Nonlinear State-Space Models with State-Dependent Variances

Jonathan R. Stroud [1]
Department of Statistics, The Wharton School, University of Pennsylvania

Peter Müller
Department of Biostatistics, MD Anderson Cancer Center, Texas Medical Center

Nicholas G. Polson
Graduate School of Business, University of Chicago

Nonlinear state-space models with state dependent variances (SDV) are commonly used in financial time series. Important examples include stochastic volatility (SV) and affine term structure models. We propose a methodology for state smoothing in this class of models. Our smoothing technique is simulation-based and uses an auxiliary mixture model. Key features of the auxiliary mixture model are the use of state dependent weights and efficient block sampling algorithms to jointly update all unobserved states given latent mixture indicators. Conditional on latent indicator variables, the auxiliary mixture model reduces to a normal dynamic linear model. We illustrate our methodology with two time series applications. First, we show how to construct the auxiliary model for a logarithmic SV model and we compare the performance of our methodology with the current literature. Second, we implement a stochastic volatility model with jumps for short-term interest rates in Hong Kong.

Keywords: State-Space Models, State-Dependent Variance, Nonlinear Time Series, Mixture Models, Smoothing, MCMC, Stochastic Volatility, Affine Term Structure Models.

[1]We wish to thank Anne Gron and Michael Johannes for useful discussions. Jonathan Stroud is Assistant Professor at the Department of Statistics, Wharton School, University of Pennsylvania, email: stroud@wharton.upenn.edu. Peter Müller is Professor at the Department of Biostatistics, MD Anderson Cancer Center at the University of Texas, and Associate Professor at the Institute for Statistics and Decision Sciences, Duke University, email: pm@odin.mdacc.tmc.edu. Nick Polson is Professor of Statistics and Econometrics at the Graduate School of Business, University of Chicago, email: ngp@gsbngp.uchicago.edu.

1 Introduction

Nonlinear state-space models with state-dependent variances (SDV) are common in many fields of application, particularly financial time series. Important examples include stochastic volatility (SV) models (Taylor, 1986) and affine term structure models (Duffie and Kan, 1996). Inference for nonlinear SDV models is complicated by the fact that the smoothing distribution of the states is unavailable in closed form. Most existing methods, such as simulated method of moments, do not provide the full smoothing distribution of the states.

Early literature on nonlinear state-space models relied on analytical approximations to provide the smoothing distribution (Anderson and Moore, 1979). Zehnwirth (1988) derived a Kalman filtering approach for models with state-dependent observation variances. More recent approaches use Monte Carlo methods for smoothing. Carlin, Polson and Stoffer (1992) provide a Markov chain Monte Carlo (MCMC) method for nonlinear state-space models with nonnormal errors. Kitagawa (1996) and Geweke and Tanizaki (2001) consider more general classes of models, and use sequential Monte Carlo and MCMC methods, respectively. Carter and Kohn (1994) and Frühwirth-Schnatter (1994) develop an efficient block sampling algorithm for conditionally Gaussian state-space models, and Shephard and Pitt (1997) introduce a related block updating scheme for models with non-Gaussian observations.

In this paper, we propose a new methodology for state smoothing and parameter estimation in nonlinear state-dependent variance (SDV) models. Our methodology draws upon the literature on conditionally Gaussian models. The approach is simulation based and requires the use of an auxiliary mixture model. The weights in the mixture model are allowed to be adaptive and dependent on the state vector. Conditional on latent mixture indicators, the auxiliary mixture model essentially reduces to a linear Gaussian state-space model (West and Harrison, 1997). This allows us to generate the state variables in block fashion with an efficient MCMC algorithm based on the methods of Carter and Kohn (1994), Shephard (1994), Frühwirth-Schnatter (1994) and Carter and Kohn (1996). A key innovation of our methodology is that the auxiliary mixture model weights are state dependent.

We illustrate our approach with two applications. First, we consider a logarithmic SV model and show how to construct the auxiliary mixture model. We compare our methodology with the approaches of Shephard and Pitt (1997) and Geweke and Tanizaki (2001). Our methodology provides the same efficiency as these other approaches, but applies more generally. For example, we also implement a square-root SV model with jumps using daily short-term interest rates in Hong Kong (HIBOR). The methods in Shephard and Pitt (1997) do not directly apply due to the state-dependent variance in the evolution equation, and the methods of Geweke and Tanizaki (2001) can lead to poor approximations due to the jumps in the model. Whereas in our methodology, allowing for the possibility of jumps requires

only one additional mixture component in the auxiliary model. The algorithm provides the smoothing distribution for three state variables, namely, the volatility of the interest rate and its jump sizes and times.

The rest of the paper is outlined as follows. Section 2 describes our methodology for estimation and smoothing of SDV models. We describe how to determine the auxiliary mixture model and its use in our MCMC algorithm. Sampling from the full model is a two-step procedure requiring simulation from a proposal distribution using conditionally Gaussian state-space methodology and then a Metropolis step to re-weight these samples. The proposal distribution is based on the auxiliary mixture model. Section 3 discusses implementation of the proposed methods in SDV models. Section 4 illustrates our approach with two applications. Finally, Section 5 concludes.

2 Nonlinear SDV Models

Nonlinear state-dependent variance models take the form

$$y_t = f_t(x_t) + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}[0, V(x_t)], \tag{1}$$

$$x_{t+1} = g_t(x_t) + \omega_t, \qquad \omega_t \sim \mathcal{N}[0, W(x_t)], \tag{2}$$

for times $t = 1, \ldots, T$. Here y_t is the observation vector, x_t is the unobserved state vector, and $\mathcal{N}(\mu, V)$ denotes a normal distribution with mean μ and variance V. The *observation equation* (1), and the *evolution equation* (2) have variance functions $V(x_t)$ and $W(x_t)$, respectively, that depend on the unobserved state x_t, and possibly on some unknown static parameter vector λ. The mean functions $f_t(x_t)$ and $g_t(x_t)$ are nonlinear in the states and may also involve unknown static parameters. The error sequences $\{\epsilon_t\}$ and $\{\omega_t\}$ are mutually and serially independent given the states x_t.

Denote the sampling distribution defined in (1) by $p(y_t|x_t)$ and use $x = (x_0, \ldots, x_T)$ and $y = (y_1, \ldots, y_T)$ to denote the entire vector of state parameters and data, respectively. The full posterior smoothing distribution for the nonlinear SDV model (1) and (2) is

$$p(x|y) \;\propto\; p(x_0) \prod_{t=1}^{T} p(x_t|x_{t-1})\, p(y_t|x_t), \tag{3}$$

where $p(y_t|x_t) = \mathcal{N}[f_t(x_t), V(x_t)]$, and $p(x_t|x_{t-1}) = \mathcal{N}[g_t(x_{t-1}), W(x_{t-1})]$. Here $p(x_0) = \mathcal{N}(m_0, C_0)$ describes the distribution on the initial state.

This joint posterior distribution is typically not available in closed form, even when the functions $f(x_t)$, $V(x_t)$, $g(x_t)$ and $W(x_t)$ are known. Moreover, standard MCMC simulation techniques are not applicable due to nonlinearities in the mean and the state dependence in the variance function. Next we will describe a new algorithm which addresses this issue.

2.1 An Auxiliary Mixture Model

The key idea is to approximate the observation and evolution equations by an auxiliary mixture model. The mixture model is denoted by $p^a(y_t|x_t, z_t)$, and is assumed to be conditionally linear and Gaussian given a state-dependent mixture indicator z_t. It is easier to discuss the cases in which nonlinearity and state dependent variance occur in either the observation or in the evolution equations. First, suppose that the state dependence only occurs in the observation equation,

$$\begin{aligned} y_t &= f(x_t) + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}[0, V(x_t)] \\ x_{t+1} &= Gx_t + \omega_t, \qquad \omega_t \sim \mathcal{N}[0, W], \end{aligned} \tag{4}$$

$t = 1, \ldots, T$, where $f(x_t)$ is nonlinear and $V(x_t)$ is state dependent.

We now describe how the auxiliary mixture model. Let $z = (z_1, \ldots, z_T)$ be a vector of mixture indicators, $z_t \in \{1, \ldots, K\}$. The auxiliary mixture model $p^a(y_t|x_t)$ is defined as

$$p^a(y_t|x_t) = \sum_{k=1}^{K} p^a(y_t|x_t, z_t = k)\ \pi_k(x_t). \tag{5}$$

where $\pi_k(x_t) = p^a(z_t = k|x_t)$ are state-dependent weights. The special case of fixed weights has been successfully used in recent literature, for example in Kim, Shephard and Chib (1998). But state-dependent weights provide more flexibility.

Conditional on the mixture indicator, z_t, we assume that the observation equation in the auxiliary model is a linear regression with constant variance, namely

$$p^a(y_t|x_t, z_t = k) = \mathcal{N}(\alpha_k + \beta_k x_t, \tau_k^2). \tag{6}$$

For the mixture weights, $p^a(z_t|x_t)$, we choose standardized Gaussian kernels

$$p^a(z_t = i|x_t) = \frac{\phi(x_t; \mu_i, \sigma_i)}{\sum_{k=1}^{K} \phi(x_t; \mu_k, \sigma_k)}, \tag{7}$$

where $\phi(x; m, s)$ denotes a normal density with mean m and variance s^2, and the weights are standardized to sum to 1 for all values of x_t. The choice of the kernels is important. Later, we exploit the Gaussian form of the kernel to define a dynamic linear model to generate proposals in a Metropolis-Hastings scheme.

The choice of $(\alpha_k, \beta_k, \tau_k, \mu_k, \sigma_k)$ is guided by the aim that $p^a(y_t|x_t)$ should provide a good approximation of $p(y_t|x_t)$. For example, $\{\mu_k\}$ could be an evenly spaced grid and (α_k, β_k) could correspond to linear expansions of $f(x_t)$ at μ_k. A linear Taylor series approximation allows automated construction of the auxiliary model. This becomes important when the mean function $f(x_t, \lambda)$ or the variance $V(x_t, \lambda)$ depends on a static parameter λ, and the

auxiliary mixture model has to adapt with λ. To achieve this, we allow the regression parameters $(\alpha_k, \beta_k, \tau_k)$ to depend on λ. For given knots μ_k we could define

$$\beta_k(\lambda) = \frac{\partial f(\mu_k, \lambda)}{\partial x_t}, \quad \alpha_k(\lambda) = f(\mu_k, \lambda) - \beta_k(\lambda)\,\mu_k \quad \text{and} \quad \tau_k^2(\lambda) = V(\mu_k, \lambda).$$

We use the auxiliary mixture model to define a proposal distribution in the Markov chain Monte Carlo (MCMC) posterior simulation for model (4). Details are described in Section 2.3.

2.2 Mixtures in the Evolution Equation

Our approach is equally applicable for models with nonlinearities or SDVs in the evolution equation. For example, assume a state-space model that is linear in the observation equation but with state-dependent variances in the evolution equation

$$\begin{aligned} y_t &= Fx_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}[0, V], \\ x_{t+1} &= g(x_t) + \omega_t, \qquad \omega_t \sim \mathcal{N}[0, W(x_t)]. \end{aligned} \tag{8}$$

To implement inference in (8) we define an auxiliary mixture model $p^a(x_t|x_{t-1})$ and latent indicators as in (5)–(7). Details of the posterior simulation are described in Section 2.3.

For problems where both $f(x_t)$ and $g(x_t)$ are nonlinear, or $V(x_t)$ and $W(x_t)$ are state-dependent two auxiliary mixture models. In Section 4.2, we consider an example with SDVs in both the observation and the evolution equation.

2.3 Simulating from the Smoothing Distribution

We describe posterior inference for model (4), with nonlinearity and SDV in the observation equation only. Inference for (8) and for the general nonlinear SDV model (1)–(2) proceeds analogously. The following three steps define one iteration in the simulated chain. Assuming currently imputed parameter values x, we first generate latent indicators z (step 1) for the auxiliary mixture model. Given the indicators we propose new values for x by recognizing $p^a(x|z, y)$ as essentially a Gaussian linear state space model (step 2). Finally, step 3 accepts the candidate $\tilde{x}$ generated in step 2 with appropriate Metropolis-Hastings acceptance probability, defined to maintain the smoothing distribution $p(x|y)$ as the stationary distribution of the Markov chain. The mixture indicators z are generated in step 1, and dropped from the state vector again at the end of step 3.

1. Generating mixture indicators $z = (z_1, \ldots, z_T)$. We augment the state vector x with a vector of mixture indicators z_t, using $p^a(z|x, y)$. The full conditional posterior

distribution of the mixture indicators under the auxiliary model is

$$p^a(z|x,y) \;\propto\; \prod_{t=1}^{T} p^a(y_t|z_t,x_t)\; p^a(z_t|x_t).$$

Given the state vector x, the indicator variables $z_1,\ldots,z_T$ are conditionally independent, and can be sampled independently from multinomial distributions with probabilities $p^a(z_t|x_t,y_t) \propto p^a(y_t|z_t,x_t)\, p^a(z_t|x_t)$.

2. Generating a proposal $\tilde{x}$. Consider the full conditional distribution $p^a(x|z,y)$ under the auxiliary mixture model

$$p^a(x|z,y) \;\propto\; p(x_0) \prod_{t=1}^{T} p(x_t|x_{t-1})\; p^a(y_t|z_t,x_t)\; p^a(z_t|x_t).$$

To devise an efficient proposal distribution for the state variables, we factor this distribution into two parts. The first part will include all terms that are linear in the states. This will be used as the proposal distribution in a Metropolis-Hastings simulation step (Tierney, 1994). The second part will be used in the acceptance probability in step 3. Let $c(x_t) = \sum_k \phi(x_t;\mu_k,\sigma_k)$ denote the denominator in (7). Substituting (7) for $p^a(z_t|x_t)$ we get

$$p^a(x|z,y) \;\propto\; \underbrace{p(x_0) \prod_{t=1}^{T} p(x_t|x_{t-1})\; p^a(y_t|z_t,x_t)\; \phi(x_t;\mu_{z_t},\sigma_{z_t})}_{q(x|z,y)}\; \frac{1}{c(x_t)}. \tag{9}$$

The first factor, denoted by $q(x|z,y)$, corresponds to the smoothing distribution in another linear, Gaussian state space model. The probability model $q(x|z,y)$ will serve as a proposal distribution in a Metropolis-Hastings step. Details are explained below. The importance of augmenting our model by a mixture of linear regressions with Gaussian kernel weights now becomes clear. We can use the efficient block sampling algorithms of Carter and Kohn (1994) and Frühwirth-Schnatter (1994) to generate candidate values $\tilde{x}$ of the state vector, $\tilde{x} \sim q(x|z,y)$. The algorithm is known as *forward-filtering, backward-sampling* (FFBS).

The following state-space model has a smoothing distribution given by $q(x|z,y)$:

$$\begin{aligned} \tilde{y}_t &= x_t + \epsilon_t, & \epsilon_t &\sim \mathcal{N}[0,\tilde{V}_t] \\ x_{t+1} &= Gx_t + \omega_t, & \omega_t &\sim \mathcal{N}[0,W], \end{aligned} \tag{10}$$

where $\tilde{y}_t = \tilde{V}_t\,[(y_t-\alpha_{z_t})\beta_{z_t}/\tau^2_{z_t} + \mu_{z_t}/\sigma^2_{z_t}]$ and $\tilde{V}_t = (\beta^2_{z_t}/\tau^2_{z_t} + 1/\sigma^2_{z_t})^{-1}$. The value $\tilde{y}$ is chosen so that the likelihood in (10) equals the last two factors in the definition of

$q(x|z,y)$; i.e., the observation equation in (10) is formed by combining the linearized observation equation (6) with the Gaussian weighting kernel (7).

Since (10) is a linear state-space model, the smoothing distribution $q(x|z,y)$ can be sampled directly using FFBS. We use it to generate a proposal for the state variables

$$\tilde{x} \sim q(x|z,y).$$

3. Metropolis-Hastings rejection step. Evaluate the acceptance probability

$$a(x,\tilde{x}) = \min\left\{1, \prod_{t=1}^{T} \frac{p(y_t|\tilde{x}_t)}{c(\tilde{x}_t)\; p^a(y_t|\tilde{x}_t)} \frac{c(x_t)\; p^a(y_t|x_t)}{p(y_t|x_t)}\right\}. \tag{11}$$

Here $p^a(y_t|x_t)$ is the approximation to the likelihood $p(y_t|x_t)$ implied in the auxiliary mixture model, i.e., $c(x_t)\; p^a(y_t|x_t) = \sum_{k=1}^{K} \phi(x_t;\mu_k,\sigma_k)\, p^a(y_t|x_t, z_t = k)$. With probability $a(x,\tilde{x})$ replace the currently imputed state parameters x by $\tilde{x}$. Otherwise discard the proposal $\tilde{x}$ and leave x unchanged.

Use of the acceptance probability $a(x,\tilde{x})$ ensures an ergodic distribution equal to $p(x|y)$, as desired. This is seen by considering an augmentation of $p(x|y)$ to $p(x,z|y) \equiv p(x|y) \cdot p^a(z|x,y)$, i.e., add z to the probability model $p(x|y)$ by defining the conditional distribution for z given x and y as in model p^a. Steps 1 through 3 define a Markov chain with ergodic distribution $p(x,z|y)$. Step 1 replaces z by sampling from the complete conditional distribution $p(z|x,y) = p^a(z|x,y)$. Step 2 generates a Metropolis-Hastings proposal $\tilde{x} \sim q(\tilde{x}|z,y)$. Step 3 accepts the proposal with the correct Metropolis-Hastings acceptance probability $\min\{1, p(\tilde{x}|z,y)\, q(x|z,y)/[p(x|z,y)\, q(\tilde{x}|z,y)]\}$. To verify the acceptance probability (11) note that

$$\begin{aligned} p(x|z,y) &\propto p(x|y)\, p(z|x,y) = p(x|y)\, p^a(z|x,y) \\ &\propto p(x_0) \prod_{t=1}^{T} p(x_t|x_{t-1})\; p(y_t|x_t)\; \frac{p^a(y_t|z_t,x_t)\, \phi(x_t;\mu_{z_t},\sigma_{z_t})}{\sum_{k=1}^{K} p^a(y_t|z_t=k,x_t)\, \phi(x_t;\mu_k,\sigma_k)}. \end{aligned}$$

If the model includes static parameters λ, an additional step is used to update λ, given the currently imputed states x_t. This is typically carried out using a Metropolis-Hastings step. Details depend on the specific model.

2.4 Extensions

An important part of the proposed algorithm is the choice of the knots $\{\mu_k\}$. Although in principle arbitrary, a good choice is important for a computationally efficient implementation. Typically, the state process x_t is stationary, and we can choose the knot locations and

kernels based on the marginal prior $p(x_t)$. This is true, for example, for the SV model discussed in Section 4.1. However, when the support of the state vector changes significantly over time, the set of knots μ_k needs to be adaptive.

To fix ideas, consider model (4) with non-linearity and SDV in the observation equation only. Assume that some approximation of marginal posterior means and standard deviations for the states x_t is available. Such approximations could be obtained, for example, using methods proposed in Shephard and Pitt (1997). Alternatively, any ad-hoc estimation of the unknown states x_t could be used. Based on these approximate moments we now define an adaptive grid of knots $\{\mu_{tk}\}$, for example, as a grid over mean plus/minus several posterior standard deviations. We complete the construction of an auxiliary mixture model for each period t by defining linear approximations with parameters $\beta_{tk} = f'(\mu_{tk})$, $\alpha_{tk} = f(\mu_{tk}) - \mu_k \beta_{tk}$ and $\tau^2_{tk} = V(\mu_{tk})$. The auxiliary mixture model (5) with $p^a(y_t|z_t = k, x_t) = \mathcal{N}(\alpha_{tk} + \beta_{tk}x_t, \tau^2_{tk})$ is then used to proceed as in Section 2.3.

In many applications, it is necessary to do the updating of the state vector x in subblocks to achieve reasonable acceptance probabilities in (11). Let $x = (x^{(1)}, \ldots, x^{(J)})$ denote a partition of the states into subvectors. We repeat steps 1 through 3 of the proposed algorithm J times, proposing at each iteration new values $\tilde{x}^{(j)}$ for one subvector only. The choice of the block size is a trade-off between attaining a reasonable Metropolis acceptance probability and the computational efficiency obtained by block updating. As a rule of thumb, the average acceptance probability should be between 5% and 95%. Similar blocking strategies are proposed in Shephard and Pitt (1997).

Finally, for a high dimensional state vector x_t we recommend a partitioning strategy (Cargnoni, Müller and West, 1997). Partition the state vector into $x_t = (x_{t1}, x_{t2})$. At each iteration of the MCMC we first update $x_1 = (x_{t1},\, t = 1 \ldots T)$ conditional on $x_2 = (x_{t2},\, t = 1 \ldots T)$ proceeding as before. Then we follow up with updating x_2 conditioning on the currently imputed values x_1. In practice, partitioning might be advantageous for state vectors beyond three dimensions.

3 Stochastic Volatility Modeling and Jumps

3.1 A Standard SV Model

SV models are commonly used to describe the evolution of equity returns (Ghysels, Harvey and Renault, 1996). A standard approach is to assume that equity prices $S(t)$ follow a geometric Brownian motion with volatility $v(t)$ that is modeled as a mean-reverting process:

$$\begin{aligned} d\log S(t) &= (\mu - v(t)/2)\, dt \;+\; \sqrt{v(t)}\, dW(t) \\ d\log v(t) &= \kappa_v(\theta_v - \log v(t))\, dt \;+\; \sigma_v\, dW_1(t). \end{aligned}$$

Here $W(t)$ and $W_1(t)$ are independent Brownian motions, κ_v governs the speed of mean reversion, θ_v is the long-run mean of log-volatility, and σ_v is the volatility of volatility. Discretizing on a unit time step, and setting $y_t = \log(S_{t+1}/S_t)$ and $x_t = \log v_t$, we obtain a nonlinear SDV model:

$$\begin{aligned} y_t &= (\mu - e^{x_t}/2) + e^{x_t/2}\,\epsilon_t \\ x_{t+1} &= x_t + \kappa_v(\theta_v - x_t) + \sigma_v\,\omega_t. \end{aligned} \tag{12}$$

Here we have nonlinearity in the mean and state-dependence in the variance of the observation equation, whereas the evolution equation for log-volatility is a standard normal linear regression model with constant variance. We now show how to form the auxiliary mixture model for the observation equation.

The choice of K, the number of mixture components, is a trade-off between a fast algorithm for the auxiliary mixture model with a slower algorithm but higher acceptance probability in the Metropolis step. Clearly, the choice of K is problem specific and is related to the degree of departure from linearity of the mean and the nature of the state dependence in the variance function. For example, when the mean and variance functions are exponentials as in (12), we can achieve a good approximation using a small number of mixture components.

Figure 1 shows an example with mean and variance function given by $f(x_t) = e^{x_t}$ and $V(x_t) = e^{x_t}$. To approximate this model, we choose a mixture of size $K = 3$, with $\mu_k \in \{-1, 0, 1\}$ and $\sigma_k \equiv 1$. Figures 1(a) and 1(b) show the mean and variance functions, $f(x_t)$ and $V(x_t)$, respectively, along with the approximating linear regression lines and variance levels. The regression lines are defined by a linear Taylor expansion around μ_k

$$f_k^a(x_t) = \alpha_k + \beta_k x_t, \quad \text{where} \quad \beta_k = f'(\mu_k) \quad \text{and} \quad \alpha_k = f(\mu_k) - \mu_k\beta_k.$$

The stepwise variance levels are defined by $\tau_k^2 = V(\mu_k)$. The bottom panels in Figure 1 show the Gaussian weighting kernels $\phi(x_t; \mu_k, \sigma_k)$, along with the knots, μ_k, denoted by triangles.

Hence, at each time period, we have a locally-weighted mixture of linear regressions with constant variance. The mixture weights are the Gaussian kernels at the bottom of Figure 1, and the linear regression models are given by the approximating mean and variance functions in the top panel of Figure 1. We can now simulate from the smoothing distribution as described in Section 2.3.

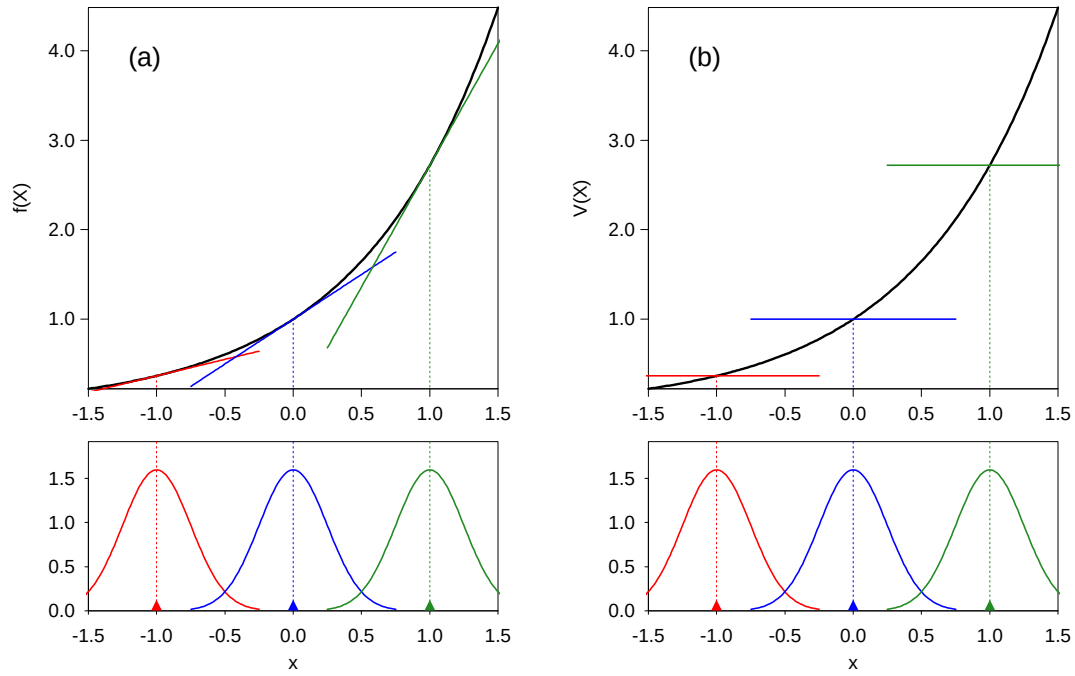

Figure 1: Construction of the auxiliary mixture model. Top panels: (a) Mean function $f(x_t) = e^{x_t}$ and approximating linear regression lines, $\alpha_k + \beta_k x_t$. (b) Variance function $V(x_t) = e^{x_t}$ and approximating variance levels τ_k^2. Bottom panels: Gaussian weight kernels $\phi(x_t; \mu_k, \sigma_k)$ and knots μ_k, denoted by triangles.

3.2 A Two-Factor SV Model with Jumps

A commonly used model for short-term interest rates is the Cox-Ingersoll-Ross (CIR) model (Cox, Ingersoll and Ross, 1985). It is a two-factor SV model where the future interest rate depends on the current short rate $r(t)$ and its volatility $v(t)$. We consider the following extension of the CIR model allowing for jumps:

$$\begin{aligned} d\log r(t) &= \kappa_r(\theta_r - \log r(t))dt + \sqrt{v(t)r(t)^{-1}}\, dW(t) + \xi(t)\, dJ(t) \\ dv(t) &= \kappa_v(\theta_v - v(t))dt + \sigma_v \sqrt{v(t)}\, dW_1(t). \end{aligned} \tag{13}$$

Here $r(t)$ is the short-term interest rate, $v(t)$ is its volatility and $W(t)$ and $W_1(t)$ are uncorrelated Brownian motions. An exogenous jump shock is incorporated in the term $\xi(t)\, dJ(t)$ where $J(t)$ is a counting process for the jumps, and $\xi(t)$ denotes the corresponding jump size. The first equation defines the sampling model ("observation equation") for the observable data $r(t)$ conditional on unobservable dynamic state parameters $v(t)$ and additional static parameters κ_r and θ_r. The second equation defines the evolution of the state variable $v(t)$.

To implement inference in this model, we use an Euler discretization of the continuous-time model. As Pritsker (1998) shows, for daily interest rate data the corresponding discretization error is insignificant compared to the estimation error. See more discussion in Elerian, Shephard and Chib (2001) and Eraker (2001).

Discretizing equation (13) gives an SDV model of the form

$$\begin{aligned} y_{t+1} &= y_t + \kappa_r(\theta_r - y_t) + \sqrt{x_t e^{-y_t}}\,\epsilon_t + \xi_t J_t, \\ x_{t+1} &= x_t + \kappa_v(\theta_v - x_t) + \sigma_v\sqrt{x_t}\,\omega_t, \end{aligned} \tag{14}$$

where $y_t = \log r_t$, $x_t = v_t$, and ϵ_t and ω_t are i.i.d. standard normal errors. The observational error term comprises two components, the stochastic volatility term and the jump term. We assume that jumps are i.i.d., $J_t \sim Ber(\lambda)$, and that the jump sizes are normal $\xi_t \sim \mathcal{N}(\mu_\xi, \sigma_\xi^2)$. In addition to the observed data y_t and the state parameters x_t the model includes static parameters $(\kappa_r, \theta_r, \kappa_v, \theta_v, \sigma_v, \lambda, \mu_\xi, \sigma_\xi)$, and the jumps J_t and jump sizes ξ_t. Implementing the algorithm outlined in Section 2.3 we require additional steps to update the static parameters and the jumps $\xi_t\, J_t$. All static parameters and $\xi_t\, J_t$ can be updated by generating from the appropriate complete conditional posterior distributions, each of which takes the form of a well-known distribution. Model (14) does not explicitly constrain x_t to positive values. In the application reported later in Section 4.2, the posterior is centered safely away from negative volatilities. With different data, it might be necessary to introduce a constraint in the evolution equation.

If we marginalize with respect to the jump term in the observation equation, model (14) becomes a SDV model with non-normal errors. The marginal error distribution, integrating over the jump process, is given by a discrete mixture of normals

$$(1-\lambda)\mathcal{N}[0, x_t e^{-y_t}] \;+\; \lambda\mathcal{N}[\mu_\xi, x_t e^{-y_t} + \sigma_\xi^2].$$

The addition of the jump component allows the model to capture "outlying" behavior in the series that cannot be explained by a change in the stochastic volatility state variable x_t. A further extension of these models is to also include jumps in the volatility equation, see Eraker, Johannes and Polson (2001).

4 Examples

First we discuss a simulation study. In a simple setting of a logarithmic SV model we compare our approach with those of Shephard and Pitt (1997) and Geweke and Tanizaki (2001). Second, we model the movements in the daily short-term interest rate in Hong Kong from 1986-2000 using a SV model with jumps. The short-term interest rate in Hong Kong is known as the HIBOR (Hong Kong Dollar Interbank Offered Rates). Allowing for

jumps or discontinuities is critical for modeling periods such as the Asian financial crisis. In Nov. 1997 and again in Jan. 1998 the short interest rate jumped from its average level of 6% to over 50%. Such rapid movements cannot be explained by a standard stochastic volatility model. Alternative approaches are proposed in recent work by Barndorff-Nielsen and Shephard (2001).

4.1 A Logarithmic SV Model

To study the efficiency of our algorithm we simulated $T = 1500$ observations from a standard logarithmic stochastic volatility model:

$$\begin{aligned} y_t &= \mu + \epsilon_t, & \epsilon_t &\sim \mathcal{N}[0, e^{x_t}], \\ x_{t+1} &= \alpha + \beta x_t + \omega_t, & \omega_t &\sim \mathcal{N}[0, W], \end{aligned}$$

with $\mu = 0$, $\alpha = -.085$, $\beta = .99$, $W = .04$, and $x_0 = -8.5$. These are typical parameter values for daily equity returns (see Jacquier, Polson and Rossi, 1994). Figure 2(a) shows simulated observations y_t under these parameter values.

For the analysis reported below, we fixed the parameters μ, α, β, and W at their given values, and assumed a diffuse prior on the initial state: $x_0 \sim \mathcal{N}(-8.5, 100)$. We then fitted the model using three different MCMC algorithms: the locally-weighted mixture method (LWM) introduced in Section 2; the Laplace block updating algorithm of Shephard and Pitt (1997) (SP); and the single-state updating method of Geweke and Tanizaki (2001) (GT). For each algorithm, we simulated 5000 MCMC samples and collected the last $M = 4000$ for inference.

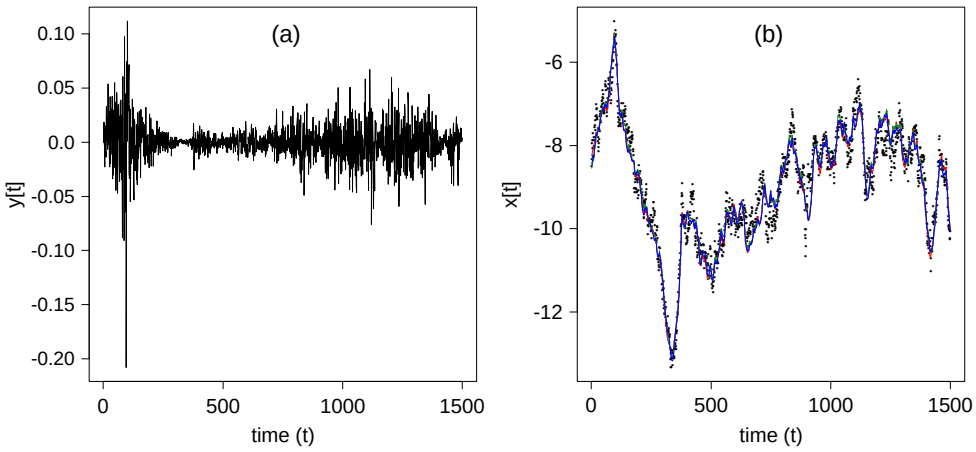

Figure 2: SV model. (a) Simulated observations y_t. (b) Simulated states x_t, and smoothed means using three MCMC algorithms: LWM, SP, and GT. The three lines are practically indistinguishable, indicating that the algorithms have converged to the same smoothed mean estimates.

To ensure reasonable acceptance probabilities for the LWM algorithm, we used 10 blocks to update the state vector. For the LWM method, we used $K = 7$ equally-space knots between $\mu_1 = -3$ and $\mu_7 = -14$, and a kernel width of $\sigma_k = 1$ for $k = 1, \ldots, 7$. Figure 2(b) shows the simulated state variables, x_t, along with the smoothed means for each of the three algorithms. The three lines are nearly indistinguishable, indicating that all three algorithms have converged to the same smoothed mean estimates.

Next, we examined graphical summaries of convergence. Figure 3 shows trace plots and autocorrelation plots of x_{500} for the three algorithms. For LWM, the autocorrelation function decays rapidly to zero, indicating a fast-mixing Markov chain.

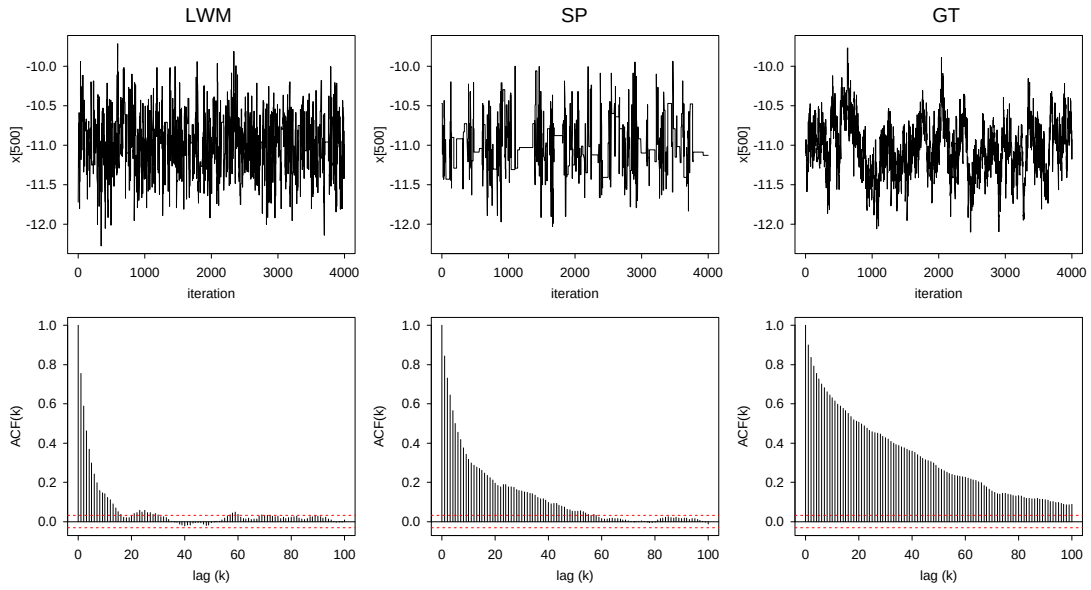

Figure 3: SV model. Trace plots and autocorrelation plots for x_t, $t = 500$ for three MCMC algorithms: LWM (left); SP (center); and GT (right). The average acceptance probabilities for the two blocking schemes were about 30% for LWM and 10% for SP. Taking into account that the implementation of SP and GT could likely be further fine tuned, we consider the performance of the three algorithms approximately comparable.

We also computed summary diagnostics proposed by Raftery and Lewis (1992). Table 1 shows dependence factors for four selected state variables, x_t, $t \in \{1, 500, 1000, 1500\}$, obtained using the software BOA (Smith, 2001). Overall, the dependence factors for LWM are smaller than for SP and GT, indicating faster convergence of the LWM algorithm. (A smaller dependence factor indicates a faster mixing Markov chain.)

On the basis of these results and taking into account that SP could likely be further fine-tuned for this application, we conclude that the performance of LWM and SP is roughly comparable, and both are more efficient than GT. The conclusion is not surprising. GT

state	Mixture (LWM)			Laplace (SP)			Single-State (GT)		
variable	5%	50%	95%	5%	50%	95%	5%	50%	95%
x_1	15	17	13	55	46	13	52	97	19
x_{500}	7	13	10	72	54	19	13	50	18
x_{1000}	21	24	34	131	46	48	20	54	10
x_{1500}	16	24	14	94	45	64	15	34	22

Table 1: SV model. Raftery and Lewis (1992) convergence diagnostics for three algorithms: LWM, SP, and GT. Convergence was monitored for the 5th, 50th, 95th percentiles for four selected state variables, x_t, $t \in \{1, 500, 1000, 1500\}$.

is very general but does not exploit any of the special time series structure in the model as does LWM. The important advantage of LWM over SP is the general structure of the evolution equation and the possibility to include state-dependent variances. This aspect is critical in finance applications where models often include nonlinearities and SDVs.

4.2 Hong Kong Interest Rates (HIBOR)

Hong Kong short interest rates provide a good illustration of the need to model jumps with the stochastic volatility model. Figure 4 shows the daily HIBOR rate for the period 1986 to 2000. The period of dramatic jumps started in Nov. 1997, when the short rate moved from a level of 6% on Nov. 17 to 50% on Nov. 23, and continued on into the beginning of 1998. One possible reason for jumps in this series is that the Hong Kong dollar operates under a currency board and the currency is pegged to the U.S. Dollar. Therefore, in periods of market stress, when there is an increased probability of a currency devaluation, the short rate has to be raised to very high levels.

We model the interest rate in two subsamples: first from Jan. 1986 to Dec. 1993 and then from Jan. 1994 to Jan. 2000. We fit a square-root stochastic volatility model (14) with jumps in the level of interest rates. For comparison we also analyze a two-factor stochastic volatility CIR model without jumps. Figure 4 plots the data in these periods. Clearly, jumps are present in the later subsample. All inference reported below is based on the second subsample for 1994–2000. We implemented inference using the proposed LWM algorithm described in Section 2. For both models, we chose 10 equally-spaced knots for the observation and evolution equations based on the marginal prior distribution $p(x_t)$.

The first row of Table 2 shows the posterior means and standard deviations of the model parameters for the no-jump model. These results are based on $M = 4000$ MCMC samples after a burn-in period of 1000. The model is forced to estimate a high level of σ_v. The residual plots (not shown) are highly non-normal indicating a very poor description of the underlying series. The second row in Table 2 shows the posterior moments under the

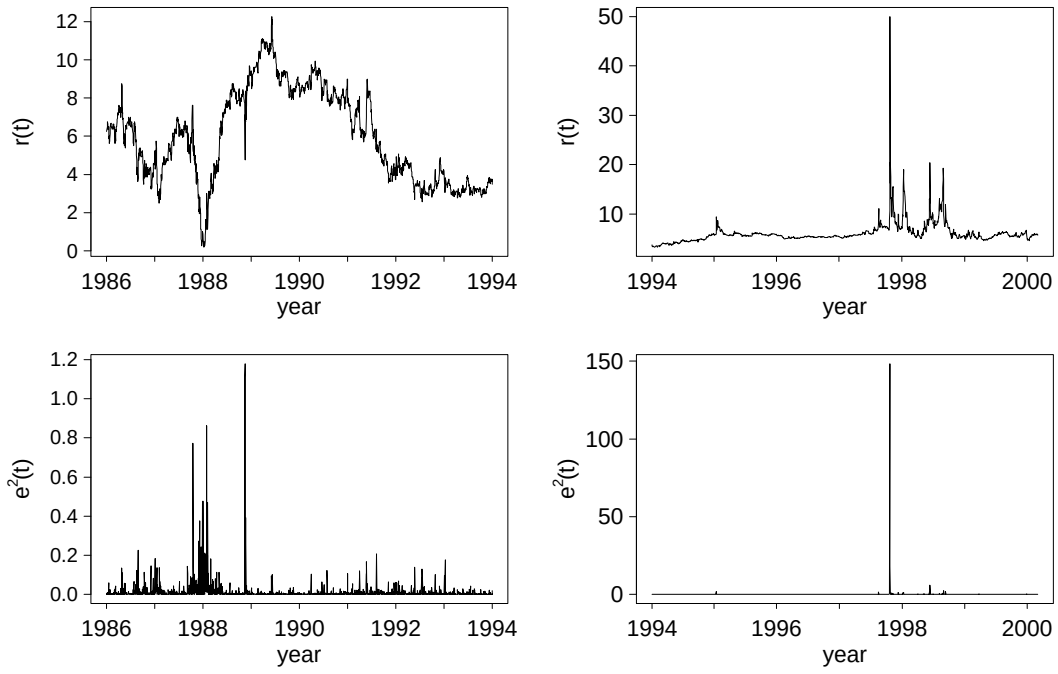

Figure 4: HIBOR data, 1986–2000. Top panels: Short term interest rates, using different scales before and after 1994. Bottom panels: Squared residuals (proxies for volatility) from a linear regression model. The plot illustrates the need to allow for jumps in the model.

model with jumps. The major difference is that, after allowing for jumps, the estimate of σ_v decreases, whereas our estimate of the mean reversion parameter κ_v increases, implying more persistence in the volatility sequence. This leads to very different pricing implications under the two models. Our methodology also provides inference for the jump component of the model. Figure 5 shows the posterior distributions of the jump probability λ and mean μ_ξ and standard deviation σ_ξ of the jump sizes. The prior specification represents our initial beliefs that the jump component of the model is infrequent and captures large jump sizes relative to the underlying stochastic volatility term. We choose a prior for the jump frequency, $\lambda \sim Beta(1, 75)$, giving a mean of 1.3% per year. To complete the

Model	$\kappa_r\theta_r$	$1-\kappa_r$	$\kappa_v\theta_v$	$1-\kappa_v$	σ_v^2
No Jumps	.0200	.990	.0010	.900	.0250
	(.09)	(.03)	(.008)	(.103)	(.005)
Jumps	.0218	.987	.0001	.935	.0002
	(.007)	(.004)	(.00001)	(.011)	(.00003)

Table 2: HIBOR data. Posterior means (and standard deviations) of model parameters.

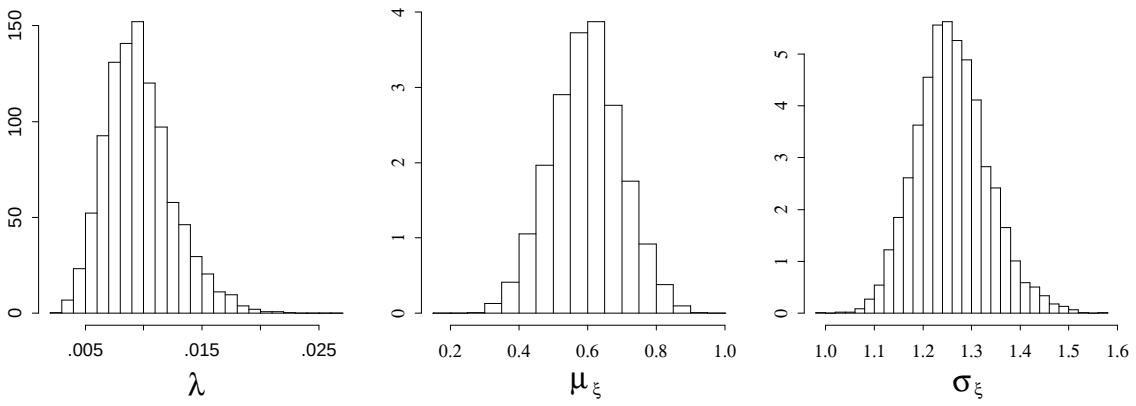

Figure 5: HIBOR data. Posterior distribution of jump process parameters: jump probability λ, and mean μ_ξ and standard deviation σ_ξ of the jump sizes.

analysis, Figure 6 shows the smoothed states: the posterior probability of jump occurrences $p(J_t = 1|y)$ and the volatility states, $E(v_t|y)$. The observed log-rate series y_t is also provided for comparison.

5 Conclusions

This paper develops a general likelihood-based approach for inference in nonlinear dynamic models with state-dependent variances. This class of models is very flexible, including both stochastic volatility (SV) and affine term structure models. To carry out parameter inference and smoothing, we provide a novel block sampling MCMC algorithm. A key step in the the algorithm is the specification of an auxiliary mixture model with state-dependent weights. This allows the algorithm to exploit the conditionally Gaussian structure embedded in the mixture model. This leads to an efficient MCMC simulation scheme for this general class of models.

One advantage of our methodology is that it applies to nonlinearities and SDVs in both the observation and evolution equations. It also is straightforward to incorporate jumps and mixture of normals in the innovation and error terms. On the other hand, the researcher has to construct an auxiliary mixture for the problem at hand. This requires some a priori knowledge of the support of the state vector. One direction for future research is to provide methods for filtering. As in Pitt and Shephard (1999) our block sampling approach lends itself naturally to provide an algorithm for tackling the computationally demanding problem of filtering in these models.

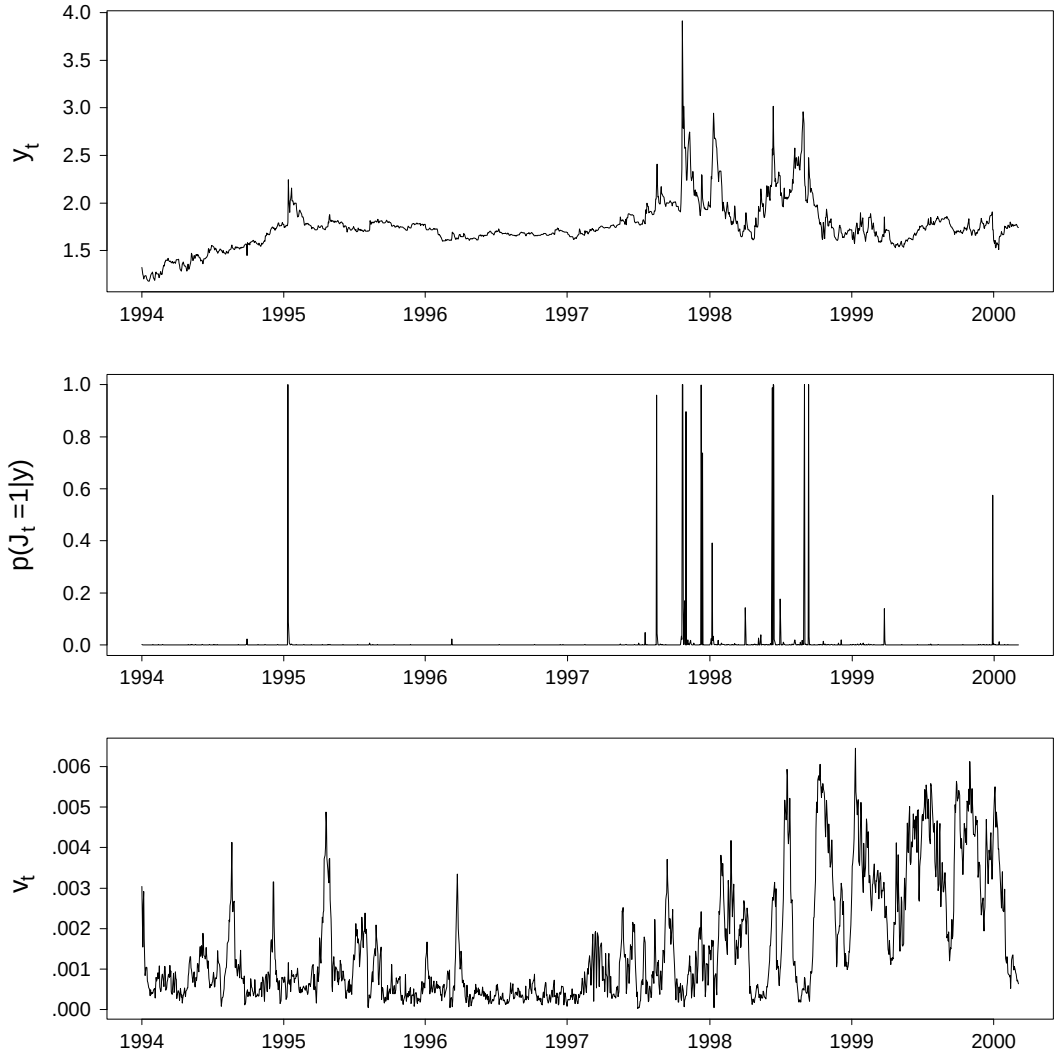

Figure 6: HIBOR data. (a) Log HIBOR rates, y_t, (1994-2000). (b) Posterior probability of jump occurrences $p(J_t = 1|y)$. (c) Smoothed volatility estimates $E(v_t|y)$.

References

Anderson, B. and Moore, J. (1979) *Optimal Filtering*. Englewood Cliffs, New Jersey: Prentice-Hall.

Barndorff-Nielsen, O. and Shephard, N. (2001) Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial econometrics (with discussion). *Journal of the Royal Statistical Society, Series B*, **63**, 167–241.

Cargnoni, C., Müller, P. and West, M. (1997) Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association*, **92**, 640–647.

Carlin, B., Polson, N. and Stoffer, D. (1992) A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, **87**, 493–500.

Carter, C. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika*, **81**, 541–553.

— (1996) Markov chain Monte Carlo in conditionally Gaussian state-space models. *Biometrika*, **83**, 589–601.

Cox, J., Ingersoll, J. and Ross, S. (1985) An intertemporal general equilibrium model of asset prices. *Econometrica*, **53**, 363–384.

Duffie, D. and Kan, R. (1996) A yield-factor model of interest rates. *Mathematical Finance*, **6**, 379–406.

Elerian, O., Shephard, N. and Chib, S. (2001) Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, **69**, 959–993.

Eraker, B. (2001) MCMC analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics*, **19**, 177–191.

Eraker, B., Johannes, M. and Polson, N. (2001) Jumps to returns and volatility in continuous time: Time series evidence. *Tech. rep.*, Graduate School of Business, University of Chicago.

Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**, 183–202.

Geweke, J. and Tanizaki, H. (2001) Bayesian estimation of state-space model using the Metropolis-Hastings algrorithm within Gibbs sampling. *Computational Statistics and Data Analysis*, **37**, 151–170.

Ghysels, E., Harvey, A. and Renault, E. (1996) Stochastic volatility. In *Statistical Methods in Finance* (eds. G. Maddala, C. Rao and H. Vinod), 119–91. North Holland.

Jacquier, E., Polson, N. and Rossi, P. (1994) Bayesian analysis of stochastic volatility (with discussion). *Journal of Business and Economic Statistics*, **12**, 371–417.

Kim, S., Shephard, N. and Chib, S. (1998) Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, **65**, 361–393.

Kitagawa, G. (1996) Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Jounal of Computation and Graphical Statistics*, **5**, 1–25.

Pitt, M. and Shephard, N. (1999) Filtering via simulation: Auxiliary particle filter. *Journal of the American Statistical Association*, **94**, 590–599.

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

Raftery, A. and Lewis, S. (1992) How many iterations in the Gibbs sampler? In *Bayesian Statistics 4* (eds. J. Bernardo, J. Berger, A. Dawid and A. Smith), 763–774. Oxford: Oxford University Press.

Shephard, N. (1994) Partial non-Gaussian state space. *Biometrika*, **81**, 115–131.

Shephard, N. and Pitt, M. (1997) Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, **84**, 653–667.

Smith, B. (2001) *Bayesian Output Analysis Program (Version 1.0.0).* Available on-line at http://www.public-health.uiowa.edu/BOA.

Taylor, S. (1986) *Modelling Financial Time Series.* New York: John Wiley.

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

West, M. and Harrison, P. (1997) *Bayesian Forecasting and Dynamic Models.* New York: Springer, 2nd edn.

Zehnwirth, B. (1988) A generalization of the Kalman filter for models with state-dependent observation variance. *Journal of the American Statistical Association*, **83**, 164–167.