero. Comparing Fig. 3 to the sample path of interest rates shown in Fig. 1, one can see that when interest rates are close to the minimum value of 0.02915, the conditioned expected change $m(x(t), t, t + \delta t \mid G(t, T))$ is slightly positive. When rates are close to their maximum value of 0.24333, the conditioned expected change is negative and sometimes strikingly so, with the most negative value obtained at the maximum of the interest rate process, at a time index of 1,942 (February 3, 1981). These results make clear that the conditioning can have an important effect on method-of-moments estimation via the moment conditions involving the expected change. While this paper does not pursue maximum likelihood estimation in the presence of conditioning, the effect on the expected change implies a similar effect on the location of the transition density, and thus has a potentially significant impact on the likelihood function.

5.2. Impact of conditioning on parameter estimates

To assess the effect of conditioning on parameter estimation, we apply the GMM procedure outlined in the previous section to the interest rate data to estimate the CHLS model both without conditioning and with conditioning on the event G. In order to account for the conditioning, we use the conditioned expected change $m(x_i, t_i, t_i + \delta t \mid G(t_i, T))$ in place of $\mathrm{E}[\delta x_i \mid x_i]$; that is, we replace the quantity

$$u_i = x_i^{-\beta_2}(\delta x_i) - x_i^{-\beta_2}\mathrm{E}[\delta x_i \mid x_i] \tag{73}$$

with

$$u_i \equiv x_i^{-\beta_2}(\delta x_i) - x_i^{-\beta_2} m(x_i, t_i, t_i + \delta t | G(t_i, T)), \tag{74}$$

where $m(x_i, t_i, t_i + \delta t \mid G(t_i, T))$ is computed as in Section 3.1. Using this expression for u_i, we minimize the objective function $H(\theta) = 1/(N-1)h(\theta)^\top W(\theta)h(\theta)$ over the vector $\theta = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)$ using a Nelder-Mead simplex method. As above, the numerical scheme is implemented with the mesh constructed such that there are 500 mesh points between the minimum and the maximum of the AS data. Consequently, the interval between mesh points is 4.2836 basis points. The CHLS parameters are estimated conditioning on $G(t, T)$ with a and b displaced the same number of mesh points from, respectively, the minimum and maximum of the AS interest rate data. We compute the estimates for five different levels of the displacement.

Fig. 4 plots the drift coefficients computed using the five sets of parameter estimates for various levels of this displacement. Each drift coefficient is plotted between the minimum and the maximum values of the data. The conditioning appears to have a large impact on the estimation of the drift coefficient, and the impact is largest when the interest rate is near its maximum. For large interest rates, conditioning increases the estimate of the unconditioned drift coefficient. Furthermore, the change in the estimated drift coefficient is increasing in the severity of the conditioning (i.e., the change increases as the number of mesh points the process is permitted to go below the minimum of 0.02915 and above the maximum of 0.24333 decreases.) Accounting for conditioning increases the estimate of the drift of the unconditioned process for large interest rates because without conditioning, a process with the CHLS and AS drift specification can go either up or down near the maximum value of the data. In the data, however, the process decreases from the maximum value. As a result, when conditioning is not taken into account, the estimation procedure chooses drift parameters that make it difficult for the process to increase near the maximum of the data. When we condition on remaining within a and b, however, the conditioning (i.e., the conditioned drift) prevents the process from increasing beyond the upper bound upon which we are conditioning, and thus makes it unlikely that the process increases beyond the maximum in the data. As a result, the drift parameters do not need to be set to make it hard for the unconditioned process to increase beyond the maximum. As the conditioning is

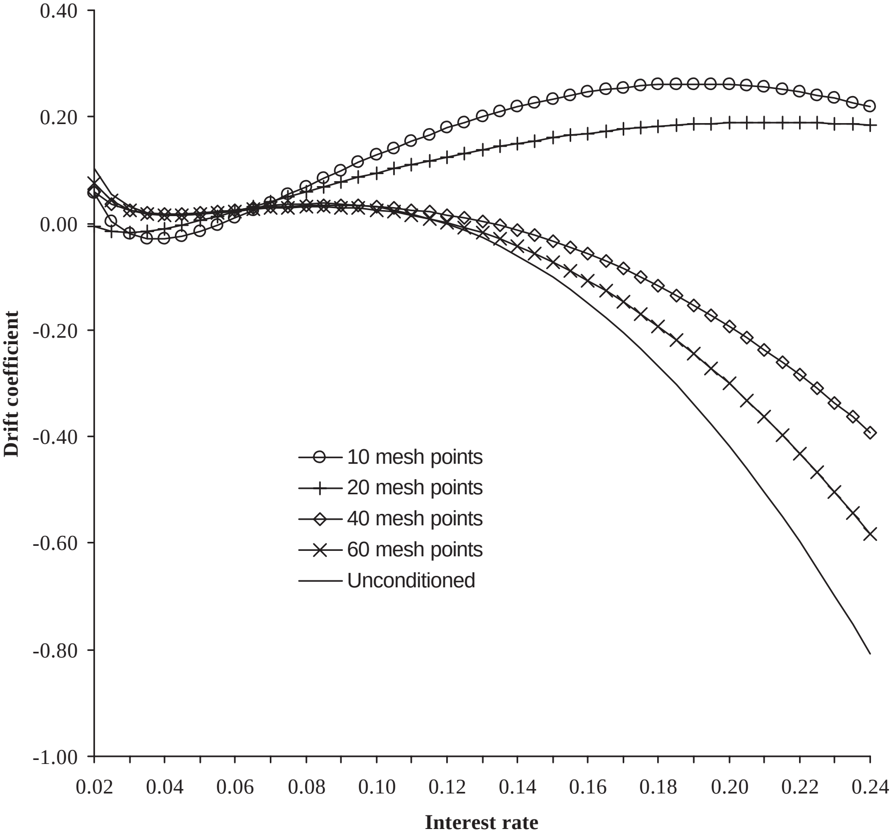

Fig. 4. Drift coefficients estimated from seven-day Eurodollar rates, June 1, 1973 to February 24, 1995, conditioning on the event that the process remains in the interval (a, b) until February 24, 1995, for various choices of a and b. The drift coefficient $\mu(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x$ and diffusion coefficient $\sigma^2(x) = \beta_1 x^{2\beta_2}$ are estimated from the Eurodollar rate data using the generalized method of moments, and the moment conditions are constructed using one-trade-date conditioned expected changes. The solid curve is the estimate of the unconditioned drift coefficient when on each trade date there is no assumption made about the minimum and maximum values the process will attain until February 24, 1995. The other curves are drift estimates obtained when on each trade the one-trade-date expected change is computed conditional on the process remaining between $a = 0.02915 - \text{displacement}$ and $b = 0.24333 + \text{displacement}$ for the remainder of the time until February 24, 1995, where 0.02915 and 0.24333 are the minimum and maximum interest rates observed in the sample. The displacement is expressed in terms of the number of spatial steps, each of size $\Delta x = 4.2836$ basis points. The curve with symbols $\times$ is the drift estimate when the conditioning is done using a displacement of $60\Delta x$ from the observed minimum and maximum. The curve with diamonds is for a displacement of $40\Delta x$, the curve with symbols $+$ is for a displacement of $20\Delta x$, and the curve with circles is for a displacement of $10\Delta x$. Rates are expressed on an annual basis.

made stricter (i.e., as the displacement decreases), the conditioning does more of the work of keeping the process within its observed range and the estimated drift parameters do less of the work. Indeed, when the displacement is as little as 10 or 20

Table 2

Unrestricted conditional estimates Unrestricted parameter estimates obtained using the sample of seven-day Eurodollar rates, June 1, 1973 to February 24, 1995, conditioning on the event that the process remains in the interval (a, b) until February 24, 1995, for various choices of a and b. The lower boundary a is $0.02915 -$ displacement, the upper boundary b is $0.24333 +$ displacement, and 0.02915 and 0.24333 are the minimum and maximum interest rates observed in the sample. The displacement is expressed in terms of the number of spatial steps, each of size $\Delta x = 4.2836$ basis points. The first column lists the displacement, and the next six columns show the parameter estimates obtained by minimizing the quadratic form H in Eq. (71). Numbers in parentheses are the corresponding standard errors. At the point estimates, the values of the quadratic form H are all less than 10^{-15}. For comparison, the table also shows the parameter estimates for the case in which there is no conditioning.

Displacement	α_0	α_1	α_2	α_3	β_1	β_2
Unconditioned	−0.430	7.678	−39.511	0.00789	2.084	1.336
	(0.34)	(5.01)	(21.79)	(0.0065)	(0.61)	(0.061)
$60\Delta x$	−0.297	5.503	−28.312	0.00548	2.061	1.334
	(0.44)	(6.94)	(33.59)	(0.0083)	(0.60)	(0.061)
$40\Delta x$	−0.208	4.035	−20.280	0.00388	2.048	1.333
	(0.50)	(8.09)	(42.95)	(0.0093)	(0.60)	(0.061)
$20\Delta x$	−0.205	3.554	−8.259	0.00264	2.025	1.331
	(0.89)	(12.36)	(28.28)	(0.0141)	(0.59)	(0.061)
$10\Delta x$	−0.506	7.366	−18.725	0.00846	2.024	1.331
	(0.59)	(8.99)	(28.89)	(0.0098)	(0.59)	(0.061)

mesh points, the estimated drift coefficient becomes positive near the maximum in the data.

The parameter estimates for various levels of this displacement are recorded in Table 2. Table 2 reveals that the signs of the parameter estimates do not change from the case in which there is no conditioning to the cases in which there is conditioning. The estimates of the drift parameters α_1 and α_3 are always positive, while those of α_0 and α_2 are always negative. The magnitudes of some of the parameter estimates, however, change considerably. For example, the estimate of α_2, the coefficient on x^2, changes from -39.51 in the case of no conditioning to -8.26 when we condition on the event that the interest rate process stays between $a = 0.02915 - 20\Delta x$ and $b = 0.24333 + 20\Delta x$, where 0.02915 and 0.24333 are the minimum and maximum interest rates in the sample. While the changes in the other parameter estimates are not as large, all of the estimates change by a factor of more than two. (However, some caution is warranted in interpreting these results, as the standard errors of the estimates of the drift parameters α_i are about the same magnitudes as the parameter estimates.)

If the specification of the drift were linear (i.e., if α_2 and α_3 were forced to be equal to zero), then we would expect a negative α_1 parameter so that when the process becomes either large or small, the drift would pull it back to intermediate values. The signs of the α_2 and α_3 parameters, however, ensure that when the process becomes large or small, the drift pulls the process towards intermediate values. For large values of the interest rate, the negative α_2 parameter dominates the drift coefficient and pulls the process lower. For small values of the interest rate, the positive α_3 parameter dominates the drift and pushes the process higher.

The estimates of the diffusion parameters β_1 and β_2 change very little as a result of conditioning. This finding is expected, since the unconditioned and conditioned diffusion coefficients are the same (see Eq. (5)). Also, the standard errors indicate that these parameters are estimated reasonably precisely, with the estimate of β_1 always greater than three times its standard error and that of β_2 always greater than 20 times its standard error.

Many standard interest rate models assume that the drift coefficient is linear (see, e.g., Vasicek, 1977; Cox et al., 1985; CKLS), and an important strand of recent literature (AS; CHLS; Stanton, 1997; Pritsker, 1998; Chapman and Pearson, 2000) is concerned with the validity of this assumption. Consequently, we re-estimate the CHLS specification without conditioning and with conditioning under the restriction that $\alpha_2 = \alpha_3 = 0$ so that the drift function is linear. When performing this estimation, we still use the full set of six moment conditions but now define $\theta_R \equiv (\alpha_0, \alpha_1, 0, 0, \beta_1, \beta_2)$. We estimate θ_R by solving the minimization problem

$$H_R = \frac{1}{N-1} \min_{\theta_R} h(\theta_R)^\top W(\theta_R) h(\theta_R), \tag{75}$$

where the h and W functions are defined in Section 2., and H_R is a restricted version of the H which solves the minimization problem (71).

Table 3 reports the results for the case of no conditioning and for conditioning with displacements of 60, 40, 20, and 10 mesh points. The final column of Table 3 reports the difference between the restricted and unrestricted objective functions, $H_R - H$, where the unrestricted objective function H is given by (71) and is identically zero at the unrestricted parameter estimates. The asymptotic distribution of this statistic is Chi-squared with two degrees of freedom. Subject to the limits of the asymptotic approximation, this statistic can be used to test for the linearity of the drift of the unconditioned process; i.e., $\alpha_2 = \alpha_3 = 0$.

The unconditioned results in the first line of Table 3 reveal that the estimate of α_0 is greater than zero and that of α_1 is less than zero, consistent with mean reversion in the interest rate process. The estimates of α_0 and α_1 are both relatively imprecise; the estimate of α_0 is less than twice its standard error, while the absolute value of the estimate of α_1 is less than its standard error. The $H_R - H$ statistic testing linearity, distributed (asymptotically) χ_2^2, has a value of 4.650. Larger values of $H_R - H$ indicate that it is less likely that the unconditioned drift coefficient is linear, because they imply that there is relatively more difficulty satisfying the moment conditions when the drift of the unconditioned process is forced to be linear. The 5% and 10% critical values of the χ_2^2 distribution are 5.991 and 4.605, and 4.650 has a p-value of 0.0978, or 9.78%. Thus, the unconditioned results indicate a "marginal" rejection of linearity using conventional significance levels.

The conditioned results are markedly different, for all four displacements. Now the point estimates of α_0 are negative and those of α_1 are positive, suggesting an explosive process. However, these estimates are imprecise, with most having absolute values less than twice their standard errors. Even if the estimated parameters do

Table 3

Restricted conditional estimates Restricted parameter estimates obtained using the sample of seven-day Eurodollar rates, June 1, 1973 to February 24, 1995, conditioning on the event that the process remains in the interval (a, b) until February 24, 1995, for various choices of a and b. The lower boundary a is 0.02915 − displacement, the upper boundary b is 0.24333 + displacement, and 0.02915 and 0.24333 are the minimum and maximum interest rates observed in the sample. The displacement is expressed in terms of the number of spatial steps, each of size $\Delta x = 4.2836$ basis points. The first column lists the displacement, and other columns show the parameter estimates obtained by imposing the restriction $\alpha_2 = \alpha_3 = 0$ and minimizing the quadratic form H_R in equation (75). Numbers in parentheses are the corresponding standard errors. For comparison, the table also shows the parameter estimates for the case in which there is no conditioning. The last column shows the test statistic $H_R - H$, which is asymptotically χ^2 with two degrees of freedom.

Displacement	α_0	α_1	α_2	α_3	β_1	β_2	$H_R - H$
Unconditioned	0.0313	−0.221	0.0	0.0	2.040	1.335	4.650
	(0.017)	(0.30)			(0.59)	(0.060)	
$60\Delta x$	−0.0168	1.339	0.0	0.0	2.084	1.336	0.280
	(0.022)	(1.15)			(0.59)	(0.059)	
$40\Delta x$	−0.0206	1.523	0.0	0.0	2.062	1.334	0.076
	(0.018)	(1.13)			(0.58)	(0.059)	
$20\Delta x$	−0.0347	2.011	0.0	0.0	2.022	1.331	0.006
	(0.027)	(1.50)			(0.57)	(0.059)	
$10\Delta x$	−0.0498	2.563	0.0	0.0	2.087	1.340	0.807
	(0.020)	(1.14)			(0.63)	(0.061)	

correspond to processes that are explosive when unconditioned, the conditioning under which they are estimated ensures that they are non-explosive within the interval (a, b). There is, of course, no guarantee that the processes are properly defined if they are estimated when conditioning on the interval (a, b) and then employed over a different interval (a', b').

Perhaps the most striking feature of the results in Table 3 is that, for all of the conditioned results, the test statistics $H_R - H$ are very small, indicating that there is no difficulty satisfying the moment conditions when the drift of the unconditioned process is forced to be linear. Therefore, there is no evidence of nonlinearity. The p-value of the largest of the test statistics, 0.807, is 0.668, indicating that there is a 66.8% chance of obtaining a larger test statistic even if the null hypothesis of linearity is true. The p-values of the other test statistics are between 86.9% and 99.7%. While the χ^2 distribution of the test statistic $H_R - H$ is an asymptotic result and the small sample properties of this statistic are unknown in the present context, these test statistics are so small that it seems unlikely that this caveat is important. Thus, when one conditions on the event that the process remains between upper and lower barriers a and b, there is no evidence of nonlinearity.

These results fail to reject linearity in the drift of the unconditioned process because the conditioned drift is nonlinear even when the unconditioned drift is linear. As noted above, in the data the process decreased from its maximum value and increased from its minimum value. As a result, when conditioning is not taken into account, the estimation procedure "wants" the drift to be nonlinear in order to accommodate the behavior of the data near the minimum and maximum. When we condition on remaining within a and b, however, the conditioned drift accommodates the behavior of the data without making the unconditioned drift nonlinear.

The different estimates of α_0 and α_1 obtained for different displacements (i.e., different levels of a and b) in Table 3 indicate that the conditioning can have important effects on the parameter estimates. Fig. 5 illustrates this by plotting the five drift coefficients computed using the five sets of parameter estimates in Table 3. As in the unrestricted case, conditioning appears to have a substantial impact on the estimation of the drift coefficient that is restricted to be linear. Without conditioning, the estimate of the unconditional drift has a negative slope. With even mild conditioning (i.e., a displacement of 60 mesh points), the slope of the drift coefficient becomes positive, and increases as the displacement is reduced to 40 mesh points or fewer. This might seem problematic, since it implies an explosive process. For the cases plotted in Fig. 5, the slope of the drift coefficient is strictly increasing in the severity of the conditioning. This occurs for the same reason that in the unrestricted case the value of the drift coefficient increases with the severity of the conditioning near the maximum of the interest rate data. For both the unrestricted and the restricted cases, conditioning produces more of a change near the maximum of the data than it does near the minimum of the data. This is because in both cases the diffusion coefficient is greater near the maximum than near the minimum of the data.

Figs. 4 and 5 make it clear that the strength of the conditioning can have an important impact on the estimated drift coefficient, and it should be kept in mind that the strength of the conditioning involves a subjective choice on the part of the researcher. In the case at hand, the results in Chapman and Pearson (2000) argue for tighter conditioning. In other contexts, the researcher may have less guidance. It should also be kept in mind, however, that some conclusions may not depend on the strength of the conditioning. For example, the results in Table 3 fail to reject linearity regardless of which of the four conditioning events is chosen.

5.3. *Is the drift coefficient linear or nonlinear?*

Once we condition on the event that the sample path remains between the lower and upper bounds a and b, the test statistics in Table 3 fail to reject the linearity of the drift coefficient by a wide margin. Should we interpret this as evidence that the drift coefficient is linear?

Even aside from the obvious point that failing to reject a hypothesis is not the same as accepting it, the answer is no. The analysis in Section 5.1 above shows that once one conditions on the lower and upper bounds a and b, even if the drift coefficient of the original unconditioned process is linear, the drift of the conditioned process is nonlinear and similar to the nonlinear drift coefficients that have been estimated by other authors. The analysis in Section 2 also suggests that if the drift coefficient of the original unconditioned process is nonlinear, then the drift of the conditioned process will be less affected by the conditioning. In particular, Eq. (4) indicates that the effect of conditioning on the drift depends on the probability that the unconditioned process satisfies the conditioning event and the spatial derivative

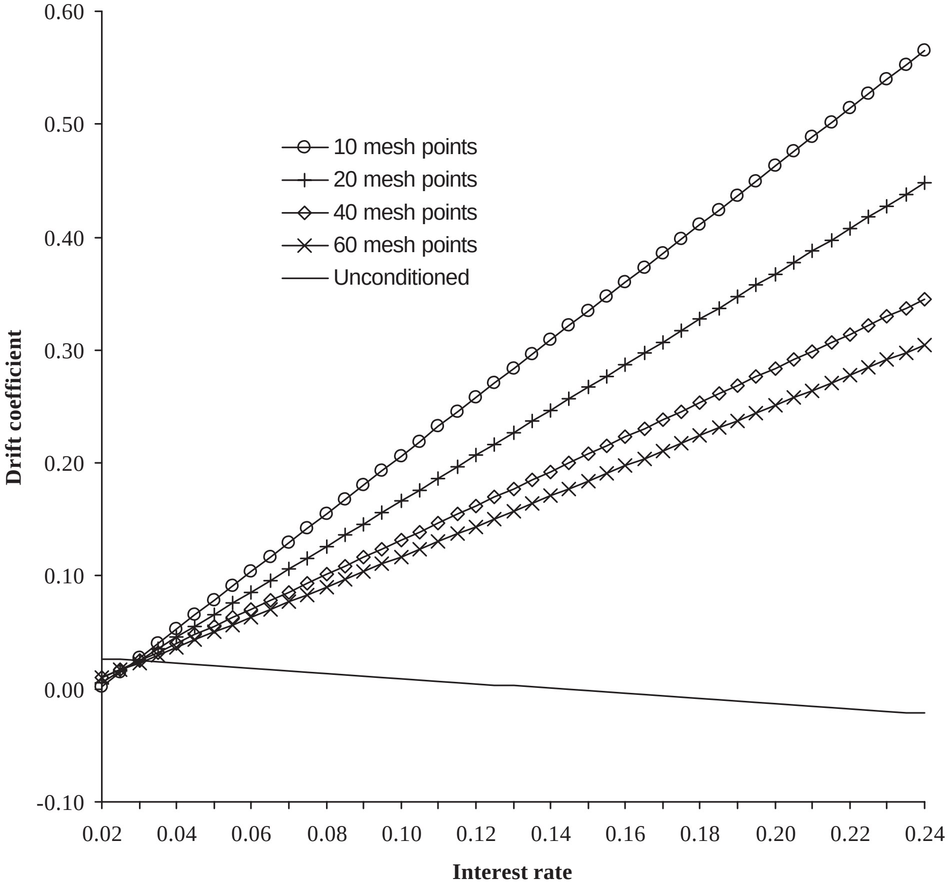

Fig. 5. Linear drift coefficients estimated from seven-day Eurodollar rates, June 1, 1973 to February 24, 1995, conditioning on the event that the process remains in the interval (a, b) until February 24, 1995, for various choices of a and b. The drift coefficient $\mu(x) = \alpha_0 + \alpha_1 x$ and diffusion coefficient $\sigma^2(x) = \beta_1 x^{2\beta_2}$ are estimated from the Eurodollar rate data using the generalized method of moments, and the moment conditions are constructed using one-trade-date conditioned expected changes. The solid line is the estimate of the unconditioned drift coefficient when on each trade date there is no assumption made about the minimum and maximum values the process will attain until February 24, 1995. The other lines are drift estimates obtained when on each trade the one-trade-date expected change is computed conditional on the process remaining between $a = 0.02915 -$ displacement and $b = 0.24333 +$ displacement for the remainder of the time until February 24, 1995, where 0.02915 and 0.24333 are the minimum and maximum interest rates observed in the sample. The displacement is expressed in terms of the number of spatial steps, each of size $\Delta x = 4.2836$ basis points. The line with symbols $\times$ is the drift estimate when the conditioning is done using a displacement of $60\Delta x$ from the observed minimum and maximum. The line with diamonds is for a displacement of $40\Delta x$, the line with symbols $+$ is for a displacement of $20\Delta x$, and the line with circles is for a displacement of $10\Delta x$. Rates are expressed on an annual basis.

of that probability. An unconditioned process with a nonlinear drift coefficient that creates strong mean reversion will have a high probability of remaining within the interval (a, b), and the spatial derivative will typically be smaller in magnitude than when the unconditioned process is linear. This implies that when the original unconditioned process has a strongly nonlinear drift, the conditioning will have a smaller effect, and also suggests that after conditioning, processes with linear and nonlinear drift coefficients might look very similar. As a result, it would not be surprising if it is difficult for statistical tests to distinguish between linear and nonlinear drifts in the original unconditioned process.

A small Monte Carlo study shows that this is indeed the case. Specifically, we generate 100 simulated paths from a process with drift parameters $\alpha = [-0.4296, 7.6782, -39.5109, 0.007892]$ and $\beta = [2.0839, 1.3363]$. These parameters are the estimates reported in the first row of Table 2 which are obtained from the actual short-rate path when we do not condition on any interval (a, b), and produce the unconditioned drift coefficient shown in Fig. 4. This drift coefficient is similar to the nonlinear drift coefficients estimated by other authors. For each simulated sample path and each level of displacement used in Table 2 and Fig. 4, we estimate the restricted model with $\alpha = [\alpha_0, \alpha_1, 0, 0]$ and $\beta = [\beta_1, \beta_2]$ and compute the χ^2_2 test statistic for the linearity restriction $\alpha_2 = \alpha_3 = 0$. For the displacement of $100\Delta x \approx 428$ basis points (that is, conditioning on the event that the sample path remains between $a = x^{\min} - 100\Delta x$ and $b = x^{\max} + 100\Delta x$, where $x^{\min}$ and $x^{\max}$ are the minimum and maximum along the sample path), only 10 (that is, 10%) of the χ^2_2 statistics exceed the 5% critical value of 5.991, and only two (2%) exceeded the 1% critical value of 9.210. For displacements of $60\Delta x$, $40\Delta x$, $20\Delta x$, and $10\Delta x$, only seven, five, six, and five of the χ^2_2 statistics exceed the 5% critical value, and only two, two, one, and zero exceed the 1% critical value. These results indicate that it is very difficult to make inferences about whether the drift is nonlinear.[3] If one estimates the diffusion process without conditioning, the estimates are subject to the bias documented in Chapman and Pearson (2000) and Section 4 above; however once one conditions the test has little power to detect the sort of nonlinearity in the drift that has been suggested by other authors.

Given this difficulty in determining from the short-rate data alone whether the drift coefficient is nonlinear, it is worth pointing out that the nonlinear drifts proposed by Stanton (1997), AS, and CHLS are compatible with a model in which the central bank acts to return the short rate back to intermediate levels whenever it gets very low or very high. See AS for a discussion of this point. In addition, these nonlinear drifts have the empirically appealing property of being locally nonstationary but globally stationary.

6. Conclusion

There are a number of circumstances in finance in which it is desirable to estimate a data-generating process conditional on the occurrence of some event. One context where conditional estimation should be performed is situations in which only data that satisfy some criteria are accessible even though the underlying process is capable

---

[3] Jones (2003) reaches a similar conclusion within a Bayesian framework.

of generating data that violate the criteria. The survivorship literature begun by Brown et al. (1995) is an important example of such a context. A second context where performing conditional estimation is appropriate is situations in which any finite data sample misrepresents in some specific and important way the underlying population from which it is drawn. A recent example of this is the estimation of short-rate models from time series of interest rate data, as explained by Chapman and Pearson (2000).

This paper develops tools that allow a researcher to estimate a multivariate diffusion process conditional on the event that the process remains in an open, connected, and bounded region $\mathcal{G} \subset \Re^d$. We show that the probabilities of the conditioning event as well as the finite time-conditioned expected change satisfy parabolic partial differential equations, and we investigate several univariate examples. Computations for these examples reveal that near the boundaries there can be large differences between the conditioned and unconditioned expected changes. This fact is important because the true conditioned expected change in the value of the process over a time period equal to the interval between observations of the data at hand is needed for moment-based estimation procedures. The effect on the expected change also has implications for the transition density, and thus has implications for maximum likelihood estimation.

To illustrate our procedure, we use our tools to estimate a flexibly specified, time-homogeneous univariate diffusion model of the short rate conditional on the process remaining between lower and upper bounds. The results indicate that in this context the conditioning has an important impact on the estimated drift coefficient but little effect on the estimated diffusion coefficient. For values of the interest rate near the lower bound, the conditioning decreases the estimate of the drift coefficient; for values of the interest rate near the upper bound, the conditioning increases the estimate of the drift coefficient. Thus, the conditioning produces an estimate of the underlying drift coefficient that is more linear than that which is obtained in the absence of conditioning. Since recently there has been considerable interest in the linearity of the drift coefficient of univariate diffusion term-structure models, the estimation is also repeated with the underlying drift coefficient restricted to a linear specification. A test statistic fails by a wide margin to reject the linearity of the diffusion coefficient of the underlying process.

References

Abhyankar, A., Basu, D., 2001. Does conditioning information matter in estimating continuous time interest rate diffusions? Journal of Financial and Quantitative Analysis 36, 335–344.

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

Aït-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. Journal of Finance 54, 1361–1395.

Aït-Sahalia, Y., 2002a. Maximum-likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica 70, 223–262.

Aït-Sahalia, Y., 2002b. Closed-form likelihood expansions for multivariate diffusions. Unpublished working paper, Princeton University.

Berger, J.O., 1986. Commentary on papers [67], [69], [70], [71]. In: Brown, L.D., Olkin, I., Sacks, J., Wynn, H.P. (Eds.), Jack Carl Kiefer Collected Papers: Supplementary Volume. Springer, New York, pp. 48–56.

Berger, J.O., Boukai, B., Wang, Y., 1997. Unified frequentist and Bayesian testing of a precise hypothesis. Statistical Science 12, 133–160.

Brown, L.D., 1978. A contribution to Kiefer's theory of conditional confidence procedures. Annals of Statistics 6, 59–71.

Brown, S.J., Goetzmann, W.N., Ross, S.A., 1995. Survival. Journal of Finance 50, 853–873.

Brownie, C., Kiefer, J., 1977. The ideas of conditional confidence in the simplest setting. Communications in Statistics Part A – Theory and Methods 6, 691–751.

Chan, K.C., Karolyi, A., Longstaff, F.A., Sanders, A.B., 1992. An empirical comparison of alternative models of the short-term interest rate. Journal of Finance 47, 1209–1227.

Chapman, D.A., Pearson, N.D., 2000. Is the short rate drift actually nonlinear? Journal of Finance 55, 355–388.

Conley, T.C., Hansen, L.P., Luttmer, E.G.J., Scheinkman, J., 1997. Short-term interest rates as subordinated diffusions. Review of Financial Studies 10, 525–577.

Cox, J.C., Ingersoll, J., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407.

Douady, R., 1998. Closed-form formulas for exotic options and their lifetime distribution. International Journal of Theoretical and Applied Finance 2, 17–42.

Goetzmann, W., Jorion, P., 1993. Testing the predictive power of dividend yields. Journal of Finance 48, 663–679.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Jones, C.S., 2003. Nonlinear mean reversion in the short-term interest rate. Review of Financial Studies 16, 793–843.

Karlin, S., Taylor, H.M., 1975. A First Course in Stochastic Processes. Academic Press, New York.

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

Kiefer, J., 1975. Conditional confidence approach in multidecision problems. In: Krishnaian, P.R. (Ed.), Multivariate Analysis IV. Academic Press, New York.

Kiefer, J., 1976. Admissibility of conditional confidence procedures. Annals of Mathematical Statistics 4, 836–865.

Kiefer, J., 1977. Conditional confidence statements and confidence estimators (with discussion). Journal of the American Statistical Association 72, 789–827.

Kunitomo, N., Ikeda, M., 1992. Pricing options with curved boundaries. Mathematical Finance 2, 275–298.

Pinsky, R.G., 1985. On the convergence of diffusion processes conditioned to remain in a bounded region for large time to limiting positive recurrent diffusion processes. The Annals of Probability 13, 363–378.

Pritsker, M., 1998. Nonparametric density estimation and tests of continuous time interest rate models. Review of Financial Studies 11, 449–487.

Stanton, R., 1997. A nonparametric model of term structure dynamics. Journal of Finance 52, 1973–2002.

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.