

Nonlinear Mean Reversion in the Short-Term Interest Rate

Christopher S. Jones

University of Southern California

Using a new Bayesian method for the analysis of diffusion processes, this article finds that the nonlinear drift in interest rates found in a number of previous studies can be confirmed only under prior distributions that are best described as informative. The assumption of stationarity, which is common in the literature, represents a nontrivial prior belief about the shape of the drift function. This belief and the use of “flat” priors contribute strongly to the finding of nonlinear mean reversion. Implementation of an approximate Jeffreys prior results in virtually no evidence for mean reversion in interest rates unless stationarity is assumed. Finally, the article documents that nonlinear drift is primarily a feature of daily rather than monthly data, and that these data contain a transitory element that is not reflected in the volatility of longer-maturity yields.

The drift of the short-term interest rate is an important determinant of a wide variety of asset prices, both inside and outside the boundaries of what is often called fixed income. While volatilities may be estimated relatively accurately using high-frequency observations of the short rate, the short rate’s extreme persistence makes identifying the true shape of the drift function a particularly elusive goal.

If our goal is merely to fit prices, then the difficulty in estimating a drift function can be avoided by backing out implied drifts using a no-arbitrage approach such as Hull and White (1990) or Heath, Jarrow, and Morton (1992). These methods bypass the true distribution entirely, focusing on solving for the risk-neutral drift function that is consistent with the cross section of bond prices. If our goal, however, is to learn from prices and to be able to assess theories, such as the expectations hypothesis, that link short and long rates, then estimating the drift function under the true measure is unavoidable.

In several recent articles, a variety of sophisticated econometric techniques have been brought to bear on the problem. In particular,

I am grateful to Yacine Aït-Sahalia, Geert Bekaert, Michael Brandt, Dave Chapman, Frank Diebold, Eric Jacquier, Ron Kaniel, Craig MacKinlay, Luboš Pástor, Matt Pritsker, Krishna Ramaswamy, and especially Robert Stambaugh for many helpful comments and discussions. In addition, the article substantially benefited from the suggestions of the editor, John Heaton, and two anonymous referees. Comments from seminar participants at Berkeley, Columbia, Duke, NYU, Rochester, Wharton, the Federal Reserve Board of Governors, and the 1998 meetings of the Western Finance Association are gratefully acknowledged. Finally, I thank Yacine Aït-Sahalia for providing his data. All errors are my own. Address correspondence to Christopher S. Jones, Marshall School of Business, 701 Hoffman Hall, University of Southern California, Los Angeles, CA 90089, or e-mail: christopher.jones@marshall.usc.edu

Aït-Sahalia (1996a) and Stanton (1997) propose nonparametric-based methods for estimating nonlinear drift and diffusion functions of the short rate. Both articles find that the nonlinearity in the drift function is important. In fact, Aït-Sahalia (1996b, p. 385) concludes, in reference to the linear drift models of Chan et al. (1992) and others, that “the principal source of rejection of existing models is the strong nonlinearity of the drift.” He finds that interest rates behave like a random walk over nearly their entire historical range, reverting toward the middle of this range only when they become very high or very low. In a fully nonparametric analysis, Stanton (1997) estimates a comparable drift function, with very little mean reversion for all rates below 15% but substantial negative drift for higher rates.

Similar results are reported by Conley et al. (1997; hereafter CHLS), who estimate a drift function that is nonzero only for rates below 3% or above 11%. Jiang and Knight (1997) find a comparable pattern of nonlinear mean reversion in a sample of Canadian interest rates.

Figure 1 plots the drift functions—the expected change in the short rate per year as a function of the level of the rate—estimated by Aït-Sahalia and CHLS. While there is general agreement that higher interest rates tend to drift downward and low rates upward, how high and how low rates must be for this to happen remains a point of contention, with Aït-Sahalia assigning random walk-like behavior for a much wider range of interest rates.

One possible criticism of all of these articles is that each assumes the stationarity of the interest rate process, a characteristic that has a great deal of economic appeal but which fails to receive strong support in formal tests.^{1,2} Aït-Sahalia’s estimator, since it requires the nonparametric estimation of the marginal density of the spot rate, is undefined if rates are nonstationary. The CHLS approach relies on the moment conditions of Hansen and Scheinkman (1995), which can loosely be interpreted as statements of the fact that functions of stationary processes have an unconditionally zero drift. Because imposing stationarity of the short rate puts restrictions on the possible shape of its drift function, any analysis that imposes this restriction runs the risk of mechanically assuming away the question of interest, no matter how appealing the restriction seems.

¹ The exception is Stanton’s (1997) method, which Bandi and Phillips (2001) show is applicable to processes that are recurrent, a significantly weaker condition than stationarity.

² Pagan, Hall, and Martin (1996) report Dickey–Fuller statistics for a variety of short-term interest rates that are generally between 0 and -2 , not negative enough to reject the unit root at a 95% confidence level. They further report evidence that the presence of a levels effect in variance reduces the Dickey–Fuller critical value, making the presence of a unit root even more difficult to reject. The Dickey–Fuller statistic for Aït-Sahalia’s data set is calculated to be -2.29 , also higher than its 5% critical value of -2.88 .

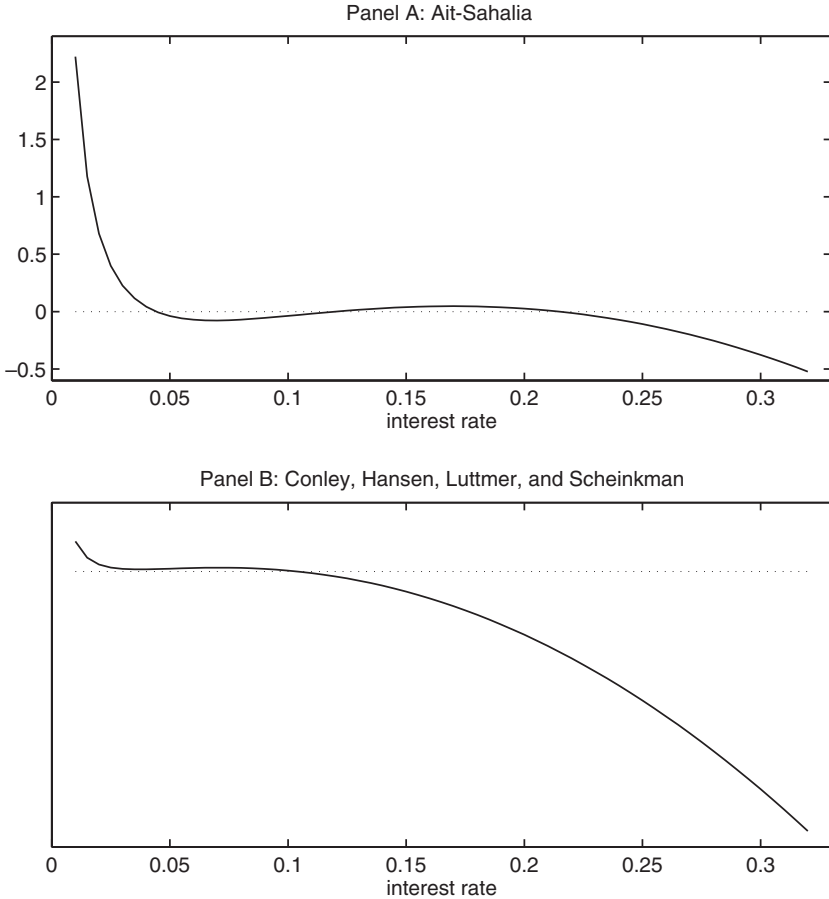


Figure 1
Drift function estimates

Panel A contains the drift function estimated by Ait-Sahalia (1996b). Units on the vertical axis denote the expected annualized change in the short rate. Panel B graphs the drift function estimated by Conley et al. (1997) under the assumption that the variance elasticity parameter γ is equal to 1.5, which is close to the values estimated in this article. Because of their treatment of the short rate as a subordinated process, the scale of the vertical axis is unidentified.

Even if the short rate is stationary, its high degree of persistence may make small-sample inference problematic. Several recent Monte Carlo studies have examined the finite sample performance of estimators used in the previous articles and have concluded that this performance can be deficient.

Pritsker (1998) finds that the asymptotic significance levels of Ait-Sahalia's (1996b) specification test are often inappropriate in finite samples. He notes that nonparametric procedures have been predominantly

studied in an *i.i.d.* setting, and that little is known about optimal implementation of these procedures (particularly the choice of bandwidth) when the data generation process is highly persistent, as is the case with interest rates. Persistence is also unrecognized by Aït-Sahalia's test statistic, since it is not a concern in large samples. In a careful consideration of the case of Vasicek (1977) interest rates, Pritsker finds that the asymptotic test rejects the true null approximately 50% of the time in some cases.

More relevant to the current study, Chapman and Pearson (2000) find that both Stanton's and Aït-Sahalia's estimators display a finite sample bias toward finding nonlinearity in a drift that is actually linear. Chapman and Pearson attribute this bias to the nonparametric procedures that underlie both of these articles' estimation methods and contend that the evidence provided by Stanton and Aït-Sahalia is insufficient to conclude that nonlinear drift is a "robust stylized fact."

While both Pritsker (1998) and Chapman and Pearson (2000) suggest that nonparametric methods may be unreliable in the detection of nonlinearities, there is a more general concern in estimating time-series models that has nothing to do with nonparametric methods. The problem is that standard estimators such as ordinary least squares and maximum likelihood are generally biased for time-series models. In the first-order autoregressive model, for example, it is wellknown that in finite samples the autoregressive coefficient is biased toward zero. How this bias generalizes to more complicated nonlinear models is unknown.

In spite of these problems, analysis of nonlinear mean reversion remains important simply because of its relevance for so many economic issues. Nonlinear drift offers potential improvements in fixed income pricing, as Ahn and Gao (1999) have recently demonstrated, and is also compelling because it has the potential to explain, at least in part, a number of the outstanding puzzles about the term structure. Bekaert, Hodrick, and Marshall (2000) propose to explain empirical findings of the expectations hypothesis using a regime-switching model that Ang and Bekaert (2002) have shown is capable of capturing nonlinear behavior in the short rate. Pfann, Schotman, and Tschernig (1996) observe that the nonlinear relations that exist between short and long yields and also between their volatilities are also consistent with nonlinear models of the short rate. Finally, nonlinearity in the short-rate drift might explain why standard tests of stationarity generally do not reject the unit root. Because Dickey-Fuller tests are based on the assumption of a linear autoregressive model, data generated by a stationary nonlinear drift model could have little power to reject the unit root.

In order to remain consistent with previous literature, I focus on representations of the short-term interest rate as a continuous-time diffusion process. This decision reflects the fact that diffusions are the modeling

framework of choice for much of modern asset pricing. With this asset pricing theory, prices of related fixed-income securities can be calculated without resorting to linear or log-linear approximations that may not hold accurately. More importantly, diffusions provide a parsimonious framework for examining data of different frequencies, since a single diffusion model automatically determines conditional distributions of the process at all time horizons.

Because of the problems that have been attributed to the use of asymptotic frequentist methods, particularly those which rely on the stationarity of the process under consideration, this article takes a Bayesian perspective. A primary task of the article is therefore to introduce a new method for the Bayesian analysis of diffusion processes that will generate exact finite-sample inferences even for nonstationary models.

Using this method I reassess the evidence for the drift nonlinearities first identified by Ait-Sahalia using a time series of short-term interest rates. Robustness of the results will be evaluated by comparing results generated under a variety of priors, where each is chosen to represent some notion of prior ignorance. This type of Bayesian analysis, suggested by Leamer (1985) and Poirier (1995), interprets the sensitivity of results to the specification of the prior as evidence against the availability of an “objective” conclusion.

The results of this article demonstrate that fully efficient parametric analysis may be no less problematic than nonparametric analysis, and that conclusions in favor of nonlinear drift may largely be driven by implicit prior beliefs that contain a nontrivial amount of information about the shape of the drift function. This article shows that the priors that generate nonlinear drift may reasonably be interpreted as informative, and that under other priors the result disappears completely.

Lastly, the article identifies that the evidence favoring nonlinear drift is primarily a feature of high-frequency data, and that these data contain a transitory noise component that accounts for roughly half the daily variation in the short rate. The analysis reveals an obvious misspecification of the one-factor model, so I propose a simple two-factor extension with a latent nonlinear stochastic mean process. The generalized model reconciles the different sampling interval results and provides further evidence against the nonlinearities identified previously.

The article proceeds as follows: Section 1 reviews previous work in modeling nonlinear interest rate processes. Section 2 develops a Bayesian method for estimating parameters of discretely observed diffusion processes. The method is applied in section 3 to analyze nonlinear mean reversion under the different prior distributions. Section 4 checks for model misspecification and introduces the stochastic mean model. Section 5 concludes.

1. Modeling Nonlinear Drift in the Short-Term Interest Rate

Within the class of one-factor models, the interest rate process has traditionally been modeled as having a linear drift, often with a constant elasticity of variance. As a diffusion, the process is written as

$$dr_t = \kappa(\mu - r_t)dt + \sigma r_t^\gamma dB_t. \quad (1)$$

In this form, the parameters of the process each have intuitive meaning. The long-run mean of the process, toward which rates drift, is given by μ , and the speed of this drift is given by κ . Volatility is measured by σ , and the variance elasticity is given by 2γ . With $\gamma=0$, this is the model of Vasicek (1977), while for $\gamma=.5$, it is the specification used by Cox, Ingersoll, and Ross (1985). A thorough examination of this class of models was carried out by Chan et al. (1992).

For simple linear models such as these, estimating the drift may be as simple as running a least squares regression. These same models, however, have often been found to be unsatisfactory in their description of short-rate dynamics and their implications for other security prices. The alternatives that have been proposed are often a great deal more complex. Gray (1996), Pfann, Schotman, and Tschernig (1996), and Naik and Lee (1993), for example, have generalized standard models to include regime shifts, Das (2002) and Johannes (2002) add jumps, while Andersen and Lund (1997), Balduzzi, Das, and Foresi (1998), and Jegadeesh and Pennacchi (1996) consider multifactor models in which volatility is stochastic or interest rates revert in a linear fashion toward a stochastic attractor. Articles too numerous to mention have explored other generalizations.

The primary model considered in this article, while more general than those first proposed by Vasicek (1977) and Cox, Ingersoll, and Ross (1985), remains in the single-factor class. This choice reflects a belief that this class of models has not yet been fully explored. At the very least, it seems natural to ask how much of the dynamics of both short- and long-term yields can be explained by a more general one-factor model before considering multifactor models.

Because there is little reason a priori to assume particular specifications of either the drift or diffusion functions, Aït-Sahalia (1996b) advocates the use of flexible functional forms to approximate their true unknown shapes. He proposes the following model of the short rate process:

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 / r_t)dt + \sqrt{\beta_0 + \beta_1 r_t + \beta_2 r_t^{\beta_3}} dB_t. \quad (2)$$

CHLS adopt the same drift parameterization as Aït-Sahalia but keep the constant elasticity of variance (CEV) diffusion used by Chan et al. (1992):

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 / r_t)dt + \sigma r_t^\gamma dB_t. \quad (3)$$

This is the primary model considered in the remainder of the article.

Because it is a characteristic that has generally been assumed in previous work, it is useful to consider the parameter restrictions that are required to generate stationarity. In fact, stationarity of the nonlinear drift model can be achieved in several ways. A simple sufficient condition is that $\alpha_2 < 0$ and $\alpha_3 > 0$. In Aït-Sahalia's (1996b) treatment of this model, these are the parameter restrictions he employs. CHLS note, however, that the restriction on α_2 is unnecessary when $\gamma > 1.5$. In this case the stationarity of the process may be "volatility induced" rather than "drift induced." I will examine the implications of imposing each type of stationarity in the estimation of the model.

2. A Bayesian Method for the Analysis of Diffusion Processes

The primary difficulty in estimating diffusion processes stems from the intractability of their transition densities and hence likelihood functions.³ Because the Bayesian posterior distribution is typically attained as the normalized product of the prior distribution and the likelihood function, the unknown form of the likelihood impedes Bayesian analysis as well. I address this problem by using a combination of simple numerical techniques: the Euler approximation, the Gibbs sampler, and the Metropolis–Hastings algorithm. By combining these tools appropriately, posterior distributions of the parameters of the diffusion process can be generated to any desired degree of accuracy. This is accomplished by generating thousands of draws from these multivariate posteriors. Given a large enough set of such draws, moments, confidence intervals, and marginal densities of the parameters can be computed easily.

The econometric approach of this article is based in a strand of statistics known as Markov chain Monte Carlo (MCMC). Appendix A provides a brief introduction to the Gibbs sampler, perhaps the simplest example of MCMC.⁴ The following sections assume a casual familiarity with this technique.

2.1 Augmenting with high-frequency data

The approach of this article to estimating diffusions is based on a simple intuition: if the diffusion,

$$dr_t = \mu(r_t, \phi)dt + \sigma(r_t, \phi)dB_t, \tag{4}$$

³ Aït-Sahalia (2002) has shown how to construct analytical approximations of the likelihood function of a univariate diffusion process. Even in the univariate case, however, the likelihood function is of a nonstandard form, making the derivation of marginal posterior densities for a subset of the parameters problematic when there are more than a few parameters.

⁴ Casella and George (1992) provide a much more detailed introduction to the Gibbs sampler.

is observed often enough, then the biases caused by the estimation of the discretized version (the Euler approximation),

$$r_t - r_{t-1} = \mu(r_{t-1}, \phi) + \sigma(r_{t-1}, \phi)\epsilon_t, \quad (5)$$

should be negligible.

In theory, therefore, we could avoid the elaborate econometrics of continuous-time processes by simply restricting our analysis to high-frequency data. Unfortunately high-frequency data are not always available, particularly for less recent historical periods. And even if, say, daily data were available, would it be of sufficiently high frequency to render discretization bias insignificant? There is in general no way to answer this question except through empirical investigation with a method that can be used to account for this bias.

Even if the data were available, there are a number of reasons why the use of high-frequency asset price data may be undesirable. Price discreteness, infrequent trading, intraday volatility periodicity, bid-ask bounce, and periodic market closure are all difficult to reconcile with the simple and elegant properties of the diffusion process. Although some of these problems can be corrected for using simple modifications of the procedure proposed here, each would invalidate the simple estimation of the discretized process of Equation(5).

The resolution proposed in Jones (1999) is to use Tanner and Wong's (1987) data augmentation algorithm to augment the observed data with paths of much higher frequency data — for example, augmenting monthly with daily data.⁵ As these augmented data are added at closer and closer intervals, the likelihood of the discretized approximation will converge to that of the true diffusion likelihood, following the results of Pedersen (1995) and Brandt and Santa-Clara (2002). In practice, the frequency of the Euler approximation will be chosen to be high enough so that it will have approximately the same distribution as the diffusion of interest. This will generally mean that the observed data are of a lower frequency than the frequency at which the Euler approximation operates. For example, we may be working with month-end data, but a reasonable diffusion approximation might require 10 discrete time transitions per month, making 9 out of every 10 data points unobserved.

Conditional on this unobserved high-frequency data in addition to the observable low-frequency data, a distribution for the model parameters may usually be obtained quite easily. We then integrate out, using a Gibbs sampler-like Markov chain, the dependence on particular paths of unobserved data to get posteriors conditional on only the observed data.

⁵ Similar methods have been proposed independently by Elerian, Chib, and Shephard (2000) and Eraker (2001), although the former's method is applied only to univariate processes.

The idea of augmenting with high-frequency data may be considered the Bayesian counterpart of the simulation-based classical literature on continuous-time econometrics, which typically uses the Euler approximation to compute by simulation objective functions that are analytically intractable. Examples include Duffie and Singleton (1993), Gouriéroux, Monfort, and Renault (1993), Pedersen (1995), Gallant and Tauchen (1996), and Brandt and Santa-Clara (2002). These approaches use the Euler approximation to simulate forward paths of artificial data. Simulated moment-based procedures, for example, use the Euler approximation to simulate long paths of the diffusion which are then used to calculate unconditional moments of the model. Simulated maximum likelihood uses the Euler approximation to compute each one-period transition density numerically, requiring a large number of short simulated paths.

In contrast, the simulations in this article merely “bridge” the observed low-frequency data with short paths of high-frequency data. Each simulation is entirely consistent with the low-frequency data, automatically preserving many of the stylized facts observable in the original data: the general historical shape, the patterns of volatility, and the degree of persistence, for example. Figure 2 illustrates the comparison of high-frequency data augmentation with two classical methods, the simulated method of moments [e.g., Duffie and Singleton (1993)] and simulated maximum likelihood [e.g., Brandt and Santa-Clara (2002)]. It is clear that by pinning down both ends of the simulated paths the variance of the latent high-frequency data can be reduced dramatically relative to other methods. Since all methods require some form of Monte Carlo integration, the lower variance of augmented data results in greater computational efficiency.

It should be emphasized that the purpose of augmenting with high-frequency data is to reduce discretization bias, not add information to the sample. Although each path of high-frequency data will add information to the relatively scarce low-frequency data, by integrating out the dependence on particular high-frequency paths, this information is washed out of the final posterior distribution.

2.2 Details of the Markov chain

To explain the details of the procedure it is necessary to have a more precise statement of the Euler approximation. For maximum intuition, the procedure is described for a univariate process r , although a multivariate generalization is simple and is pursued later in the article. A discrete time process operating on a unit of time of length h , the Euler approximation of Equation (4) may be written as

$$r_{(k+1)h} - r_{kh} = h\mu(r_{kh}, \phi) + \sqrt{h}\sigma(r_{kh}, \phi)\epsilon_k, \quad (6)$$

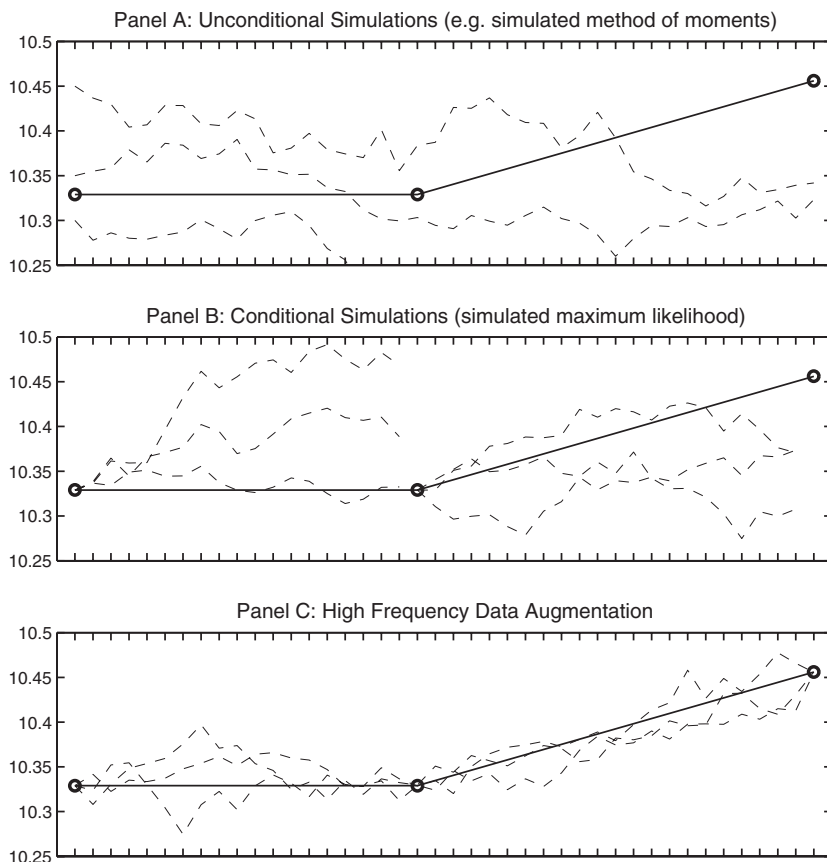


Figure 2
Simulation of high-frequency data

The figure depicts three different schemes for simulating at a higher frequency than the observed data. In panel A, simulations are being used to compute unconditional moments and may have little or no relation to the observed data. The simulations in panel B could be used [see Brandt and Santa-Clara (2002)] to compute numerical approximations of transition densities of the process. Panel C shows how simulated high-frequency data may be used to “bridge” the observed low frequency data.

where $\epsilon_k \sim i.i.d. N(0, 1)$ and ϕ is a vector of parameters. I will assume that data are observed at equally spaced intervals of unit length, and that the interval endpoints correspond to the integer values of kh . In other words, the Euler approximation breaks up the observation interval into $1/h$ subperiods, each of length h . When the dependence on a particular value of h is implicit, it is convenient to let r_k denote r_{kh} .

From Pedersen (1995) or Brandt and Santa-Clara (2002) we know that under regularity conditions the likelihood of the Euler approximation converges to that of the diffusion as $h \rightarrow 0$. The approach will therefore allow h to be arbitrarily small regardless of the frequency of the observed data.

Let \mathbf{R}^o denote the set of all the observed low-frequency data, corresponding to integer values of kh . Let \mathbf{R}^u denote the unobserved high-frequency data, corresponding to noninteger kh . Following the intuition of the Gibbs sampler, the Markov chain will alternate between drawing from the conditional distributions $p(\phi|\mathbf{R}^o, \mathbf{R}^u)$ and $p(\mathbf{R}^u|\phi, \mathbf{R}^o)$.

We draw from the distribution of the model parameters conditional on both observed and augmented data. From Bayes' rule,

$$p(\phi|\mathbf{R}^o, \mathbf{R}^u) \propto L(\phi; \mathbf{R}^o, \mathbf{R}^u)p(\phi), \tag{7}$$

where L is the likelihood function and $p(\phi)$ is the prior. The Euler approximation allows us to compute $L(\phi; \mathbf{R}^o, \mathbf{R}^u)$ as the product of Gaussian transition densities, allowing the computation of the conditional density $p(\phi|\mathbf{R}^o, \mathbf{R}^u)$ up to a constant of proportionality. Frequently this density is a highly tractable form, and standard Gaussian methods may be used to draw ϕ . At the very least, knowledge of the distribution makes it possible to draw the parameter vector ϕ using a numerical procedure such as the Metropolis–Hastings algorithm. This step is described in more detail for a specific example in the next section.

If it were possible to draw directly from the distribution $p(\mathbf{R}^u|\phi, \mathbf{R}^o)$, then the specification of the Markov chain would be complete. In even the simplest cases, however, this high-dimensional distribution is of unknown form, meaning that an additional numerical technique must be applied.

I adapt a technique proposed by Jacquier, Polson, and Rossi (1994) for the analysis of discrete-time stochastic volatility models. It is termed a cyclic Metropolis chain because it “cycles” through the individual elements of \mathbf{R}^u , drawing values of \mathbf{R}^u point by point using the Metropolis–Hastings algorithm at each step. In essence, we make each element of \mathbf{R}^u a separate block in the Markov chain. Thus if there were 1000 elements of \mathbf{R}^u , we would have 1001 block draws in the Markov chain: 1000 draws of high-frequency data points and one draw of ϕ .⁶

Appendix B describes the data augmentation procedure in greater detail.

3. Estimating the Short-Rate Model

The primary model of nonlinear drift considered in the remainder of the article is

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3/r_t)dt + \sigma r_t^\gamma dB_t. \tag{8}$$

⁶ Although it is most convenient to illustrate a Markov chain that “cycles” through the elements of \mathbf{R}^u drawing each element r_k individually, this is not generally the most computationally efficient method. In computer languages that are optimized for matrix operations (such as Matlab) we can increase speed by performing multiple draws simultaneously. Because of the Markovian nature of the problem, this is not problematic as long as adjacent elements of \mathbf{R}^u are not drawn simultaneously. By drawing every other element of \mathbf{R}^u at the same time, the Markov chain can be reduced to just three “blocks.”

Special cases of the model include Vasicek; Cox, Ingersoll, and Ross; and the linear drift class considered by Chan et al.

The Euler approximation of the nonlinear drift model is given by

$$r_{k+1} - r_k = h(\alpha_0 + \alpha_1 r_k + \alpha_2 r_k^2 + \alpha_3 / r_k) + \sqrt{h} \sigma r_k^\gamma \epsilon_{k+1}. \tag{9}$$

It is important to note that the usual sufficient conditions for convergence of the Euler approximation are not met by this model. Specifically, growth and Lipschitz are conditions are violated as $r \rightarrow 0$ and as $r \rightarrow \infty$. Because the minimum and maximum interest rates observed in the sample are about 3% and 24%, respectively, it is possible that failure of these conditions in regions far from where the data were actually realized is unimportant. Because of these concerns, however, several tests of convergence are performed in Appendix C, with the results highly supportive of convergence.

In particular, the appendix shows that augmenting with high-frequency data is particularly important when looking at monthly data, as interest rates generated by a naive discretization ($h = 1$) can easily be rejected as coming from the corresponding diffusion process. By reducing the discretization interval to .05 or .2, however, the tests no longer result in rejections. In the simulation of daily data, discretization bias is not detected, implying that discretization bias may not be very important for these data. With the support these results provide, we proceed with the use of the discretization scheme.

The heteroscedasticity in Equation (9) may be eliminated by rearranging the Euler approximation as

$$\begin{aligned} \frac{r_{k+1} - r_k}{\sqrt{h} r_k^\gamma} &= \alpha_0 \sqrt{h} r_k^{-\gamma} + \alpha_1 \sqrt{h} r_k^{1-\gamma} + \alpha_2 \sqrt{h} r_k^{2-\gamma} \\ &+ \alpha_3 \sqrt{h} r_k^{-1-\gamma} + \sigma \epsilon_{k+1} \end{aligned} \tag{10}$$

were γ considered a known constant rather than a parameter to be estimated, this rearrangement falls under the standard homoscedastic linear regression framework. The “flat” prior $p(\alpha, \sigma) = 1/\sigma$ and the normality of the ϵ_k would therefore lead to a Student-*t*/inverted gamma distribution for the parameter vector (α, σ) conditional on the full set of actual and augmented data. In order to estimate γ , an additional Metropolis–Hastings step must be added. This step is described in Appendix D.

3.1 The data

The time series used to proxy for the short-term interest rate is the same seven-day Eurodollar rate series used by Ait-Sahalia (1996b). The data are graphed in Figure 3. This daily series, with 5505 observations, covers the period from June 1, 1973, to February 25, 1995.

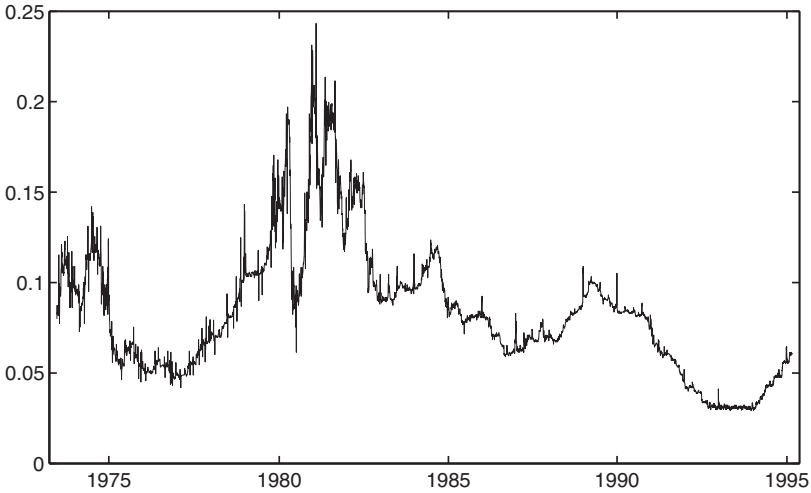


Figure 3
The one-week Eurodollar rate
The figure depicts daily observations of the one-week Eurodollar rate from June 1, 1973, to February 25, 1995.

One goal of this article is to determine the robustness of nonlinear mean reversion to different sampling intervals. In addition to estimating the model using the entire daily sample, I will repeat the estimations using only the 261 month-end observations. While the daily data have the potential of adding additional information, they appear to be very noisy with many highly transitory shocks, especially in the first half of the sample. Part of this noise appears to be microstructure-related, since the reported rates are usually approximate multiples of one-sixteenth of 1%. Monthly data should allow us to mitigate the effects of this predominately high-frequency noise. In any case, if our primary concern is to learn about the drift of the process, it is likely that monthly and daily data will yield similar results, as higher-frequency observation tends to add little information about parameters of the drift.

3.2 Prior distributions

I will consider several prior distributions with the goal of determining how different prior beliefs affect conclusions about the shape of the drift function. The two classes of priors are considered—the flat prior and an approximate Jeffreys prior—are both chosen to represent different notions of prior ignorance. Within each class I will consider differing prior beliefs about stationarity. The first is a prior that is not informative about stationarity. The second is a prior that contains a belief that the process is stationary with probability one. The last is a belief that the process is stationary, and furthermore, that the stationary is drift induced,

corresponding to the parameter restrictions imposed by Aït-Sahalia (1996b). Differences in conclusions across the six priors will be taken as evidence of a Bayesian small sample problem, in which no “objective” Bayesian inference is possible.

The first class, the flat prior, is particularly easy to work with and is interesting for a variety of reasons. Flat priors allow us to examine most directly the shape of the likelihood function. Since the flat prior mode is typically very close to the maximum-likelihood estimate, sometimes identical to it, flat prior results have a frequentist interpretation. In addition, the flat prior is also a natural choice since it is the prior that is often favored by applied Bayesian researchers.

In the case of exogenous regressors, the flat prior has a more theoretical grounding as well, since it is synonymous with the Jeffreys prior, which is known to have many desirable properties. One such property is that the Jeffreys prior is invariant to reparameterizations of the model—two models parameterized differently will yield the same results if each is analyzed under the Jeffreys prior derived under its own parameterization. Another is the fact that the Jeffreys prior is the prior distribution that minimizes Shannon’s commonly used measure of information, giving a more formal justification for the view that the Jeffreys prior is maximally ignorant.

When regressors are endogenous, the flat and Jeffreys priors no longer coincide.⁷ As has been argued forcefully by Phillips (1991a, 1991b), flat priors can be quite informative for time-series models. In his analysis of the first-order autoregressive model, $y_t = \rho y_{t-1} + \epsilon_t$, Phillips notes that the data should be expected to do a better job distinguishing nearby values of the autoregressive parameter ρ when the true value of ρ is close to or within the explosive region $|\rho| \geq 1$. Intuitively, if y explodes then the ratio of signal to noise about ρ goes to infinity, since the mean is level dependent but the variance is not. In a frequentist setting this behavior leads to the superconsistency and downward bias of the MLE estimator.

Phillips argues that the flat prior, by ignoring this property of the model, effectively imposes a prior view that explosive behavior is improbable. Mechanically the MLE estimate of ρ is identical to its Bayesian posterior mean. By not anticipating and correcting for the bias of the MLE estimator, the researcher is implicitly taking an informed view that this bias is somehow desirable.

Proposing to use the Jeffreys prior as a better representation of prior ignorance, Phillips derives the Jeffreys prior for the AR(1) model and finds that it assigns much higher prior densities to values of ρ in the explosive region than to nonexplosive values of ρ . In effect, the Jeffreys

⁷ Stambaugh (1998) is a recent case in which flat and Jeffreys priors are different both in their form and in the inferences drawn using them.

prior offsets the finite sample bias of MLE. Phillips finds that the conclusions that result from using the Jeffreys prior are similar to those made using frequentist unit root econometrics. Namely, the rejections of unit roots that result from flat prior Bayesian analysis are generally overturned when using the Jeffreys prior.

Whether or not the short rate actually has a unit root, its high degree of persistence makes concerns about the flat prior relevant for our analysis. I therefore consider an approximation of the Jeffreys class of priors as an alternative to flat priors.⁸ Again, I consider the case in which the prior belief contains no information about stationarity and the case in which parameter combinations that generate stationary or drift-stationary behavior are viewed as having zero prior one.

Without a prior belief about stationarity, the flat prior is given by

$$p_F(\alpha, \sigma, \gamma) \propto \frac{1}{\sigma}, \tag{11}$$

while the stationary flat prior is⁹

$$p_{FS}(\alpha, \sigma, \gamma) \propto \begin{cases} \frac{1}{\sigma} & \text{for } \alpha_2 < 0 \ \& \ \alpha_3 > 0, \text{ or } \gamma > 1.5 \ \& \ \alpha_3 > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

The flat prior that imposes drift-induced stationarity is

$$p_{FD}(\alpha, \sigma, \gamma) \propto \begin{cases} \frac{1}{\sigma} & \text{for } \alpha_2 < 0 \ \& \ \alpha_3 > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

The Jeffreys prior, as discussed in Appendix E, does not have a closed-form representation and must be computed by simulation. If we let p_J denote the Jeffreys prior that does not impose stationarity, then the corresponding stationary prior is given by

$$p_{JS}(\alpha, \sigma, \gamma) \propto \begin{cases} p_J(\alpha, \sigma, \gamma) & \text{for } \alpha_2 < 0 \ \& \ \alpha_3 > 0, \text{ or } \gamma > 1.5 \ \& \ \alpha_3 > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

The Jeffreys prior that imposes drift-induced stationarity is then

$$p_{JS}(\alpha, \sigma, \gamma) \propto \begin{cases} p_J(\alpha, \sigma, \gamma) & \text{for } \alpha_2 < 0 \ \& \ \alpha_3 > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

In all cases, σ must be positive.

⁸ The approximation arises because it is assumed in the derivation of the Jeffreys prior that the discretized process is observed continuously, so that data augmentation is not required in the computation of the prior. The implications of this assumption are discussed in Appendix E.

⁹ The case in which $\alpha_0 > 0$, $\alpha_1 < 0$, and $\alpha_2 = \alpha_3 = 0$ is also consistent with stationarity but has zero prior probability (since the prior contains no point masses) and can therefore be ignored. Assigning a point mass to $\alpha_2 = \alpha_3 = 0$ in the prior should be expected to tilt posteriors away from nonlinearity.

These “restricted” priors used to impose stationarity are particularly easy to work with. Following Box and Tiao (1973, p. 67–69), it can be shown that, in the region in which the restricted prior is nonzero, a posterior which incorporates a restricted prior is proportional to the corresponding posterior using an unrestricted prior. Where the prior probability is zero, so must be the posterior probability. This result suggests the simple approach of accept/reject as a way of drawing the parameters in the restricted case: draw the vector of parameters as if the prior were unrestricted and accept only those parameter vectors for which the stationarity restrictions hold.

3.3 Results

Markov chains were simulated to length 110,000 and the first 10,000 draws were discarded to negate the effects of initial conditions. To facilitate numerical computations only 1 out of every 10 iterations of the chain were saved, leaving 10,000 draws from the posterior distribution for each prior. A natural concern in any Markov chain Monte Carlo method is that the posterior draws are too highly autocorrelated, an indication that the chain may be slow to converge to its invariant distribution. The autocorrelation of the 10,000 draws saved is not high, however. In fact, the first-order autocorrelations of the drift parameter draws are nearly identically zero. The draws of σ and γ have first-order autocorrelations of about .5, declining to about .02 at the 10th lag, values that should not raise concerns about convergence.

Given the results in Appendix C, discretization bias is eliminated by setting h equal to .2 for all analysis with daily data and .05 for analysis with monthly data. Smaller values of h have no noticeable impact on any of the results.

Table 1 lists descriptive statistics on the posterior draws for the annualized parameters for both sampling frequencies and each of the six priors. Specifically, I report the means, standard deviations, and 95% highest posterior intervals.¹⁰

Comparison of panels A and B reveals major differences between the parameter values implied by the daily and monthly data. First, the drift parameter posterior means are significantly closer to zero for the monthly data than they are for the daily data. Surprisingly, the monthly data generate much lower standard deviations for the drift parameter posteriors than do the daily data.

The obvious cause of this difference is the much higher annualized volatility of the daily Eurodollar rates. Using the posterior mean values of σ and γ obtained under the flat prior, I compute and plot the time series

¹⁰ For unimodal distributions such as these, the 95% highest posterior density (HPD) interval is calculated numerically as the shortest interval that contains at least 95% of the posterior draws of a given parameter. It may be interpreted as a Bayesian confidence interval.

Table 1
Summary statistics for nonlinear drift model posteriors

	Flat prior	Stationary flat prior	Drift-stationary flat prior	Jeffreys prior	Stationary Jeffreys prior	Drift-stationary Jeffreys prior
Panel A: Daily Data						
Posterior means						
$\alpha_0 \times 10$	-3.62	-4.14	-4.14	0.75	-0.66	-0.66
$\alpha_1 \times 10$	6.91	7.66	7.66	0.43	2.83	2.83
$\alpha_2 \times 10^{-1}$	-3.74	-4.05	-4.05	-0.96	-2.04	-2.04
$\alpha_3 \times 10^3$	6.40	7.40	7.40	-2.21	0.08	0.08
σ	1.55	1.55	1.55	1.56	1.62	1.62
γ	1.36	1.36	1.36	1.36	1.38	1.38
Posterior standard deviations						
$\alpha_0 \times 10$	2.60	2.19	2.19	1.34	0.37	0.37
α_1	3.95	3.38	3.38	2.36	1.16	1.16
$\alpha_2 \times 10^1$	1.77	1.55	1.55	1.23	0.80	0.80
$\alpha_3 \times 10^3$	4.98	4.17	4.17	2.29	0.29	0.29
σ	0.09	0.09	0.09	0.08	0.07	0.07
γ	0.02	0.02	0.02	0.02	0.02	0.02
Posterior 95% HPD intervals						
$\alpha_0 \times 10$	(-8.51, 1.61)	(-8.19, -0.18)	(-8.19, -0.18)	(-1.18, 3.62)	(-1.18, -0.17)	(-1.18, -0.17)
α_1	(-1.13, 14.27)	(1.59, 14.23)	(1.59, 14.23)	(-4.21, 4.56)	(0.99, 4.50)	(0.99, 4.50)
$\alpha_2 \times 10^1$	(-7.35, -0.44)	(-7.07, -1.16)	(-7.07, -1.16)	(-3.09, 1.51)	(-3.01, -0.70)	(-3.01, -0.70)
$\alpha_3 \times 10^3$	(-3.71, 15.71)	(0.00, 14.79)	(0.00, 14.79)	(-7.12, 0.14)	(0.00, 0.17)	(0.00, 0.17)
σ	(1.39, 1.72)	(1.40, 1.73)	(1.40, 1.73)	(1.40, 1.71)	(1.49, 1.71)	(1.49, 1.71)
γ	(1.32, 1.40)	(1.32, 1.40)	(1.32, 1.40)	(1.32, 1.40)	(1.34, 1.40)	(1.34, 1.40)

Table 1
(continued)

	Flat prior	Stationary flat prior	Drift-stationary flat prior	Jeffreys prior	Stationary Jeffreys prior	Drift-stationary Jeffreys prior
Panel B: Monthly Data						
Posterior Means						
$\alpha_0 \times 10$	-1.14	-1.56	-1.58	0.31	-0.16	-0.15
α_1	2.11	2.75	2.80	-0.29	0.46	0.48
$\alpha_2 \times 10^{-1}$	-1.11	-1.38	-1.41	0.05	-0.31	-0.31
$\alpha_3 \times 10^3$	1.95	2.73	2.77	-0.77	0.02	0.01
σ	1.49	1.50	1.50	1.63	1.84	1.84
γ	1.63	1.64	1.64	1.67	1.72	1.72
Posterior standard deviations						
$\alpha_0 \times 10$	1.23	0.95	0.94	0.47	0.08	0.07
α_1	1.95	1.54	1.51	0.89	0.29	0.28
$\alpha_2 \times 10^{-1}$	0.92	0.75	0.74	0.53	0.28	0.27
$\alpha_3 \times 10^3$	2.28	1.75	1.74	0.79	0.03	0.02
σ	0.33	0.33	0.33	0.34	0.23	0.23
γ	0.08	0.08	0.08	0.08	0.05	0.05
Posterior 95% HPD intervals						
$\alpha_0 \times 10$	(-3.67, 1.17)	(-3.41, -0.02)	(-3.41, -0.07)	(-0.27, 1.25)	(-0.32, -0.10)	(-0.31, -0.09)
α_1	(-1.62, 6.05)	(0.09, 5.81)	(0.19, 5.69)	(-2.10, 1.06)	(-0.22, 1.16)	(0.22, 1.17)
$\alpha_2 \times 10^{-1}$	(-2.96, 0.65)	(-2.92, -0.02)	(-2.82, -0.10)	(-0.81, 1.18)	(-0.84, -0.03)	(-0.84, -0.03)
$\alpha_3 \times 10^3$	(-2.37, 6.61)	(0.00, 5.97)	(0.00, 5.98)	(-2.15, 0.04)	(0.00, 0.03)	(0.00, 0.03)
σ	(0.87, 2.14)	(0.85, 2.13)	(0.85, 2.13)	(1.07, 2.25)	(1.27, 2.17)	(1.27, 2.17)
γ	(1.48, 1.81)	(1.48, 1.80)	(1.48, 1.81)	(1.52, 1.81)	(1.60, 1.80)	(1.60, 1.79)

The table reports means, standard deviations, and 95% highest posterior density intervals (the shortest interval containing 95% of all posterior mass) for each of the six parameters of the model

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 / r_t) dt + \sigma r_t^\gamma dB_t.$$

Posterior distributions are generated by data augmentation using $h = .2$ for daily data and $h = .05$ for monthly data. The dataset consists of all monthly seven-day Eurodollar rates recorded from June 1973 to February 1995.

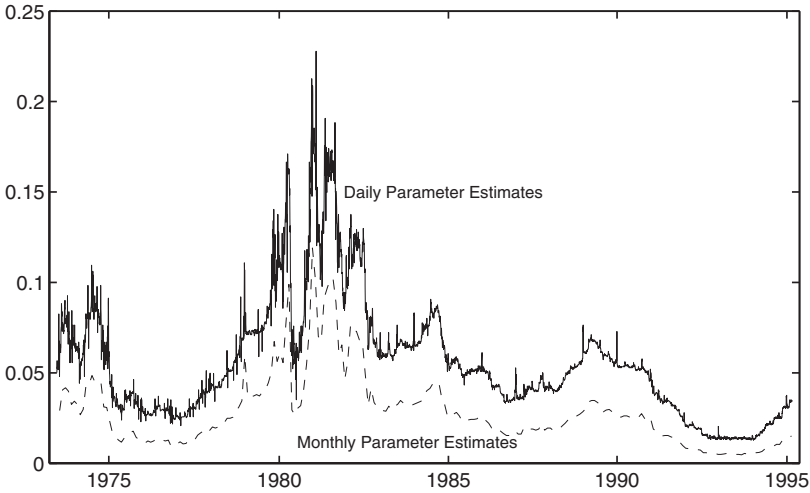


Figure 4
Model-implied Eurodollar volatility

The solid line represents the daily time series of spot rate volatilities implied by the nonlinear drift model using posterior means of σ and γ computed using daily data. The dashed line shows the corresponding monthly time series computed using monthly posterior means.

of σr_t^γ . Figure 4 compares the resulting annualized volatility paths that result from using daily and monthly posterior means. The differences are striking, with daily data implying an average annualized volatility of about 5.6%, compared with just 2.7% implied by the monthly data.

In addition to simply being higher overall, volatility in the daily data is less level dependent than it is for monthly data. For daily rates, the posteriors of γ for different prior distributions are tight around means between 1.35 and 1.4, slightly lower than the values reported by Chan et al. That monthly rates imply a somewhat higher γ is consistent with the presence of transitory noise that is less level dependent.

One possibility is that this noise is simply a product of a bid-ask effect or the existence of a discrete grid on which rates or prices are quoted. A preliminary version of this article calculated that such a grid would have to be fairly coarse for this to be a plausible explanation. A quick calculation yields a similar result: Suppose the observed interest rate, r_t , is the sum of some “true” unobserved rate, r_t^* , and an *i.i.d.* error, η_t , that is normally distributed with mean zero. The variance of the change in observed rates is therefore equal to

$$\text{var}(\Delta r_t^*) + 2\text{var}(\eta_t). \tag{16}$$

As the sampling frequency decreases, the first term comes to dominate the overall variance, making the observation error η_t irrelevant. For higher-frequency data, however, this term should be more important.

Rough calculations reveal that raising the annualized volatility from 2.7% to 5.6% as the sampling frequency increases from once per month to once per day would require the standard deviation of η_t to be around 0.2 percentage points, which would seem to be a large amount in the liquid Eurodollar market.

While the choice of prior has little impact on the posteriors of the variance parameters σ and γ , the prior has a major effect on the posterior means of the drift parameters $(\alpha_0, \alpha_1, \alpha_2, \alpha_3)$. Posterior standard deviations are also affected by the prior, with larger differences in the monthly results. In general, the flat prior results in posteriors for the drift parameters that are further away from zero than those of the Jeffreys prior, although with somewhat higher standard deviations.

Because they impose sign restrictions, it is not surprising that the priors that impose stationarity result in posteriors that are more conclusive about the signs of the drift parameters. Nevertheless, for both sampling frequencies and for each prior, the large dispersion of the posteriors often makes inferences about the exact magnitudes of individual parameters difficult, especially the parameters of the drift function.

Because the multivariate posterior distribution exhibits strong correlations, sometimes above .95 in absolute value, looking at marginal posteriors may understate the informativeness of the joint posterior. In addition, the parameters $(\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ have, individually, little economic interpretation, so a more illuminating viewpoint of the posterior is desirable. A natural quantity of interest is the drift function itself,

$$\mu(r) = \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3 / r, \quad (17)$$

evaluated over a variety of values of r . Using the parameter draws of the Markov chain, a posterior of the drift function can be evaluated for the range of r observed in the data (2.9% to 24.3%). Figures 5 and 6 show the medians and 95% HPD confidence intervals (dashed lines) of these distributions for each prior and sampling frequency.

Panel A of Figure 5, for example, reveals a pattern of nonlinear mean reversion similar to that reported in previous studies. Little positive or negative drift is found for rates between 3% and 15%, while very strong negative drift is found for higher rates. The magnitude of the effect is striking. When the short rate is at 20%, its posterior median drift is -45% per year. Even the upper bound of the 95% confidence interval is about -10% per year.

In comparison, the same data, when analyzed under the Jeffreys prior, produces much weaker evidence for nonlinear drift. Panel B of Figure 5 shows that the drift posterior computed under the Jeffreys prior has substantial mass above zero even for interest rates above 15%.

Given that the flat prior analysis suggests highly stationary parameter values, imposing stationarity does not substantially affect any results, as is

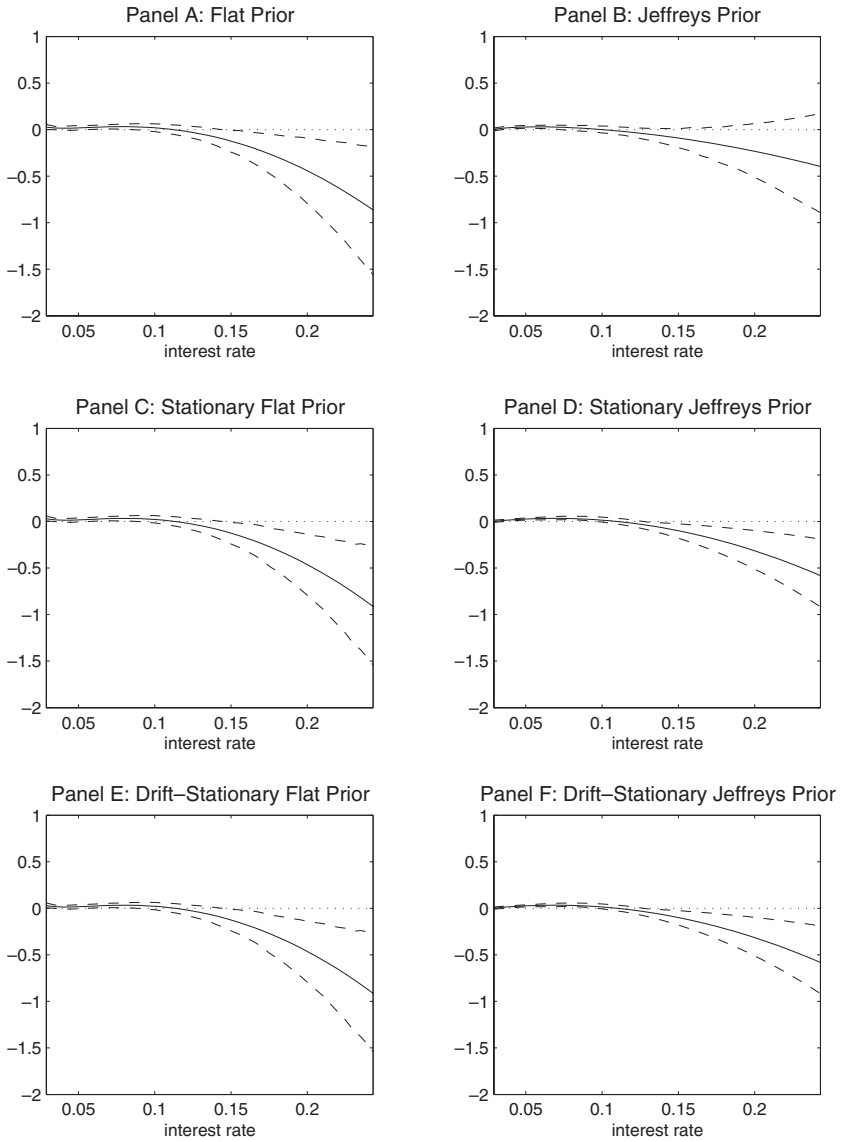


Figure 5
Interest rate drift posteriors for daily data

The figure reports posterior means and 95% highest posterior density intervals (the shortest interval containing 95% of all posterior mass) for the drift function,

$$\alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3 / r,$$

evaluated over the range of r observed in the sample. All posteriors were estimated using daily Eurodollar data from June 1, 1973, to February 25, 1995.

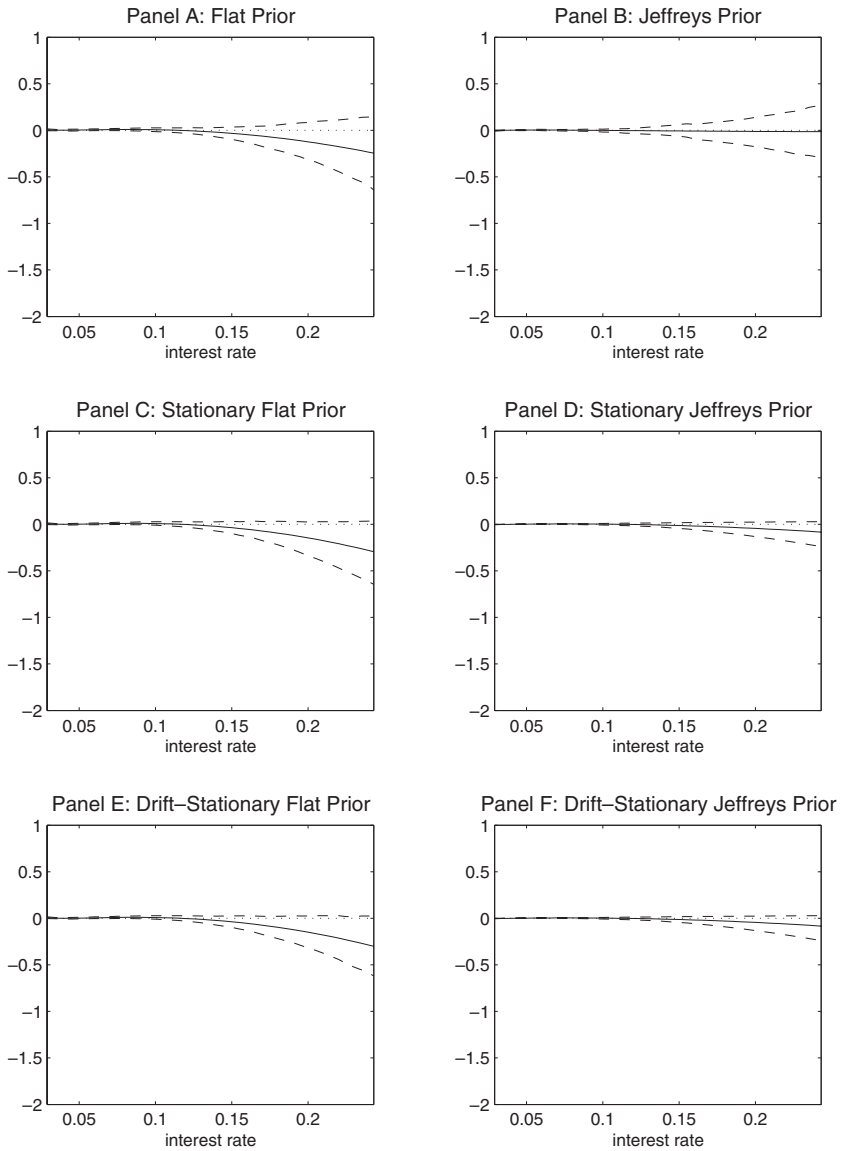


Figure 6

Interest rate drift posteriors for monthly data

The figure reports posterior means and 95% highest posterior density intervals (the shortest interval containing 95% of all posterior mass) for the drift function,

$$\alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3 / r,$$

evaluated over the range of r observed in the sample. All posteriors were estimated using month-end Eurodollar data from June 1973 to February 1995.

apparent in Table 1 and panels C and E of Figure 5. Under the Jeffreys prior, however, stationarity is no longer as obvious, so a prior belief that imposes stationarity has a significant impact. In panels D and F, we see that nonlinear drift is restored even under the Jeffreys prior.

Given the form of the parameter restrictions imposed by drift-induced stationarity, the nonlinearity found in panel F is not totally unexpected, since the restriction that $\alpha_2 > 0$ ensures a negative drift for sufficiently high levels of the interest rate. What is interesting is that this negative drift is inferred for values of r that are not too extreme, with reliably negative drifts for interest rates as low as 15%. Because the posterior distributions of γ lie below 1.5, stationarity must be induced by the drift rather than the volatility of interest rates. Therefore there is little difference between the results generated by the two different types of stationarity restriction.

Comparing the daily results of Figure 5 with the monthly results of Figure 6 reveals a relation similar to that found in the parameter estimates themselves: nonlinear mean reversion appears much stronger in daily data than it does in monthly data, despite the fact that confidence intervals are larger for daily data. As measured by the width of the 95% HPD intervals, the monthly data are actually more informative, and they suggest that nonlinear drift, if it exists, is not as large as one would conclude after looking only at higher-frequency data.

As with daily data, monthly data support more nonlinear drift more strongly under the flat prior than the Jeffreys prior. In fact, panel B of Figure 6 shows that monthly data provide no evidence of any drift when viewed under the Jeffreys prior, generating a drift posterior that is almost perfectly centered around zero. When a stationarity restriction is added to either prior, whether that stationarity is drift induced or volatility induced, nonlinear drift is again observed, but with a magnitude far below that implied by daily data.

These results suggest that the finding of nonlinear drift is highly dependent on the choice of the sampling frequency, the type of prior—flat or Jeffreys—and the prior belief about whether interest rates are stationary. Only for daily data under a flat prior can this negative drift in high interest rates be inferred without imposing stationarity.

For both daily and monthly data, discretization bias is evident when comparing the above results with those generated under the naive discretization ($h = 1$). For daily data analyzed under the flat prior, for example, the posterior mean of γ rises from 1.31 when $h = 1$ to 1.36 when $h = .2$, a movement of more than two posterior standard deviations. Reducing h even further to .05, however, does not further change this mean. For monthly data, the mean of γ under the flat prior rises from 1.56 with $h = 1$ to 1.63 with $h = .05$, and then rises slightly to 1.64 as h is decreased further to .01.

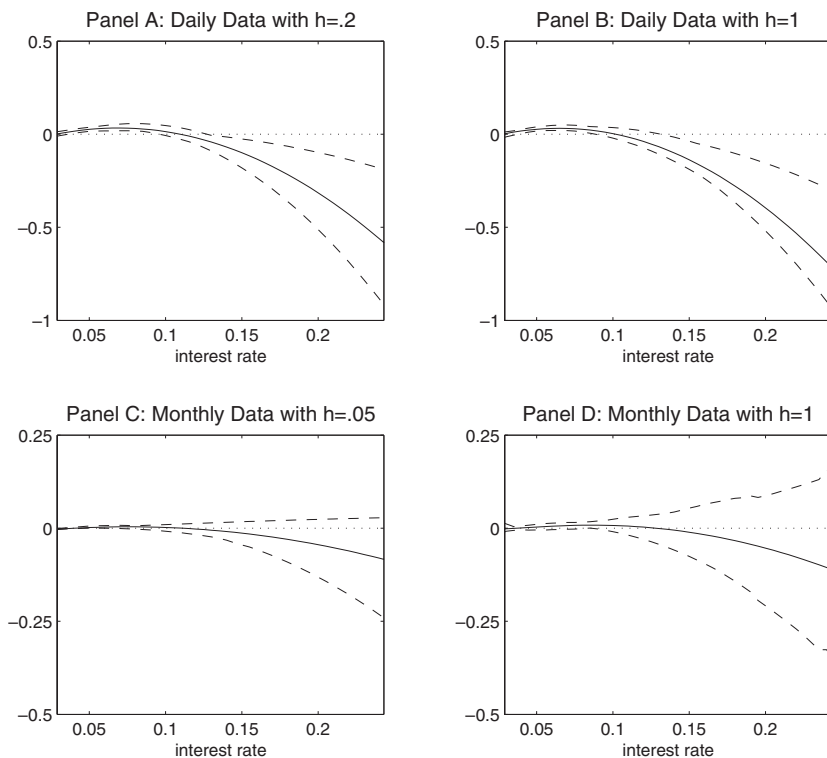


Figure 7
Discretization bias in interest rate drift posteriors

The figure reports posterior means and 95% highest posterior density intervals (the shortest interval containing 95% of all posterior mass) for the drift function,

$$\alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3 / r,$$

computed using different values of the discretization parameter h . For daily data, the results for $h = .2$ correspond to the results reported previously for the flat prior. For monthly data, the $h = .05$ results are identical to the previous flat prior results. Choosing $h = 1$ sets the Euler approximation frequency equal to the frequency of the data, making data augmentation unnecessary but inducing discretization bias.

Drift inferences change and discretization bias is reduced through data augmentation. Panels A and B of Figure 7 show the drift posteriors obtained using daily data under the stationary Jeffreys prior with h set either to 1 or to .2. While the differences are not large, the drift nonlinearity is slightly more severe with $h = 1$, and the posterior variance appears smaller as well. Differences are much more pronounced in monthly data, as evident in panels C and D, where a conclusion of drift nonlinearity appears to hinge on the value of h chosen, with smaller h now making nonlinear drift significantly more likely.

3.4 What belief does the flat prior represent?

Although the flat prior is intuitively appealing and has a natural interpretation as being similar to maximum likelihood, it cannot be justified formally as uninformative. As Phillips (1991a, 1991b) argued for the first-order autoregressive model, the flat prior for the nonlinear drift model is likely to represent an informed belief about the probabilities of different parameter vectors that are near the boundaries of the stationary parameter space.

A natural question is whether a bias like that found in the simple AR(1) model might appear in the more complicated model considered here. I will then ask how results generated under the Jeffreys prior should be expected to differ.

Suppose interest rates are generated according to the linear drift model

$$dr_t = (\alpha_0 + \alpha_1 r_t)dt + \sigma r_t^\gamma dB_t \tag{18}$$

but when we estimate the model we include the nonlinear drift parameters as well:

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 / r_t)dt + \sigma r_t^\gamma dB_t. \tag{19}$$

What will be the sampling distribution of the posterior mean of the vector $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$?

While sampling distributions are of obvious interest to the frequentist econometrician, they are useful to the Bayesian as well. Because sampling distributions are known a priori, they are revealing about the properties of the prior. In particular, biases can be interpreted as evidence of a prior that is not completely uninformative.

The Monte Carlo experiment performed is designed to capture some characteristics of the daily sample of Eurodollar data. One thousand 5505-day samples were simulated under the parameter values $\alpha_0 = .0072$, $\alpha_1 = -.12$, $\sigma = 1.55$, and $\gamma = 1.36$. While the values of σ and γ are equal to their posterior means from Table 1, α_0 and α_1 are chosen to generate a highly persistent process that slowly reverts to a long-run mean of 6%.¹¹ Because discretization bias appears to be negligible for daily data, the process was both simulated and estimated with $h = 1$, meaning that no data augmentation was used.

Posterior distributions were computed under both the flat and Jeffreys priors used above. With posterior means chosen as point estimates, the top half of Table 2 contains bias and root mean squared error summaries for the six parameters of the model. The bottom half addresses the frequencies with which the true parameters are within the top 5% or

¹¹ The properties of the linear drift function, $.0072 - .12 r_t$, become more apparent by rewriting the function as $.12 (.06 - r_t)$.

bottom 5% of the posterior distributions, where values near 5% are clearly desirable for each.

While the volatility parameters are precisely estimated under both priors, the results show substantial bias under the flat prior for all four parameters of the drift. Using the Jeffreys prior results in biases that are uniformly smaller, in some cases by wide margins. For instance, under the flat prior the parameter α_2 is on average estimated to be equal to -25.48 , even though its true value is zero. The Jeffreys prior results are much better behaved, with a bias of just -1.11 for the same parameter. Root mean squared errors are also much lower under the Jeffreys prior, generally around half of their values under the flat prior.

It is also interesting to look at the frequencies with which the true parameter values lie in the tails of the posterior distributions. In this dimension, both priors exhibit difficulties. Ideally an uninformative prior would have the property that true parameter value would be contained in the upper 5% of the posterior mass in approximately 5% of the Monte Carlo samples. Table 2 shows, however, that in 1000 Monte Carlo samples, the true value of α_3 , zero, was *never* in the upper in the upper tail of the posteriors computed under the flat prior, while it was in 60% of the upper tails using the Jeffreys prior. Less extreme but still problematic results are obtained for other parameters. Neither prior therefore adequately represents a completely uninformed view.

More important is how these biases are translated into biases about the drift as a whole. Following the procedure in Section 3.3, I compute

Table 2
Monte Carlo simulation results

	α_0	α_1	α_2	α_3	σ	γ
True parameters	0.0072	-0.12	0	0	1.55	1.36
	Bias					
Flat prior	-0.0979	2.740	-25.48	0.00117	0.0136	0.00124
Jeffreys prior	0.0288	-0.452	-1.11	-0.00042	0.0178	0.00197
	Root mean squared error					
Flat prior	0.1406	4.049	41.66	0.00185	0.1057	0.01981
Jeffreys prior	0.0759	1.818	15.67	0.00108	0.1105	0.02053
	Probability that true parameter is in top 5% of posterior					
Flat prior	0.277	0.003	0.246	0.000	0.053	0.055
Jeffreys prior	0.035	0.096	0.047	0.607	0.089	0.093
	Probability that true parameter is in bottom 5% of posterior					
Flat prior	0.000	0.269	0.001	0.284	0.074	0.070
Jeffreys prior	0.174	0.051	0.048	0.031	0.124	0.119

The table reports results from the Monte Carlo simulation of the nonlinear drift model

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 / r_t) dt + \sigma r_t^\gamma dB_t$$

under parameter values that produce linear drift. Bias is defined as the average difference between posterior means and true parameter values, with root mean squared error defined similarly. The bottom two panels report how often the true parameters fall in the upper and lower tails of the posterior distributions.

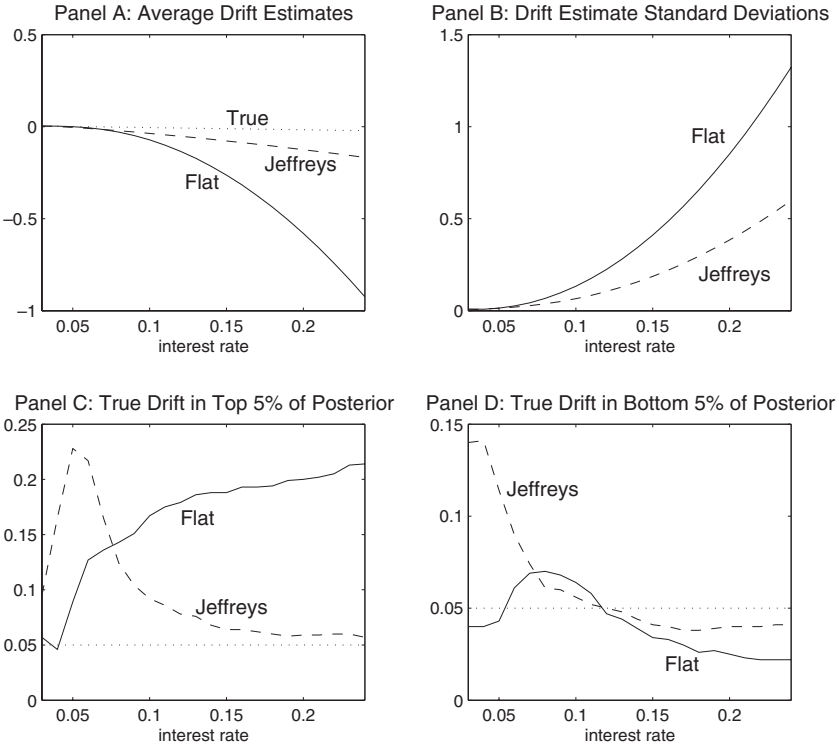


Figure 8
Monte Carlo distributions of the estimated drift

The figure summarizes the results of 1000 Monte Carlo simulations of interest rate paths generated by a model with linear drift. For each simulated sample, posterior distributions of the parameters of the nonlinear drift model,

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 / r_t) dt + \sigma r_t^\gamma dB_t$$

were computed under both the flat and Jeffreys priors. Panel A plots the average of the posterior means of the drift function in addition to the true drift function (the dotted line), while panel B summarizes the standard deviation of this estimator as a function of the level of the interest rate. Panels C and D report how often the true drift falls in the upper and lower 5% of the posterior distributions.

a posterior mean for the drift function for each of the Monte Carlo samples. The average of the drifts computed under each prior, as well as the true drift, are plotted in panel A of Figure 8.

The graph reveals that the biases apparent in the elements of α under the flat prior generate strong biases toward nonlinear drift. Furthermore, the magnitudes of the nonlinear drift typically estimated using the flat prior are not unlike those estimated by Ait-Sahalia (1996b), CHLS (1997), and Stanton (1997), as well as in the current article. Panel A also shows that while the Jeffreys prior does not completely eliminate this sort of bias, it reduces it considerably. Panel B shows that the standard deviation of the Jeffreys prior “estimator” is less than half of that of the flat prior.

Panels C and D report the frequency with which the true drift falls in the upper 5% and lower 5% of the posterior distribution, respectively. Ideally, if a prior is truly uninformative these frequencies should each be close to 5%. Unfortunately the figure shows that both can be far from that value for both priors.

For the flat prior, panel C shows that the true drift for interest rates above 10% is in the posterior distribution's upper tail in 15% to 22% of all samples. Meanwhile, the probability that the true drift is in the lower tail of the posterior distribution is too low for the flat prior. Furthermore, summing these frequencies reveals that the true drift, for high interest rates, is in the middle 90% of the posterior distribution less than 80% of the time. The holder of a flat prior, in addition to exhibiting bias, therefore shows a tendency to be overly confident in his conclusions.¹²

Panels C and D show that drift posteriors computed under the Jeffreys prior are comparatively well behaved for high interest rates, but that they have deficiencies at low to moderate rates. Specifically, the frequencies with which the true drift lies in the upper and lower tails of the posterior distribution are both far too high. At an interest rate of 5%, for instance, there is roughly a one in three chance that the true drift will lie outside the middle 90% of the posterior. Since both the bias and estimator standard deviations are very small in this region, the result can only be explained by the Jeffreys prior generating inferences that are overly sharp. When using the Jeffreys prior, inferences that small but significant positive or negative drift exists in low to moderate rates should therefore be discounted.

Before looking at the data, the holder of a flat prior expects to conclude in favor of the existence of nonlinear drift even when it is not a true feature of the data. As in the autoregressive model, the flat prior therefore represents an informative prior belief that the model is stationary. In particular, the flat prior in this case corresponds to a belief that the drift is nonlinear.

We can find some intuition for the directions of these biases in an analogy with linear time-series models. In the case of the AR(1), finite sample bias tends to make the process appear more mean reverting than it actually is, with the magnitude of this bias decreasing as the sample size grows [see, e.g., Marriott and Pope (1954)]. Since drift nonlinearity is a feature of the tails of the empirical distribution of short rates, the parameters that determine the degree of nonlinearity in the model, α_2 and α_3 , are effectively estimated with less data than the parameters α_0 and α_1 , which substantially affect the drift of the short rate throughout that

¹² Since flat prior analysis generates conclusions similar to maximum-likelihood estimation, a natural question from the frequentist perspective is how often the null hypothesis that $\alpha_2 = \alpha_3 = 0$ is rejected. Likelihood ratio statistics computed from the 1000 Monte Carlo draws show that the probability of rejecting at the 95% level is 9.3%, while the probability of rejecting at the 90% level is 17%.

distribution. As with the AR(1), finite sample biases lead us to find spurious mean reversion, but with biases most severe in the nonlinear parameters α_2 and α_3 , we also incorrectly characterize this mean reversion as nonlinear.

These results provide a natural interpretation of the drift posteriors graphed in Figures 5 and 6. In panels A and C of Figure 5, we saw that adding a belief in stationarity to the flat prior resulted in few changes. The Monte Carlo exercise suggests that this is because the flat prior is already informative about stationarity. The same effect is present in panels A and B of Figure 6, although not as strongly. The Jeffreys prior, meanwhile, represents less of a belief in stationarity, so the addition of this information to the prior has large effects. In Figure 6, panels B and D, for example, assuming stationarity leads to the conclusion that nonlinear drift is highly probable, even when no drift was evident without that assumption.

4. Specification Analysis and an Extension to the Model

The very different inferences drawn using daily and monthly data are compelling evidence for model misspecification, since for diffusion models, all sampling frequencies should generate similar parameter estimates, although possibly of differing precision. In this section I present additional evidence of model misspecification and explore an alternative model that reconciles some earlier results.

4.1 A specification check

A direct specification analysis may be performed by examining the normalized residuals that are generated in the estimation process at each step of the Markov chain. In the Euler approximation,

$$r_{(k+1)h} - r_{kh} = h\mu(r_{kh}, \phi) + \sqrt{h}\sigma(r_{kh}, \phi)\epsilon_k, \quad (20)$$

the normalized residuals ϵ_k are assumed to be independent standard normal random variables.

Following Zellner (1975), we may view ϵ as a parameter vector and compute the posterior distributions of various functions of it. For model diagnostic purposes, these functions should include moments and autocorrelations. Violations of either independence or normality is indicative of model misspecification.

Posterior distributions of these functions are obtained similarly to posteriors of the model parameters. At each iteration of the Markov chain, given the current draw of the parameter vector and the augmented data, the time series of ϵ_k may be calculated.¹³ The mean, standard

¹³ Results in this section are based on the flat prior, although other priors lead to similar conclusions.

Table 3
Specification analysis

Panel A: $r_{(k+1)h} - r_{kh} = h\mu(r_{kh}, \phi) + \sqrt{h}\sigma(r_{kh}, \phi)\epsilon_k$						
	Mean(ϵ_k)	StDev(ϵ_k)	Skew(ϵ_k)	Kurt(ϵ_k)	$\rho_1(\epsilon_k)$	$\rho_{1/h}(\epsilon_k)$
Daily data	0.0000 (0.0061)	1.0000 (0.0042)	0.0656 (0.0255)	4.7368 (0.4501)	-0.0169 (0.0060)	-0.0639 (0.0056)
Monthly data	-0.0002 (0.0138)	0.9999 (0.0096)	0.0013 (0.0339)	3.0036 (0.0678)	0.0002 (0.0137)	0.0025 (0.0135)
Panel B: $r_{(k+1)h} - r_{kh} = h\mu^r(r_{kh}, \theta_{kh}, \phi) + \sqrt{h}\sigma^r(r_{kh}, \theta_{kh}, \phi)\epsilon_k^r$ $\theta_{(k+1)h} - \theta_{kh} = h\mu^\theta(\theta_{kh}, \phi) + \sqrt{h}\sigma^\theta(\theta_{kh}, \phi)\epsilon_k^\theta$						
	Mean(ϵ_k^r)	StDev(ϵ_k^r)	Skew(ϵ_k^r)	Kurt(ϵ_k^r)	$\rho_1(\epsilon_k^r)$	$\rho_{1/h}(\epsilon_k^r)$
Daily data	-0.0003 (0.0061)	1.0000 (0.0042)	-0.0779 (0.0169)	3.3503 (0.0698)	-0.0016 (0.0060)	-0.0003 (0.0060)
Monthly data	0.0019 (0.0277)	0.9999 (0.0198)	-0.0002 (0.0678)	3.0059 (0.1370)	-0.0012 (0.0277)	-0.0022 (0.0279)
	Mean(ϵ_k^θ)	StDev(ϵ_k^θ)	Skew(ϵ_k^θ)	Kurt(ϵ_k^θ)	$\rho_1(\epsilon_k^\theta)$	$\rho_{1/h}(\epsilon_k^\theta)$
Daily data	-0.0001 (0.0060)	1.0001 (0.0043)	-0.0001 (0.0147)	3.0123 (0.0301)	0.0003 (0.0060)	0.0005 (0.0060)
Monthly data	-0.0000 (0.0278)	0.9998 (0.0198)	-0.0125 (0.0679)	3.0556 (0.1463)	-0.0038 (0.0278)	-0.0025 (0.0283)
	$\rho(\epsilon_k^r, \epsilon_k^\theta)$					
Daily data	0.0001 (0.0059)					
Monthly data	-0.0008 (0.0275)					

The table reports posterior means and standard deviations (in parentheses) of various moments of the residuals of the one- and two-factor models. A correct specification implies that the average residual, Mean(ϵ_k), should be zero. The residual standard deviations, StDev(ϵ_k), should be one, Skew(ϵ_k) should be zero and Kurt(ϵ_k) should be three (since it represents total rather than excess kurtosis). Within-period order autocorrelation, $\rho_1(\epsilon_k)$, between-period autocorrelation, $\rho_{1/h}(\epsilon_k)$, and cross-equation correlation, $\rho(\epsilon_k^r, \epsilon_k^\theta)$, should all equal zero.

deviation, skewness, and kurtosis of the ϵ vector are then computed for comparison with their theoretical values of 0, 1, 0, and 3, respectively. In addition, the first and 1/hth order autocorrelations are calculated to detect violations of independence, where the first-order autocorrelation primarily captures within-period dependence and the 1/hth order autocorrelation captures dependence between adjacent periods.

Panel A of Table 3 lists the posterior means and standard deviations of these functions of ϵ . For daily data, only the mean and standard deviation of the standardized residuals appear to be close to their theoretical values. Residuals exhibit positive skewness and pronounced excess kurtosis, and their autocorrelations appear to be negative, particularly between adjacent days. Taken together, these observations suggest there is a transient and fat-tailed component of interest rates that is not captured by the current model specification.

Results from monthly data reveal none of these problems, as the *i.i.d.* normal assumption appears to be well satisfied. This further supports the notion that the source of the model misspecification is a transient component that ceases to be relevant at a one-month horizon. The possible sources of such a component include bid-ask bounce and feedback from

the reserve requirement cycle effects in the Federal funds market identified by Hamilton (1996).

The existence of this noisy component of high-frequency interest rates casts strong doubt on the relevance of some of the previous results and those of the studies that use the same data. As the data come from Aït-Sahalia (1996b), the criticisms are relevant for this article in particular, but they are also applicable to some of the parametric analysis of Chapman and Pearson (2000), which also uses the daily Eurodollar data to estimate a nonlinear one-factor model.

4.2 A nonlinear stochastic mean model of interest rates

Durham (2002) also finds that interest rate drift nonlinearity is more associated with noisy interest rate data, and he has suggested that the apparent transitory component not currently captured by the model motivates the adoption of a stochastic mean model of interest rates. These models posit that interest rates are driven by a persistent process, but that rates deviate from this process in a random but highly transient way. Examples of stochastic mean models may be found in the articles by Andersen and Lund (1997), Balduzzi, Das, and Foresi (1996), Jegadeesh and Pennacchi (1996), and Piazzesi (2001), among others.

The stochastic mean model considered in this article,

$$dr_t = \kappa(\theta_t - r_t)dt + \xi\theta_t^\delta dB_t^{(1)} \tag{21}$$

$$d\theta_t = (\alpha_0 + \alpha_1\theta_t + \alpha_2\theta_t^2 + \alpha_3/\theta_t)dt + \sigma\theta_t^\gamma dB_t^{(2)}, \tag{22}$$

has the interest rate r_t mean revert to the stochastic mean process θ_t in a linear fashion, putting all drift nonlinearities in the stochastic mean equation.¹⁴ The volatility elasticity is allowed to differ between the two processes, since earlier results suggested that more transient dynamics may have a lower elasticity. As simplifying assumptions, both variances are assumed to depend on θ_t only, rather than on both θ_t and r_t , and the two Brownian motions are assumed to be independent.

The stochastic mean model is somewhat more difficult to estimate since the θ_t process is latent. Nevertheless, the econometric approach described previously and in Appendix B is easily extended to such models. Following this algorithm, parameter estimates were obtained using daily data by again setting $h = .2$ for daily data and $h = .05$ for monthly data.

Panel B of Table 3 reveals that the two-factor stochastic mean model shows much less evidence of misspecification than did the previous

¹⁴ Ideally, one might like to make the drift functions of both equations nonlinear. The decision to focus on the stochastic mean equation reflects a desire for parsimony and the literature’s focus on nonlinear mean reversion as a long-run phenomenon.

one-factor model. While the interest rate equation [Equation (21)] generates some excess kurtosis in its standardized residuals when estimated from daily data, it is far less than that reported for the original model. No violations of *i.i.d.* normality are apparent for the stochastic mean equation or for either equation when estimated with monthly data.

Parameter posterior statistics for the stochastic mean model estimated with daily data are reported in panel A of Table 4. Figure 9 contains the corresponding drift posterior graphs, where the drift shown is now the drift of the stochastic mean process, θ_t , rather than the interest rate. As before, a variety of priors are used, with the flat prior now given by $p(\kappa, \xi, \delta, \alpha, \sigma, \gamma) \propto 1/\xi\sigma$. Instead of deriving a new Jeffreys prior on the combined set of drift parameters κ and α , I use the approximate Jeffreys prior on α derived for the univariate process. While this is not the true Jeffreys prior for the two-factor model, it is the Jeffreys prior for the drift parameters in α conditional on the vector $(\kappa, \xi, \delta, \sigma, \gamma)$. The high precision of the posterior distributions of these parameters suggests that conditioning on these parameters should be relatively harmless.

Prior robustness is also checked by varying the prior belief about stationarity. Since it turns out that $\kappa > 0$ with posterior probability one for all prior distributions, stationarity of the joint process depends in practice solely on the conditions on α and γ described in Section 1.

Both Figure 9 and panel A of Table 4 reveal that the dynamics of the stochastic mean process estimated from daily data are almost identical to the dynamics of the original nonlinear interest rate process when estimated using *monthly* data. The transient dynamics identified earlier therefore appear to be well captured in the difference between the r_t and θ_t processes.

Given the unobservability of θ_t , estimating the stochastic mean model using 261 monthly observations should be imprecise, at best. In addition, the transitory nature of the deviations of r_t from θ_t induces a potential *aliasing* problem, since high-frequency dynamics or r_t should be difficult, if not impossible, to estimate using low-frequency data. Nevertheless, for completeness, parameter estimates obtained using monthly data are reported in panel B of Table 4. The corresponding drift plots are in Figure 10.

The table shows that there are large differences between the values of κ and ξ supported by daily and monthly data, although this is somewhat to be expected due to aliasing. In the frequency domain, it is known that cycles of higher frequency than that of the observed data will be incorrectly attributed to lower frequency cycles [see Hamilton (1994)]. Since the deviations of r_t from θ_t implied by daily data have half-lives well under one month, the aliasing problem is likely to be severe here.

The monthly parameter estimates of the stochastic mean process, however, capture a much lower frequency dynamic and are almost identical to

Table 4
Summary statistics for stochastic mean model posteriors

	Flat prior	Stationary flat prior	Drift-stationary flat prior	Jeffreys prior	Stationary Jeffreys prior	Drift-stationary Jeffreys prior
Panel A: Daily data						
Posterior means						
$\epsilon \times 10^{-3}$	3.15	3.14	3.14	3.13	3.03	3.03
ξ	1.66	1.66	1.66	1.65	1.69	1.69
δ	1.35	1.35	1.35	1.35	1.36	1.36
$\alpha_0 \times 10$	-1.08	-1.51	-1.54	0.34	-0.19	-0.19
α_1	2.01	2.68	2.73	-0.36	0.65	0.65
$\alpha_2 \times 10^{-1}$	-1.06	-1.35	-1.39	0.10	-0.48	-0.48
$\alpha_3 \times 10^3$	1.83	2.63	2.68	-0.80	0.01	0.01
σ	1.62	1.62	1.62	1.66	1.63	1.63
γ	1.67	1.68	1.68	1.69	1.69	1.69
Posterior standard deviations						
$\epsilon \times 10^{-3}$	0.23	0.23	0.23	0.23	0.16	0.16
ξ	0.12	0.12	0.12	0.13	0.09	0.09
δ	0.03	0.03	0.03	0.03	0.02	0.02
$\alpha_0 \times 10$	1.22	0.93	0.92	0.46	0.09	0.09
α_1	1.94	1.52	1.48	0.87	0.35	0.35
$\alpha_2 \times 10^{-1}$	0.92	0.75	0.73	0.53	0.32	0.32
$\alpha_3 \times 10^3$	2.25	1.71	1.70	0.77	0.05	0.05
σ	0.27	0.27	0.27	0.28	0.22	0.22
γ	0.06	0.06	0.06	0.06	0.06	0.06
Posterior 95% HPD intervals						
$\epsilon \times 10^{-3}$	(2.71, 3.63)	(2.68, 3.60)	(2.68, 3.60)	(2.68, 3.61)	(2.64, 3.34)	(2.64, 3.34)
ξ	(1.43, 1.90)	(1.43, 1.90)	(1.43, 1.90)	(1.41, 1.90)	(1.55, 1.90)	(1.55, 1.90)
δ	(1.29, 1.41)	(1.29, 1.41)	(1.29, 1.41)	(1.28, 1.40)	(1.33, 1.41)	(1.33, 1.41)
$\alpha_0 \times 10$	(-3.55, 1.23)	(-3.31, 0.02)	(-3.31, -0.06)	(-3.37, 1.23)	(-0.36, -0.09)	(-0.36, -0.09)
α_1	(-1.75, 5.84)	(0.12, 5.81)	(0.32, 5.73)	(-2.23, 1.07)	(0.38, 1.36)	(0.38, 1.36)
$\alpha_2 \times 10^{-1}$	(-2.87, 0.77)	(-2.94, -0.03)	(-2.72, -0.03)	(-0.83, 1.18)	(-1.10, -0.10)	(-1.10, -0.10)
$\alpha_3 \times 10^3$	(-2.54, 6.31)	(0.00, 5.78)	(0.00, 5.81)	(-2.39, 0.04)	(0.00, 0.04)	(0.00, 0.04)
σ	(1.12, 2.18)	(1.12, 2.18)	(1.12, 2.18)	(1.24, 2.31)	(1.20, 1.82)	(1.20, 1.82)
γ	(1.56, 1.81)	(1.56, 1.80)	(1.56, 1.80)	(1.59, 1.81)	(1.57, 1.74)	(1.57, 1.74)

Table 4
(continued)

	Flat prior	Stationary flat prior	Drift-stationary flat prior	Jeffreys prior	Stationary Jeffreys prior	Drift-stationary Jeffreys prior
Panel B: monthly data						
Posterior means						
$\kappa \times 10^{-3}$	0.29	0.29	0.29	0.27	0.21	0.19
ξ	0.45	0.44	0.44	0.46	0.58	0.58
δ	1.31	1.30	1.30	1.28	1.36	1.36
$\alpha_0 \times 10$	-1.27	-1.68	-1.71	0.29	-0.17	-0.19
α_1	2.32	2.95	3.01	-0.29	0.62	0.69
$\alpha_2 \times 10^{-1}$	-1.21	-1.49	-1.52	0.06	-0.49	-0.56
$\alpha_3 \times 10^3$	2.19	2.95	2.99	-0.70	0.03	0.04
σ	1.95	1.97	1.97	2.53	2.97	3.14
γ	1.70	1.71	1.71	1.82	1.89	1.92
Posterior standard deviations						
$\kappa \times 10^{-3}$	0.09	0.09	0.09	0.10	0.05	0.04
ξ	0.35	0.35	0.35	0.33	0.18	0.19
δ	0.26	0.26	0.26	0.24	0.13	0.13
$\alpha_0 \times 10$	1.30	1.03	1.02	0.48	0.16	0.16
α_1	2.09	1.70	1.66	0.90	0.33	0.28
$\alpha_2 \times 10^{-1}$	1.01	0.85	0.82	0.54	0.25	0.17
$\alpha_3 \times 10^3$	2.40	1.89	1.88	0.81	0.30	0.32
σ	0.69	0.69	0.69	0.83	0.74	0.57
γ	0.13	0.13	0.13	0.13	0.11	0.08
Posterior 95% HPD intervals						
$\kappa \times 10^{-3}$	(0.12, 0.47)	(0.12, 0.46)	(0.12, 0.46)	(0.13, 0.48)	(0.14, 0.30)	(0.14, 0.23)
ξ	(0.07, 1.15)	(0.07, 1.15)	(0.07, 1.15)	(0.08, 1.14)	(0.27, 0.74)	(0.27, 0.74)
δ	(0.94, 1.68)	(0.94, 1.68)	(0.94, 1.68)	(0.94, 1.65)	(1.13, 1.51)	(1.13, 1.48)
$\alpha_0 \times 10$	(-3.81, 1.27)	(-3.66, 0.02)	(-3.65, -0.05)	(-0.34, 1.36)	(-0.28, -0.04)	(-0.28, -0.12)
α_1	(-1.73, 6.47)	(0.12, 6.42)	(0.19, 6.20)	(-2.34, 1.16)	(0.04, 1.13)	(0.49, 1.13)
$\alpha_2 \times 10^{-1}$	(-3.19, 0.75)	(-3.20, 0.04)	(-3.04, -0.03)	(-0.82, 1.26)	(-0.96, 0.09)	(-0.96, -0.42)
$\alpha_3 \times 10^3$	(-2.55, 6.86)	(0.00, 6.40)	(0.00, 6.42)	(-2.52, 0.02)	(0.00, 0.01)	(0.00, 0.01)
σ	(0.86, 3.40)	(0.86, 3.40)	(0.86, 3.40)	(1.32, 3.85)	(1.43, 3.85)	(2.40, 3.85)
γ	(1.46, 1.96)	(1.46, 1.96)	(1.46, 1.96)	(1.59, 2.02)	(1.64, 2.02)	(1.84, 2.02)

The table reports means, standard deviations, and 95% highest posterior density intervals (the shortest interval containing 95% of all posterior mass) for each of the nine parameters of the model

$$dr_t = \kappa(\theta_t - r_t)dt + \xi\theta_t^{\delta}dB_t^{(1)}$$

$$d\theta_t = (\alpha_0 + \alpha_1\theta_t + \alpha_2\theta_t^2 + \alpha_3\theta_t^3)dt + \sigma\theta_t^{\gamma}dB_t^{(2)}$$

Posterior distributions are generated by data augmentation using $h = 2$ for daily data and $h = .05$ for monthly data. The data set consists of all monthly seven-day Eurodollar rates recorded from June 1973 to February 1995.

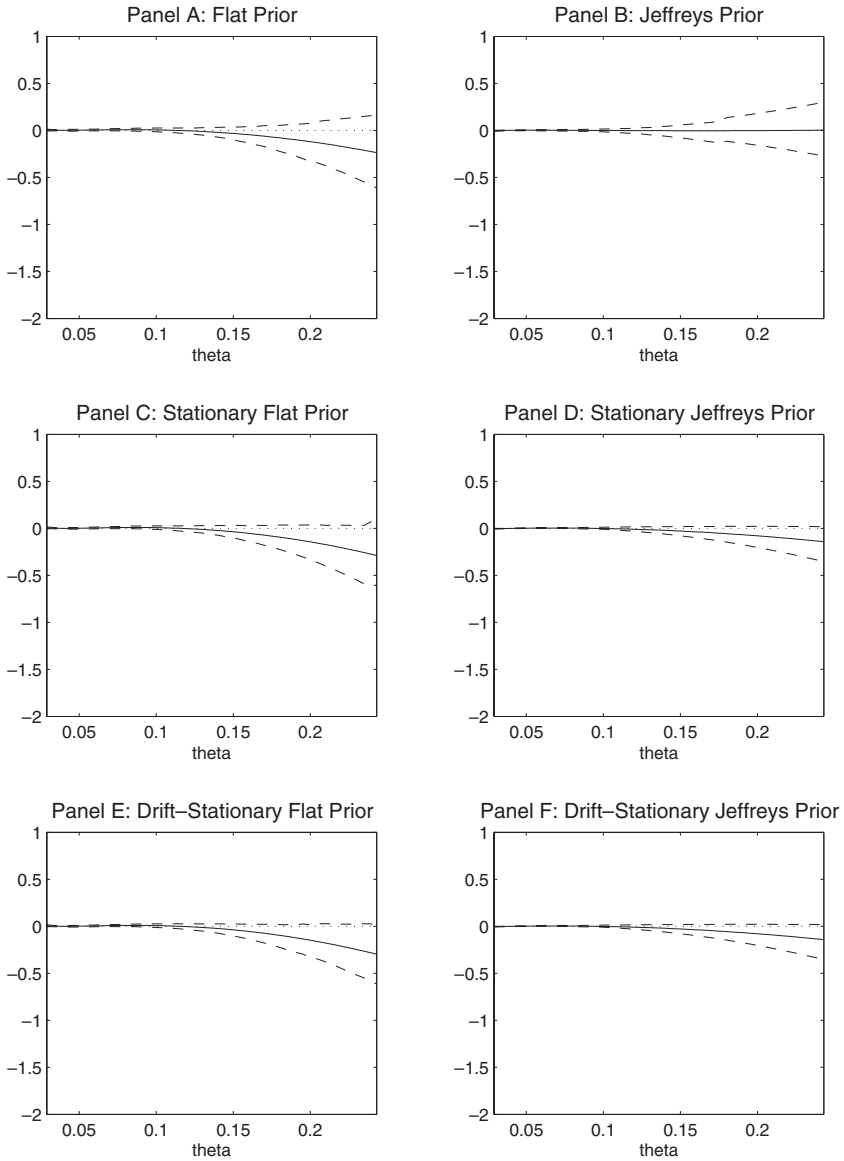


Figure 9
Stochastic mean drift posteriors for daily data

The figure reports posterior means and 95% highest posterior density intervals (the shortest interval containing 95% of all posterior mass) for the drift function of the stochastic mean process,

$$\alpha_0 + \alpha_1\theta + \alpha_2\theta^2 + \alpha_3/\theta.$$

All posteriors were estimated using daily Eurodollar data from June 1, 1973, to February 25, 1995.

the daily estimates. In addition, the precision of the posterior distributions of the drift parameters are almost identical to the precision obtained with daily data, suggesting again that high-frequency data have little information to add over monthly data about the shape of the drift function.

Figures 9 and 10 confirm earlier results that nonlinear drift is primarily a feature of a misspecified model of high-frequency data. While sufficient stationarity assumptions can be imposed through the prior to generate nonlinear drift, the magnitude of this nonlinearity is much less than that found under the original model using daily data. Using monthly data, conclusions are largely unaffected by the choice of model, as the short-term deviations from the stochastic mean process become irrelevant.

4.3 An economic specification test

Because one of the primary reasons for estimating the short-rate process is to be able to use that process to price other fixed-income securities, a natural evaluation of a model might therefore be based on how well the model describes the prices or price dynamics of these securities.

Specifically I consider whether the models and parameter estimates reported above are consistent with the observed volatility of three-month interest rates. While this maturity is relatively short, it is still substantially longer than the seven-day rates used to estimate the model.

Model bond prices are calculated under the local expectations hypothesis. Longstaff (2000) has argued that the expectations hypothesis is an accurate characterization of three-month repo rates, but may fail to hold for Treasury bills because of institutional demand for the high liquidity they provide. Duffee (1996) documents liquidity-driven volatility in Treasuries bills that appears absent from other short-term debt. Because repo rates are difficult to obtain over a long sample, I use Eurodollar loan rates instead, which should similarly be unaffected by liquidity effects. Although these Eurodollar rates contain a credit risk component, if this component is relatively smooth it should not influence the calculation of daily interest rate volatilities.¹⁵

Given a level, r_0 , of the current short rate, three-month bond prices $B(r_0)$ for the one-factor model are obtained by simulating 10,000 three-month paths of the short rate and then calculating the Monte Carlo estimate

$$B(r_0) = \frac{1}{10,000} \sum_{i=1}^{10,000} \exp\left(-\int_{t=0}^{.25} r_t dt\right), \quad (23)$$

where a trapezoidal approximation is used to compute the integral. The three-month interest rate is then computed as $R(r_0) = -(1/.25)\log B(r_0)$.

¹⁵ A constant term premia would not affect these calculations either.

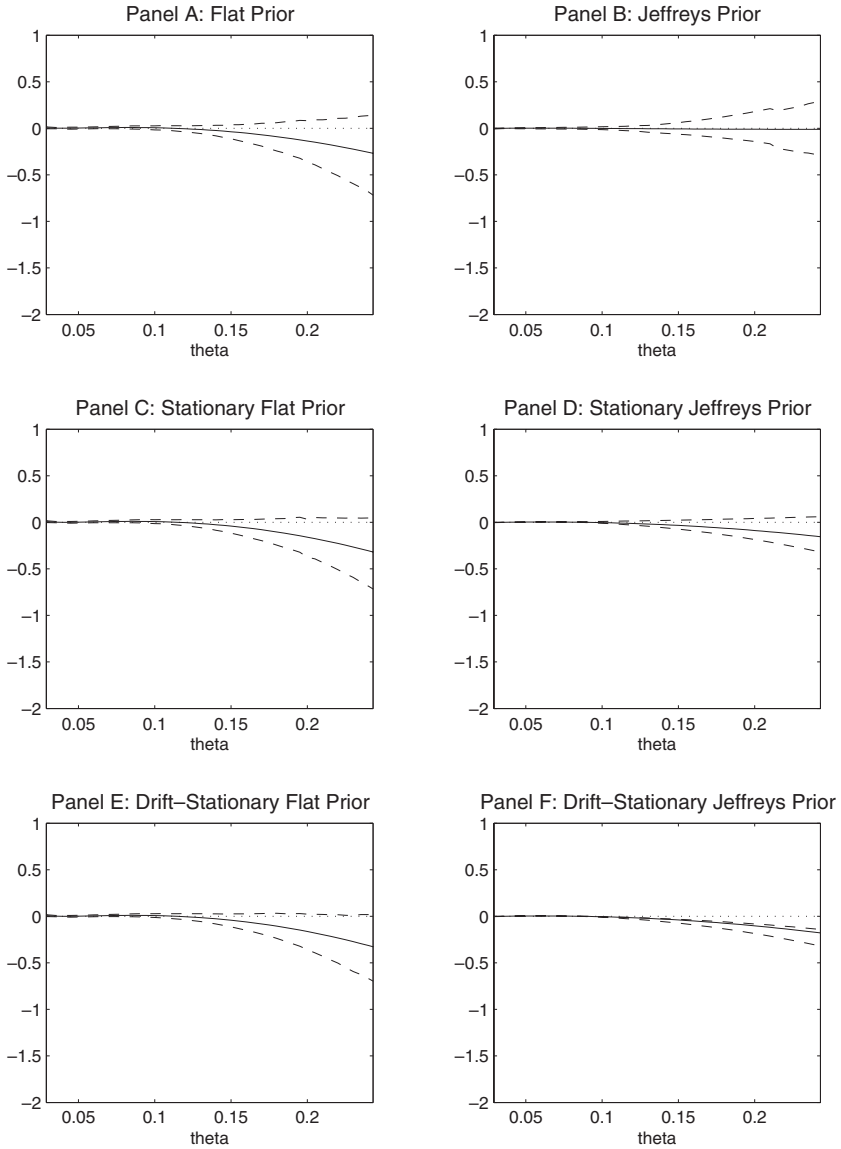


Figure 10

Stochastic mean drift posteriors for monthly data

The figure reports posterior means and 95% highest posterior density intervals (the shortest interval containing 95% of all posterior mass) for the drift function of the stochastic mean process,

$$\alpha_0 + \alpha_1\theta + \alpha_2\theta^2 + \alpha_3/\theta.$$

All posteriors were estimated using monthly Eurodollar data from June 1973 to February 1995.

Following Ito's lemma, the volatility of the three-month rate is then obtained by numerically differentiating $R(r_0)$ and multiplying this derivative by the instantaneous volatility of the short rate, yielding $(\partial R/\partial r_0)\sigma r_0^\gamma$.

Volatilities for the two-factor model are calculated similarly under the assumption that $r_0 = \theta_0$.¹⁶ As before, simulation is used to obtain $R(r_0, \theta_0)$, $R(r_0 + \epsilon, \theta_0)$, and $R(r_0, \theta_0 + \epsilon)$ for each value of r_0 . Together, these may be used to numerically calculate the partial derivatives $\partial R/\partial r_0$ and $\partial R/\partial \theta_0$. Because of the independence of the processes for r_t and θ_t , the three-month interest rate variance is given by

$$\left(\frac{\partial R}{\partial r_0}(r_0, \theta_0)\xi\theta_0^\delta\right)^2 + \left(\frac{\partial R}{\partial \theta_0}(r_0, \theta_0)\sigma\theta_0^\gamma\right)^2. \quad (24)$$

These calculations were performed using both models and both daily and monthly posterior distributions computed under the flat prior. From each posterior, 500 sets of model parameters were drawn at random to construct posterior distributions of the three-month rate's volatility as a function of its level. Figure 11 plots the mean of this distribution, along with its 5th and 95th, percentiles, as solid lines.

These model-implied volatilities are compared to a locally linear nonparametric regression estimate of the daily volatility of changes in the three-month rate. This regression estimate is calculated using the Federal Reserve's time series of three-month Eurodollar rates over the same time period used to estimate the models. Figure 11 plots these curves as dashed lines, along with the 5th and 95th percentiles for the nonparametric estimate calculated from 5000 draws of the Künsch (1989) block bootstrap with a block size of 100.¹⁷

While comparison of Bayesian posteriors to frequentist confidence intervals is somewhat informal, the top left panel shows clearly that the volatility in daily Eurodollar rates is largely absent from three-month Eurodollar rates, as the one-factor model fitted to daily data grossly overpredicts the level of volatility of this longer-maturity yield. When fitted to monthly data (top right panel), the model-implied volatilities come very close to matching the nonparametric estimates, implying again that transient movements in the short rate do not impact the three-month rate.

¹⁶ Since the deviations of r_t from θ_t are short-lived, the initial state of r_t is fairly unimportant.

¹⁷ Given the time series of $r_{3M}(t)$, I run a locally linear regression of realized squared changes $y(t) \equiv (r_{3M}(t) - r_{3M}(t-1))^2$ on a constant and $r_{3M}(t-1)$. In local linear regression, the fitted value at \bar{r}_{3M} is obtained as the intercept of the weighted least squares regression of $y(t)$ on a constant and $r_{3M}(t-1) - \bar{r}_{3M}$. The weights are proportional to the normal density with mean zero and standard deviation .0125 (the bandwidth parameter) evaluated at $r_{3M}(t-1) - \bar{r}_{3M}$. The Künsch (1989) bootstrap algorithm resamples the time series $r_{3M}(t)$ in blocks of consecutive observations in order to account for the persistence of the data. Local linear regression is applied to each resampled dataset to generate the 5th and 95th percentiles graphed.

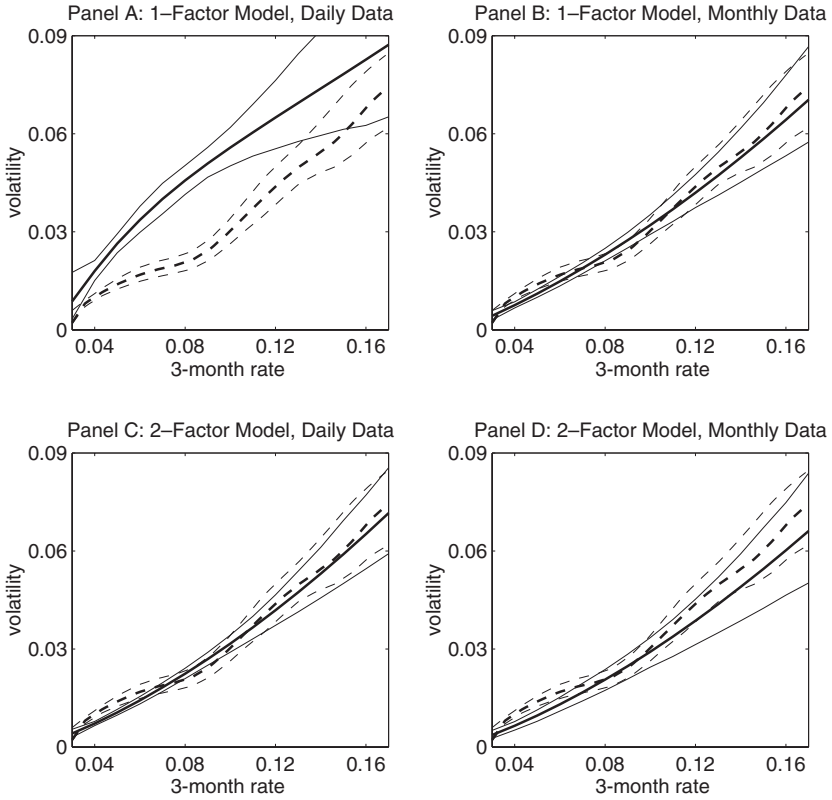


Figure 11
Nonparametric versus model-implied three-month interest rate volatility
Solid lines represent the mean, 5th, and 95th percentiles of the posterior distribution of the three-month rate's volatility, calculated under the model listed and assuming the local expectations hypothesis. The heavy dashed line depicts three-month volatilities calculated using a locally linear nonparametric regression, while other dashed lines show 5th and 95th percentiles from the bootstrap distribution of the nonparametric regression line.

The bottom panels contain results for the stochastic mean model. This model produces three-month volatilities that come very close to matching the nonparametric estimates regardless of what sampling interval is used to estimate the model, suggesting that it is better specified than the one-factor model.

Given the problems with transient noise in the seven-day Eurodollar rate, one might argue for using a different short-term rate, such as the Federal funds rate. Unfortunately other very short-term rates generate similar results, not reported here. The Federal funds rate and, to a lesser extent, the 30-day Eurodollar rate are both more volatile and have much stronger nonlinear mean reversion when sampled daily versus monthly. Another alternative would be to use a longer maturity rate, such as the

three-month Treasury bill rate, to proxy for the short rate. Chapman, Long, and Pearson (1999) argue, however, that the three-month yield is a poor substitute for the “instantaneous” rate of interest when the model under consideration is nonlinear, as is ours. Using the longer yield “can significantly affect both estimates of the diffusion function and discount bond prices.” The use of noisy short-maturity rates may therefore be unavoidable.

5. Conclusion

Taken together, the results of the article combine to suggest that objective evidence for nonlinear mean reversion in the short-term interest rate is weak. The conclusion that high interest rates exhibit strong negative drift is extremely sensitive to the choice of prior, even when the choice is made between priors that could all be defended as representing relatively uninformed views. Results are also sensitive to the sampling frequency and model, with daily data implying much stronger nonlinearities and a level of interest rate volatility almost twice that apparent in monthly data.

Results in Chapman and Pearson (2000) suggest that nonparametric methods are biased toward finding nonlinear mean reversion even when it is not present. This article establishes that fully efficient parametric inference (such as maximum likelihood) may be just as vulnerable to such false inferences. From a frequentist perspective, this vulnerability arises in the form of biases similar to those found for simpler linear time-series models. In the nonlinear drift model, however, these biases affected nonlinear terms most severely, often generating spurious nonlinearity.

From a Bayesian perspective, we may attribute the tendency to find spurious nonlinearity to the selection of an informative prior distribution, possibly one that does not accurately reflect the investigator’s actual prior belief. This view suggests that alternative priors be considered, and the article considered a number of variations. While it is impossible to say which prior is the “correct” one, several characteristics of the prior distributions are important to note.

- The flat prior effectively represents a prior belief that the drift function is nonlinear, with the same shape and possibly the same magnitude as the drift function that is estimated in the data. Similar to the AR(1) model, the posterior means of the drift parameters are biased in repeated samples. A flat prior, by not anticipating and correcting for this bias, is implicitly taking an informed view that this bias is desirable.
- The Jeffreys prior, which Phillips (1991b) argues is the best representation of true prior ignorance, suggests no evidence for nonlinear drift unless stationarity is imposed.

- Imposing stationarity in a prior distribution represents a nontrivial amount of prior information. While such a prior is not unreasonable, it must be recognized that conclusions drawn about nonlinear drift under this prior are not entirely data based. For the monthly interest rate sample and also for the two-factor stochastic mean model, a stationarity restriction was required to generate the conclusion that high rates have a negative drift.

It was also shown that changing the sampling frequency can result in very different inferences about both the drift and volatility of interest rates. Specifically, daily data appear to contain a volatile transitory component that is unreflected in longer-term dynamics or the volatilities of the three-month Eurodollar rate. A nonlinear stochastic mean model of interest rates appears to fit the data much better and suggests that it is the unmodeled transitory component of short rates that is largely responsible for the finding of nonlinear drift.

While many of the problems with high-frequency data could be avoided, without losing much sample information, by looking solely at month-end observations, the data augmentation procedure was crucial for eliminating discretization bias in these estimates. Under some priors, discretization bias was sufficiently severe to substantially change one's inferences about drift nonlinearity.

Although a definitive conclusion about the existence of nonlinear drift cannot be made solely by observing the short rate itself, there exists a variety of information in long-term yields and interest rate options that may be much more revealing than the short rate itself. While incorporating these data into the analysis of nonlinear drift remains a challenge, it is called for by the fact that although more than 5000 observations of daily data are available, the current data sample is effectively small. With these data alone, precise statements about the shape of the drift function—statements that different individuals with different prior beliefs can agree on—are impossible to make.

Appendix A: An Introduction to the Gibbs Sampler and Data Augmentation

The Gibbs sampler is motivated by the frequent need to draw from intractable multivariate distributions. For simplicity, consider the bivariate case in which we desire to draw from the distribution $p(\alpha, \beta|X)$, where X represents the observed data. In many cases the density $p(\alpha, \beta|X)$ is of an unknown form, while the conditional densities $p(\alpha|\beta, X)$ and $p(\beta|\alpha, X)$ are of standard forms.

A Gibbs sampling chain is formed as follows:

1. Choose some arbitrary value for α and label it α_0 .
2. Draw β_0 from the distribution $p(\beta|\alpha_0, X)$.
3. Draw α_1 from the distribution $p(\alpha|\beta_0, X)$.
4. Repeatedly draw β_n from $p(\beta|\alpha_n, X)$ and α_{n+1} from $p(\alpha|\beta_n, X)$.

Under very mild conditions, the pairs (α_n, β_n) converge in distribution to $p(\alpha, \beta|X)$. Posterior means, for example, may therefore be calculated by simulating a long chain of (α_n, β_n) , discarding the values at the beginning of the chain (the “burn-in period”), and then averaging the remaining draws.

A simple example of the usefulness of the Gibbs sampler is provided by the following discrete time version of Vasicek’s interest rate model:

$$r_t - r_{t-1} = \kappa(\mu - r_{t-1}) + \sigma\epsilon_t \tag{25}$$

While this model is linear in the data, it is nonlinear in parameters (because of the interaction of κ and μ), making standard linear regression analysis inapplicable.

Note, however, that were μ a known constant, then the equation would conform to the standard linear regression framework, with the quantity $\mu - r_{t-1}$ filling the role of the regressor. Under the flat prior $p(\kappa, \sigma) \propto 1/\sigma$, the posterior distribution of κ and σ is well known; κ is distributed as a student’s t and σ as an inverted gamma.

Similarly, if κ and σ were known, then Equation (25) could be rearranged as

$$r_t - (1 - \kappa)r_{t-1} = \kappa\mu + \sigma\epsilon_t. \tag{26}$$

With κ assumed known, the model is linear in its lone unknown parameter μ . Since σ is known, the flat prior $p(\mu) \propto 1$ implies that the posterior of μ is normal.

By alternately drawing from the conditional distributions $p(\kappa, \sigma|\mu, R)$ and $p(\mu|\kappa, \sigma, R)$, the Gibbs sampler may be used to obtain draws from the joint posterior, $p(\kappa, \sigma, \mu|R)$.¹⁸ Averaging these draws, for example, would produce an estimate of the posterior mean.

In principle, the Gibbs sampler may be used to draw from any distribution $p(\theta^1, \theta^2, \dots, \theta^k|X)$ in which the conditional distributions $p(\theta^j|\theta^i, j \neq i, X)$ are of standard forms. Furthermore, each parameter “block,” θ^j , may be uni- or multivariate. The Gibbs sampler may therefore be used to analyze very complex posteriors when decomposition into simpler conditionals is possible. A variety of examples may be found in Chib and Greenberg (1996).

A particularly powerful incarnation of the Gibbs sampler has been coined “data augmentation” by Tanner and Wong (1987). This approach is motivated by the fact that many posterior distributions could be calculated more easily if some unobserved variable was in the researcher’s dataset. Although the researcher does not observe this latent data, he may know (or be able to draw from) their distribution conditional on the observed data and the unobserved model parameters. The solution is to form a Gibbs sampling chain, alternately drawing from the conditional distribution of the model parameters given the observed and augmented data, and the conditional distribution of the augmented data given the real data and the model parameters.

Jacquier, Polson, and Rossi (1994) used this technique in a well-known analysis of stochastic volatility models. In this case, estimation of the price and volatility equations would be straightforward were volatility an observed variable, that is,

$$p(\text{parameters}|\text{prices, volatilities})$$

is a known density. Since volatility is latent, it must be integrated out to obtain the true posterior

$$p(\text{parameters}|\text{prices}).$$

This is accomplished using a Gibbs sampling chain that alternates between the two conditionals¹⁹

$$p(\text{parameters}|\text{prices, volatilities}) \quad \text{and} \quad p(\text{volatilities}|\text{parameters, prices}).$$

¹⁸ In this example it is easy to see that the “blockings” of the Gibbs sampler are often not unique. It would have been equally straightforward to alternate between $p(\kappa|\mu, \sigma, R)$ and $p(\mu, \sigma|\kappa, R)$.

¹⁹ The conditional distribution of the volatility paths turns out to be quite complicated, requiring the use of additional tools similar to the ones used in later sections of the current article. For purposes of illustration, I skip the details here.

In essence, the latent, or “augmented,” data are treated as a high-dimensional parameter vector. The data augmentation scheme therefore generates the joint posterior distribution of the parameters and the augmented data given the observed data. This makes it possible to construct marginal posteriors not only of the parameters, but for the latent variables as well. In applications such as stochastic volatility, this may be a very useful by-product of the estimation scheme.

Appendix B: Details of the Data Augmentation Procedure

Let X_t denote an L -dimensional diffusion process satisfying the stochastic differential equation

$$dX_t = \mu(X_t, \phi)dt + \sigma(X_t, \phi)dB_t, \quad (27)$$

where $\mu(x, \phi): \mathcal{R}^L \times \Phi \rightarrow \mathcal{R}^L$ and $\sigma(x, \phi): \mathcal{R}^L \times \Phi \rightarrow \mathcal{R}^L \times \mathcal{R}^D$ satisfy regularity conditions, B_t is a D -dimensional standard Brownian motion, and ϕ is a vector of parameters.

The Euler approximation of this model is given by

$$X_{(k+1)h} = X_{kh} + h\mu(X_{kh}, \phi) + \sqrt{h}\sigma(X_{kh}, \phi)\epsilon_{k+1}, \quad (28)$$

where $\epsilon_k \sim i.i.d. N(0, I_D)$, I_D is the D -dimensional identity matrix, and h is the discretization interval length. For brevity, we will write X_{kh} as X_k , making dependence on a particular value of h implicit.

As stated above, the approach followed in this article will be to estimate the discretized process of Equation (28) while allowing h to be arbitrarily small. If the discretization interval h is smaller than the frequency of the observed data, Tanner and Wong’s (1987) data augmentation algorithm will be used to augment the observed low-frequency data with unobserved high-frequency data.

Suppose the vector X_k represents the time kh realization of the L -dimensional process generated by the Euler approximation of Equation (28). Divide the L -dimensional vector X_k into subvectors, X_k^o and X_k^u , based on whether the realization of the component of the process at that time is observed (X_k^o) or unobserved (X_k^u). If kh is a noninteger, then X_k is completely unobserved, implying $X_k^u = X_k$ and $X_k^o = \phi$. In other cases, X_k may be partially observed, as in a stochastic volatility model, where a price may be observed while volatility remains latent.

To perform the data augmentation, the Markov chain cycles through all k for which X_k^u is nonempty and uses the Metropolis–Hastings algorithm to replace old values of X_k^u with new ones.

To draw the new value of X_k^u , let \mathbf{X}_{-k}^u denote the set of all unobserved realizations save X_k^u , the unobserved part of the process realized at time kh . Let \mathbf{X}^o denote the set of all observed data.

Our goal is to draw from the conditional distribution, $p(X_k^u | \mathbf{X}_{-k}^u, \mathbf{X}^o, \phi)$. Because the Euler approximation is a Markov process (reflecting our assumption about the underlying diffusion), only the contemporaneous and adjacent observations are relevant conditioning variables, meaning that

$$p(X_k^u | \mathbf{X}_{-k}^u, \mathbf{X}^o, \phi) = p(X_k^u | X_{k-1}, X_k^o, X_{k+1}, \phi). \quad (29)$$

Bayes’ rule and the Markov property can be applied to show that this density is proportional to

$$\pi(X_k^u) \equiv p(X_{k+1} | X_k^u, X_k^o, \phi) p(X_k^u | X_{k-1}, X_k^o, \phi), \quad (30)$$

a product of two Gaussian kernels, one for X_{k+1} and one for X_k^u . While $\pi(X_k^u)$ is generally not proportional to a Gaussian density for X_k^u , the second term, $p(X_k^u | X_{k-1}, X_k^o, \phi)$, is a Gaussian density corresponding to the conditional distribution of X_k^u from the Euler discretization.

We therefore use this density as a candidate generator for drawing X_k^u with the Metropolis–Hastings algorithm.

For every candidate-generating density, the Metropolis–Hastings algorithm specifies the acceptance probability required for convergence. The acceptance probability depends on both the target and candidate-generating densities evaluated at both the current and candidate draws. This probability is higher for candidate draws that have higher probability under the target density, but is lessened for draws that are generated too frequently by the candidate distribution. If $q(X_k^u)$ denotes the density of the candidate generator and $\pi(X_k^u)$ the target density (up to a constant of proportionality), then the acceptance probability (the probability of replacing the current draw X_k^u with a new draw X_k^{u*}) is equal to

$$\min \left\{ \frac{\pi(X_k^{u*})q(X_k^u)}{\pi(X_k^u)q(X_k^{u*})}, 1 \right\}. \tag{31}$$

The main advantage of the candidate-generating density proposed is simply that it reduces the number of calculations required to implement the algorithm, since the candidate density cancels out one of the kernels in the target density of Equation (30). The candidate density $p(X_k^u|X_{k-1}, X_k^o, \phi)$, along with a target density that is proportional to Equation (30), therefore results in a very simple implementation of Metropolis–Hastings:

- Draw a candidate value, X_k^{u*} , from $p(X_k^u|X_{k-1}, X_k^o, \phi)$ as a possible replacement of the current value, X_k^u .
- Replace the current value, X_k^u , with the new draw, X_k^{u*} , with probability

$$\min \left\{ \frac{p(X_{k+1}|X_k^{u*}, X_k^o, \phi)}{p(X_{k+1}|X_k^u, X_k^o, \phi)}, 1 \right\}. \tag{31}$$

Otherwise, retain the old value.

Essentially we simulate the process forward from time $(k - 1)h$ to time kh to generate the candidate draw X_k^{u*} , then accept X_k^{u*} over the current draw X_k^u depending on how likely each one is to have preceded X_{k+1} .

One of the important characteristics of a candidate-generating density is that its tails dominate those of the target. If this is not the case, then the algorithm may display high rejection rates or even become “stuck” for many draws. The candidate generator chosen naturally has fatter tails than the target because it is conditioned on less information (X_{k-1} and X_k^o) than the target density (X_{k-1} , X_k^o , and X_{k+1}), so we do not experience such problems here. Typically the acceptance rate for draws in the univariate case is about .6, while for the bivariate case it is about .4.

Lastly, in the interest rate diffusions considered in the article, negative interest rates are prohibited. Because the interest rate volatility rapidly declines as $r \rightarrow 0$ for the models considered, it is extremely rare for the candidate generator to produce a negative candidate draw for the interest rate. In these rare cases we simply reject the draw.

Appendix C: Convergence of the Euler Approximation

Because Lipschitz and grown conditions are not satisfied by the drift or diffusion functions of the nonlinear model, standard sufficient conditions for the convergence of the Euler approximation are not met, raising the possibility that the approximation may not converge. In this appendix I briefly consider two moment-based tests of convergence intended to provide some validation of the Euler discretization.

For a given set of parameter values satisfying stationarity conditions, N hundred year-long paths of the interest rate process are simulated using the Euler discretization of Equation (9). The terminal value of the i th simulation, $r_{T,i}$ is taken as a single draw from the unconditional distribution of the discretized process.

I first test whether the unconditional first through fourth moments of the discretized process match those of the corresponding diffusion. Following the work of Ait-Sahalia (1996a and 1996b), tools for solving for stationary densities have become well known. In particular, we know that the stationary distribution of any diffusion process $dr_t = \mu(r_t) dt + \sigma(r_t) dB_t$ is proportional to

$$\frac{1}{\sigma(r)^2} \exp\left(\int_{r_0}^r \frac{2\mu(x)}{\sigma(x)^2} dx\right). \tag{32}$$

Once this density is computed, moments may be calculated by quadrature.

Letting M_j denote the j th uncentered moment of r_T calculated from Equation (32), define the vector h_i as

$$h_i = \begin{bmatrix} r_{T,i} - M_1 \\ r_{T,i}^2 - M_2 \\ r_{T,i}^3 - M_3 \\ r_{T,i}^4 - M_4 \end{bmatrix}. \tag{33}$$

If the unconditional density of the Euler approximation, computed by simulation, matches the diffusion process density, then $E[h_i] = 0$. Given the independence of the $r_{T,i}$, these moment restrictions may be tested, following Hansen (1982), by defining

$$g = \frac{1}{N} \sum_{i=1}^N h_i \quad \text{and} \quad S = \frac{1}{N} \sum_{i=1}^N h_i h_i' \tag{34}$$

and computing the statistic

$$q_1 = Tg'S^{-1}g. \tag{35}$$

If $E[h_i] = 0$, then as $N \rightarrow \infty$ the limiting distribution of q_1 is χ^2 with four degrees of freedom. (Since there are no parameters that must be estimated, the number of degrees of freedom matches the number of moment conditions.)

The second test uses moment restrictions derived by Hansen and Scheinkman (1995) for stationary diffusion processes. Let $\phi(x) = x$ and $\phi^*(x) = x^2$ denote two “test functions” and \mathcal{A} the infinitesimal generator of the interest rate diffusion process for a given set of parameter values satisfying stationarity conditions.²⁰ Hansen and Scheinkman’s results may be applied to show that if the values $r_{T,i}$ are generated by the diffusion process corresponding to \mathcal{A} , then the random vector

$$z_i = \begin{bmatrix} \mathcal{A}\phi(r_{T,i}) \\ \mathcal{A}\phi^*(r_{T,i}) \\ \mathcal{A}\phi(r_{T,i})\phi(r_{T-1,i}) - \phi(r_{T,i}) \mathcal{A}\phi(r_{T-1,i}) \\ \mathcal{A}\phi^*(r_{T,i})\phi(r_{T-1,i}) - \phi^*(r_{T,i}) \mathcal{A}\phi(r_{T-1,i}) \\ \mathcal{A}\phi(r_{T,i})\phi^*(r_{T-1,i}) - \phi(r_{T,i}) \mathcal{A}\phi^*(r_{T-1,i}) \\ \mathcal{A}\phi^*(r_{T,i})\phi^*(r_{T-1,i}) - \phi^*(r_{T,i}) \mathcal{A}\phi^*(r_{T-1,i}) \end{bmatrix} \tag{36}$$

²⁰ Hansen and Scheinkman’s (1995) results are stated in terms of the infinitesimal generators or the forward (\mathcal{A}) and reverse-time (\mathcal{A}^*) processes. For stationary one-factor diffusions, $\mathcal{A} = \mathcal{A}^*$, simplifying the presentation here.

Table 5
Euler discretization convergence tests

	Daily data			
	$h = 1$	$h = .5$	$h = .33$	$h = .2$
Test 1 (p -value)	8.61 (0.07)	0.42 (0.98)	4.30 (0.37)	8.44 (0.08)
Test 2 (p -value)	4.07 (0.67)	1.19 (0.98)	10.69 (0.10)	5.26 (0.51)
	Monthly data			
	$h = 1$	$h = .5$	$h = .2$	$h = .05$
Test 1 (p -value)	107.62 (0.00)	15.34 (0.00)	3.21 (0.52)	5.99 (0.20)
Test 2 (p -value)	77.00 (0.00)	44.98 (0.00)	8.14 (0.23)	6.23 (0.40)

The table reports test statistics and p -values for two tests of the convergence of the Euler approximation of the model

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 / r_t) dt + \sigma r_t^\gamma dB_t.$$

In the top panel, results are reported for data simulated under the parameter values reported in the flat prior column of Table 1, panel A, while the bottom panel uses parameters from panel B.

Test 1 checks that the first four unconditional moments of the discretized process match those of the diffusion and is distributed asymptotically as a $\chi^2(4)$ under the null hypothesis that the discretized and continuous-time processes produce the same stationary distribution. Test 2 checks moments from Hansen and Scheinkman (1995) and is distributed as a $\chi^2(6)$ under the null.

has mean zero. Note for this test both the terminal value and next-to-terminal values $r_{T,i}$ and $r_{T-1,i}$ must be collected from each simulation. A test statistic q_2 may be calculated similarly to q_1 , but q_2 will have six degrees of freedom instead of four.

While the first test is used to check that the Euler approximation and diffusion process produce the same marginal distribution for r_T , the second test, since it relies on the joint distribution of r_{T-1} and r_T , should also detect discrepancies between the transition probabilities of the Euler approximation and that of the diffusion.

Each test was implemented using both the flat prior posterior means from daily and monthly data to simulate data and construct the g_i and z_i variables. One hundred thousand independent simulations were performed using values of h ranging from 1 to .05. Test statistics and p -values are displayed in Table 5.

Overall, convergence does not seem to be much of an issue for data sampled at a daily frequency, as none of the test statistics are large enough to reject the null hypothesis that the simulated data are equivalent to data generated by the limiting diffusion process. Even the daily simulations with $h = 1$ produce a distribution that is indistinguishable from the true diffusion.

With monthly data, discretization bias is clearly evident, as both tests easily reject the null that the Euler approximation, simulated with either $h = 1$ or $h = .5$, generates the same distribution as the diffusion process. Convergence appears extremely likely though, since the same tests do not result in rejections for smaller values of h . I conclude that concerns about the validity of the Euler approximation for this model are not large enough to avoid its use, though h should preferably be set equal to a number smaller than .2 for monthly data.

Appendix D: Drawing the Variance Parameters of the Short-Rate Model

The Euler approximation of the nonlinear spot rate model is given by

$$r_{k+1} - r_k = h(\alpha_0 + \alpha_1 r_k + \alpha_2 r_k^2 + \alpha_3 / r_k) + \sqrt{h} \sigma r_k^\gamma \epsilon_{k+1}. \tag{37}$$

Because γ is unknown, a closed-form conditional distribution for the full parameter vector $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \sigma, \gamma)$ given the augmented dataset does not exist.

Were γ known, however, the Euler approximation could be rearranged as

$$\frac{r_{k+1} - r_k}{\sqrt{hr_k^\gamma}} = \alpha_0 \sqrt{hr_k^{-\gamma}} + \alpha_1 \sqrt{hr_k^{1-\gamma}} + \alpha_2 \sqrt{hr_k^{2-\gamma}} + \alpha_3 \sqrt{hr_k^{-1-\gamma}} + \sigma \epsilon_{k+1}. \quad (38)$$

In this standard regression form, flat priors imply an inverted gamma distribution for σ and a multivariate Student's t distribution for $(\alpha_0, \alpha_1, \alpha_2, \alpha_3)$. If σ were known as well, $(\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ would be multivariate normal.

Given this distribution for

$$p(\alpha_0, \alpha_1, \alpha_2, \alpha_3 | \mathbf{R}^o, \mathbf{R}^u, \sigma, \gamma), \quad (39)$$

the full parameter posterior could be drawn from in two separate blocks if the conditional distribution

$$p(\sigma, \gamma | \mathbf{R}^o, \mathbf{R}^u, \alpha_0, \alpha_1, \alpha_2, \alpha_3) \quad (40)$$

were known as well. While this distribution is of an unknown form, its density is known up to a constant of proportionality as the product of the Gaussian Euler likelihood and the prior distribution. Drawing σ and γ using the Metropolis–Hastings algorithm is therefore feasible.

Construction of a Metropolis step requires the specification of a candidate-generating density for (σ, γ) . Our choice of candidate generator is driven by the availability of analytical draws from

$$q(\sigma | \gamma) = p(\sigma | \mathbf{R}^o, \mathbf{R}^u, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \gamma). \quad (41)$$

Following standard linear regression techniques, this distribution of σ is an inverted gamma.

Given a candidate-generating density for γ , say $q(\gamma)$, a joint candidate generator is given by

$$q(\sigma, \gamma) = q(\sigma | \gamma)q(\gamma). \quad (42)$$

For $q(\gamma)$ we will choose a Gaussian density with mean M_γ and variance V_γ^2 , which are chosen by trial and error to minimize serial correlation in the draws of σ and γ .

The Metropolis–Hastings acceptance probability, the probability of moving from one draw (σ, γ) to a new draw (σ^*, γ^*) is therefore equal to

$$\min \left\{ \frac{L(\sigma^*, \gamma^*)q(\sigma | \gamma)q(\gamma)}{L(\sigma, \gamma)q(\sigma^* | \gamma^*)q(\gamma^*)}, 1 \right\}, \quad (43)$$

where $L(\sigma, \gamma)$ is the data likelihood for σ and γ holding the other parameters fixed. The Markov chain will therefore tend to move from (σ, γ) to (σ^*, γ^*) when the latter yields a higher likelihood, with this tendency tempered by the probabilities of (σ^*, γ^*) versus (σ, γ) as draws from the candidate-generating distribution.

Appendix E: Implementing the Jeffreys Prior

As a preliminary step, we discuss the calculation of the Jeffreys prior. While the full Jeffreys prior is formulated as the square root of the determinant of the information matrix, for multiparameter models it is common to define the Jeffreys prior for a subset of the parameters of the model. What is called the “Jeffreys prior” in this article actually consists of a flat prior on σ and γ , or $p(\sigma, \gamma) \propto 1/\sigma$, multiplied by the square root of the determinant of the block of the information matrix that pertains to α .

Using the Euler approximation likelihood, we calculate the 4×4 information matrix for the α vector, whose (i, j) element is given by

$$-E \left[\frac{\partial^2 \log L}{\partial \alpha_i \partial \alpha_j} \right], \quad (44)$$

where $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$.

Evaluation of these expressions is problematic for several reasons. First, although these partial derivatives can be evaluated easily, it is not possible to compute the expectations of these expressions analytically. We must therefore resort to simulation to take expectations.

Second, the likelihood of the process is only computable after augmenting with high-frequency data, while the expression above is an expectation of a function of the observed data only. In order to maintain tractability, an approximate Jeffreys prior is therefore derived under the assumption that the discretized process is observed continuously rather than only once per period. Since more frequent observation of a process does not generally result in sharper inference about mean parameters, the effect of assuming more frequent observation is most likely unimportant.

The Jeffreys prior is defined as the square root of the determinant of the information matrix. Given observations observed at intervals of length h , the likelihood function may be written as

$$L = \prod_{k=0}^{K-1} \frac{1}{\sqrt{2h\pi\sigma r_k^\gamma}} \exp\left(-\frac{1}{2} \frac{(r_{k+1} - r_k - h\alpha_0 - h\alpha_1 r_k - h\alpha_2 r_k^2 - h\alpha_3/r_k)^2}{h\sigma^2 r_k^{2\gamma}}\right). \tag{45}$$

Taking logs and calculating second derivatives, we find

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \alpha_0^2} &= -\frac{h}{2\sigma^2} \sum_{k=0}^{K-1} r_k^{-2\gamma} & \frac{\partial^2 \log L}{\partial \alpha_0 \partial \alpha_1} &= -\frac{h}{2\sigma^2} \sum_{k=0}^{K-1} r_k^{1-2\gamma} \\ \frac{\partial^2 \log L}{\partial \alpha_0 \partial \alpha_2} &= -\frac{h}{2\sigma^2} \sum_{k=0}^{K-1} r_k^{2-2\gamma} & \frac{\partial^2 \log L}{\partial \alpha_0 \partial \alpha_3} &= -\frac{h}{2\sigma^2} \sum_{k=0}^{K-1} r_k^{-1-2\gamma} \\ \frac{\partial^2 \log L}{\partial \alpha_1^2} &= -\frac{h}{2\sigma^2} \sum_{k=0}^{K-1} r_k^{2-2\gamma} & \frac{\partial^2 \log L}{\partial \alpha_1 \partial \alpha_2} &= -\frac{h}{2\sigma^2} \sum_{k=0}^{K-1} r_k^{3-2\gamma} \\ \frac{\partial^2 \log L}{\partial \alpha_1 \partial \alpha_3} &= -\frac{h}{2\sigma^2} \sum_{k=0}^{K-1} r_k^{-2\gamma} & \frac{\partial^2 \log L}{\partial \alpha_2^2} &= -\frac{h}{2\sigma^2} \sum_{k=0}^{K-1} r_k^{4-2\gamma} \\ \frac{\partial^2 \log L}{\partial \alpha_2 \partial \alpha_3} &= -\frac{h}{2\sigma^2} \sum_{k=0}^{K-1} r_k^{1-2\gamma} & \frac{\partial^2 \log L}{\partial \alpha_3^2} &= -\frac{h}{2\sigma^2} \sum_{k=0}^{K-1} r_k^{-2-2\gamma} \end{aligned} \tag{46}$$

Using the fact that $T = Kh$, define

$$N(p) = \frac{1}{\sigma^2} E \left[\frac{T}{K} \sum_{k=0}^{K-1} r_k^{p-2\gamma} \right], \tag{47}$$

and note that as $h \rightarrow 0$ $N(p)$ converges to the expected path integral of some power of r_t . The α block of the information matrix is therefore proportional to

$$I = \begin{bmatrix} N(0) & N(1) & N(2) & N(-1) \\ N(1) & N(2) & N(3) & N(0) \\ N(2) & N(3) & N(4) & N(1) \\ N(-1) & N(0) & N(1) & N(-2) \end{bmatrix}, \tag{48}$$

and the Jeffreys prior on (α, σ, γ) ,

$$p_J(\alpha, \sigma, \gamma) \propto \sqrt{|I|}/\sigma. \tag{49}$$

To compute the Jeffreys prior in practice, the expectations in $N(p)$ must be computed by simulation. To evaluate the prior for a given set of parameters, 1000 interest rate paths, based on 500 paths of standard normal deviates, were simulated using antithetic random variables. To prevent nonnegativity, paths were truncated at .1%.²¹

²¹ Changing the truncation to .01% did not substantially affect any results.

Rather than redoing the parameter draws under the Jeffreys prior, we can make use of the 10,000 parameter draws made for the flat prior. Let the subscript J denote the Jeffreys prior and F the flat prior, so

$$p_J(\alpha, \sigma, \gamma | \mathbf{R}^0) \propto p(\mathbf{R}^0 | \alpha, \sigma, \gamma) p_J(\alpha, \sigma, \gamma) \quad (50)$$

$$p_F(\alpha, \sigma, \gamma | \mathbf{R}^0) \propto p(\mathbf{R}^0 | \alpha, \sigma, \gamma) \frac{1}{\sigma}. \quad (51)$$

Substituting the second expression into the first, we have

$$p_J(\phi | \mathbf{R}^0) \propto p_F(\phi | \mathbf{R}^0) p_J(\alpha, \sigma, \gamma) \sigma \propto p_F(\phi | \mathbf{R}^0) \sqrt{|I|}. \quad (52)$$

From our flat prior analysis, we already have many draws from $p_F(\phi | \mathbf{R}^0)$. To “convert” these draws into draws from $p_J(\phi | \mathbf{R}^0)$, we turn once again to the Metropolis–Hastings algorithm. Using our empirical distribution of $p_F(\phi | \mathbf{R}^0)$ as the candidate generator, the Metropolis acceptance probability of moving from $\phi = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \sigma, \gamma)$ to $\phi^* = (\alpha_0^*, \alpha_1^*, \alpha_2^*, \alpha_3^*, \sigma^*, \gamma^*)$ takes a simple form:

$$\alpha(\phi, \phi^*) = \max \left\{ \frac{\sigma^* p_J(\phi^*)}{\sigma p_J(\phi)}, 1 \right\}. \quad (53)$$

Parameters that have greater probability under the Jeffreys prior will be favored, while those with a low Jeffreys prior probability will be accepted with lower probability.

References

- Ahn, D., and B. Gao, 1999, “A Parametric Nonlinear Model of Term Structure Dynamics,” *Review of Financial Studies*, 12, 721–762.
- Ait-Sahalia, Y., 1996a, “Nonparametric Pricing of Interest Rate Derivative Securities,” *Econometrica*, 64, 527–560.
- Ait-Sahalia, Y., 1996b, “Testing Continuous-Time Models of the Spot Interest Rate,” *Review of Financial Studies*, 9, 385–426.
- Ait-Sahalia, Y., 2002, “Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach,” *Econometrica*, 70, 223–262.
- Andersen, T. G., and J. Lund, 1997, “Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate,” *Journal of Econometrics*, 77, 343–377.
- Ang, A., and G. Bekaert, 2002, “Short Rate Nonlinearities and Regime Switches,” *Journal of Economic Dynamics and Control*, 26, 1243–1274.
- Balduzzi, P., S. R. Das, and S. Foresi, 1996, “The Central Tendency: A Second Factor in Bond Yields,” *Review of Economics and Statistics*, 80, 62–72.
- Bandi, F., and P. C. B. Phillips, 2001, “Fully Nonparametric Estimation of Scalar Diffusion Models,” Discussion Paper 1332, Cowles Foundation.
- Bekaert, G., R. J. Hodrick, and D. A. Marshall, 2000, “Peso Problem Explanations for Term Structure Anomalies,” forthcoming in *Journal of Monetary Economics*.
- Box, G. E. P., and G. C. Tiao, 1973, *Bayesian Inference in Statistical Analysis*, Wiley, New York.
- Brandt, M. W., and P. Santa-Clara, 2002, “Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets,” *Journal of Financial Economics*, 63, 161–210.
- Casella, G., and E. I. George, 1992, “Explaining the Gibbs Sampler,” *American Statistician*, 46, 167–174.
- Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders, 1992, “An Empirical Comparison of Alternative Models of the Short-Term Interest Rate,” *Journal of Finance*, 47, 1209–1227.

- Chapman, D. A., J. B. Long, Jr., and N. D. Pearson, 1999, "Using Proxies for the Short Rate: When are Three Months Like an Instant?" *Review of Financial Studies*, 12, 763–806.
- Chapman, D. A., and N. D. Pearson, 2000, "Is the Short Rate Drift Actually Nonlinear?" *Journal of Finance*, 55, 355–388.
- Chib, S., and E. Greenberg, 1996, "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, 12, 409–431.
- Conley, T. G., L. P. Hansen, E. G. J. Luttmer, and J. A. Scheinkman, 1997, "Short-Term Interest Rates as Subordinated Diffusions," *Review of Financial Studies*, 10, 525–578.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407.
- Das, S. R., 2002, "The Surprise Element: Jumps in Interest Rates," *Journal of Econometrics*, 106, 27–65.
- Duffee, G., 1996, "Idiosyncratic Variation of Treasury Bill Yields," *Journal of Finance*, 51, 527–552.
- Duffie, D., and K. J. Singleton, 1988, "Simulated Moments Estimation of Diffusion Models of Asset Prices," working article, Stanford University.
- Duffie, D., and K. J. Singleton, 1993, "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.
- Durham, G. B., 2002, "Likelihood-Based Specification Analysis of Continuous-Time Models of the Short-Term Interest Rate," forthcoming in *Journal of Financial Econometrics*.
- Elerian, O., S. Chib, and N. Shephard, 2000, Likelihood Inference for Discretely Observed Non-Linear Diffusions," *Econometrica* 69, 959–993.
- Eraker, B., 2001, "MCMC Analysis of Diffusion Models with Application to Finance," *Journal of Business and Economic Statistics*, 19, 177–191.
- Gallant, R., and G. Tauchen, 1996, "Which Moments to Match?," *Journal of Econometric Theory*, 12, 657–681.
- Gray, S., 1996, "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27–62.
- Gourieroux, C., A. Monfort, and E. Renault, 1993, "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.
- Hamilton, J. D., 1994, *Time Series Analysis*, Princeton University Press, Princeton, NJ.
- Hamilton, J. D., 1996, "The Daily Market for Federal Funds," *Journal of Political Economy*, 104, 26–56.
- Hansen, L. P., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.
- Hansen, L. P., and J. A. Scheinkman, 1995, "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63, 767–804.
- Heath, D., R. Jarrow, and A. Morton, 1992, "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation," *Econometrica*, 60, 77–105.
- Hull, J., and A. White, 1990, "Pricing Interest Rates Derivative Securities," *Review of Financial Studies*, 3, 573–592.
- Jacquier, E., N. G. Polson, and P. E. Rossi, 1994, "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371–389.
- Jegadeesh, N. and G. G. Pennacchi, 1996, "The Behavior of Interest Rates Implied by the Term Structure of Eurodollar Futures," *Journal of Money, Credit and Banking*, 28, 426–446.

- Jiang, G. J., and J. L. Knight, 1997, "A Nonparametric Approach to the Estimation of Diffusion Processes, with an Application to a Short-Term Interest Rate Model," *Econometric Theory*, 13, 615–645.
- Johannes, M. S., 2002, "The Statistical and Economic Role of Jumps in Interest Rates," forthcoming in *Journal of Finance*.
- Jones, C. S., 1999, "Bayesian Estimation of Continuous-Time Finance Models," working article, University of Rochester.
- Kloeden, P. E., and E. Platen, 1992, *Numerical Solution of Stochastic Differential Equation*, Springer-Verlag, New York.
- Künsch, H. R., 1989, "The Jackknife and Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.
- Leamer, E. E., 1985, "Sensitivity Analysis Would Help," *American Economic Review*, 75, 300–313.
- Longstaff, F. A., 2000, "The Term Structure of Very Short-Term Rates: New Evidence for the Expectations Hypothesis," *Journal of Financial Economics*, 58, 397–415.
- Marriott, F. H. C., and J. A. Pope, 1954, "Bias in the Estimation of Autocorrelations," *Biometrika*, 41, 390–402.
- Naik, V., and M. H. Lee, 1993, "The Yield Curve and Bond Option Prices with Discrete Shifts in Economic Regimes," working paper, University of British Columbia.
- Pagan, A. R., A. D. Hall, and V. Martin, 1996, "Modeling the Term Structure," in G. S. Maddala and C. R. Rao (eds.), *Handbook of Statistics*, 14, North-Holland, Amsterdam, 91–118.
- Pedersen, A. R., 1995, "A New Approach to Maximum-Likelihood Estimation of Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, 22, 55–71.
- Pfann, G. A., P. C. Schotman, and T. Tschernig, 1996, "Nonlinear Interest Rate Dynamics and Implications for the Term Structure," *Journal of Econometrics*, 74, 149–176.
- Phillips, P. C. B., 1991a, "Bayesian Routes and Unit Roots: De Rebus Prioribus Semper Est Disputandum," *Journal of Applied Econometrics*, 6, 435–473.
- Phillips, P. C. B., 1991b, "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics*, 6, 333–364.
- Piazzesi, M., 2001, "An Econometric Model of the Yield Curve with Macroeconomic Jump Effects," Working Paper 8246, NBER.
- Poirier, D. J., 1995, *Intermediate Statistics and Econometrics: A Comparative Approach*, MIT Press, Cambridge, MA.
- Pritsker, M., 1998, "Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models," *Review of Financial Studies*, 11, 449–487.
- Stambaugh, R. F., 1998, "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421.
- Stanton, R., 1997, "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance*, 52, 1973–2002.
- Tanner, M. A., and W. H. Wong, 1987, "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–549.
- Vasicek, O., 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.
- Zellner, A., 1975, "Bayesian Analysis of Regression Error Terms," *Journal of the American Statistical Association*, 70, 138–144.