Bayesian Estimation of Continuous-Time Finance Models

Christopher S. Jones*
Simon School of Business
Carol Simon Hall 3-110P
University of Rochester
Rochester, NY 14627
Tel: (716) 275-3491
c_s_jones@ssb.rochester.edu
http://www.ssb.rochester.edu/fac/c_s_jones

First draft: September 1997
This revision: December 1998

J.E.L. classifications: C11, C15, G12
Key words: diffusion process, Bayesian analysis, Markov chain Monte Carlo, price discreteness, stochastic volatility, jump-diffusion

Abstract

A new Bayesian method is proposed for the analysis of discretely sampled diffusion processes. The method, which is termed high frequency augmentation (HFA), is a simple numerical method that is applicable to a wide variety of univariate or multivariate diffusion and jump-diffusion processes. It is furthermore useful when observations are irregularly observed, when one or more elements of the multivariate process are latent, or when microstructure effects add error to the observed data. The Markov chain-Monte Carlo-based procedure can be used to attain the posterior distributions of the parameters of the drift and diffusion functions as well as the posteriors of missing or latent data. Several examples are explored. First, posteriors of the parameters of a geometric Brownian motion are attained using HFA and compared with those obtained using standard analytical methods in a short Monte Carlo study. Second, a stochastic volatility model is estimated on a sample of S&P 500 returns, a problem for which posteriors are analytically intractable. Third, it is shown how the method can be used to estimate an interest rate process using data that suffer from severe rounding. Finally, extension of the method to jump-diffusions is described and applied to the analysis of the U.S dollar/German mark exchange rate.

*I thank Michael Brandt, Valentina Corradi, Frank Diebold, Bjorn Eraker, Eric Jacquier, Craig MacKinlay, Robert Stambaugh, and participants at the 1998 North American summer meetings of the Econometric Society for their comments and help. All errors remain my responsibility.

1 Introduction

Stochastic differential equations have become indispensable tools of finance and economics in the last twenty-five years. Advances in the theory of continuous time finance, for example, have made the most complicated problems in derivative pricing relatively simple, causing the usage of diffusion models to become widespread. While the theoretical developments generally presuppose knowledge of the parameters that underlie these diffusions, in reality these parameters are unknown and must be estimated.

Many methods currently exist to estimate diffusion processes, although few can handle the generality of processes considered in the theoretical literature. Multivariate models, especially ones with latent variables such as stochastic volatility, are commonplace in finance, yet many existing econometric methods are not capable of estimating such models (e.g Aït-Sahalia, 1996, Pedersen, 1995, and Santa-Clara, 1995). Equally common is the requirement by many methods that the process under consideration be stationary (e.g. Hansen and Scheinkman, 1995), an assumption that is certainly violated by many economic variables, most importantly stock prices.

Another concern is that many methods do not make full use of the information contained in the sample, either because they match only certain moments or even just stationary distributions (e.g. Aït-Sahalia, 1996). The heavier reliance on large-sample asymptotics that are necessitated by such methods may make inferences unreliable in finite samples. Such issues are critical when dealing with highly persistent data, such as interest rates, or the short sample sizes available for recently introduced securities.

The method proposed in this paper is Bayesian and is applicable to a large class of discretely observed multivariate diffusion and jump-diffusion processes, including those for which observations are irregularly spaced, those in which some elements of the process are latent (such as stochastic volatility), and those in which the data are observed with error due to factors

such as market microstructure. As the method is Bayesian, there is no need for large sample approximations; inference will be exact in finite samples. Likewise, stationarity is not required, although it may be imposed through the prior if desired.

There are practical reasons why these advantages may be important. Because diffusion-based models are frequently used in derivatives pricing, an exact characterization of finite sample uncertainty is critical for the risk management of firms that enter into derivative contracts. The Bayesian framework is particularly useful because it can link uncertainty about parameters and latent variables to the predictive uncertainty of the process. For example, in a stochastic volatility model, it is this joint distribution of parameters and the current and future values of volatility and stock prices which determines the predictive distribution of gains and losses. The frequentist methods available are typically capable of making only a large sample approximation of parameter uncertainty while providing little connection to predictive uncertainty and uncertainty about unobserved variables.

Because the method is Bayesian, it enjoys several more advantages over frequentist methods. One is the ability to incorporate prior information, such as stationarity, if such information is available. Also, the posteriors of arbitrary functions of the model parameters and latent variables can be computed, at least numerically, allowing us to analyze in a straightforward manner the finite sample uncertainty in quantities such as equilibrium prices or optimal hedge ratios.

The method proposed in this paper employs two recent innovations in time series econometrics. The computational framework is provided by Jacquier, Polson, and Rossi (1994, hereafter JPR), who apply the Bayesian technique of data augmentation to the study of stochastic volatility models. Although the analysis of JPR deals with discrete time models, the computational backbone of their paper, which they term a cyclic Metropolis chain, is adopted here in a somewhat different form. The convergence of the likelihood approximations used relies on the results of Pedersen (1995) and Santa-Clara (1995).

Independently of this paper, Eraker (1998) and Elerian, Chib, and Shephard (1998) have proposed methods similar to the one presented here for the Bayesian analysis of diffusion processes. This paper differs from both in its consideration of jump-diffusions and measurement

error. In addition, Elerian, Chib, and Shephard (1998) consider only univariate models, making applications to many common finance models such as stochastic volatility impossible. The method we propose is simpler, and it is this simplicity that makes the method more flexible and easily generalized to more complex problems, such as the stochastic volatility/jump models of Bakshi, Cao, and Chen (1998).

The paper will proceed as follows. Section two contains the intuition behind the HFA procedure, while section three describes the method more formally and discusses the convergence results relevant to it. Three examples are considered in section four: univariate geometric Brownian motion, a stochastic volatility model, and an interest rate process with rounded data. Section five extends the method to jump-diffusion processes, and section six concludes.

2 High frequency augmentation

The primary difficulty in estimating diffusion processes stems from the intractability of their transition densities and hence likelihood functions.[1] 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.

Similar to the simulation-based frequentist literature, the solution of this paper is to work with a high frequency discrete time model rather than the diffusion process itself. The particular discretization scheme chosen is the Euler approximation, which has the advantage that its increments are conditionally Gaussian. Under regularity conditions, as the time interval of the discretized process shrinks to zero the likelihood function of the Euler approximation will converge to that of the diffusion.

In practice, it is sometimes assumed that the discrete time approximation *is* the true model of prices. If the goal of the analysis is to obtain inferences about the parameters of the *diffusion*

[1] Aït-Sahalia (1998) has shown how to construct analytical approximations of the likelihood function of a univariate diffusion process. Even in this univariate case, however, the likelihood function is of a nonstandard form, making the derivation of marginal posteriors problematic when there are more than a few parameters.

process, however, this approach will generally result in inconsistent estimates. It is imperative, therefore, that any use of discrete approximations allows the discretization interval to be arbitrarily small. In particular, the frequency of the discretized model must be allowed to be arbitrarily higher than the frequency of the data.

To resolve this asymmetry between model and data frequencies, this paper proposes to augment the observed data with much higher frequency data - for example augmenting monthly with daily data. Using the Euler approximation's Gaussian likelihood function, distributions for the model parameters conditional on the actual and augmented data may then be obtained, often by using standard analytical methods. We then integrate out, using Markov chain Monte Carlo, the uncertainty introduced by the introduction of unobserved data to get posteriors conditional on only the observed data.

Specifically, a Markov chain similar to a Gibbs sampler is employed to alternatively augment the observed low frequency data with paths of high frequency data, and then use this augmented sample to generate a conditional distribution of the parameters of the model. We emphasize that by adding data we do not seek to decrease the variance of the posteriors, since the effect of having a higher sample size will be offset by uncertainty about what unobserved higher frequency data was actually realized. Rather, data augmentation merely makes the discrete time approximation bias vanish, so that the posteriors being computed correspond to those of the true diffusion.

It is worth noting the differences between the use of the Euler approximation in this paper and its use in simulation-based classical methods. In Duffie and Singleton (1993), Gallant and Tauchen (1996), and Gourieroux, Monfort, and Renault (1993), among others, the Euler approximation is used to simulate long paths of data which are then used to numerically compute moments. In the simulated maximum likelihood methods of Pedersen (1995) and Santa-Clara (1995), the Euler approximation is used to simulate paths forward from one observation to the next, with the terminal values of the simulations used to approximate the one-period transition density.

Figure 1 illustrates these methods and shows that the simulations in this paper, in contrast,

merely bridge the observed low frequency data with short paths of high frequency data. In essence, data augmentation is nothing more than "filling in the gaps." From the picture 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 methods that pin down one or no ends. Since all these methods require some form of Monte Carlo integration, lower variance of the augmented data should translate into in greater computational efficiency.

3 Formalization of HFA

3.1 The Euler approximation

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

$$dX_t = \mu(X_t, \phi)dt + \sigma(X_t, \phi)dB_t, \tag{1}$$

where $\mu(x, \phi) : \Re^L \times \Phi \to \Re^L$ and $\sigma(x, \phi) : \Re^L \times \Phi \to \Re^L \otimes \Re^D$ satisfy growth and Lipschitz conditions, B_t is a D-dimensional standard Brownian motion, and ϕ is a vector of parameters.

It is a well-known result of the theory of stochastic differential equations (see Kloeden and Platen, 1992, p. 473) that under regularity conditions a variety of discretized approximations converge weakly to Itô processes as the time discretization step goes to zero. One approximation that is frequently employed is the Euler approximation, which is given by

$$\hat{X}_{(k+1)h} = \hat{X}_{kh} + h\mu(\hat{X}_{kh}, \phi) + \sqrt{h}\sigma(\hat{X}_{kh}, \phi)\epsilon_{k+1}, \tag{2}$$

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 frequently write X_{kh} as X_k, making dependence on a particular value of h implicit.

In addition to weak convergence, Pedersen (1995) shows that the likelihood function of the Euler approximation converges to that of the diffusion as well. In particular, under Lipschitz and growth conditions and the existence of continuous derivatives for μ and σ, the likelihood of the Euler approximation converges in probability for all parameter values ϕ to that of the

diffusion as the discretization interval h goes to zero. These conditions also guarantee the existence of a weak solution to the stochastic differential equation (1).

3.2 Data augmentation

As stated in the previous section, the approach followed in this paper will be to estimate the discretized Gaussian process (2), while allowing h to be arbitrarily small. The process being estimated will therefore be at a higher frequency than that of the observed data. We will, in effect, be estimating a Gaussian model with a large number of missing data points. The Bayesian tool used to resolve this asymmetry is a version of the Metropolis-Hastings algorithm. The particular variant used borrows the intuition of both the Gibbs sampler and and Tanner and Wong's (1987) data augmentation algorithm.

The motivation behind data augmentation is that many posterior distributions could be calculated more easily if some extra set of data was available Although we do not observe this extra "augmented" data, we may know (or be able to draw from) its distribution conditional on the observed data and the unobserved model parameters. Treating the augmented data as random variables equivalent to the model parameters, we form a Gibbs sampler or other Markov chain that alternates between drawing from the conditional distribution of the model parameters given both the observed and augmented data, and the conditional distribution of the augmented data given the observed data and the model parameters. In the current problem, this augmented data set corresponds to the set of paths of unobserved data connecting the observed data.

For the remainder of this section we ignore the distinction between diffusion and Euler approximation and proceed as though the data are generated by the latter, relying on the fact that the distribution of the two may be made arbitrarily close through the choice of h. We define a time grid T_h based on the choice of the Euler approximation's discretization interval, h, as

$$T_h = \{kh : k = 1, 2, \ldots, K\} \tag{3}$$

and assume that all realizations of the Euler process occur at times in T_h.

Suppose the vector X_k represents the time kh realization of the L-dimensional process generated by the Euler approximation (2). In general, the majority of components of X_k for a given k will be unobserved. This is primarily because the Euler approximation operates at a higher frequency than the observed data, so that for many k the vector X_k is entirely unobserved. It is also because some elements of the vector process X may *never* be observed, a leading example being volatility in a stochastic volatility model. In some cases, one element of X may be observed more frequently than another, or not over the same time interval.

We 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). The lengths of these subvectors vary through time, with the length of X_k^o denoted by m and the length of X_k^u denoted by n. (Note that m and n are time-varying and that L is not, although this is not made notationally explicit.) In all periods it is the case that $m+n=L$ and $X_k^u \cup X_k^o = X_k$.

Let $\mathbf{X^o}$ represent the union across time of the vectors X_k^o, and let $\mathbf{X^u}$ be the union of all X_k^u. If it were possible, we would alternate between drawing from the following conditional distributions:

$$p(\phi|\mathbf{X^o},\mathbf{X^u}) \quad \text{and} \quad p(\mathbf{X^u}|\phi,\mathbf{X^o}),$$

Under weak conditions, this Markov chain would converge to an invariant distribution $p(\phi,\mathbf{X^u}|\mathbf{X^o})$, whose marginal distribution on ϕ, $p(\phi|\mathbf{X^o})$, is frequently our object of interest. As we will see, however, the draw of $\mathbf{X^u}$ is nontrivial and must be decomposed further.

3.3 Drawing the parameters/choice of prior

We first consider the draw from $p(\phi|\mathbf{X^o},\mathbf{X^u})$, the conditional distribution of the model parameters given the data, both actual and augmented. From Bayes rule,

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

where L is the likelihood function and $p(\phi)$ is the prior. The advantage of the Euler approximation is that it allows us to compute $L(\phi;\mathbf{X^o},\mathbf{X^u})$ as the product of Gaussian transition densities. It is therefore, at the very least, possible to compute the conditional distribution of ϕ up to a

constant of proportionality. This fact alone makes it possible, in theory, to draw the parameter vector ϕ using a numerical procedure such as the Metropolis-Hastings algorithm.

In a number of situations, however, including many of the popular diffusions used in finance, the Gaussian nature of the Euler likelihood function makes the draw from $p(\phi|\mathbf{X^o}, \mathbf{X^u})$ straightforward. If we partition the parameter vector as $\phi = (\beta, \overline{\sigma})$, then this will be the case under flat ($p(\phi) \propto 1/\overline{\sigma}$) or normal/inverted Wishart priors when $\mu(X, \beta)$ is linear in β and when $\sigma(X, \overline{\sigma}) = \overline{\sigma} g(X)$, where $\overline{\sigma}$ is a parameter and g is a function that does not depend on ϕ.

In this case there exist functions $f_i(\cdot)$ independent of ϕ such that the Euler approximation may be written as

$$X_{k+1} = X_k + h\left(\sum_{j=1}^{J} \beta_j f_i(X_k)\right) + \sqrt{h}\overline{\sigma} g(X_k)\epsilon_{k+1},$$

which can be rearranged as

$$\frac{X_{k+1} - X_k}{\sqrt{h} g(X_k)} = \sum_{j=1}^{J} \beta_j \frac{\sqrt{h} f_i(X_k)}{g(X_k)} + \overline{\sigma}\epsilon_{k+1}.$$

Rescaling by $g(X_k)$ eliminates the heteroskedasticity of the disturbance term, putting the problem in a standard linear regression framework. It is well-known in this case (see Zellner, 1971, p. 60-61) that the posterior for β is a multivariate student-t and that σ is distributed as an inverted gamma. Standard methods for drawing from these distributions make more sophisticated methods such as Metropolis-Hastings unnecessary. The examples of section four all fall into this category.

In addition to simplifying the draws in this stage of the Markov chain, the availability of the analytical conditional distributions also makes it possible to compute the marginal posterior densities or moments using Rao-Blackwellization (see Tanner and Wong, 1987, or Gelfand and Smith, 1990). This method for post-processing the output of the Markov chain can substantially reduce the chain's Monte Carlo integration error, as we will see in the numerical examples.

An easy way to approximate posterior moments is simply to compute sample moments from the output draws of the Markov chain. Posterior densities are often approximated by computing a histogram of the same draws. Essentially, these basic methods use the parameter draws

themselves but ignore the information in the full conditional distribution of the parameters available at each iteration of the Markov chain.

The simplest version of Rao-Blackwellization computes the posterior mean as the average of $E[\phi|\mathbf{X^o}, \mathbf{X^u}]$ rather than as the average of the draws of ϕ themselves. Not surprisingly, the smoothing of the expectations operator generally causes a reduction in variance (see Liu, Wong, and Kong, 1994). This expectation will be known, for example, in the linear case discussed above. Posterior densities, too, can be computed by Rao-Blackwellization when the conditional densities $p(\phi|\mathbf{X^o}, \mathbf{X^u})$ are known in closed form. An approximation to $p(\phi|\mathbf{X^o})$ is computed by averaging $p(\phi|\mathbf{X^o}, \mathbf{X^u})$ across the draws of $\mathbf{X^u}$ generated by the Markov chain.

3.4 Drawing the augmented data

In general, the most challenging task in the implementation of HFA is to draw high frequency data that is consistent with both the current draw of the parameters and the observed data $\mathbf{X^o}$. If it were possible, we would draw from the distribution $p(\mathbf{X^u}|\phi, \mathbf{X^o})$ directly. It appears, however, that in almost every situation this complicated multivariate distribution is unknown. We therefore adopt a tool conceived by Jacquier, Polson, and Rossi (1994) for the analysis of a discrete time stochastic volatility model.

3.4.1 The cyclic Metropolis chain

The tool, which JPR call a cyclic Metropolis chain, breaks down the draw of $p(\mathbf{X^u}|\phi, \mathbf{X^o})$ into a large number of univariate draws. These univariate draws are made for each element of $\mathbf{X^u}$ conditional on ϕ, $\mathbf{X^o}$, and all the other elements in $\mathbf{X^u}$. By iterating over these conditional draws, JPR forms a Markov chain chain that "cycles" through all the elements of $\mathbf{X^u}$ as well as the parameter vector ϕ. Because even these univariate conditional densities are not of a known form, JPR must make draws using the Metropolis-Hastings algorithm. (For a review of the Metropolis-Hastings algorithm, see Chib and Greenberg, 1995.)

While JPR found it most convenient to break up the draw of $p(\mathbf{X^u}|\phi, \mathbf{X^o})$ into a number of

univariate draws, there is in general a variety of "blockings" that can be used to break up this multivariate draw. The one considered here – breaking up $\mathbf{X^u}$ into its components X_k^u – is the most convenient given the notation we have developed.

Let $\mathbf{X^u_{-k}}$ denote the set of all unobserved realizations save X_k^u, the unobserved part of the process realized at time kh. When all elements of $X_k = (X_k^o, X_k^u)$ are observed, then X_k^u is null and no draw needs to be made. When X_k is not fully observed, then X_k^u contains at least one element and our goal is to draw from its conditional distribution, $p(X_k^u|\mathbf{X^u_{-k}}, \mathbf{X^o}, \phi)$. Because the Euler approximation is a Markov process (reflecting our assumption about the underlying diffusion), only the adjacent observations are relevant, meaning that

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

Bayes rule can be applied to show that this density is proportional to

$$p(X_k^u, X_k^o, X_{k+1}|X_{k-1}, \phi).$$

and again to show proportionality with

$$p(X_{k+1}|X_k^u, X_k^o, X_{k-1}, \phi)p(X_k^u|X_{k-1}, X_k^o, \phi).$$

The Markov property further implies that this product of densities simplifies to

$$p(X_{k+1}|X_k^u, X_k^o, \phi)p(X_k^u|X_{k-1}, X_k^o, \phi). \tag{4}$$

This product of normal densities, which we denote in shorthand as $\pi(X_k^u)$, is easily computed and is therefore amenable to the Metropolis-Hastings algorithm. Only occasionally is this product itself proportional to a known standard density, making the Metropolis algorithm unnecessary. One example of this is found in section 4.2.

Construction of a Metropolis chain requires the specification of a candidate-generating density $q(x, x^*)$ governing the probability of x^* being chosen as a candidate to replace the existing draw x. Repeated draws from $q(x, x^*)$ are accepted with probability

$$\alpha(x, x^*) = \min\{\frac{\pi(x^*)q(x^*, x)}{\pi(x)q(x, x^*)}, 1\}. \tag{5}$$

If x^* is rejected, the current draw x is repeated. The Metropolis-Hastings result is that the resulting Markov chain has an invariant distribution whose density is proportional to $\pi(x)$.

There is a natural candidate generator that can be used to simplify the acceptance probability $\alpha(X_k^u, X_k^{u*})$ significantly. We saw previously that

$$p(X_k^u | \mathbf{X^u_{-k}}, \mathbf{X^o}, \phi) \propto p(X_{k+1} | X_k^u, X_k^o, \phi) p(X_k^u | X_k^o, X_{k-1}, \phi).$$

The second density on the right hand side, taken in isolation, implies a Gaussian distribution for X_k^u. We therefore define the candidate generating density as

$$q(X_k^u, X_k^{u*}) = q(X_k^{u*}) = p(X_k^{u*} | X_k^o, X_{k-1}, \phi).$$

Note that the current value of the Markov chain, X_k^u, does non enter the candidate generator – q generates a so-called independent Metropolis chain.

Of course, elements of the chain are not truly independent because they are linked by the acceptance probability α. This probability is simplified greatly by this choice of q, since q cancels out one of the two densities that make up π. We are left with

$$\alpha(X_k^u, X_k^{u*}) = \min\{\frac{p(X_{k+1} | X_k^{u*}, X_k^o, \phi)}{p(X_{k+1} | X_k^u, X_k^o, \phi)}, 1\}. \tag{6}$$

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 previous draw X_k^u depending on how likely each one is to have preceded X_{k+1}. Typically, the acceptance rate for draws in the univariate case is about .6, while for the bivariate case it is about .4.

The chief advantage of this forward simulation candidate generator is its simplicity. The Metropolis chain constructed using the forward simulator requires a fraction of the calculations required to construct more complicated Metropolis chains, reducing computer run time significantly.

3.4.2 Cycling versus simultaneous draws

Although it is most intuitive to think of the Markov chain as "cycling" through the elements of $\mathbf{X^u}$ drawing each element X_k^u 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. It is not problematic to adopt this strategy here as long as some care is taken.

Divide the time grid

$$T_h = \{kh : k = 0, 1, 2, ..., K\}$$

into two sets of non-adjacent times (assume K is even):

$$T_h^1 = \{kh : k = 0, 2, 4, ..., K\} \text{ and } T_h^2 = \{kh : k = 1, 3, 5, ..., K-1\}$$

As we have argued, because the process is Markovian only adjacent observations ($k-1$ and $k+1$) are relevant for each draw X_k^u. We can therefore draw all elements in T_h^1 simultaneously conditional on the observations at times in T_h^2, then draw observations in T_h^2 conditioning on those in T_h^1. By alternating between just *two* matrix draws rather than K individual draws we can sometimes improve computational performance dramatically.

3.5 A review of the algorithm

(1) Choose some initial values for the parameters ϕ and the unobserved data $\mathbf{X^u}$.
(2) Redraw the unobserved paths, or "bridges", in between the observed data and fill in any latent variables by cycling, point by point, through the elements of $\mathbf{X^u}$.
(3) Draw new parameters conditional on the augmented data set.
(4) Go back to **(2)** or terminate the Markov chain if convergence has been determined.

4 Three applications

In this section we will try both to assess basic statistical properties of the method and to demonstrate its usefulness in a variety of situations. We first perform a brief Monte Carlo study to compare the method with a more standard approach. A stochastic volatility model is then estimated using actual S&P 500 returns. Last, we consider a problem in which an interest rate process is observed in an environment of microstructure-induced measurement error.

4.1 Geometric Brownian motion

This section applies the algorithm described above to determine the posteriors of the parameters of a univariate geometric Brownian motion (GBM) process. GBM is examined because it has the property of having normally-distributed log differences, making it possible to compute analytically the posteriors for the model parameters without relying on data augmentation. A brief Monte Carlo exercise is therefore able to check directly whether the posteriors generated by this paper's high frequency augmentation (HFA) procedure are the same as those of the analytical method.

4.1.1 A standard approach

A geometric Brownian motion x is described by the stochastic differential equation

$$dx = \mu x dt + \sigma x dB.$$

By Itô's lemma the solution of this SDE can be written as

$$x_{t+1} = x_t \exp\left(\mu - \frac{1}{2}\sigma^2 + \sigma(B_{t+1} - B_t)\right), \tag{7}$$

which implies the well-known result that continuously compounded returns are normally distributed, or

$$\log(\frac{x_{t+1}}{x_t}) \sim N(\mu - \frac{1}{2}\sigma^2, \sigma^2).$$

Defining $\beta = \mu - \frac{1}{2}\sigma^2$, this can be rewritten as

$$\log(\frac{x_{t+1}}{x_t}) \sim N(\beta, \sigma^2).$$

We place a standard diffuse prior, $p(\mu, \sigma) \propto \frac{1}{\sigma}$, on μ and σ. Because the determinant of the Jacobian of the transformation from (β, σ) to (μ, σ) is a constant, the same diffuse priors are implied for β and σ, or $p(\beta, \sigma) \propto \frac{1}{\sigma}$.

This case is particularly easy to analyze, since the posteriors of β and σ are known analytically. In particular, the following results are attained (see Zellner, 1971, p. 60-61):

- Let $\hat{\sigma}^2$ be the sample variance of $\log(\frac{x_{t+1}}{x_t})$ and let T be the sample size. Then σ is distributed as an inverted gamma with gamma parameter equal to $\frac{2}{(T-2)\hat{\sigma}^2}$ and alpha parameter equal to $\frac{T-2}{2}$.

- Conditional on σ, β is normally distributed with mean equal to the sample mean of $\log(\frac{x_{t+1}}{x_t})$ and variance equal to $\frac{\sigma^2}{T-1}$.

Numerical integration may be used to solve for the marginal density $p(\mu|\mathbf{x})$ given the bivariate density $p(\beta, \sigma|\mathbf{x})$. Alternatively, by repeatedly drawing values from the joint posterior distribution of β and σ and using the relation $\mu = \beta + \frac{1}{2}\sigma^2$, we can similarly approximate the posterior of μ to any desired level of accuracy.

4.1.2 The HFA approach

The HFA procedure proposed in this paper does not make use of the exact Gaussian transformation that GBM allows. Instead, we approximate geometric Brownian motion by its Euler approximation,

$$x_{k+1} = x_k + h\mu x_k + \sqrt{h}\sigma x_k \epsilon_{k+1}. \tag{8}$$

As before, x_k denotes the value of the Euler approximation at time kh. Rescaling to eliminate heteroskedasticity, we have

$$\frac{x_{k+1} - x_k}{\sqrt{h}x_k} = \sqrt{h}\mu + \sigma\epsilon_{k+1}. \tag{9}$$

The Euler approximation can therefore be interpreted as saying that over very small intervals, *simple* returns are normally distributed, although with a different mean than for continuously compounded returns.

To draw the augmented data in this section we follow the procedure outlined in section three. The unobserved data, specifically the $\frac{1}{h} - 1$ high frequency values of x that lie between each two observations of the process, are individually drawn using Metropolis-Hastings. The candidate generating distribution, $p(x_k|x_{k-1}, \mu, \sigma)$, is normal with mean $(1 + h\mu)x_{k-1}$ and variance $h\sigma^2 x_{k-1}^2$. The probability of accepting a new draw x_k^* over the previous draw x_k is

equal to

$$\min\{\frac{x_k}{x_k^*}\exp\left(-\frac{1}{2}\frac{(x_{k+1}-(1+h\mu)x_k^*)^2}{(\sigma x_k^*)^2}+\frac{1}{2}\frac{(x_{k+1}-(1+h\mu)x_k)^2}{(\sigma x_k)^2}\right),1\}.$$

We again use the standard diffuse prior $p(\mu,\sigma)\propto\frac{1}{\sigma}$ in order to make the results comparable with those of the analytical method. Conditional on the augmented and actual data, we obtain simple expressions for the distributions of μ and σ:

- Let $\hat{\sigma}^2$ be the sample variance of $\frac{x_{k+1}-x_k}{\sqrt{h}x_k}$ and let $K=\frac{T-1}{h}$ be the size of the augmented sample. Then σ has an inverted gamma distribution with gamma parameter $\frac{2}{(K-1)\hat{\sigma}^2}$ and alpha parameter $\frac{K-1}{2}$.
- Conditional on σ, μ is normally distributed with mean equal to the sample mean of $\frac{x_{k+1}-x_k}{\sqrt{h}x_k}$ and variance equal to $\frac{\sigma^2}{Kh}$.

4.1.3 Monte Carlo results

In this section we conduct a modest Monte Carlo experiment to determine the properties of the algorithm proposed. Because the log differences of geometric Brownian motion are independent Gaussian random variables, the sampling distributions of the true posterior means and variances in this case are well-known. Our focus will therefore be on the performance of our numerical method *relative* to the analytical method. In essence, therefore, we are asking whether relatively short chains with reasonably small values of h are sufficient to generate accurate results.

One hundred paths of one hundred discrete time observations of geometric Brownian motion were simulated using the exact discretization (7) under the parameter values $\mu=.025$ and $\sigma=.25$, values that were deliberately chosen to magnify the impact of discretization bias.[2] A representative series is plotted in figure 2. Markov chains for the numerical method were then

[2]The value $\sigma=.25$ represents a higher volatility than that apparent in annual S&P 500 returns, for example. The value $\mu=.025$, meanwhile, is relatively small, making the difference between the log difference mean $\alpha=\mu-\frac{\sigma^2}{2}$ and the discretized mean μ large.

run to 11000 iterations, with the first 1000 thrown out to negate the effects of initial conditions. The discretization interval h is set to .1.

Posterior means for μ and σ were computed three different ways for each series of simulated data. First, we compute them using the analytical posterior $p(\beta, \sigma|\mathbf{x^o})$, where the means of μ were computed by numerical integration. For the numerical method, posterior means were computed once by simply taking averages of the Markov chain parameter draws. They were then computed using Rao-Blackwellization, in which the posterior mean is computed as the average across draws

Table 1: Monte Carlo Results for Geometric Brownian Motion

Let $\hat{\mu}_a$ and $\hat{\sigma}_a$ denote posterior means computed using the analytical method, $\hat{\mu}_{mc}$ and $\hat{\sigma}_{mc}$ the means computed by averaging the Markov chain draws, and $\hat{\mu}_{rb}$ and $\hat{\sigma}_{rb}$ the means computed by Rao-Blackwellization of those draws.

Panel A: analysis of posterior means

	$\hat{\mu}_a$	$\hat{\mu}_{mc}$	$\hat{\mu}_{rb}$	$\hat{\sigma}_a$	$\hat{\sigma}_{mc}$	$\hat{\sigma}_{rb}$
True parameter	0.0250	0.0250	0.0250	0.2500	0.2500	0.2500
Average posterior mean	0.0231	0.0229	0.0229	0.2525	0.2513	0.2514
(Standard error)	(0.0025)	(0.0026)	(0.0026)	(0.0019)	(0.0018)	(0.0018)
Root mean squared error	0.0255	0.0256	0.0256	0.0189	0.0184	0.0184

Panel B: deviations from analytical method

	$\hat{\mu}_{mc} - \hat{\mu}_a$	$\hat{\mu}_{rb} - \hat{\mu}_a$	$\hat{\sigma}_{mc} - \hat{\sigma}_a$	$\hat{\sigma}_{rb} - \hat{\sigma}_a$
Average	-0.0002	-0.0002	-0.0012	-0.0011
(Standard error)	(0.0001)	(0.0000)	(0.0002)	(0.0002)
Standard deviation	0.0005	0.0005	0.0020	0.0020

of the conditional means $E[\sigma|\mathbf{x^o}, \mathbf{x^u}]$ and $E[\mu|\mathbf{x^o}, \mathbf{x^u}]$, which are both known in closed form as the means of inverted gamma and student-t random variables, respectively.

The top panel of table 1 reports the average and root mean squared error of the posterior means when computed by each of the three methods. In every case but one ($\hat{\sigma}_a$), the average posterior mean is within one standard error of the true parameter value, providing no evidence that any of the methods appears to generate significant bias. Rao-Blackwellization, a variance reduction technique, not surprisingly does not affect these numbers. Comparing the root mean squared errors across the three methods, we furthermore find no evidence that either Markov chain-based method results in less precisely estimated posterior means.

It is perhaps more informative, however, to look at the *deviations* of the Markov chain posterior means from the analytical means. The lower panel of table 1 shows that the average

deviation, while statistically significant, is extremely small, in all cases less than one percent of the true parameter value. The last row of the table measures the variability of this deviation. For μ, this variation is minute. Furthermore, although too small to be visible in the table, Rao-Blackwellization reduces this standard deviation by about ten percent. For σ, the variation of the deviations from the analytical posterior means is somewhat larger, although still of little economic importance. Nevertheless, if this error is considered nontrivial, then a longer chain could be simulated with little additional difficulty.[3]

As an additional illustration, we consider one of the simulated paths of data (the same one plotted in figure 2) in more detail. To assess more informally how different choices of h affect the parameter posteriors, various degrees of data augmentation were considered, ranging from $h = 1$ (no data augmentation) to $h = .05$. No data augmentation amounts to the estimation of the Euler approximation (8) with $h = 1$, an approach which may perhaps be considered naive, but is nevertheless used in some empirical literature. The choice $h = .05$ is within the typical range used in the simulation-based classical literature.

Figures 3 and 4 plot the posterior densities of μ and σ attained using different levels of discretization.[4] Densities were computed using Rao-Blackwellization by averaging $p(\mu|\mathbf{x^o}, \mathbf{x^u})$ and $p(\sigma|\mathbf{x^o}, \mathbf{x^u})$ across draws of $\mathbf{x^u}$. From the top panels it is clear that most of the effect of data augmentation occurs by adding just a single augmented data point between observations ("h=.5"). Between $h = .1$ and $h = .05$, the posteriors appear to "settle down," although a bit of approximation error is clearly still evident.

The bottom panels compare the correct analytical posteriors to the most data-augmented posteriors and the "naive" posteriors that use no data augmentation. For both μ and σ it is clear that the naive method ("h=1") yields faulty inference. While the other two posteriors ("h=.05" and "analytical") are not indistinguishable, they are, in comparison to the naive posteriors, very

[3]The total run time of this Monte Carlo experiment was four hours on a Pentium II 266. All computations were performed in Matlab.

[4]For additional accuracy, the Markov chain was extended to 21000 iterations, instead of the 11000 used in the Monte Carlo study.

similar. We conclude that even with parameters designed to overstate the magnitude of the discretization bias likely to be found in real data, a modestly small value of h is sufficient to eliminate almost all of that bias.

4.2 Stochastic volatility

A more challenging statistical problem is to infer the parameters of a process that is not directly observable. In a typical stochastic volatility model, the price of an asset is assumed to follow a generalized geometric Brownian motion in which the mean parameter is constant but the variance parameter is itself a separate stochastic process. This second diffusion, which is assumed to be unobservable, is driven by a Brownian motion that may or not be correlated with that of the price process.

This section will examine an approximation of the model of Scott (1987), who models price and volatility dynamics as the following bivariate diffusion process:

$$dS_t = \mu S_t dt + e^{Y_t} S_t dB_t^{(1)}$$

$$dY_t = (\alpha + \delta Y_t)\, dt + \sigma dB_t^{(2)}$$

By Itô's lemma, we may rewrite the price dynamics in terms of log prices as

$$d(\log S_t) = (\mu - \frac{1}{2} e^{2Y_t}) dt + e^{Y_t} dB_t^{(1)}. \tag{10}$$

Following Duffie and Singleton (1988), we work with the simplification

$$d(\log S_t) = \mu dt + e^{Y_t} dB_t^{(1)}.$$

Although not pursued here, estimation of the correct form (10) is not problematic, requiring a slight adjustment to the calculations below.

The Euler approximation of this model may be written as

$$\log S_{k+1} = \log S_k + h\mu + \sqrt{h} e^{Y_k} \epsilon_{k+1}^{(1)} \tag{11}$$

$$Y_{k+1} = Y_k + h\,(\alpha + \delta Y_k) + \sqrt{h}\sigma \epsilon_{k+1}^{(2)}, \tag{12}$$

where $\epsilon_{k+1}^{(1)}$ and $\epsilon_{k+1}^{(2)}$ are standard normals.[5]

For simplicity, we assume independence of the Brownian motions $B^{(1)}$ and $B^{(2)}$. Then conditional on the observation of the complete data set, which includes the high frequency values of S_k and all values of Y_k, the parameters of each equation may be estimated independently. Consideration of correlated processes can be handled in the manner suggested by Jacquier, Polson, and Rossi (1998) for discrete time models.

The log price equation in (11) has μ as its only parameter. Given the augmented data set, we can rewrite the equation in standard regression form as

$$\frac{\log S_{k+1} - \log S_k}{\sqrt{h}\exp(Y_k)} = \frac{\sqrt{h}}{\exp(Y_k)}\mu + \epsilon_{k+1}^{(1)}. \tag{13}$$

Given the diffuse prior $p(\mu) \propto 1$, the conditional distribution of μ is normal with mean and variance identical to those of the sampling distribution of the ordinary least squares estimator of μ.

We can similarly rearrange the log volatility equation as

$$\frac{Y_{k+1} - Y_k}{\sqrt{h}} = \sqrt{h}\alpha + \sqrt{h}Y_k\delta + \sigma\epsilon_{k+1}^{(2)}. \tag{14}$$

Standard diffuse priors $p(\alpha, \delta, \sigma) \propto \frac{1}{\sigma}$ imply the following conditional distributions for α, δ, and σ:

- Let $\hat{\alpha}$ and $\hat{\delta}$ denote the OLS slope coefficients of a regression of $\frac{Y_{k+1}-Y_k}{\sqrt{h}}$ on $\sqrt{h}$ and $\sqrt{h}Y_k$, and let $\hat{\sigma}$ be the standard error of that regression. Then σ has an inverted gamma posterior with gamma parameter $2/\left((K-1)\hat{\sigma}^2\right)$ and alpha parameter $\frac{K-1}{2}$, where K is the size of the augmented sample.

[5] As Duffie and Singleton (1988) note, the process does not satisfy Lipschitz conditions due to the unbounded derivative of e^Y as $Y \to \pm\infty$. A standard sufficient condition for the weak convergence of the Euler approximation is therefore not met. Our solution is similar to the one they propose: truncate the volatility term e^Y at high and low values of Y, i.e., let the instantaneous standard deviation of $d(\log S)$ be equal to $e^{\underline{Y}}$ if $Y \le \underline{Y}$, e^Y if $\underline{Y} < Y < \overline{Y}$, and $e^{\overline{Y}}$ if $Y > \overline{Y}$. If $\underline{Y}$ and $\overline{Y}$ are chosen, respectively, low and high enough, then the truncation is irrelevant.

- Conditional on σ, (α, δ) has a bivariate normal posterior with mean $(\hat{\alpha}, \hat{\delta})$ and covariance equal to the standard OLS covariance matrix multiplied by $\frac{\sigma^2}{\hat{\sigma}^2}$.

Rather than using a bivariate candidate generating density to draw $\log S_k$ and Y_k simultaneously, as suggested in section three, we further increase the blocking of the Markov chain and draw $\log S_k$ and Y_k separately for each k. This choice is motivated by the availability of an analytic representation of the conditional distribution

$$p(\log S_k | \log S_{k-1}, \log S_{k+1}, Y_{k-1}, Y_k, Y_{k+1}, \phi)$$

derived in appendix A, where $\phi = (\mu, \alpha, \delta, \sigma)$.

For drawing elements of the path of the log volatility process Y, the Metropolis-Hastings algorithm is still a necessity. We follow the previous sections in drawing a candidate Y_k based on the previous point Y_{k-1}. For $k = 1$, however, we must draw from

$$p(Y_1 | \log S_1, \log S_2, Y_2, \phi) \tag{15}$$

instead (since S_1 should be observed), which is proportional to product of the unconditional distribution of the process and a one-period transition probability, or

$$p(Y_1 | \phi) p(\log S_2, Y_2 | \log S_1, Y_1, \phi).$$

The unconditional distribution $p(Y|\phi)$ is therefore used as the candidate generating density for Y_1. Because Y follows an Ornstein-Uhlenbeck process, its stationary distribution is Gaussian with mean $-\frac{\alpha}{\delta}$ and variance $-\frac{\sigma^2}{2\delta}$.

4.2.1 Application to S&P Composite Returns

The model was estimated on 521 weekly S&P 500 Composite returns constructed from Wednesday closing prices between January 1987 and December 1996. Various degrees of discretization were considered. First, we generate posteriors by augmenting only with volatility and not with higher frequency price data ("h=1"). Next, we increase the frequency of the Euler approximation, which requires us to additionally augment using up to five subperiods per observed data point ("h=.2").

One million iterations of the Markov chain were performed, with the first 50,000 discarded to negate the effects of initial conditions. Posterior means and standard deviations are reported in

Table 2: Stochastic Volatility Parameters

	μ	α	δ	σ
Posterior mean	0.00331	-0.842	-0.203	0.273
Posterior standard deviation	0.00068	0.528	0.126	0.097
Standard error of posterior mean	0.00001	0.017	0.004	0.004

table 2 for the case of $h = .2$. To assess numerical accuracy, batch means of 190 blocks of size 5,000 were computed. The batch means are plotted in figure 5, and it is evident that the block size is large enough so that these batch means are approximately serially uncorrelated.[6] The standard deviation of these means may therefore be used to calculate the standard error of the posterior mean in the usual way.

Table 2 reports posterior means, standard deviations, and standard errors of the posterior means. Consistent with the vast amount of research in time-varying volatility, we find volatility to be highly persistent, with the posterior mean for δ implying a daily autocorrelation of around .96. The posterior means are estimated with great accuracy, with numerical standard errors at least 30 times smaller than the respective posterior standard deviations.

The full posterior densities of the four parameters of the model are shown in figure 6. Again, posteriors are plotted using the Rao-Blackwell procedure described previously. Three out of the four posteriors appear highly non-normal, with pronounced skewness, implying that classical asymptotic approximations could be inaccurate.

In addition, the graphs reveal that discretization bias does exist – the "naive" posteriors ("h=1") differ noticeably from the HFA posteriors (such as "h=.2"), except for the parameter

[6]The batch means of α, δ, and σ have serial correlations less than .02 in absolute value. The batch means of μ have a serial correlation of .19, implying that its estimated standard error is likely somewhat understated.

μ, the mean return. The posteriors constructed using augmented higher frequency data are generally more dispersed than those that do not augment with more frequent data, and this dispersion is towards larger absolute values. Furthermore, a relatively small number of subintervals appears to erase most of this bias.[7]

Figure 7 plots the mean and 95% highest posterior density intervals of the posteriors for the latent volatility process (e^Y) for each week of the sample.[8] We find the time series average of the weekly volatility posterior means to be about 1.75 percent per week, or .8 percent daily. Time variation is clearly present (the October 1987 crash is very prominent), but the width of the 95% HPD intervals is large, making precise statements about the location of the latent volatility process difficult. In fact, on 485 out of the 521 weeks of the sample, the 95% HPD interval for volatility includes the time series average of 1.75 percent. We conclude that although time variation is clearly present in the data, identifying particular periods in which volatility is relatively high or low is not as straightforward as one might imagine.

4.3 An interest rate process with rounding

In the presence of bid-ask spreads, price discreteness, or other microstructure effects observable in high frequency data, it is apparent that either the model or the data must not be taken literally. Some seven-day Eurodollar rates, for example, are typically quoted in sixteenths, even when annualized rates are as low as three percent, as they were in the early 1990s. The biases induced by rounding and the construction of more robust classical estimators are the subjects of papers by Cho and Frees (1988), Gottlieb and Kalay (1985), and Ball (1988), among others. These papers largely deal with the estimation of simple and geometric Brownian motion, however, and none provide a way to avoid rounding bias for more general processes. In this section we suggest a general approach to estimating diffusion processes in the presence of rounding.

[7] Using a subsample we verify that higher frequency Euler approximations to not noticeably shift the posteriors any further.

[8] 95% HPD intervals of $\exp(Y_k)$ are calculated as follows. Let Y_k^i, $i = 1, \ldots, N$ denote N posterior draws of Y_k sorted in ascending order. Then (Y_k^i, Y_k^j) will be the HPD interval if i and j solve $\min_{i,j} Y_k^j - Y_k^i$ s.t. $j - i \geq .95N$.

Furthermore, it should be possible to extend the approach to account for more complicated models that account for a bid-ask spread, such as those proposed by Hasbrouck (1998).

4.3.1 A rounding correction

Instead of taking recorded interest rates as observations, without error, of a diffusion process, we may alternatively interpret them as specifying the *approximate* location of the diffusion at that time. A simple working hypothesis might be that the observed rate is simply the tick that is closest to the true value of the process. Suppose that the tick size is known to be $\frac{1}{16}$. Then we may believe that an observed rate of $4\frac{1}{16}$ percent actually implies that the true value of the process fell somewhere inside the interval $4\frac{1}{32}$ to $4\frac{3}{32}$. The HFA procedure can then be used to augment the observed rounded data with the unobserved "true" diffusion. This requires only a slight generalization of the Markov chain used thus far.

Let $\mathbf{R}$ denote the rounded process, $\mathbf{X}$ the true unobserved process, and ϕ the set of model parameters. Our goal is to construct a Markov chain for ϕ and $\mathbf{X}$ whose marginal distribution on ϕ will converge to $p(\phi|\mathbf{R})$. Since the diffusion is never actually observed, $\mathbf{X^o}$ is empty and $\mathbf{X} = \mathbf{X^u}$.

Let T_R denote the set of times for which the interest rate is observed (i.e. for which R_k is not null). For times kh in T_R, $\mathbf{R}$ provides a "prior" about the location of the diffusion X_k. In this case we know by the Markov property that

$$p(X_k|\mathbf{X_{-k}}, \mathbf{R}, \phi)$$

simplifies to[9]

$$p(X_k|X_{k-1}, X_{k+1}, R_k, \phi).$$

By Bayes rule, this is proportional to

$$p(X_{k+1}|X_k, R_k, \phi)p(X_k|X_{k-1}, R_k, \phi).$$

[9]Previous and future values of R are irrelevant since they are just noisy versions of previous and future values of X, and the Markov property implies that the only relevant values of X are X_{k-1} and X_{k+1}.

The first conditional density is simplified by noting the deterministic relationship between X_k and R_k. Then applying Bayes rule to the second conditional density and again noting the deterministic relationship between X_k and R_k, we conclude with the result that

$$p(X_k|\mathbf{X_{-k}}, \mathbf{R}, \phi) \propto p(X_{k+1}|X_k, \phi)p(X_k|X_{k-1}, \phi)p(R_k|X_k).$$

The result is identical to that of the previous sections except for the addition of the "prior" $p(R_k|X_k)$, which is proportional to unity where $|R_k - X_k|$ is less than half of one tick and is zero elsewhere. It is easily incorporated into the Metropolis-Hastings chain as follows: Draw a candidate X_k^* to replace the current draw X_k from the distribution $p(X_k|X_{k-1}, \phi)$. Accept this draw with probability

$$\alpha(X_k, X_k^*) = \min\{\frac{p(X_{k+1}|X_k^*, \phi)p(R_k|X_k^*)}{p(X_{k+1}|X_k, \phi)p(R_k|X_k)}, 1\}.$$

Since the acceptance probability will be equal to zero if X_k^* and R_k are not within half a tick, we are guaranteed that the augmented diffusion process will be consistent with the discretely observed rounded data $\mathbf{R}$. This also implies that it will always be the case that for the previous draw $p(R_k|X_k) = 1$. This term is therefore superfluous.

For times kh that are not in T_R, the same logic holds except that R_k is null. We still draw candidates X_k^* to replace the current draw X_k from the distribution $p(X_k|X_{k-1}, \phi)$. The acceptance probability, however, simplifies to

$$\alpha(X_k, X_k^*) = \min\{\frac{p(X_{k+1}|X_k^*, \phi)}{p(X_{k+1}|X_k, \phi)}, 1\},$$

which of an identical form to that of the previous sections.

4.3.2 The interest rate process

Chan, Karolyi, Longstaff, and Sanders (1992) estimated models of the short rate of the form

$$dx_t = (\alpha + \beta x_t)dt + \sigma x_t^{\gamma} dB_t \tag{16}$$

This specification encompasses a number of interest rate models, most notably those of Vasicek (1977) and Cox, Ingersoll, and Ross (1985).

The Euler approximation of this process[10] is given by

$$x_{k+1} = x_k + h\alpha + h\beta x_k + \sqrt{h}\sigma x_k^{\gamma}\epsilon_{k+1}.$$

Although it is possible to estimate γ, for simplicity of exposition we fix γ at the value 1.5, close to the estimate of Chan *et al.* The discretization then suggests a straightforward approach. After some rearrangement of terms, we are left with

$$\frac{x_{k+1} - x_k}{\sqrt{h}x_k^{1.5}} = \alpha\sqrt{h}x_k^{-1.5} + \beta\sqrt{h}x_k^{-.5} + \sigma\epsilon_{k+1}, \tag{17}$$

again putting us in a standard linear regression framework. If we assume the flat prior

$$p(\alpha, \beta, \sigma) \propto \frac{1}{\sigma},$$

then, as before, σ will have an inverted gamma distribution and (α, β) will have a bivariate normal distribution that is conditional on the draw of σ.

4.3.3 Experimental results

One thousand days of interest rate data were constructed by simulating the Euler approximation of the interest rate diffusion process (16) with $h = .04$ and then rounding off to the nearest sixteenth of one percent. The annualized parameter values used were $\alpha = .008$, $\beta = -.1$, and $\sigma = .75$, with 244 days per year. The parameter γ is fixed at and assumed known equal to 1.5. The simulated data are plotted in figure 8.

We consider both the rounded and the underlying unrounded data and seek to answer two questions: (1) How does the presence of rounding shift the parameter posteriors? (2) Will the rounding correction described above correct the bias induced by rounding, or will the loss of information that rounding entails make inference about the true parameters difficult even with the correction?

[10] Again, the process does not satisfy Lipschitz conditions due to the unbounded derivative of x^{γ}. If $\gamma > 1$, as we will have in this case, the derivative is unbounded as $x \to \infty$. We solve the problem by truncating the diffusion function, but at a sufficiently high level (interest rates of 500 percent) such that the truncation is irrelevant.

Figure 9 shows posteriors for each of the three parameters. Posteriors were calculated with $h = .1$, which is larger than the value used in the simulation but is sufficiently small to eliminate the discretization bias. In each panel, three posteriors are shown: one based on the unrounded diffusion data (solid line), one based "naively" on the rounded data (dashed line), and one that uses the rounded data but implements the rounding correction (dotted line).

For α and β, rounding has a relatively minor effect on posteriors, shifting both slightly away from zero. The posteriors that correct for rounding are virtually indistinguishable from those based on the unrounded data – rounding causes no loss of information relevant for inference about α and β.

For σ, rounding causes a major bias. The posterior that is based on rounded data is shifted towards significantly higher values, consistent with the observations of Cho and Frees (1988) and Ball (1988), among others, for the case of geometric Brownian motion. The rounding correction does not restore the "idealized" posterior perfectly, as it did for α and β, suggesting that rounding does cause a loss of information about the location of σ. Nevertheless, the rounding correction yields a vast improvement over the naive processing of rounded data.

In the previous examples, convergence of the Markov chain was investigated informally, either through the use of a Monte Carlo experiment (table 1) or an inspection of batch means (figure 5). While no evidence pointed to the absence of convergence, a more formal diagnostic may be desired. We may assess convergence using the method of Gelman and Rubin (1992), who propose to use a variance ratio as an indicator of convergence.

The method requires that a number of independent Markov chains be simulated. We simulate 30 chains in parallel up to a length of 4500 iterations. To check for convergence by iteration n, the within-chain variance of the first n iterations is compared to the between-chain variance of the same draws. The precise construction of this variance ratio may be found in Gelman and Rubin (1992).

A variance ratio near one indicates that the chain has approximately converged, as each chain is providing the same estimate of the target density. For the current example, these ratios are plotted in figure 10, and the results are encouraging. For the drift parameters α and

β, convergence appears to be very quick, with only minute deviations from one after about 1000 iterations. For σ, convergence is somewhat slower, with the variance ratio decreasing from about 1.4 to 1.02 after 3500 iterations. The results suggest that the apparent convergence in the previous examples could be established more formally were multiple Markov chains to be simulated.

5 Extension to jump-diffusions

It is well-documented (see, for example, Jorion, 1988, or Ball and Torous, 1985) that the price continuity implied by diffusion models is a poor description of some asset price time series. An alternative to the pure diffusion process is the jump-diffusion. This section will illustrate how Poisson-type jumps can be added to the diffusion models considered previously.

For simplicity, however,we focus on univariate processes in which the process represents the price of a traded asset and hence must be non-negative. The jump diffusion can be represented by the following stochastic differential equation:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t + X_t dq_t, \tag{18}$$

where $\mu(X_t)$ and $\sigma(X_t)$ are the same as before and dq_t is the increment of a Poisson jump process with lognormal jumps[11] and constant jump intensity λ. Conditional on a jump occurring, the increment dq_t is of size Y_t, making the impact on the increment dX_t equal to $X_t Y_t$. We make the standard assumption that

$$\log(1 + Y_t) \sim N(\theta, \delta^2).$$

This combination of assumptions insures that the process X_t will remain nonnegative.

Following Ball and Torous (1983), we use a Bernoulli jump process to generate a convergent discrete approximation of the Poisson component. Combining this with the Euler approximation considered thus far, our discrete-time approximation of the Poisson jump-diffusion model

[11] In cases in which X_t is not a traded asset price but is instead some other process that may become negative, jumps that are normal, not lognormal, may be considered.

is given by

$$X_{k+1} = X_k + h\mu(X_k) + \sqrt{h}\sigma(X_k)\epsilon_{k+1} + \iota_{k+1}X_kY_{k+1}, \tag{19}$$

where $Y_{k+1} = \exp(\theta + \delta\eta_{k+1}) - 1$, ϵ_k and η_k are independent standard normals, and ι_k is an independent Bernoulli random variable that takes the value 1 with probability λh and 0 otherwise.

The last term, $\iota_{k+1}X_kY_{k+1}$, represents the potential jump in the process X, with $\iota_k = 1$ indicating that there was a jump between observations $k-1$ and k. Conditional on there being a jump, it's magnitude does not depend on the discretization interval h; only the probability of jumping declines with h.

The simplifying assumption of the Bernoulli over the Poisson jump process is the restriction that no more than one jump can occur over any period of length h in the Bernoulli model, while multiple jumps are possible in the Poisson model. As Ball and Torous note, as $h \to 0$ the probability of more than one jump per interval shrinks to zero at a faster rate than the probability of zero or one jumps per interval. Because h is a parameter under our control, the difference between the Bernoulli and Poisson models may therefore be made arbitrarily small.

5.1 The jump-diffusion Markov chain

The Markov chain used will be a modification of the one developed previously. As before, we will adopt an algorithm of drawing one period of latent data at a time. The jump component adds another dimension to this draw, however, so that instead of drawing from

$$p(X_k|X_{k-1}, X_{k+1}, \phi)$$

when X_k is a latent data point, we will instead draw from

$$p(X_k, J_k|X_{k-1}, J_{k-1}, X_{k+1}, J_{k+1}, \phi),$$

where $J_k = (\iota_k, Y_k)$ and ϕ denotes the vector of parameters of both the diffusion and jump components.

As before, this expression can be rearranged using the Markov property of the jump-diffusion model as the product of two transition densities:

$$p(X_k, J_k | X_{k-1}, J_{k-1}, \phi) p(X_{k+1}, J_{k+1} | X_k, J_k, \phi)$$

Furthermore, because the jump arrivals are independent of one another, this distribution simplifies to

$$p(X_k, J_k | X_{k-1}, \phi) p(X_{k+1}, J_{k+1} | X_k, \phi).$$

As in the pure diffusion case, this distribution is generally intractable, but the rearrangement again suggests a simple procedure: draw X_k and $J_k = (\iota_k, Y_k)$ from $p(X_k, J_k | X_{k-1}, \phi)$, a normal/Bernoulli/lognormal draw, and accept the draw in a Metropolis-Hastings step with an acceptance probability based on the remaining term, $p(X_{k+1}, J_{k+1} | X_k, \phi)$.

By Bayes rule, this remaining term is proportional to $p(X_{k+1} | J_{k+1}, X_k, \phi) p(J_{k+1} | X_k, \phi)$. Since J_{k+1} is independent of X_k, it reduces to $p(X_{k+1} | J_{k+1}, X_k, \phi)$, a Gaussian density. The probability of accepting a new draw (X_k^*, J_k^*) over the previous draw (X_k, J_k) will therefore be equal to

$$\min\{\frac{p(X_{k+1} | J_{k+1}, X_k^*, \phi)}{p(X_{k+1} | J_{k+1}, X_k, \phi)}, 1\}.$$

This provides a method for drawing X_k and J_k when both are unobserved. When X_k is observed, however, it is still the case that J_k is not – even if we observe the process we still don't observe the jumps. Therefore, some draws of the latent data will be from

$$p(J_k | X_{k-1}, J_{k-1}, X_k, X_{k+1}, J_{k+1}, \phi).$$

Because J_k describes the jump between X_{k-1} and X_k, most of these conditioning arguments are irrelevant, reducing the distribution to

$$p(J_k | X_{k-1}, X_k, \phi),$$

which is proportional to

$$p(X_k | X_{k-1}, J_k, \phi) p(J_k | X_{k-1}, \phi).$$

Since J_k is independent of X_{k-1}, this further reduces to

$$p(X_k|X_{k-1}, J_k, \phi)p(J_k|\phi).$$

As before, we use a Metropolis-Hastings step and draw J_k from $p(J_k|\phi)$ with a Bernoulli/lognormal draw. The remaining term, $p(X_k|X_{k-1}, J_k, \phi)$, is a normal density, and it is used to construct the probability of accepting J_k^* over J_k, which is given by

$$\min\{\frac{p(X_k|J_k^*, X_{k-1}, \phi)}{p(X_k|J_k, X_{k-1}, \phi)}.$$

The remaining block draw of the Markov chain is the draw of the parameters given the observed and augmented data. Given the full time series of both X_k and J_k, changes in X_k can be decomposed into a diffusion component and jump component, and the parameters of each may be considered separately. The diffusion component, $X_{k+1} - X_k - \iota_k X_k Y_{k+1}$, can be addressed as in the pure diffusion case previously developed.

Because jump arrivals are independent sources of randomness, the jump intensity parameter may be considered separately from other parameters, with a simple Binomial likelihood based on the time series of ι_k. Let K equal the augmented sample size minus one, which is the maximum number of jumps possible. Let N_1 equal the number of jumps, or $N_1 = \sum_{k=2}^{K} \iota_k$. Then the likelihood of λh is proportional to $(\lambda h)^{N_1}(1-\lambda h)^{K-N_1}$, which is itself proportional to a standard beta density for λh. A conjugate beta prior on λh generates a corresponding beta posterior.

Inference about the parameters underlying the jump mean and variance parameters θ and δ^2 follows straightforward analysis, since the time series of $\log(1+Y_k)$ is i.i.d. normal. (Of course, Y_k is only known conditional on a jump occurring, or when $\iota_k = 1$.) Flat or normal/inverted gamma priors on θ and δ^2 result in the standard student-t/inverted gamma posterior for these parameters.

It is important to recognize a danger in likelihood-based analysis of jump-diffusions, as pointed out by Honoré (1998). It is that the likelihood function is potentially unbounded if δ or $\sigma(X_t)$ are allowed to become arbitrarily small. Honoré alleviates this problem by restricting his maximum likelihood parameter grid search to a compact interval that does not include

parameters that would allow either δ or $\sigma(X_t)$ to reach zero. We may effectively follow the same prescription here by imposing zero prior probability on parameter values that allow δ or $\sigma(X_t)$ to become smaller than some prespecified values. This will result in a truncated posterior that may easily be sampled from using accept-reject (see Box and Tiao, 1973, p. 67-69).

5.2 An application to the dollar/mark exchange rate

In an analysis of jumps in the foreign exchange market, Jorion (1988) found that the presence of jumps in the U.S dollar/German mark exchange rate could explain a significant amount of the leptokurtosis observed in exchange rates. In this section we will reestimate his model over the same sample period, 1974-85, as well as an updated sample period, 1974-1997. Jorion found that jumps were most beneficial in explaining weekly movements, so that is the data frequency examined here as well.

The model considered by Jorion is a simple combination of geometric Brownian motion and lognormally distributed Poisson jumps:

$$dX_t = \mu X_t + \sigma X_t dB_t + X_t dq.$$

Given a discretization interval h, we chose the following uninformative prior for the five parameters:

$$p(\mu, \sigma, \theta, \delta, \lambda) \propto \left(\sqrt{\lambda h(1-\lambda h)}\sigma\delta\right)^{-1}$$

This prior combines standard flat priors on (μ, σ) and (θ, δ) with the uninformative prior for λh suggested by Box and Tiao (1973, p. 34-35). To avoid unboundedness in the likelihood functions, we modify the prior to assign zero probability to values of σ or δ below 10^{-4}, well below the values reported by Jorion.

Under these priors, conditional on the full augmented sample of X_k and J_k, we have the following conditional distributions for the five parameters:

- Let $\hat{\mu}$ be the slope and $\hat{\sigma}$ the standard error of a regression of

$$\frac{X_{k+1} - X_k - \iota_{k+1} X_k Y_{k+1}}{\sqrt{h} X_k}$$

on the constant $\sqrt{h}$. Then σ has an inverted gamma distribution (truncated at 10^{-4}) with gamma parameter $2/\left((K-1)\hat{\sigma}^2\right)$ and alpha parameter $\frac{K-1}{2}$.

- Conditional on σ, μ is normally distributed with mean $\hat{\mu}$ and variance $\frac{\sigma^2}{Kh}$.
- Let $\hat{\theta}$ be the sample mean and $\hat{\delta}^2$ the sample variance of $\log(1+Y_k)$. Then δ has an inverted gamma distribution with gamma parameter $2/\left((N_1-1)\hat{\delta}^2\right)$ and alpha parameter $\frac{N_1-1}{2}$, also truncated at 10^{-4}.
- Conditional on δ, θ is normally distributed with mean $\hat{\theta}$ and variance $\frac{\delta^2}{N_1}$.
- The jump probability λh has a standard beta distribution with parameters $N_1 + \frac{1}{2}$ and $K - N_1 + \frac{1}{2}$.

Table 3: Dollar/DM Exchange Rate Estimation

Posterior means and standard deviations (in parentheses)

	$\mu(\times 10^3)$	$\sigma(\times 10^2)$	λ	$\theta(\times 10^3)$	$\delta(\times 10^2)$
Subsample	-0.05	0.39	1.58	0.01	1.03
1974-85	(1.10)	(0.21)	(0.44)	(0.95)	(0.15)
Full sample	0.02	1.20	0.21	-2.44	2.27
1974-97	(0.50)	(0.23)	(0.25)	(4.38)	(0.67)

Results for the 1974-85 sample, reported in table 3, generally confirm those of Jorion (1988), although we find the jump intensity to be somewhat higher and the two mean parameters to be indistinguishable from zero. Parameters differ markedly between the subsample and full sample. The jump process implied by the full sample has much less frequent but significantly larger jumps, undoubtedly reflecting the fact that excess kurtosis is 1.83 for the whole sample, compared with just .91 for the 1974-85 subsample.

There appears to be a problem separating the effects of the different parameters influencing the mean return; μ and θ, for example, have a posterior correlation of $-.84$. An advantage of the present Bayesian procedure is that if we were interested in the total expected return per unit time, which is $\mu + \lambda\left(\exp(\theta + \delta^2/2) - 1\right)$, we could construct a posterior distribution of this quantity as well. It turns out that this total expected return has a mean of 2.56×10^{-4} and a standard deviation of 5.53×10^{-4} for the subsample and a mean of -2.40×10^{-4} and a standard deviation of 4.28×10^{-4} for the full sample. Although still indistinguishable from zero, inference about these total expected returns is still sharper than it is about the individual parameters.

It is also clear that many of the marginal posteriors are highly non-normal. In the full sample period, for instance, λ, which is strictly positive, has a mean of .21 and a standard deviation of .25, indicating a large amount of positive skewness in the posterior. One implication is that the asymptotic Gaussian sampling distribution relied on by frequentist estimators such as MLE may be inappropriate in finite samples. The Bayesian approach's finite sample inference may therefore offer a more realistic assessment of parameter uncertainty.

The main advantage of the method presented here, however, is that diffusions other than simple geometric Brownian motion may be combined with Poisson jump processes. The method may therefore be useful in estimating the more complicated jump-diffusion processes currently being studied in the options pricing literature.

6 Conclusion

The HFA procedure is a simple and general method for the analysis of a wide variety of diffusion processes commonly found in continuous time finance theory. It was shown to be easily extendible to jump-diffusions as well. In part because of its Bayesian nature, HFA is robust to a variety of features common in economic and financial data, such as nonstationarity, latent variables, small samples, and microstructure-induced measurement error.

The main complication of the method is the need to augment with many points of high fre-

quency data. The Gaussian nature of this discretization, however, makes the method tractable. This tradeoff results in a computational requirement that can be met in reasonable time on a desktop PC. This is in contrast to some computationally infeasible simulation-based frequentist methods (see Danielsson, 1994).

The application of HFA to several examples generates results that are consistent with conventional methods and in line with other empirical studies. A Monte Carlo study demonstrated that the method has good performance even when chains are relatively short. One of the advantages of the Bayesian nature of the procedure is demonstrated when the latent S&P volatility process is extracted from the time series of prices. This link between the uncertainty about parameters an volatility may be helpful in options hedging and risk management in general. Together with its robustness to certain microstructure effects and its relative ease in implementation, it is hoped that the method proposed in this paper will provide a more practical way to operationalize the use of continuous time models.

A The conditional distribution of $\log S_k$

In working with the Euler approximation of the stochastic volatility model of section 4.2, the advantage of working in terms of log prices is that the distribution

$$p(\log S_k | \log S_{k-1}, \log S_{k+1}, X_{k-1}, X_k, X_{k+1})$$

is Gaussian, while

$$p(S_k | S_{k-1}, S_{k+1}, X_{k-1}, X_k, X_{k+1})$$

is not. This section will derive the former distribution for the general case in which the innovations in the Euler approximation (11) have correlation ρ.

Section three demonstrated that

$$\begin{aligned} &p(\log S_k | \log S_{k-1}, \log S_{k+1}, X_{k-1}, X_k, X_{k+1}) \\ &\quad \propto p(\log S_k, X_k | \log S_{k-1}, X_{k-1}) p(\log S_{k+1}, X_{k+1} | \log S_k, X_k). \end{aligned}$$

Applying Bayes rule again, this is equal to

$$\begin{aligned} &p(\log S_k | \log S_{k-1}, X_{k-1}, X_k) p(X_k | \log S_{k-1}, X_{k-1}) \\ &\quad \cdot p(\log S_{k+1} | \log S_k, X_k, X_{k+1}) p(X_{k+1} | \log S_k, X_k). \end{aligned}$$

Now note that both $p(X_k | \log S_{k-1}, X_{k-1})$ and $p(X_{k+1} | \log S_k, X_k)$ are independent of $\log S_k$. For the former distribution, this observation is obvious. For the latter we rely on the fact that S_{k-1} does not enter either the mean or the variance of X_k. This implies that

$$\begin{aligned} &p(\log S_k | \log S_{k-1}, \log S_{k+1}, X_{k-1}, X_k, X_{k+1}) \\ &\quad \propto p(\log S_k | \log S_{k-1}, X_{k-1}, X_k) p(\log S_{k+1} | \log S_k, X_k, X_{k+1}). \end{aligned}$$

Now let

$$\begin{aligned} m_1 &= h\mu + \rho \frac{e^{X_{k-1}}}{\sigma} (X_k - X_{k-1} - h\alpha - h\delta X_{k-1}) \\ m_2 &= h\mu + \rho \frac{e^{X_k}}{\sigma} (X_{k+1} - X_k - h\alpha - h\delta X_k) \\ v_1 &= h e^{2X_{k-1}} (1 - \rho^2) \\ v_2 &= h e^{2X_k} (1 - \rho^2) \end{aligned}$$

Using standard results for bivariate normal random variables, we can see that

$$\begin{aligned}\log S_k|(\log S_{k-1}, X_{k-1}, X_k) &\sim N(\log S_{k-1} + m_1, v_1)\\ \log S_{k+1}|(\log S_k, X_k, X_{k+1}) &\sim N(\log S_k + m_2, v_2)\end{aligned}$$

The density of $\log S_k$ can therefore be written as

$$\begin{aligned}&f(\log S_k = y|\log S_{k-1}, \log S_{k+1}, X_{k-1}, X_k, X_{k+1})\\ &\quad\propto \exp\left(-\frac{1}{2}\frac{(y - \log S_{k-1} - m_1)^2}{v_1} - \frac{1}{2}\frac{(\log S_{k+1} - y - m_2)^2}{v_2}\right),\end{aligned}$$

which is proportional to

$$\exp\left(-\frac{1}{2}(Ay^2 + By)\right),$$

where

$$\begin{aligned}A &= \frac{1}{v_1} + \frac{1}{v_2}\\ B &= 2\left(-\frac{m_1 + Y_{k-1}}{v_1} + \frac{m_2 - Y_{k+1}}{v_2}\right)\end{aligned}$$

Completing the square, we conclude that

$$\log S_k|(\log S_{k-1}, \log S_{k+1}, X_{k-1}, X_k, X_{k+1}) \sim N\left(-\frac{B}{2A}, \frac{1}{A}\right).$$

References

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

[2] Aït-Sahalia, Y., 1998, "Maximum-Likelihood Estimation of Discretely Sampled Diffusions," working paper, University of Chicago.

[3] Bakshi, G., C. Cao, and Z. Chen, 1997, "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, 52, 2003-2049.

[4] Ball, C. A., 1988, "Estimation Bias Induced by Discrete Security Prices," *Journal of Finance*, 43, 841-865.

[5] Ball, C. A., and W. N. Torous, 1983, "A Simplified Jump Process for Common Stock Returns," *Journal of Financial and Quantitative Analysis*, 18, 53-65.

[6] Ball, C. A., and W. N. Torous, 1985, "On Jumps in Common Stock Prices and Their Impact on Call Option Pricing," *Journal of Finance*, 40, 155-173.

[7] Box, G. E. P., and G. C. Tiao, 1973, *Bayesian Inference in Statistical Analysis*, John Wiley, New York.

[8] Casella, G., and E. I. George, 1992, "Explaining the Gibbs Sampler," *American Statistician*, 46, 167-174.

[9] 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.

[10] Chib, S., and E. Greenberg, 1995, "Understanding the Metropolis-Hastings Algorithm," *American Statistician*, 49, 327-335.

[11] Chib, S., and E. Greenberg, 1996, "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, 12, 409-431.

[12] Cho, D. C., and E. W. Frees, 1988, "Estimating the Volatility of Discrete Stock Prices," *Journal of Finance*, 43, 451-466.

[13] Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-407.

[14] Danielsson, J., 1994, "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood," *Journal of Econometrics*, 64, 375-400.

[15] Duffie, D., and K. J. Singleton, 1988, "Simulated Moments Estimation of Diffusion Models of Asset Prices," working paper, Stanford University.

[16] Duffie, D., and K. J. Singleton, 1993, "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929-952.

[17] Elerian, O., S. Chib, and N. Shephard, 1998, "Likelihood Inference for Discretely Observed Non-Linear Diffusions," working paper.

[18] Eraker, B., 1998, "MCMC Analysis of Diffusion Models with Application to Finance," working paper, Norwegian School of Economics and Business Administration.

[19] Gallant, R., and G. Tauchen, 1996, "Which moments to match?," *Journal of Econometric Theory*, 12, 657-681.

[20] Gallant, A. R., and G. Tauchen, 1997, "Estimation of Continuous Time Models for Stock Returns and Interest Rates," working paper, University of North Carolina.

[21] Gelfand, A. E., and A. F. M. Smith, 1990, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398-409.

[22] Gelman, A., and D. B. Rubin, 1992, "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457-511.

[23] Geweke, J., 1996, "Simulation-Based Bayesian Inference for Economic Time Series," working paper, Federal Reserve Bank of Minneapolis.

[24] Gottlieb, G., and A. Kalay, 1985, "Implications of the Discreteness of Observed Stock Prices," *Journal of Finance*, 40, 135-153.

[25] Gourieroux, C., A. Monfort, and E. Renault, 1993, "Indirect Inference," *Journal of Applied Econometrics*, 8, S85-S118.

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

[27] Hasbrouck, J., 1998, "Security Bid/Ask Dynamics with Discreteness and Clustering: Simple Strategies for Modeling and Estimation," working paper.

[28] Honoré, P., 1998, "Pitfalls in Estimating Jump-Diffusion Models," working paper.

[29] Hull, J. C., and A. White, 1987, "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, 42, 281-300.

[30] 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.

[31] Jacquier, E., N. G. Polson, and P. E. Rossi, 1998, "Stochastic Volatility: Univariate and Multivariate Extensions," working paper.

[32] Jorion, P., 1988, "On Jump Processes in the Foreign Exchange and Stock Markets," *Review of Financial Studies*, 1, 427-445.

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

[34] Kloeden, P. E., and E. Platen, 1992, *Numerical Solution of Stochastic Differential Equation*, Springer-Verlag, New York.

[35] Liu, J. S., W. H. Wong, and A. Kong, 1994, "Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes," *Biometrika*, 81, 27-40.

[36] Pedersen, A. R., 1995, "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, 22, 55-71.

[37] Santa-Clara, P., 1995, "Simulated Maximum Likelihood Estimation of Diffusions with an Application to the Short-Term Interest Rate," working paper.

[38] Scott, L. O., 1987, "Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application," *Journal of Financial and Quantitative Analysis*, 22, 419-438.

[39] Tanner, M. A., 1996, *Tools for Statistical Inference*, Springer-Verlag, New York.

[40] 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.

[41] Tierney, L., 1994, "Markov Chains for Exploring Posterior Distributions," *Annals of Statistics*, 22, 1701-1762.

[42] Vasicek, O., 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177-188.

[43] Zellner, A., 1971, *An Introduction to Bayesian Inference in Econometrics*, John Wiley, New York.

Figure 1

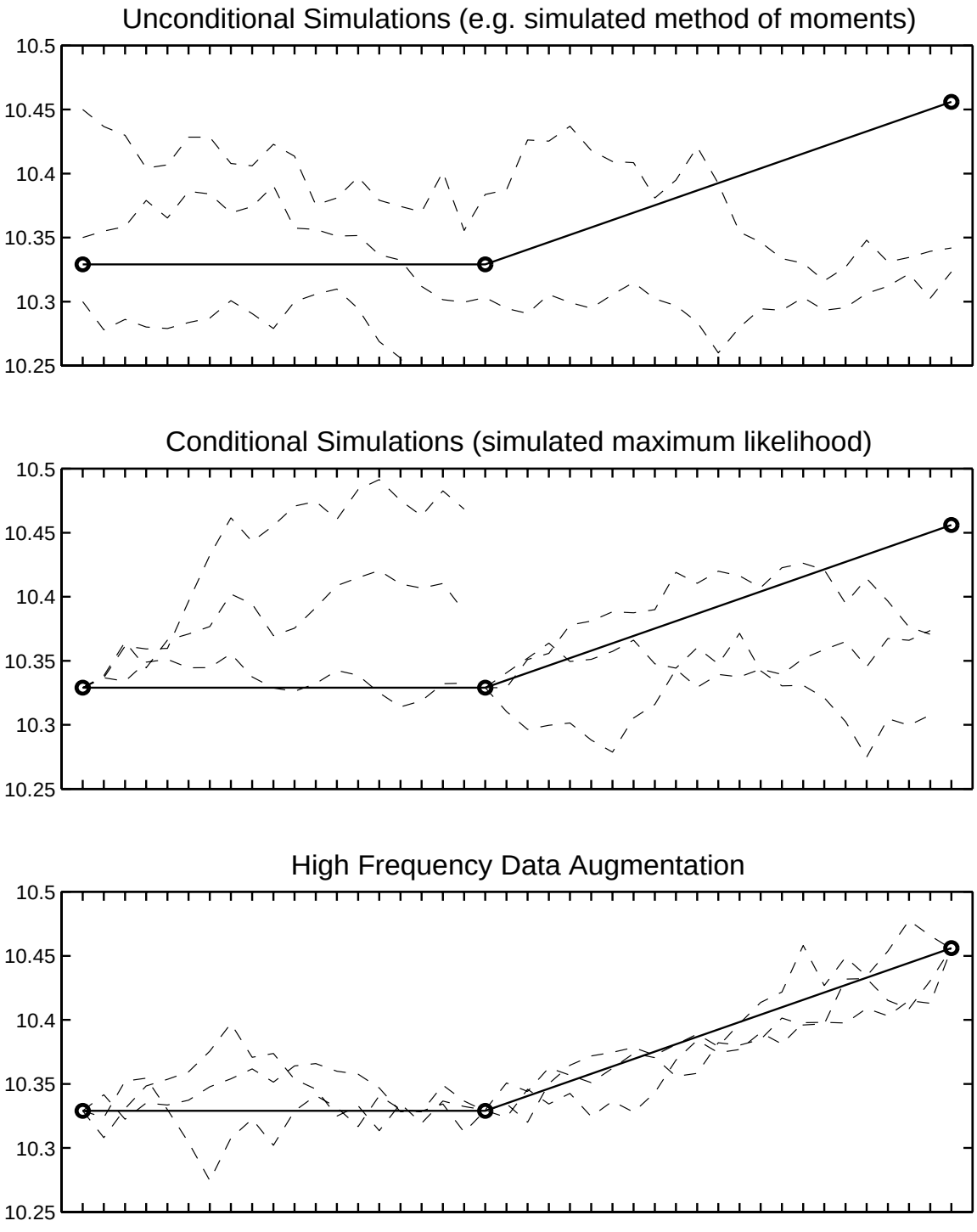
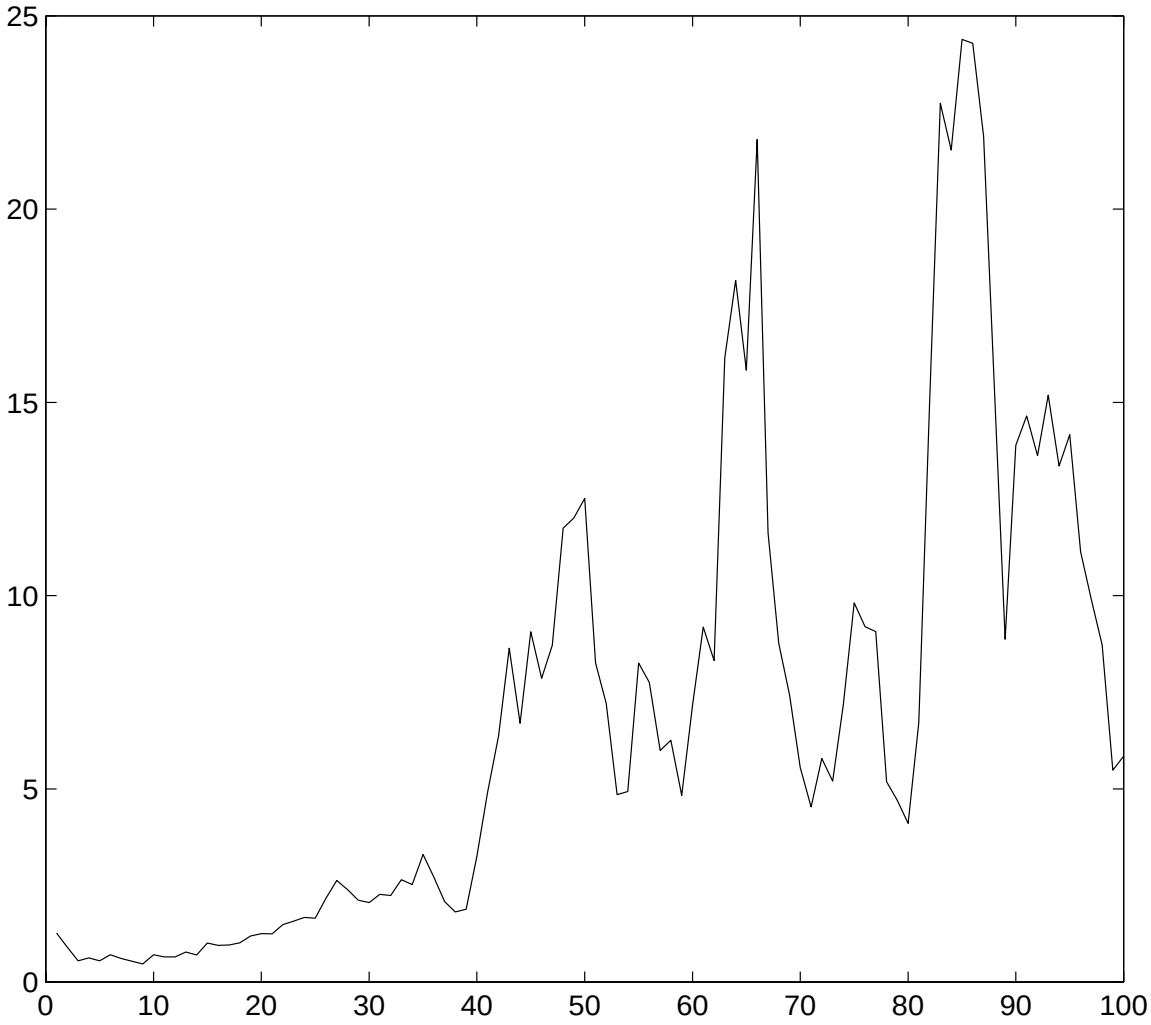
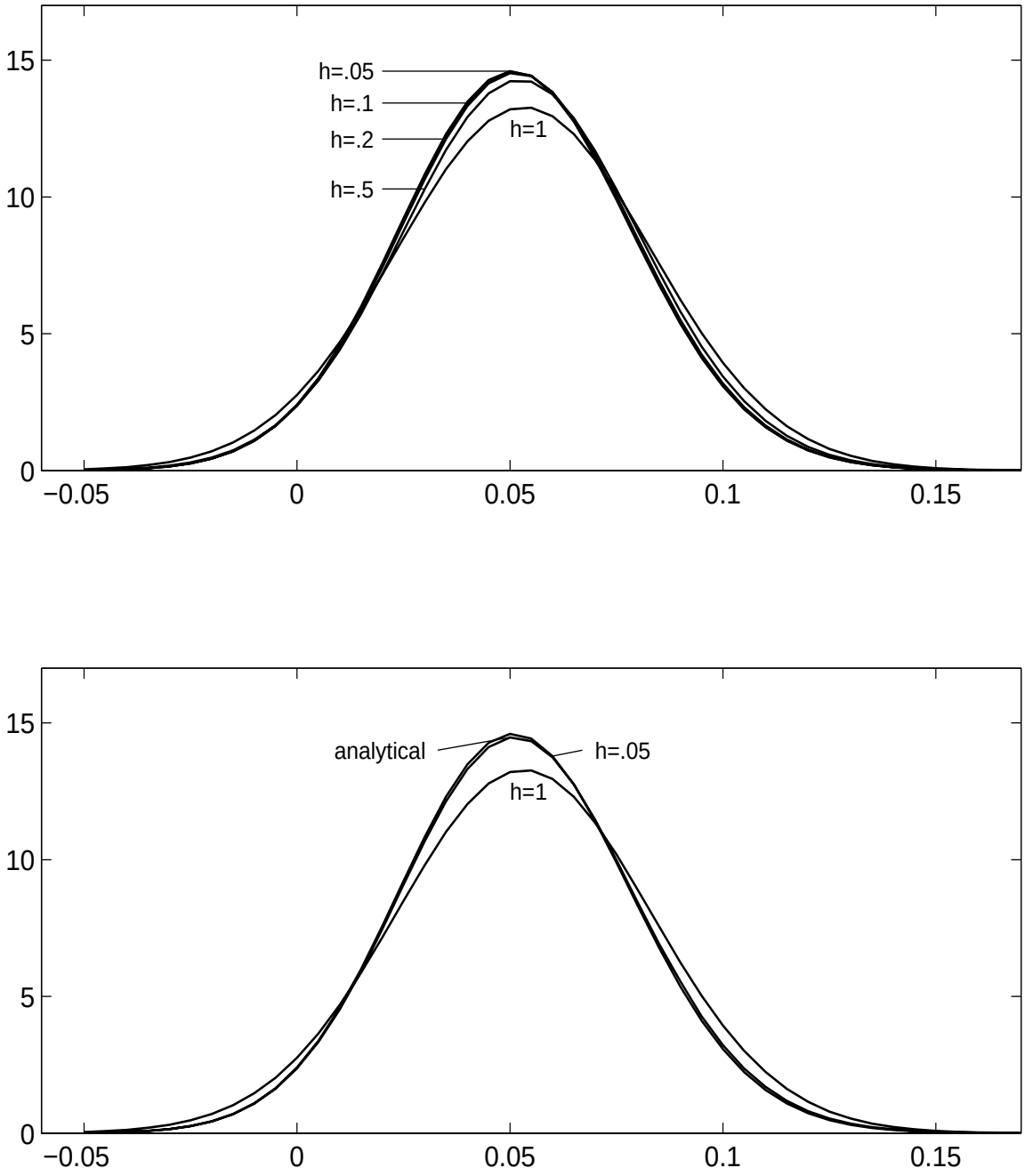

Unconditional Simulations (e.g. simulated method of moments)
10.5
10.45
10.4
10.35
10.3
10.25
Conditional Simulations (simulated maximum likelihood)
10.5
10.45
10.4
10.35
10.3
10.25
High Frequency Data Augmentation
10.5
10.45
10.4
10.35
10.3
10.25

Figure 2: Simulated Geometric Brownian Motion



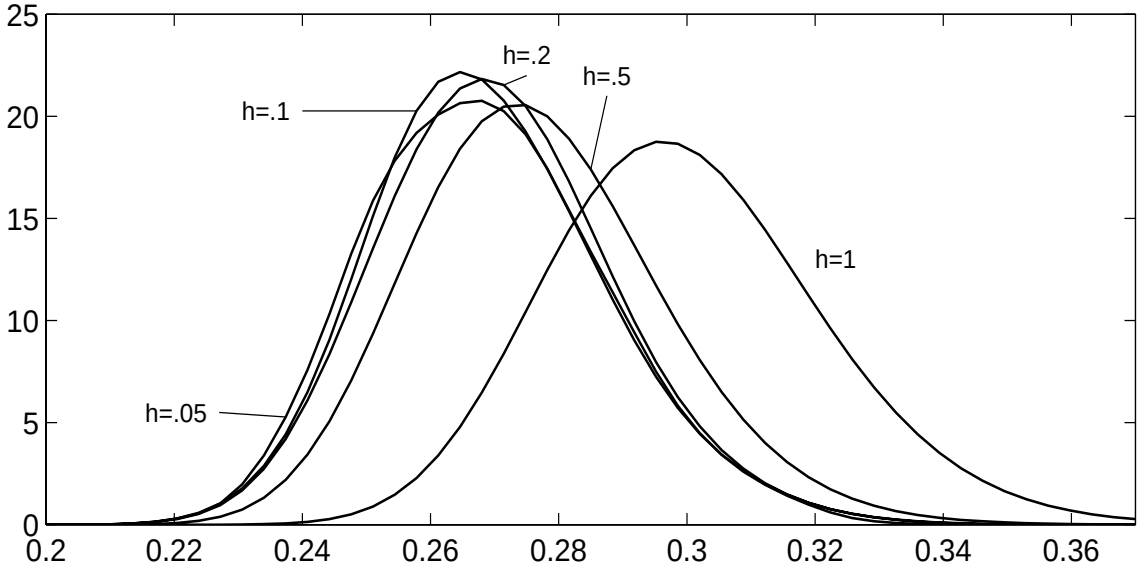
simulated with mu=.025 and sigma=.25

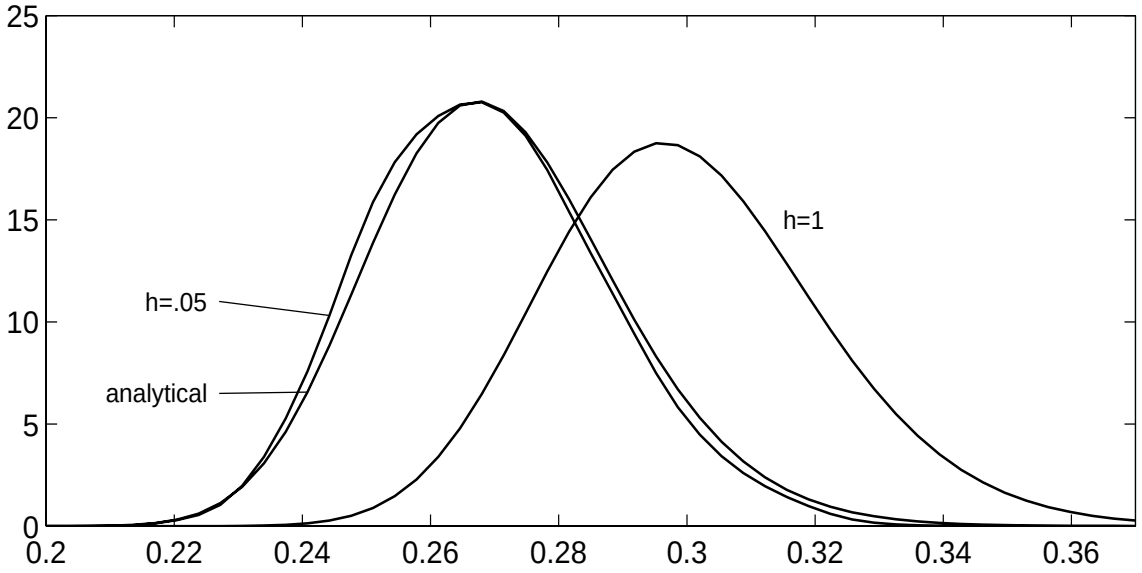
Figure 3: Posteriors for Mu (GBM)



The top panel plots posterior distributions of mu under different Euler discretizations, ranging from h=1 (no data augmentation) to h=.05. The bottom panel compares the finest and coursest discretizations with the posterior constructed using the analytical procedure.

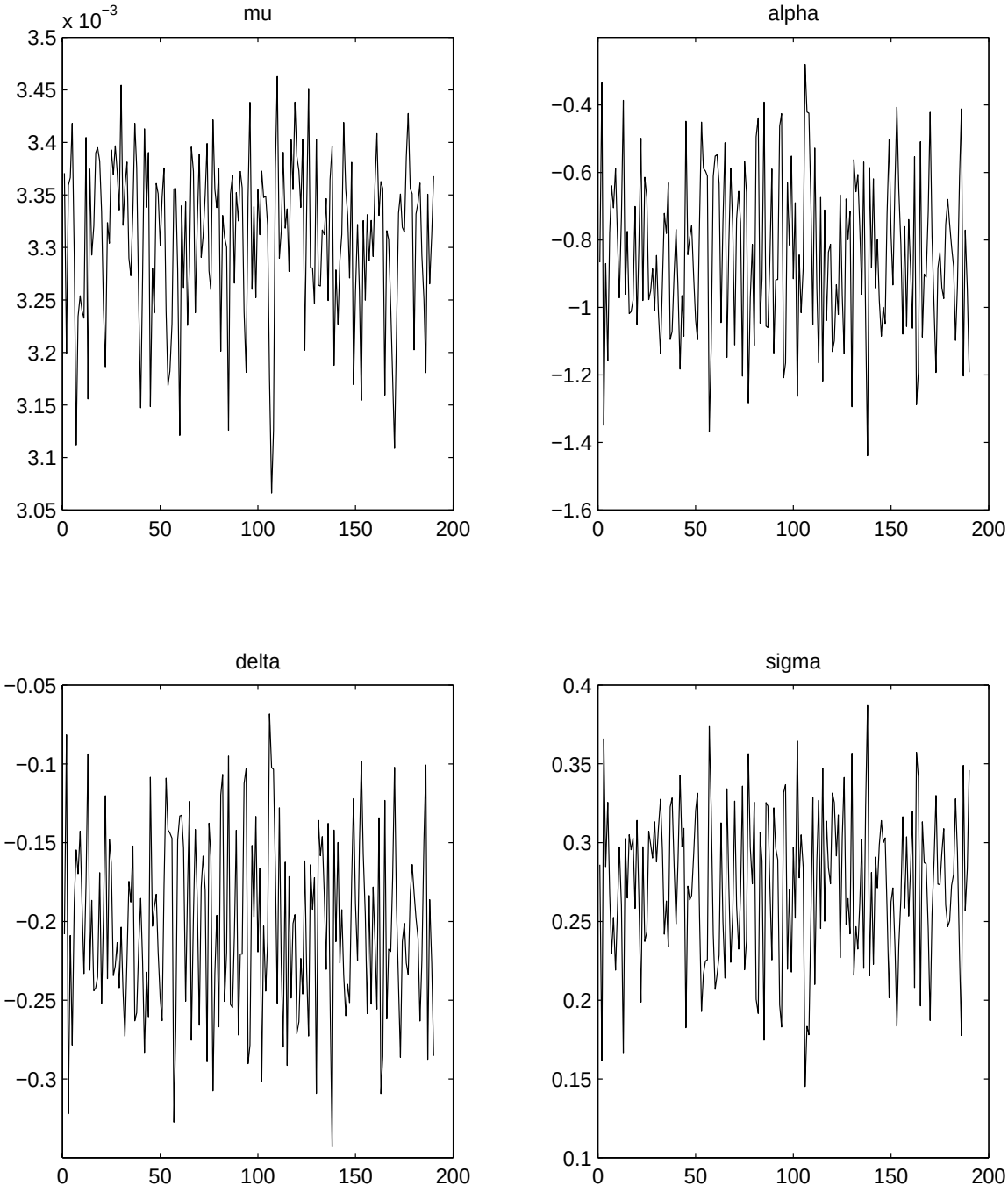
Figure 4: Posteriors for Sigma (GBM)





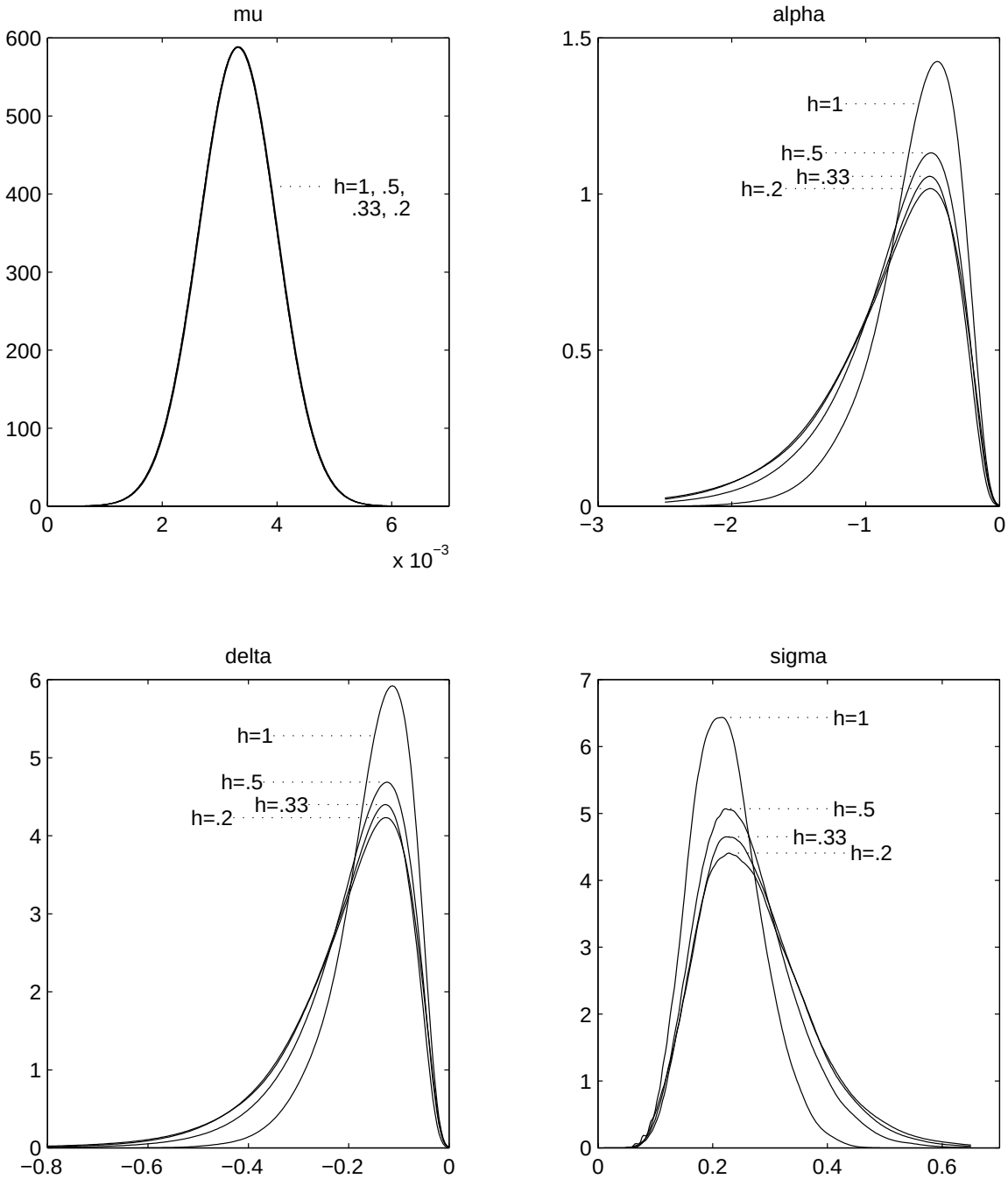
The top panel plots posterior distributions of sigma under different Euler discretizations, ranging from h=1 (no data augmentation) to h=.05. The bottom panel compares the finest and coursest discretizations with the posterior constructed using the analytical procedure.

Figure 5: Stochastic Volatility Parameter Batch Means



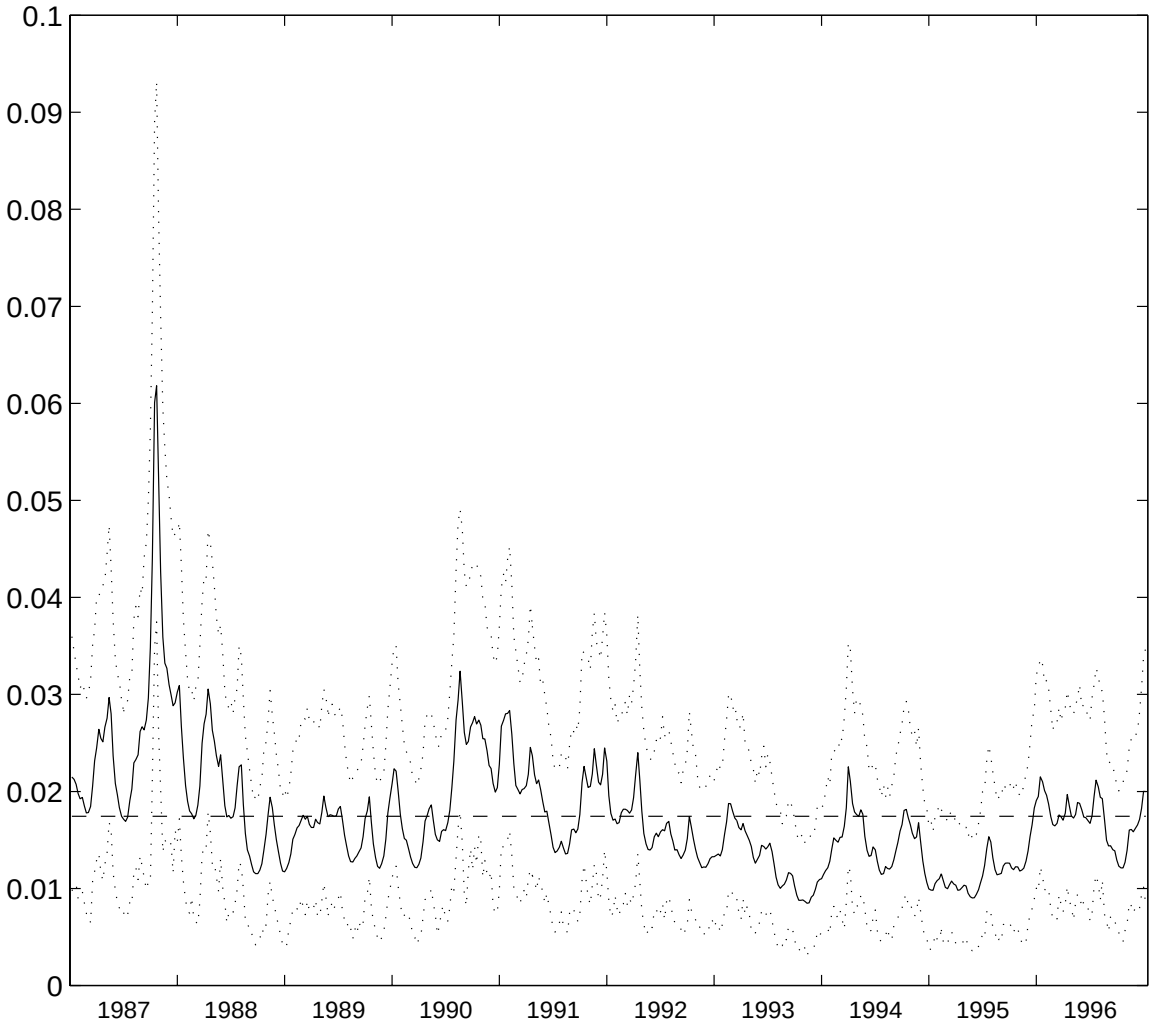
For each parameter, the Markov chain output draws were divided into batches of size 5000 and a mean was computed for each batch. These batch means are plotted in each panel.

Figure 6: Stochastic Volatility Posteriors



Each panel plots posterior distributions of the stochastic volatility model parameters under varying degrees of high frequency augmentation, ranging from augmentation with volatility data only (h=1) to augmentation with 5 subperiods per observed data point (h=.2).

Figure 7: S&P 500 Volatility Posteriors



For each week in the sample the mean of the posterior of the latent volatilty process is plotted (solid line) along with the upper and lower 95% HPD bounds (dotted lines). The dashed line plots the time series average of the weekly posterior means.

Figure 8A: Simulated Interest Rate Diffusion Data

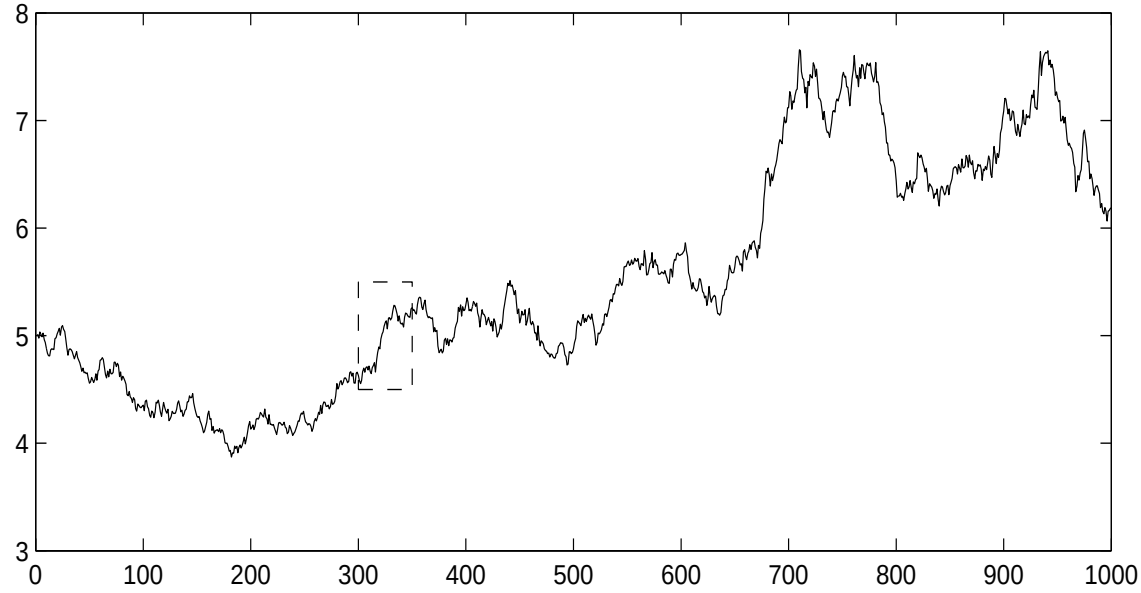
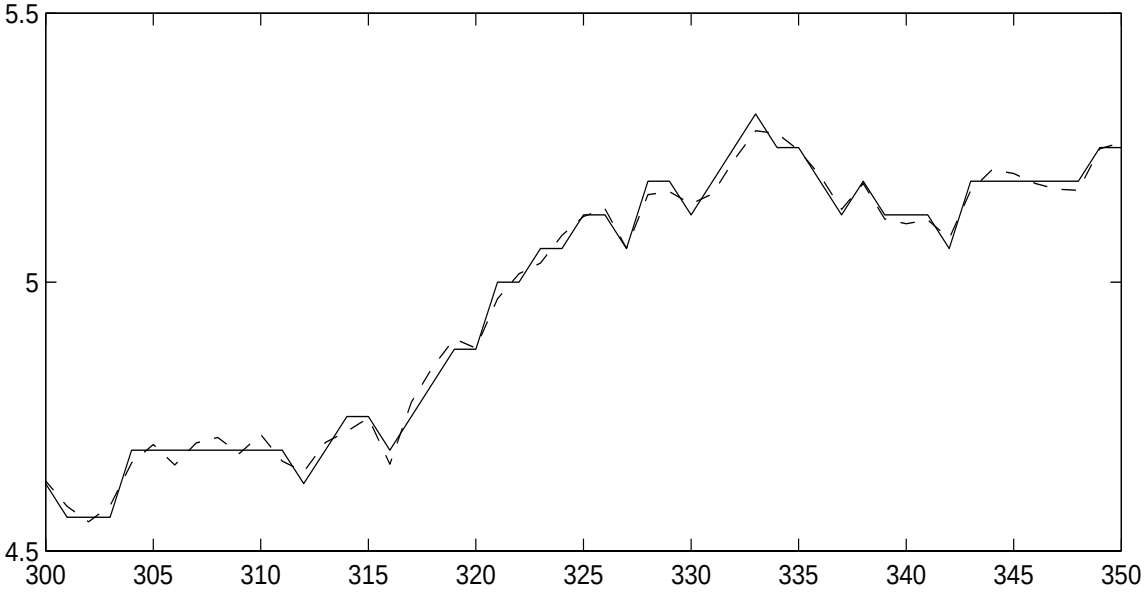
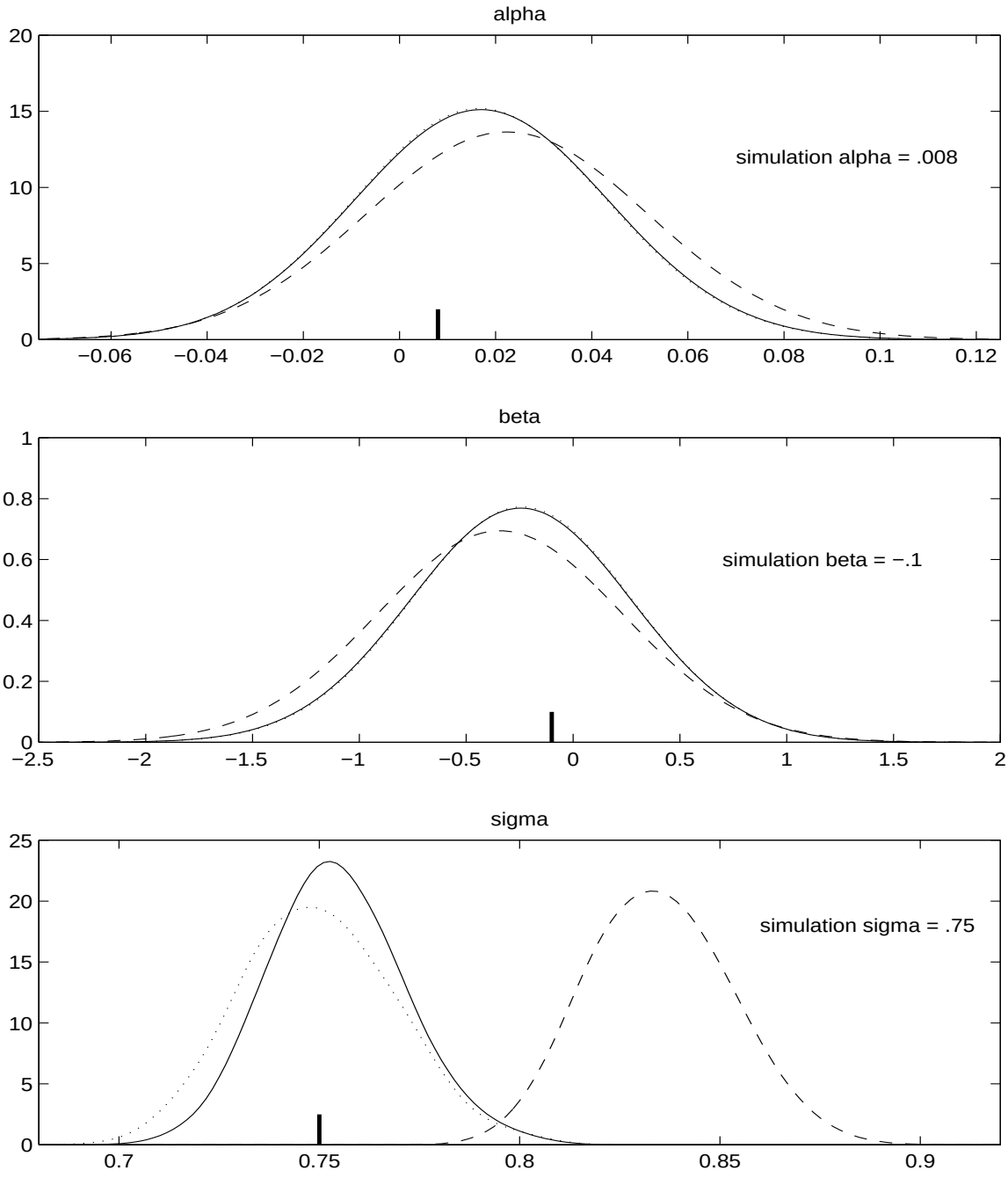


Figure 8B: Unrounded vs. Rounded Data



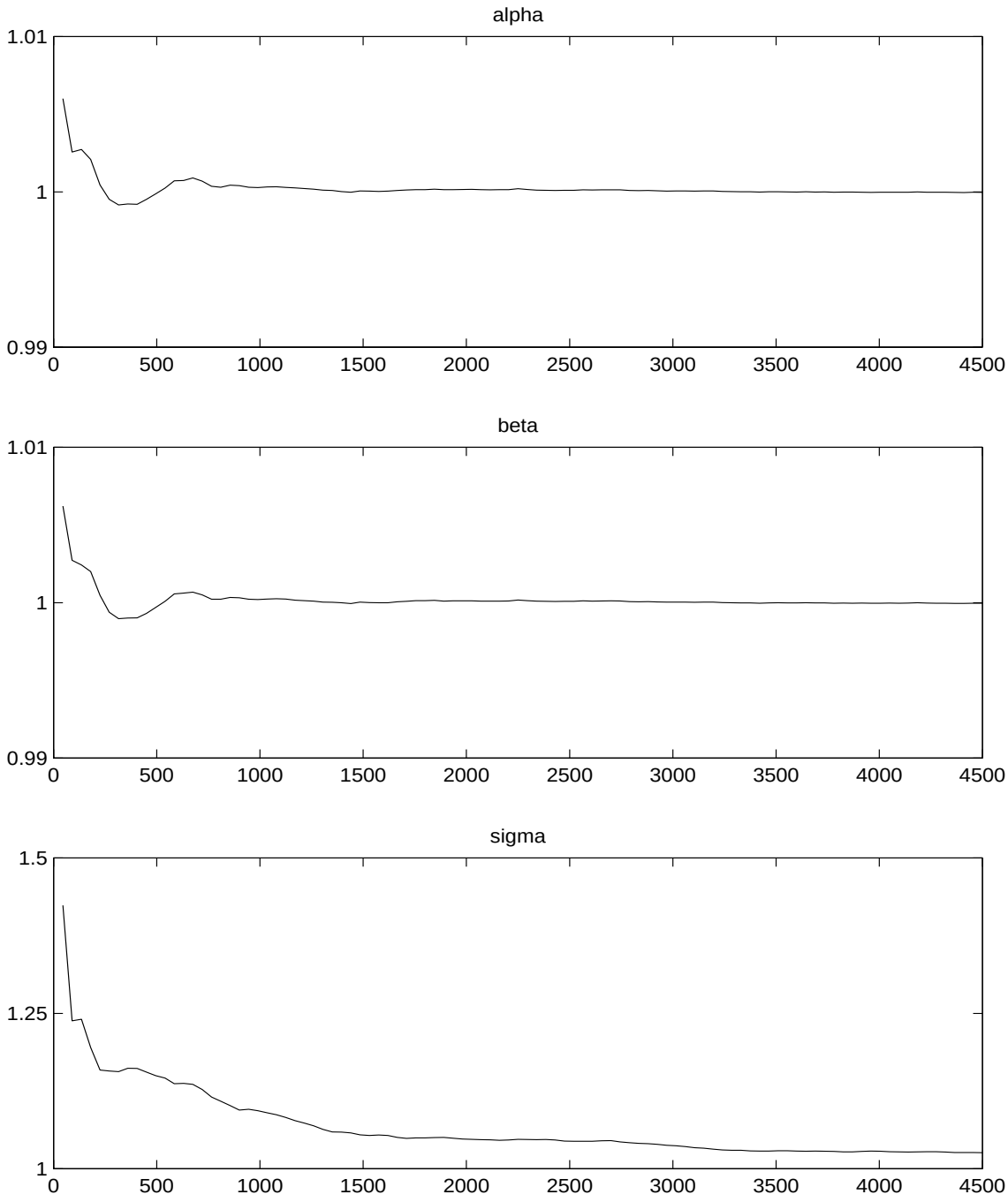
The dashed line in the lower panel shows the data within the box in the upper panel in closer detail. The solid line plots the rounded version of the same data.

Figure 9: Interest Rate Process Posteriors



The solid lines plot annualized posteriors constructed using the unrounded diffusion data. Dashed lines result from data that are rounded to the nearest sixteenth of one percent. The dotted lines use the rounded data but implement the rounding correction described in section 4.3.

Figure 10: Ratio of Between−Chain to Within−Chain Variance for CEV Model



Thirty Markov chains were simulated in parallel up to 4500 iterations. Following Gelman and Rubin (1992), each panel plots the ratio of the variance between chains to the variance within chains for a particular parameter as a function of the length of the Markov chain. Values close to 1 indicate approximate convergence.