MCMC Analysis of Diffusion Models with Application to Finance

Bjørn Eraker*
Norwegian School of Economics and Business Adm.,
Helleveien 30, 5035 Bergen - Sandviken
Norway

First draft: August 1997
This draft: March 1998

Abstract

In this paper a new method is proposed for estimation of parameters in diffusion processes from discrete observations. The proposed simulation based MCMC methodology applies to a wide class of models including systems with unobservable state variables and non-linearities. We apply the method to the estimation of parameters in one-factor interest rate models of the CEV class and to a generalization of this model to a two-factor model with a stochastic volatility component. The small sample properties of the estimator are studied trough sampling experiments for the stochastic volatility model and the results indicate that the method provides accurate estimates at moderate sample sizes.

Keywords: Diffusion process, stochastic volatility, Gibbs sampler, MCMC, Interest Rates
JEL Classification: C11, C15, G12

*Correspondance: E-mail: Bjorn.Eraker@nhh.no. This work has received support from the Norwegian Supercomputer Committee through a grant of computing time. I am grateful for helpful comments from Jørgen Haug, Bernt A. Ødegaard, Ola Elerian, seminar participants at the Norwegian School of Economics and Business Administration, Centre for Analytical Finance, Aarhus, CORE, Lovain-la-Neuve, Belgium, University of Copenhagen, and in particular, Dag Tjøstheim and Gunnar Stensland. Remaining mistakes are mine.

1 Introduction

The elegant and rigorous theory of the absence of arbitrage in securities markets coupled with Ito calculus make very powerful tools for financial analysis. Yet, most empirical research in finance has with a few exceptions focused on discrete time models. The shortage of empirical research adopting continuous specifications has one obvious reason; estimating continuous time diffusion models from discrete time samples is rather involved and until recently, no general estimation methods allowing for arbitrary drift and diffusion functions have been available. For instance, even for simple one factor models of the short term interest rate, the transition density and hence the likelihood function, may be impossible to formulate explicitly.

In the following, we give a brief review of different methods for estimating parameters in diffusion models. As indicated, the problem of evaluating the likelihood function is essentially that of a very general form for the transition density for discrete observations of the process. In some cases, notably for the Ornstein-Uhlenbeck process and the Cox-Ingersoll-Ross square root diffusions, exact ML estimation is feasible because the respective transition densities are known. Although the transition density of more general diffusion processes is typically unknown for large sampling intervals, $\triangle t$, say, it is conditionally Gaussian over finely spaced intervals. Therefore, a quasi maximum likelihood (QML) estimator is obtained by discretization using the Euler scheme. The QML estimator converges to the true parameter vector as $\triangle t \to 0$. A number of papers propose estimators converging to the true parameter more rapidly than the QML estimator (i.e., Prakasa Rao (1988), Yoshida (1992) and Kessler (1997)). For instance, in Kessler (1997) the transition density is approximated by a normal density, however with means and variances more carefully matching those of the true transition density. This is proven to give estimates converging as $(\triangle t)^3 n \to 0$ where n is the sample size.

Another branch of the literature is concerned with developing so-called estimating functions. Estimating functions are equations that, when equated to zero, approximate the first order conditions corresponding to the usual MLE problem. A simple example is quadratic equations which can be interpreted as moment conditions. See Sørensen (1995) and Bibby and Sørensen (1996) for a review of these developments.

The problems associated with formulating the likelihood function have led researchers to consider method of moments estimation for diffusions. Examples of moment conditions are given in Chan, Karolyi, Longstaff, and Sanders (1992) and He (1990). A general framework for deriving appropriate moment conditions to be used in method of moments estimation is discussed in Hansen and Scheinkman (1995) and Duffie and Glynn (1996). For an example of how this theory applies, see Conley, Hansen, Luttmer, and Scheinkman (1997).

Aït-Sahalia (1996a, 1996b) considers non-parametric approaches to the estimation prob-

lem. One such approach is to match the unconditional density of the process as it is implied by the particular specification of drift and diffusion functions, to a corresponding non-parametric one estimated from sample observations. Aït-Sahalia also considers a metric exploiting the information in the transitions of the process. One advantage of non-parametric methods is that they potentially allow for flexible, non-parametric estimation of the drift and/or the diffusion functions. For example, Aït-Sahalia (1996b) estimates a general diffusion function in a one factor model for the short term interest rate by parameterizing the drift function. Stanton (1997) develops a framework allowing for a general form for both the drift and diffusion functions. As in Kessler (1997), this approach utilizes a Taylor type expansion of the drift and diffusion functions to approximate the discrete mean and variances of the transitions of the process. From corresponding non-parametric estimates of the discrete means and variances, the Taylor type expansions can then be inverted to find estimates of the local drift and diffusions.

Like the estimator considered in this paper, some previous work has focused on simulation based inference. Typically these methods work with a discretized approximation to the process and use *unconditional* simulation in the sense that the simulated paths of the process are conditioned only on its initial values. Pedersen (1995) points out that if some m missing points on the path connecting two consecutive discrete observations are used as an approximation to this path, then the transition density can be approximated by simulation using the Euler scheme. The resulting approximation can be made arbitrarily accurate by increasing m and the number of Monte-Carlo experiments allowing for a simulated maximum likelihood approach to be carried out.

The method of indirect inference (Gourieroux, Monfort, and Renault (1993)) and the "Efficient Method of Moments" (EMM) (Gallant and Tauchen (1996)) are general methods for solving difficult estimation problems. The essential idea in these methods is to first fit some "auxiliary" discrete time model to the data. This model should be sufficiently general to approximate the actual dynamics of the discrete observations. The next step consists of choosing the parameter vector in the diffusion model in such a way that when the diffusion model is simulated using a fine grid, the simulated sample paths matches the properties of the discrete time model fitted in the first step. The EMM method the criterion function is based on the scores of the auxiliary model, which consequently, is referred to as the "score generator". An attractive feature of indirect inference methods is that they apply in general settings such as when the diffusion model includes unobserved variables. This is not the case for the aforementioned methods.

In this paper, we study a simulation based approach that relies on a Bayesian formulation of the problem. Our approach utilizes the recent and yet rapidly growing research on conditional simulation strategies in Bayesian statistics, known as Markov Chain Monte Carlo (MCMC). These methods have previously proven particularly useful for handling estimation

problems involving latent variables or missing values. For example, MCMC methods have been successfully applied to discrete time stochastic volatility models (i.e., Shephard (1993), Jacquier, Polson, and Rossi (1994) and Kim, Shephard, and Chib (1996) among others). Kim et al. also point out that the methodology applies to observations missing in the time domain.

Our setup is similar to that of Pedersen (1995) in that the estimation problem is treated as a missing values problem. Hence, some m evenly spaced points on the trajectory between consecutive observations are simulated. These missing data have to be integrated out of joint posterior to obtain the marginal posterior of the parameter vector. The principal difference between the two approaches is that while Pedersen uses the Euler scheme to carry out these simulations, we sample the missing data from the posterior density. Because the posterior density is defined conditional upon on all the observations of the process, the two methods are distinctly different: Loosely speaking, when the sample paths are simulated conditional upon the observations, they "connect" with the observations in a Brownian Bridge sense (see for instance Karlin and Taylor (1981) for a discussion on the Brownian bridge). By contrast, the Euler scheme is likely to generate "jumps" between the simulated sample paths and the observations.

A principal advantage in the use of the conditional simulation strategy outlined above, is that it applies to models incorporating unobserved "state" variables. Examples of such models include the family of continuous time stochastic volatility models for financial assets. For these models, unconditional simulation strategies such as the Euler scheme, do not apply for the same reasons that they do not apply in discrete time (see Danielsson (1994) for a discussion of the discrete time case).

A general algorithm is proposed to sample the unobserved paths of the process: Subject to minimal conditions, it applies regardless of the functional forms of the drift and diffusion functions. Moreover, it applies equally well when the system includes unobservable state variables. This contrasts the usual discrete time case where algorithms for handling unobserved components have to be designed on a problem by problem basis. The design of a general algorithm is possible because the conditional posterior density of one particular element of the diffusion process is approximately Gaussian over a sufficiently small time interval. Hence, the move from discrete to continuous time has certain appeal, even from an econometric viewpoint.

The simulated paths of the process can be used in two ways to construct estimates of the unknown parameter vector. First, if a large number of sample paths are generated conditional upon one value of the parameter vector, one can obtain the marginalized posterior simply by averaging. A new value for the parameter vector is obtained by maximizing this estimate of the marginal posterior. As this process is repeated, it results in a simulated EM estimator. A second approach, which is the one taken here, is to update the parameter vector

conditional upon each draw of the simulated trajectories to form a Gibbs sampler. This step is typically equivalent to conducting inference in a regression model. Consequently, whenever the parameters enter into the drift and diffusion functions in a linear way, standard linear time series methods are applicable.

The remainder of the paper is organized as follows: In the next 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 constant elasticity of variance model of the short term interest rate as well as a generalization of this model to allow for a stochastic volatility component. Section three contains the empirical estimates obtained from fitting the example models to the short term US interest rate. We also present some evidence from sampling experiments designated at assessing the accuracy of the MCMC solution for the stochastic volatility model. Section four 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. We assume that the process

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

generates discrete time observations and our overall objective is to conduct inference for the parameter vector, ρ. 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

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

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

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

A2 Y is stationary and ergodic.

A3 $\sigma(x;\rho)\sigma(x;\rho)'$ is positive definite for all $x \in \Re^d$ where $'$ denotes the matrix transpose.

A4 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) (see Karatzas and Shreve (1991) theorem 5.2.9 or Øksendal (1995), theorem 5.5). A2 can sometimes be relaxed by a simple reformulation of the problem, for instance, by removing a constant drift and running the analysis on the transformed series. Notice however, that the usual discrete time approach of taking the first order difference may not yield a modified process that can be represented as a stochastic differential equation. In accordance with A4, 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 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 shall work with the discretized version of (1) given by

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

where $\triangle W_t$ are i.i.d. $N(0, \triangle t)$ variates. 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 missing values problem. 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 $d_1 T(m-1)$ data points are missing from the observable part of the system while $d_2 Tm$ 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.,

$$\hat{Y} = \begin{bmatrix} X_{1,t_0} & \hat{X}_{1,t_1} & \cdots & X_{1,t_m} & \hat{X}_{1,t_{m+1}} & \cdots & X_{1,t_n} \\ X_{2,t_0} & \hat{X}_{2,t_1} & & X_{2,t_m} & \hat{X}_{2,t_{m+1}} & \cdots & X_{2,t_n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ X_{d_1,t_0} & \hat{X}_{d_1,t_1} & \cdots & X_{d_1,t_m} & \hat{X}_{d_1,t_{m+1}} & \cdots & \hat{X}_{d_1,t_n} \\ \hat{Z}_{1,t_1} & \hat{Z}_{1,t_2} & \cdots & \hat{Z}_{1,t_m} & \hat{Z}_{1,t_{m+1}} & & \hat{Z}_{1,t_n} \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ \hat{Z}_{d_2,t_1} & \hat{Z}_{d_2,t_2} & \cdots & \hat{Z}_{d_2,t_m} & \hat{Z}_{d_2,t_{m+1}} & \cdots & \hat{Z}_{d_2,t_n} \end{bmatrix}.$$

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

The following additional assumption is made:

A5 The econometrician holds uninformative priors for the parameter ρ and the missing data points.

The assumption of uninformative priors for ρ is easily relaxed. For instance, for the example models considered, the relevant conditional distributions for sub-blocks of ρ are Gaussian and Gamma respectively, implying that standard conjugate priors are available. [1] Conditioning on the first observation the joint posterior density is then given by

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

where

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

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 $\mu_i := \mu(\hat{Y}_i; \rho)$, $\sigma_i := \sigma(\hat{Y}_i; \rho)$ 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, $\pi(\rho \mid \hat{Y})$ are proportional to (5).

[1] In some rare cases, it might also be advisable to work with informative priors for the missing data. For instance, if the model implies that the diffusion process is positive (say), the specification of a prior with support on the positive real line may prevent the simulated missing paths from becoming negative which could happen as a result of discretization. An example where this theoretically possible, is the CEV models considered in later sections.

2.1 Gibbs sampling

Like with typical applications of MCMC, our goal is to obtain a sequence of Monte-Carlo samples, $\{\rho^{(h)}\}_{h=1}^{M}$ of the parameter vector from the marginal posterior density. We follow a commonly used notational shorthand and henceforth let $\pi(.)$ generically denote all posterior densities (i.e., π will generally take on different forms depending on its arguments). Similarly, we let $p()$ generically denote $\pi()$ 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, observed and unobserved data, and then to integrate the unobservables out of this joint density. To perform this integration, it is sufficient that one is able to compute a simulation sequence, say $\{\rho^{(h)},\ \textit{missing observations}^{(h)}\}_{h=1}^{M}$ from the posterior (5). Then, the sequence $\{\rho^{(h)}\}_{h=1}^{M}$ is implicitly a sample from the marginal posterior $\pi(\rho \mid \textit{observations})$. 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 $\rho^{(h)} \sim \pi(\rho \mid \textit{missing observations}^{(h-1)}, \textit{observations})$ in one operation, one could in principle sample $\textit{missing observations}^{(h-1)} \sim \pi(\textit{missing observations} \mid \textit{observations}, \rho^{(h)})$ in a second step. By repeating this process, the sequence of simulated parameters and missing data forms a Markov chain with stationary distribution being 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

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

where

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

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 - a consequence of the Markov property of the 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 quadratic terms entering into the exponent have roots that, loosely speaking, lie 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 the fortunate property that it has approximately the shape of a normal density for $\triangle t$ sufficiently small.

It is important to distinguish between the conditional density $p(\hat{Y}_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}; \rho)$ and the transition density $p(\hat{Y}_i \mid \hat{Y}_{i-1}; \rho)$. 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

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

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 part of the system. Notice also that it is sufficient to simulate these elements from $\pi(\hat{Y}_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}; \rho)$ since

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

As the Gibbs sampler iterates, is produces a trajectory that "passes through" the successive discrete observations. If, on the other hand, we were to simulate from $p(\hat{Y}_i \mid \hat{Y}_{i-1}; \rho)$ we would most likely obtain a large jump from $\hat{X}_{i-1}$ to the observation X_i.

The next step in setting up a Gibbs sampler for this problem is to device some method to simulate from $\pi(\hat{Y}_i \mid \hat{Y}_{i-1}, \hat{Y}_{i+1}; \rho)$. 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 $\rho = \{\mu, \sigma\}$ and the following result obtains

Proposition 2.1. *For a scalar process, $Y_t \in \Re$ with constant drift and diffusion functions,*

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

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

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

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

□

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 (see Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) or Albert and Chib (1996) for a recent treatment) was successfully applied to one factor models for the interest rate. The results obtained with this algorithm however, showed signs of slow convergence. We consequently suggest to use the hybrid accept/reject Metropolis Hastings algorithm (see Tierney (1994)). This algorithm is very general and requires only 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 A for a brief review. We suggest to use 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, $q(x \mid Y_{i-1}, Y_{i+1}, \rho)$ in the hybrid rejection Metropolis Hastings 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 2.2. *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 $\triangle t \to 0$.

Proof. First, note that while the limit $\lim_{\triangle t \to 0} \hat{Y}_i/\sqrt{\triangle t}$ does not exists, 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,

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

Similarly,

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

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

$$\begin{aligned} p(Y_i \mid .) \propto \left|\sigma_{i-1}^{-2}\right|^{\frac{1}{2}} \left|\sigma_i^{-2}\right|^{\frac{1}{2}} \exp\Bigg\{ & -\frac{1}{2}\left\|\left(DY_i - \mu_{i-1}\sqrt{\triangle t}\right)\sigma_{i-1}^{-1}\right\| \\ & -\frac{1}{2}\left\|\left(DY_{i+1} - \mu_i\sqrt{\triangle t}\right)\sigma_i^{-1}\right\|\Bigg\} \\ = \left|\sigma_{i-1}^{-2}\right|^{\frac{1}{2}} \Bigg|\left(\sigma_{i+1} + O_p(\sqrt{\triangle t})\right)^2\Bigg|^{\frac{1}{2}} \exp\Bigg\{ & -\frac{1}{2}\left\|\left(DY_i - O_p(\triangle t)\right)\sigma_{i-1}^{-1}\right\| \\ & -\frac{1}{2}\left(DY_{i+1} - O_p(\triangle t)\right)\left(\sigma_{i-1} + O_p(\sqrt{\triangle t})\right)^{-1}\left(DY_{i+1} - O_p(\triangle t)\right)'\Bigg\} \\ \to \left|\sigma_{i-1}^{-2}\right| \exp\Bigg\{ & -\frac{1}{2}\left\|DY_i\sigma_{i-1}^{-1}\right\| - \frac{1}{2}\left\|DY_{i+1}\sigma_{i-1}^{-1}\right\|\Bigg\} \\ \propto \left|(\frac{1}{2}\sigma_{i-1})^{-2}\right|^{\frac{1}{2}} \exp\Bigg\{ & -\left\|\left((\hat{Y}_i - \frac{1}{2}(\hat{Y}_{i-1} + \hat{Y}_{i+1}))\triangle t^{-\frac{1}{2}}\right)\sigma_{i-1}^{-1}\right\|\Bigg\} \end{aligned} \quad (8)$$

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

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:

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

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 $x = m_i$.

For the example applications considered here, this proposal density provided good results and the acceptance probability ranged $0.9 - 0.98$ for the various one-factor models considered. Moreover, the approximation generally proved better for small $\triangle t$ which is in line with expectations in light of 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, 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 this 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 $\rho^{(h+1)} \sim \pi(\rho \mid \hat{Y}^{(h)})$. The general form of this conditional (unnormalized) density is again given by (5). Conditioning on the data, this density corresponds to a general non-linear, heteroscedastic system. Thus, the exact form of this density has to be determined on a problem by problem basis. In general, whenever this density is not recognized as one with a standard form, we might impose a Metropolis Hastings step, sampling $\rho^{(h+1)}$ conditional upon $\rho^{(h)}$.

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. We now present two examples for illustration and subsequently discuss the principle in a more general setting.

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

$$dY_t = (\rho_1 + \rho_2 Y_t)dt + \rho_3 Y_t^{\beta} dW_t \tag{9}$$

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 (9) into (5) we get

$$\pi(\rho \mid \hat{Y}) \propto \prod_{i=1}^{n} \frac{1}{\rho_2 \hat{Y}_{i-1}^{\beta}} \exp\left\{ -\frac{1}{2} \frac{\left(\hat{Y}_i - \hat{Y}_{i-1} - (\rho_1 + \rho_2 \hat{Y}_{i-1})\triangle t\right)^2}{\rho_2^2 \hat{Y}_{i-1}^{2\beta} \triangle t} \right\}. \tag{10}$$

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 (10) as the likelihood function corresponding to a linear regression of y on X. Consequently

$$(\rho_1, \rho_2) \quad \mid \quad \rho_3, \hat{Y} \sim N(\overline{\rho}, \rho_3^2 (X'X)^{-1}) \tag{11}$$

$$\rho_3^{-2} \quad \mid \quad \hat{Y} \sim IG(n-2, \overline{s}^2) \tag{12}$$

where $\overline{\rho} = (X'X)^{-1}(X'y)$ and $\overline{s}^2 = \frac{1}{n}\sum_i (y_i - X_i\overline{\rho})^2$ and where $N()$ and $IG()$ denote multivariate normal and inverted Gamma densities, respectively. Sampling from these densities is easily accomplished. 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 $\hat{Y}_i^{(h)} \mid \hat{Y}_{\setminus i}^{(h-1)}, \rho$ 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 $(\rho_1, \rho_2)^{(h)}$ using (11)
4. Draw $\rho_5^{(h)}$ using (12)
5. Increase h and return to step 2.

Notice the obvious difficulty obtained if we let β be a free parameter to be estimated. Since this parameter enters into the posterior in a non-linear way, it is difficult to obtain the exact posterior density of this parameter. If β is to be treated as a free parameter in this model, an alternative strategy is to use a Random Walk Metropolis Hastings step to sample this parameter. This adds an additional step in the Gibbs sampler. Notice however that this parameter is likely to be highly correlated with other parameters entering the model. As a

consequence, modeling β as a free parameter is likely to slow down the convergence of the Gibbs sampler.

Example 2: Continuous time stochastic volatility

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

$$\begin{aligned} dX_t &= (\rho_1 + \rho_1 X_t)\, dt + \exp\left(\frac{1}{2} Z_t\right) X_t^{\beta} dW_{1,t} \\ dZ_t &= (\rho_3 + \rho_4 Z_t)\, dt + \rho_5 dW_{2,t} \end{aligned} \tag{13}$$

where X_t is assumed to be observed at discrete intervals $t = 0, 1, .., T$ 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 process, X_t has been proposed as a model of the short term interest rate by, among others, Andersen & Lund (1997a,1997b). If we disregard the proportionality term, X_t^{β} in the diffusion for X_t(i.e., set $\beta = 0$), it represents a basic extension of the usual discrete time stochastic volatility model to a continuous time setting.

In obtaining samples of ρ 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 $\{\rho_1, \rho_2\}$ and the log-volatility process $\{\rho_3, \rho_4\}$ and ρ_5. As we have

$$\mu(Y_i; \rho) = \begin{bmatrix} \rho_1 + \rho_2 X_i \\ \rho_3 + \rho_4 Z_i \end{bmatrix}$$

$$\sigma(Y_i; \rho) = \begin{bmatrix} \exp\left(\frac{1}{2} Z_i\right) X_i^{\beta} & 0 \\ 0 & \rho_5 \end{bmatrix}$$

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

$$\pi(\rho \mid \hat{Y}) \propto p(\rho_1, \rho_2 \mid \hat{Y}) p(\rho_3, \rho_4, \rho_5 \mid \hat{Y})$$

where

$$\begin{aligned} p(\rho_1, \rho_2 \mid \hat{Y}) &= \prod_{i=1}^{n} \exp\left(-\frac{1}{2} Z_i\right) X_i^{-\beta} \triangle t^{-1/2} \exp\left\{-\frac{(\triangle X_i - (\rho_1 + \rho_2 X_i)\, \triangle t)^2}{2 \exp(Z_t) X_i^{2\beta} \triangle t}\right\} \\ p(\rho_3, \rho_4, \rho_5 \mid \hat{Y}) &= \prod_{i=1}^{n} \rho_5\ \triangle t^{-1/2} \exp\left\{-\frac{(\triangle Z_i - (\rho_3 + \rho_4 Z_i)\, \triangle t)^2}{2 \rho_5^2 \triangle t}\right\} \end{aligned}$$

which have forms similar to that of the likelihoods 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 $\{\rho_1, \rho_2\}, \{\rho_3, \rho_4\}$ and ρ_5 as separate blocks.

We conclude the discussion on the continuous time stochastic volatility model with some notes on the model in the case of $\beta = 0$. Since the log-volatilities have a first order autoregressive structure, a different Gibbs sampling scheme can then be set up to simulate the missing paths: Suppose we simulate the missing values of X in one step of the Gibbs sampler, conditional upon the volatilities using the results of the previous section. We can then sample the volatilities conditional upon the data using the same algorithm, or, we can apply the results of Shephard and Pitt (1997) to use rejection sampling for the volatilities. This algorithm runs at least twice as fast as the hybrid Rejection Metropolis Hastings algorithm and almost always accepts the candidates. Hence, for this particular model it is possible to set up a Gibbs sampler which is considerably more efficient than the general algorithm given above. The cost, however, is that of limited applicability.

A final comment on the case for $\beta = 0$: Suppose we run the algorithm above using $\triangle t = 1$ and hence $m = 1$. The number of missing observations of X is then $m - 1 = 0$ and we can simulate only the missing log-volatilities, Z_t, at $t = 0, 1, 2, ...$ Thus, the whole algorithm reduces to the algorithm used by Jacquier, Polson, and Rossi (1994) with the exception of slightly different means and variances of the proposal densities in the volatility simulations. As a consequence, the computer program used to generate the empirical results reported in the next section, could be used to estimate the standard discrete time stochastic volatility model by just setting $\triangle t = 1$.

2.4 Drift and diffusions linear in the parameters

As we have seen for the CEV model and the stochastic volatility model, the posterior can sometimes be recognized as the likelihood function of a general linear model. Both models are special cases of the following general model: Suppose that the drift and diffusion functions take on forms

$$\begin{aligned} \mu(x;\rho) &= \rho_1 h_1(x) \\ \sigma(x;\rho) &= h_2(x)\rho_2 \end{aligned}$$

where $\rho = \{\rho_1, \rho_2\}$ and $h_1 : \Re^d \to \Re^{q\times d}$, $h_2 : \Re^d \to \Re^{d\times d}$ and ρ_i are $1\times q$ and $d\times d$ dimensional matrices respectively. Assuming that $h_2(x)$ and ρ_2 is invertible for all x, we have

$$\begin{aligned} \sigma_{i-1}^2 &= \sigma(x;\rho)\sigma(x;\rho)' = h_2(x)\rho_2\rho_2' h_2(x)' \\ \sigma_{i-1}^{-2} &= \left[\sigma(x;\rho)\sigma(x;\rho)'\right]^{-1} = \left[h_2(x)\right]^{-1}\left[\rho_2\rho_2'\right]^{-1}\left[h_2(x)'\right]^{-1} \end{aligned}$$

If we define the $1 \times d$ dimensional vector $\triangle \hat{Y}_i^* := (\hat{Y}_i - \hat{Y}_{i-1}) h_2(\hat{Y}_{i-1})^{-1}$ and a $q \times d$ dimensional matrix $h_1^*(\hat{Y}_{i-1}) = h_1(\hat{Y}_{i-1}) h_2(\hat{Y}_{i-1})^{-1}$ the conditional posterior (5) can be written

$$p(\rho \mid \hat{Y}) = \left| \sum_{i=1}^{n} \left[h_2(\hat{Y}_{i-1}) \right]^{-1} \left[\rho_2 \rho_2' \right]^{-1} \left[h_2(\hat{Y}_{i-1})' \right]^{-1} \right|^{\frac{1}{2}} \times \exp\left[\sum_{i=1}^{n} -\frac{1}{2} \left(\triangle \hat{Y}_i - \rho_1 h_1(\hat{Y}_{i-1}) \triangle t \right) \left[h_2(\hat{Y}_{i-1}) \right]^{-1} \left[\rho_2 \rho_2' \right]^{-1} \left[h_2(\hat{Y}_{i-1})' \right]^{-1} (\triangle t)^{-1} \left(\triangle \hat{Y}_i - \rho_1 h_1(\hat{Y}_{i-1}) \triangle t \right)' \right].$$

Since h_2 is a function of the data only,

$$\pi(\rho \mid \hat{Y}) \propto \left| n \left[\rho_2 \rho_2' \right]^{-1} \right|^{\frac{1}{2}} \exp\left\{ \sum_{i=1}^{n} -\frac{1}{2} \left(\triangle \hat{Y}_i^* - \rho_1 h_1^*(\hat{Y}_{i-1}) \triangle t \right) \left[\rho_2 \rho_2' \right]^{-1} (\triangle t)^{-1} \left(\triangle \hat{Y}_i^* - \rho_1 h_1^*(\hat{Y}_{i-1}) \triangle t \right)' \right\}$$

This equation corresponds to the likelihood of a standard linear multivariate system and the conditional posteriors are accordingly given by multivariate normal and Wishart densities, respectively.

2.5 Convergence and starting values

In general, convergence of the Gibbs sampler is difficult to assess. In typical applications, convergence is assessed 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 serious 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 all expressions for the conditional posteriors collapses to Dirac delta functions in the limit. Consequently, the sampler does not move from its starting 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.

One strategy that has proved successful in the empirical applications 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 it does not seem to have any effect on the empirical distribution of the output from the chain.

3 Empirical application to US interest rates

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 filings of daily market data[2], and are essentially the same as those used by Andersen & Lund (1997a, 1997b) with the exception of 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-dept exploratory analysis of these data is given in Andersen and Lund (1997b).

Table 1
Descriptive Statistics

	mean	median	st.d	skewness	kurtosis
level	0.05552	0.05140	0.02857	1.08645	4.59365
differenced	0.00002	0.00000	0.00251	-0.48679	22.40646
			autocorrelations		
lag	1	2	3	4	5
level	0.99614	0.99158	0.98672	0.98141	0.97565
differenced	0.09184	0.03790	0.05679	0.05819	0.05416

3.1 Empirical results for the CEV model

We now turn to the results obtained from fitting the constant elasticity of variance model (CEV) to the interest rate data. We report the results for different fixed values of the CEV parameter, β. The results were obtained by following the strategy outlined in the previous section: We ran the sampler for some 1200 iterations with $\triangle t = 1/2$. Then the sampler was run with $\triangle t = 1/4$ (and hence one missing observation) for another 400 iterations before $\triangle t$ was reduced 1/8 at iteration 1500. We ran a total of 10,000 Gibbs iterations for this model corresponding to about a half hour of Origin 2000 CPU time. This figure is comparable to 1

[2]The data are available at URL: http://www.bog.frb.fed.us/Releases/H15/data/h15.zip

Table 2

Posterior moments for the CEV model

The table reports posterior means and variances for parameters in the CEV diffusion for fixed β. Function values are means of the unnormalized log-posteriors. $\Delta t = 1/8$. The results are based on 10,000 Gibbs iterations with 3,000 discarded for "burn in". Nummerical standard errors (NSE) for the posterior means are calculated from a Bartlett-window with maximum lag of 1000.

function		ρ_1	ρ_2	ρ_3
		$\beta = 1$		
105091.4	mean	0.000092	−0.000858	0.040256
	st.dev.	0.000043	0.001354	0.000576
	NSE	0.000000	0.000019	0.000033
		$\beta = 0.75$		
105803.2	mean	0.000134	−0.001995	0.018158
	st.dev.	0.000055	0.001391	0.000275
	NSE	0.000001	0.000033	0.000021
		$\beta = 0.5$		
105041.3	mean	0.000186	−0.003024	0.008873
	st.dev.	0.000071	0.001511	0.000125
	NSE	0.000001	0.000004	0.000006
		$\beta = 0$		
100435.3	mean	0.000257	−0.004323	0.002509
	st.dev.	0.000114	0.001817	0.000039
	NSE	0.000001	0.000026	0.000003

1/2 hours of Intel P233 CPU time. The code was not optimized for these particular models, indicating that lower CPU times are attainable.

The output from the Gibbs sampler in the case of $\beta = 0.5$ is plotted in figure 3.1. The figure shows the so-called batch means (means of successive Gibbs iterations in batch sizes of 100). There is no visual evidence of a trend in the data for ρ_1 and ρ_2. For the diffusion parameter, ρ_3, there seem to be some trend for at least the first 2000 iterations. As a consequence, the posterior means reported in table 2 are constructed by discarding the first 3000 iterations. Notice also that the magnitude of variations in the batch means are small, indicating that each sample mean is a relatively good estimator of the posterior expectations.

Table 2 presents the posterior means and standard deviations of the CEV model. The estimates of ρ_2 implies that the data are near non-stationary. Another aspect of the results in this table is that as β increases, the mean reversion parameter, ρ_2, decreases. These results are consistent with the findings of Tauchen (1997) who uses EMM to estimate this model with a sample of similar data.

In the case where the CEV coefficient, β, is zero, the process is identical to the O-U process. The parameters can easily be estimated by maximum likelihood in this case and we present these estimation results in table 3 for comparison. Clearly, the Bayesian MCMC estimates reported for the O-U process ($\beta = 0$) in table 2 are very close to those estimated by MLE reported in table 3; the parameters and standard errors obtained with the two different methods differ only in the fifth or sixth decimal place.

FIGURE 1
BATCH MEANS FROM GIBBS SAMPLER FOR SQUARE ROOT DIFFUSION ($\beta = .5$)

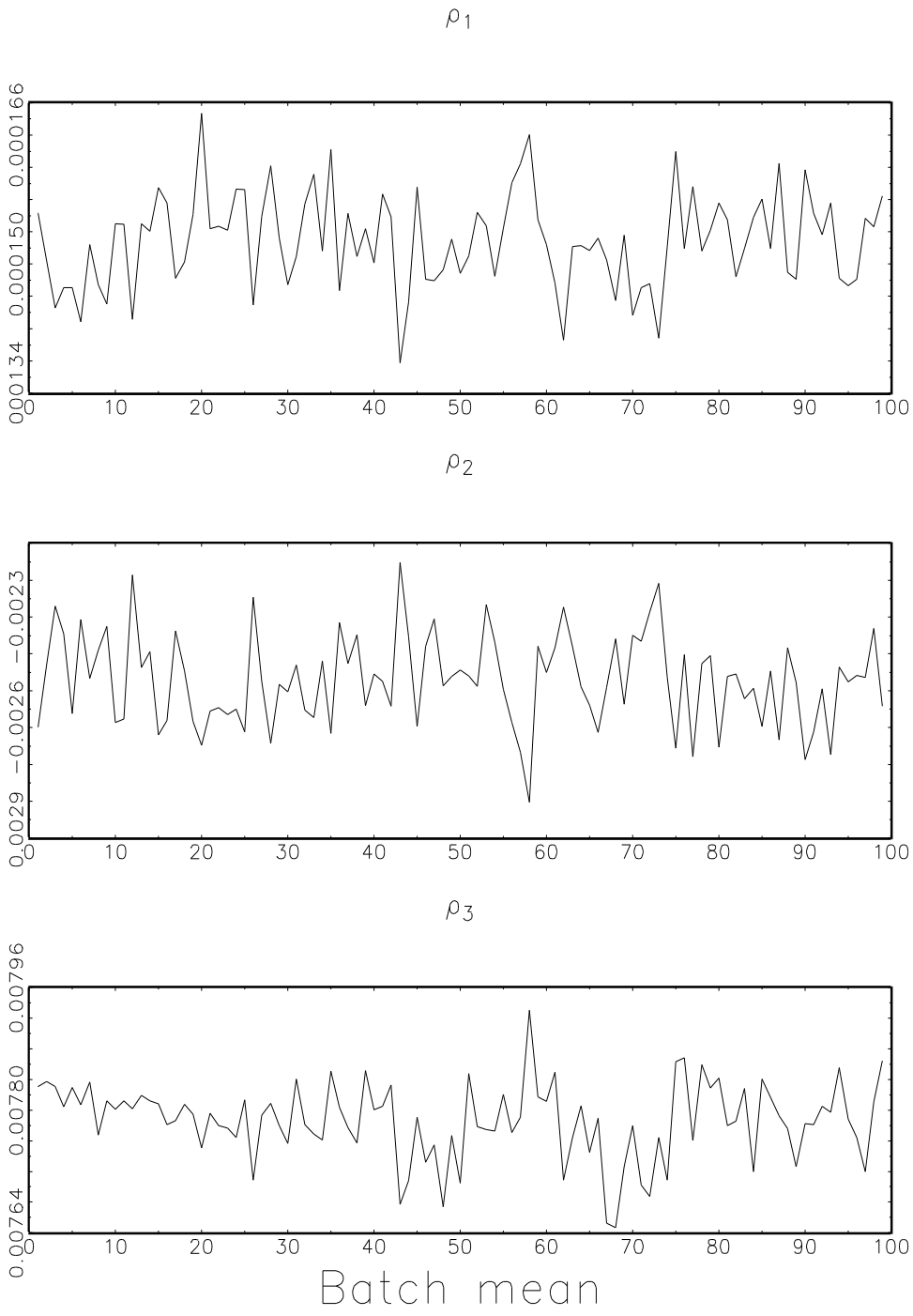

TABLE 3

MLE ESTIMATES OF PARAMETERS FOR THE O-U PROCESS

The table reports the exact maximum likelihood estimates of the parameters in the Ornstein-Uhlenbeck process estimated on the US Tbill rates. Standard errors are shown in parentheses.

ρ_1	ρ_2	ρ_3
0.000258	-0.004346	0.002500
(0.000114)	(0.001828)	

3.2 Empirical results for the stochastic volatility model

Following the procedures outlined in the previous sections, we fitted the stochastic volatility model in (13) to the US Tbill rates. The Gibbs sampler was initialized by setting the log-volatility process to $log(y_t/100)$. Other starting values, such as the Garch estimate of conditional volatility and initial constant conditional volatility, provided identical results. A total of 100,000 Gibbs iterations were computed corresponding to some 8 1/2 hours of Origin 2000 CPU time for each model. Batch means from the Gibbs sampler are plotted in figure 2. By visually inspecting the output from the Gibbs sampler in figure 2, it is clear that the sampler provides highly dependent samples (notice that this figure covers a posterior sample size 10 times that of the previous results). The high dependence is also evident from other diagnostics, such as autocorrelations. For example, the first order autocorrelations for the parameters in the volatility process are all above 0.99. Another characteristic of the plots in figure 2 is the high correlation between the parameters. For example, ρ_3 and ρ_4 are almost perfectly correlated (correlation coefficient is 0.9997). This has no impact on the dependency of the successive Gibbs iterations because these parameters are drawn jointly from the (bivariate) conditional posterior density. Nevertheless, the high dependence experienced with the SV model, suggests the usefulness of acceleration techniques such as re-parameterization and variance reduction techniques for this particular model.

Table 4 presents the estimation results for the stochastic volatility model. Judging from the estimates of ρ_2 reported, the results suggest that the data are close to non-stationary. Also, the estimates of ρ_4 imply that the volatility process is highly persistent. This is in line with evidence from the discrete time literature; GARCH and discrete time SV models applied to these data, typically imply highly persistent estimates of conditional volatility. On the other hand, the estimates reported in table 4 differ markedly from those reported by Andersen and Lund (1997a). In their application of EMM to a slightly shorter sample of the same data, they obtain an estimate of ρ_4 close to -1 - a large difference from the results obtained here.

So what is the better model for the short term US interest rate? At this point, we wish to remain deliberately vague on this point. The comparison of various models in our Bayesian setting calls for the calculation of Bayes factors. Bayes factors are the ratio of the normal-

FIGURE 2
BATCH MEANS FROM GIBBS SAMPLING FOR THE SV MODEL, β=.5

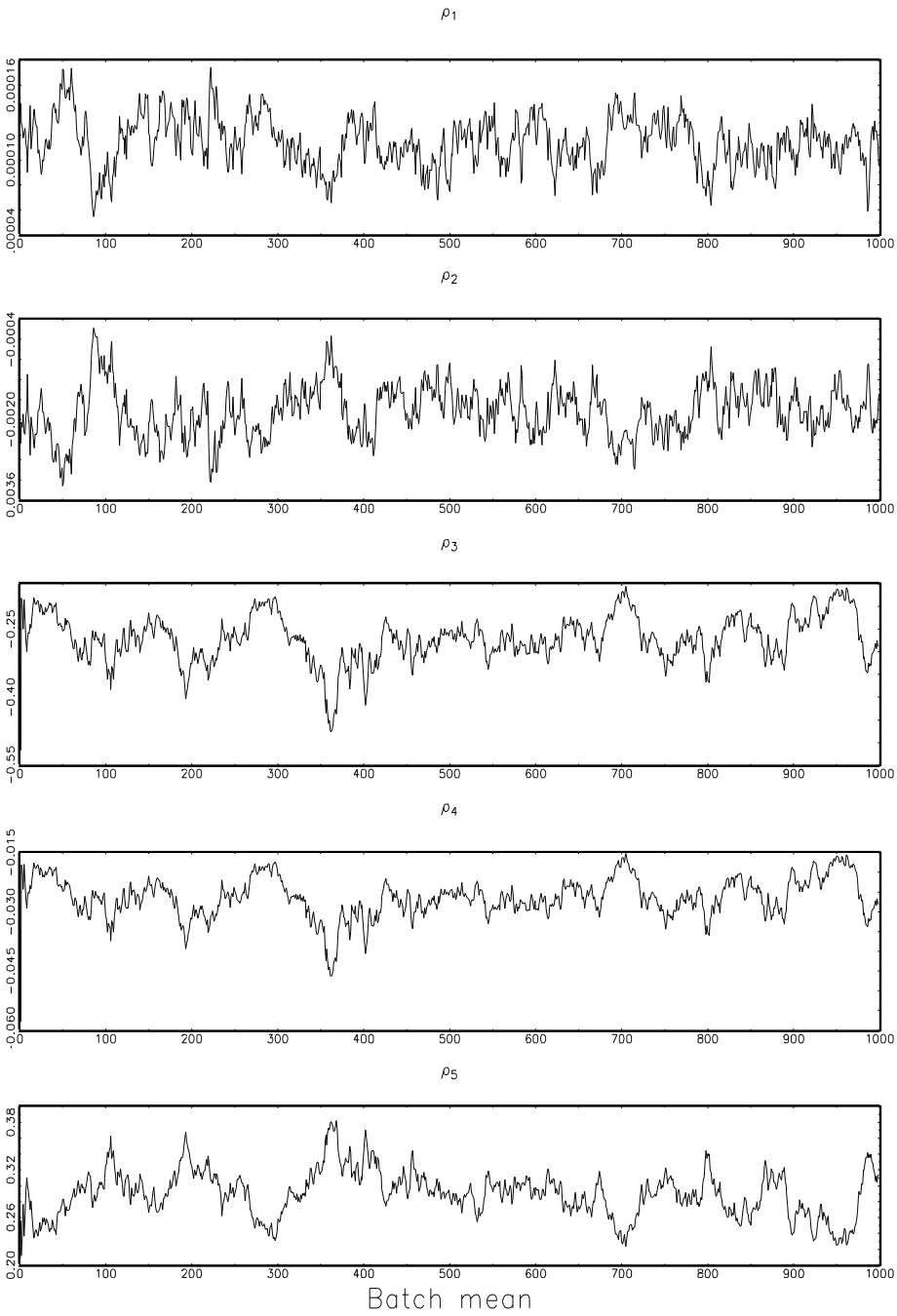

TABLE 4

POSTERIOR MOMENTS FOR THE STOCHASTIC VOLATILITY MODEL

The table reports posterior means and variances for parameters in the stochastic volatility model for fixed β. Function values are means of the unnormalized log-posteriors. $\triangle t = 1/8$. The results are based on 100,000 Gibbs iterations with 10,000 discarded for "burn in". Nummerical standard errors (NSE) for the posterior means are calculated from a Bartlett-window with maximum lag of 5000.

function		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
			$\beta = 1$			
125448.5	mean	0.00011	−0.00165	−0.20887	−0.02867	0.30260
	st.dev.	0.00002	0.00051	0.03851	0.00528	0.02743
	NSE	0.00000	0.00006	0.00650	0.00089	0.00455
			$\beta = 0.75$			
127520.3	mean	0.00012	−0.00198	−0.21465	−0.02446	0.26845
	st.dev.	0.00002	0.00049	0.03942	0.00452	0.02616
	NSE	0.00000	0.00006	0.00664	0.00076	0.00442
			$\beta = 0.5$			
126053.3	mean	0.00011	−0.00183	−0.27284	−0.02658	0.29141
	st.dev.	0.00002	0.00051	0.05213	0.00506	0.02963
	NSE	0.00000	0.00006	0.00848	0.00082	0.00487
			$\beta = 0$			
127396.1	mean	0.00011	−0.00198	−0.21521	−0.01623	0.26941
	st.dev.	0.00002	0.00049	0.03884	0.00293	0.02574
	NSE	0.00000	0.00006	0.00702	0.00053	0.00476

izing constants for the posterior densities of two models. Unfortunately, these normalizing constants are not easily estimable when the model setup contains a large number of unobservables such as here. The issues of specifications testing, model selection and comparisons with discrete time models, are consequently left for future work.

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 huge computational demands, the analysis is limited to one set of parameter values suggestive of the results reported in table 4 corresponding to $\beta = 0.5$. Each experiment was repeated 200 times and the resulting parameter estimates are summarized in table 5. 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, ρ_2, which measures the speed of the mean reversion of the (artificial) interest rate process, has "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, ρ_4, the estimates are also slightly downward biased, while the estimate of

TABLE 5

SAMPLING EXPERIMENTS FOR THE STOCHASTIC VOLATILITY MODEL

The table reports means, medians and root mean squared errors (RMSE) for parameters estimated on data generated according to the stochastic volatility model with $\beta = 0.5$. 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 $10,000 - 20,000$.

	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
			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

the diffusion coefficient for the log-volatility process, ρ_5, is virtually unbiased using a sample size of 2000.

When interpreting the results in table 5, it should be kept in mind that the reversion coefficients can be interpreted as approximately "autocorrelations - 1". This, coupled with the well known fact that sample autocorrelation coefficients are downward biased for processes with near unit roots may, at least partially, explain 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. This does not affect the biases of our estimates since these errors have zero mean provided that the sampler has been run with a sufficiently long "burn in"-period. It does, however, influence the dispersion of the posterior means, as measured by the RMSE. Hence, the results in table 5 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 5 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 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 result is driven by 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 (1996) hybrid rejection Metropolis Hastings algorithm to simulate one point on the unobserved paths by proposing from a normal density. 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, this problem typically reduces to simulating from the posterior of a regression model.

Applications of the proposed methodology include the constant elasticity of variance model (CEV) and stochastic volatility (SV) models. For the latter model, the successive Gibbs iterations contain a large amount of serial dependence. This is undesirable from a practical viewpoint because it increases the number of iterations necessary to obtain accurate estimates in the sense that they are not influenced by Monte Carlo error. On the other hand, comparisons between MCMC and MLE estimates of parameters in the Ornstein-Uhlenbeck process as well as the sampling experiments conducted for the SV model, are two benchmarks measuring the accuracy of the method. These benchmarks both suggest that the MCMC solution is largely successful.

Appendix

A 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 unnormalized 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.

B Computational issues

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.

References

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

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

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

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

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

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

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

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

Cox, J. C., J. E. Ingersoll, and S. A. Ross (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.

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

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

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

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

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

He, H. (1990): "Moment Approximation and Estimation of Diffusion Models of Asset Prices," *Working Paper, University of California, Berkley*.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994): "Bayesian Analysis of Stochastic Volatility Models (with discussion)," *Journal of Business and Economic Statistics*, 12, 371–417.

KARATZAS, I., AND S. E. SHREVE (1991): *Brownian Motion and Stochastic Calculus.* Springer.

KARLIN, S., AND H. M. TAYLOR (1981): *A Second Course in Stochastic Processes.* Academic Press, New York.

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

KESSLER, M., AND M. SØRENSEN (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., N. SHEPHARD, AND S. CHIB (1996): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies (forthcomming).*

METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, N. M. TELLER, AND E. TELLER (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., S. A. TEUKOLSKY, W. T. VETTERLING, AND B. P. FLANNERY (1992): *Numerical Recepes in C.* Cambridge University Press.

SHEPHARD, N. (1993): "Fitting Nonlinear Time Series Models with Application to Stochastic Variance Models," *Journal of Applied Econometrics*, 8, S135–S152.

SHEPHARD, N., AND M. K. PITT (1997): "Likelihood analysis of non-Gaussian parameter-driven models," *Biometrika, forthcomming.*

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*, forthcomming.

TANNER, M. A., AND W. H. WONG (1987): "The calculation of posterior distributions by data augmentation (with discussion)," *Journal of the American Statistical Association*, 82, 528–550.

TAUCHEN, G. (1997): "New Minimum Chi-Square Methods in Empirical Finance," in *Advances in Economics and Econometrics: Theory and Applications. Vol III, Econometric Society Monographs no. 28*. D. M. Kreps and K. F. Wallis.

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.

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