Facing Up to Conditioned Diffusions

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

Minqiang Li
Department of Finance
University of Illinois at Urbana-Champaign

Neil D. Pearson
Department of Finance
University of Illinois at Urbana-Champaign

and

Allen M. Poteshman
Department of Finance
University of Illinois at Urbana-Champaign

OFOR Paper Number 01-01
April 2001

Facing Up to Conditioned Diffusions

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

30 March 2001

Abstract

Most data used in finance are generated naturally rather than experimentally. While researchers are typically interested in estimates of model parameters that are not conditional on the particular sample, actual estimates are necessarily conditional on the data. Recent research on survivorship bias in equity returns and the estimation of term structure models from time-series of interest rate data suggests that failing to account for the implicit conditioning can seriously bias the results of empirical research. This paper develops theoretical and numerical tools that make it possible to account for the implicit conditioning when the underlying data are generated by a time-homogeneous univariate diffusion, and carries out a detailed analysis for three specific conditioning events that are of interest in finance. The techniques are illustrated by obtaining estimates of the drift and diffusion coefficients of a term-structure model from a standard time-series of interest rate data both with and without conditioning on these three events. The estimates indicate that the conditioning events have an important impact on the estimated drift coefficient but little effect on the estimated diffusion coefficient. A test statistic fails to reject linearity of the drift coefficient of the short rate process regardless of which of the conditioning events is assumed.

*Ph.D. candidate, Associate Professor, and Assistant Professor, respectively, University of Illinois, Department of Finance, 1206 South Sixth Street, Champaign, Illinois 61820. E-mail mli1@uiuc.edu, pearson2@uiuc.edu, and poteshma@uiuc.edu, respectively. We thank Yacine Aït-Sahalia for making available the interest rate data used in this paper, and Reza Mahani for performing some of the preliminary computations. The bulk 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.

Facing Up to Conditioned Diffusions

Abstract

Most data used in finance are generated naturally rather than experimentally. While researchers are typically interested in estimates of model parameters that are not conditional on the particular sample, actual estimates are necessarily conditional on the data. Recent research on survivorship bias in equity returns and the estimation of term structure models from time-series of interest rate data suggests that failing to account for the implicit conditioning can seriously bias the results of empirical research. This paper develops theoretical and numerical tools that make it possible to account for the implicit conditioning when the underlying data are generated by a time-homogeneous univariate diffusion, and carries out a detailed analysis for three specific conditioning events that are of interest in finance. The techniques are illustrated by obtaining estimates of the drift and diffusion coefficients of a term-structure model from a standard time-series of interest rate data both with and without conditioning on these three events. The estimates indicate that the conditioning events have an important impact on the estimated drift coefficient but little effect on the estimated diffusion coefficient. A test statistic fails to reject linearity of the drift coefficient of the short rate process regardless of which of the conditioning events is assumed.

1 Introduction

A central challenge for the great majority of empirical work in finance is that the data are generated naturally rather than experimentally. Typically, the researcher observes one historical draw from some data generating process which is then used to estimate a model of the process. While estimates of the parameters of the model which are not conditional on the particular historic sample are often of primary interest, the actual estimates produced by the researcher are necessarily conditional on the observed data. The conditional estimates will deviate from the unconditional estimates insofar as: (1) the observed sample is not fully representative of the underlying population from which it is drawn; and (2) the researcher does not adjust for the less than fully representative nature of the data at hand.

Survivorship bias is one example of this problem. Brown, Goetzmann, and Ross (1995) analyze the case of survivorship bias in equity returns in which a historical sample of equity returns is not representative of the underlying population, because it only includes firms or stock markets which have managed to survive until a certain point in time. In their analysis they assume that the (unconditioned) stock price process is a geometric Brownian motion, but that the analyst or econometrician observes only the process conditioned on survival. They find the relation between the drifts (i.e., expected returns) of the conditioned and unconditioned processes, and argue that the conditioning bias can be significant in interpreting such diverse phenomenon as the equity premium, long-term autocorrelation studies, "post-announcement drift" following earnings announcements, and stock split studies. More recently, simulations in Goetzmann and Jorion (1999) suggest that the survival bias can be significant in estimating the expected returns of "emerging markets."

It also appears that failing to take conditioning into account when analyzing time-series of interest rate data can lead to significant biases. Recently, an important strand of the large literature on estimating term structure models from interest rate data has employed models that are univariate diffusions with drift coefficients that are specified either as flexible parametric forms (Aït-Sahalia (1996, henceforth AS), Conley, Hansen, Luttmer, and Scheinkman (1997, henceforth CHLS)) or nonparametrically (Stanton (1997)). These papers suggest that there are important non-linearities in the drift coefficient. However, Pritsker (1998) raises questions about the finite sample performance of the test for nonlinearity used in AS, and the simulations in Chapman and Pearson (2000) provide evidence that the apparent nonlinearities may arise from features of the time-series of interest rate data that are essentially historical accidents such as the particular minimum and maximum observed. The same type of conditioning bias can also arise in more standard term structure models where the drift coefficient is less flexibly specified. For example, Abhyankar and Basu (2000) 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 $\bar{b}$, and also

the drift of a Brownian motion process conditioned to remain in an interval $(\bar{a}, \bar{b})$. They find that the conditioned drifts are nonlinear, even though the original unconditioned processes have linear drifts.

Since empirical work in finance is largely non-experimental, conditioning biases may well play an important role not just in the study of equity returns and interest rates but across nearly the entire discipline. Over the last thirty years continuous-time diffusion models have come to occupy an important place in the theory and practice of finance (see, e.g., Sundaresan (2000)). In this paper we show how to compute for various conditioning events the conditioned drift and diffusion coefficients as well as the mean over a finite interval (which is important for estimation) for general univariate diffusion processes. When the unconditioned process is geometric Brownian motion, these quantities can be computed explicitly for simple forms of conditioning. We show that, in general, the computations require the solution of parabolic partial differential equations subject to various boundary conditions, and describe the numerical tools to solve these equations. The ability to compute these quantities makes it possible to remove the effect of the types of conditioning that we consider when estimating univariate diffusion processes from observed historical data. The general method we develop can be applied to a wide variety of conditioning events.

We illustrate our analysis by estimating a flexibly specified diffusion model from a standard time-series of interest rate data subject to several types of conditioning that are of particular interest for finance, namely, that over a specific period of time: (1) a continuously monitored process stays between upper and lower boundaries, (2) a continuously monitored process has specific minimum and maximum values, and (3) a process monitored at a set of discrete dates remains between minimum and maximum values. We find that each of these conditions has an important impact on the estimate of the unconditioned drift coefficient but that none has much of an influence on the estimate of the diffusion coefficient. A test statistic fails to reject linearity of the drift coefficient by a wide margin for all three types of conditioning.

The remainder of the paper is organized as follows. Section 2 of the paper carries out two small Monte Carlo experiments in order to illustrate the bias that can result from implicitly conditioning either on staying within certain minimum and maximum values or achieving specific minimum and maximum values when estimating a model from a time series of data. The first experiment illustrates the severe bias that can result from the fact that the data that comprise any historical sample necessarily are greater than or equal to its minimum and less than or equal to its maximum, even though over a fixed time interval the underlying process can produce paths with lesser minima and/or greater maxima. The second experiment shows that a significant bias also can arise from the fact that any historical sample necessarily has its least value at its minimum and its greatest value at its maximum even though the underlying process can produce paths with a wide variety of minima and maxima. Section 3 derives expressions for the conditioned drift and diffusion coefficients

of a general univariate 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 conditioned mean satisfy parabolic partial differential equations with boundary conditions that depend upon the conditioning events. Sections 4 through 6 then analyze three specific conditioning events that arise in finance. Section 4 develops the numerical tools to solve the partial differential equations when the conditioning event is that the diffusion remains between an upper and lower boundary for a specified amount of time, and applies the analysis to estimate specifications of the drift and diffusion coefficients of an interest rate process, given that the econometrician only observes a time-series that satisfies the conditioning event. Because the upper boundary may be infinity, this conditioning event covers the cases considered in the survival literature. Section 5 develops the necessary tools when the conditioning event is that the process has a particular minimum and maximum over some specified amount of time, and again applies the analysis to the estimation of an interest rate process. Section 6 treats the case where the conditioning event is that the process observed at a set of discrete times remains between upper and lower boundaries. Section 7 briefly concludes.

2 Conditioning Bias when Estimating a Univariate Diffusion

A problem that has received a great deal of attention in the term-structure literature is that of estimating a univariate, time-homogeneous diffusion process from a time-series of interest rate data. These diffusions can be specified by

$$dx(t) = \mu(x(t))dt + \sigma(x(t))dB(t), \tag{1}$$

where μ and σ 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 AS (1996) and CHLS (1997). The drift and diffusion coefficients of the CHLS specification are given by

$$\begin{aligned} \mu(x) &= \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x, & (2)\\ \sigma^2(x) &= \beta_1 x^{2\beta_2}, & (3) \end{aligned}$$

where the 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 severe conditioning bias that arises because any observed time series of interest rate data necessarily does not go below its minimum value or above its maximum value,

we perform the following simulation experiment. First, artificial interest rate series are generated from the CHLS specification with α_2 and α_3 restricted to be zero so that the drift coefficient is linear. The first 1000 of these paths whose minimum and maximum stay within two predetermined values are retained. Next, these 1000 paths are used to estimate the CHLS specification without any restriction on α_2 or α_3. Finally, the estimated drift and diffusion coefficients are compared to the drift and diffusion coefficients which generated the data.

We calibrate the CHLS specification using Hansen's (1982) generalized method of moments procedure from the daily time-series of seven day Eurodollar spot rates (bid-ask midpoints) used in AS. This time-series is 5505 days long and covers the period from 1973 to 1995. Its minimum value is 0.02915 (which occured in early 1993) and its maximum value is 0.24333 (which occured in 1981). Figure (1) shows the actual interest rate path we used.

Suppose we have data at equally spaced times t_i, $i = 1, \ldots, N$, and 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 zero, a first-order Euler approximation of the CHLS specification is given by

$$\delta x_i = (\alpha_0 + \alpha_1 x(t_i))(\delta t) + \sqrt{\beta_1} x(t_i)^{\beta_2} \sqrt{\delta t}\, \epsilon_i \tag{4}$$

where $\epsilon_i \sim N(0,1)$. Dividing both sides by $x(t_i)^{\beta_2}$ (which corresponds to a correction for heteroscedasticity) results in

$$x(t_i)^{-\beta_2}(\delta x_i) = x(t_i)^{-\beta_2}(\alpha_0 + \alpha_1 x(t_i))(\delta t) + \sqrt{\beta_1}\sqrt{\delta t}\, \epsilon_i. \tag{5}$$

Using x_i to denote $x(t_i)$ and defining $\tilde{u}_i \equiv x_i^{-\beta_2}(\delta x_i) - x_i^{-\beta_2}(\alpha_0 + \alpha_1 x_i)(\delta t)$, this suggests that the parameter vector $\tilde{\theta} = (\alpha_0, \alpha_1, \beta_1, \beta_2)$ be estimated from the moment conditions

$$E \begin{bmatrix} \tilde{u}_i \\ x_i \tilde{u}_i \\ \tilde{u}_i^2 - \beta_1(\delta t) \\ x_i \tilde{u}_i^2 - x_i \beta_1 (\delta t) \end{bmatrix} = 0. \tag{6}$$

These moment conditions are those used by Chan, Karolyi, Longstaff, and Sanders (1992, henceforth CKLS), except that CKLS do not divide by $x_i^{\beta_2}$.

To implement the moment conditions, we define

$$\tilde{g}_i(\tilde{\theta}) \equiv \begin{bmatrix} \tilde{u}_i \\ x_i \tilde{u}_i \\ \tilde{u}_i^2 - \beta_1(\delta t) \\ x_i \tilde{u}_i^2 - x_i \beta_1 (\delta t) \end{bmatrix} \tag{7}$$

and

$$\tilde{h}(\tilde{\theta}) \equiv \frac{1}{N-1} \sum_{i=1}^{N-1} \tilde{g}_i(\tilde{\theta}), \tag{8}$$

so that $\tilde{h}$ is the sample analogue of the left-hand side of (6). The standard GMM approach is that the estimates are the solution of

$$\widetilde{G} = \frac{1}{N-1} \min_{\tilde{\theta}} \tilde{h}(\tilde{\theta})' \widetilde{W} \tilde{h}(\tilde{\theta}), \tag{9}$$

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

$$\widetilde{S}(\tilde{\theta}) = \frac{1}{N-2} \sum_{i=1}^{N-1} \left[(\tilde{g}_i(\tilde{\theta}) - h(\tilde{\theta}))(\tilde{g}_i(\tilde{\theta}) - h(\tilde{\theta}))' \right]. \tag{10}$$

Applying this GMM procedure to the AS data yields parameter estimates of $\alpha_0 = 0.063$, $\alpha_1 = -0.74$, $\beta_1 = 2.08$, and $\beta_2 = 1.34$.

The initial value in the AS data is 0.07984. Paths are repeatedly simulated from the CHLS specification using a Milstein scheme from the starting value 0.07984 using the parameter estimates above (and with $\alpha_2 = 0$ and $\alpha_3 = 0$). Each path is 5,505 trade dates long, and the simulation is continued until 1000 paths are obtained which have a minimum greater than the AS data minimum of 0.02915 and a maximum less than the AS data maximum of 0.24333.

The CHLS specification is estimated for each of the 1000 paths using a GMM procedure similar to the one just described except that α_2 and α_3 are not restricted to be zero. Accordingly the parameter vector to be estimated becomes $\theta = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)$ and the moment conditions become:

$$E \begin{bmatrix} u_i \\ x_i u_i \\ x_i^2 u_i \\ x_i^{-1} u_i \\ u_i^2 - \beta_1(\delta t) \\ \\ x_i u_i^2 - x_i \beta_1(\delta t) \end{bmatrix} = 0, \tag{11}$$

where now $u_i \equiv x_i^{-\beta_2}(\delta x_i) - x_i^{-\beta_2}(\alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \alpha_3 / x_i)(\delta t)$. Defining $g_i(\theta)$ to be

[1] Since the number of moment conditions is equal to the number of parameters, the same estimates would result if $\widetilde{W}$ were set equal to the identity matrix.

$$g_i(\theta) \equiv \begin{bmatrix} u_i \\ x_i u_i \\ x_i^2 u_i \\ x_i^{-1} u_i \\ u_i^2 - \beta_1(\delta t) \\ x_i u_i^2 - x_i \beta_1(\delta t) \end{bmatrix}, \tag{12}$$

the sample analogue of the left-hand side of (11) becomes

$$h(\theta) \equiv \frac{1}{N-1} \sum_{i=1}^{N-1} g_i(\theta). \tag{13}$$

The weighting matrix is changed accordingly into a 6×6 matrix W. The objective function is now:

$$G = \frac{1}{N-1} \min_{\theta} h(\theta)' W h(\theta). \tag{14}$$

Except for these changes each of the 1000 simulated paths is estimated using the GMM procedure described above.

Figure 2 shows the results of the estimation exercise. The top graph reports on the estimation of the drift coefficient and the bottom graph reports on the estimation of the diffusion coefficient. In both of these graphs, the solid line is the true coefficient from which the simulated data were generated. The dashed line is the pointwise mean of the 1,000 estimated coefficients at each level of the interest rate. The dotted lines are the 25th and 75th percentiles across the 1,000 estimated coefficients at each level of the interest rate.

The top graph in Figure 2 shows that conditioning on the minimum and maximum values of the interest rate process leads to severe bias in the estimate of the drift coefficient. For example, at the maximum value of 0.24333, the true data-generating process has a drift coefficient with a value of about -0.1, but the mean of the estimated drift coefficients is less than -1.2. The 75th percentile of the estimated drift coefficients is still less than -0.8 at the 0.24333 interest rate level. The reason for the bias is that the true unconditional data-generating process can exceed 0.24333, but each of the simulated interest rate paths is conditioned to have a maximum less than 0.24333. Consequently, unconditionally the process can increase or decrease when it is at a value of 0.24333, but the conditional process can only decrease at its maximum value, which will be less than 0.24333. As a result, the conditional drift coefficient is negatively biased relative to the unconditional drift coefficient when the level of the interest rate is near the maximum. Conversely, when the process is at its minimum it can either increase or decrease, but the conditioned process can only increase. This produces the positive bias of the conditional drift coefficient relative to the unconditional drift coefficient seen on the left hand side of the top graph of Figure 2 when the level of interest rates

are low. The bias near the minimum value of the process is smaller than that 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 bottom graph in Figure 2 shows that conditioning so that the minimum value of the interest rate process is greater than 0.02915 and the maximum is less than 0.24333 leads to no appreciable bias in the estimate of the diffusion coefficient. In addition, the 25th and 75th percentile bands show that the diffusion coefficient can be estimated quite precisely from a daily time-series that is 5505 trade dates long. These results on estimating the diffusion coefficient are consistent with previous research such as Chapman and Pearson (2000).

Next a Monte Carlo experiment is performed where the paths are conditioned to have a specific minimum and maximum value. In this case, paths are generated from the same specification as in the previous experiment, but the first 1000 paths which have a minimum within 10 basis points of 0.02915 and a maximum within 10 basis points of 0.24333 are retained. The CHLS specification is estimated from these 1000 paths using the same GMM procedure as in the previous experiment. The results of the experiment are presented in Figure 3. The bias in the drift coefficient is substantial and in the same direction as in the previous experiment. At the same time, it is approximately half as large. To understand why the bias is larger in the previous experiment, consider the right hand side of Figure 3 where the interest rate is large, say 0.24. The previous experiment has a number of paths whose maximum values are less than 0.24. For these paths, the bias at the maximum value will be extrapolated out to 0.24. This extrapolation magnifies the bias which produces the difference between the drift coefficient plots in Figure 2 and Figure 3. As in the previous case, there is little bias in the estimation of the diffusion coefficient, and it is measured quite precisely.

3 Conditioning on a general event

This section of the paper develops the tools needed to estimate the unconditional drift and diffusion coefficients of a time-homogeneous univariate diffusion from a time-series of data that is observed conditional on some event A. (The event A could, for example, be that the process stays in between some lower and upper boundaries or that it has a particular minimum and maximum value.) A natural approach to this estimation problem is to implement a GMM procedure like that used in Section 2 but with the unconditional drift and diffusion coefficients in the Euler approximation replaced with their conditional counterparts. Following this approach, the Euler approximation becomes

$$\delta x_i = \mu(x(t_i), t_i \,|A)(\delta t) + \sigma(x(t_i), t_i \,|A)\sqrt{\delta t}\,\epsilon_i \tag{15}$$

where $\mu(x(t_i), t_i \,|A)$ and $\sigma(x(t_i), t_i \,|A)$ are, respectively, the drift and diffusion of the process conditional on the occurrence of event A. If the time-series of data consists of daily observations, experience indicates that the discretization bias introduced by the Euler approximation is small when performing unconditional estimation of the process.

We proceed by computing the transition density, drift coefficient, and diffusion coefficient of the conditioned process. The result for the conditioned drift shows how the conditioning bias is related to the diffusion coefficient and the probability of the conditioning event. The conditioned drift, however, is not directly useful for estimation, because the Euler approximation based on the conditioned drift performs poorly even at a daily time interval. As a result, we show how to compute the conditioned expected change in the process by solving a particular partial differential equation, and then carry out estimation using these conditioned expected changes.

3.1 The conditioned transition density

In order to compute the conditioned drift and diffusion coefficients of the process, it is necessary first to obtain the formula for the process' conditioned transition density. Let $f(x, t, y, s)$ be the 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 $f(x, t, y, s \,|A(t_1, t_2))$ be the 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 some event $A(t_1, t_2)$ between times t_1 and $t_2 > t_1$. (The event $A(t_1, t_2)$ could, for example, be that between the times t_1 and t_2 the process never reaches some level.) Then for $\delta t > 0$

$$f(x, t, y, t + \delta t \,|A(t, T))dy = P\left[x(t + \delta t) \in dy \mid x(t) = x, A(t, T)\right], \tag{16}$$

where the notation $x(t + \delta t) \in dy$ denotes $y \leq x(t + \delta t) \leq y + dy$. Recall that Bayes rule states

$$P[A \mid B, C] = \frac{P[A \mid B] \cdot P[C \mid B, A]}{P[C \mid B]}. \tag{17}$$

Applying Bayes rule to (16) yields

$$f(x, t, y, t+\delta t \,|A(t, T))dy = \frac{P[x(t + \delta t) \in dy \mid x(t) = x] \cdot P[A(t, T) \mid x(t) = x, x(t + \delta t) = y]}{P[A(t, T) \mid x(t) = x]}, \tag{18}$$

or

$$f(x, t, y, t + \delta t \,|A(t, T))dy = \frac{f(x, t, y, t + \delta t)P[A(t, T) \mid x(t) = x, x(t + \delta t) = y]dy}{P[A(t, T) \mid x(t) = x]}. \tag{19}$$

To simplify the term $P[A(t, T) \mid x(t) = x, x(t+\delta t) = y]$, we impose the following two properties on the conditioning event $A(t, t')$:

1. *$\mathcal{F}_{t'}$ measurability:* $A(t,t') \in \sigma(x(t''); t \le t'' \le t')$, i.e. at time t', it is known whether $A(t,t')$ is true.

2. *Semigroup property:* For any event B, $P[A(t,t') \mid B] = P[A(t,t'') \cap A(t'',t') \mid B]$ for all t'' such that $t \le t'' \le t'$.

Then when $T - t \ge \delta t$ these two properties imply

$$\begin{aligned}
&P[A(t,T) \mid x(t) = x, x(t+\delta t) = y] \\
&\quad = \; P[A(t,t+\delta t) \cap A(t+\delta t, T) \mid x(t) = x, x(t+\delta t) = y] \\
&\quad = \; P[A(t,t+\delta t) \mid x(t) = x, x(t+\delta t) = y] \\
&\qquad \times P[A(t+\delta t, T) \mid x(t) = x, x(t+\delta t) = y, A(t,t+\delta t)] \\
&\quad = \; P[A(t,t+\delta t) \mid x(t) = x, x(t+\delta t) = y] P[A(t+\delta t, T) \mid x(t+\delta t) = y], \qquad (20)
\end{aligned}$$

where in the second step we have used Bayes rule and in the last step we have used the Markov property of diffusion processes.

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

$$\bar{\pi}(x,t,y,t+\delta t; A(t,t+\delta t)) = P[A(t,t+\delta t) \mid x(t) = x, x(t+\delta t) = y], \qquad (21)$$

and also define $\pi(x,t; A(t_1,t_2))$ to be the probability that $A(t_1,t_2)$ is satisfied between times t_1 and $t_2 > t_1$ given that the process is at the value of x at time t:

$$\pi(x,t; A(t_1,t_2)) \equiv P[A(t_1,t_2) \mid x(t) = x]. \qquad (22)$$

Then the second term of the numerator of (19) becomes

$$P[A(t,T) \mid x(t) = x, x(t+\delta t) = y] = \bar{\pi}(x,t,y,t+\delta t; A(t,t+\delta t)) \pi(y,t+\delta t; A(t+\delta t,T)). \qquad (23)$$

Recognizing that the denominator of (19) is just $\pi(x,t; A(t,T))$ and substituting this and (23) into (19) yields

$$f(x,t,y,t+\delta t \,|A(t,T))\, dy = \frac{\bar{\pi}(x,t,y,t+\delta t; A(t,t+\delta t))\, \pi(y,t+\delta t; A(t+\delta t,T))\, f(x,t,y,t+\delta t)\, dy}{\pi(x,t; A(t,T))}. \qquad (24)$$

Note that by the definition of $f(x,t,y,t+\delta t \,|A(t,T))$, we have

$$\int_0^\infty f(x,t,y,t+\delta t\,|A(t,T))\,dy = 1. \tag{25}$$

Equation (24) differs from equation (9.17) on page 267 of Karlin and Taylor (1981) insofar as the term $\bar{\pi}(x,t,y,t+\delta t; A(t,t+\delta t))$ is missing from (9.17).[2] This omission does not affect the computation of the conditioned drift in Karlin and Taylor, because in the diffusion limit $\delta t \to 0$ the term $\bar{\pi}(x,t,y,t+\delta t; A(t,t+\delta t)) \to 1$. However, it is crucial to retain this term in the computation of the conditioned mean over a finite time step δt which we carry out below.

Finally, one more fact about the quantity $\pi(x,t;A(t,T))$ is needed below. The law of total probability can be used to write the probability $\pi(x,t;A(t,T))$ as the expected value of a function of the process (1) conditional on an initial value. In particular, if 1_A is the indicator function taking the value one if the event A occurs and zero otherwise, then $\pi(x,t;A(t,T)) = E[1_{A(t,T)}\,|x(t)=x]$. Such expectations satisfy the Kolmogorov backward differential equation (see Section 5 of chapter 15 of Karlin and Taylor (1981)). Hence, $\pi(x,t;A(t,T))$ obeys

$$\frac{1}{2}\sigma^2(x)\frac{\partial^2\pi(x,t;A(t,T))}{\partial x^2} + \mu(x)\frac{\partial\pi(x,t;A(t,T))}{\partial x} + \frac{\partial\pi(x,t;A(t,T))}{\partial t} = 0. \tag{26}$$

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

$$\begin{aligned} d\pi(x(t),t;A(t,T)) \;=\; & \left[(1/2)\sigma^2(x(t))\left.\frac{\partial^2\pi(x,t;A(t,T))}{\partial x^2}\right|_{x=x(t)} + \mu(x(t))\left.\frac{\partial\pi(x,t;A(t,T))}{\partial x}\right|_{x=x(t)}\right. \\ & \left. +\left.\frac{\partial\pi(x,t;A(t,T))}{\partial t}\right|_{x=x(t)}\right] dt + \sigma(x(t))\left.\frac{\partial\pi(x,t;A(t,T))}{\partial x}\right|_{x=x(t)} dB(t). \end{aligned} \tag{27}$$

If (27) is a martingale then its drift must be zero, which implies that the function π satisfies the partial differential equation (26). The boundary conditions for this equation depend on the particular conditioning event $A(t,T)$.

3.2 Conditioned drift and diffusion coefficients

The drift coefficient conditional on the event $A(t,T)$ is given by

$$\mu(x,t\,|A(t,T)) = \lim_{\delta t\to 0}\frac{1}{\delta t}\int_\Omega (y-x) f(x,t,y,t+\delta t\,|A(t,T))\,dy, \tag{28}$$

[2]For the case of geometric Brownian motion for which we present explicit formulas below, it can be directly verified that the conditioned density given in equation (9.17) of Karlin and Taylor (1981) does not integrate to one.

where Ω is the domain of the conditioned process. For many of the standard stock price and interest rate models, Ω will be a subset of the interval $[0, \infty)$.

Assume that the process (1) and the conditioning event $A(t,T)$ are such that $\pi(x,t;A(t_1,t_2))$ has sufficient regularity to permit the Taylor expansion

$$\begin{aligned}\pi(y,t+\delta t;A(t_1,t_2)) &= \pi(x,t;A(t_1,t_2)) + (y-x)\frac{\partial \pi(x,t;A(t_1,t_2))}{\partial x} \\ &+ \delta t \frac{\partial \pi(x,t;A(t_1,t_2))}{\partial t} + o(y-x) + o(\delta t). \end{aligned} \tag{29}$$

Assume further that the process (1) and the conditioning event $A(t,T)$ are such that as $\delta t \to 0$, $\pi(x,t;A(t,t+\delta t)) \to 1$ and $\bar{\pi}(x,t,x,t+\delta t;A(t,t+\delta t)) \to 1$. Intuitively, this assumption says that as the time interval shrinks to zero, there is no time for the conditioning to be violated. Substituting (24) and (29) into (28) gives

$$\begin{aligned}\mu(x,t\,|A(t,T)) &= \lim_{\delta t \to 0} \frac{1}{\delta t} \int_\Omega (y-x) f(x,t,y,t+\delta t)\, dy \\ &+ \frac{1}{\pi(x,t;A(t,T))} \frac{\partial \pi(x,t;A(t,T))}{\partial x} \lim_{\delta t \to 0} \frac{1}{\delta t} \int_\Omega (y-x)^2 f(x,t,y,t+\delta t)\, dy \\ &+ \frac{1}{\pi(x,t;A(t,T))} \frac{\partial \pi(x,t;A(t,T))}{\partial t} \lim_{\delta t \to 0} \int_\Omega (y-x) f(x,t,y,t+\delta t)\, dy. \end{aligned} \tag{30}$$

Now

$$\lim_{\delta t \to 0} \int_\Omega (y-x) f(x,t,y,t+\delta t)\, dy = \int_\Omega (y-x)\delta(y-x)\, dy = 0, \tag{31}$$

where $\delta(y-x)$ is the Dirac delta function. Substituting (31) into (30) and using the definitions of the unconditioned drift and diffusion coefficients results in

$$\mu(x,t\,|A(t,T)) = \mu(x,t) + \sigma^2(x,t) \frac{1}{\pi(x,t;A(t,T))} \frac{\partial \pi(x,t;A(t,T))}{\partial x}. \tag{32}$$

An argument similar to the one that leads to (32) shows that the conditioned diffusion coefficient is equal to the unconditioned diffusion coefficient

$$\sigma^2(x,t\,|A(t,T)) = \sigma^2(x,t). \tag{33}$$

3.3 Conditional expected change over a finite time interval

Equation (32) provides the expected rate of change at time t of the conditional process over an infinitesimal length of time dt. In order for this expression to be used successfully with the Euler

approximation given by equation (15), the expected change of the conditioned process at time t must be approximately linear in time over a period of length δt. Although such linearity is often a reasonable assumption for the unconditioned process when δt is a short interval like one trade date, the type of conditioning we consider can cause the approximation to deteriorate to the point where it is not useful even for such short intervals.

To illustrate this point, assume that an unconditioned process obeys geometric Brownian motion

$$dx(t) = \mu x(t)dt + \sigma x(t)dB(t), \tag{34}$$

where $\mu = 0.05$ and $\sigma = 0.20$, and assume that the current value is $x(t) = 799$.[3] Now suppose we condition on the event that this process remains between 300 and 800 for four more time units (e.g., four more years), and consider the expected change over a small time interval of length η. Starting from the current value of 799, Figure 4 plots the expected change of the conditioned process as a function of the length of the interval η, as η increases from 0 to $\delta t = 1/250$ (one trade date).[4] The plot demonstrates that the expected change is highly nonlinear in the time interval η. Indeed, when $\eta = \delta t$ (one trade date) the expected change is about -15, while using the Euler approximation $\mu(x(t), t\,|A(t,T))\delta t$ the expected change is approximately -40. The intuition for this discrepancy is clear. When the process is very close to the upper boundary of 800, the instantaneous drift rate must be highly negative to prevent the boundary from being crossed. As a result, the process typically moves quickly away from the boundary, so that in a fraction of a trade date it will be an appreciable distance from the boundary. But once the process is an appreciable distance from the boundary, the (absolute value) of the instantaneous drift will decrease markedly because there is no longer an imminent danger of crossing the boundary. Hence, when the process is near a boundary, the expected instantaneous drift rate changes considerably over the next time interval. Since the Euler approximation assumes that the instantaneous drift rate remains constant, when the process is near a boundary it overstates the expected change over even short finite horizons.

We now turn to developing a framework for computing the conditional expected change over a finite time interval δt. This quantity can be written in terms of the conditional transition density. Let $m(x, t, t+\delta t\,|A(t,T))$ be the expected change in the process (1) over a finite time-step δt when it is currently at a level $x(t) = x$ conditional on some event $A(t,T)$. Then $m(x, t, t+\delta t\,|A(t,T))$ is given by

$$m(x, t, t+\delta t\,|A(t,T)) \;\; = \;\; \int_\Omega (y-x) f(x, t, y, t+\delta t\,|A(t,T))\, dy$$

[3]The process $\{x(t)\}$ can be interpreted as a stock index, in which case 799 is the number of index points. Alternatively, it might be an interest rate process, measured in basis points.

[4]In Section 4 we show how to compute these expected changes explicitly in the case of geometric Brownian motion.

$$= \int_\Omega y f(x,t,y,t+\delta t\,|A(t,T))\,dy - x. \tag{35}$$

Substituting into this expression from equation (24) gives

$$\begin{aligned} & m(x,t,t+\delta t\,|A(t,T)) \\ &= \frac{1}{\pi(x,t;A(t,T))} \times \qquad (36) \\ &\quad \int_\Omega \bar{\pi}(x,t,y,t+\delta t;A(t,t+\delta t))\,\pi(y,t+\delta t;A(t+\delta t,T))\,f(x,t,y,t+\delta t)\,y\,dy - x \\ &= \frac{1}{\pi(x,t;A(t,T))} \times \qquad (37) \\ &\quad E\left[\bar{\pi}(x,t,y,t+\delta t;A(t,t+\delta t))\,\pi(y,t+\delta t;A(t+\delta t,T))\,y|x(t)=x\right] - x. \end{aligned}$$

Next define the quantity

$$v(x,t,t+\delta t\,|A(t,T)) \equiv E\left[\bar{\pi}(x,t,y,t+\delta t;A(t,t+\delta t))\,\pi(y,t+\delta t;A(t+\delta t,T))\,y|x(t)=x\right], \tag{38}$$

so that

$$m(x,t,t+\delta t\,|A(t,T)) = \frac{v(x,t,t+\delta t\,|A(t,T))}{\pi(x,t;A(t,T))} - x. \tag{39}$$

Standard arguments on the relations between expectations and partial differential equations[5] imply that on the interval $[t,t+\delta t]$ the function v satisfies the backward equation

$$\frac{1}{2}\sigma^2(x)\frac{\partial v(x,t,t+\delta t\,|A(t,T))}{\partial x^2} + \mu(x)\frac{\partial v(x,t,t+\delta t\,|A(t,T))}{\partial x} + \frac{\partial v(x,t,t+\delta t\,|A(t,T))}{\partial t} = 0, \tag{40}$$

together with a terminal boundary condition at time $t+\delta t$ given by

$$\begin{aligned} & v(x,t+\delta t,t+\delta t\,|A(t+\delta t,T)) \\ &\quad = E\left[\bar{\pi}(x,t+\delta t,x,t+\delta t;A(t+\delta t,t+\delta t))\,\pi(x,t+\delta t;A(t+\delta t,T))\,x|x(t)=x\right] \qquad (41) \\ &\quad = E\left[1\times\pi(x,t+\delta t;A(t+\delta t,T))\,x|x(t)=x\right] \qquad (42) \\ &\quad = \pi(x,t+\delta t;A(t+\delta t,T)) \qquad (43) \end{aligned}$$

and spatial boundary conditions determined by the particular conditioning event $A(t,T)$. Below we compute $m(x,t,t+\delta t\,|A(t,T))$ by (numerically) solving this equation over the interval $[t,t+\delta t]$

[5] See, e.g., section 3 of chapter 15 of Karlin and Taylor (1981) or section 3.1 above.

to obtain $v(x,t,t+\delta t\,|A(t,T))$, solving (26) over the interval $[t,T]$ to obtain $\pi(x,t;A(t,T))$, and then combining $v(x,t,t+\delta t\,|A(t,T))$ and $\pi(x,t;A(t,T))$ using (39). We emphasize that (40) (with appropriate boundary conditions) applies over any interval $[t,t+\delta t]$, allowing the computation of $m(x,t,t+\delta t\,|A(t,T))$ over any such interval. By repeatedly solving (40) and using (39) for every interval of the form $[t_i,t_{i+1}]$ over which we have data, we are able to compute the conditioned means $m(x(t_i),t_i,t_{i+1}\,|A(0,T))$ which appear in the moment conditions used for estimation.

4 $A_3(t,T)$: The process stays above $\bar{a}$ and below $\bar{b}$ from time t to time T

4.1 The event $A_3(t,T)$

We now consider the event that the process is confined within a box $(\bar{a},\bar{b})\times[t,T]$, or that the process never reaches a lower boundary $\bar{a}$ or an upper boundary $\bar{b}$ from time t to T. The event is defined by

$$A_3(t,T)\equiv\left[\min_{u\in[t,T]}x(u)>\bar{a}\right]\bigcap\left[\max_{u\in[t,T]}x(u)<\bar{b}\right]. \tag{44}$$

It is easily seen that this event satisfies the measurability and semigroup properties from Section 3. For reasons that will become apparent in the next section, we use the subscript 3 to indicate this event.

In this section we compute the conditional expected change in the process $m(x,t,t+\delta t\,|A_3(t,T))$ $\equiv E[x(t+\delta t)|x(t)=x,A_3(t,T)]-x(t)$ over a finite time interval δt for the conditioning event $A_3(t,T)$. For convenience, we sometimes use the shorthand m_3 to refer to this conditional expected change. As indicated above, for estimation we will need to compute m_3 for every interval $[t_i,t_i+\delta t]=[t_i,t_{i+1}]$ for which we have data. We begin in subsection 4.2 by deriving an explicit formula for $m_3=m(x,t,t+\delta t\,|A_3(t,T))$ when the underlying process is geometric Brownian motion. Then in subsection 4.3 we show how to apply the results from Section 3 for a general conditioning event to the event $A_3(t,T)$ in order to compute m_3 numerically when the underlying process is a general univariate diffusion.

4.2 Explicit calculations for geometric Brownian motion

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

$$dx(t)=\mu x(t)dt+\sigma x(t)dB(t), \tag{45}$$

where $B(t)$ is a standard Brownian motion, μ is a constant drift parameter, and σ is a constant volatility parameter. This is a standard model for the evolution of stock prices, and is also a special case of the interest rate models used by AS (1996) and CHLS (1997).

To obtain an expression for m_3, we proceed by deriving in the following order expressions for the density $f(x,t,y,t+\delta t)$, the probabilities $\pi_3 \equiv \pi(x,t;A_3(t,T))$ and $\bar{\pi}_3 \equiv \bar{\pi}(x,t,y,t+\delta t;A_3(t,t+\delta t))$, and the conditioned density $f_3 \equiv f(x,t,y,t+\delta t\,|A_3(t,T))$.

4.2.1 Transition density $f(x,t,y,t+\delta t)$

Because the process $\{x(t)\}$ has non-constant drift and diffusion coefficients, it is more convenient to work with the transformed process $\{\ln x(t)\}$:

$$d\ln x(t) = \Big(\mu - \frac{1}{2}\sigma^2\Big)dt + \sigma dB(t). \tag{46}$$

Since $\ln x(t)$ follows a Brownian motion with nonzero drift, if $g(\ln x,t,\ln y,t')$ denotes the unconditioned transition density for $\ln x(t)$ going from $\ln x(t) = \ln x$ to $\ln x(t') = \ln y$, then

$$g(\ln x,t,\ln y,t') = \frac{1}{\sqrt{2\pi(t'-t)}\sigma}\exp\left(-\frac{(\ln y - \ln x - (\mu - \frac{1}{2}\sigma^2)(t'-t))^2}{2\sigma^2(t'-t)}\right). \tag{47}$$

Now let $f(x,t,y,t+\delta t)$ be the unconditioned transition density for the process to go from $x(t) = x$ to $x(t' = t+\delta t) = y$. Then it follows immediately from equation (47) that

$$f(x,t,y,t+\delta t)\,dy = \frac{1}{\sigma y\sqrt{2\pi\delta t}}\exp\left(-\frac{\left(\ln(y/x) - (\mu - \frac{1}{2}\sigma^2)\delta t\right)^2}{2\sigma^2\delta t}\right)dy. \tag{48}$$

Note that $f(x,t,y,t+\delta t)$ depends on t and $t+\delta t$ only through the difference δt.

4.2.2 Probability of the event $A_3(t,T)$

We turn next to finding a "closed form" expression for the probability $\pi_3 = \pi(x,t;A_3(t,T))$ of the process staying within $(\bar{a},\bar{b})$ during the time interval $[t,t+\delta t]$ given the current level $x(t) = x$.

Kunotomo and Ikeda (1992) consider pricing options with curved boundaries when the underlying asset follows 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 π_3.

Considering the special case of constant boundaries and setting the option strike price to the value of our lower boundary, we can modify Kunitomo and Ikeda equation (3.2) to obtain the following "closed form" expression for $\pi(x,t;A_3(t,T))$:

$$\pi(x,t;A_3(t,T)) = \sum_{n=-\infty}^{\infty}\left\{\Big(\frac{\bar{b}^n}{\bar{a}^n}\Big)^c\cdot\Big[N(\bar{d}_{1n}) - N(\bar{d}_{2n})\Big] - \Big(\frac{\bar{a}^{n+1}}{\bar{b}^n x}\Big)^c\cdot\Big[N(\bar{d}_{3n}) - N(\bar{d}_{4n})\Big]\right\}, \tag{49}$$

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

$$\bar{d}_{1n} = \frac{\ln\left(x\bar{b}^{2n}/\bar{a}^{2n+1}\right) + (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \tag{50}$$

$$\bar{d}_{2n} = \frac{\ln\left(x\bar{b}^{2n-1}/\bar{a}^{2n}\right) + (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \tag{51}$$

$$\bar{d}_{3n} = \frac{\ln\left(\bar{a}^{2n+1}/x\bar{b}^{2n}\right) + (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \tag{52}$$

$$\bar{d}_{4n} = \frac{\ln\left(\bar{a}^{2n+2}/x\bar{b}^{2n+1}\right) + (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \tag{53}$$

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 $\bar{a}$. In this case, we can obtain the probability of satisfying this condition by letting $\bar{b} \to \infty$ in (49). This probability might be useful for analyzing survivorship bias in the estimation of diffusion models of stock prices, for example. Combined with equation (24) and the other formulas in this section, it allows for exact maximum likelihood estimation when the underlying unconditioned process is geometric Brownian motion.

Below we use the above expression to check the numerical scheme we develop for computing π_3 when the underlying process follows a general univariate diffusion. In our calculation, the series turns out to converge quickly as claimed by Kunitomo and Ikeda (1992).

4.2.3 The probability $\bar{\pi}_3 = \bar{\pi}(x, t, y, t + \delta t; A_3(t, t + \delta t))$

By definition, $\bar{\pi}(x, t, y, t + \delta t; A_3(t, t + \delta t))$ is the probability that the geometric Brownian bridge with fixed value x at starting time t and y at ending time $t + \delta t$ never hits either of the two spatial boundaries, that is

$$\bar{\pi}(x, t, y, t + \delta t; A_3(t, t + \delta t)) = P\Big[\max_{u\in[t,t+\delta t]} x(u) < \bar{b}, \min_{u\in[t,t+\delta t]} x(u) > \bar{a} \mid x(t) = x, x(t + \delta t) = y\Big]. \tag{54}$$

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

$$X(t) \equiv \frac{1}{\sigma}\ln x(t) = \lambda t + B(t), \tag{55}$$

where $\lambda \equiv \frac{1}{\sigma}(\mu - 1/2\sigma^2)$. If we now define

$$\psi(x, t, y, t + \delta t)\, dy \equiv P\Big[X(t + \delta t) \in dy, \max_{u\in[t,t+\delta t]} X(u) - X(t) < h, \min_{u\in[t,t+\delta t]} X(u) - X(t) > k \mid X(t) = x\Big], \tag{56}$$

it follows from Douady (1998) that

$$\begin{aligned}\psi(x,t,y,t+\delta t) &= \mathbf{1}_{[k<x,y<h]}\cdot\exp\Big[-\frac{\lambda^2\delta t}{2}+\lambda(y-x)\Big]\\ &\times \sum_{n=-\infty}^{\infty}\Big\{g_{\delta t}\Big[(y-x)+2n(h-k)\Big]-g_{\delta t}\Big[2(h-y)-(y-x)+2n(h-k)\Big]\Big\},\end{aligned}\tag{57}$$

where

$$g_{\delta t}(z)=\frac{1}{\sqrt{2\pi\delta t}}\exp\Big(-\frac{z^2}{2\delta t}\Big)\tag{58}$$

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

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

$$P[X(t+\delta t)\in dy\mid X(t)=x]=\frac{1}{\sqrt{2\pi\delta t}}\exp\Big(-\frac{(y-x-\lambda\delta t)^2}{2\delta t}\Big)\,dy,\tag{59}$$

we have

$$\begin{aligned}&P\Big[\max_{u\in[t,t+\delta t]}X(u)-X(t)<h,\min_{u\in[t,t+\delta t]}X(u)-X(t)>k\mid X(t)=x,X(t+\delta t)=y\Big]\\ &= \frac{P\Big[X(t+\delta t)\in dy,\max_{u\in[t,t+\delta t]}X(u)-X(t)<h,\min_{u\in[t,t+\delta t]}X(u)-X(t)>k\mid X(t)=x\Big]}{P[X(t+\delta t)\in dy\mid X(t)=x]}\\ &= \mathbf{1}_{[k<x,y<h]}\cdot\sqrt{2\pi\delta t}\exp\big(\frac{(y-x)^2}{2\delta t}\big)\\ &\quad\times\sum_{n=-\infty}^{\infty}\Big\{g_{\delta t}\Big[(y-x)+2n(h-k)\Big]-g_{\delta t}\Big[2(h-y)-(y-x)+2n(h-k)\Big]\Big\},\end{aligned}\tag{60}$$

where for the last step we have substituted the expressions (57) and (59) for the numerator and denominator, and simplified the result.

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

$$\begin{aligned}&\bar\pi(x,t,y,t+\delta t;A_3(t,T))\\ &= P\Big[\max_{u\in[t,t+\delta t]}x(u)<\bar b,\min_{u\in[t,t+\delta t]}x(u)>\bar a\,\Big|\,x(t)=x,x(t+\delta t)=y\Big]\\ &= P\Big[\max_{u\in[t,t+\delta t]}\ln x(u)-\ln x(t)<\sigma h\equiv\ln\bar b-\ln x(t),\\ &\qquad\min_{u\in[t,t+\delta t]}\ln x(u)-\ln x(t)>\sigma k\equiv\ln\bar a-\ln S(t)\,\Big|\,x(t)=x,x(t+\delta t)=y\Big]\\ &= P\Big[\max_{u\in[t,t+\delta t]}X(u)-X(t)<h,\min_{u\in[t,t+\delta t]}X(u)-X(t)>k\,\Big|\,X(t)=\frac{1}{\sigma}\ln x,X(t+\delta t)=\frac{1}{\sigma}\ln y\Big]\\ &= \mathbf{1}_{[\bar a<x,y<\bar b]}\cdot\sqrt{2\pi\delta t}\exp\Big(\frac{(\ln y/x)^2}{2\sigma^2\delta t}\Big)\end{aligned}$$

$$\times \sum_{n=-\infty}^{\infty} \Big\{ g_{\delta t}\Big[\frac{1}{\sigma} \ln(y/x) + \frac{2n}{\sigma} \ln(\bar{b}/\bar{a}) \Big] - g_{\delta t}\Big[\frac{2}{\sigma} \ln(\bar{b}/x) - \frac{1}{\sigma} \ln(y/x) + \frac{2n}{\sigma} \ln \bar{b}/\bar{a} \Big] \Big\}, \tag{61}$$

where in the last step we have used equation (60) and substituted in the expressions for h and k.

Using the fact that $g_{\delta t}(z)$ is symmetric around $z = 0$, it is easy to verify that $\bar{\pi}_3$ satisfies the boundary conditions

$$\begin{aligned} \bar{\pi}(\bar{a}, t, y, t+\delta t; A_3(t, t+\delta t)) &= \bar{\pi}(\bar{b}, t, y, t+\delta t; A_3(t, t+\delta t)) = 0, & (62) \\ \bar{\pi}(x, t, \bar{a}, t+\delta t; A_3(t, t+\delta t)) &= \bar{\pi}(x, t, \bar{b}, t+\delta t; A_3(t, t+\delta t)) = 0, & (63) \\ \bar{\pi}(x, t+\delta t, x, t+\delta t; A_3(t+\delta t, t+\delta t)) &= 1. & (64) \end{aligned}$$

Expression (61) 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.

4.2.4 The conditioned density $f_3 = f(x, t, y, t+\delta t \,|A_3(t, T))$ and the conditioned mean $m_3 = m(x, t, t+\delta t \,|A_3(t, T))$

Let $f(x, t, y, t+\delta t \,|A_3(t, T))$ be the transition density for the process $\{x(t)\}$ to go from $x(t) = x$ to $x(t+\delta t) = y$ conditional on $A_3(t, T)$. From (24) we have

$$f(x, t, y, t+\delta t \,|A_3(t, T))\, dy = \frac{\bar{\pi}(x, t, y, t+\delta t; A_3(t, T))\, \pi(y, t+\delta t; A_3(t+\delta t, T))\, f(x, t, y, t+\delta t)}{\pi(x, t; A_3(t, T))}\, dy. \tag{65}$$

Equations (48), (49), and (61) provide expressions for all of the quantities on the right hand side of (65). Simply substituting these quantities into equation (65) provides one method for computing f_3. However, when δt is small we can simplify the calculation by approximating $\bar{\pi}(x, t, y, t+\delta t; A_3(t, T))$. In particular, the unconditioned transition density $f(x, t, y, t+\delta t)$ is always concentrated around x. This observation justifies the following approximation:

$$\bar{\pi}(x, t, y, t+\delta t; A_3(t, t+\delta t)) \approx \begin{cases} P\Big[\max_{u \in [t, t+\delta t]} x(u) < \bar{b} \mid x(t) = x, x(t+\delta t) = y\Big] & \text{if } x(t) \text{ is near } \bar{b}, \\ P\Big[\min_{u \in [t, t+\delta t]} x(u) > \bar{a} \mid x(t) = x, x(t+\delta t) = y\Big] & \text{if } x(t) \text{ is near } \bar{a}, \\ 1 & \text{otherwise.} \end{cases} \tag{66}$$

One way to operationalize the word 'near' in this approximation is to use the volatility of the process as a measure of nearness. For example, we could let $\bar{\pi}_3$ be 1 when $x(t)$ is more than $d = 20$

standard deviations away from both the spatial boundaries. In practice, the approximation above is excellent when δt is small. For instance, when we consider a stock price process and let δt be one trade date, the above approximation is almost exact (without making d very big, the error in the approximation, which is essentially exponential in $-d$, can be made much smaller than the machine accuracy of a typical computer workstation). In the next section, we will use the approximation (66) when we check the accuracy of our numerical scheme.

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

$$\begin{aligned}
\psi^{\bar{b}}(x,t,y,t+\delta t)\,dy &\equiv P\Big[X(t+\delta t)\in dy, \max_{u\in[t,t+\delta t]} X(u)-X(t)<h \mid X(t)=x\Big] && (67)\\
&= \frac{1}{\sqrt{2\pi\delta t}}\times\Big[\exp\Big(-\frac{(y-x-\lambda\delta t)^2}{2\delta t}\Big) && (68)\\
&\qquad -\exp\Big(2\lambda h-\frac{(2h+\lambda\delta t-(y-x))^2}{2\delta t}\Big)\Big]\,dy.
\end{aligned}$$

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

$$P\Big[\min_{u\in[t,t+\delta t]} x(u)<b \mid x(t)=x, x(t+\delta t)=y\Big] = 1-\exp\left(-\frac{(2\ln(b/x)-\ln(y/x))^2-\ln^2 y/x}{2\sigma^2\delta t}\right). \quad (69)$$

The second probability in the approximation scheme can be worked out by using the following reflection principle:

$$P[\lambda t+B(t)>k] = P[-\lambda t+B(t)<-k], \quad (70)$$

and reducing the problem to the known probability $\psi^{\bar{b}}$. The result is

$$\bar{\pi}(x,t,y,t+\delta t;A_3(t,T)) \approx \begin{cases} 1-\exp\Big(-\frac{(2\ln\bar{b}/x-\ln y/x)^2-\ln^2 y/x}{2\sigma^2\delta t}\Big) & \text{if } x(t) \text{ is near } \bar{b},\\ 1-\exp\Big(-\frac{(2\ln\bar{a}/x-\ln y/x)^2-\ln^2 y/x}{2\sigma^2\delta t}\Big) & \text{if } x(t) \text{ is near } \bar{a},\\ 1 & \text{otherwise.}\end{cases} \quad (71)$$

Alternatively, one can start from equation (61) and take the limit $\bar{a}\to 0$ when x is close to $\bar{b}$ and the limit $\bar{b}\to\infty$ when x is close to $\bar{a}$. When $\bar{b}\to\infty$, we need only consider $n=0$ in the first term of the infinite summation and $n=-1$ in the second term (similarly for $\bar{a}\to 0$.) As a result, this approach is also computationally straightforward.

Finally, the expected change m_3 over a finite time interval can be calculated from

$$m(x,t,t+\delta t\,|A_3(t,T)) = \int_0^\infty (y-x) f(x,t,y,t+\delta t\,|A_3(t,T))\,dy \tag{72}$$

$$= \frac{1}{\pi(x,t;A_3(t,T))}\int_0^\infty f(x,t,y,t+\delta t)\bar{\pi}(x,t,y,t+\delta t;A_3(t,T)) \tag{73}$$
$$\times\pi(y,t+\delta t;A_3(t,T))y\,dy - x.$$

We use an adaptive recursive Newton-Cotes 8 panel rule to integrate m_3 numerically.

4.3 Numerical approach

Although explicit formulas are available when the underlying process obeys geometric Brownian motion, for most other specifications this is not the case. As a result, we now turn to the task of computing the expected change over a finite time interval conditional on the event A_3 when the underlying process (unconditionally) follows a general time-homogeneous univariate diffusion.

The starting point for this computation is equation (36) specialized to the particular conditioning event A_3:

$$m(x,t,t+\delta t\,|A_3(t,T)) = \frac{1}{\pi(x,t;A_3(t,T))} \tag{74}$$
$$\times E\left[\bar{\pi}(x,t,y,t+\delta t;A_3(t,t+\delta t))\,\pi(y,t+\delta t;A_3(t+\delta t,T))\,y|x(t)=x\right] - x.$$

Unlike in the previous subsection, the strategy to calculate this will not be separately to compute each of the components $\pi(x,t;A_3(t,T))$, $\bar{\pi}(x,t,y,t+\delta t;A_3(t,t+\delta t))$, and $\pi(y,t+\delta t;A_3(t+\delta t,T))$. Rather, the two terms $\pi(x,t;A_3(t,T))$ and $E[\bar{\pi}(x,t,y,t+\delta t;A_3(t,t+\delta t))\ \pi(y,t+\delta t;A_3(t+\delta t,T))\,y|x(t)=x]$ will be computed as the (numerical) solutions of two different partial differential equations, (26) and (40), with appropriate boundary conditions. These numerical computations will then be combined to obtain $m(x,t,t+\delta t\,|A_3(t,T))$.

A tricky feature of the problem is that the partial differential equations for $\pi(x,t;A_3(t,T))$ and $E\left[\bar{\pi}_3(x,t,y,t+\delta t;A_3(t,t+\delta t))\,\pi(y,t+\delta t;A_3(t+\delta t,T))\,y|x(t)=x\right]$ hold in different regions. Equation (74) above is the expected change over the interval $[t,t+\delta t]$, and the differential equation (40) used to compute $E\left[\bar{\pi}(x,t,y,t+\delta t;A_3(t,t+\delta t))\,\pi(y,t+\delta t;A_3(t+\delta t,T))\,y|x(t)=x\right]$ holds on the interval $[t,t+\delta t]$, with terminal condition (41). We will need separately to solve this differential equation for each interval over which the expected change is required. In contrast, the probability $\pi(x,t;A_3(t,T))$ which appears in the denominator of (74) is the probability of satisfying the condition on the entire interval $[t,T]$; we need to solve the differential equation (26) only once on the interval $[0,T]$, and then refer to the single solution at the various points (x,t) in order to obtain $\pi(x,t;A_3(t,T))$.

4.3.1 Computation of π_3

Substitution into (26) indicates that π_3 obeys the partial differential equation

$$\frac{1}{2}\sigma^2(x)\frac{\partial^2\pi(x,t;A_3(t,T))}{\partial x^2}+\mu(x)\frac{\partial\pi(x,t;A_3(t,T))}{\partial x}+\frac{\partial\pi(x,t;A_3(t,T))}{\partial t}=0. \tag{75}$$

The boundary conditions are:

$$\begin{aligned}\pi(\bar{b},t;A_3(t,T)) &= 0 \text{ for } t\in[0,T],\\ \pi(\bar{a},t;A_3(t,T)) &= 0 \text{ for } t\in[0,T],\\ \pi(x,T;A_3(t,T)) &= 1 \text{ for } \bar{a}<x<\bar{b}.\end{aligned} \tag{76}$$

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 $(\bar{a},\bar{b})\times[0,T]$. Except near the terminal boundary, we use a grid with elements of identical size $\Delta x\times\Delta t$ and the Crank-Nicholson scheme, and we generally set the time-step Δt to be one trade date.[6]

We check the accuracy of our numerical scheme by comparing the π_3 obtained from it when the underlying process follows geometric Brownian motion with the explicit solution for this case derived in the preceding subsection. The geometric Brownian motion is described by

$$dx(t)=\mu x(t)dt+\sigma x(t)dB(t), \tag{77}$$

and is a special case of the specifications used by AS (1996) and CHLS (1997). As in our earlier example, $\mu=0.05$, $\sigma=0.20$, and the process is confined within $(300,800)\times[0,12]$, i.e. , $\bar{a}=300$, $\bar{b}=800$, and $T=12$ years. We set $\Delta x=1$ and set $\Delta t=1/250$ or approximately one trade date in most of the region, but use a smaller time-step near the terminal boundary. Figure 5 illustrates the accuracy of the Crank-Nicholson scheme by showing the π_3 vectors for both the Crank-Nicholson scheme and the explicit solution when $t=8$, so that 4 years remain.[7] The two curves lie right on

[6]Difficulties arise at the terminal boundary because π_3 is discontinuous at the points $(\bar{a},T)$ and $(\bar{b},T)$ where the spatial and terminal boundaries meet, i.e. $\pi(\bar{a},T;A_3(T,T))=\pi(\bar{b},T;A_3(T,T))=0$ but $\pi(\bar{a}+\eta,T;A_3(T,T))=\pi(\bar{b}-\eta,T;A_3(T,T))=1$ for any $\eta>0$. As a result, the function π_3 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.

[7]The magnitudes of the errors depend on the time t, and the errors for $t=8$ are typical of those in most of the region $(300,800)\times[0,12]$. The largest 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_3 is highly curved and the errors in the finite difference scheme are as large as 10^{-4}. (Near the

top of one another indicating that the solution from the Crank-Nicholson scheme is very accurate. Figure 6 plots the difference vector of the two approaches which is on the order of 10^{-7}. Although geometric Brownian motion is only a special case of the CHLS specification and we do not have an explicit formula for the general case, we have no reason to believe that the performance of the Crank-Nicholson scheme is any poorer in the general case.

We next use the Crank-Nicholson scheme to compute π_3 for the AS short rate data described above. Once again the process is restricted to be within a box $(a, b) \times [0, T]$. The minimum and maximum of the actual interest rate process are $a = 0.02915$ and $b = 0.24333$, and we use 500 spatial steps Δx to cover this range, so $\Delta x = (a - b)/500 = 4.2836$ basis points. We set $\bar{a}$ and $\bar{b}$, respectively, to be 10 mesh points (42.836 basis points) below and above the minimum and maximum, so that the interval $[\bar{a} = a - 10\Delta x = 0.02487, \bar{b} = b + 10\Delta x = 0.24761]$ is covered by 520 spatial steps. The AS data consists of 5,505 observations covering a period of length 21.73424 years, and we use 5,504 time steps each of length $\Delta t = 21.73424/5,504 = 0.0039488$ year or one trading day.[8] The drift and diffusion coefficients of the underlying diffusion process are set to:

$$\mu(x) = \alpha_0 + \alpha_1 x = 0.03400 - 0.2834x, \tag{78}$$

$$\sigma^2(x) = \beta_1 x^{2\beta_2} = 2.0511 x^{2(1.3333)}. \tag{79}$$

Figure 7 presents the numerical solution for π_3, as a function of the time and spatial indices. Here $\pi_3 = \pi(x, t; A_3(t, T))$ is the probability that the process stays between $\bar{a} = 0.02487$ and $\bar{b} = 0.24761$ through the end of the sample period, as a function of the current interest rate level x and time t. Time index 5,504 (the "back" edge of the graph) is the terminal boundary at which $\pi_3 = 1$, while time index 0 is the beginning of the sample period. The probability of staying between $\bar{a} = 0.02487$ and $\bar{b} = 0.24761$ is zero at these spatial boundaries, and greater than zero in between them. At time zero this probability is approximately 0.10, of course increasing to 1 at the terminal boundary. The asymmetry between low and high interest rates is caused by the fact that the diffusion process has considerably greater volatilty for higher interest rates. As a result, at a particular point in time the process is more likely to hit the upper than the lower boundary when starting at a fixed distance from either.

Figure 7 helps us understand the difference between the conditioned and unconditioned drift coefficients, and when this difference will be large. Looking back to equation (32), the difference between the conditioned and unconditioned drifts is

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 where the spatial and terminal boundaries meet turn out to be unimportant for the estimation we carry out below, because the interest rate path does not enter these neighborhoods.

[8] As discussed in footnote 6, the time step closest to the terminal boundary is subdivided into 10 smaller time steps.

$$\mu(x,t\,|A_3(t,T)) - \mu(x) = \sigma^2(x)\frac{1}{\pi(x,t;A_3(t,T))}\frac{\partial\pi(x,t;A_3(t,T))}{\partial x}, \tag{80}$$

and is large when $\partial\pi_3/\partial x$ is large and π_3 is small. The derivative $\partial\pi_3/\partial x$ is approximately zero and π_3 is not very small for a reasonably wide range of rates, especially near the terminal boundary. In this region the difference between the conditioned and unconditioned drifts will be small. Figure 7 suggests that this difference will be large when the process is very close to the lower spatial boundary, but that even a small distance away from the lower boundary the difference between the conditioned and unconditioned drifts will be small. The difference will also be large near the upper spatial boundary; even though $\partial\pi_3/\partial x$ does not appear to be as large as near the lower boundary, the volatility $\sigma(x)$ is much larger (given the parameter estimates) near the upper than the lower boundary. Moreover, because $\partial\pi_3/\partial x$ does not change as rapidly near the upper boundary, the difference will be large for a larger range of interest rates.

Figure 8 plots π_3 along the actual interest rate path. As the time index increases from 0 to 5,505 (as time increases from 0 to T), π_3 is generally increasing. This is reasonable since the actual interest rate path remained inside of the defined box. A comparison with Figure 1 indicates that, as one would expect, π_3 dips whenever the interest rate process approaches either boundary.

4.3.2 Computation of $E[\bar{\pi}(x,t,y,t+\delta t;A_3(t,t+\delta t))\ \pi(y,t+\delta t;A_3(t+\delta t,T))\,y|x(t)=x]$

To compute m_3 using equation (74) we must first calculate $E[\bar{\pi}_3\pi_3\,|x(t)=x]$, i.e. we must calculate

$$E\left[\bar{\pi}(x,t,y,t+\delta t;A_3(t,t+\delta t))\,\pi(y,t+\delta t;A_3(t+\delta t,T))\,y|x(t)=x\right].$$

Following equation (38), define

$$v(x,t,t+\delta t\,|A_3(t+\delta t,T)) \equiv E\left[\bar{\pi}(x,t,y,t+\delta t;A_3(t,t+\delta t))\,\pi(y,t+\delta t;A_3(t+\delta t,T))\,y|x(t)=x\right]. \tag{81}$$

Then from equation (40), v_3 satisfies the partial differential equation

$$\begin{aligned}&\frac{1}{2}\sigma^2(x)\frac{\partial v(x,t,t+\delta t\,|A_3(t+\delta t,T))}{\partial x^2}\\ &\quad+\mu(x)\frac{\partial v(x,t,t+\delta t\,|A_3(t+\delta t,T))}{\partial x}+\frac{\partial v(x,t,t+\delta t\,|A_3(t+\delta t,T))}{\partial t}=0,\end{aligned} \tag{82}$$

together with the appropriate boundary conditions. To compute the conditioned means m_3 for each interval of the form $[t,t+\delta t]$ for which we have data, we must first solve this partial differential equation for each such interval.

To do this, we need boundary conditions for various intervals of the form $[t, t+\delta t]$. For the event $A_3(t,T)$, the function v_3 satisfies the boundary conditions

$$\begin{aligned} & v(\bar{a}, t, t+\delta t \,|A_3(t+\delta t, T)) \\ &= E\left[\bar{\pi}(\bar{a}, t, y, t+\delta t; A_3(t, t+\delta t))\, \pi(y, t+\delta t; A_3(t+\delta t, T))\, y | x(t) = \bar{a}\right] \\ &= E\left[0 \times \pi(y, t+\delta t; A_3(t+\delta t, T))\, y | x(t) = \bar{a}\right] \\ &= 0, \end{aligned} \tag{83}$$

$$\begin{aligned} & v(\bar{b}, t, t+\delta t \,|A_3(t+\delta t, T)) \\ &= E\left[\bar{\pi}(\bar{b}, t, y, t+\delta t; A_3(t, Tt+\delta t))\, \pi(y, t+\delta t; A_3(t+\delta t, T))\, y | x(t) = \bar{b}\right] \\ &= E\left[0 \times \pi(y, t+\delta t; A_3(t+\delta t, T))\, y | x(t) = \bar{b}\right] \\ &= 0, \end{aligned} \tag{84}$$

$$\begin{aligned} & v(x, t+\delta t, t+\delta t \,|A_3(t+\delta t, T)) \\ &= E\left[\bar{\pi}(x, t+\delta t, x, t+\delta t; A_3(t+\delta t, t+\delta t))\, \pi(x, t+\delta t; A_3(t+\delta t, T))\, x | x(t+\delta t) = x\right] \\ &= E\left[1 \times \pi(x, t+\delta t; A_3(t+\delta t, T))\, x | x(t+\delta t) = x\right] \\ &= \pi(x, t+\delta t; A_3(t+\delta t, T)) x \text{ for } x \in (\bar{a}, \bar{b}). \end{aligned} \tag{85}$$

For the most part, we solve these partial differential equations using the Crank-Nicholson scheme with just one time step. That is, the time increment Δt in the finite difference scheme is chosen to be equal to the interval δt between data points, so that the region $[t, t+\delta t]$ over which the partial differential equation holds is traversed using just one finite difference time step. However, Figure 4 discussed above reveals that near the boundaries the expected change m_3 and thus v_3 are nonlinear in the time horizon. This suggests that using the Crank-Nicholson scheme with a single time increment $\Delta t = \delta t$ near the boundary may provide a poor approximation to v_3 and m_3. Thus, we use a finer mesh for times when the interest rate is near either of the two boundaries $\bar{a}$ or $\bar{b}$. As Figure 4 suggests, the use of a finer mesh makes a significant difference in the computation of m_3.[9]

4.3.3 Computation of m_3

We use equation (74) to compute m_3 by substituting in the numerical solutions of π_3 and m_3 described in the previous two subsections. We again check the accuracy of our numerical approach

[9]However, our use of the finer mesh turns out not to make a significant difference in the results of the estimation 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, and thus little impact on the estimates.

by comparing it to the explicit solution that is available in the case of geometric Brownian motion, using the same parameters as above to check the accuracy of the computation of π_3. Figure 9 plots the conditional expected change m_3 over a time interval of $\delta t = 1/250$ (one trade date) as a function of x at time $t = 8$, using both the explicit formula and our finite difference approximation. The figure reveals that the two curves overlap and cannot be distinguished, suggesting that the Crank-Nicholson finite difference approximation of m_3 is accurate.

Figure 10 provides a more detailed examination of the errors in our finite difference approximation. In particular, it shows the difference between the conditional expected change m_3 computed using the Crank-Nicholson scheme and that obtained from the explicit formula. As the figure shows, the errors are essentially zero about 20 or more grid points (recall that here $\Delta x = 1$) away from the boundaries. Even when we are within 20 grid points of a boundary, the relative error is only on the order of 0.01 percent of the value of x. Since in the estimations we carry out below errors in computing m_3 of this magnitude occur at only a few data points, they will have essentially no effect on our moment conditions.

4.3.4 Discussion

This checking simply verifies that we are able accurately to compute the conditional expected change m_3 using the Crank-Nicholson scheme. But is the conditioning important? In other words, is the difference between the conditioned change m_3 and the unconditioned change large? Figure 9 also answers this question, revealing the conditioning to be very important near the boundaries. Given the drift $\mu(x) = 0.05x$ used to construct the figure, the unconditioned expected change is a straight line, in marked contrast to the conditioned change shown in the figure. The conditioning makes the expected change m_3 very negative near the upper boundary and positive near the lower boundary, as one would expect.

We next apply our numerical approach to AS's interest rate data using the linear unconditioned drift shown in equations (78) and (79) of section 4.3.1. Using the parameter vector $(\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (0.0340, -0.2834, 0.0000, 0.0000)$, Figure 11 graphs the expected one trade date change for different interest rates at time $t = 0$ with conditioning (i.e., m_3) and without conditioning. The expected changes without conditioning are computed using the Euler approximation $\mu(x, t)\delta t$. The curves at other times t are similar to these. We see that conditioning has a large effect on the expected change vector. Thus, it will not be surprising if conditioning has an important effect on parameter estimation.

The value of m_3 along the sample path is plotted in Figure 12. When interest rates are close to the lower boundary, m_3 is positive and when they are close to the upper boundary, m_3 is negative. The most negative value of m_3 is obtained at the maximum of the interest rate process, at a time index of 1942.

4.3.5 Estimation

To assess the effect of conditioning on parameter estimation, we apply the GMM procedure outlined in Section 2 to AS's interest rate data to estimate the CHLS model both without conditioning and with conditioning on A_3. In order to account for the conditioning we replace the quantity

$$u_i = x_i^{-\beta_2}(\delta x_i) - x_i^{-\beta_2}(\alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \alpha_3/x_i)(\delta t) \tag{86}$$

with

$$u_i \equiv x_i^{-\beta_2}(\delta x_i) - x_i^{-\beta_2} m_A(x_i, t_i, t_i + \delta t; A_3(t_i, T)). \tag{87}$$

Using this expression for u_i, we minimize the objective function $G(\theta) = \frac{1}{N-1} h'(\theta) W(\theta) h(\theta)$ over the vector $\theta = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)$ using a Nelder-Mead simplex method. This method is sometimes called downhill direct search and is well known for its robustness. The numerical scheme is implemented with the mesh constructed so that there are 500 grid points between the minimum and the maximum of the AS data. Consequently, the interval between grid points is 4.2836 basis points. The CHLS parameters are estimated conditioning on A_3 with $\bar{a}$ and $\bar{b}$ displaced the same number of grid points from, respectively, the minimum and maximum of the AS interest rate data. The estimation results for various levels of this displacement are recorded in Table 1.

Table 1 reveals that the signs of the parameter estimates do not change from the case where there is no conditioning to the cases where 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 $\bar{a} = a - 20\Delta x$ and $\bar{b} = b + 20\Delta x$, where a and b are set equal to 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 estimate.)

If the specification of the drift were linear (i.e., if α_2 and α_3 were forced to be zero), then we would expect a negative α_1 coefficient 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 parameters dominate 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 equation (33)). 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.

Figure 13 plots the five drift coefficients computed using the five sets of parameter estimates in Table 1. Each drift coefficient is plotted between the minimum and the maximum values of the AS data. As expected given the changes in the parameter estimates mentioned above, 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 grid points the process is permitted to go below a and above b 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 $\bar{a}$ and $\bar{b}$, however, the conditioning (i.e., the conditioned drift) prevents the process from increase beyond the upper bound upon which we are conditioning, and 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 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 grid points, the estimated drift coefficient becomes positive near the maximum in the data.

Many standard interest rate models assume that the drift coefficient is linear (Vasicek (1977), Cox, Ingersoll, and Ross (1985), and 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 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

$$G_R = \frac{1}{N-1} \min_{\theta_R} h'(\theta_R) W(\theta_R) h(\theta_R) \tag{88}$$

where the h and W functions are defined in Section 2. G_R is a restricted version of the G which solves the minimization problem (14).

Table 2 reports the results for the case of no conditioning and for conditioning with displacements of 60, 40, 20, and 10 grid points. The final column of Table 2 reports the difference between the restricted and unrestricted objective functions, $G_R - G$, where the unrestricted objective function G is given by (14) 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. that $\alpha_2 = \alpha_3 = 0$.

The unconditioned results in the first line of Table 2 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 $G_R - G$ statistic testing linearity, distributed (asymptotically) χ^2_2, has a value of 4.650. Larger values of $G_R - G$ indicate that it is less likely that the unconditioned drift coefficient is linear, because they imply that there was relatively more difficulty satisfying the moment conditions when the drift of the unconditioned process is forced to be linear. The five and ten percent critical values of the χ^2_2 distribution are 5.99 and 4.605, and the value 4.650 has a p-value of 0.0978 or 9.78 percent. 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 of them having absolute values less than twice their standard errors. Most strikingly, for all of the conditioned results the test statistics $G_R - G$ 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 and therefore providing no evidence of non-linearity. The p-value of the largest of them, 0.807, is 0.668, indicating that there is a 66.8 percent chance of obtaining a larger test statistic even if the null hypothesis of linearity is true. The p-values of the others are between 86.9 and 99.7 percent. While the Chi-squared distribution of the test statistic $G_R - G$ is an asymptotic result and the small sample properties of this statistic are unknown in the present context, these test statistics are so small it seems unlikely that this caveat is important. Thus, when one conditions on the event that the process remains between upper and lower barriers $\bar{a}$ and $\bar{b}$, there is no evidence of non-linearity.

These results fail to reject linearity in the drift of the unconditioned process because the conditioned drift is non-linear even when the unconditioned drift is linear. As mentioned 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 $\bar{a}$ and $\bar{b}$, however, the conditioned drift accommodates the behavior of the data, without making the unconditioned drift be nonlinear.

The different estimates of α_0 and α_1 obtained for different displacements (i.e., different levels of $\bar{a}$ and $\bar{b}$) in Table 2 indicate that the conditioning can have important effects on the parameter estimates. Figure 14 illustrates this by plotting the five drift coefficients computed using the five sets of parameter estimates in Table 2. 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 grid points), the slope of the drift coefficient becomes positive, and increases as the displacement is reduced to 40 grid points or fewer. This might seem problematic, since it implies an explosive process. For the cases plotted in Figure 14, 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.

5 $A_0(0,T)$: The process has a minimum of a and a maximum of b from time 0 to time T

We argued above that conditioning is especially important at the minimum and the maximum of the observed data. The reason is that typically the unconditioned process can either increase or decrease at both the minimum and maximum observed values. In the actual data, however, the process only increases at the minimum value and only decreases at the maximum value. As a result, the underlying process is necessarily misrepresented by the observed data at a sample's minimum and maximum. This fact suggests that it is important to provide an analysis of conditioning on reaching particular maximum and minimum values to supplement the analysis of conditioning on staying between a maximum and minimum value that was provided in the previous section.

5.1 The event $A_0(0,T)$

We next consider the event that the process has a minimum value of a and a maximum value of b from time 0 to time T. More formally, the event is defined by

$$A_0^*(0,T) \equiv \left[\min_{u\in[0,T]} x(u) = a\right] \bigcap \left[\max_{u\in[0,T]} x(u) = b\right]. \tag{89}$$

This event satisfies the measurability and semigroup properties from Section 3.

One issue is that the event (89) has zero probability, which is problematic for several reasons. We overcome this difficulty by introducing $\bar{a} < a$ and $\bar{b} > b$ and considering the event

$$A_0(0,T) \equiv \left[\min_{u\in[0,T]} x(u) > \bar{a}\right] \bigcap \left[\min_{u\in[0,T]} x(u) \leq a\right] \bigcap \left[\max_{u\in[0,T]} x(u) \geq b\right] \bigcap \left[\max_{u\in[0,T]} x(u) < \bar{b}\right]. \tag{90}$$

Then as $\bar{a} \to a$ and $\bar{b} \to b$ the event (90) approximates the event (89).

5.2 Strategy for computing m_0

The computation of the GMM moment conditions requires the calculation of the finite time conditional expected change for each data point. The behavior of the process up until the time t dictates the event that must occur over the interval $[t,T]$ in order for the event $A_0(0,T)$ to occur. Let t be the time of a generic data point. It is helpful to classify the data point observed at time t into one of four categories depending upon whether up to time t:

- Neither a nor b has been reached.
- a has been reached but neither $\bar{a}$ nor b has been reached.
- b has been reached but neither a nor $\bar{b}$ has been reached.
- a and b have been reached, but neither $\bar{a}$ nor $\bar{b}$ have been reached.

For each of these possibilities, we define an event which must be satisfied over the time interval $[t,T]$ in order for the event $A_0(0,T)$ to occur:

$$\begin{aligned}
A_0(t,T) &\equiv \left[\min_{u\in[t,T]} x(u) > \bar{a}\right] \bigcap \left[\min_{u\in[t,T]} x(u) \leq a\right] \bigcap \left[\max_{u\in[t,T]} x(u) \geq b\right] \bigcap \left[\max_{u\in[t,T]} x(u) < \bar{b}\right]; \\
A_1(t,T) &\equiv \left[\min_{u\in[t,T]} x(u) > \bar{a}\right] \bigcap \left[\max_{u\in[t,T]} x(u) \geq b\right] \bigcap \left[\max_{u\in[t,T]} x(u) < \bar{b}\right]; \\
A_2(t,T) &\equiv \left[\min_{u\in[t,T]} x(u) > \bar{a}\right] \bigcap \left[\min_{u\in[t,T]} x(u) \leq a\right] \bigcap \left[\max_{u\in[t,T]} x(u) < \bar{b}\right]; \\
A_3(t,T) &\equiv \left[\min_{u\in[t,T]} x(u) > \bar{a}\right] \bigcap \left[\max_{u\in[t,T]} x(u) < \bar{b}\right].
\end{aligned}$$

Corresponding to these four events we define four probabilities:

$$\begin{aligned}
\pi(x,t;A_0(t,T)) &\equiv P[A_0(t,T) \mid x(t)=x];\\
\pi(x,t;A_1(t,T)) &\equiv P[A_1(t,T) \mid x(t)=x];\\
\pi(x,t;A_2(t,T)) &\equiv P[A_2(t,T) \mid x(t)=x];\\
\pi(x,t;A_3(t,T)) &\equiv P[A_3(t,T) \mid x(t)=x].
\end{aligned}$$

We will compute the conditional expected change over a finite time interval for each of the four types of data points by applying equation (36). Letting $i \in \{0,1,2,3\}$ this equation becomes

$$\begin{aligned}
m(x,t,t+\delta t\,|A_i(t,T)) &= \frac{1}{\pi(x,t;A_i(t,T))} \times \\
&\quad E\left[\bar{\pi}(x,t,y,t+\delta t;A_i(t,t+\delta t))\,\pi(y,t+\delta t;A_i(t+\delta t,T))\,y|x(t)=x\right]-x.
\end{aligned} \tag{91}$$

The next two subsections will show how to compute numerically for $i \in \{0,1,2,3\}$ the quantities $\pi(x,t;A_i(t,T))$ and $v(x,t,t+\delta t\,|A_i(t,T)) = E[\bar{\pi}(x,t,y,t+\delta t;A_i(t,t+\delta t))\ \pi(y,t+\delta t;A_i(t+\delta t,T))\,y|x(t)=x]$ as the solutions to backward partial differential equations with appropriate boundary conditions. The results of these computations will then be combined to obtain $m(x,t,t+\delta t\,|A_i(t,T))$. Where convenient, we use the notation π_i, $\bar{\pi}_i$, v_i, and m_i as shorthand for the probabilities of the events A_i and the expected values conditional on the events A_i.

5.3 Computation of the π_i's

This subsection describes the computation of the π_i probabilities when the dynamics of x are described by equation (1). The main task will be to show that each of these probabilities satisfies a (backward) Kolmogorov equation subject to particular boundary conditions. Once this is accomplished, each π_i will be calculated by solving the backward equation with its boundary conditions using a numerical scheme similar to the one described above. It will turn out that the solutions of some of the π_i's will serve as part of the boundary conditions for other π_i's. Consequently, the order in which the π_i's are considered is important.

In order to obtain the partial differential equations for the π_i's it is useful to introduce a stopped process $\{\hat{x}(t^*)\}$ defined by

$$\hat{x}(t^*) = \begin{cases} 0 & \text{if } t^* \geq \tau_{\bar{a}} \\ 0 & \text{if } t^* \geq \tau_{\bar{b}} \\ x(t^*) & \text{otherwise} \end{cases} \tag{92}$$

That is, $\{\hat{x}(t^*)\}$ is the process $\{x(t^*)\}$ killed at $\tau_{\bar{a}}$ and $\tau_{\bar{b}}$, where $\tau_{\bar{a}}$ and $\tau_{\bar{b}}$ are the first hitting times after t of the sets $\{x(t^*) \leq \bar{a}\}$ and $\{x(t^*) \geq \bar{b}\}$, respectively.

5.3.1 The partial differential equation for π_3

We begin with $\pi_3 = \pi(x,t;A_3(t,T)) \equiv P\left[A_3(t,T) \mid x(t)=x\right]$. From this definition it follows that π_3 is given by the expectation

$$\pi(x,t;A_3(t,T)) = E\left[1_{\{\bar{a}<\hat{x}(T)<\bar{b}\}}|\hat{x}(t)=x\right]. \tag{93}$$

where 1_A is the indicator function taking on the value 1 if the event A is realized.

From the characterization of the probability as an expected value in (93), it follows that the process $\{\pi_3\}$ is a martingale. Consequently, as demonstrated in section 3.1 above, an application of Itô's lemma shows that π_3 satisfies the partial differential equation

$$(1/2)\sigma^2(x)\frac{\partial^2\pi_3}{\partial x^2} + \mu(x)\frac{\partial\pi_3}{\partial x} + \frac{\partial\pi_3}{\partial t} = 0.$$

This is of course the backward Kolmogorov equation. It is clear that at the lower boundary $\bar{a}$ and the upper boundary $\bar{b}$ the value of π_3 is zero, because touching either of the boundaries kills the process which guarantees that the expectation in equation (93) will be zero. On the other hand, if the terminal boundary is reached at time T and the process has not been killed, then the expectation will be one. Hence, the boundary conditions are

$$\begin{aligned}
\pi(\bar{b},t^*;A_3(t,T)) &= 0,\\
\pi(\bar{a},t^*;A_3(t,T)) &= 0,\\
\pi(x,T;A_3(t,T)) &= 1 \text{ for } \bar{a}<x<\bar{b}.
\end{aligned}$$

5.3.2 The partial differential equation for π_2

We next consider $\pi_2 = \pi(x,t;A_2(t,T)) \equiv P\left[A_2(t,T) \mid x(t)=x\right]$. From this definition it follows that π_2 is given by the expectation

$$\pi(x,t;A_2(t,T)) = E\left[1_{\{\bar{a}<\hat{x}(T)<\bar{b}\}\cap\{\tau_a\le T\}}|\hat{x}(t)=x\right]. \tag{94}$$

Since π_2 can be written as an expectation, applying Itô's lemma once again shows that π_2 satisfies

$$(1/2)\sigma^2(x)\frac{\partial^2\pi_2}{\partial x^2} + \mu(x)\frac{\partial\pi_2}{\partial x} + \frac{\partial\pi_2}{\partial t} = 0.$$

The boundary condition at $\bar{b}$ is zero, because touching the boundary at $\bar{b}$ kills the process. The boundary condition at time T in the spatial region $a<x<\bar{b}$ is zero, because once T is reached there is no time left for the process to satisfy $\tau_a \le T$ by hitting a. Finally, if the process first hits a at a time $t^*=\tau_a, t<t^*<T$, then from time t^* forward satisfying $A_2(t,T)$ is equivalent to

satisfying $A_3(t^*, T)$. Accordingly, π_2's boundary condition at spatial level a is equal to the solution of π_3 at a. The boundary conditions for π_2 then can be written

$$\begin{aligned}
\pi(\bar{b}, t^*; A_2(t,T)) &= 0, \\
\pi(a, t^*; A_2(t,T)) &= \pi(a, t^*; A_3(t,T)), \\
\pi(x, T; A_2(t,T)) &= 0 \text{ for } a < x < \bar{b}.
\end{aligned}$$

5.3.3 The partial differential equation for π_1

The probability π_1 is just like the probability π_2 with the roles of the lower and upper boundaries exchanged. Accordingly,

$$\pi(x, t; A_1(t,T)) = E\left[1_{\{\bar{a} < \hat{x}(T) < \bar{b}\} \cap \{\tau_b \leq T\}} | \hat{x}(t) = x\right]. \tag{95}$$

and an application of Itô's lemma shows that π_1 satisfies

$$(1/2)\sigma^2(x)\frac{\partial^2 \pi_1}{\partial x^2} + \mu(x)\frac{\partial \pi_1}{\partial x} + \frac{\partial \pi_1}{\partial t} = 0.$$

The boundary condition at $\bar{a}$ is zero, because touching the boundary at $\bar{a}$ kills the process. The boundary condition at time T in the spatial region $\bar{a} < x < b$ is zero, because once T is reached there is no time left for the process to satisfy $\tau_b < T$ by hitting b. Finally, if the process first hits b at a time $t^* = \tau_b, t < t^* < T$, then from time t^* forward satisfying $A_1(t,T)$ is equivalent to satisfying $A_3(t^*, T)$. Accordingly, π_1's boundary condition at spatial level b is equal to the solution of π_3 at b. The boundary conditions for π_1 then can be written

$$\begin{aligned}
\pi(b, t^*; A_1(t,T)) &= \pi(b, t^*; A_3(t,T)), \\
\pi(\bar{a}, t^*; A_1(t,T)) &= 0, \\
\pi(x, T; A_1(t,T)) &= 0 \text{ for } \bar{a} < x < b.
\end{aligned}$$

5.3.4 The partial differential equation for π_0

Finally, we consider $\pi_0 = \pi(x, t; A_0(t,T)) \equiv P\left[A_0(t,T) \mid x(t) = x\right]$. In order to analyze this case, we will break the event $A_0(t,T)$ into two exhaustive and disjoint events. Let $A_0^L(t,T)$ be the subset of $A_0(t,T)$ where the lower boundary a is first hit before the upper boundary b, and let $A_0^U(t,T)$ be the subset of $A_0(t,T)$ where the upper boundary b is first hit before the lower boundary a

- $A_0^L(t,T) \equiv A_0(t,T) \bigcap [\tau_a < \tau_b]$;
- $A_0^U(t,T) \equiv A_0(t,T) \bigcap [\tau_b < \tau_a]$.

Corresponding to these events we define the probabilities

- $\pi(x,t;A_0^L(t,T)) \equiv P[A_0^L(t,T) \mid x(t)=x]$;
- $\pi(x,t;A_0^U(t,T)) \equiv P[A_0^U(t,T) \mid x(t)=x]$.

From these definitions it follows that the probabilities π_0^L and π_0^U can be expressed as the following expectations

$$\pi(x,t;A_0^L(t,T)) = E\left[1_{\{\bar{a}<\hat{x}(T)<\bar{b}\}\cap\{\tau_a<\tau_b\leq T\}}|\hat{x}(t)=x\right], \quad (96)$$

$$\pi(x,t;A_0^U(t,T)) = E\left[1_{\{\bar{a}<\hat{x}(T)<\bar{b}\}\cap\{\tau_b<\tau_a\leq T\}}|\hat{x}(t)=x\right]. \quad (97)$$

Since these probabilities can be written as expectations, applying Itô's lemma shows that they satisfy

$$\begin{aligned}\frac{1}{2}\sigma^2(x)\frac{\partial^2\pi_0^L}{\partial x^2}+\mu(x)\frac{\partial\pi_0^L}{\partial x}+\frac{\partial\pi_0^L}{\partial t} &= 0,\\ \frac{1}{2}\sigma^2(x)\frac{\partial^2\pi_0^U}{\partial x^2}+\mu(x)\frac{\partial\pi_0^U}{\partial x}+\frac{\partial\pi_0^U}{\partial t} &= 0.\end{aligned}$$

Consider the boundary conditions for π_0^L first. For this case, the process first hits a before it first hits b. When the process first hits a at a time $t^*=\tau_a, t<t^*<T$, then from time t^* forward satisfying $A_0^L(t,T)$ is equivalent to satisfying $A_1(t^*,T)$. Accordingly, π_0^L's boundary condition at spatial level a is equal to the solution of π_1 at a. The terminal boundary condition at time T in the spatial region $a<x<b$ is zero, because once T is reached there is no time left for the process to satisfy $\tau_a<\tau_b\leq T$ by first hitting a and then hitting b. Finally, if the process first hits b before it first hits a, then the expectation in equation (96) will be zero, because it will be impossible to satisfy $\tau_a<\tau_b\leq T$. Hence, the boundary condition at the spatial level b is zero. The boundary conditions for π_0^L then can be written

$$\begin{aligned}\pi(b,t^*,A_0^L(t,T)) &= 0,\\ \pi(a,t^*;A_0^L(t,T)) &= \pi(a,t^*;A_1(t,T)),\\ \pi(x,T;A_0^L(t,T)) &= 0 \text{ for } a<x<b.\end{aligned}$$

A similar analysis of π_0^U indicates that its boundary conditions can be written

$$\begin{aligned}
\pi(b, t^*, A_0^U(t,T)) &= \pi(b, t^*; A_2(t,T)), \\
\pi(a, t^*; A_0^U(t,T)) &= 0, \\
\pi(x, T; A_0^U(t,T)) &= 0 \text{ for } a < x < b.
\end{aligned}$$

The probabilities π_0^L and π_0^U both obey the Kolmogorov backward equation with distinct sets of boundary conditions. Since π_0 is the sum of π_0^L and π_0^U, the linearity of the backward equation entails that π_0 obeys the backward equation with boundary conditions which are the sum of the boundary conditions for π_0^L and π_0^U. Hence, π_0 obeys

$$\frac{1}{2}\sigma^2(x)\frac{\partial^2 \pi_0}{\partial x^2} + \mu(x)\frac{\partial \pi_0}{\partial x} + \frac{\partial \pi_0}{\partial t} = 0 \tag{98}$$

with boundary conditions

$$\begin{aligned}
\pi(b, t^*, A_0(t,T)) &= \pi(b, t^*; A_2(t,T)), \\
\pi(a, t^*; A_0(t,T)) &= \pi(a, t^*; A_1(t,T)), \\
\pi(x, T; A_0(t,T)) &= 0 \text{ for } a < x < b.
\end{aligned}$$

At this point we have shown that each of the probabilities π_i for $i \in \{0, 1, 2, 3\}$ satisfies the Kolmogorov backward partial differential equation with certain boundary conditions. For any given data point, the solution of only one of these backward equations and associated boundary conditions is needed to compute the conditional expected change over a finite time interval via equation (91). Which backward equation and set of boundary conditions are solved for the data point observed at time t depends on the history of the process up until time t.

A further complication is that part of the boundary conditions for π_1 and π_2 comes from the solution of π_3. Hence, if the past history of the process dictates that for a given data point the finite time conditional expected return should be computed from π_1 or π_2, then π_3 must be solved first with its boundary conditions in order to get the boundary conditions to solve for π_1 or π_2. Similarly, parts of the boundary conditions for π_0 come from the solutions of π_1 and π_2. Hence, if the past history of the process dictates that for a given data point the finite time conditional expected return should be computed from π_0, then π_3 must first be solved to provide the boundary conditions for π_1 and π_2. Then π_1 and π_2 must be solved to provide the boundary conditions for π_0. Only then can π_0 be solved. The partial differential equations for the various π_i's are computed using a numerical scheme like the one described in Section 4.

5.3.5 Solutions for the π_i

Figures 15–18 show the solutions for π_3, π_2, π_1, and π_0. All four solutions were computed using the unrestricted parameter estimates reported in Table 3 below.

5.4 Computation of $v(x,t,t+\delta t\,|A_i(t,T)) = E[\bar{\pi}(x,t,y,t+\delta t; A_i(t,t+\delta t)) \times \pi(y,t+\delta t; A_i(t+\delta t,T))\, y|x(t)=x]$

In order to use equation (91) to calculate the finite time expected change, in addition to $\pi(x,t;A_i(t,T))$ we must compute

$$v(x,t,t+\delta t\,|A_i(t,T)) = E\left[\bar{\pi}(x,t,y,t+\delta t; A_i(t,t+\delta t))\,\pi(y,t+\delta t; A_i(t+\delta t,T))\,y|x(t)=x\right]. \tag{99}$$

This subsection explains how to compute this quantity.

From equation (38) we get

$$v(x,t,t+\delta t\,|A_i(t,T)) \equiv E\left[\bar{\pi}(x,t,y,t+\delta t; A_i(t,t+\delta t))\,\pi(y,t+\delta t; A_i(t+\delta t,T))\,y|x(t)=x\right]. \tag{100}$$

Then from equation (40), the $v(x,t,t+\delta t\,|A_i(t,T))$ obey the backward equation

$$\frac{1}{2}\sigma^2(x)\frac{\partial v(x,t,t+\delta t\,|A_i(t,T))}{\partial x^2} + \mu(x)\frac{\partial v(x,t,t+\delta t\,|A_i(t,T))}{\partial x} + \frac{\partial v(x,t,t+\delta t\,|A_i(t,T))}{\partial t} = 0. \tag{101}$$

We now derive the boundary conditions for the v_i. As in the case of the π_i, the solutions to some of the v_i serve as boundary conditions for other v_i. The first line of each of the three line derivations follows directly from the definition of the v_i given by equation (38). We begin with the boundary conditions for v_3. At the spatial boundary $\bar{a}$,

$$\begin{aligned} v(\bar{a},t,t+\delta t\,|A_3(t,T)) &= E\left[\bar{\pi}(\bar{a},t,y,t+\delta t; A_3(t,t+\delta t))\,\pi(y,t+\delta t; A_3(t+\delta t,T))\,y|x(t)=\bar{a}\right] \\ &= E\left[0\times\pi(y,t+\delta t; A_3(t+\delta t,T))\,y|x(t)=\bar{a}\right] \\ &= 0. \end{aligned}$$

The second line follows from the first, because at time t the process is at $\bar{a}$ which violates the condition $A_3(t,t+\delta t)$. At the spatial boundary $\bar{b}$,

$$\begin{aligned} v(\bar{b},t,t+\delta t\,|A_3(t,T)) &= E\left[\bar{\pi}(\bar{b},t,y,t+\delta t; A_3(t,t+\delta t))\,\pi(y,t+\delta t; A_3(t+\delta t,T))\,y|x(t)=\bar{b}\right] \\ &= E\left[0\times\pi(y,t+\delta t; A_3(t+\delta t,T))\,y|x(t)=\bar{b}\right] \\ &= 0. \end{aligned}$$

Now the second line follows from the first, because at time t the process is at $\bar{b}$ which also violates the condition $A_3(t, t+\delta t)$. At the terminal boundary,

$$\begin{aligned} v(x, t+\delta t, t+\delta t \,|A_3(t,T)) &= E[\bar{\pi}(x, t+\delta t, x, t+\delta t; A_3(t+\delta t, t+\delta t)) \\ &\quad \times \pi(x, t+\delta t; A_3(t+\delta t, T))\, x | x(t+\delta t) = x] \\ &= E\left[1 \times \pi(x, t+\delta t; A_3(t+\delta t, T))\, x | x(t+\delta t) = x\right] \\ &= \pi(x, t+\delta t; A_3(t+\delta t, T))\, x, \text{ for } x \in (\bar{a}, \bar{b}). \end{aligned}$$

Here the second line follows from the first because there is no time for the conditioning event A_3 to be violated.

For v_2 at the spatial boundary a,

$$\begin{aligned} v(a, t, t+\delta t \,|A_2(t,T)) &= E\left[\bar{\pi}(a, t, y, t+\delta t; A_2(t, t+\delta t))\, \pi(y, t+\delta t; A_2(t+\delta t, T))\, y | x(t) = a\right] \\ &= E\left[\bar{\pi}(a, t, y, t+\delta t; A_3(t, t+\delta t))\, \pi(y, t+\delta t; A_3(t+\delta t, T))\, y | x(t) = a\right] \\ &= v(a, t, t+\delta t; A_3(t, T)). \end{aligned}$$

In this derivation, the second line follows from the first, because at time t the process is at a and once a is reached the conditioning event A_2 collapses to the conditioning event A_3. The third line follows from the second by the definition in equation (38).

At the spatial boundary $\bar{b}$,

$$\begin{aligned} v(\bar{b}, t, t+\delta t \,|A_2(t,T)) &= E\left[\bar{\pi}(\bar{b}, t, y, t+\delta t; A_2(t, t+\delta t))\, \pi(y, t+\delta t; A_2(t+\delta t, T))\, y | x(t) = \bar{b}\right] \\ &= E\left[0 \times \pi(y, t+\delta t; A_2(t+\delta t, T))\, y | x(t) = \bar{b}\right] \\ &= 0. \end{aligned}$$

Here the second line follows from the first, because at time t the process is at $\bar{b}$ which violates the condition $A_2(t, t+\delta t)$. At the terminal boundary,

$$\begin{aligned} v(x, t+\delta t, t+\delta t \,|A_2(t,T)) &= E[\bar{\pi}(x, t+\delta t, x, t+\delta t; A_2(t+\delta t, t+\delta t)) \\ &\quad \times \pi(x, t+\delta t; A_2(t+\delta t, T))\, x | x(t+\delta t) = x] \\ &= E\left[1 \times \pi(x, t+\delta t; A_2(t+\delta t, T))\, x | x(t+\delta t) = x\right] \\ &= \pi(x, t+\delta t; A_2(t+\delta t, T))\, x, \text{ for } x \in (a, \bar{b}). \end{aligned}$$

The boundary conditions for v_1 are obtained in a similar way to those for v_2 except the roles of the lower and upper boundaries are switched. We have

$$\begin{aligned} v(\bar{a}, t, t+\delta t \,|A_1(t,T)) &= E\left[\bar{\pi}(\bar{a}, t, y, t+\delta t; A_1(t, t+\delta t))\, \pi(y, t+\delta t; A_1(t+\delta t, T))\, y | x(t) = \bar{a}\right] \\ &= E\left[0 \times \pi(y, t+\delta t; A_1(t+\delta t, T))\, y | x(t) = \bar{a}\right] \\ &= 0, \end{aligned}$$

$$\begin{aligned} v(b, t, t+\delta t \,|A_1(t,T)) &= E\left[\bar{\pi}(b, t, y, t+\delta t; A_1(t, t+\delta t))\, \pi(y, t+\delta t; A_1(t+\delta t, T))\, y | x(t) = b\right] \\ &= E\left[\bar{\pi}(b, t, y, t+\delta t; A_3(t, t+\delta t))\, \pi(y, t+\delta t; A_3(t+\delta t, T))\, y | x(t) = b\right] \\ &= v(b, t, t+\delta t; A_3(t,T)), \end{aligned}$$

$$\begin{aligned} v(x, t+\delta t, t+\delta t \,|A_1(t,T)) &= E[\bar{\pi}(x, t+\delta t, x, t+\delta t; A_1(t+\delta t, t+\delta t)) \\ &\quad \times \pi(x, t+\delta t; A_1(t+\delta t, T))\, x | x(t+\delta t) = x] \\ &= E\left[1 \times \pi(x, t+\delta t; A_1(t+\delta t, T))\, x | x(t+\delta t) = x\right] \\ &= \pi(x, t+\delta t; A_1(t+\delta t, T))\, x, \text{ for } x \in (\bar{a}, b). \end{aligned}$$

Finally, the boundary conditions for v_0 are

$$\begin{aligned} v(a, t, t+\delta t \,|A_0(t,T)) &= E\left[\bar{\pi}(a, t, y, t+\delta t; A_0(t, t+\delta t))\, \pi(y, t+\delta t; A_0(t+\delta t, T))\, y | x(t) = a\right] \\ &= E\left[\bar{\pi}(a, t, y, t+\delta t; A_1(t, t+\delta t))\, \pi(y, t+\delta t; A_1(t+\delta t, T))\, y | x(t) = a\right] \\ &= v(a, t, t+\delta t; A_1(t,T)). \end{aligned}$$

The second line follows from the first, because at time t the process is at a and once a is reached the conditioning event A_0 collapses to the conditioning event A_1. The third line follows from the second by the definition in equation (38).

At b,

$$\begin{aligned} v(b, t, t+\delta t \,|A_0(t,T)) &= E\left[\bar{\pi}(b, t, y, t+\delta t; A_0(t, t+\delta t))\, \pi(y, t+\delta t; A_0(t+\delta t, T))\, y | x(t) = b\right] \\ &= E\left[\bar{\pi}(b, t, y, t+\delta t; A_2(t, t+\delta t))\, \pi(y, t+\delta t; A_2(t+\delta t, T))\, y | x(t) = b\right] \\ &= v(b, t, t+\delta t; A_2(t,T)). \end{aligned}$$

In this derivation, the second line follows from the first, because at time t the process is at b and once b is reached the conditioning event A_0 collapses to the conditioning event A_2. The third line follows from the second by the definition in equation (38).

$$\begin{aligned} v(x, t+\delta t, t+\delta t \,|A_0(t,T)) &= E[\bar{\pi}(x, t+\delta t, x, t+\delta t; A_0(t+\delta t, t+\delta t)) \\ &\quad \times \pi(x, t+\delta t; A_0(t+\delta t, T))\, x | x(t+\delta t) = x] \\ &= E\left[1 \times \pi(x, t+\delta t; A_0(t+\delta t, T))\, x | x(t+\delta t) = x\right] \\ &= \pi(x, t+\delta t; A_0(t+\delta t, T))\, x, \text{ for } x \in (a, b). \end{aligned}$$

To compute the quantity $E\left[\bar{\pi}(x, t, y, t+\delta t; A_i(t, t+\delta t))\, \pi(y, t+\delta t; A_i(t+\delta t, T))\, y | x(t) = x\right]$ which we need to calculate the finite time conditioned means for the GMM moment conditions, we solve the backward equations for v_i using the appropriate boundary conditions. This is accomplished using a numerical scheme like the one described in Section 4. As before, when we are near a boundary we use a finer mesh.

5.5 Discussion

Figure 19 shows the conditioned expected change m_3 at time index 5,250, along with the unconditioned expected change. As with the π_i's above, the solution is computed using the unrestricted estimates in Table 3 below, which explains why the unconditioned expected change is a nonlinear function of the interest level. The time index 5,250 is close to the end of the sample period (there are a total of 5,505 data points), and at this point in the sample period both the minimum and maximum have already been reached, so that the conditioning event has "collapsed" to A_3 and the expected change is given by m_3. The figure shows that the conditioned and unconditioned expected changes are considerably different for high levels of the interest rate.

Figure 20 shows the conditioned expected change m_2 and unconditioned expected change at time index 3,000. This time index is after the maximum has been reached but before the minimum, so that the conditioning event is A_2. The figure shows that the conditioned and unconditioned expected changes are considerably different for high levels of the interest rate.

Figure 21 shows the conditioned expected change m_0 and unconditioned expected change at time index 1,000. This time index is before either the minimum or the maximum have been reached, so the conditioning event is A_0. In this region the conditioned and unconditioned expected changes agree for a wider range of interest rates, but still differ for large values of the interest rate.

5.6 Estimation

To determine how conditioning on reaching particular minimum and maximum values impacts parameter estimation, we apply the GMM procedure to AS's interest rate data to estimate the

CHLS specification both without conditioning and with conditioning on A_0. Now

$$u_i \equiv x_i^{-\beta_2}(\delta x_i) - x_i^{-\beta_2} m(x_i, t_i, t_i + \delta t \,|A_0(t_i, T)) \tag{102}$$

and we minimize $G(\theta) = (1/(N-1))h'(\theta)W(\theta)h(\theta)$ over the vector $\theta = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)$ using the Nelder-Mead simplex method. Once again the numerical scheme is implemented with the mesh constructed so that there are 500 grid points between the minimum and the maximum of the AS data which results in an interval of 4.2836 basis points between grid points. The parameters are estimated conditioning on A_0 with $\bar{a}$ and $\bar{b}$ displaced the same number of grid points from, respectively, below the minimum and above the maximum of the AS interest rate data.

Table 3 shows the parameter estimates for various levels of this displacement. For comparison, it also shows the unconditioned results previously reported in Table 1. As before, the drift parameters α_i are estimated imprecisely, with the absolute values of the parameter estimates always less than twice their standard errors, and often less than their standard errors. Also as before, the estimates of α_1 are positive but the process is nonetheless non-explosive because of the negative signs on the estimates of α_2 and the positive signs on the estimates of α_3. Figure 22 plots the five drift coefficients computed using the five sets of parameter estimates in Table 3.

The results in Table 3 and Figure 22 indicate that conditioning on A_0 affects the estimated drift coefficients. Focusing initially only on the four estimates of the drift coefficient in which we condition on the event A_0 using the four different displacements, it appears that stronger conditioning (i.e., smaller displacements) reduces but does not eliminate the estimated non-linearity. However, the dependence of the drift parameters α_i and estimated drift $\mu(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x$ on the displacement is less than in Table 1 and Figure 13.

To understand the behavior of the drift coefficients in Figure 22, recall that at the upper boundary, conditioning on A_0 entails that the process reaches the level b but does not reach the greater level $\bar{b}$. Consider first the situation when the process has already touched b and is currently at a high level. Then fulfilling A_0 is equivalent to fulfilling A_3 until time T. Hence the results from Section 4 apply. Referring back to Figure 13, we see that in this situation conditioning increases the level of the drift for large interest rate levels and that the magnitude of the increase is increasing in the severity of the conditioning (i.e., is decreasing in the displacement.) This is the first effect at work in Figure 22. Consider now the situation when the current level of the process is high, but b has not yet been touched. The actual data goes on to hit b before T. This fact about the actual data increases the drift coefficient for large values of the interest rate in the absence of conditioning on eventually hitting b. However, conditioning on the fact that the process will go on to hit b makes this increase in the drift coefficient unnecessary. Hence, the data where b has not yet been hit but the level of the process is high will decrease conditioned drift coefficients for high levels of the process relative to the unconditioned case. This decrease should be relatively invariant to the

level of displacement, because the displacement effects only $\bar{b}$ which is of secondary importance in this situation. This is the second effect at work in Figure 22. The two effects together explain the drift coefficients for large values of the interest rate in Figure 22. The first effect increases the conditioned drifts relative to the unconditioned drifts with the increase being greater when the displacement is smaller. The second effect decreases all of the conditioned drifts relative to the unconditioned drift.

Unsurprisingly, and similar to the earlier results, the estimates of the diffusion parameters β_1 and β_2 indicates that the conditioning has very little effect on the diffusion coefficient. Also similar to the earlier results, these parameters are estimated with reasonable precision.

We next re-estimate the CHLS specification without and with conditioning on A_0 under the restriction that $\alpha_2 = \alpha_3 = 0$, so that the drift coefficient is linear. We continue to use the full set of six moment conditions, and minimize the quadratic form over $\theta_R \equiv (\alpha_0, \alpha_1, 0, 0, \beta_1, \beta_2)$. As in Section 4, θ_R is the solution to the minimization problem

$$G_R \equiv \frac{1}{N-1} \min_{\theta_R} h'(\theta_R) W(\theta_R) h(\theta_R). \tag{103}$$

The results are recorded in Table 4 for conditioning with displacements of 60, 40, 20, and 10 grid points. The final column of Table 4 records the value of $G_R - G$ which, as before, is asymptotically distributed Chi-squared with two degrees of freedom. As in the other tables, the first line repeats the results for the unconditioned case previously shown in Table 2.

Examining the rightmost column of the table, the test statistics $G_R - G$ for the displacements of 60, 40, 20, and 10 grid points are larger than the corresponding test statistics in Table 2, but still far too small to reject the null hypothesis of linearity at conventional levels of significance. The p-value of the largest of these for test statistics, 0.979, is 0.613, indicating that there is a 61.3 percent chance of obtaining a larger test statistic even if the null hypothesis of linearity is true. The p-values of the others are between 68.4 and 86.2 percent. As was the case with the earlier results discussed in Section 4, these test statistics fail to reject linearity in the drift of the unconditioned process because the conditioned drift is non-linear even when the unconditioned drift is linear.

Figure 23 plots the drift coefficients from Table 4. Conditioning on A_0 appears to have a large impact on the estimation of the drift coefficient when it is restricted to be linear. The downward pull on the drift coefficient for data where the level of the process is high but b has not been reached yet does not appear to be as important when the drift coefficient is restricted to be linear.

6 $C_3(t,T)$: The discrete-time process $\{x(t_i)\}$ is above $\bar{a}$ and below $\bar{b}$ for times t_i, $i = 1, \ldots, N$

6.1 The event $C_3(t,T)$

So far we have considered events defined in terms of the entire sample path of the continuous-time process $x(u), u \in [0,T]$. For example, equation (44) defined the event $A_3(t,T)$ in terms of the minimum and maximum values $\min_{u\in[t,T]} x(u)$ and $\max_{u\in[t,T]} x(u)$. Events such as A_3 defined in terms of the entire sample path are interesting because observing the entire sample path is a natural idealization of observing the process at a set of frequently spaced dates $\{t_i\}$, and it can be easier to work with events defined in terms of the entire sample path. Further, some events such as the survival of an equity market are naturally defined as the entire sample path remaining above some lower boundary (see, e.g. Brown, Goetzman, and Ross (1995)). Nonetheless, the econometrician or other analyst typically does not observe the entire sample path, but instead observes the process at a set of dates $\{t_i\}$, i.e. he or she observes the discrete-time process $\{x(t_i)\}$. Conditioning on the minimum, maximum, or other features of the process then means conditioning on features of the process $\{x(t_i)\}$.

In this section we consider the discrete-time analog of the event A_3 analyzed in Section 4. We suppose that the process is observed at times t_i, $i = 1, \ldots, N$, and condition on the event that the discrete-time process $\{x(t_i)\}$ never reaches a lower boundary $\bar{a}$ or an upper boundary $\bar{b}$ from time $t_1 = 0$ to $t_N = T$. Specifically, the event is defined by

$$C_3(t,T) \equiv \left[\min_{i\in\{1,2,\ldots,N\}} x(t_i) > \bar{a}\right] \bigcap \left[\max_{i\in\{1,2,\ldots,N\}} x(t_i) < \bar{b}\right]. \tag{104}$$

Below we describe the modifications needed to handle this more complicated event. The principal difference is that the computation of $\pi(x,t;C_3(t,T))$, i.e. the probability that the process satisfies the conditioning event $C_3(t,T)$, becomes more difficult.

6.2 Numerical approach

We assume that the times t_i are equally spaced with $t_{i+1} = t_i + \delta t$, though the equal spacing is not essential. As in section 4.3, the starting point for this computation is equation (36), now specialized to the particular conditioning event C_3:

$$\begin{aligned} m(x,t_i,t_i+\delta t\,|C_3(t_i,T)) &= \frac{1}{\pi^i(x,t_i;C_3(t_i,T))} \\ &\times E\left[\bar{\pi}(x,t_i,y,t_i+\delta t;C_3(t_i,t_i+\delta t))\,\pi^i(y,t_i+\delta t;C_3(t_i+\delta t,T))\,y|x(t_i)=x\right] - x, \end{aligned} \tag{105}$$

where $\pi^i(x,t;C_3(t,T))$ denotes the solution on the interval $[t_i,t_{i+1}]$. We index the probabilities by the superscripts $i=1,\ldots,N-1$ because the solution for one time interval provides the boundary condition for the next, and we need to distinguish between them. Again, the approach will be separately to compute the two terms $\pi^i(x,t_i;C_3(t_i,T))$ and $E[\bar{\pi}(x,t_i,\,y,t_i+\delta t;C_3(t_i,t_i+\delta t))\,\pi^i(y,t_i+\delta t;C_3(t_i+\delta t,T))\,y|x(t_i)=x]$, and then combine the results to obtain $m(x,t_i,t_i+\delta t\,|C_3(t_i,T))$.

6.2.1 Computation of $\pi^i(x,t_i;C_3(t_i,T))$

The new difficulty that arises with the event C_3 is that the probabilities $\pi^i(x,t;C_3(t,T))$ do not satisfy a single partial differential equation on a region of the form $(\bar{a},\bar{b})\times[0,T]$, but instead satisfies a set of N (in our case, 5,504) differential equations on the N regions $(0,+\infty)\times[t_i,t_{i+1}]$. The partial differential equation on the time interval $[t_i,t_{i+1}]=[t_i,t_i+\delta t]$ is linked to the solution on the next time interval $[t_{i+1},t_{i+2}]$ through a terminal boundary condition imposed at time t_{i+1}.

To understand this, start with the last interval $[t_{N-1},t_N]=[T-\delta t,T]$. For times t in this interval the probability is just $\pi^{N-1}(x,t;C_3(t,T))=E[1_{\{\bar{a}<x(T)<\bar{b}\}}|x(t)=x]$. It satisfies the backward equation (26) for $t\in[t_{N-1},T]$, and at time T satisfies the terminal boundary condition

$$\pi^{N-1}(x,T;C_3(T,T))=1_{\{\bar{a}<x<\bar{b}\}}.$$

Spatial boundary conditions are not imposed at $\bar{a}$ and $\bar{b}$ because the event C_3 does not place any restriction on the path the process takes at times between (but not equal to) $t_{N-1}=T-\delta t$ and $t_N=T$; it places restrictions only on the $\{x(t_i)\}$.

Now step back δt time units, and consider the interval $[t_{N-2},t_{N-1}]=[T-2\delta t,T-\delta t]$. In this interval, the probability $\pi^{N-2}(x,t;C_3(t,T))=E[1_{\{\bar{a}<y=x(t_{N-1})<\bar{b}\}}\times\pi^{N-1}(y,t_{N-1};C_3(t_{N-1},T))\,|\,x(t)=x]$. Clearly it satisfies the backward equation (26) for $t\in[t_{N-2},t_{N-1}]$, and at time t_{N-1} satisfies the terminal boundary condition

$$\pi^{N-2}(x,t_{N-2};C_3(t_{N-1},T))=1_{\{\bar{a}<x<\bar{b}\}}\times\pi^{N-1}(x,t_{N-1};C_3(t_{N-1},T)).$$

Similar reasoning establishes that, for $t\in[t_i,t_{t+1}]$, the probability $\pi^i(x,t;C_3(t_i,T))$ satisfies the backward equation

$$\frac{1}{2}\sigma^2(x)\frac{\partial^2\pi^i(x,t;C_3(t,T))}{\partial x^2}+\mu(x)\frac{\partial\pi^i(x,t;C_3(t,T))}{\partial x}+\frac{\partial\pi^i(x,t;C_3(t,T))}{\partial t}=0, \tag{106}$$

and at time t_{i+1} satisfies the terminal boundary condition

$$\pi^i(x,t_{i+1};C_3(t_{i+1},T))=1_{\{\bar{a}<x<\bar{b}\}}\times\pi^{i+1}(x,t_{i+1};C_3(t_{i+1},T)). \tag{107}$$

We compute the probabilities $\pi^i(x, t_i; C_3(t_i, T))$ by solving the set of N partial differential equations (106) "backward" from time $t_N = T$ to time $t_1 = 0$. Within each interval $[t_i, t_{i+1}]$ we use a fully implicit finite difference scheme,[10] and at the boundaries between the intervals we use the boundary condition (107) connecting the solution in each interval to that in the next. Since each of the equations (106) holds on a semi-infinite region $(0, +\infty) \times [t_i, t_{i+1}]$, we truncate the mesh at points much smaller and larger than $\bar{a}$ and $\bar{b}$, respectively, and set the probability equal to zero at these points.

We solve these partial differential equations using the fully implicit scheme with a spatial grid size of $\Delta x = 4.2836$ basis points. For the most part, the time increment in the finite difference scheme Δt is set equal to the interval δt between data points, so that the region $[t_i, t_i + \delta t]$ over which each of the partial differential equations (106) hold is traversed using just one finite difference time step. However, we suspect that this will provide a poor approximation of the solution when $x \approx \bar{a}$ or $\bar{b}$. Thus, we use a finer mesh when either of the two adjacent interest rate data points is near either of the two boundaries $\bar{a}$ or $\bar{b}$.

6.2.2 Computation of $E[\bar{\pi}(x, t_i, y, t_i + \delta t; C_3(t_i, t_i + \delta t))\ \pi^i(y, t_i + \delta t;\ C_3(t_i + \delta t, T))\, y | x(t_i) = x]$ and $m(x, t_i, t_i + \delta t\,|C_3(t_i, T))$

To compute $m(x, t_i, t_i + \delta t\,|C_3(t_i, T))$ using equation (105) we must first compute $E[\bar{\pi}(x, t_i, y, t_i + \delta t; C_3(t_i, t_i + \delta t))\ \pi^i(y, t_i + \delta t;\ C_3(t_i + \delta t, T))\, y | x(t_i) = x]$. This computation is very similar to the computation of $E[\bar{\pi}_3 \pi_3 |\, x(t) = x]$ in section 4.3.2.

Following equation (38), define

$$v(x, t, t_i + \delta t\,|C_3(t_i + \delta t, T)) \equiv E\left[\bar{\pi}(x, t, y, t_i + \delta t; C_3(t, t_i + \delta t))\, \pi^i(y, t_i + \delta t; C_3(t_i + \delta t, T))\, y | x(t) = x\right], \tag{108}$$

which is analogous to (81). Then $v(x, t, t_i + \delta t\,|C_3(t_i + \delta t, T))$ satisfies the partial differential equation

$$\begin{aligned} &\frac{1}{2}\sigma^2(x)\frac{\partial v(x, t, t_i + \delta t\,|C_3(t_i + \delta t, T))}{\partial x^2} \\ &\quad + \mu(x)\frac{\partial v(x, t, t_i + \delta t\,|C_3(t_i + \delta t, T))}{\partial x} + \frac{\partial v(x, t, t_i + \delta t\,|C_3(t_i + \delta t, T))}{\partial t} = 0, \end{aligned} \tag{109}$$

together with the appropriate boundary conditions. For the interval $[t_i, t_{i+1}]$, the terminal boundary condition is

[10] We compute the solutions using a fully implicit scheme instead of the Crank-Nicholson scheme used earlier due to the discontinuities in the boundary condition (107) at $\bar{a}$ and $\bar{b}$. The Crank-Nicholson scheme uses the function values on the terminal boundary in its approximation of the spatial derivatives, which makes no sense when there is a discontinuity in the terminal boundary condition.

$$
\begin{aligned}
& v(x, t_i + \delta t, t_i + \delta t \,|C_3(t_i + \delta t, T)) \\
&= E\left[\bar{\pi}(x, t_i + \delta t, x, t_i + \delta t; C_3(t_i + \delta t, t_i + \delta t))\, \pi^i(x, t_i + \delta t; C_3(t_i + \delta t, T))\, x | x(t_i + \delta t) = x\right] \quad (10) \\
&= E\left[1_{\{\bar{a} < x < \bar{b}\}} \times \pi^i(x, t_i + \delta t; C_3(t_i + \delta t, T))\, x | x(t_i + \delta t) = x\right] \\
&= \pi^i(x, t_i + \delta t; C_3(t_i + \delta t, T)) x \times 1_{\{\bar{a} < x < \bar{b}\}}.
\end{aligned}
$$

Spatial boundary conditions are not needed, because the partial differential equation holds on the region $(0, +\infty) \times [t_i, t_i + \delta t]$ and 0 and $+\infty$ cannot be reached with positive probability.

We compute the conditioned means $m(x, t_i, t_i + \delta t \,|C_3(t_i, T))$ for each interval of the form $[t_i, t_i + \delta t]$ for which we have data by solving this partial differential equation for each such interval. As in the computation of the probability π^i, we use a fully implicit scheme, and for the most part, we solve these partial differential equations using just one time step to cover the interval $[t_i, t_i + \delta t]$. Again, we use a spatial grid size of $\Delta x = 4.2836$ basis points and truncate the mesh at points well below and above $\bar{a}$ and $\bar{b}$, respectively.

6.3 Estimation

Using this new event C_3, we repeat the estimation exercise reported in section 4.3.5 for the event A_3. The results are presented in Tables 5 and 6, which parallel Tables 1 and 2 discussed in section 4.3.5. The estimates of the drifts are graphed in Figures 24 and 25, which parallel Figures 13 and 14. The results for the two different events are broadly similar, which is to be expected because the event C_3 is the discrete-time analogue of A_3. In particular, examining the rightmost column of Table 6, the test statistics $G_R - G$ for the displacements of 60, 40, 20, and 10 grid points are still far too small to reject the null hypothesis of linearity at conventional levels of significance.

The principal difference between the results for events C_3 and A_3 is that, with the same displacement, conditioning on the event C_3 has a smaller effect on the estimates of the drift parameters α_i and drift $\mu(x)$ than does conditioning on the event A_3. This is unsurprising, because C_3 involves "less" conditioning, in that with the same displacement A_3 rules out paths that C_3 does not. In fact, comparing Figures 24 and 25 showing results for the event C_3 to Figures 13 and 14 showing results for the event A_3, we see that for any given level of displacement the drift conditioned on C_3 is closer to the unconditioned drift than the drift conditioned on A_3.

7 Conclusion

The majority of empirical work in finance analyzes data that are generated naturally rather than experimentally. The researcher often observes one historical draw from some data generating process which is then used to estimate a model of the process. Although estimates of the parameters

which are not conditional on the particular historic sample are often desired, the actual estimates are necessarily conditional on the observed data. Recent research on survivorship bias in equity returns and the estimation of term structure models from time-series of interest rate data suggests that failing to account for the conditional nature of observed data can seriously bias the results of research in these areas. In addition to this literature, two Monte Carlo experiments presented above show that the estimate of the drift coefficient of a univariate diffusion from a time series of data can be severely biased, because the underlying process can produce paths with a wide variety of minima and maxima over a time period equal to the length of the sample but the historical sample necessarily has a particular minimum and maximum value. Hence, there is a need for tools that make it possible to adjust for various types of conditioning when estimating models used in finance.

This paper develops theoretical and numerical tools that make it possible to account for the conditional nature of observed data when the underlying data generating process is assumed to be a time-homogeneous univariate diffusion. In particular, it derives expressions for the conditioned drift and diffusion coefficients in terms of the probabilities that a conditioning event will be satisfied. It also shows that the probabilities of the conditioning event as well as the finite time conditioned mean satisfy parabolic partial differential equations with boundary conditions that depend upon the conditioning event. Numerical techniques for solving these partial differential equations and associated boundary conditions are devised for three specific conditioning events that are of interest in finance, namely, that over a specific period of time: (1) a continuously monitored process stays between upper and lower boundaries, (2) a continuously monitored process has specific minimum and maximum values, and (3) discrete observations of a process remain between minimum and maximum values.

The paper also derives explicit expressions for the conditioned drift and diffusion coefficients and the finite time conditioned mean for the case where the unconditioned process obeys geometric Brownian motion and the conditioning event is that the process stays in between an upper and a lower boundary for some specified length of time. These expressions indicate that near the boundaries there can be large differences between the true expected one trade date change in the value of the process and the approximation to this change formed by multiplying the conditioned drift coefficient by a one trade date time interval. 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 GMM type estimation procedures.

In order to illustrate the techniques that are developed in the paper, estimates are computed for the parameters of flexibly specified drift and diffusion coefficients of a term-structure model from a standard time-series of interest rate data. The estimates are computed without conditioning and also with conditioning on the process (1) staying between continuously monitored upper and lower

boundaries, (2) having specific continuously monitored minimum and maximum values, and (3) remaining between minimum and maximum values at discrete dates. The estimates suggest that in this context all three types of conditioning have an important impact on the estimated drift coefficient but little effect on the estimated diffusion coefficient. Since there has recently been a lot of interest in the linearity of the drift coefficient of univariate diffusion term-structure models, the estimation is also repeated with the drift coefficient restricted to a linear specification. A test statistic fails to reject the linearity of the diffusion coefficient of the underlying process for all of the various conditioning events.

References

Abhyankar, Abhay and Devraj Basu, 2000, Does Conditioning Information Matter in Estimating Continuous Time Interest Rate Difffusions?, working paper, Warwick Business School.

Aït-Sahalia, Yacine, 1996, Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies* 9, No. 2, pp. 385–426.

Brown, Stephen J., William N. Goetzmann, and Stephen A. Ross, 1995, Survival, *Journal of Finance* L, No. 3 (July), pp. 853–873.

Chan, K. C., Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders, 1992, An Empirical Comparison of Alternative Models of the Short-Term Interest Rate, *Journal of Finance* 47, No. 3 (July), pp. 1209–1227.

Chapman, David A. and Neil D. Pearson, 2000, Is the Short Rate Drift Actually Nonlinear, *Journal of Finance* 55, No. 1 (February), pp. 355–388.

Conley, Timothy C., Lars Peter Hansen, Erzo G. J. Luttmer, and José A. Scheinkman, 1997, Short-Term Interest Rates as Subordinated Diffusions, *Review of Financial Studies* 10, No. 3, pp. 525–577.

Courtadon, Georges, 1982, A More Accurate Finite Difference Approximation for the Valuation of Options, *Journal of Financial and Quantative Analysis* 17, 5, pp. 697–703

Cox, John C., Jonathon Ingersoll, and Stephen A. Ross, 1985, A Theory of the Term Structure of Interest Rates, *Econometrica* 53, pp. 385–407.

Douady, Raphaël, 1998, Closed Form Formulas for Exotic Options and Their Lifetime Distribution, *International Journal of Theoretical and Applied Finance*, 2, No. 1, pp. 17–42

Goetzmann, William and Philippe Jorion, 1999, Re-Emerging Markets, *Journal of Financial and Quantitative Analysis* 34, 1–32.

Hansen, Lars Peter, 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* 50, pp. 1029–1054.

Karlin, Samuel and Howard M. Taylor, 1981, *A Second Course on Stochastic Processes*, Academic Press.

Kunitomo, Naoto and Ikeda Masayuki, 1992, Pricing Options with Curved Boundaries, *Mathematical Finance* 2, No. 4, pp. 275–298.

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

Stanton, Richard, 1997, Nonparametric Model of Term Structure Dynamics, *Journal of Finance*, 52, No. 5, pp. 1973–2002

Sundaresan, Suresh, 2000, Continuous-Time Methods in Finance: A Review and Assessment, *Journal of Finance* 55, No. 4 (August), 1515–1567.

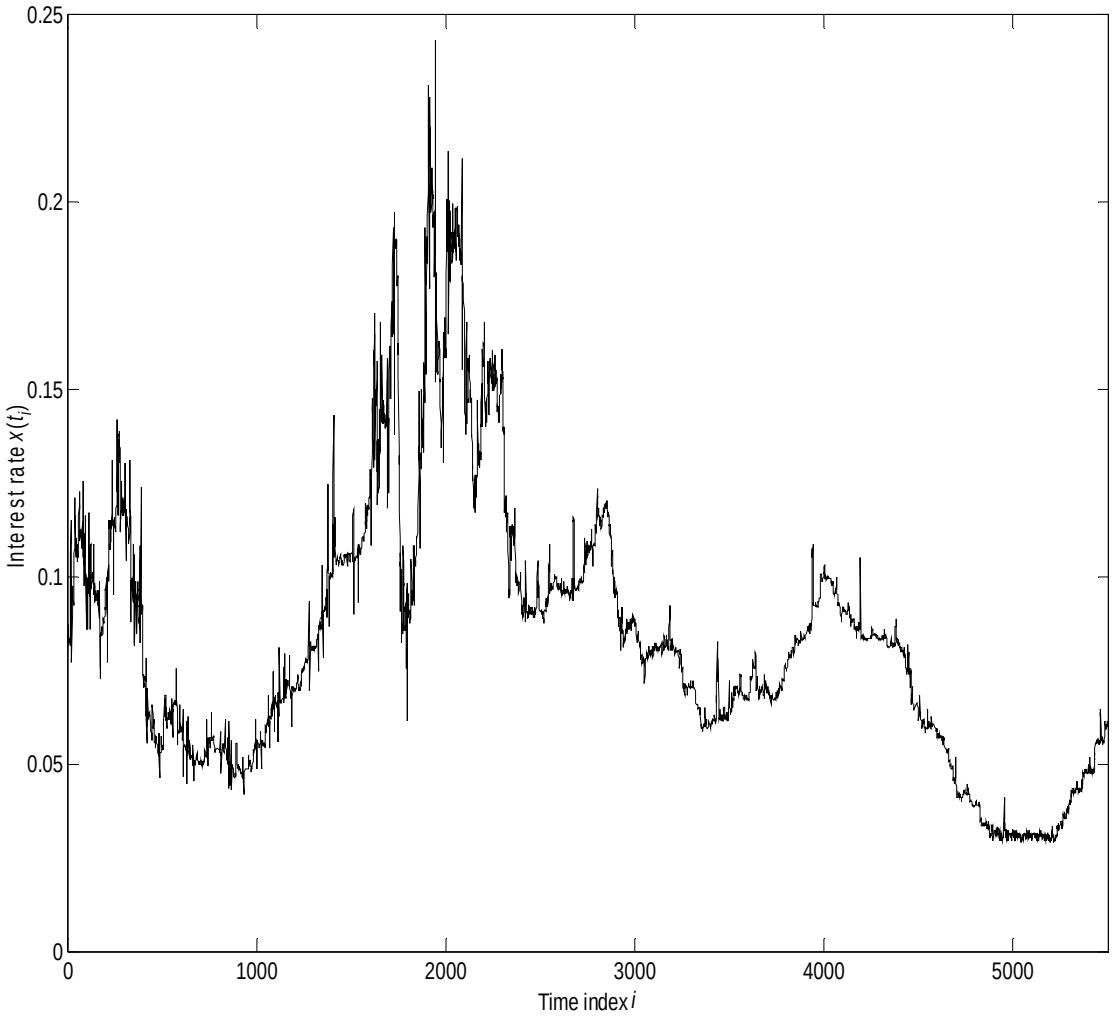

Figure 1: Seven day Eurodollar rates, June 1, 1973 to February 25, 1995. Seven day Eurodollar deposit spot rate bid-ask midpoint, June 1, 1973 to February 25, 1995.

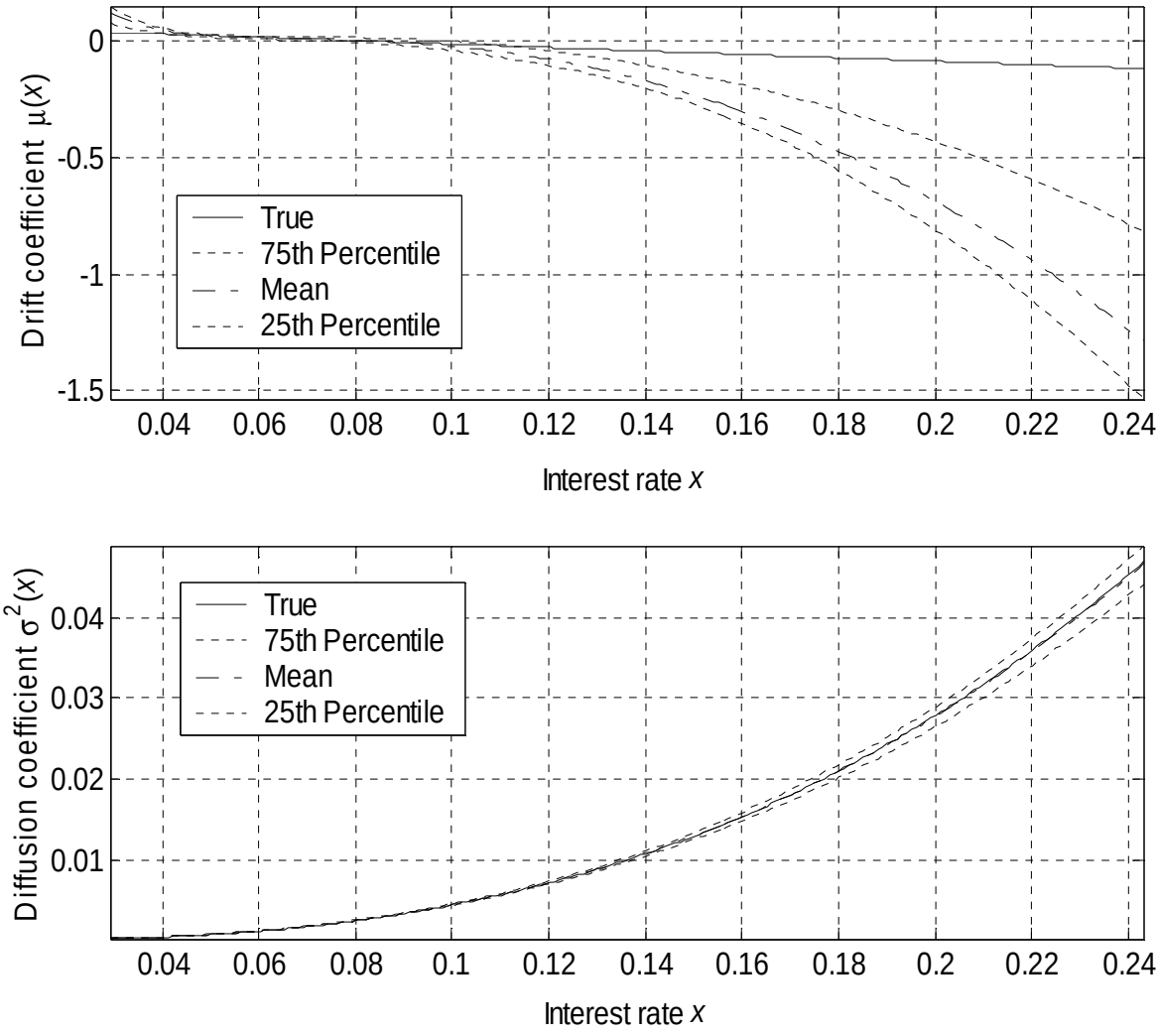

Figure 2: Drift and diffusion coefficients estimated from paths generated by linear drift coefficient $\mu(x) = 0.063 - 0.74x$ that are conditioned on minimum being greater than 0.02915 and maximum being less than 0.24333. The diffusion coefficient that is used to generate the paths is $\sigma^2(x) = 2.08x^{2(1.34)}$. A Milstein scheme is used to generate paths 5505 trade dates long with a starting value of 0.07984 from a univariate diffusion with the given drift and diffusion coefficients. The first 1000 of these paths which have a minimum value greater than 0.02915 and a maximum value less than 0.24333 are retained. The CHLS specification with drift coefficient $\mu(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x$ and diffusion coefficient $\sigma^2(x) = \beta_1 x^{2\beta_2}$ is estimated from each of these paths using the generalized method of moments. The solid line is the true value of the drift or diffusion coefficient. The dot-dashed line is the pointwise mean of the 1000 estimated coefficients, and the dotted lines are the pointwise 25th and 75th percentiles of the coefficients.

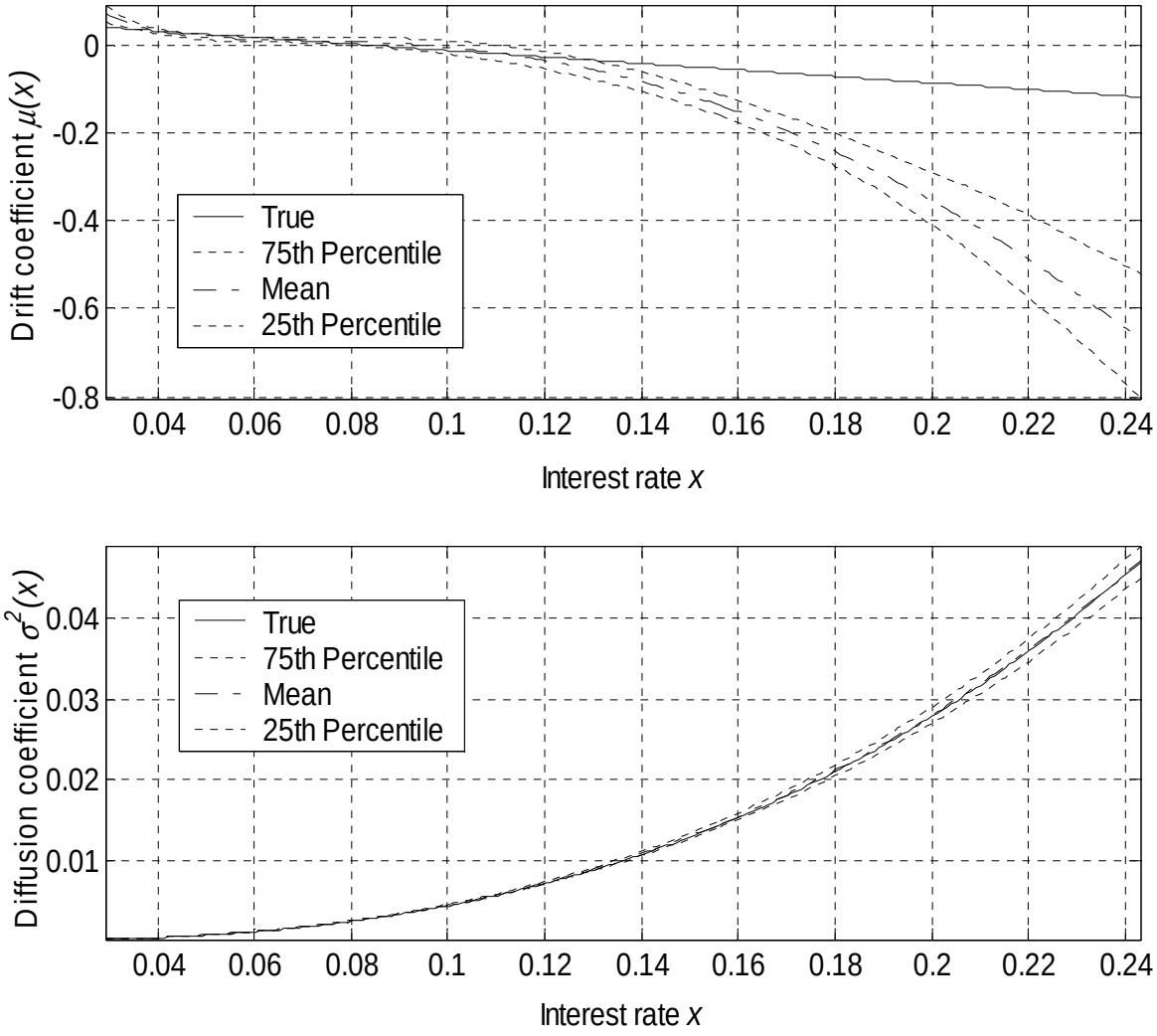

Figure 3: Drift and diffusion coefficients estimated from paths generated by linear drift coefficient $\mu(x) = 0.063 - 0.74x$ that are conditioned on minimum being within ten basis points of 0.02915 and maximum being within ten basis points of 0.24333. The diffusion coefficient that is used to generate the paths is $\sigma^2(x) = 2.08x^{2(1.34)}$. A Milstein scheme is used to generate paths 5505 trade dates long with a starting value of 0.07984 from a univariate diffusion with the given drift and diffusion coefficients. The first 1000 of these paths which have a minimum value within ten basis points of 0.02915 and a maximum value within ten basis points of 0.24333 are retained. The CHLS specification with drift coefficient $\mu(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x$ and diffusion coefficient $\sigma^2(x) = \beta_1 x^{2\beta_2}$ is estimated from each of these paths using the generalized method of moments. The solid line is the true value of the drift or diffusion coefficient. The dot-dashed line is the pointwise mean of the 1000 estimated coefficients, and the dotted lines are the pointwise 25th and 75th percentiles of the coefficients.

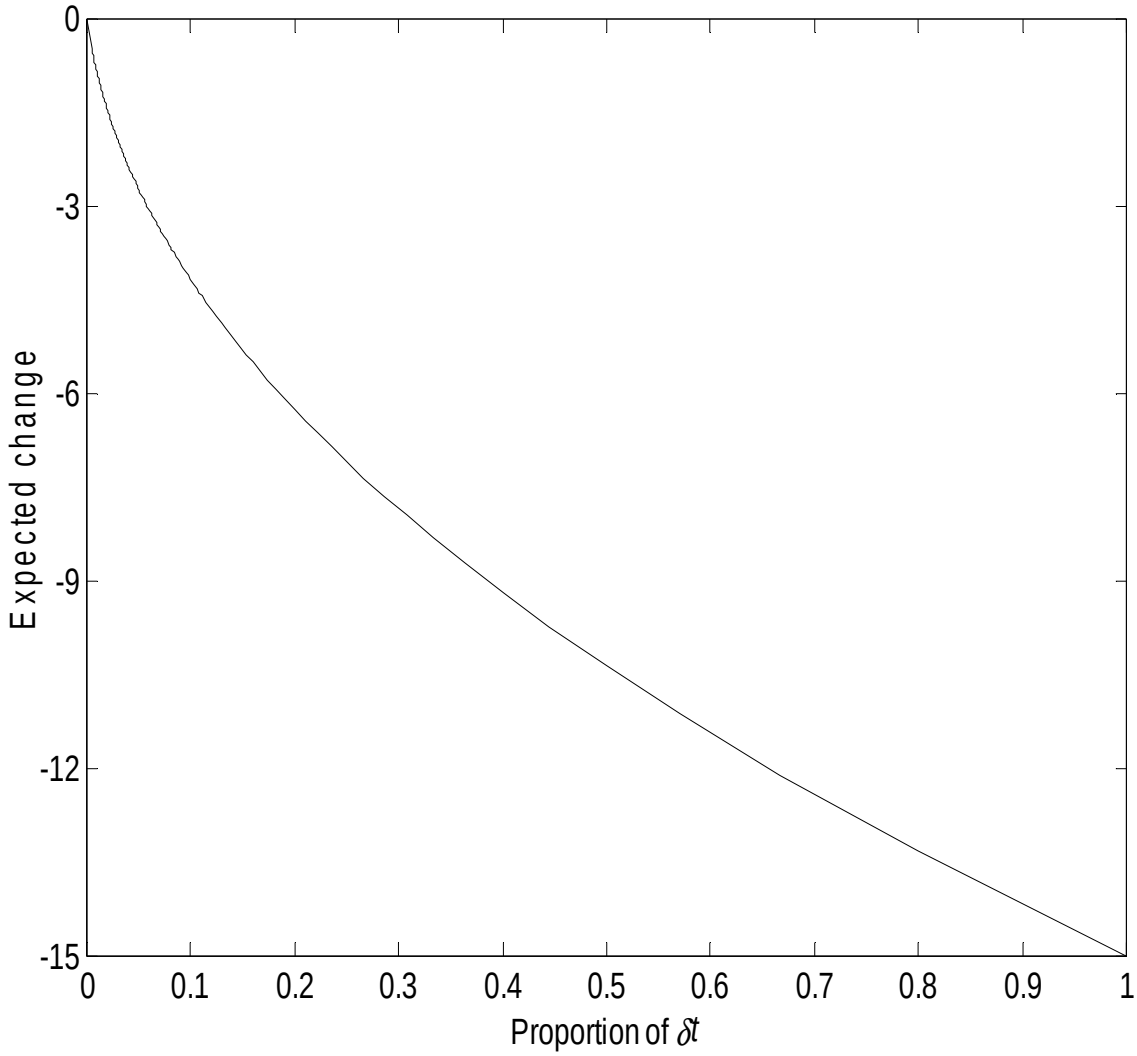

Figure 4: Expected change over the course of one trade date of a process currently at a value of 799 that unconditionally obeys geometric Brownian motion and is conditioned to remain in between 300 and 800 for the next four years. The drift coefficient of the unconditioned process is $\mu(x) = 0.05x$ and the diffusion coefficient is $\sigma^2(x) = 0.04x^2$. Each point on the plot is the expected change in the value of the process from the initial value of 799 from time zero to the fraction of one trade date indicated on the horizontal axis.

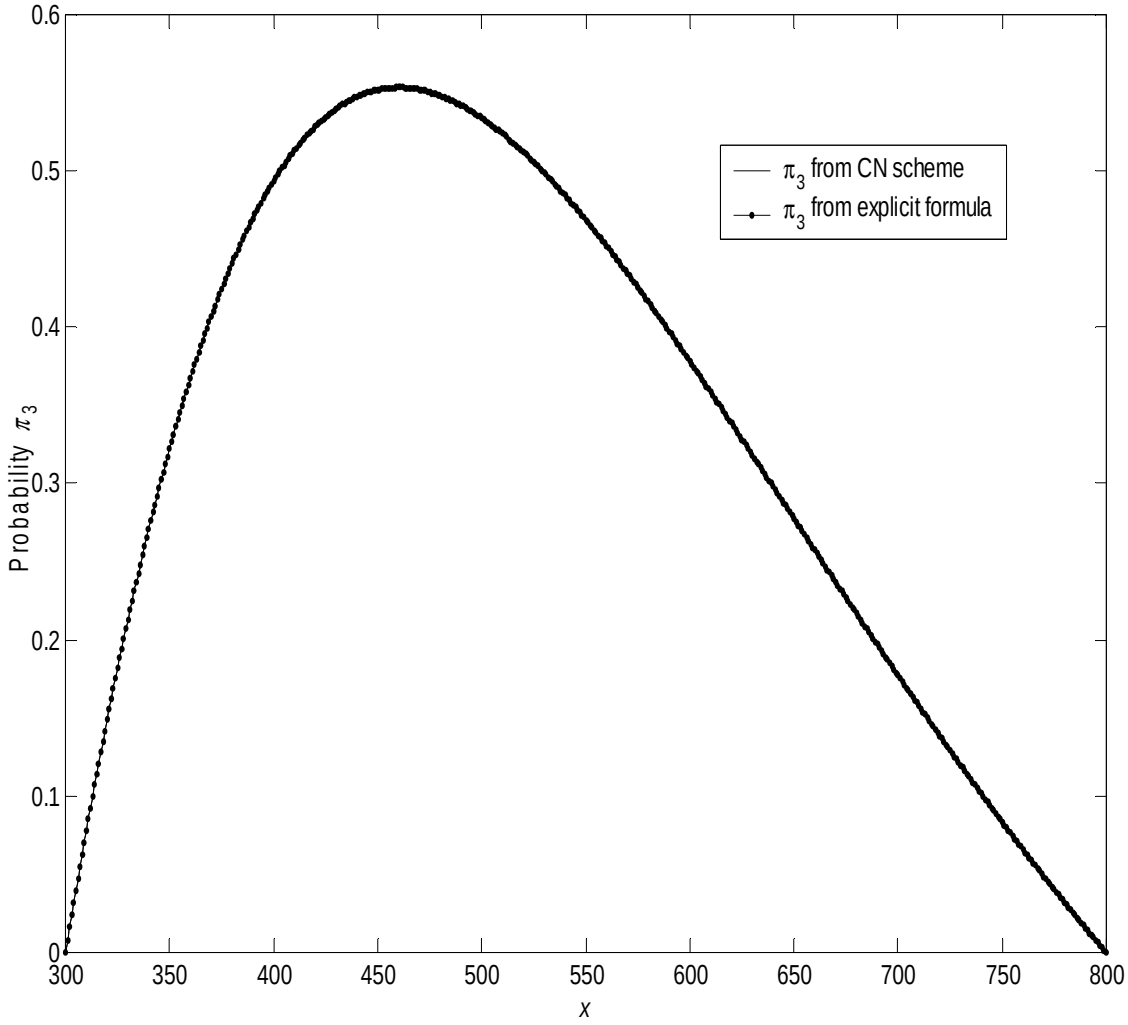

Figure 5: The probability that the process remains in between 300 and 800 for four more years when it follows geometric Brownian motion as a function of its current level x. The drift coefficient of the unconditioned process is $\mu(x) = 0.05x$ and the diffusion coefficient is $\sigma^2(x) = 0.04x^2$. The solid line shows the probabilities computed using the explicit formula and the line with dots shows the probabilities computed using the Crank-Nicholson scheme.

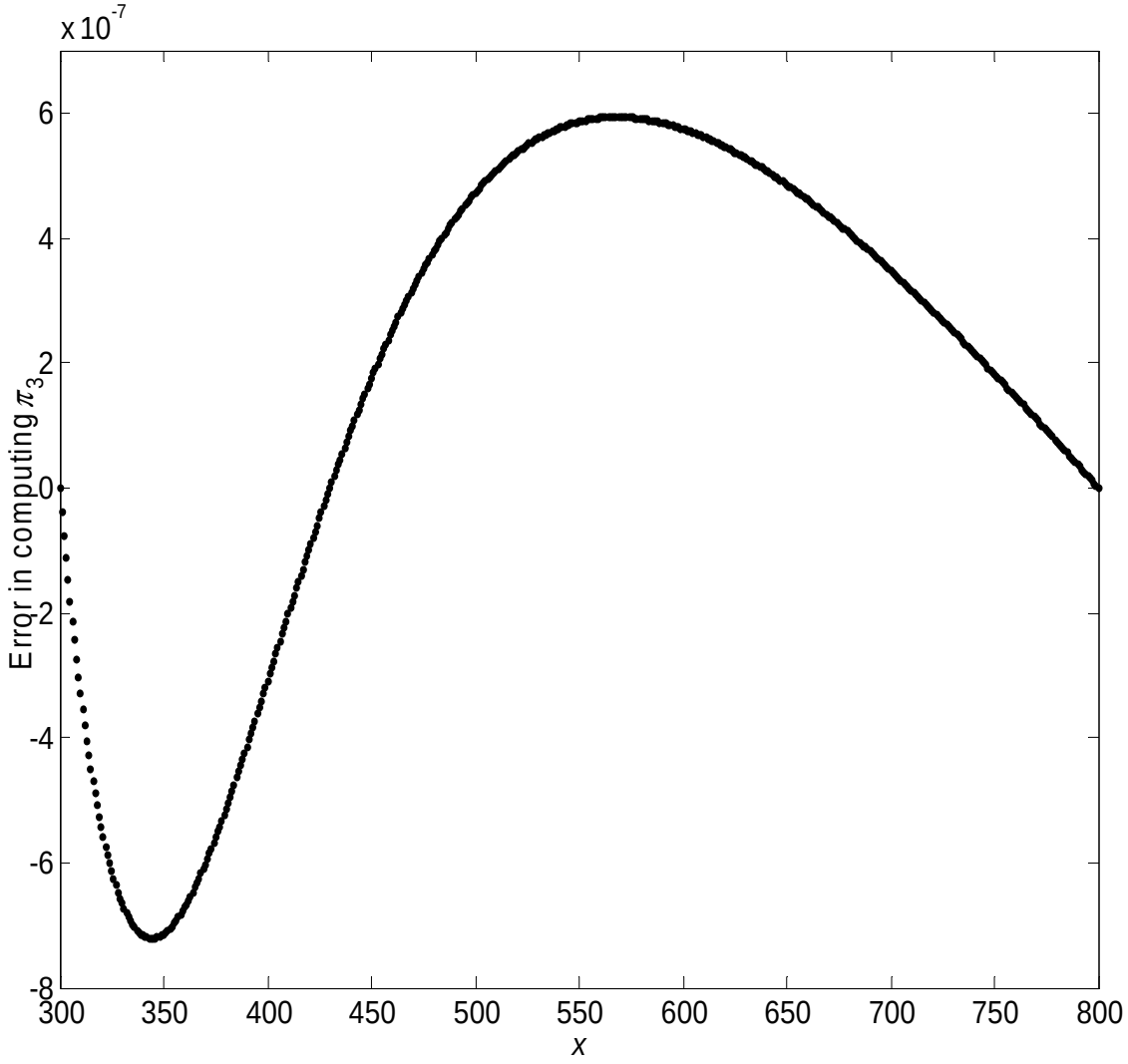

Figure 6: The difference between the numerically and analytically computed probability that the process remains in between 300 and 800 for four more years when it follows geometric Brownian motion as a function of its current level. The drift coefficient of the unconditioned process is $\mu(x) = 0.05x$ and the diffusion coefficient is $\sigma^2(x) = 0.04x^2$. The curve is the probability computed from a Crank-Nicholson scheme minus the probability from the explicit expression (49).

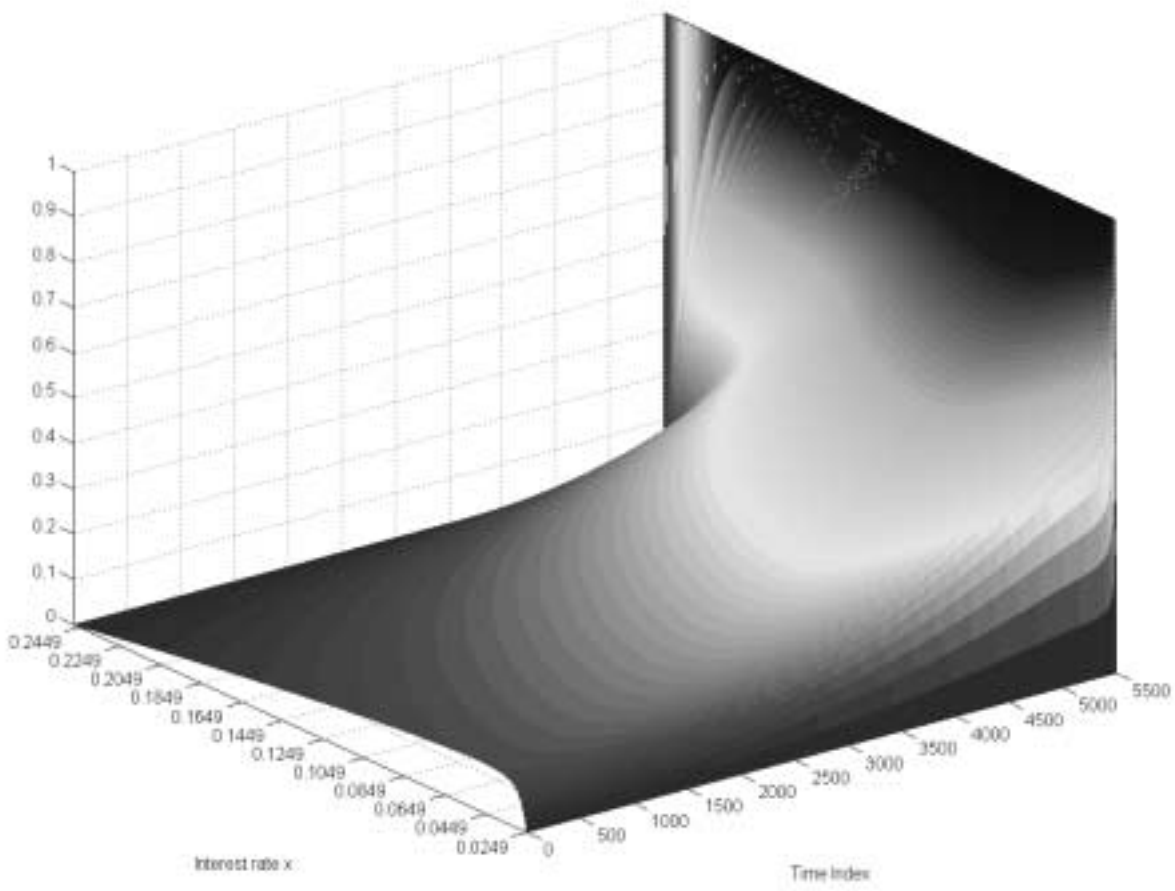

Figure 7: The probability $\pi_3 = \pi(x, t \,|\, A_3(t, T))$ that the interest rate $x(t)$ remains between $\bar{a} = 0.02487$ and $\bar{b} = 0.24761$ until time T (time index 5,505, February 25, 1995) as a function of the time index and the interest rate. The probabilities are computed using a Crank-Nicholson scheme, the drift coefficient of the unconditioned Eurodollar rate process is $\mu(x) = 0.03400 - 0.02834x$, and the diffusion coefficient is $\sigma^2(x) = 2.0511x^{2(1.3333)}$. The interval between trade dates is $\Delta t = 21.73424/5,504 = 0.0039488$ years.

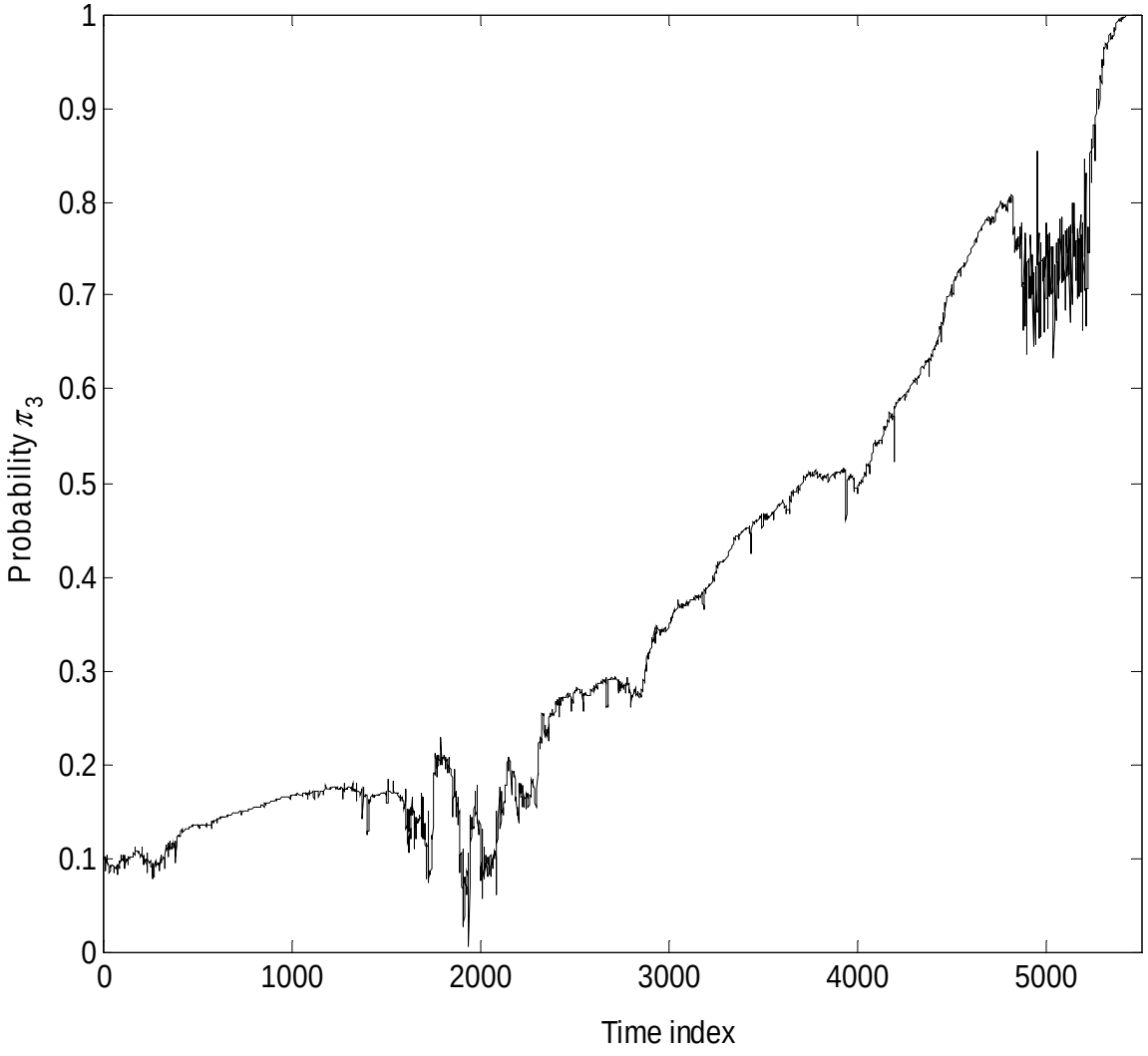

Figure 8: The probability $\pi(x(t), t \mid A_3(t, T))$ that the seven day Eurodollar rate process remains between $\bar{a} = 0.02487$ and $\bar{b} = 0.24761$ until time T (time index 5,505, February 25, 1995) along the actual interest rate path for each trade date from time index 0 (June 1, 1973) to time index 5,505 (time T, February 25, 1995). The probabilities are computed using a Crank-Nicholson scheme, the drift coefficient of the unconditioned Eurodollar rate process is $\mu(x) = 0.03400 - 0.02834x$, and the diffusion coefficient is $\sigma^2(x) = 2.0511x^{2(1.3333)}$. The interval between trade dates is $\Delta t = 21.73424/5,504 = 0.0039488$ years.

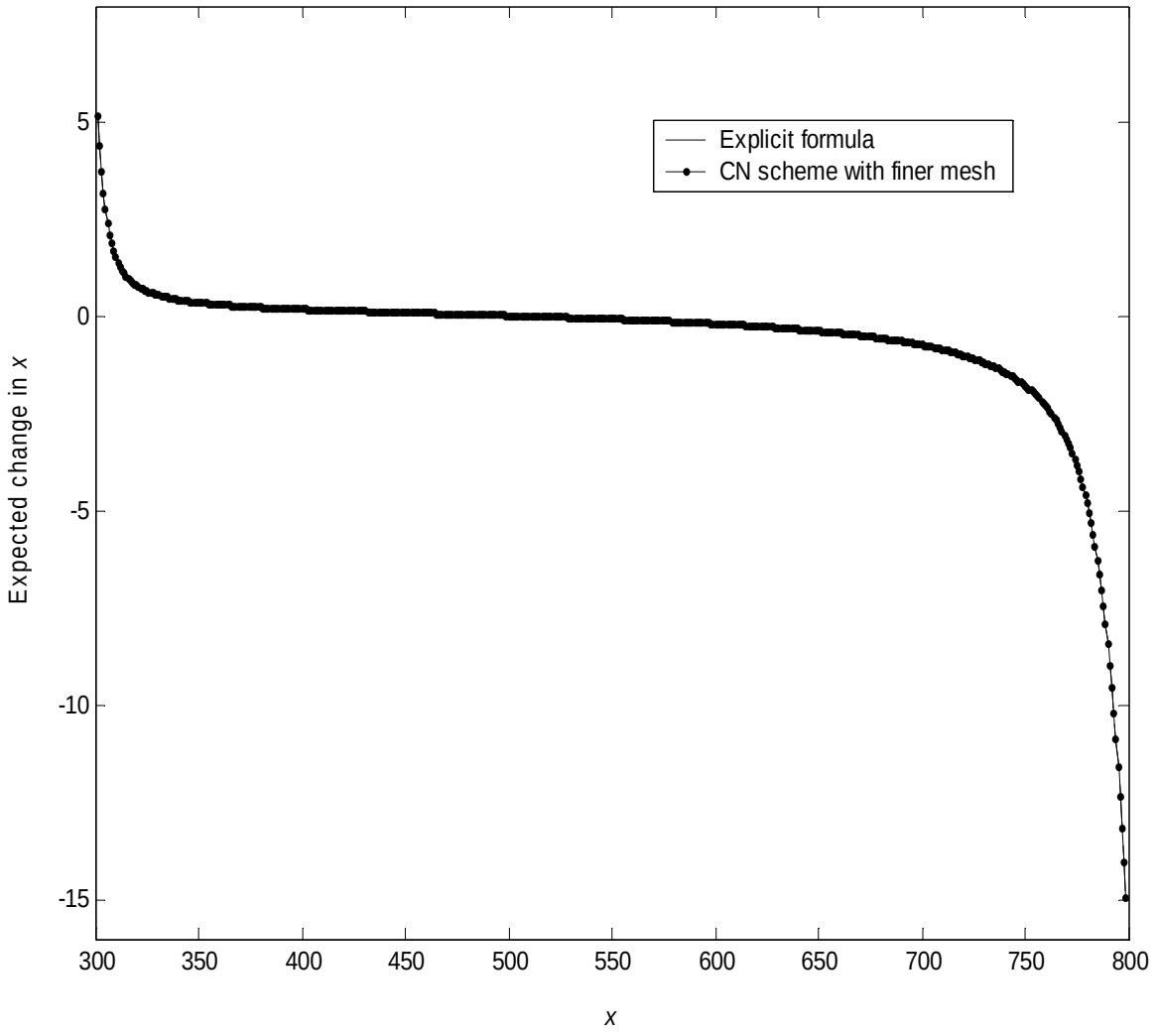

Figure 9: The one trade date expected change for a process that follows geometric Brownian motion conditional on it remaining in between 300 and 800 for four more years as a function of its current level x. The drift coefficient of the unconditioned process is $\mu(x) = 0.05x$ and the diffusion coefficient is $\sigma^2(x) = 0.04x^2$. The dashed line computes the expected changes from an analytical approximation and the line with dots computes the expected change numerically using the Crank-Nicholson scheme.

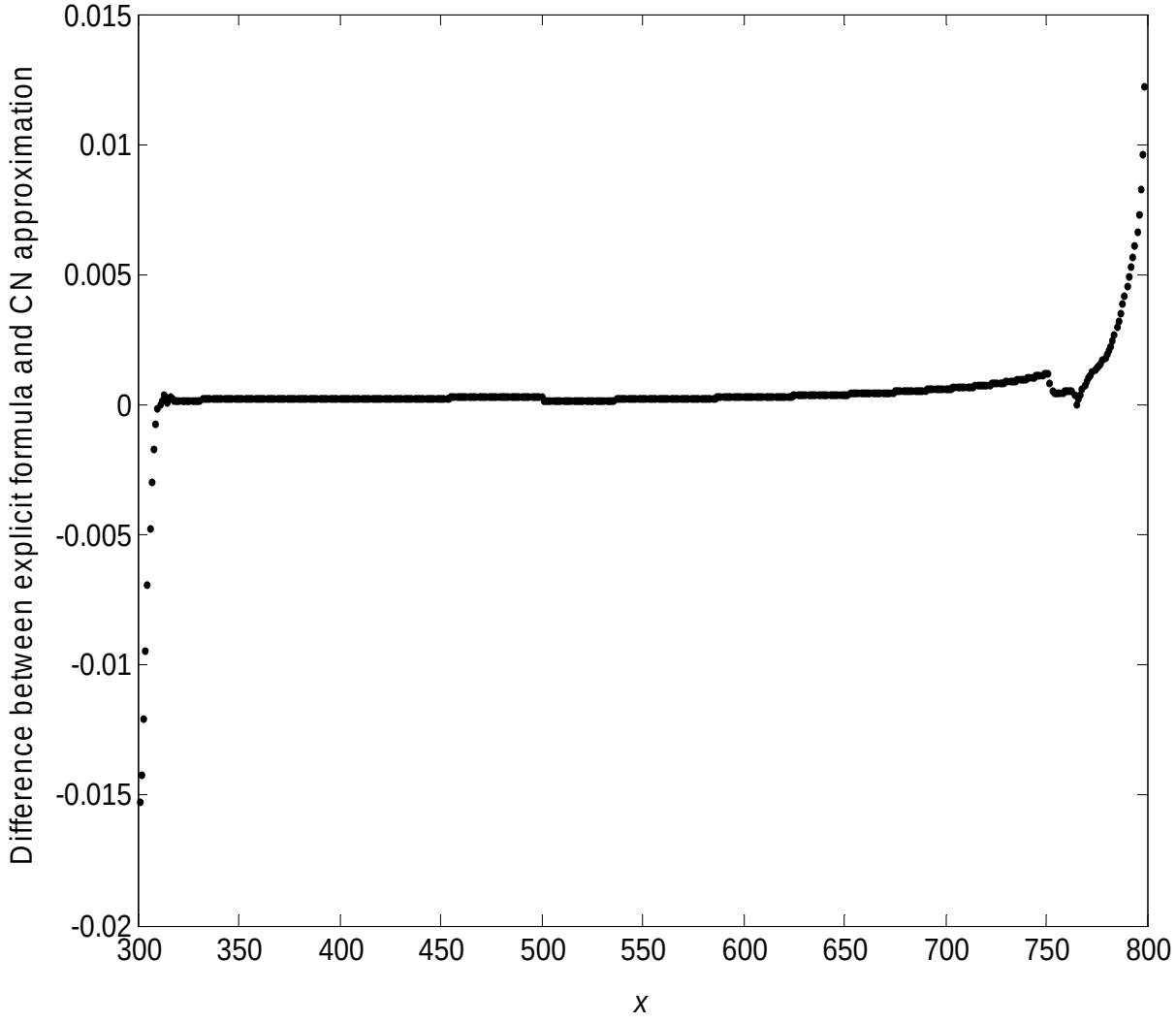

Figure 10: The difference between the one trade date expected change computed using the Crank-Nicholson scheme and the explicit formula for a process that follows geometric Brownian motion conditional on it remaining between 300 and 800 for four more years as a function of its current level x. The drift coefficient of the unconditioned process is $\mu(x) = 0.05x$ and the diffusion coefficient is $\sigma^2(x) = 0.04x^2$.

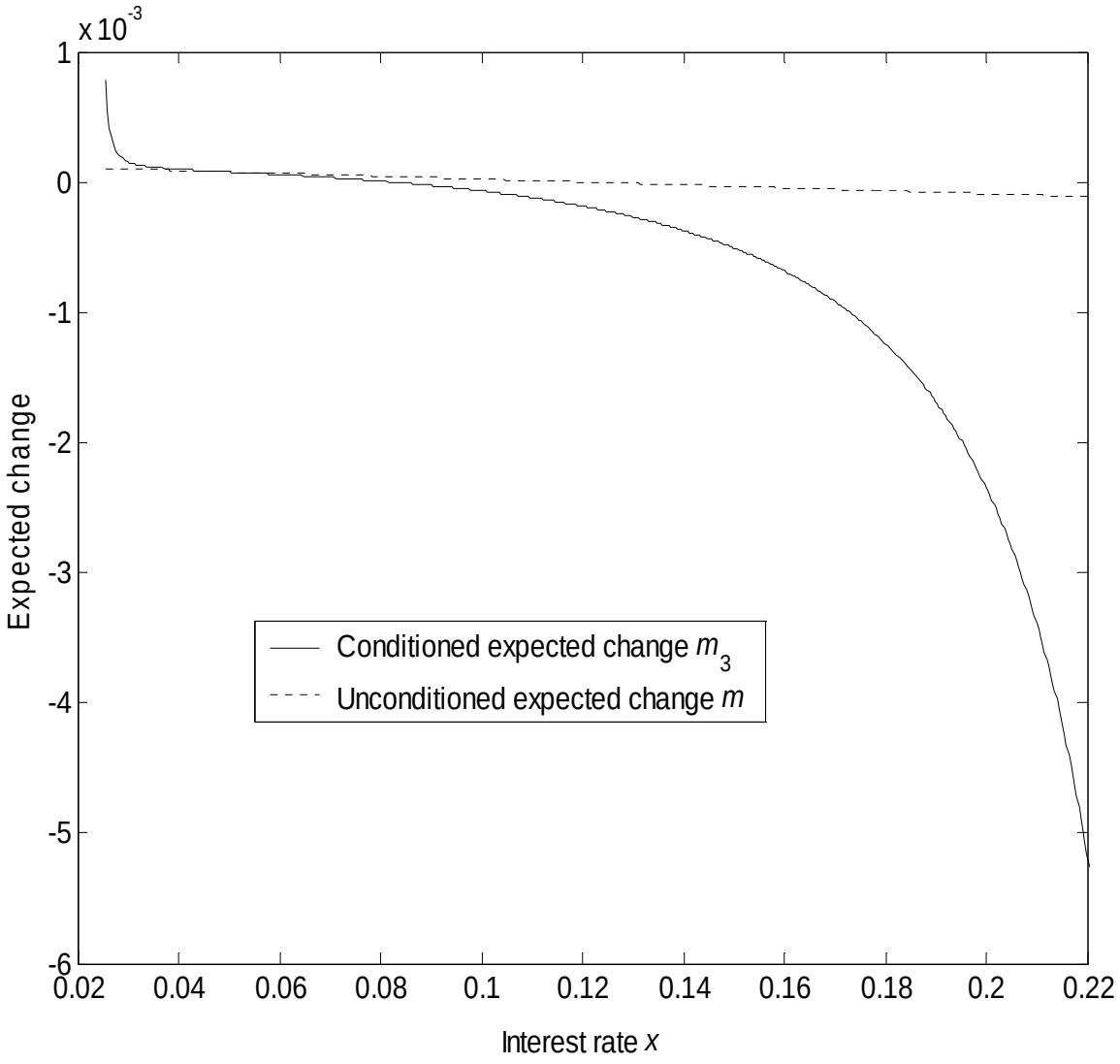

Figure 11: The one trade date expected change for a process with and without conditioning on the process remaining in between $\bar{a} = 0.02487$ and $\bar{b} = 0.24761$ for the next twelve years as a function of the current interest rate level. The drift coefficient of the unconditioned Eurodollar rate 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.

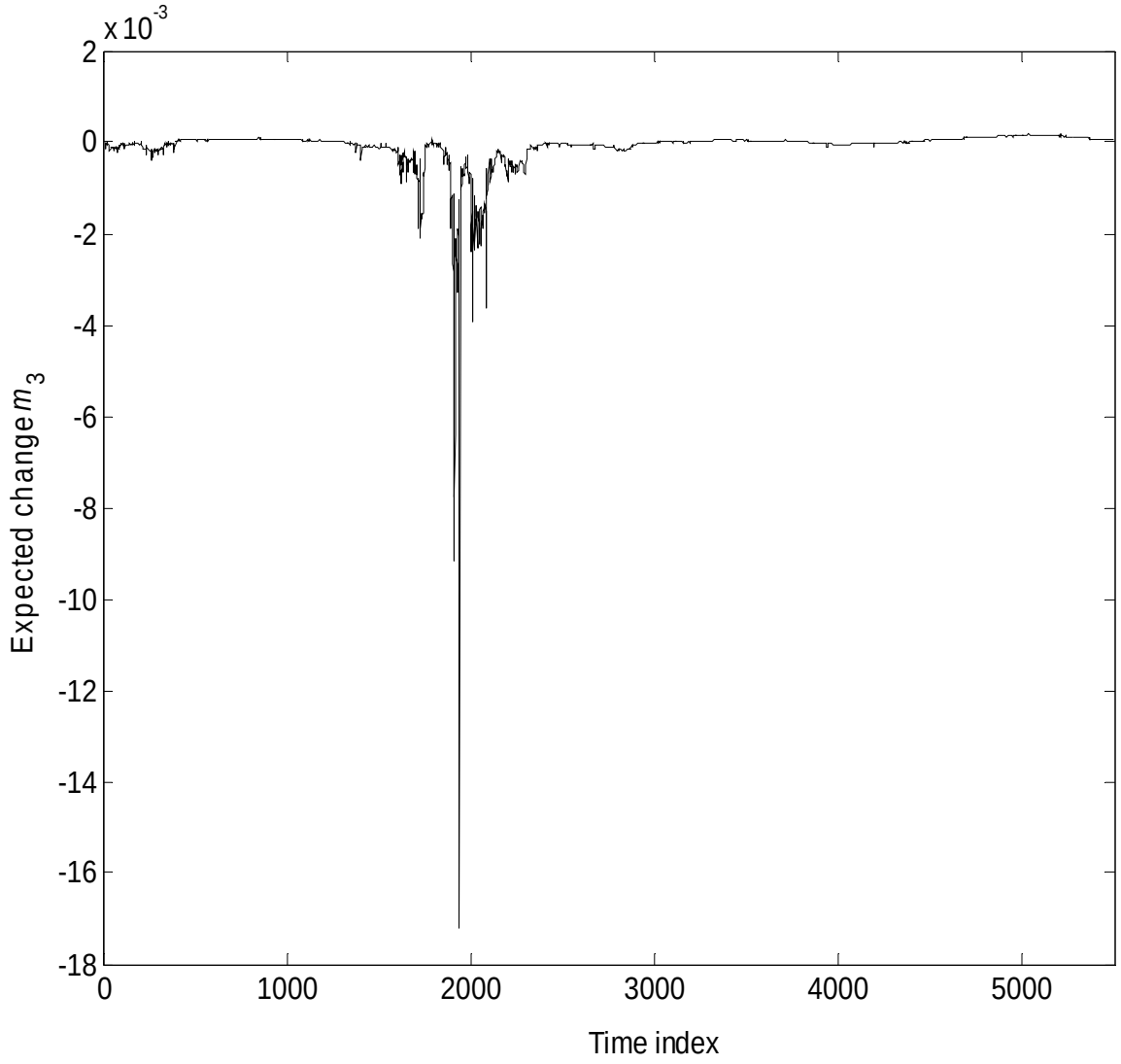

Figure 12: The expected one trade date change in the seven day Eurodollar rate on each trade date from June 1, 1973 to February 25, 1995 conditional on it remaining in between $\bar{a} = 0.02487$ and $\bar{b} = 0.24761$. The drift coefficient of the unconditioned process is assumed to be $\mu(x) = 0.03400 - 0.02834x$ 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 25, 1995. The interval between trade dates is $\Delta t = 21.73424/5,504 = 0.0039488$ years.

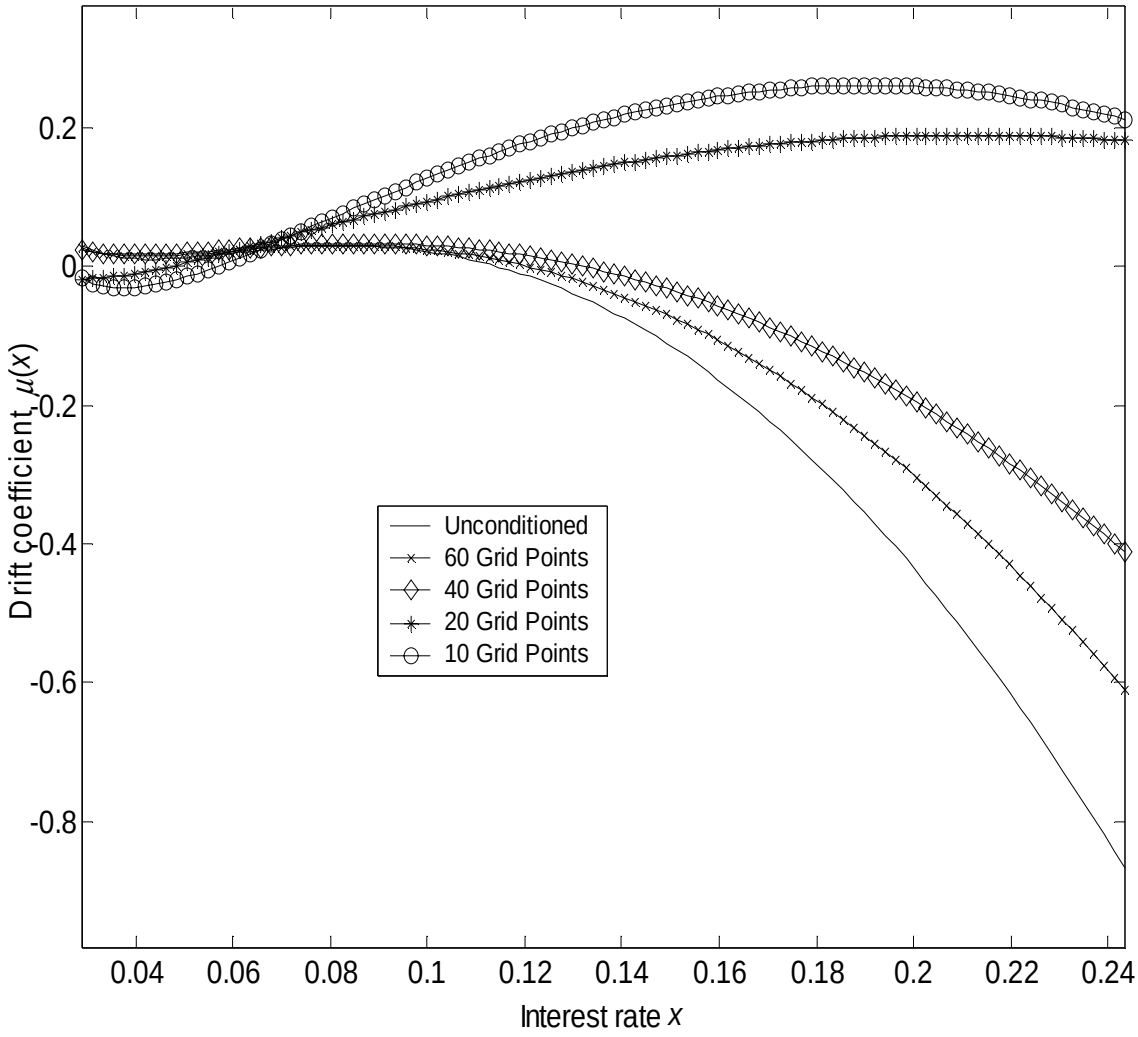

Figure 13: Drift coefficients estimated from seven day Eurodollar rates, June 1, 1973 to February 25, 1995, conditioning on the event $A_3(t,T)$. The drift and diffusion coefficients $\mu(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x$ and $\sigma^2(x) = \beta_1 x^{2\beta_2}$ were estimated from the Eurodollar rate data using the generalized method of moments, and the moment conditions were constructed using one trade date conditional 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 25, 1995. The line with crosses $\times$ is the drift estimate obtained when on each trade the one trade date expected change is computed conditional on the process remaining between $\bar{a} = a -$ displacement $\times \Delta x$ and $\bar{b} = b +$ displacement $\times \Delta x$ for the remainder of the time until February 25, 1995, where and a and b 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 diamonds is the drift estimate when the conditioning is done using a displacement of $40\Delta x$ from the observed minimum and maximum. The line with asterisks is for a displacement of $20\Delta x$, and the line with circles is for a displacement of $10\Delta x$.

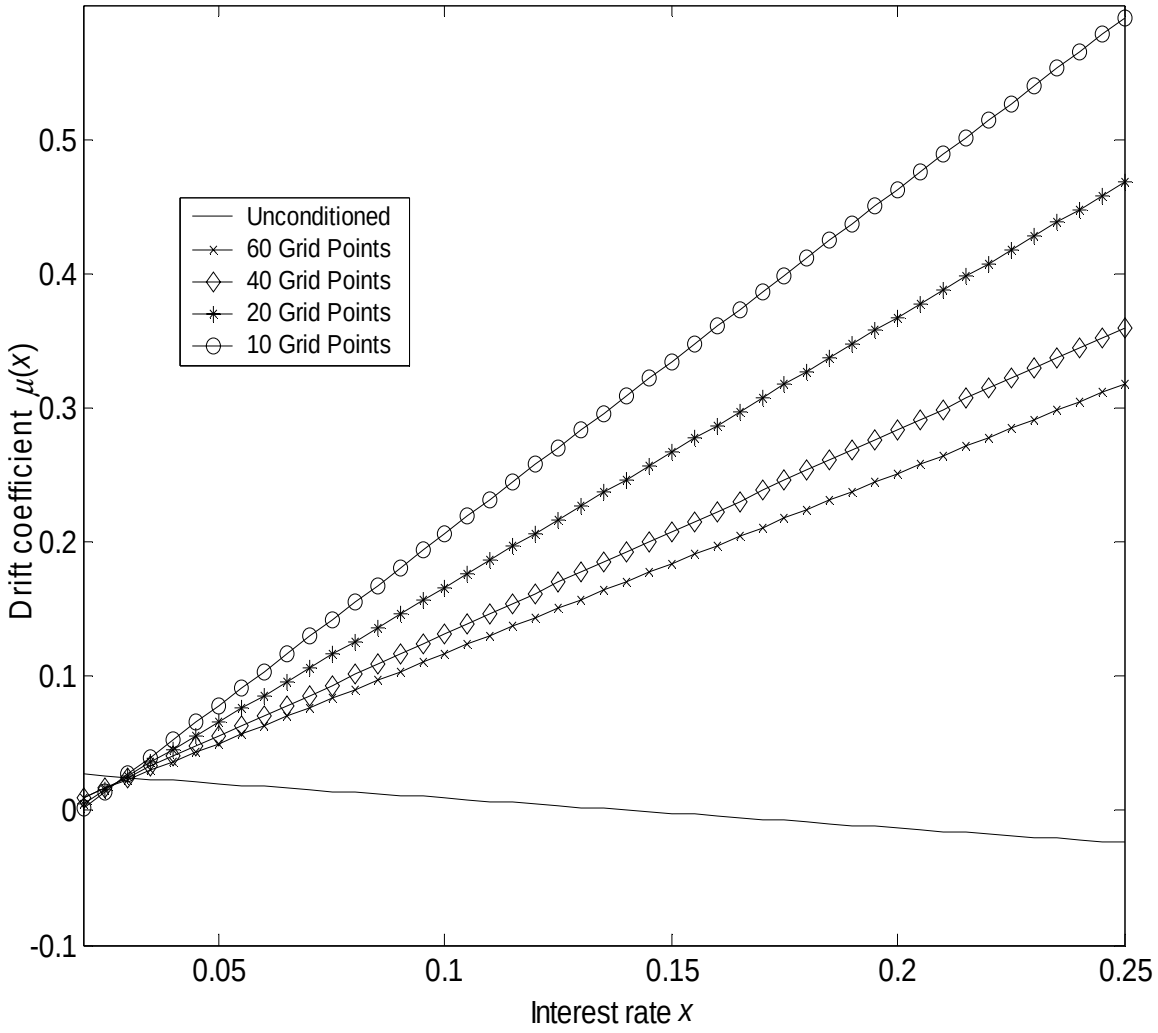

Figure 14: Linear drift coefficients estimated from seven day Eurodollar rates, June 1, 1973 to February 25, 1995, conditioning on the event $A_3(t,T)$. The drift and diffusion coefficients $\mu(x) = \alpha_0 + \alpha_1 x$ and $\sigma^2(x) = \beta_1 x^{2\beta_2}$ were estimated from the Eurodollar rate data using the generalized method of moments, and the moment conditions were constructed using one trade date conditional 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 25, 1995. The line with crosses $\times$ is the drift estimate obtained when on each trade the one trade date expected change is computed conditional on the process remaining between $\bar{a} = a -$ displacement $\times \Delta x$ and $\bar{b} = b +$ displacement $\times \Delta x$ for the remainder of the time until February 25, 1995, where and a and b 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 diamonds is the drift estimate when the conditioning is done using a displacement of $40\Delta x$ from the observed minimum and maximum. The line with asterisks is for a displacement of $20\Delta x$, and the line with circles is for a displacement of $10\Delta x$.

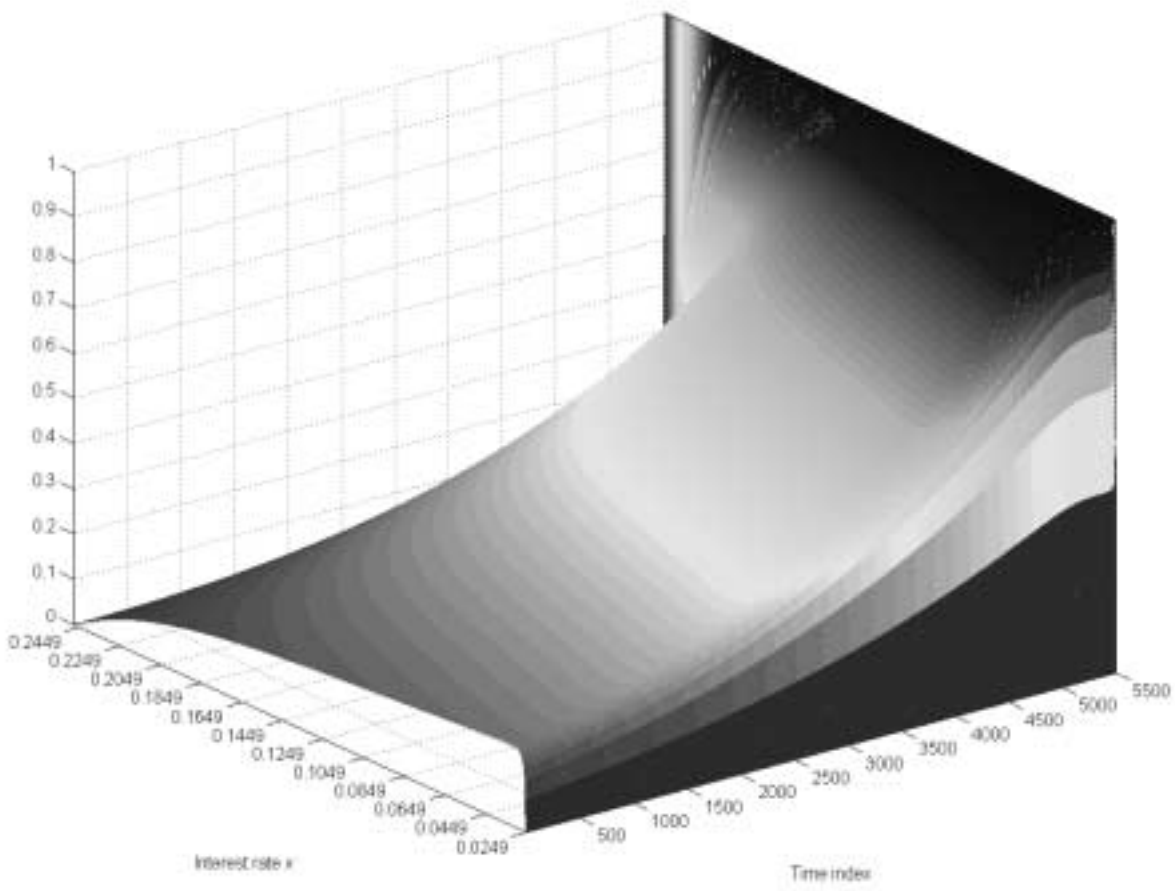

Figure 15: The probability $\pi_3 = \pi(x, t|\, A_3(t, T))$ that the process remains in between $\bar{a} = 0.02487$ and $\bar{b} = 0.24761$ from the current time step until the end of 5504 time steps as a function of the number of time steps that have elapsed to the current time and the current level of the process. The drift coefficient of the unconditioned process is $\mu(x) = -0.2684 + 5.0549x - 26.1224x^2 + 0.005897/x$ and the diffusion coefficient is $\sigma^2(x) = 2.0145x^{2(1.3299)}$. Each time step corresponds to $\Delta t = 21.73424/5,504 = 0.0039488$ or one trade date.

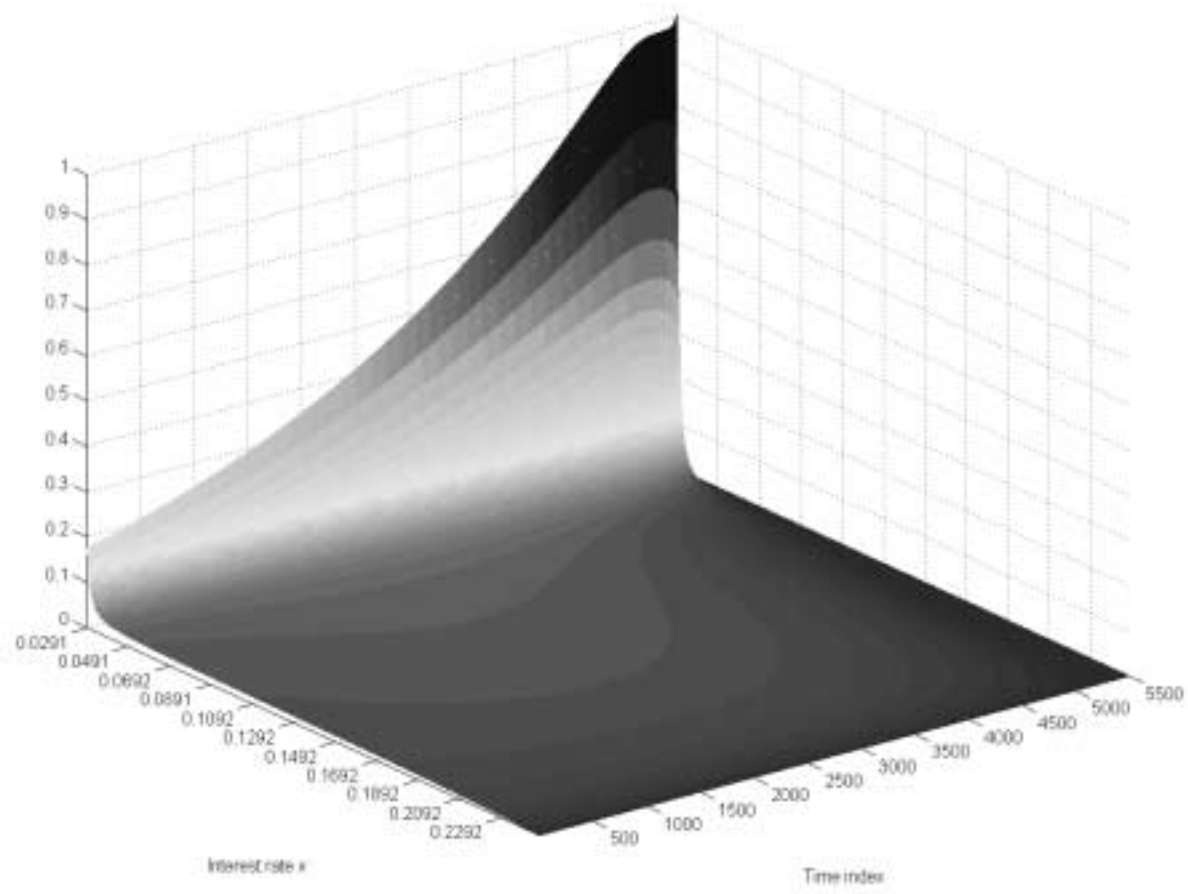

Figure 16: The probability $\pi_2 = \pi(x, t|\, A_2(t, T))$ that the process reaches $a = 0.02915$ but remains in between $\bar{a} = 0.02487$ and $\bar{b} = 0.24761$ from the current time step until the end of 5504 time steps as a function of the number of time steps that have elapsed to the current time and the current level of the process. The drift coefficient of the unconditioned process is $\mu(x) = -0.2684 + 5.0549x - 26.1224x^2 + 0.005897/x$ and the diffusion coefficient is $\sigma^2(x) = 2.0145x^{2(1.3299)}$. Each time step corresponds to $\Delta t = 21.73424/5,504 = 0.0039488$ or one trade date.

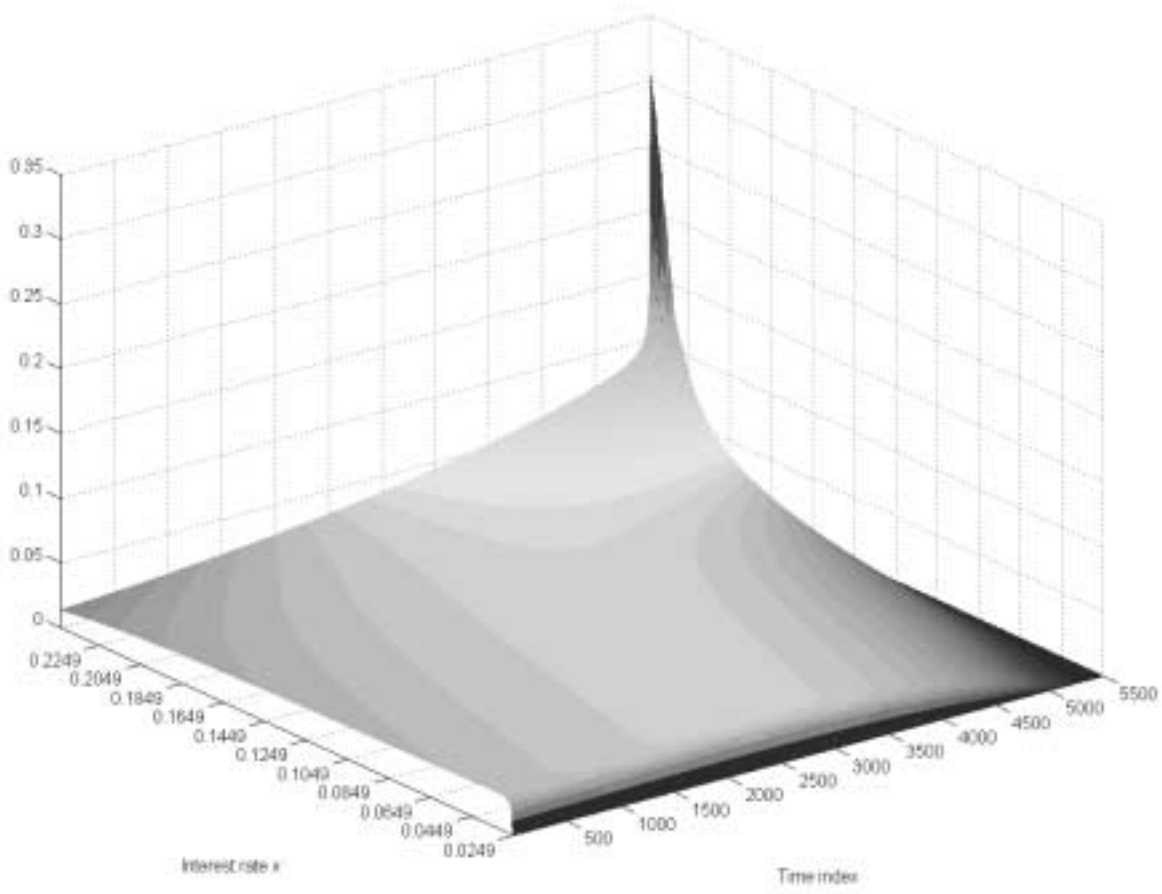

Figure 17: The probability $\pi_1 = \pi(x, t|\, A_1(t, T))$that the process reaches 0.24333 but remains in between 0.02487 and 0.24761 from the current time step until the end of 5504 time steps as a function of the number of time steps that have elapsed to the current time and the current level of the process. The drift coefficient of the unconditioned process is $\mu(x) = -0.2684 + 5.0549x - 26.1224x^2 + 0.005897/x$ and the diffusion coefficient is $\sigma^2(x) = 2.0145x^{2(1.3299)}$. Each time step corresponds to $\Delta t = 21.73424/5,504 = 0.0039488$ or one trade date.

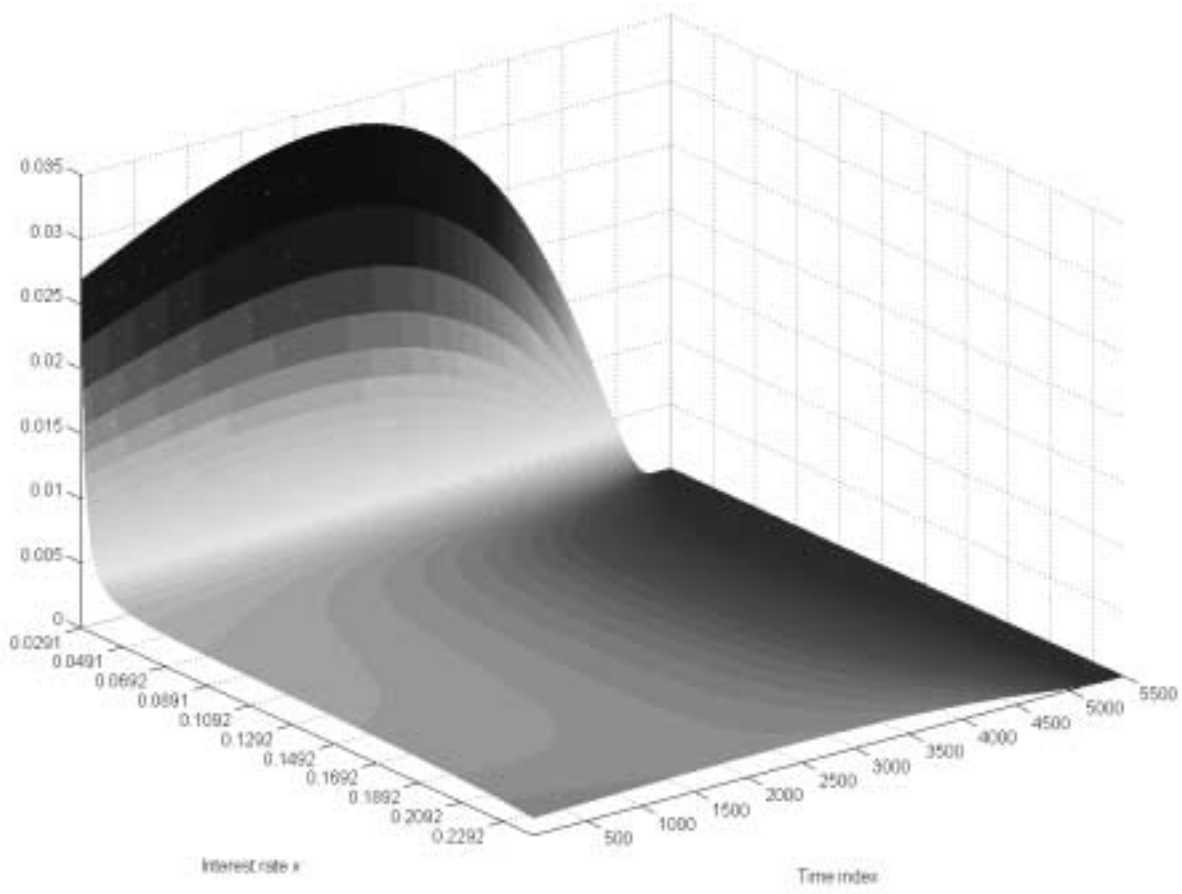

Figure 18: The probability $\pi_0 = \pi(x, t|\, A_0(t, T))$ that the process reaches 0.02915 and 0.24333 but remains in between 0.02487 and 0.24761 from the current time step until the end of 5504 time steps as a function of the number of time steps that have elapsed to the current time and the current level of the process. The drift coefficient of the unconditioned process is $\mu(x) = -0.2684 + 5.0549x - 26.1224x^2 + 0.005897/x$ and the diffusion coefficient is $\sigma^2(x) = 2.0145x^{2(1.3299)}$. Each time step corresponds to $\Delta t = 21.73424/5,504 = 0.0039488$ or one trade date.

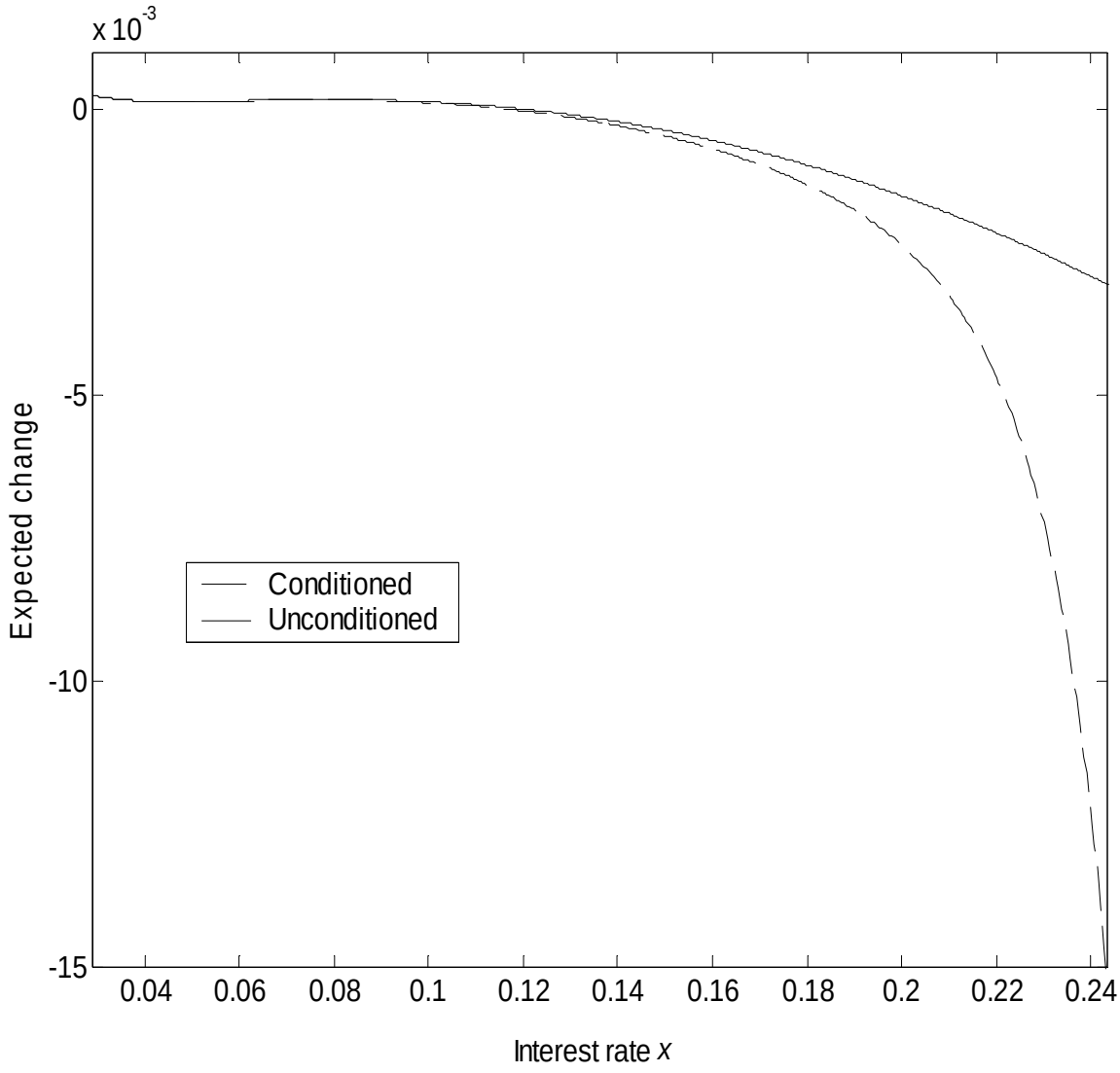

Figure 19: The one trade date expected change $m_3 = m(x, t, t + \delta t | \, A_3(t, T))$ for a process with and without conditioning on the event A_3. The event A_3 is that the process remains between $\bar{a} = 0.02487$ and $\bar{b} = 0.24761$ for the next 255 trade dates as a function of the current interest rate level. The drift coefficient of the unconditioned process is $\mu(x) = -0.2684 + 5.0549x - 26.1224x^2 + 0.005897/x$ and the diffusion coefficient is $\sigma^2(x) = 2.0145x^{2(1.3299)}$. The dashed line is the conditioned one trade date expected change computed using a Crank-Nicholson scheme, and the solid line is the unconditioned one trade date expected change.

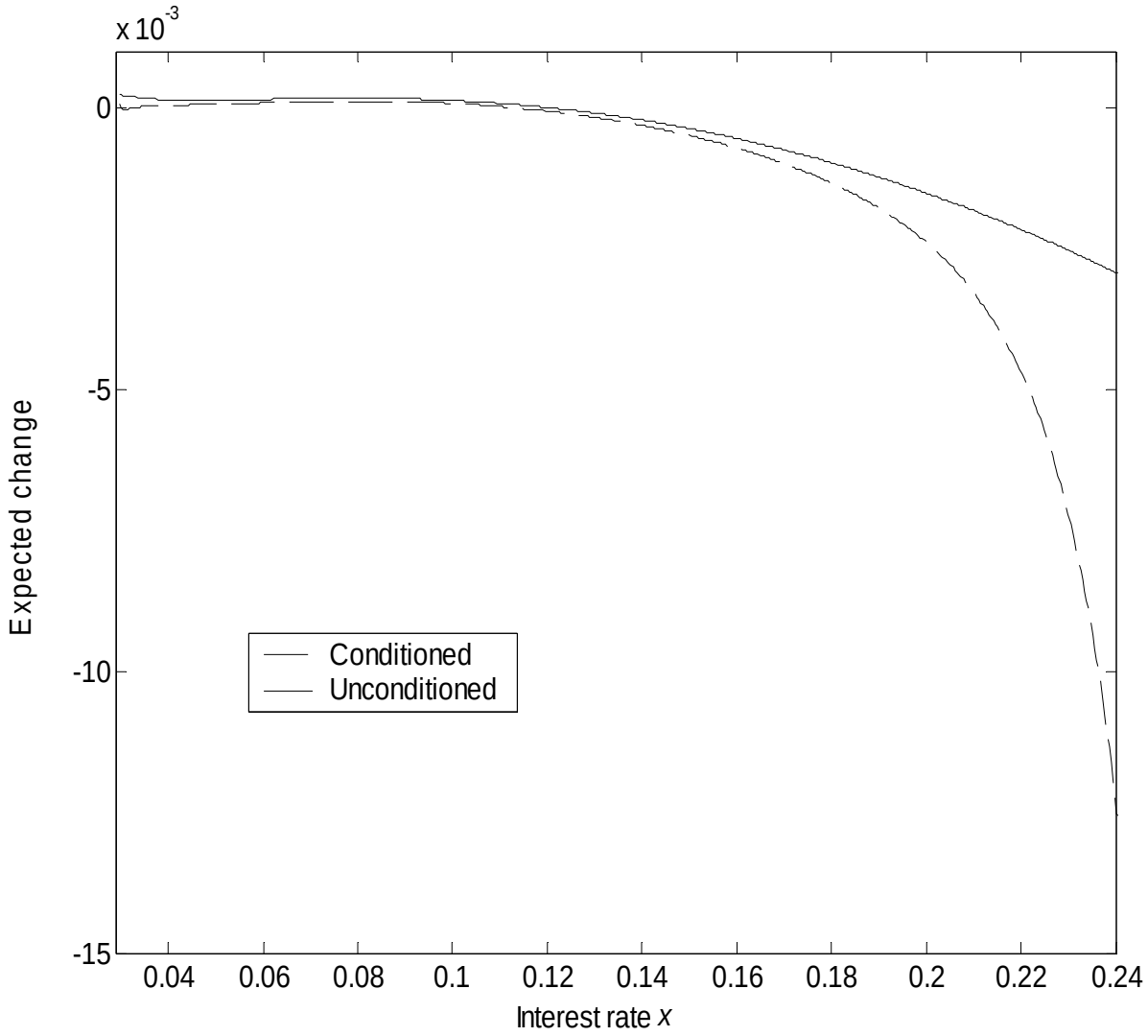

Figure 20: The one trade date expected change $m_2 = m(x, t, t + \delta t | \, A_2(t, T))$ for a process with and without conditioning on the event A_2. The event A_2 is that the process reaches $a = 0.02915$ and remaining in between $\bar{a} = 0.02487$ and $\bar{b} = 0.24761$ for the next 2505 trade dates as a function of the current interest rate level. The drift coefficient of the unconditioned process is $\mu(x) = -0.2684 + 5.0549x - 26.1224x^2 + 0.005897/x$ and the diffusion coefficient is $\sigma^2(x) = 2.0145x^{2(1.3299)}$. The dashed line is the conditioned one trade date expected change computed using a Crank-Nicholson scheme, and the solid line is the unconditioned one trade date expected change.

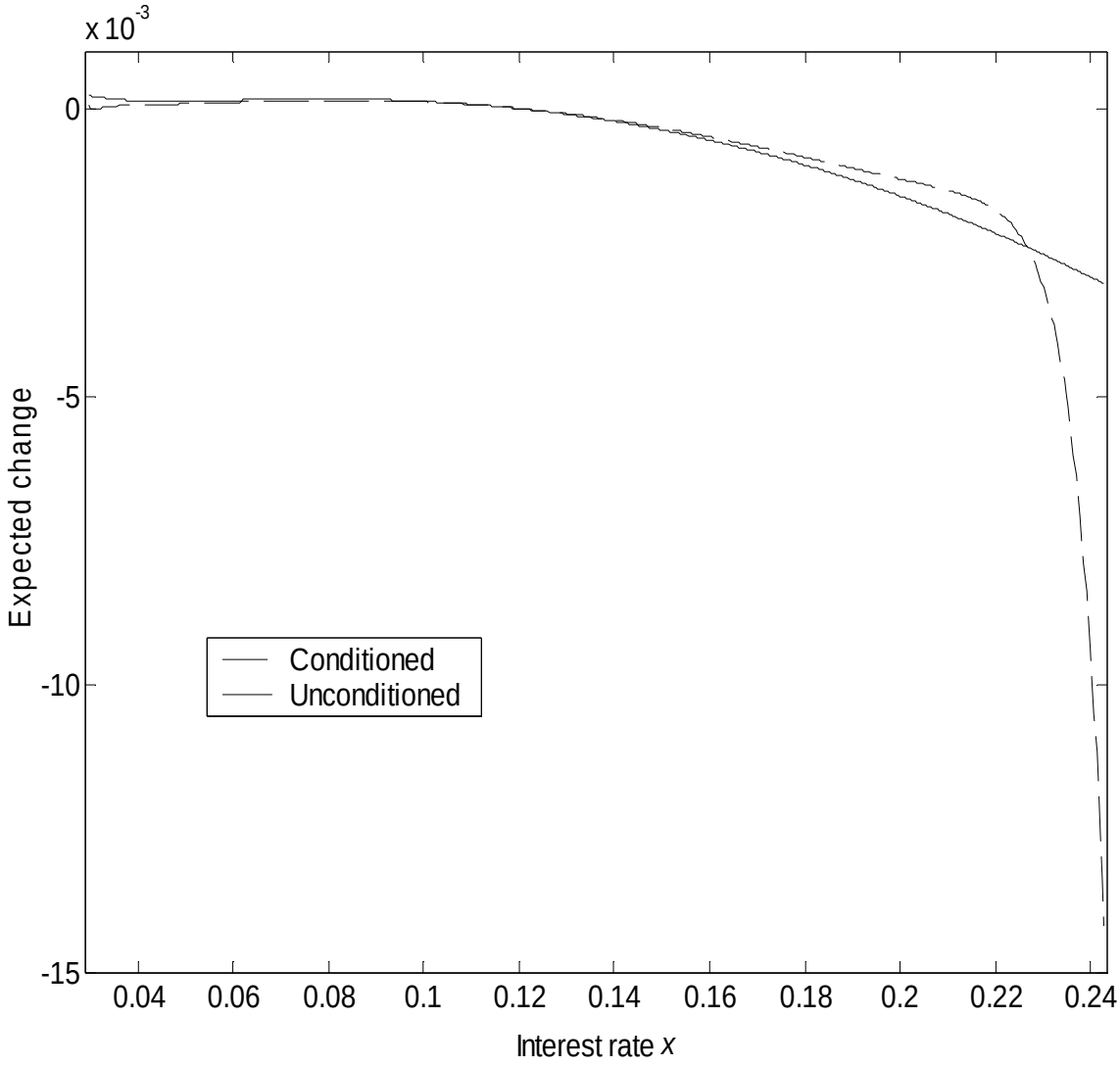

Figure 21: The one trade date expected change $m_0 = m(x, t, t + \delta t |\, A_0(t, T))$ for a process with and without conditioning on the event A_0. The event A_0 is that the process reaches $a = 0.02915$ and $b = 0.24333$ and also remains between $\bar{a} = 0.02487$ and $\bar{b} = 0.24761$ for the next 4505 trade dates as a function of the current interest rate level. The drift coefficient of the unconditioned process is $\mu(x) = -0.2684 + 5.0549x - 26.1224x^2 + 0.005897/x$ and the diffusion coefficient is $\sigma^2(x) = 2.0145x^{2(1.3299)}$. The dashed line is the conditioned one trade date expected change computed using the Crank-Nicholson scheme, and the solid line is the unconditioned one trade date expected change.

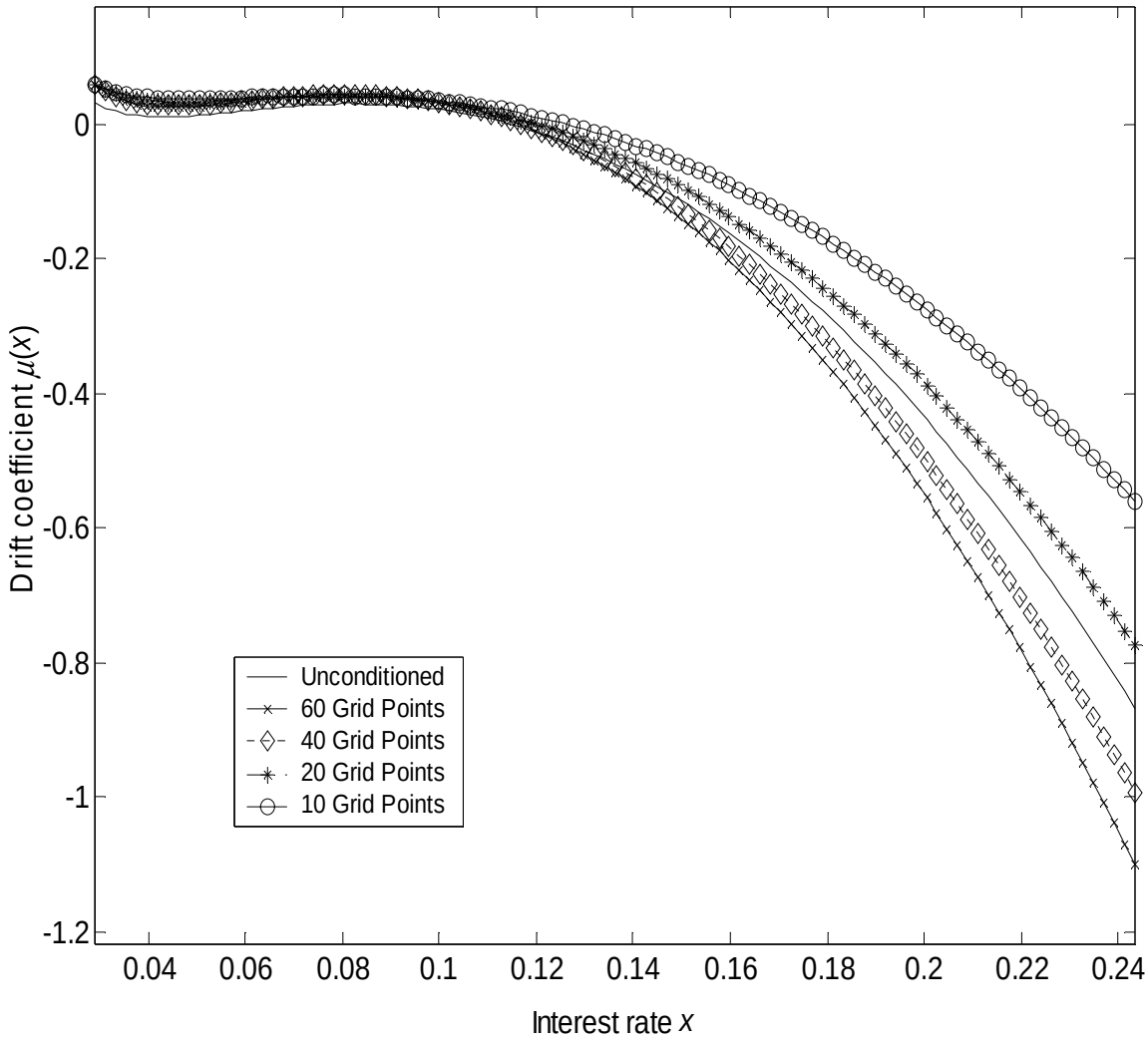

Figure 22: Drift coefficients estimated from seven day Eurodollar rates, June 1, 1973 to February 25, 1995, conditioning on the occurrence of $A_0(t,T)$. The event $A_0(t,T)$ occurs if the process reaches $a = 0.02915$ and $b = 0.24333$ in the time interval from June 1, 1973 to February 25, 1995 but does not reach $a -$ displacement $\times\, \Delta x$ or $b +$ displacement $\times\, \Delta x$ over the time interval, where $\Delta x = 4.2836$ basis points. The drift and diffusion coefficients $\mu(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x$ and $\sigma^2(x) = \beta_1 x^{2\beta_2}$ were estimated using the generalized method of moments, and the moment conditions were constructed using one trade date conditional expected changes. The solid line is the estimate of the drift coefficient when on each trade date there is no conditioning. The line with crosses $\times$ is the drift estimate when the conditioning is done using a displacement of $60\Delta x$. The line with diamonds is for a displacement of $40\Delta x$, the line with asterisks is for a displacement of $20\Delta x$, and the line with circles is for a displacement of $10\Delta x$.

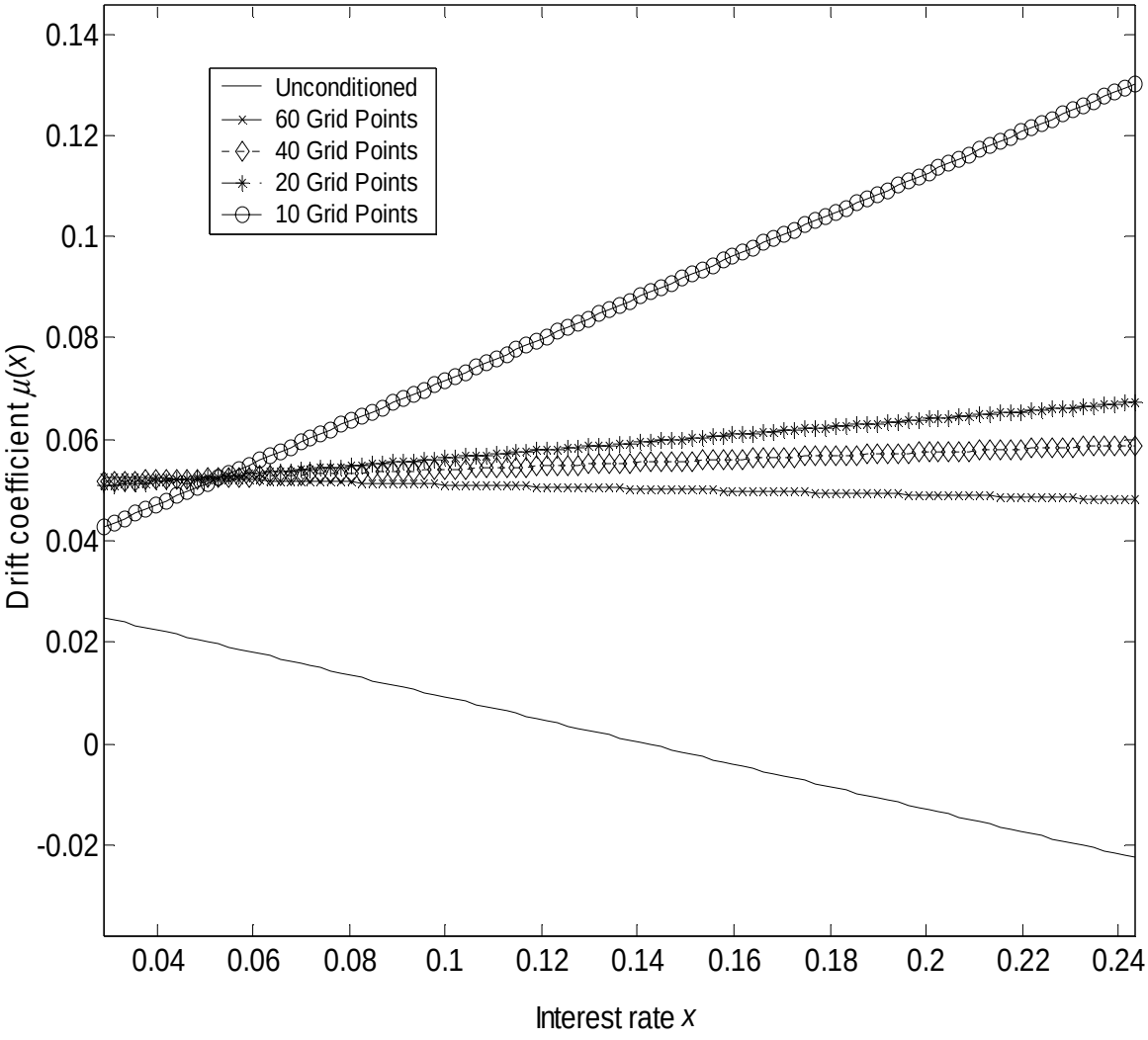

Figure 23: Linear drift coefficients estimated from seven day Eurodollar rates, June 1, 1973 to February 25, 1995, conditioning on each trade date on the occurrence of $A_0(t,T)$. The event $A_0(t,T)$ occurs if the process reaches $a = 0.02915$ and $b = 0.24333$ in the time interval from June 1, 1973 to February 25, 1995 but does not reach $a-\text{displacement}\times\Delta x$ or $b+\text{displacement}\times\Delta x$ over the time interval, where $\Delta x = 4.2836$ basis points. The drift and diffusion coefficients $\mu(x) = \alpha_0 + \alpha_1 x$ and $\sigma^2(x) = \beta_1 x^{2\beta_2}$ were estimated using the generalized method of moments, and the moment conditions were constructed using one trade date conditional expected changes. The solid line is the estimate of the drift coefficient when on each trade date there is no conditioning. The line with crosses $\times$ is the drift estimate when the conditioning is done using a displacement of $60\Delta x$. The line with diamonds is for a displacement of $40\Delta x$, the line with asterisks is for a displacement of $20\Delta x$, and the line with circles is for a displacement of $10\Delta x$.

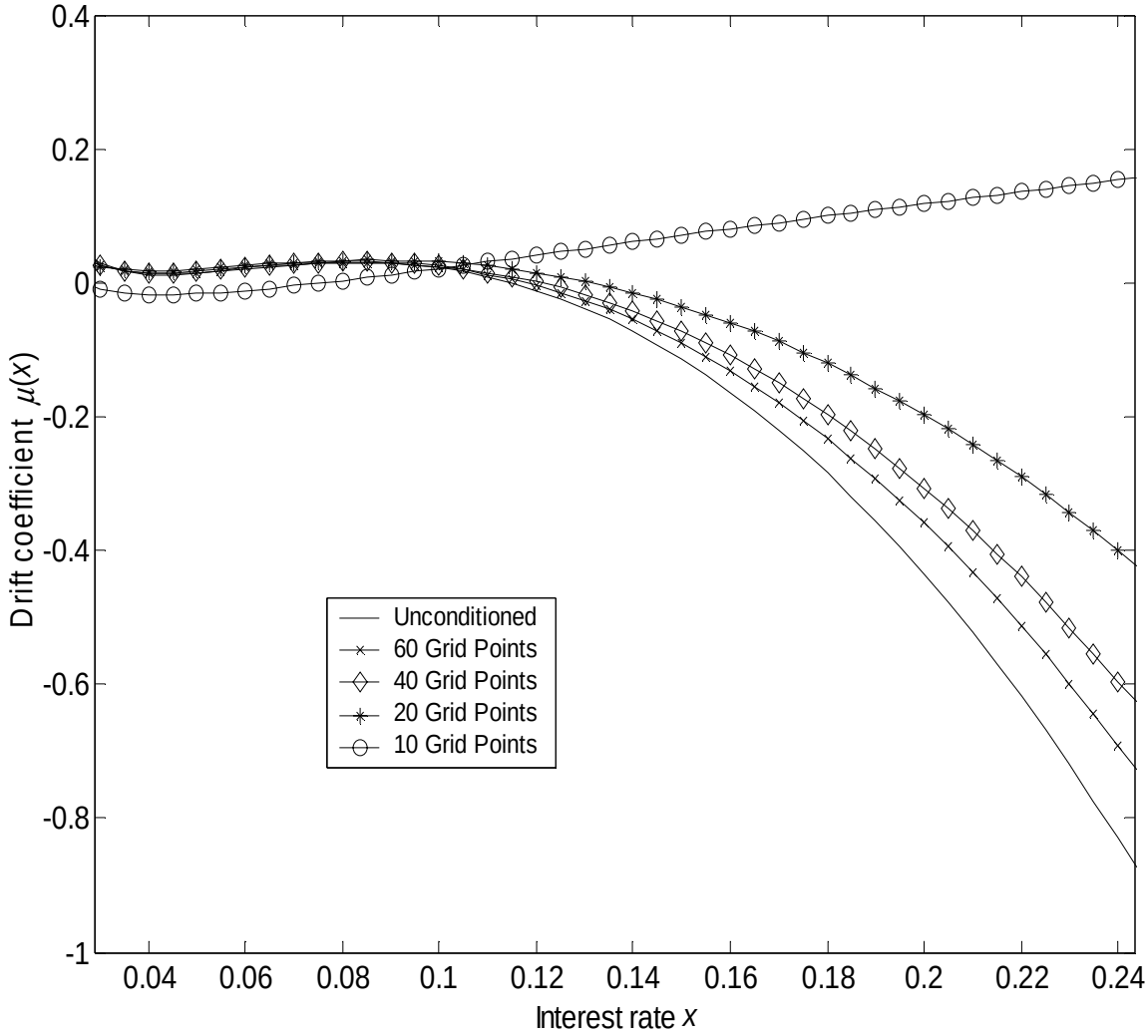

Figure 24: Drift coefficients estimated from seven day Eurodollar rates, June 1, 1973 to February 25, 1995, conditioning on each trade date on the event $C_3(t,T)$. The event $C_3(t,T)$ is the event that the minimum and maximum of the discrete-time process $\{x(t_i)\}$ remain between $\bar{a}$ and $\bar{b}$ until time T, February 25, 1995. The drift and diffusion coefficients $\mu(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x$ and $\sigma^2(x) = \beta_1 x^{2\beta_2}$ were estimated from the Eurodollar rate data using the generalized method of moments, and the moment conditions were constructed using one trade date conditional 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 25, 1995. The line with crosses $\times$ is the drift estimate obtained when on each trade the one trade date expected change is computed conditional on the process remaining between $\bar{a} = a - \text{displacement} \times \Delta x$ and $\bar{b} = b + \text{displacement} \times \Delta x$ for the remainder of the time until February 25, 1995, where and a and b 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 diamonds is the drift estimate when the conditioning is done using a displacement of $40\Delta x$ from the observed minimum and maximum. The line with asterisks is for a displacement of $20\Delta x$, and the line with circles is for a displacement of $10\Delta x$.

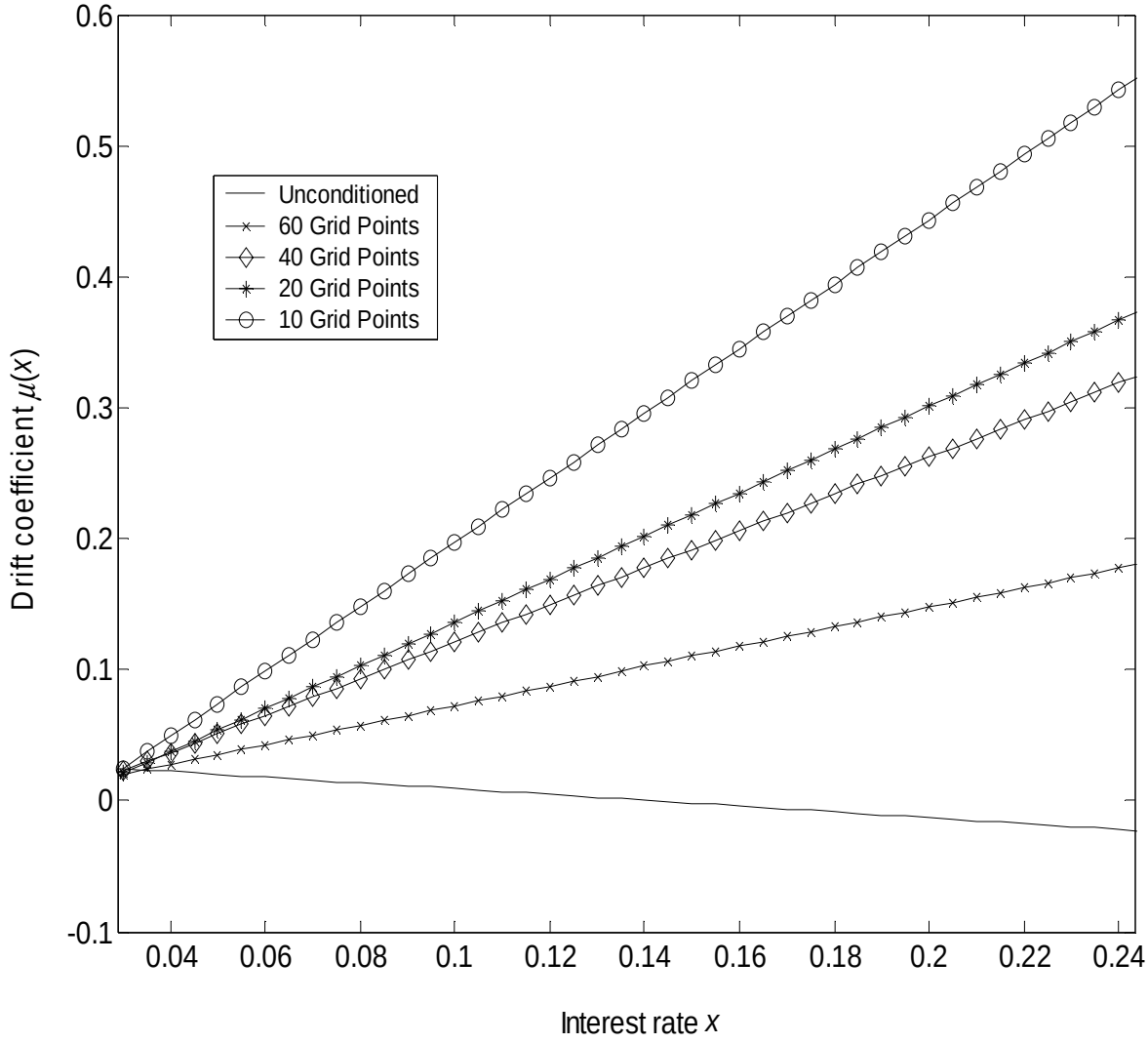

Figure 25: Linear drift coefficients estimated from seven day Eurodollar rates, June 1, 1973 to February 25, 1995, conditioning on each trade date on the event $C_3(t,T)$. The event $C_3(t,T)$ is the event that the minimum and maximum of the discrete-time process $\{x(t_i)\}$ remain between $\bar{a}$ and $\bar{b}$ until time T, February 25, 1995. The drift and diffusion coefficients $\mu(x) = \alpha_0 + \alpha_1 x$ and $\sigma^2(x) = \beta_1 x^{2\beta_2}$ were estimated from the Eurodollar rate data using the generalized method of moments, and the moment conditions were constructed using one trade date conditional 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 25, 1995. The line with crosses $\times$ is the drift estimate obtained when on each trade the one trade date expected change is computed conditional on the process remaining between $\bar{a} = a - \text{displacement} \times \Delta x$ and $\bar{b} = b + \text{displacement} \times \Delta x$ for the remainder of the time until February 25, 1995, where and a and b 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 diamonds is the drift estimate when the conditioning is done using a displacement of $40\Delta x$ from the observed minimum and maximum. The line with asterisks is for a displacement of $20\Delta x$, and the line with circles is for a displacement of $10\Delta x$.

Table 1

Unrestricted Estimates for the Event A_3

The event A_3 is that the process remains in a box $(\bar{a}, \bar{b}) \times [0, T]$, where the lower boundary $\bar{a} = a -$ displacement $\times \Delta x$, the upper boundary $\bar{b} = b +$ displacement $\times \Delta x$, and a and b 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 G in equation (14). Numbers in parentheses are the corresponding standard errors. At the point estimates, the values of the quadratic form G are all less than 10^{-15}. For comparison, the table also shows the parameter estimates when there is no conditioning.

Displacement	α_0	α_1	α_2	α_3	β_1	β_2
Unconditioned	-0.4296	7.6782	-39.5109	0.007892	2.0839	1.3363
	(0.34)	(5.01)	(21.79)	(0.0065)	(0.61)	(0.061)
60	-0.2966	5.5026	-28.3116	0.005478	2.0608	1.3342
	(0.44)	(6.94)	(33.59)	(0.0083)	(0.60)	(0.061)
40	-0.2081	4.0348	-20.2803	0.003877	2.0476	1.3330
	(0.50)	(8.09)	(42.95)	(0.0093)	(0.60)	(0.061)
20	-0.2046	3.5538	-8.2587	0.002639	2.0254	1.3309
	(0.89)	(12.36)	(28.28)	(0.0141)	(0.59)	(0.061)
10	-0.5060	7.3658	-18.7245	0.008456	2.0240	1.3309
	(0.59)	(8.99)	(28.89)	(0.0098)	(0.59)	(0.061)

Table 2

Restricted Estimates for the Event A_3

The event A_3 is that the process remains in a box $(\bar{a}, \bar{b}) \times [0, T]$, where the lower boundary $\bar{a} = a -$ displacement $\times \Delta x$, the upper boundary $\bar{b} = b +$ displacement $\times \Delta x$, and a and b 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 four columns show the parameter estimates obtained by imposing the restriction $\alpha_2 = \alpha_3 = 0$ and minimizing the quadratic form G_R in equation (88). Numbers in parentheses are the corresponding standard errors. For comparison, the table also shows the parameter estimates when there is no conditioning. The last column shows the test statistic $G_R - G$, which is asymptotically χ^2 with two degrees of freedom.

Displacement	α_0	α_1	α_2	α_3	β_1	β_2	$G_R - G$
Unconditioned	0.03125	-0.2211	0	0	2.0403	1.3346	4.650
	(0.017)	(0.30)			(0.59)	(0.060)	
60	-0.01676	1.3390	0	0	2.0836	1.3356	0.280
	(0.022)	(1.15)			(0.59)	(0.059)	
40	-0.0206	1.5225	0	0	2.0620	1.3339	0.076
	(0.018)	(1.13)			(0.58)	(0.059)	
20	-0.0347	2.0107	0	0	2.0216	1.3305	0.006
	(0.027)	(1.50)			(0.57)	(0.059)	
10	-0.0498	2.5626	0	0	2.0865	1.3401	0.807
	(0.020)	(1.14)			(0.63)	(0.061)	

Table 3

Unrestricted Estimates for the Event A_0

The event A_0 is that the process: (1) achieves the minimum a and maximum b; and (2) remains in a box $(\bar{a}, \bar{b}) \times [0, T]$, where the lower boundary is $\bar{a} = a - \text{displacement} \times \Delta x$ and the upper boundary is $\bar{b} = b + \text{displacement} \times \Delta x$. 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 G in equation (14). Numbers in parentheses are the corresponding standard errors. At the point estimates, the values of the quadratic form G are all less than 10^{-15}. For comparison, the table also shows the parameter estimates when there is no conditioning.

Displacement	α_0	α_1	α_2	α_3	β_1	β_2
Unconditioned	-0.4296	7.6782	-39.5109	0.007892	2.0839	1.3363
	(0.34)	(5.01)	(21.79)	(0.0065)	(0.61)	(0.061)
60	-0.5498	9.7904	-50.2898	0.01064	2.0668	1.3348
	(0.59)	(9.73)	(47.60)	(0.0102)	(0.60)	(0.061)
40	-0.4891	8.7744	-45.2503	0.009620	2.0526	1.3335
	(0.58)	(9.45)	(46.48)	(0.0099)	(0.60)	(0.062)
20	-0.3740	6.8253	-35.3517	0.007716	2.0292	1.3313
	(0.56)	(9.28)	(46.51)	(0.0097)	(0.60)	(0.062)
10	-0.2684	5.0549	-26.1224	0.005897	2.0145	1.3299
	(0.58)	(9.71)	(50.59)	(0.0100)	(0.60)	(0.063)

Table 4

Restricted Estimates for the Event A_0

The event A_0 is that the process: (1) achieves the minimum a and maximum b; and (2) remains in a box $(\bar{a}, \bar{b}) \times [0, T]$, where the lower boundary is $\bar{a} = a - \text{displacement} \times \Delta x$ and the upper boundary is $\bar{b} = b + \text{displacement} \times \Delta x$. 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 four columns show the parameter estimates obtained by imposing the restriction $\alpha_2 = \alpha_3 = 0$ and minimizing the quadratic form G_R in equation (88). Numbers in parentheses are the corresponding standard errors. For comparison, the table also shows the parameter estimates when there is no conditioning. The last column shows the test statistic $G_R - G$, which is asymptotically χ^2 with two degrees of freedom.

Displacement	α_0	α_1	α_2	α_3	β_1	β_2	$G_R - G$
Unconditioned	0.03125	-0.2211	0	0	2.0403	1.3346	4.650
	(0.017)	(0.30)			(0.59)	(0.060)	
60	0.05317	-0.0206	0	0	2.0567	1.3319	0.979
	(0.064)	(1.13)			(0.59)	(0.061)	
40	0.05063	0.03403	0	0	2.0467	1.3311	0.760
	(0.065)	(1.16)			(0.59)	(0.061)	
20	0.04858	0.0770	0	0	2.0228	1.3290	0.550
	(0.065)	(1.20)			(0.58)	(0.061)	
10	0.03077	0.4088	0	0	2.0130	1.3286	0.297
	(0.062)	(1.28)			(0.59)	(0.061)	

Table 5

Unrestricted Estimates for the Event C_3

The event C_3 is that observations of the discretely sampled process $\{x(t_i)\}$ remain in a box $(\bar{a}, \bar{b}) \times [0, T]$, where the lower boundary is $\bar{a} = a -$ displacement $\times \Delta x$, the upper boundary is $\bar{b} = b +$ displacement $\times \Delta x$, and a and b 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 G in equation (14). Numbers in parentheses are the corresponding standard errors. At the point estimates, the values of the quadratic form G are all less than 10^{-15}. For comparison, the table also shows the parameter estimates when there is no conditioning.

Displacement	α_0	α_1	α_2	α_3	β_1	β_2
Unconditioned	-0.4296	7.6782	-39.5109	0.007892	2.0839	1.3363
	(0.34)	(5.01)	(21.79)	(0.0065)	(0.61)	(0.061)
60	-0.3540	6.4395	-33.1488	0.006521	2.0700	1.3350
	(0.41)	(6.37)	(29.60)	(0.0078)	(0.61)	(0.061)
40	-0.3031	5.6065	-28.8354	0.005598	2.0617	1.3343
	(0.45)	(6.95)	(33.46)	(0.0084)	(0.60)	(0.061)
20	-0.2117	4.0932	-20.6111	0.003938	2.0487	1.3331
	(0.51)	(8.18)	(46.06)	(0.0095)	(0.60)	(0.061)
10	-0.1364	1.4454	-1.1776	0.002524	2.0243	1.3328
	(1.31)	(24.71)	(81.60)	(0.0202)	(0.60)	(0.061)

Table 6

Restricted Estimates for the Event C_3

The event C_3 is that observations of the discretely sampled process $\{x(t_i)\}$ remain in a box $(\bar{a}, \bar{b}) \times [0, T]$, where the lower boundary is $\bar{a} = a -$ displacement $\times \Delta x$, the upper boundary is $\bar{b} = b +$ displacement $\times \Delta x$, and a and b 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 four columns show the parameter estimates obtained by imposing the restriction $\alpha_2 = \alpha_3 = 0$ and minimizing the quadratic form G_R in equation (88). Numbers in parentheses are the corresponding standard errors. For comparison, the table also shows the parameter estimates when there is no conditioning. The last column shows the test statistic $G_R - G$, which is asymptotically χ^2 with two degrees of freedom.

Displacement	α_0	α_1	α_2	α_3	β_1	β_2	$G_R - G$
Unconditioned	0.03125	-0.2211	0	0	2.0403	1.3346	4.650
	(0.017)	(0.30)			(0.59)	(0.060)	
60	-0.002141	0.7479	0	0	2.0884	1.3359	0.574
	(0.045)	(1.34)			(0.59)	(0.059)	
40	-0.01951	1.4090	0	0	2.0833	1.3356	0.322
	(0.026)	(1.36)			(0.59)	(0.059)	
20	-0.02882	1.6491	0	0	2.0732	1.3351	0.105
	(0.045)	(2.06)			(0.59)	(0.059)	
10	-0.04912	2.4633	0	0	2.1516	1.3448	0.665
	(0.021)	(1.19)			(0.59)	(0.057)	