MCMC Analysis of Diffusion Models with Application to Finance

Bjørn ERAKER
University of Chicago

This paper proposes a new method for estimation of parameters in diffusion processes from discrete observations. The method is based on MCMC methodology and applies to a wide class of models including systems with unobservable state variables and non-linearities. The method is applied to the estimation of parameters in one-factor (CEV) interest rate models and a two-factor model with a latent stochastic volatility component (SV). The CEV model is found to do a poor job in capturing the time varying volatility interest rate data. The SV model provides vastly superior fit to that of the CEV model. The paper also presents a simulation study which demonstrates that the method provides accurate parameter estimates of the SV model at moderate sample sizes.

KEY WORDS: Diffusion process, Markov Chain Monte Carlo, Gibbs sampler, Bayesian analysis, Stochastic volatility

1. INTRODUCTION

Diffusion processes have become the standard tool for modeling prices in financial markets for derivative and risk management purposes. Although such continuous time processes offer analytic tractability, the parameters that govern their dynamics are often difficult to estimate from discrete time data. In a nutshell, estimation is problematic because the model is formulated in continuous time, while sample data are naturally only available at discrete frequencies. This implies that estimates obtained by naive discretizations of diffusion processes can be subject to discretization bias.

A number of methods have been proposed to estimate diffusion parameters. Working with the difference equation resulting from a Euler discretization of the process can give rise to QML or moment based estimators (e.g., Chan, Karolyi, Longstaff, and Sanders 1992). Naive discretizations, however, give estimates subject to the discretization bias mentioned above if sampling times are infrequent. Prakasa Rao (1988), Yoshida (1992), and Kessler (1997) propose estimators converging to the true parameter more rapidly as the data are sampled more frequently. An alternative strategy which relies on discretizations of the continuous time likelihood function is given in Liptser and Shiryayev (1977), and Aase (1987) among others. Other analytic methods include those of Sørensen (1995) and Bibby and Sørensen (1996) (estimating functions), Aït-Sahalia (1998) (analytic approximation to the likelihood function) while GMM based estimators are discussed in Duffie and Glynn (1996), Hansen and Scheinkman (1995) and Conley, Hansen, Luttmer, and Scheinkman (1997) among others. Non-parametric methods have been proposed in Aït-Sahalia (1996a,b), Jiang and Knight (1997), and Stanton (1997). Previously, simulation-based methods have been proposed for estimating diffusions by the method of simulated moments (Duffie and Singleton 1993), indirect inference methods (Gourieroux, Monfort, and Renault 1993) and the efficient method of moments (EMM) (Gallant and Tauchen 1996) among others. The advantage of simulation based methods is that they typically apply to more general processes than the analytic methods. For instance, Andersen and Lund (1997a,b) apply EMM in estimation of two and three factor non-linear interest rate models with unobserved factors. The level of generality offered by such methods is important because there is mounting empirical evidence of the shortcomings of simple one-factor representations of asset prices.

We propose a new method for estimating parameters in diffusion models based on Markov Chain Monte Carlo (MCMC) simulations. In short, our proposed MCMC approach does the following: since data are only available at discrete times and possibly infrequently, we introduce a set of $m-1$ auxiliary latent data points in-between each pair of observations. These latent data points are integrated out of the likelihood function by Monte Carlo integration in our MCMC sampler. The idea of augmenting the observed low-frequency data with simulated high frequency ones has previously been pursued in Pedersen (1995). Our approach is similar to his, but with an important difference: MCMC simulations are conditioned on more information than the Euler simulations upon which Pedersen's approach is based. This distinction offers crucial efficiency advantages, one of which is that it allows us to easily incorporate latent, unobserved factors.

We apply the method to the estimation of parameters in one-factor interest rate models of the CEV class and to a two-factor model with a stochastic volatility component (SV). While neither of the example models match the data perfectly, the evidence suggests that the SV model provides a vastly superior fit to that of the CEV model. The paper also presents a simulation study which indicates that the method provides accurate estimates at moderate sample sizes.

Journal of Business & Economic Statistics
???

MCMC methods are particularly suited to handle the diffusion problem. Indeed, MCMC methods have proven very useful for solving problems similar to this one where the likelihood function involves the computation of a high dimensional integral. One such example is state space models (e.g., Carlin, Polson, and Stoffer 1992). Another recurring example in financial econometrics is the stochastic volatility (SV) model for which the log-variance of the asset price follows an AR(1). An MCMC solution to this problem was initially proposed by Jacquier, Polson, and Rossi (1994a) and has since been revised and extended in a number of papers. See Jacquier, Polson, and Rossi (1994b) and Kim, Shephard, and Chib (1998) for a review. Kim et al. also point out that the methodology applies to observations missing in the time domain.

SV models provide an illuminating example of the advantages of MCMC. As shown in Jacquier et al. (1994a), taking the posterior means as point estimates of the parameters give estimates that compare favorably to GMM and quasi-likelihood methods. The method also provides a solution to the filtering problem for the unobserved volatility path. In addition, computing times required often compare favorably to other simulation based methods.

The MCMC approach to diffusion models suggested here inherits many of the attractive features of the discrete SV models. It also offers several principal advantages over previous methods. First, it applies to multivariate models. In particular, it avoids problems with an exploding number of parameters entering auxiliary models (as often will be the case with indirect inference methods) by doing away with the specification of auxiliary models altogether. Our method also provides a solution to the filtering and prediction problem for unobserved latent factors as a by-product of the parameter estimation problem. Third, the application of exact sample inference is often considered to be advantageous to that of asymptotic approximations. Although we are primarily concerned with Bayesian calculations here, MCMC simulations can also be used for maximum likelihood estimation (e.g., Geyer 1996). Finally, though not pursued in detail, the principles outlined here allow estimation of multivariate models in which variables are sampled at different times and possibly non-synchronously. This offers interesting possibilities for analyzing relations between infrequently released macroeconomic figures and higher frequency financial data.

Since the first version of this paper appeared, similar methods have been described by Jones (1998) and Elerian, Chib, and Shephard (1999). The methods suggested by these authors differ from this one in the design of MCMC algorithms for sampling the missing data. Jones proposes a somewhat simpler approach than suggested here. Elerian et al. propose a more elaborate scheme based on an independence Metropolis sampler to draw larger blocks of missing data. Their algorithm applies to univariate, one-factor diffusion processes.

The remainder of the paper is organized as follows: in the following section, we detail how MCMC methods can be used for analyzing diffusion models. Examples of how the methodology applies to particular models are provided and include the CEV model of the short term interest rate, as well as a generalization of this model to allow for a stochastic volatility component. Section 3 contains the empirical estimates and corresponding model diagnostics obtained from fitting the example models to the short term US interest rate. We also present some evidence from sampling experiments suggestive of the accuracy of the MCMC solution for the stochastic volatility model. Section 4 summarizes and discusses directions for future research.

2. METHODOLOGY

In the following, we describe an MCMC sampling scheme that allows sampling from the posterior distribution of parameters entering into the drift and diffusion functions. The process of interest is

$$dY_t = \mu(Y_t;\theta)dt + \sigma(Y_t;\theta)dW_t, \tag{1}$$

where $\mu : \Re^d \times \Theta \to \Re^d$ and $\sigma : \Re^d \times \Theta \to \Re^{d\times d}$ for some compact $\Theta \subseteq \Re^k$.

We assume that the process generates discrete time observations and our overall objective is to conduct inference for the parameter vector, θ. Both the diffusion process, Y_t, and the Brownian motion, W_t, are assumed to be d-dimensional.

We further assume that

A1 The drift and diffusion functions satisfy the Lipschitz and linear growth conditions

$$\| \mu(x;\theta) - \mu(y;\theta) \| + \| \sigma(x;\theta) - \sigma(y;\theta) \| \leq C \| x - y \| \tag{2}$$

$$\| \mu(x;\theta) \|^2 + \| \sigma(x;\theta) \|^2 \leq C^2(1+ \| x \|^2) \tag{3}$$

where C is a positive constant and $\| \cdot \|$ denotes the Euclidean norm.

A2 $\sigma(x;\theta)\sigma(x;\theta)'$ is positive definite for all $x \in \Re^d$ where $'$ denotes the matrix transpose.

A3 We have a sample of observations $Y_{t,j}, t = 0, 1, 2, .., T$, $j = 1, 2, .., d_1$ to be used in estimation where $d_1 \leq d$.

A1 is sufficient to ensure the existence of a strong (unique), square integrable solution $Y_t, t \in [0, T]$ to (1) (e.g., Øksendal 1995, p. 64). More importantly, bounded growth of the diffusion function is required for the key step in the MCMC sampling algorithm. We elaborate on this below. It is sufficient that the conditions hold with probability one. **A2** is needed to have a well defined likelihood function. In accordance with **A3**, the process, Y_t, may consist of both observable and non-observable components. If Y_t contains unobservables, the system in (1) defines a continuous time, non-linear state space model. An example of a system with an unobservable component is the continuous time stochastic volatility model. For the sake of clarity, it might be useful to write $Y_t = (X_t, Z_t)$, where X defines the observable part of the system and Z (henceforth referred to as "the state variable") denotes the

unobserved part. X_t and Z_t are assumed d_1 and d_2 dimensional, respectively, hence $d = d_1 + d_2$. We work with the discretized version of (1) given by

$$\triangle Y_t = \mu(Y_t;\theta)\triangle t + \sigma(Y_t;\theta)\triangle W_t, \tag{4}$$

where $\triangle W$ is a d dimensional i.i.d. $N(0, I_d\triangle t)$ random vector where I_d denotes a d-dimensional identity matrix. As argued, the discrete approximation in (4) implies a Gaussian transition density which is only guaranteed to be a good approximation to its continuous counterpart if the discretization interval, $\triangle t$, is sufficiently small. We have assumed that observations are available at integer times, but we do not assume that putting $\triangle t = 1$ is sufficient to approximate the transition density directly. Instead, we propose to put $\triangle t = 1/m$ for some positive integer m which is chosen by the econometrician. This ensures that the bias introduced by discretization can be made arbitrarily small. On the other hand, it introduces a problem of missing values. We divide the time interval $[0, T]$ into $n = mT$ equidistant points $0 = t_0 < t_1 < ... < t_{n-1} < t_n = T$. This implies that X_{t_i} is an observation whenever i is an integer multiple of m. For the unobserved variables Z, the entire m observations are missing in the time span $(t, t+1]$. Altogether, this setup implies that $d_1T(m-1)$ data points are missing from the observable part of the system, while d_2Tm points of the unobserved part are missing.

The essential idea is to substitute the missing data, Y_{t_i}, with simulations, $\hat{Y}_{t_i}$. We refer to the collection of simulated data and observations as the "augmented data". Let $\hat{Y}$ be the $d \times n$ matrix stacking all elements of the augmented data, i.e.,

$$\hat{Y} = \begin{bmatrix} X_{1,t_0} & \hat{X}_{1,t_1} & \cdots & X_{1,t_m} & \hat{X}_{1,t_{m+1}} & \cdots & X_{1,t_n} \\ X_{2,t_0} & \hat{X}_{2,t_1} & & X_{2,t_m} & \hat{X}_{2,t_{m+1}} & \cdots & X_{2,t_n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ X_{d_1,t_0} & \hat{X}_{d_1,t_1} & \cdots & X_{d_1,t_m} & \hat{X}_{d_1,t_{m+1}} & \cdots & X_{d_1,t_n} \\ \hat{Z}_{1,t_0} & \hat{Z}_{1,t_1} & \cdots & \hat{Z}_{1,t_m} & \hat{Z}_{1,t_{m+1}} & & \hat{Z}_{1,t_n} \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ \hat{Z}_{d_2,t_0} & \hat{Z}_{d_2,t_1} & \cdots & \hat{Z}_{d_2,t_m} & \hat{Z}_{d_2,t_{m+1}} & \cdots & \hat{Z}_{d_2,t_n} \end{bmatrix}.$$

Let $\hat{Y}_i$ denote the i'th column of $\hat{Y}$.

Conditioning on the first observation, the joint posterior density is given by

$$\pi(\hat{Y}, \theta) \propto \prod_{i=1}^{n} p(\hat{Y}_i \mid \theta)p(\theta), \tag{5}$$

where $p(\theta)$ is the prior density for the parameter and

$$p(\hat{Y}_i \mid \theta) = \left|\sigma_{i-1}^{-2}\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left\|\left(\triangle\hat{Y}_i - \mu_{i-1}\triangle t\right)\sigma_{i-1}^{-1}\left(\triangle t\right)^{-\frac{1}{2}}\right\|^2\right\},$$

where $\|\cdot\|$ denotes the usual Euclidean norm. We have defined $\triangle\hat{Y}_i := \hat{Y}_i - \hat{Y}_{i-1}$ and used the shorthand notation $\mu_i := \mu(\hat{Y}_i;\theta)$, $\sigma_i := \sigma(\hat{Y}_i;\theta)$ and $\sigma_{i-1}^{-2} := \left(\sigma_{i-1}\sigma_{i-1}'\right)^{-1}$ where $\left|\sigma_{i-1}^{-2}\right|$ defines the determinant of σ_{i-1}^{-2}. Notice that all posterior densities of interest, for instance the conditional posterior of the parameters given the observed data, $\pi(\theta \mid \hat{Y})$, are proportional to (5).

2.1 Gibbs Sampling

The main goal of the analysis is to obtain a sequence of Monte-Carlo samples, $\{\theta^{(h)}\}_{h=1}^{M}$, of the parameter vector from the marginal posterior density. We follow a commonly used notational shorthand and henceforth let π generically denote all posterior densities (i.e., π will generally take on different forms depending on its arguments). Similarly, we let p generically denote π in its unnormalized form.

The inclusion of unobserved data in our model setup needs to be dealt with specifically. One way of handling such missing data is to formulate the joint density for parameters, as well as observed and unobserved data, and then to integrate the unobserved data out of this joint density. To perform this integration, it is sufficient that one is able to compute a simulation sequence, say $\{\theta^{(h)}, \textit{ missing observations}^{(h)}\}_{h=1}^{M}$ from the posterior (5). Then, the sequence $\{\theta^{(h)}\}_{h=1}^{M}$ is implicitly a sample from the marginal posterior $\pi(\theta \mid \textit{observations})$. This principle is referred to as "data-augmentation" and is formalized in Tanner and Wong (1987). Because the number of unobservables (missing data points and parameters) is large, it is not possible to obtain independent samples of these quantities directly from (5). MCMC methods in general, and the Gibbs sampler in particular, are methods designed for sampling from such high dimensional densities. The Gibbs sampler works by sampling one element at the time from the posterior, while keeping the other elements constant as conditioning arguments. For example, by sampling $\theta^{(h)} \sim \pi(\theta \mid \textit{missing observations}^{(h-1)}, \textit{observations})$ in one operation, one could, in principle, sample $\textit{missing observations}^{(h-1)} \sim \pi(\textit{missing observations} \mid \textit{observations}, \theta^{(h)})$ in a second step. By repeating this process, the sequence of simulated parameters and missing data forms a Markov chain for which the stationary distribution is the posterior distribution of interest (see Casella and George 1992, and Albert and Chib 1996 for discussions).

2.2 Conditional Posterior for Missing Data

Unfortunately, sampling all the missing observations in one operation is impossible due to high dimensionality of the associated density. Thus, it is necessary to work with smaller blocks in the Gibbs sampler. Although various choices of blocking schemes in the Gibbs sampler is possible for this problem, we concentrate on the case where the Gibbs sampler updates one column vector of $\hat{Y}$ at a time. To accomplish this, we need the conditional posterior for the i'th column of $\hat{Y}$. This density is defined by the proportionality relationship

$$\pi(\hat{Y}_i \mid \hat{Y}_{\setminus i}, \theta) \propto p(\hat{Y}_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta), \tag{6}$$

where

$$p(\hat{Y}_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta) = \left|\sigma_{i-1}^{-2}\right|^{\frac{1}{2}} \left|\sigma_i^{-2}\right|^{\frac{1}{2}} \times \exp\left\{-\frac{1}{2}\left\|\left(\triangle\hat{Y}_i - \mu_{i-1}\triangle t\right)\sigma_{i-1}^{-1}(\triangle t)^{-\frac{1}{2}}\right\|^2 - \frac{1}{2}\left\|\left(\triangle\hat{Y}_{i+1} - \mu_i\triangle t\right)\sigma_i^{-\frac{1}{2}}(\triangle t)^{-1}\right\|^2\right\}, \quad (7)$$

and $Y_{\setminus i}$ denotes all elements of $\hat{Y}$ except the i th column. Equation (7) follows from (5) because all terms in the product sum in (5) where $\hat{Y}_i$ does not enter can be absorbed into the proportionality constant. This is a consequence of the Markov property of the diffusion process.

It is clear that the unnormalized density in (7) may have various shapes depending on the particular form of the drift and diffusion functions. For instance, if the two normal terms in (7) have roots that are far apart, the density may be bimodal. This may be the case if the discretization step, $\triangle t$, is large. (As will become clear from proposition 2.2, the conditional density in (7) has approximately the shape of a normal density for $\triangle t$ sufficiently small.)

It is important to distinguish between the conditional density $p(\hat{Y}_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}; \theta)$ and the transition density $p(\hat{Y}_i \mid \hat{Y}_{i-1}; \theta)$. The latter defines the density of $\hat{Y}_i$ conditional upon $\hat{Y}_{i-1}$, while the former is conditioned on both the preceding value $\hat{Y}_{i-1}$ and the subsequent value, $\hat{Y}_{i+1}$. At the h th iteration of the Gibbs sampler, we then draw

$$\hat{Y}_i^{(h)} \sim \pi(\hat{Y}_i \mid \hat{Y}_{i-1}^{(h)}, \hat{Y}_{i+1}^{(h-1)}; \theta)$$

for all $i = 0, 1, ..., n$. Naturally, whenever i is an integer multiple of m, we only simulate the d_2 elements corresponding to the unobserved state, Z. To accomplish this, it is sufficient to know $\pi(\hat{Y}_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}; \theta)$ since

$$p(Z_{1,i}, ..., Z_{d_2,i} \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}, X_{1,i}, X_{2,i}, .., X_{d_1,i}; \theta) \propto p(\hat{Y}_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}; \theta)$$

by Bayes theorem. Hence, we use the same expression for the target density to simulate Z_i when ever X_i is fixed at an observed value.

As the Gibbs sampler iterates, it produces a trajectory that bridges the successive discrete observations. This is the key feature of our approach that ensures an efficient Monte-Carlo estimate of the likelihood function. To see this, consider simulating from $p(\hat{Y}_i \mid \hat{Y}_{i-1}; \theta)$, as in Pedersen (1995). Then, a number of the simulated trajectories will produce large jumps from $\hat{X}_{i-1}$ to the observation X_i. Since the probability of such large jumps is zero under the diffusion specification, the simulations inducing large jumps are made redundant in the evaluation of the likelihood function. There are three situations in which the use of unconditional simulation along the lines of Pedersen (1995) will suffer from simulation inefficiencies; the inefficiency increases as the discretization becomes finer and as the dimensionality, d, of the system increases. Most noticeably however, unconditional simulation is typically inadequate for handling models with latent variables (see Danielsson 1994 for a discussion of this in the context of discrete SV models).

The next step in setting up a Gibbs sampler for this problem is to devise some method to simulate $\hat{Y}_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta$. We start by considering the case of a one-dimensional process $Y_t \in \Re$ with constant drift and diffusion functions, $\mu(x) = \mu$, $\sigma(x) = \sigma$ for all $x \in \Re$. The parameter vector is now $\theta = \{\mu, \sigma\}$ and the following result obtains:

Proposition 1. For a scalar process, $Y_t \in \Re$ with constant drift and diffusion functions,

$$\hat{Y}_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta \sim N(\frac{1}{2}(\hat{Y}_{i-1} + \hat{Y}_{i+1}), \frac{1}{2}\sigma^2\triangle t).$$

The above result generalizes straightforwardly to a multivariate system with constant drift and diffusion functions.

We now turn to the general case of arbitrary drift and diffusions. In the initial phase of this work, a random walk Metropolis Hastings algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953) was successfully applied to one-factor models for the interest rate. Still, we suggest using the hybrid accept/reject Metropolis Hastings (hereafter AR-MH) algorithm (see Tierney 1994) to be discussed below, as this algorithm appears to give more rapid convergence. The AR-MH algorithm is very general and only requires knowledge of the unnormalized target density (here: p) and a *proposal* density, q, which we can sample from and that satisfies $\text{supp}(q) \subseteq \text{supp}(p)$. This algorithm works best if the proposal density q and a positive constant C, make qC a good approximation to the target density, p, (see Appendix C for a brief review). We suggest a normal density with mean $m_i = \frac{1}{2}(\hat{Y}_{i-1} + \hat{Y}_{i+1})$ and covariance matrix $\frac{1}{2}\sigma(\hat{Y}_{i-1})\sigma(\hat{Y}_{i-1})'\triangle t$ as proposal density, $q(x \mid Y_{i-1}, Y_{i+1}, \theta)$ in the AR-MH sampler. An intuitive argument motivating this choice goes as follows: both the drift and diffusion functions have bounded growth by assumption. Since the sample paths of the process are continuous, both drift and diffusion functions are locally constant, and so the result in proposition 2.1 should hold approximately for small $\triangle t$ for some estimate of the local variance of the process such as σ_{i-1}^2.

More formally, we have:

Proposition 2. For drift and diffusion functions satisfying A1-A4,

$$\triangle t^{-\frac{1}{2}}(\hat{Y}_i - \frac{1}{2}(\hat{Y}_{i-1} + \hat{Y}_{i+1})) \Rightarrow N(0, \frac{1}{2}\sigma_{i-1}^2)$$

as $\triangle t \to 0$.

By virtue of the above result, we argue that for sufficiently small $\triangle t$ we approximately have

$$\hat{Y}_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1} \sim N(\frac{1}{2}(\hat{Y}_{i-1} + \hat{Y}_{i+1}), \frac{1}{2}\sigma_{i-1}^2\triangle t).$$

One further requirement for efficient implementation of the hybrid AR-MH algorithm is to find some constant C, such that the target p is as close to qC as possible. To

avoid costly numerical optimization, we suggest putting C to the ratio of the two densities evaluated at m_i:

$$C = \frac{p(m_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta)}{q(m_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta)} = p(m_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta) \times \left| (2\pi)^d \frac{1}{2} \sigma(\hat{Y}_{i-1}) \sigma(\hat{Y}_{i-1})' \triangle t \right|^{-\frac{1}{2}}$$

This choice is motivated by the fact that if p approximately has the shape of q, then the ratio of the two densities is constant for all x and thus for $x = m_i$.

Figure 1 illustrates the convergence of qc to p as m gets larger. The figure plots the ratio of the two densities for the CEV process to be considered in the next section at the posterior parameter mean reported below. Since the support of both q and p shrink in $\triangle t$, the figure shows the ratio of the densities for numerically evaluated quantiles in the range $(0.001, 0.999)$. As shown in the figure, the ratio of the two densities gets closer to one as m increases.

For the sample applications considered here, this proposal density provided good results, and the acceptance probability ranged $0.9 - 0.98$ for the one-factor models considered. Moreover, the approximation generally proved better for small $\triangle t$, as to be expected from the above result. It should be emphasized, however, that since practical implementation requires a positive value of $\triangle t$, the normal proposal density does not necessarily dominate the target density in the tails, essentially violating the requirement that $\text{supp}(p) \subseteq \text{supp}(q)$. This, if not accounted for, can cause the algorithm to give seriously erroneous results. A simple remedy for handling such a situation, is to scale the variance of the proposal density by some factor greater than one. This leads to a higher rejection rate in the accept/reject step of the hybrid rejection MH algorithm and implies that the MH step is invoked less frequently. A simple strategy for ensuring that domination occurs is to monitor the number of MH steps invoked to ensure that these stay at a minimum, since, if only the AR part of the algorithm is invoked, dominance is guaranteed.

2.3 Conditional Posterior for the Parameters

The second major step in the Gibbs sampler involves keeping $\hat{Y}^{(h)}$ fixed and sample $\theta^{(h+1)} \sim \pi(\theta \mid \hat{Y}^{(h)})$. The general form of this conditional is given in (5). In the example applications below, we assume a standard non-informative prior.

Conditioning on the data, this density corresponds to a general non-linear, heteroscedastic system. Although it is possible to implement a fully general algorithm to sample the parameter, one is typically better off devising a method for sampling the parameters on a case-by-case basis. Whenever the conditional density for the parameter vector is not recognized as one with a standard form, we might still impose a Metropolis Hastings step, sampling $\theta^{(h+1)}$ conditional upon $\theta^{(h)}$. An example of a non-linearity leading to a non-standard conditional density is given below in the CEV example. Here we employ a Normal approximation based on a Taylor expansion around the current state of the algorithm. This principle can be generalized to sample the entire or larger blocks of the parameter vector. Another possibility is the use of adaptive chains (e.g., Gilks, Best, and Tan 1994 and Gilks, Roberts, and Sahu 1998).

Fortunately, the general expression in (5) allows an explicit representation in terms of known densities for a rather wide class of diffusion models analogous to linear systems. Usually, this is at least the case for a subset of the parameters in the model. In such cases, one might sample from the known conditionals in one Gibbs step of the algorithm and use, for instance, a Metropolis based algorithm for the parameters which enter in a non-linear way. This is the case for the CEV model of the short term interest rate discussed next.

Example 1: CEV Model of the Short-Term Interest Rate

A number of processes that have been proposed as models for the short term interest rate fall into the class generally referred to as constant elasticity of variance (CEV) models. The CEV model is a one-factor model (i.e., Y_t is a scalar) taking on the form

$$dY_t = (\theta_r + \kappa_r Y_t)dt + \sigma Y_t^{\beta} dW_t, \tag{8}$$

where β is a constant. Clearly, $\beta = 0$ produces the O-U process used by Vasicek (1977) as a model for the short-term interest rate, while letting $\beta = 0.5$ gives the "square-root" process studied by Cox, Ingersoll, and Ross (1985). Plugging the drift and diffusion functions defined by (8) into (5), we get

$$\pi(\theta \mid \hat{Y}) \propto \prod_{i=1}^{n} \frac{1}{\sigma \hat{Y}_{i-1}^{\beta}} \exp\left\{ -\frac{1}{2} \frac{\left(\hat{Y}_i - \hat{Y}_{i-1} - (\theta_r + \kappa_r \hat{Y}_{i-1})\triangle t\right)^2}{\sigma^2 \hat{Y}_{i-1}^{2\beta} \triangle t} \right\}. \tag{9}$$

If we define y to be the vector obtained by stacking $\left(\hat{Y}_i - \hat{Y}_{i-1}\right) / \left(\hat{Y}_{i-1}^{\beta}\sqrt{\triangle t}\right)$ and X to be the matrix obtained by stacking $\left[\begin{array}{cc} \sqrt{\triangle t}\hat{Y}_{i-1}^{-\beta} & \sqrt{\triangle t}\hat{Y}_{i-1}^{1-\beta} \end{array}\right]$ for all i we can rewrite (9) as the likelihood function corresponding to a linear regression of y on X. Consequently,

$$(\theta_r, \kappa_r) \mid \sigma, \hat{Y} \sim N(\overline{\theta}, \sigma^2 (X'X)^{-1}) \tag{10}$$

$$\sigma^{-2} \mid \hat{Y} \sim IG(n-2, \overline{s}^2), \tag{11}$$

where $\overline{\theta} = (X'X)^{-1}(X'y)$ and $\overline{s}^2 = \frac{1}{n}\sum_i (y_i - X_i\overline{\theta})^2$ and where $N()$ and $IG()$ denote multivariate normal and inverted Gamma densities, respectively. Sampling from these densities is easily accomplished.

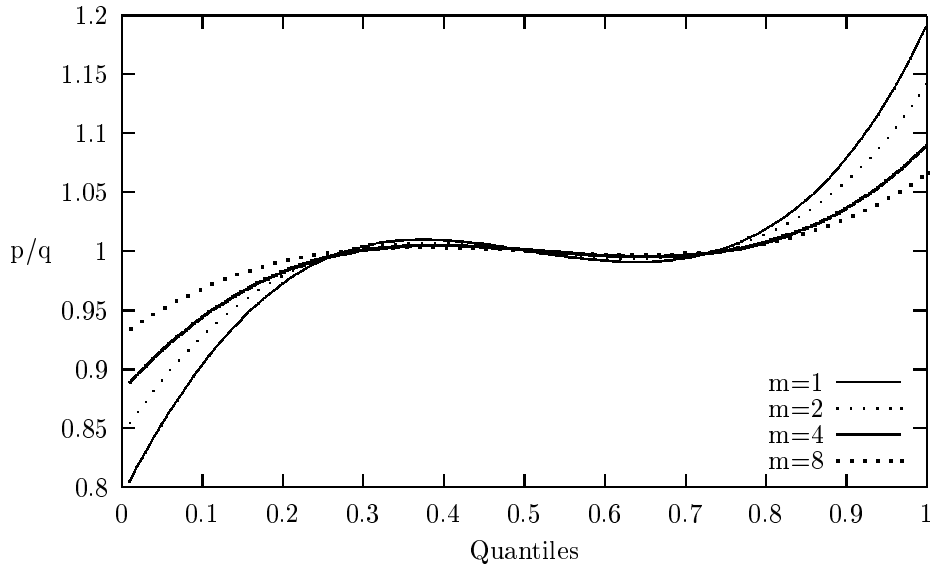

Figure 1. Convergence of the conditional posterior

To estimate β, we notice that the second derivative of the log-posterior, $p(\beta \mid \hat{Y}, \theta_r, \kappa_r, \sigma)$, with respect to β is

$$-2 \sum_i \left(\frac{(\hat{Y}_i - \hat{Y}_{i-1} - (\theta_r + \kappa_r \hat{Y}_{i-1}) \triangle t)^2}{\hat{Y}_{i-1}^{\beta}} \right)^2 \log(\hat{Y}_{i-1})^2,$$

which is negative, implying that this density is log-concave. This renders sampling of β possible by a variety of methods (see for example Wild and Gilks (1993)). We use a Metropolis step with a normal proposal density based on a second order Taylor expansion of the log-posterior. Hence, we propose from

$$N\left(\beta^{(h-1)} - \frac{f'(\beta^{(h-1)})}{f''(\beta^{(h-1)})}, -(f''(\beta^{(h-1)}))^{-1} \right), \qquad (12)$$

where $f(\beta) := \log p(\beta \mid \hat{Y}, \theta_r, \kappa_r, \sigma)$. In the empirical applications of this model, this Metropolis sampler had acceptance probabilities near one, indicating that this proposal is a very good approximation to the target density.

We now have the following sampling recipe for the CEV model:

1. Initialize all unknowns. For instance, we might use linear interpolation between observed values of X_i to initialize $\hat{X}_i$. Set $h = 1$.
2. For all $i = 0, 1, 2, .., n$ draw $\hat{Y}_i^{(h)} \mid \hat{Y}_{\setminus i}^{(h-1)}, \theta$ using the hybrid rejection Metropolis Hastings algorithm with proposal density $N(\frac{1}{2}(\hat{Y}_{i-1}^{(h)} + \hat{Y}_{i+1}^{(h-1)}), \frac{1}{2}[\sigma_{i-1}^{(h)}]^2 \triangle t)$.
3. Draw $(\theta_r, \kappa_r)^{(h)}$ using (10).
4. Draw $\sigma^{(h)}$ using (11).
5. Draw $\beta^{(h)}$ using Metropolis Hastings based on the proposal in (12).
6. Increase h and return to step 2.

Example 2: Continuous Time Stochastic Volatility

As a generalization of the basic CEV model considered in the previous section, we study the model

$$\begin{aligned} dr_t &= (\theta_r + \kappa_r r_t)\, dt + \exp\left(\frac{1}{2} Z_t\right) r_t^{\beta} dW_{1,t} \\ dZ_t &= (\theta_z + \kappa_z Z_t)\, dt + \sigma_z dW_{2,t}, \end{aligned} \qquad (13)$$

where r_t is assumed to be observed at discrete intervals (hence $X = r$ in terms of the notation in the previous section) $t = 0, 1, .., T$, and Z_t is an unobserved process. At risk of abusing the term, we henceforth refer to Z_t as the "log-volatility process" (which is accurate only if $\beta = 0$). The model has previously been studied by Andersen & Lund (1997a,b).

Downing (1999) points out that the Andersen-Lund SV model violates the growth condition (assumption A1). Indeed, this condition does not hold for all values $(r_t, Z_t) \in \Re^2$ but it does hold with probability one, if the joint process is bounded with probability one. Downing uses Euler simulations to verify that this is the case for the parameter estimates reported in Andersen and Lund (1997a,b).

In obtaining samples of θ conditional upon the augmented data, $\hat{Y}$ we suggest at least two additional blocks in the Gibbs sampler containing the parameters entering the "interest rate" process $\{\theta_r, \kappa_r\}$, the log-volatility process $\{\theta_z, \kappa_z\}$ and σ_z. As we have

$$\mu(Y_i; \theta) = \begin{bmatrix} \theta_r + \kappa_r X_i \\ \theta_z + \kappa_z Z_i \end{bmatrix}$$

$$\sigma(Y_i; \theta) = \begin{bmatrix} \exp\left(\frac{1}{2} Z_i\right) X_i^{\beta} & 0 \\ 0 & \sigma_z \end{bmatrix},$$

it is easily seen that $\pi(\theta \mid \hat{Y})$ can be written as the product of two normal densities

$$\pi(\theta \mid \hat{Y}) \propto p(\theta_r, \kappa_r \mid \hat{Y}) p(\theta_z, \kappa_z, \sigma_z \mid \hat{Y}),$$

where

$$p(\theta_r, \kappa_r \mid \hat{Y}) = \prod_{i=1}^{n} \exp\left(-\frac{1}{2} Z_i\right) r_i^{-\frac{1}{2}} \triangle t^{-1/2} \times \exp\left\{-\frac{(\triangle r_i - (\theta_r + \kappa_r r_i)\triangle t)^2}{2\exp(Z_t) r_i \triangle t}\right\}$$

$$p(\theta_z, \kappa_z, \sigma_z \mid \hat{Y}) = \prod_{i=1}^{n} \sigma_z^{-1} \triangle t^{-1/2} \times \exp\left\{-\frac{(\triangle Z_i - (\theta_z + \kappa_z Z_i)\triangle t)^2}{2\sigma_z^2 \triangle t}\right\}.$$

These expressions have forms similar to that of the (normal) likelihood function for linear regression problems and are consequently easy to sample from. A Gibbs sampler for the stochastic volatility model can now be set up by treating $\{\theta_r, \kappa_r\}, \{\sigma, \kappa_z\}$ and σ_z as separate blocks.

Notice that the dynamics of the SV model in (13) is equivalently represented by

$$\begin{aligned} dr_t &= (\theta_r + \kappa_r r_t)\, dt + \sigma_r \exp\left(\frac{1}{2} Z_t\right) r_t^{\beta}\, dW_{1,t} \quad (14) \\ dZ_t &= \kappa_z Z_t dt + \sigma_z dW_{2,t}. \end{aligned}$$

Here, the constant term θ_z in the volatility drift is replaced by a σ_r in the interest rate diffusion. For both parameterizations, these parameterizations determine the unconditional standard deviation of the interest rate changes. The parameterization in (14) is advantageous from a computational point of view, which is further elaborated upon in Appendix D. Notice that the reparameterizaton does not lead to problems in drawing the parameters, as the conditional for κ_z is normal, while the conditionals for σ_r and σ_z are inverse gamma.

Notice that for multivariate models with drift and diffusion which are linear/multiplicative in the parameters, one obtains conjugate normal/wishart posteriors with moments corresponding to general linear models.

2.4 Convergence Issues

The Gibbs sampler is generally convergent for problems with a continuous target density with non-disjoint support. As with many high dimensional problems, this is difficult to verify generally. Therefore, as for many practical applications of MCMC, we must assess convergence on the basis of empirical analysis of the output from the sampler. For instance, if the sampler is initialized at several different starting values and provides output for which the empirical distributions are indistinguishable by means of, for instance, analysis of variance tests, the process is assumed to have converged. The most common way, however, of evaluating the convergence of MCMC methods is simply to study the "time-series" plots of the output from the sampler.

Even if a Gibbs sampler converges, it may provide highly dependent samples. In practical applications, it is difficult to detect whether a sequence is non-convergent or just converging slowly. Slow convergence increases the computational burden and excessively dependent sequences may be a practical problem. In applications to stochastic differential equations, convergence slows down as $\triangle t$ gets small. In the limit, as $\triangle t \to 0$, the Gibbs sampler does not converge at all. This is easily seen by noting that certain expressions for the conditional posteriors collapse to Dirac measures in the limit. Consequently, the sampler does not move from its initial position as $\triangle t$ goes to zero. On the other hand, as $\triangle t$ gets large, the estimates are obviously worsened by the discretization bias. The empirical applications reported here and elsewhere indicate that this is not a problem in practice.

A further discussion on the issue of sampling algorithms with high serial dependence vs. block Metropolis algorithms proposed elsewhere, is given in Eraker, Jacquier, and Polson (1999a).

Regarding starting values, one strategy that has proven successful in the empirical applications given in the next section, is to start the Gibbs sampler setting $\triangle t$ large, say 1, and then subsequently decreasing it after convergence has been obtained. If, for instance, $\triangle t$ is decreased by half, then the size of the augmented data is roughly doubled. We now have twice as many missing observations as before which can be sampled using the hybrid MH algorithm discussed previously. Similarly, one can continue to decrease $\triangle t$ until further reductions have negligible effect on the empirical distribution of the chain.

3. EMPIRICAL APPLICATIONS

In this section, we present estimates of the parameters in the CEV one-factor model and the CEV/stochastic volatility models considered in the previous sections. The data used consist of 2288 weekly (Wednesday) observations of the 3 month US Treasury Bill yields, covering the period from 1/1/1954 to 10/5/1997. The data were collected from the Federal Reserve's weekly H.15 reports of daily market data. (The data are available at URL: http://www.bog.frb.fed.us/Releases/H15/data/h15.zip, and are essentially the same as in Andersen and Lund (1997a,b), but with a slightly longer sampling period.) As with most financial data, these interest rates are known to exhibit large amounts of heteroscedasticity. They are also highly autocorrelated as can be seen from the sample statistics presented in table 1. A more in-depth exploratory analysis of these data is given in Andersen and Lund (1997b).

Table 1. Descriptive Statistics

	mean	*median*	*st.dev.*	*skewness*	*kurtosis*
r_t	*0.0555*	*0.0514*	*0.0286*	*1.0865*	*4.5936*
$r_t - r_{t-1}$	*0.0000*	*0.0000*	*0.0025*	*-0.4868*	*22.4065*
			autocorrelations		
lag	*1*	*2*	*3*	*4*	*5*
r_t	*0.9961*	*0.9916*	*0.9867*	*0.9814*	*0.9757*
$r_t - r_{t-1}$	*0.0918*	*0.0379*	*0.0568*	*0.0582*	*0.0542*

3.1 Empirical Results for the CEV Model

Figure 2 plots the density estimates of the parameters in the CEV model for different orders of discretization. For computational details such as posterior sample sizes and running times, see Appendix B. There is generally little difference between the estimated densities.

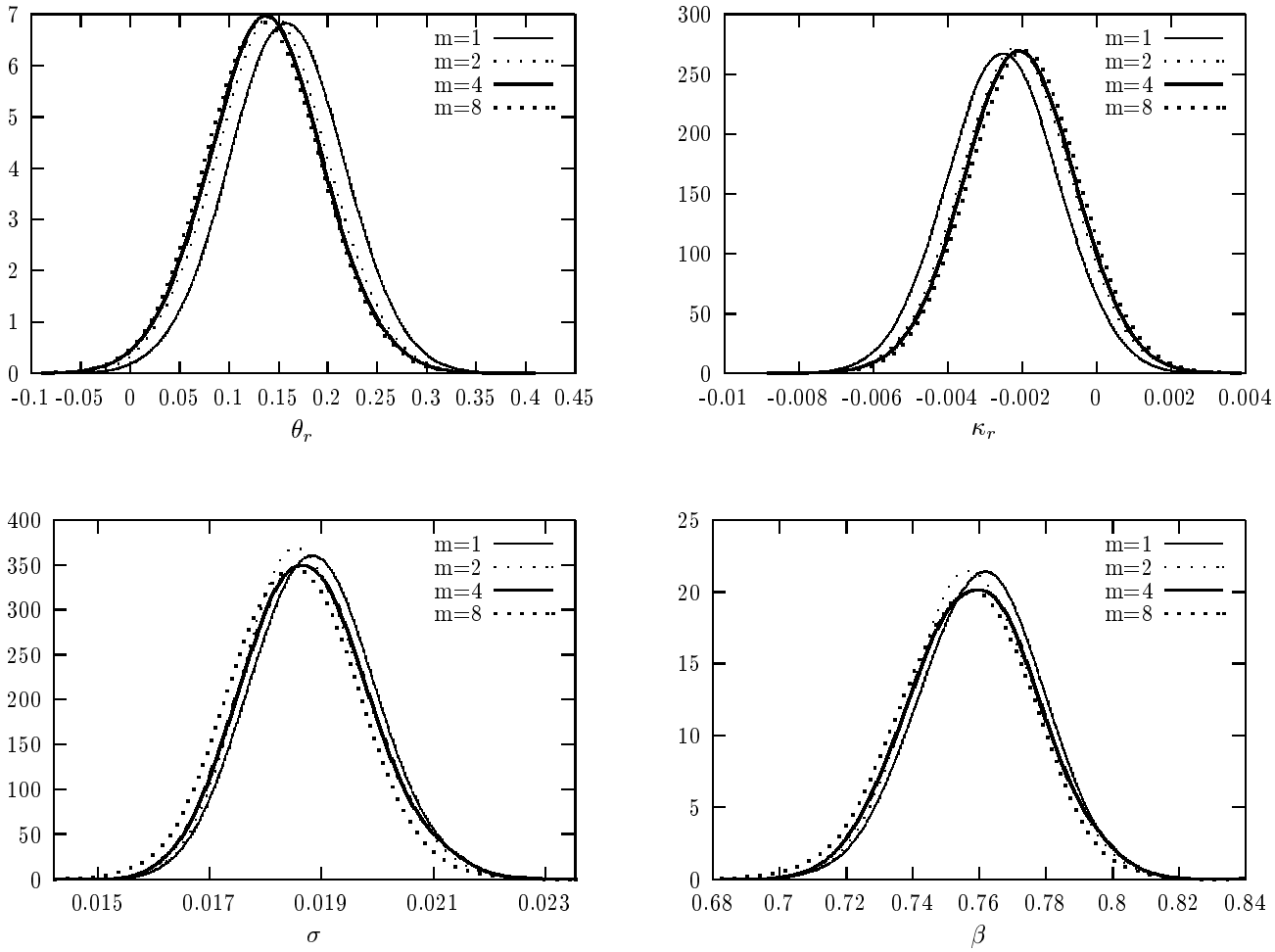

Figure 2. Posterior densities for the CEV model

Table 2. Posterior Moments for the CEV Model

	$\theta_r \times 1000$	κ	σ	β
		$\triangle t = 1$		
mean	*0.1589*	*-0.0025*	*0.0189*	*0.7608*
st.dev.	*0.0551*	*0.0014*	*0.0010*	*0.0175*
NSE	*0.0005*	*0.0000*	*0.0000*	*0.0005*
		$\triangle t = 1/2$		
mean	*0.1443*	*-0.0022*	*0.0187*	*0.7579*
st.dev.	*0.0547*	*0.0014*	*0.0011*	*0.0188*
NSE	*0.0004*	*0.0000*	*0.0001*	*0.0008*
		$\triangle t = 1/4$		
mean	*0.1367*	*-0.0020*	*0.0186*	*0.7570*
st.dev.	*0.0553*	*0.0014*	*0.0011*	*0.0169*
NSE	*0.0005*	*0.0000*	*0.0000*	*0.0007*
		$\triangle t = 1/8$		
mean	*0.1334*	*-0.0020*	*0.0185*	*0.7567*
st.dev.	*0.0549*	*0.0014*	*0.0010*	*0.0173*
NSE	*0.0003*	*0.0000*	*0.0000*	*0.0007*

NOTE: The table reports posterior means and standard deviations for parameters in the CEV diffusion. NSE denotes estimates of Monte Carlo error in the posterior means. Zero entries for the NSE indicate numerical standard errors less than 0.0001.

Indeed, for θ_r and κ, it is difficult to distinguish between the $m = 4$ and $m = 8$ cases. Generally, the largest shift in the estimated posteriors seems to occur when going from $m = 1$ to $m = 2$. Table 2

quantifies the differences in the posterior moments across different levels of discretization. As can be seen from this table, the shift in the posterior means is largest between $m = 1$ and $m = 2$ and relatively small for $m = 4$ and $m = 8$.

The parameter estimates for the CEV process indicate that the US T-bill data are highly persistent: the posterior mean-reversion parameter κ_r has about 8% of its mass above 0 for the highest discretization ($m = 8$). The similar figure for $m = 1$ is 3.7 %, indicating that a decision on the 5 % level about the positivity of this parameter is reversed by employing the higher order discretization scheme.

The CEV parameter β is estimated to about 0.75, which is roughly half of the 1.5 estimate obtained by GMM in Chan et al. (1992). Andersen and Lund (1997a) estimate this parameter to 0.67.

Figure 3 presents model diagnostics for the CEV model. The upper plot in the figure compares the distribution of observed interest rate differentials to that implied by the CEV model by means of a Q-Q plot. The model implied distribution was estimated by a long sample path (100 $\times$ the actual sample size) using the posterior means for $\triangle t = 1/8$. There is a systematic departure between the model's implied stationary distribution and the sample data evidenced by the S shape in the estimated quantiles. The discrepancy between the two densities is due to large outliers in the sample data. In short, the CEV model is incapable of reproducing the high kurtosis in the observed interest rate differentials. The inability of the CEV model to capture the tails of the distribution of interest rate differentials has previously been pointed out by Andersen and Lund (1997b), among others.

Figure 3 also plots the autocorrelations for the interest rate changes and the squared interest rate changes. The autocorrelations are compared to that of corresponding autocorrelations obtained from simulations of the same sample size. The figure plots the mean of these simulations along with its 5 th and 95 th percentiles. As can be seen, the autocorrelations in the observed interest rate changes often exceeds that of the 5 th and 95 th percentiles from the simulated series. This suggests that the simple linear drift specification does not fully capture the dynamics in the interest rates. Alternative drift specifications such as the non-linear one in Aït-Sahalia (1996b) and the central tendency factor model (e.g., Andersen and Lund 1997b) could potentially capture this model departure.

Yet, the inability of the CEV model to fit the conditional mean dynamics of the interest rate process is not the only evidence of mis-specification: the CEV model dramatically fails to capture the conditional variance in the interest rate changes. This is evidenced by the model's inability to generate sufficiently high autocorrelations in squared interest rate changes, in particular for the shorter lags. The bottom plot depicts the residuals obtained from one particular MCMC run with parameters near the mode of the posterior. These residuals are obtained by integrating over the Brownian increments over each unit of time and should be i.i.d. $N(0, 1)$ if the CEV model holds true. Notice that the posterior distribution of Bayesian residual terms generally departs from normality, but normality still holds around the mode of the likelihood function (see Zellner 1975). The residuals change minimally across the posterior simulations, so the plot is representative.

Clearly, the CEV error terms are strongly heteroscedastic. For example, in the period 1962-1965 the residual standard deviation is about 0.3, while the corresponding figure is 1.9 during 1980-1983. Notice that the high volatility periods do not necessarily correspond with high levels of the interest rate itself which is evidenced by the high residual variance over the 1958-1960 period. The average interest rate in this period was about 3.6 %. There are a number of subperiods in the data with high volatility/ low interest rates and vice versa. This renders it doubtful that any one-factor model in which the diffusion term depends on the interest rate level itself can accurately capture the dynamics in these data. Rather, the evidence of significant autocorrelation in squared residuals suggests that ARCH-like processes, such as the stochastic volatility model considered next, hold greater promise.

3.2 Empirical Results for the Stochastic Volatility Model

Figure 4 and Table 3 summarize the posterior simulations for the stochastic volatility model. Notice that the results in the table correspond to a re-parameterization of the model that improve upon the convergence speed of the algorithm. This is elaborated upon in Appendix D. The parameters in the interest rate drift are of similar magnitude as those in the CEV model. One noticeable difference, however, is the posterior standard deviation for κ_r which

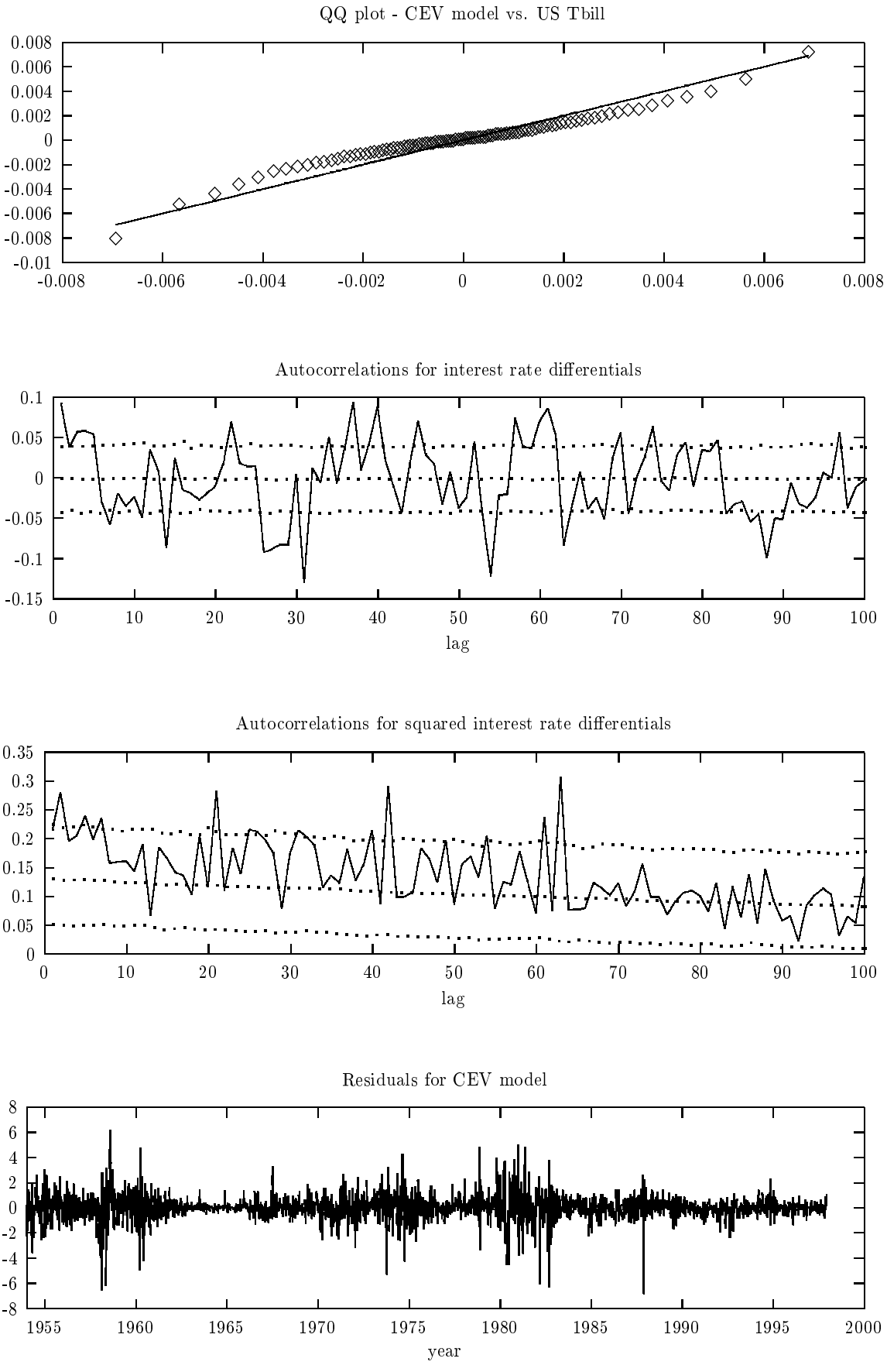

Figure 3. Diagnostics for the CEV Model

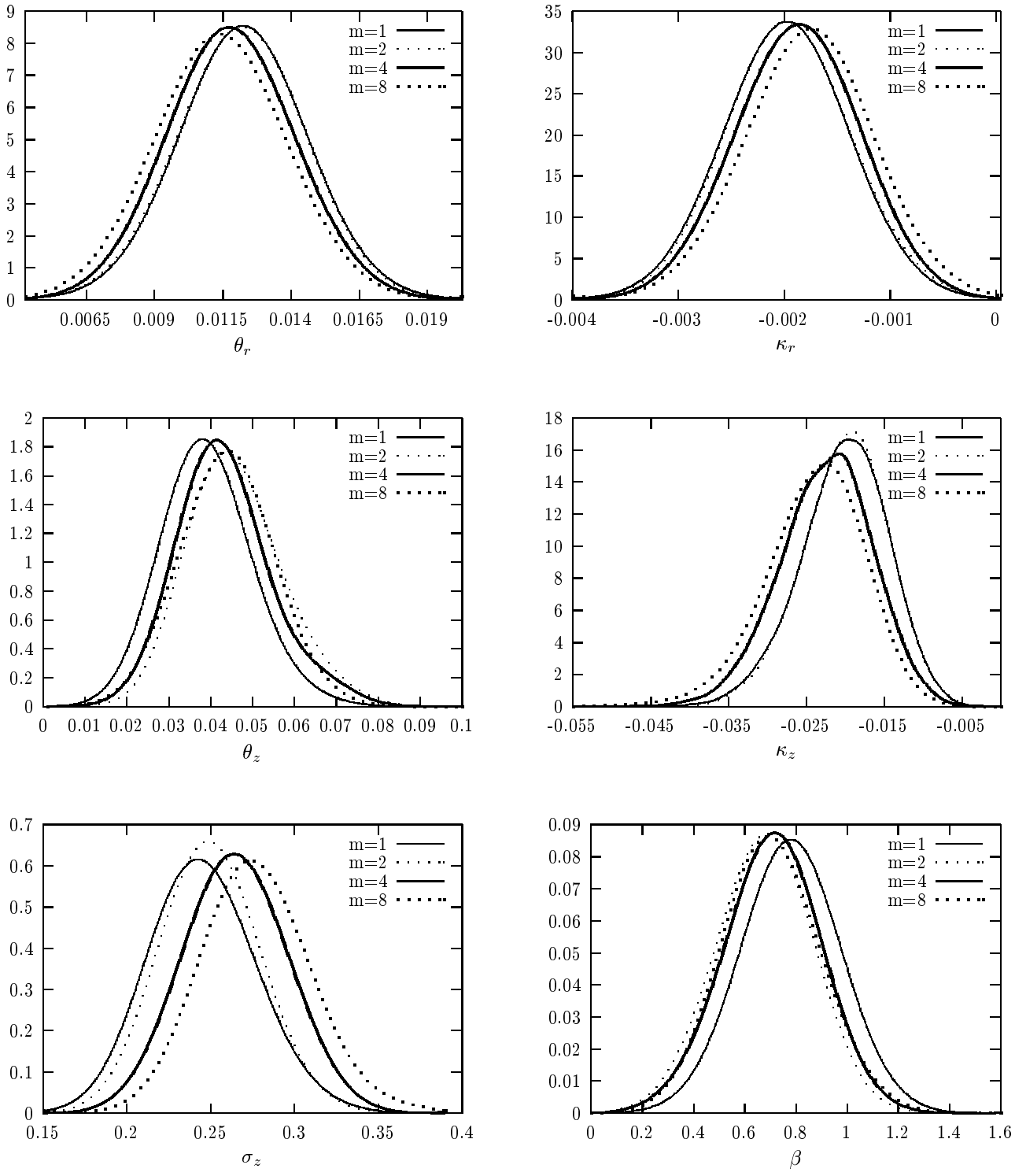

Figure 4. Posterior densities for the SV model

Table 3. Posterior Moments for the Stochastic Volatility Model

	θ_r	κ_r	σ_r	κ_z	σ_z	β
			$\triangle t = 1$			
mean	0.0123	−0.0020	0.0392	−0.0201	0.2457	0.7813
st.dev.	0.0020	0.0005	0.0079	0.0056	0.0267	0.1257
NSE	0.0000	0.0000	0.0006	0.0001	0.0008	0.0092
			$\triangle t = 1/2$			
mean	0.0122	−0.0020	0.0459	−0.0201	0.2496	0.6755
st.dev.	0.0020	0.0005	0.0088	0.0055	0.0236	0.1170
NSE	0.0001	0.0000	0.0017	0.0002	0.0015	0.0227
			$\triangle t = 1/4$			
mean	0.0118	−0.0019	0.0434	−0.0224	0.2650	0.7095
st.dev.	0.0020	0.0005	0.0087	0.0060	0.0252	0.1213
NSE	0.0000	0.0000	0.0010	0.0001	0.0009	0.0136
			$\triangle t = 1/8$			
mean	0.0114	−0.0018	0.0437	−0.0240	0.2770	0.7027
st.dev.	0.0021	0.0005	0.0083	0.0064	0.0277	0.1239
NSE	0.0001	0.0000	0.0011	0.0002	0.0020	0.0161

NOTE: The table reports posterior means and standard deviations for the stochastic volatility model for different discretizations different $\triangle t$. NSE denotes estimates of Monte Carlo error in the posterior means.

is about 1/3 that of the CEV model. None of the posterior draws for this parameter were positive, indicating that the evidence of mean reversion in interest rates is stronger in the stochastic volatility framework.

An examination of the parameters in the log-volatility process reveals the usual high volatility persistence: the κ_z estimate implies a first order autocorrelation of the log-volatility process of 0.976. This, of course, is in line with evidence reported elsewhere in the GARCH/SV literature. Notice, also, that the posterior densities of σ_z and κ_z are skewed to the right. The non-normality of these posteriors implies that inference based on normality assumptions will give misleading inference for these parameters.

Our analysis allows direct comparison with the results obtained by the method of EMM in Andersen and Lund (1997a). Of particular interest, are the estimates of the volatility related parameters, κ_z, σ_z and β. Our estimate of κ_z is close to the AL estimates: their estimate translates into -0.02 at the weekly frequency. The AL estimate of σ_z translates to 0.176 so our estimate of 0.27 implies a volatility process behaving considerably more erratically. Similarly, our estimates of β are higher and thus imply more time variation in the interest rate volatility. Both studies find that this parameter is imprecisely estimated.

Another interesting comparison between the results here and those of AL, is the precision in the parameter estimates. The posterior standard deviation for the κ_r parameter is slightly lower than that of AL. However, for the parameters in the volatility process, this is reversed: the posterior standard deviations reported here for κ_z and σ_z are about three times greater than the corresponding standard errors in AL.

Regarding the parameter posteriors for the stochastic volatility model, it is also worth noting that different values of β lead to very small changes in the posteriors for the other parameters. Results obtained with β fixed to $\frac{1}{2}$ (not reported here) give almost identical fit for the remaining parameters.

Turning to an examination of different discretization schemes, we see that the relative shift in the posterior mean across different values of the discretization parameter is largest for β. The significance of this shift is less evident because of the large posterior standard deviation for this parameter. The volatility parameters, κ_z and σ_z also change by relatively large amounts. As with β, these changes seem small relative to the estimation errors as measured by the posterior standard deviations. Taking estimation error into account, the effect of discretization error is largest for the σ_z. The posterior of the low-level discretization ($m = 1$) has about 62% of its mass in common with the posterior based on $m = 8$.

Also relating to discretization bias, it is worth noting that, in general, it is well known that discretization errors for simple OU processes increase with the level of mean reversion. Here, both processes are close to the unit root, and so the lower level discretizations tend to provide better approximations than what can be expected for data with lower amounts of mean reversion. There are many examples of financial market models with more rapidly mean reverting processes than those studied here. For example, stock market volatility is typically found to be mean reverting faster than the results reported here for interest rate volatility. Also, in interest rate models that specify a central tendency factor (see for example Andersen and Lund (1997b)), the short rate typically reverts rapidly to the central tendency factor.

Figure 5 presents some diagnostics suggestive of the fit of the SV model to the observed interest rate data. First of all, it is evident that the SV model provides a much better fit to the unconditional density of interest rate changes than the CEV model does. As seen from the QQ plot in

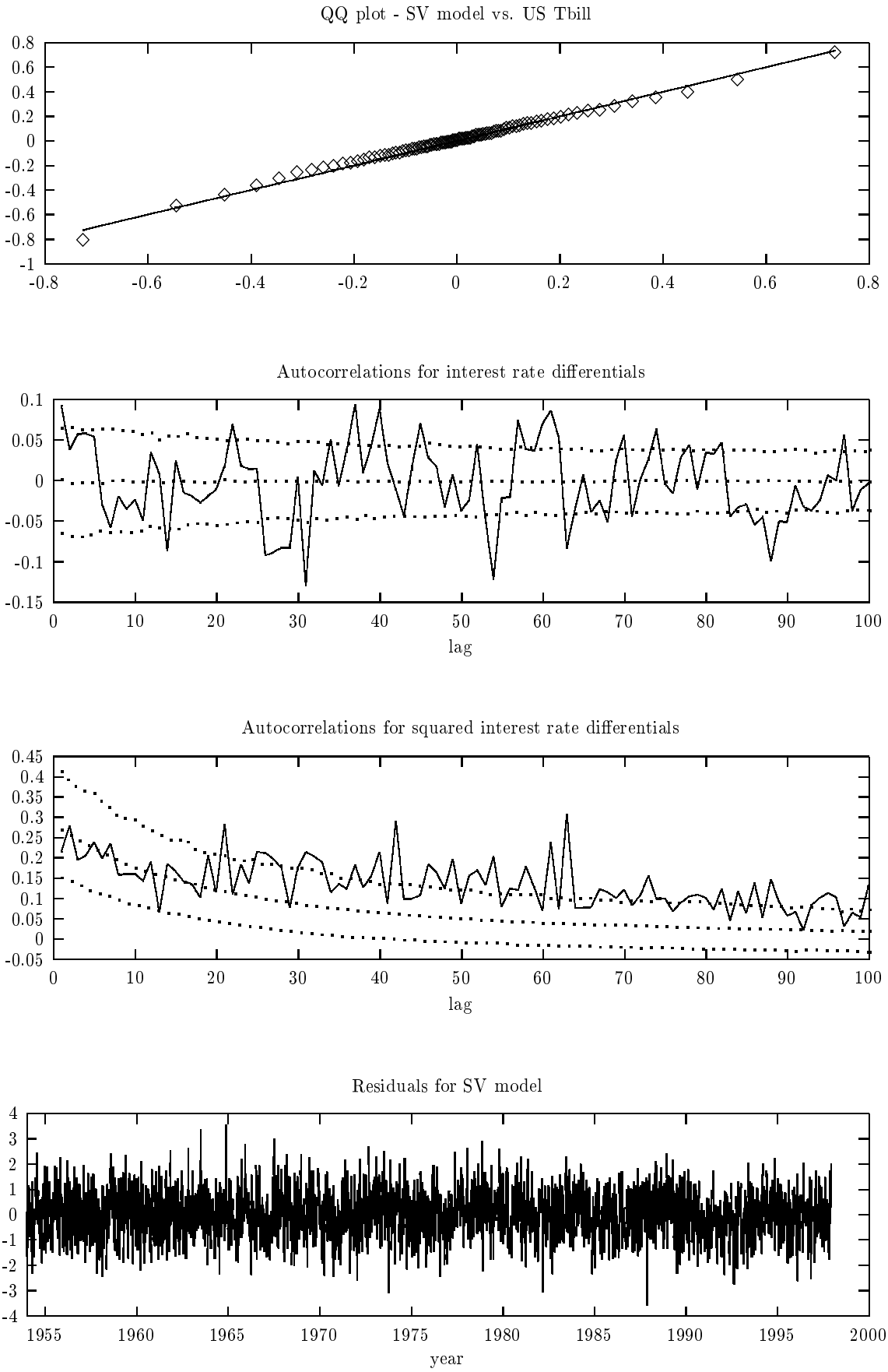

Figure 5. Diagnostics for the SV Model

the upper plot of the figure, the simulated SV path gives a good match to the empirical quantiles of the sample data.

Figure 5 also indicates that the SV model does not seem to generate enough autocorrelation in squared interest rate changes to match the observed patterns. In fact, in this respect, the CEV model rather surprisingly seems to be doing better than the SV model. The evidence suggests longer range dependence in volatility suggested by Bollerslev and Mikkelsen (1996) in stock market returns. Recent suggestions to capture these effects in the context of continuous time diffusions are given in Gallant, Hsu, and Tauchen (1998).

Still, the SV model outperforms the CEV model in terms of autocorrelations for the first 20 lags of the squared series. The residual-plot in Figure 5 also suggests dramatically less heteroscedasticity than what was the case for the CEV model. This, in our view, along with the drastic improvement in fit to the unconditional distribution of the data, indicates that the SV model fits the data markedly better than the CEV model does.

3.3 Small Sample Evidence

To assess the accuracy of the Bayesian MCMC estimator of the parameter in the stochastic volatility model, this section presents evidence on the performance of the estimator on data generated by Monte Carlo for the stochastic volatility model. Due to the computational demands, we limit the posterior draws to 20,000. Each experiment was repeated 200 times, and the resulting parameter estimates are summarized in Table 4. Also, β, which posterior distribution requires an extensive number of draws to estimate accurately, is fixed to $\frac{1}{2}$ in the simulation study. The results from estimating the model with β fixed to $\frac{1}{2}$ using the US T-bill data, are very close to those in table 3. This suggests that fixing β should have negligible effect on the remaining, estimated parameters in the simulation study.

The sampling experiments produce estimates that, on average, are close to the true parameters that generated the sample data. Naturally, the biases are larger for the smaller data sets consisting of 500 observations. For instance, κ_r, which measures the speed of the mean reversion of the (artificial) interest rate process, has a "true value" of -0.002 while it was on average estimated to -0.0127 when using 500 observations. The accuracy increases as the sample size is increased to 2000, yielding an estimate of -0.0031 on average. For the parameter measuring the reversion in the log-volatility process, κ_z, the estimates are also slightly downward biased, while the estimate of the diffusion coefficient for the log-volatility process, σ_z, is virtually unbiased using a sample size of 2000.

When interpreting the results in Table 4, it should be kept in mind that the reversion coefficients can be interpreted as approximately "autocorrelation - 1". This, coupled with the well-known fact that sample autocorrelation coefficients are downward biased for processes with near unit roots may explain, at least partially, the direction of the bias in the sampling experiments.

It should be noted that the relatively small posterior sample size of $10,000$ used to calculate the posterior means in our sampling experiments may be expected to give estimates of the posterior means that are influenced by Monte Carlo errors. Since these errors have mean zero, provided that the sampler has been run with a sufficiently long "burn in"-period, this does not affect the biases of our estimates. It does, however, influence the dispersion of the posterior means, as measured by the RMSE. Hence, the results in Table 4 are likely to underestimate the efficiency of the MCMC estimator.

Overall, the performance of the MCMC method for estimating the parameters in the stochastic volatility model is difficult to assess. Arguably, the relevant measure of its accuracy would be in comparing the numbers in Table 4 with that of competing methods. For the stochastic volatility model, the competition narrows down to indirect inference methods such as EMM. Unfortunately, the small sample performance of EMM for diffusion processes currently remains an unexplored area of research.

4. CONCLUDING REMARKS

This paper has been concerned with the estimation of parameters entering the drift and diffusion functions of stochastic differential equations. The proposed estimator is based on simulation-based inference through Markov Chain Monte Carlo methods (MCMC). Such methods have proven to be particularly well suited for solving the high dimensional integrals associated with the calculation of marginal posterior densities in previous applications involving latent, or missing, observations. In particular, this results from the fact that the posterior density can be written as the joint posterior density of the "augmented data" consisting of both observable and unobservable parts of the continuous system.

Our Gibbs sampling approach consists of two major steps: first, we simulate points on the unobserved paths between discrete time observations of the partially observed system and we simulate the entire path of the unobserved part. The essential point is that this simulation should be conducted conditional upon all observed values of the process. As argued, these simulations can be carried out by sampling from the sequence of conditional densities in a Gibbs sampler. We propose a sampling scheme based on Tierney's hybrid rejection Metropolis Hastings algorithm to simulate missing vectors of observations. The second major step in the Gibbs sampler is to simulate the parameter vector conditional upon the "augmented data." Since the data are kept constant at this step, the problem is typically reduced to simulating from the posterior of a regression model.

Applications of the proposed methodology include the constant elasticity of variance model (CEV) and a stochastic volatility (SV) model. The empirical examples suggest that the MCMC method is largely successful in capturing the discretization bias associated with the weekly data. This is particularly evident from the sampling experiments carried out for the SV model. Neither of these models

Table 4. Sampling Experiments for the Stochastic Volatility Model

	θ_r	κ_r	θ_z	κ_z	σ_z
			True values		
	0.00020	−0.00200	−0.30000	−0.03000	0.30000
			Simulation Results		
			T=500		
mean	0.00127	−0.01271	−0.38174	−0.03873	0.24297
median	0.00083	−0.00875	−0.33569	−0.03385	0.24876
RMSE	0.00154	0.01468	0.21844	0.02209	0.07173
			T=2000		
mean	0.00031	−0.00316	−0.32770	−0.03262	0.29906
median	0.00028	−0.00281	−0.31568	−0.03166	0.29800
RMSE	0.00017	0.00176	0.07807	0.00765	0.02880

NOTE: The table reports means, medians and root mean squared errors (RMSE) for parameters estimated on data generated according to the stochastic volatility model. The experiments were repeated 200 times using the Euler approximation with $\Delta t = 1/20$ and repeated with sample sizes of 500 and 2000 observations. The estimation results are based on $20,000$ Gibbs iterations with $\Delta t = 1/8$. Parameter-estimates are given as the posterior means of iterations $10,000 - 20,000$.

seem to fit the observed interest data perfectly. In particular, the CEV model seems incapable of capturing the highly lepto-kurtic distribution of interest rate differentials as well as the high serial correlation in squared interest changes at lower lags. Not surprisingly, the SV model alleviates both these problems, but still has some problems in capturing the long run autocorrelations in the squared interest rate changes. Nevertheless, our evidence indicates that the SV model captures the predominant features of the interest rate data.

There are several interesting extensions to the current modeling framework. Applications to jump diffusion models have been been considered by Johannes, Kumar, and Polson (1999). Their framework relies on a discretized diffusion component with normally distributed jump sizes and allows for possible state dependence in the jump frequencies. The assumption of normally distributed jump sizes allows for generalizations into models with arbitrary drift and diffusion functions, as shown here. The case of non-normal jump sizes has been considered in an application to affine jump diffusion models with jumps to both returns and stochastic volatility in Eraker, Johannes, and Polson (1999b).

ACKNOWLEDGMENTS

I am grateful for the helpful suggestions from an anonymous associate editor and two referees. I have also received helpful comments from Jørgen Haug, Bernt A. Ødegaard, Ola Elerian, Hans-Peter Bermin, Jesper Lund, Nicholas G. Polson, Michael S. Johannes and seminar participants at the Norwegian School of Economics and Business Administration, Centre for Analytical Finance in Aarhus, CORE (Lovain-la-Neuve), University of Copenhagen, Nordic Symposium on Contingent Claims Analysis (Copenhagen Business School), Stanford University, and the University of Chicago. This paper is based on my thesis at the Norwegian School of Economics and Business Administration (Eraker (1997a)), supervised by Dag Tjøstheim and Gunnar Stensland. Their helpful suggestions are also greatly appreciated. I also thank the Norwegian Supercomputer Committee for a grant of computing time.

APPENDIX A: PROOFS

Proof of Proposition 2.1

From (7) we have that $x \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta$ has density

$$
\begin{aligned}
p(x \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta) &\propto \exp\left\{-\frac{(x-\hat{Y}_{i-1}-\mu\triangle t)^2+(\hat{Y}_{i+1}-x-\mu\triangle t)^2}{2\sigma^2\triangle t}\right\} \\
&= \exp\left\{-\frac{2x^2+\hat{Y}_{i-1}^2+\hat{Y}_{i+1}^2+2(\mu\triangle t)^2-2x(\hat{Y}_{i-1}+\hat{Y}_{i+1})+2(\hat{Y}_{i-1}+\hat{Y}_{i+1})\mu\triangle t}{2\sigma^2\triangle t}\right\} \\
&\propto \exp\left\{-\frac{2x^2-2x(\hat{Y}_{i-1}+\hat{Y}_{i+1})}{2\sigma^2\triangle t}\right\} \\
&\propto \exp\left\{-\frac{(x-\frac{1}{2}(\hat{Y}_{i-1}+\hat{Y}_{i+1}))^2}{\sigma^2\triangle t}\right\}
\end{aligned}
$$

which is recognized as the kernel of a normal density with mean $\frac{1}{2}(Y_{i-1}+Y_{i+1})$ and variance $\frac{1}{2}\sigma^2\triangle t$.

Proof of Proposition 2.2

First, note that while the limit $\lim_{\triangle t\to 0}\hat{Y}_i/\sqrt{\triangle t}$ does not exist, it can be seen from (4) that $(\hat{Y}_i-\hat{Y}_{i-1})/\sqrt{\triangle t}$ is $O_p(1)$. Let ∂ denote the derivative operator applied to the first argument. By the boundedness of derivatives of the drift and diffusion functions implied by A1,

$$\begin{aligned}\mu(\hat{Y}_i;\theta) &= \mu(\hat{Y}_{i-1};\theta)+\partial\mu(\hat{Y}_{i-1};\theta)(\hat{Y}_i-\hat{Y}_{i-1})+O_p(\triangle t)\\ &= \mu(\hat{Y}_{i-1};\theta)+O_p(\sqrt{\triangle t}).\end{aligned}$$

Similarly,

$$\sigma(\hat{Y}_i;\theta)=\sigma(\hat{Y}_{i-1};\theta)+O_p(\sqrt{\triangle t}).$$

Now set $DY_i := (\hat{Y}_i-\hat{Y}_{i-1})\triangle t^{-\frac{1}{2}}$. We can then rewrite (7) as

$$\begin{aligned}p(Y_i\mid .)\propto\ & \left|\sigma_{i-1}^{-2}\right|^{\frac{1}{2}}\left|\sigma_i^{-2}\right|^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\left\|\left(DY_i-\mu_{i-1}\sqrt{\triangle t}\right)\sigma_{i-1}^{-1}\right\|^2-\frac{1}{2}\left\|\left(DY_{i+1}-\mu_i\sqrt{\triangle t}\right)\sigma_i^{-1}\right\|^2\right\}\\ =\ & \left|\sigma_{i-1}^{-2}\right|^{\frac{1}{2}}\left|\left(\sigma_{i+1}+O_p(\sqrt{\triangle t})\right)^2\right|^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\left\|\left(DY_i-O_p(\triangle t)\right)\sigma_{i-1}^{-1}\right\|^2\right.\\ & \left.-\frac{1}{2}\left(DY_{i+1}-O_p(\triangle t)\right)\left(\sigma_{i-1}+O_p(\sqrt{\triangle t})\right)^{-2}\left(DY_{i+1}-O_p(\triangle t)\right)'\right\}\\ \to\ & \left|\sigma_{i-1}^{-2}\right|\exp\left\{-\frac{1}{2}\left\|DY_i\sigma_{i-1}^{-1}\right\|^2-\frac{1}{2}\left\|DY_{i+1}\sigma_{i-1}^{-1}\right\|^2\right\}\\ \propto\ & \left|(\frac{1}{2}\sigma_{i-1})^{-2}\right|^{\frac{1}{2}}\exp\left\{-\left\|\left((\hat{Y}_i-\frac{1}{2}(\hat{Y}_{i-1}+\hat{Y}_{i+1}))\triangle t^{-\frac{1}{2}}\right)\sigma_{i-1}^{-1}\right\|^2\right\},\end{aligned}\quad (A.15)$$

where the proportionality relationship in the last expression follows by multiplying with the constant necessary to complete the quadratic terms and normalize.

APPENDIX B: COMPUTATIONAL ISSUES

To obtain posterior estimates for the CEV model, the MCMC sampler was run for 20,000 for $m=1,2,4$ and for 40,000 iterations for $m=8$. Each run with different m, the additional missing points were initiated by interpolating the values from the lower level discretization. The first 5,000 iterations with a new value of m were discarded for "burn-in" leaving effective posterior sample sizes of 15,000 and 35,000 respectively. A run using 10,000 iterations with $\triangle t=1/8$ took about 30 minutes of Origin 2000 computer time, which translates to about 1.5 hours on a Pentium II 200 mhz. Notice however that lower level approximations with fewer MCMC runs can give Pentium 200 running times down to about 30 minutes with satisfactorily low numerical errors. The computer-program did not take advantage of the scalar arithmetic available for the one-factor example, indicating that drastically lower running times are available for such models. With respect to computing times for one factor model, we also compared this algorithm with the one proposed in Elerian et al. (1999). In a previous version of their paper, Elerian et al. report a simulation experiment using 500 iterations with $m=4$ to take 258 seconds running on a pentium 133 processor. The algorithm proposed here in used 4.5 seconds to run the same experiment on a similar processor.

For the stochastic volatility example, we ran the sampler for 200,000 iterations with $m=1,2,4$ and for 500,000 iterations using $m=8$. The large number of posterior draws are necessary to produce the smooth density estimates in Figure 4 for β. The remaining posteriors can be accurately estimated with 100,000 iterations and $m=8$. Every 10 th simulation was recorded for posterior analysis. The Origin 2000 running time for 100,000 iterations with $m=8$ is about 8.5 hours.

Numerical standard errors reported in Tables 2 and 3 were based on methods suggested in Geweke (1992) using a Parzen window with a maximum lag of 1000.

The empirical results reported in this paper were obtained using computer programs written in the C language by the author and with auxiliary routines from Press, Teukolsky, Vetterling, and Flannery (1992). The simulation results reported were obtained by running independent processes in parallel on various IBM RS 9000 580/590 and a Cray Origin 2000 at the University of Bergen, Norway.

APPENDIX C: THE HYBRID ACCEPT/REJECT METROPOLIS HASTINGS ALGORITHM

Assume that the aim is to sample a parameter θ_i from a posterior density, $\pi(\theta_i)$ available only in its un-normalized form denoted $p(\theta_i)$. Suppose we have available proposal densities $q(\theta_i)$ (for every i th element of θ) which are non-dependent on the current state of the Markov chain. To provide a sample from the target density π, the hybrid rejection MH algorithm works as follows: At the h th iteration, 1) Use rejection sampling to sample a candidate θ_i^* by proposing from q and with accept probability $\min\left\{1, \frac{p(\theta_i^*)}{q(\theta_i^*)c}\right\}$.

2) Write $\theta_i^{(h)} = \theta_i^*$ with probability

$$\alpha = \begin{cases} 1 & \text{if } p(\theta_i^{(h-1)}, \theta_{\setminus i}^{(h)}) < cq_i(\theta_i^{(h-1)}); \\ p(\theta_i^{(h-1)}, \theta_{\setminus i}^{(h)})/cq(\theta_i^{(h-1)}) & \text{if } p(\theta_i^{(h-1)}, \theta_{\setminus i}^{(h)}) > cq(\theta_i^{(h-1)}) \text{ and } p(\theta_i^*, \theta_{\setminus i}^{(h)}) < cq(\theta_i^*); \\ \min\left[\frac{p(\theta_i^*, \theta_{\setminus i}^{(h)})q(\theta_i^{(h-1)})}{p(\theta_i^{(h-1)}, \theta_{\setminus i}^{(h)})q(\theta_i^*)}, 1\right] & \text{if } p(\theta_i^{(h-1)}, \theta_{\setminus i}^{(h)}) > cq(\theta_i^{(h-1)}) \text{ and } p(\theta_i^*, \theta_{\setminus i}^{(h)}) > cq(\theta_i^*). \end{cases}$$

otherwise, write $\theta_i^{(h)} = \theta_i^{(h-1)}$. That is, the i'th element of θ is unchanged.

Notice that in the first step, we continue until we accept a proposal. In the second step, we keep the previous value of θ_i if the uniformly distributed number u is less than the acceptance probability.

APPENDIX D: REPARAMETERIZING THE SV MODEL

MCMC samples for the stochastic volatility model given in (13) will typically have high serial dependence. For example, consider the conditional sample mean of the volatility process $\bar{Z}^{(g)} = \frac{1}{n}\sum_{i=1}^n Z_i^{(g)}$. The successive draws of $\bar{Z}^{(g)}$ are likely to be close, and so the MC draws for any of the remaining parameters in the model that are highly correlated with $\bar{Z}^{(g)}$ are also likely to be highly serially dependent. This is primarily a concern with the CEV parameter, β. In the following we consider ways to reduce this problem to reduce the running times of the sampler.

First, we can write

$$\begin{aligned} dr_t &= \kappa_r(\theta_r - r_t)dt + \exp(\frac{1}{2}k)\exp(\frac{1}{2}z_t - \frac{1}{2}k)r_t^\beta dW_{1,t} \\ &= \kappa_r(\theta_r - r_t)dt + \sigma_r \exp(\frac{1}{2}Z_t^*)r_t^\beta dW_{1,t}, \end{aligned}$$

where $\sigma_r = \exp(\frac{1}{2}k)$ for some constant k. This might lead us to consider the parameterization,

$$\begin{aligned} dr_t &= \kappa_r(\theta_r - r_t)dt + \sigma_r \exp(\frac{1}{2}z_t^*)r_t^\beta dW_{1,t} \\ dZ_t^* &= \kappa_z Z_t^* dt + \sigma_z dW_{2,t} \end{aligned}$$

Now the log volatility process, Z_t^*, has mean zero, and the average volatility is determined by the scaling parameter σ_r^2, which conditional distribution is of the usual inverted Gamma form. Again, the parameters σ_r and β will be highly correlated, but this problem can effectively be handled by drawing these parameters jointly (e.g., by MH) if their correlation seriously slows the convergence of the Gibbs sampler. Notice also that we cannot simultaneously identify σ_r and θ_z (other than through informative priors) since both determine the average volatility.

Unfortunately, this parameterization has certain undesirable characteristics: the mean of the conditional distribution for κ_z is given by

$$\frac{\sum \nabla Z_i^* Z_i^*}{\sum Z_i^{*2}\nabla t} = \frac{cov(\nabla Z_i^*, Z_i^*)}{var(Z_i^*) + \bar{Z}^{*2}}.$$

This expression introduces a downward bias in the simulated values of κ_z for any value of $\bar{Z}$ different from its true value of 0. This leads to undesirable properties in the MCMC chain, as the simulated values of κ_z tend toward zero. Hence, this parameterization cannot be recommended.

To eliminate this problem, consider the system

$$\begin{aligned} dr_t &= \kappa_r(\theta_r - r_t)dt + \sigma_r \exp(\frac{1}{2}Z_t - \frac{1}{2}k)r_t^\beta dW_{1,t} \\ dZ_t &= \kappa_z(k - Z_t)dt + \sigma_z dW_{2,t} \end{aligned}$$

Notice that the joint posterior of all unknows in the system is left unaffected by the introduction of the constant k. Hence, for each iteration of the MCMC sampler, we can add or subtract some suitable k from the system without affecting the stationary distribution of the chain. Thus, we might set $k = \bar{Z}^{(g)}$. This suggests the following MCMC sampler:

1. Draw the missing data $Z_i^{(g)}$ for $i = 0, ..., n$ and $r_i^{(g)}$ for $i = 1, .., m-1, m+1, ..., 2(m-1), 2(m+1), ..., n-1$ using the general hybrid AR-MH algorithm.
2. Sample $\kappa_r^{(g)}, \theta_r^{(g)}$ and $\sigma_r^{(g)2}$ from their respective normal/ inverted Gamma distributions.
3. Sample $\beta^{(g)}$ using the MH.
4. Write $Z_i^{(g)} \leftarrow Z_i^{(g)} - \bar{Z}^{(g)}$ and $\sigma_r^{(g)} \leftarrow \sigma_r^{(g)} \exp(\frac{1}{2}\bar{Z}^{(g)})$
5. Draw κ_z and σ_z^2 from their respective Normal/ Gamma distributions.

Notice that step 3 in the above scheme corrects the bias problem, as the mean of the conditional for κ_z is now the same as under the original parameterization.

[Received October 1998. Revised July 2000.]

REFERENCES

Aase, K. K. (1987). Estimation in Models of Security Prices. *Scandinavian Actuarial Journal*, *1*, 211–224.

Aït-Sahalia, Y. (1996a). Nonparametric Pricing of Interest Rate Derivative Securities. *Econometrica*, *64*, 527–560.

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

Aït-Sahalia, Y. (1998). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approach. *Working paper, Princeton.*

Albert, J. and Chib, S. (1996). Computation in Bayesian Econometrics: An Introduction to Markov Chain Monte Carlo. *Advances in Econometrics*, *11 A*, 3–24.

Andersen, T. G. and Lund, J. (1997a). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, *7*, 343–377.

Andersen, T. G. and Lund, J. (1997b). Stochastic Volatility and Mean Drift in the Short Rate Diffusion: Sources of Steepness, Level and Curvature in the Yield Curve. *Working Paper, Northwestern University.*

Bergstrom, A. R. (1976). *Statistical Inference in Continuous Time Economic Models.* North-Holland, Amsterdam.

Bibby, B. M. and Sørensen, M. (1996). On Estimation for Discretely Observed Diffusions: A Review. *Theory of Stochastic Processes*, *2(18)*, 49–56.

Bollerslev, T. and Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, *73*, 151–184.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling. *Journal of the American Statistical Association*, *87*, 493–500.

Casella, G. and George, E. (1992). Explaining the Gibbs sampler. *American Statistician*, *46*, 167–174.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., and Sanders, A. B. (1992). An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *Journal of Finance*, *3*, 1209–1227.

Conley, T., Hansen, L. P., Luttmer, E., and Scheinkman, J. (1997). Estimating Subordinated Diffusions from Discrete Time Data. *Review of Financial Studies*, *10*, 525–577.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, *53*, 385–407.

Danielsson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *Journal of Econometrics*, *64*, 375–400.

Downing, C. (1999). Nonparametric Estimation of Multifactor Continuous time Interest Rate Models. *Working paper, Federal Reserve (BOG).*

Duffie, D. and Glynn, P. (1996). Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals. *Manuscript, Stanford University.*

Duffie, D. and Singleton, K. (1993). Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, *61*, 929–952.

Elerian, O., Chib, S., and Shephard, N. (1999). Likelihood Analysis of Discretely Observed Diffusions. *Working paper, Nuffield College, Oxford.*

Eraker, B. (1997). MCMC Analysis of Diffusion Models with Application to Finance. *HAE Thesis, Norwegian School of Economics and Business Administration.*

Eraker, B., Jacquier, E., and Polson, N. G. (1999a). The Effects of Low Conductance Sets on MCMC Algorithms. *Working paper, University of Chicago.*

Eraker, B., Johannes, M. S., and Polson, N. G. (1999b). Return Dynamics in Continuous-Time with Jumps to Volatility and Returns. *Working paper, University of Chicago.*

Gallant, A. R., Hsu, C., and Tauchen, G. (1998). Using High/Low Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance. *Working paper, Duke University.*

Gallant, A. R. and Tauchen, G. (1996). Which Moments to Match?. *Econometric Theory*, *12*, 657–681.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In Berger, J. O., J.M. Bernado, A. D., and Smith, A. (Eds.), *Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics.* Oxford University Press.

Geyer, C. J. (1996). Estimation and Optimization of Functions. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (Eds.), *Markov Chain Monte Carlo in Practice.* Chapman & Hall.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in practice.* Chapman & Hall.

Gilks, W. R., Roberts, G. O., and Sahu, S. K. (1998). Adaptive Markov Chain Monte Carlo through Regeneration. *Working paper, Cambridge.*

Gilks, W., Best, N. G., and Tan, K. K. C. (1994). Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *Working Paper, Cambridge.*

Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect Inference. *Journal of Applied Econometrics*, *8*, S85–S118.

Hansen, L. P. and Scheinkman, J. (1995). Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes. *Econometrica*, *63*, 767–804.

Jacquier, E., Polson, N. G., and Rossi, P. (1994a). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, *12*, 371–389.

Jacquier, E., Polson, N. G., and Rossi, P. (1994b). Stochastic Volatility: Univariate and Multivariate Extensions. *Working paper, University of Chicago.*

Jiang, G. J. and Knight, J. L. (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., Kumar, R., and Polson, N. G. (1999). State Dependent Jump Models: How do US Equity Indices Jump?. *Working paper, University of Chicago.*

Jones, C. (1998). A Bayesian Method for the Analysis of Diffusion and Jump-Diffusion Processes. *Manuscript, Wharton.*

Kessler, M. (1997). Estimation of an Ergodic Diffusion from Discrete Observations. *Scandinavian Journal of Statistics*, *24*, 1–19.

Kessler, M. and Sørensen, M. (1995). Estimating Equations based on Eigenfunctions for a Discretely Observed Diffusion Process. *Research report No. 332, Department of theoretical Statistics, University of Aarhus.*

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, *65*, 361–393.

Liptser, R. S. and Shiryayev, A. N. (1977). *Statistics of Random Processes I: General Theory.* Springer.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, N. M., and Teller, E. (1953). Equations of State Calculations by Fast Computing Machine. *Journal of Chemical Physics*, *21*, 1087–1091.

Øksendal, B. (1995). *Stochastic Differential Equations: An introduction with applications.* Springer.

Pedersen, A. (1995). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics*, *22*, 55–71.

Prakasa Rao, B. L. S. (1988). Statistical Inference from Sampled Data for Stochastic Processes. *Contemporary Mathematics*, *80*, 249–284.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recepes in C*. Cambridge University Press.

Sørensen, M. (1995). Estimating Functions for Discretely Observed Diffusions: A review. *Research Report no. 348, University of Aarhus*.

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 Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, *82*, 528–550.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, *22*, 1701–1762.

Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, *5*, 177–188.

Wild, P. and Gilks, W. (1993). Algorithm AS-287 - adaptive rejection sampling from log-concave density-functions. *Journal of the Royal Statistical Society, Series C*, *42*, 701–709.

Yoshida, N. (1992). Estimation for Diffusion Processes from Discrete Observations. *Journal of Multivariate Analysis*, *41*, 220–242.

Zellner, A. (1975). Bayesian Analysis of Regression Error Terms. *Journal of the American Statistical Association*, *70*, 138–144.