

# Nonlinear State-Space Models with State-Dependent Variances

JONATHAN R. STROUD <sup>1</sup>

*Department of Statistics, University of Chicago*

PETER MÜLLER

*Institute of Statistics and Decision Sciences, Duke University*

NICHOLAS G. POLSON

*Graduate School of Business, University of Chicago*

In this paper we provide a methodology for state smoothing in nonlinear state space models with state dependent variance (SDV). This general class of models contains both stochastic volatility (SVOL) and affine term structure models (ATSMs) which are commonly used in financial time series. For our smoothing technique, we use simulation-based methods with an auxiliary mixture model. We illustrate our methodology with three time-series applications. First, we show how to construct the auxiliary model for a logarithmic SVOL model. Second, we implement a stochastic volatility model with jumps for short-term interest rates in Hong Kong. We find strong evidence for jumps and stochastic volatility in the data and we find the smoothing distribution for the jump times, sizes and volatilities. Third, we implement a two-factor affine term structure model for daily U.S. bond yields from 1996-1999. Our methodology uncovers the unobserved state vector and provides sharper estimates of the parameters of the state dynamics than a simple SVOL interest-rate model. Finally, we conclude with directions for future research.

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

---

<sup>1</sup>We wish to thank Anne Gron and Michael Johannes for many useful discussions. Jonathan Stroud is Postdoctoral Research Associate at the Department of Statistics, University of Chicago. Peter Müller is Associate Professor at the Institute for Statistics and Decision Sciences, Duke University. Nick Polson is Professor at the Graduate School of Business, University of Chicago. *Corresponding author:* Jonathan Stroud, 5734 S. University Ave., Chicago, IL 60637, U.S.A.

# 1 Introduction

In this paper we provide a methodology for state smoothing and estimation for nonlinear and state-dependent variance (SDV) models. Nonlinearity and state-dependent variances are prevalent in many fields of application, particularly financial time series. This general class of models contains both stochastic volatility (SVOL) and affine term structure models (ATSMs) which are commonly used in time series. However, current estimation methodologies do not account for these features. Our methodology provides smoothed state and parameter distributions in an efficient likelihood-based framework. Other methodologies are based on less efficient procedures such as simulated method of moments and do not provide the full distribution of the states.

Simulation-based smoothing has received little attention in SDV models. Our methodology draws on the literature on conditionally Gaussian models. Carlin, Polson and Stoffer (1992) show how to implement nonlinear state-space models with non-normal error distributions. Zehnwirth (1988) considers state dependence in the observation variance. Kitagawa (1996) and Hürzeler and Künsch (1998) consider general state space models. The latter are based on sequential Monte Carlo methods, whereas our approach has the benefit of block updating of states in an efficient MCMC algorithm; see the updating schemes of Carter and Kohn (1994), Frühwirth-Schnatter (1994) and deJong and Shephard (1995).

The purpose of this paper is to provide a filtering and smoothing algorithm for the states of nonlinear SDV models. Our methodology is simulation based and requires the use of an auxiliary mixture model. To do this, we propose a two-step procedure. First we determine an auxiliary discrete mixture model to approximate the given model. The weights in the mixture model are allowed to be adaptive and depend on the state vector. Conditional on the mixture component we assume that the auxiliary model reduces to a linear Gaussian state-space model (West and Harrison, 1997) and so implementation is based on using a block sampling Metropolis algorithm.

We implement two SDV models: a stochastic volatility model with jumps and an affine term structure model. SVOL models are used to describe the evolution of asset returns and a common approach is to assume that volatility follows an log-autoregressive or square-root process, as in Taylor (1986) and Jacquier, Polson and Rossi (1994). ATSMs are used to describe the evolution of interest rate time series and financial theory suggests that interest rates and hence the bond yields should be affine in the state variables. An ATSM assumes that the evolution of these states follows a SDV model. Typical choices for state variables are the short-term interest rate and its volatility. While these models are currently used in derivative pricing, estimation has remained problematic due to the nonlinearities inherent in the SDV model. Recent literature (Dai and Singleton, 2000; Andersen et al., 1998) uses a computationally intensive simulated method of moments approach. Our approach is an

alternative, less computationally intensive method based on full posterior simulation.

The rest of the paper is outlined as follows. Section 2 describes our methodology for estimation and smoothing of SDV models. It describes how to determine the auxiliary mixture model and its use in our MCMC algorithm. Sampling from the full model is a two-step procedure first 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. Adaptive auxiliary models can be used to allow for additional static parameters in the mean or variance functions. Section 3 discusses implementation of the proposed methods in two important classes of SDV models. Section 4 illustrates our approach with two applications. First, we model the short-term interest rate in Hong Kong using a square-root SVOL model with jumps. Allowing the possibility of jumps only requires one additional mixture component in the auxiliary model. Second, we implement a two-factor ATSM for U.S. bond yields from 1996–1999 with states given by the unobserved instantaneous short-term interest rate and the short-rate volatility. The data consist of a panel of bond yields across different maturities and the observation equation is linear in the state variables. Finally, Section 5 concludes with directions for future research.

## 2 Nonlinear SDV Models

Nonlinear state-dependent variance models take the form

$$y_t = f_t(x_t) + \epsilon_t, \quad \epsilon_t \sim N[0, V(x_t)], \quad (1)$$

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

for times  $t = 1, \dots, T$ . Here  $y_t$  is the observation vector,  $x_t$  is the unobserved state vector, and  $N(\mu, V)$  denotes a normal distribution. The *observation equation* (1), and the *evolution equation* (2) have variance functions  $V(x_t)$  and  $W(x_t)$  that depend on the unobserved state  $x_t$ , and possibly some unknown static parameters  $\lambda$ . 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 assumed to be serially and mutually independent.

To begin, we set some notation. Let the distribution of observation  $y_t$  given state  $x_t$  under model (1) be denoted by  $p(y_t|x_t)$ . We use  $x = (x_0, \dots, x_T)$  and  $y = (y_1, \dots, y_T)$  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

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

where  $p(y_t|x_t) = N[f_t(x_t), V(x_t)]$ , and  $p(x_t|x_{t-1}) = N[g_t(x_{t-1}), W(x_{t-1})]$ . Here  $p(x_0) = N(m_0, C_0)$  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, as it requires normalization. Moreover, standard MCMC simulation techniques are not applicable due to nonlinearities in the mean and the state dependence in the variance function. Our methodology will provide an algorithm that simulates from the full posterior smoothing distribution and provides estimates of the marginal posterior distribution of the individual states  $p(x_t|y)$ .

## 2.1 An Auxiliary Mixture Model

The key idea is to approximate the SDV model by an auxiliary mixture model, denoted by  $p^a(\cdot)$ , that is conditionally linear and Gaussian given a state-dependent mixture indicator  $z_t$ . The first step of our approach is to determine an auxiliary model based on locally-weighted mixtures. It is easiest to discuss the cases where nonlinearity and state dependent variance occur in either the observation or evolution equations separately. First, suppose that the state dependence only occurs in the observation equation,

$$\begin{aligned} y_t &= f(x_t) + \epsilon_t, & \epsilon_t &\sim N[0, V(x_t)] \\ x_{t+1} &= Gx_t + \omega_t, & \omega_t &\sim N[0, W], \end{aligned} \quad (3)$$

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

We now describe details of the auxiliary mixture model. Let  $z = (z_1, \dots, z_T)$  be a vector of state-dependent mixture indicators. The likelihood for the auxiliary mixture model  $p^a(y_t|x_t)$  is defined by

$$p^a(y_t|x_t) = \sum_{k=1}^K p^a(y_t|x_t, z_t = k) p^a(z_t = k|x_t). \quad (4)$$

Conditional on the mixture indicator,  $z_t = k$ , the observation equation in the auxiliary model is a regression

$$p^a(y_t|x_t, z_t = k) = N(\alpha_k + \beta_k x_t, \tau_k^2), \quad (5)$$

with linear mean function and constant variance. For the state-dependent mixture weights  $p^a(z_t|x_t)$  we choose standardized Gaussian kernels

$$p^a(z_t = i|x_t) = \phi(x_t; \mu_i, \sigma_i) / \sum_{k=1}^K \phi(x_t; \mu_k, \sigma_k), \quad (6)$$

where  $\phi(x; m, s)$  is a normal density with mean  $m$  and variance  $s^2$ , evaluated at  $x$ . The choice of  $(\alpha_k, \beta_k, \tau_k, \mu_k, \sigma_k)$  is arbitrary, with the guiding principle being 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  $(\alpha_k, \beta_k)$  could correspond to linear expansions of  $f(x_t)$  at  $\mu_k$ . We will use the auxiliary mixture model  $p^a$  to define the proposal distribution in our Markov chain Monte Carlo (MCMC) algorithm for model (3), exploiting the fact that  $p^a(z_t|x_t, y_t)$  is a multinomial distribution and  $p^a(x|z, y)$  allows random variate generation using efficient algorithms for linear Gaussian state-space models. Details are described in the following sections.

## 2.2 Simulating from the Smoothing Distribution

To simulate from the smoothing distribution  $p(x|y)$ , we use MCMC with proposal distributions based on the auxiliary mixture model. The key feature of the auxiliary model is that the whole state vector  $x$  can be simulated from in block moves given the mixture indicators  $z$ . Details are shown in step 2, below. An appropriate rejection step (step 3, below) ensures  $p(x|y)$  as ergodic distribution of the defined Markov chain. See, for example, Tierney (1994) for a discussion of MCMC simulation for posterior inference.

The following three steps define one iteration in the simulated Markov 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 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  $z = (z_1, \dots, z_T)$ . 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  $z_1, \dots, z_T$  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 (6). Substituting (6) for  $p^a(z_t|x_t)$  we get

$$p^a(x|z, y) \propto \underbrace{p(x_0) \prod_{t=1}^T p(x_t|x_{t-1}) p^a(y_t|z_t, x_t) \phi(x_t; \mu_{(t)}, \sigma_{(t)})}_{q(x|z, y)} \frac{1}{c(x_t)}. \quad (7)$$

Here, we use the notation  $\mu_{(t)}$  and  $\sigma_{(t)}$  to represent  $\mu_{z_t}$  and  $\sigma_{z_t}$ , the knot and kernel width corresponding to the value of the indicator  $z_t$ . The first factor, denoted by  $q(x|z, y)$  in (7), corresponds to the smoothing distribution in another linear, Gaussian state space model.  $q(x|z, y)$  will serve as proposal distribution in a Metropolis-Hastings step. 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  $q(x|z, y)$ :

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

where  $\tilde{y}_t = \tilde{V}_t [(y_t - \alpha_{(t)})\beta_{(t)}/\tau_{(t)}^2 + \mu_{(t)}/\sigma_{(t)}^2]$  and  $\tilde{V}_t = (\beta_{(t)}^2/\tau_{(t)}^2 + 1/\sigma_{(t)}^2)^{-1}$ . The full smoothing distribution  $q(x|z, y)$  of this model 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

$$a(x, \tilde{x}) = \min \left\{ 1, \prod_{t=1}^T \frac{p(y_t|x_t)}{c(x_t)} \frac{c(\tilde{x}_t) p^a(y_t|\tilde{x}_t)}{p(y_t|\tilde{x}_t)} \right\}. \quad (8)$$

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 expression (8) note that

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

### 2.3 Mixtures in the Evolution Equation

The proposed approach is equally applicable for models with nonlinearities or SDVs in the evolution equation. Assume we have a state-space model that is linear in the observation equation but with SDV in the evolution equation

$$\begin{aligned} y_t &= Fx_t + \epsilon_t, & \epsilon_t &\sim N[0, V], \\ x_{t+1} &= g(x_t) + \omega_t, & \omega_t &\sim N[0, W(x_t)]. \end{aligned} \tag{9}$$

The evolution equation includes a nonlinear mean function  $g(x_t)$  and state dependent variance  $W(x_t)$ . Using the same methods as in Section 2.2, we can define an evolution mixture, where the size of the mixture will depend on the degree of nonlinearity of  $g(x_t)$  and the departure of  $W(x_t)$  from a constant function.

For problems where both  $f(x_t)$  and  $g(x_t)$  are nonlinear, or  $V(x_t)$  and  $W(x_t)$  are state-dependent, the procedure follows the same lines as above, now requiring the construction of two auxiliary mixture models. In Section 4.2, we consider an example with SDVs in both the observation and the evolution equations.

### 2.4 Static Parameters and Other Extensions

When the mean function  $f(x_t, \lambda)$  or the volatility  $V(x_t, \lambda)$  in (3) depends on a static parameter  $\lambda$ , then the auxiliary mixture model has to adapt with  $\lambda$ . To achieve this, we allow the regression parameters  $(\alpha_k, \beta_k, \tau_k)$  in (5) to depend on  $\lambda$ . For given knots  $\mu_k$  we define  $\beta_k(\lambda) = \partial f / \partial x_t(\mu_k, \lambda)$ ,  $\alpha_k(\lambda) = f(\mu_k, \lambda) - \beta_k(\lambda) \mu_k$  and  $\tau_k^2(\lambda) = V(\mu_k, \lambda)$ . Again, we implement posterior simulation based on the auxiliary mixture model. Simulating the state vector  $x$  and  $z$  given  $\lambda$  is done as described in Section 2.2, with an additional step to update  $\lambda$  given the currently imputed values of state parameters  $x$  and indicators  $z$ . The details of

this step depend on how  $\lambda$  enters the model. For example, the static parameters  $\lambda = (\Omega, \Lambda)$  in the ATSM discussed in Section 3.3 are updated using random walk Metropolis proposals.

Non-Gaussianity in the error disturbances can be incorporated using either discrete or continuous scale mixtures of normals, using the methods of Carlin et al. (1992). In Section 3.1, we provide an example that allows for jump distributions in the observation equation.

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 state parameters in the model discussed in 4.2. However, when the support of the state vector changes significantly over time, the set of knots  $\mu_k$  needs to be adaptive.

To fix ideas, consider model (3) with non-linearity and SDV in the observation equation only. Assume 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  $\tau_{tk}^2 = V(\mu_{tk})$ . The auxiliary mixture model (4) with  $p^a(y_t|z_t = k, x_t) = N(\alpha_{tk} + \beta_{tk}x_t, \tau_{tk}^2)$  is then used to proceed as in Section 2.2.

Finally, in many applications, it is necessary to do the updating of the state vector  $x$  in sub-blocks to achieve reasonable acceptance probabilities in (8). Let  $x = (x^{(1)}, \dots, x^{(J)})$  denote a partition of the states into subvectors. We repeat steps 1 through 3 of the algorithm described in Section 2.2  $J$  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.

### 3 SVOL and Affine Term Structure Models

#### 3.1 A Standard SVOL Model

SVOL models are commonly used to describe the evolution of equity returns. 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:

$$\begin{aligned} d \log S(t) &= (\mu - 1/2 v(t)) dt + \sqrt{v(t)} dW(t) \\ d \log v(t) &= \kappa_v (\theta_v - \log v(t)) dt + \sigma_v dW_1(t). \end{aligned}$$

Here  $W(t)$  and  $W_1(t)$  are independent Brownian motions,  $\kappa_v$  governs the speed of mean reversion,  $\theta_v$  is the long-run mean of log-volatility, and  $\sigma_v$  is the volatility of volatility. This can be written as an SDV model by discretizing and setting  $y_t = \log(S_{t+1}/S_t)$  and  $x_t = \log(v_t)$ ,

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

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 approximating the observation equation.

First, 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 variance function. In the bottom panel of Figures 1(a) and 1(b), we have plotted three Gaussian kernels centered at the knots,  $\mu_k$ . By construction, we require that the three kernels  $\phi(x_t; \mu_k, \sigma_k)$ , are the same for both the mean and variance functions. We choose all kernels to have a common scale factor. The kernels will serve two purposes in the approximating mixture model. First, they will serve as weighting functions in the locally-weighted mixture. Secondly, they will act as additional likelihood factors in (7). Figure 1(a) shows the mean function  $f(x_t) = \exp(x_t)$ , along with three linear regression lines,

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

The three linear regression lines act as the mean functions in the approximating mixture model. The knots,  $\mu_k$ , denoted by triangles at the bottom of Figure 1(a), are chosen to be equally-spaced in the state space of  $x_t$ . Figure 1(b) shows the variance function  $V(x_t) = \exp(x_t)$ , and three variance levels,  $\tau_k^2$  defined as  $\tau_k^2 = V(\mu_k)$ .

Hence, at each time period, we have a locally-weighted mixture (4) of linear regressions with constant variance. The mixture weights are the Gaussian kernels at the bottom of Figures 1(a) and 1(b). We can now simulate from the smoothing distribution as described in Section 2.2.

### 3.2 A Two-Factor SVOL 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 SVOL model where the future interest

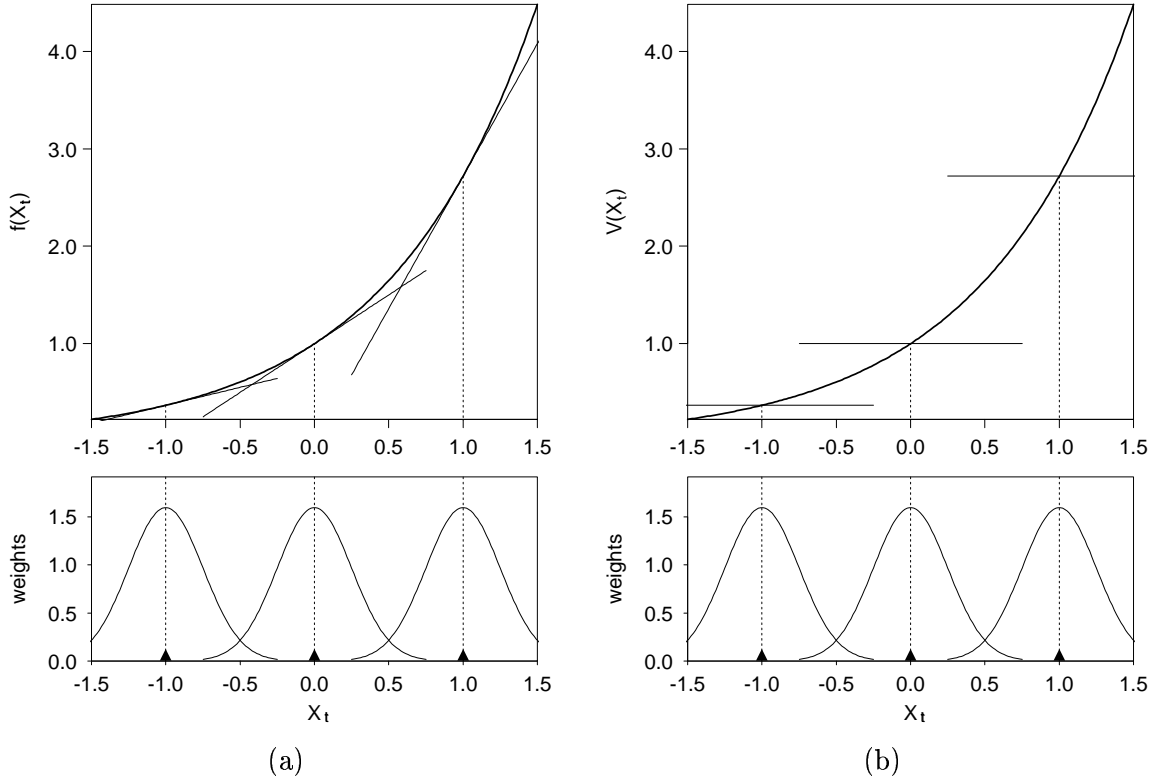


Figure 1: Construction of the auxiliary mixture model. (a) Exponential mean function  $f(x_t)$  and three approximating regression lines,  $\alpha_k + \beta_k x_t$ . (b) Exponential variance function  $V(x_t)$  and the three approximating variance levels  $\tau_k^2$ . Bottom plots in (a) and (b) show the knots  $\mu_k$ , denoted by triangles, and the Gaussian weight kernels  $\phi(x_t; \mu_k, \sigma_k)$ .

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:

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

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 an indicator of whether a jump has occurred 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  $\kappa_r$  and  $\theta_r$ . The second equation defines the evolution of the state variable  $v(t)$ .

To implement this model on discrete time data  $r_t$  we use an Euler discretization of the continuous-time model. Discretizing the above equation gives an SDV model of the form

$$\begin{aligned} y_{t+1} - y_t &= \kappa_r(\theta_r - y_t) + \sqrt{x_t e^{-y_t}} \epsilon_t + \xi_t J_t, \\ x_{t+1} - x_t &= \kappa_v(\theta_v - x_t) + \sigma_v \sqrt{x_t} \omega_t, \end{aligned} \tag{10}$$

where  $y_t = \log r_t$ ,  $x_t = v_t$ , and  $\epsilon_t$  and  $\omega_t$  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.,  $J_t \sim \text{Ber}(\lambda)$ , and that the jump sizes are normal  $\xi_t \sim N(\mu_\xi, \sigma_\xi^2)$ . 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 jump indicators  $J_t$  and jump sizes  $\xi_t$ . Implementing the algorithm outlined in Section 2.2 we require additional steps to update the static parameters and the jump indicators and 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, all of which take the form of well known distributions.

If we marginalize with respect to the jump term in the observation equation, model (10) 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

$$(1 - \lambda)N[0, x_t \exp(-y_t)] + \lambda N[\mu_\xi, x_t \exp(-y_t) + \sigma_\xi^2].$$

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, 2000).

### 3.3 Affine Term Structure Models (ATSMs)

Affine Term Structure Models (ATSMs) provide a general class of models for pricing yields of different bonds across different maturities. ATSMs were introduced by Duffie and Kan (1996) and provide a rich class of price dynamics. In this section we describe the general ATSM and show how the discretized model corresponds to a multivariate SDV model of the form (9). Model implementation requires estimation of a nonlinear SDV model for the underlying unobserved factors.

The ATSM assumes that the instantaneous short-term interest rate  $r(t)$  is unobserved and an affine function of  $N$  state variables  $x(t) = (x_1(t), \dots, x_N(t))$ ,

$$r(t) = \delta_0 + \delta'_x x(t), \tag{11}$$

where  $(\delta_0, \delta_x)$  is a vector of parameters. Typically, the first state variable  $x_1(t)$  is chosen to be  $r(t)$ , corresponding to  $\delta_0 = 0$  and  $\delta_x = (1, 0, \dots, 0)$ . The state variables,  $x(t)$ , are

assumed to follow an affine diffusion equation. Together, these assumptions imply the key property of the ATSM, namely that the yields are linear in the unobserved state variables  $x(t)$  (Dai and Singleton, 2000).

Let  $P_\tau(t)$  denote the price of a zero-coupon bond at time  $t$  that expires in  $\tau$  years, and let  $R_\tau(t) = -1/\tau \log P_\tau(t)$  denote the corresponding bond yield,

$$R_\tau(t) = -\frac{1}{\tau} [A(\tau) - B(\tau)'x(t)]. \quad (12)$$

We will define the functions  $A(\tau)$  and  $B(\tau)$  below, after discussing the  $x(t)$  process which is required in the definition of these functions.

The state vector  $x(t)$  follows an affine diffusion of the form

$$dx(t) = \mathcal{K}(\theta - x(t)) dt + \Sigma \sqrt{S(t)} dW(t). \quad (13)$$

where  $\mathcal{K}$  and  $\Sigma$  are  $N \times N$  matrices,  $\theta$  is an  $N \times 1$  vector and  $W(t)$  is a vector of  $N$  independent Brownian motions. The variance function  $S(t)$  is a  $N \times N$  diagonal matrix and is allowed to depend on the unobservable state vector  $x(t)$  in a linear fashion with elements  $[S(t)]_{ii} = \alpha_i + \beta_i'x(t)$ .

To fix ideas, suppose that bond prices are determined by a two factor ATSM driven by the unobserved short-term interest rate and its volatility, that is  $x(t) = (r(t), v(t))$  where  $dr(t) = \kappa_r(\theta_r - r(t))dt + \sqrt{v(t)} dW_1(t)$ , and  $dv(t) = \kappa_v(\theta_v - v(t))dt + \sigma_v \sqrt{v(t)} dW_2(t)$ . As Duffie and Kan (1996) note, the instantaneous short-rate  $r(t)$  is the limit of yields and due to institutional irregularities in the short rate, ATSM typically do not include the short-term bond yields in the panel dataset. Discretizing the diffusion equations for  $x(t) = (r(t), v(t))$  we get evolution equations with SDV

$$r(t+1) - r(t) = \kappa_r(\theta_r - r(t)) + \sqrt{v(t)} \omega_{1t} \quad (14)$$

$$v(t+1) - v(t) = \kappa_v(\theta_v - v(t)) + \sigma_v \sqrt{v(t)} \omega_{2t}. \quad (15)$$

The evolution equations (14) and (15) describe the dynamics of the (unobserved) instantaneous short rate and its volatility. The observed data is a multivariate panel of bond yields denoted by  $y(t) = (y_1(t), \dots, y_n(t))$ , where  $y_i(t)$  is the observed bond yield at maturity  $\tau_i$ . Assuming independent normal pricing errors  $\epsilon_i(t)$  and substituting (12) we obtain the multivariate observation equation

$$y_i(t) = -\frac{1}{\tau_i} [A(\tau_i) - B_1(\tau_i)r(t) - B_2(\tau_i)v(t)] + \sigma \epsilon_i(t), \quad i = 1, \dots, n. \quad (16)$$

The functions  $A(\tau)$  and  $B(\tau)$  in the pricing equation depend on the parameters of the unobserved state process. However, following standard financial theory the pricing formula is based on expectations under risk-neutral dynamics, rather than under the physical process

dynamics (13). The risk-neutral dynamics for the unobserved  $x(t)$  are described by an affine diffusion of the same form as (13), but with parameters  $(\tilde{\mathcal{K}}, \tilde{\theta}, \Sigma)$  replacing the physical process parameters  $(\mathcal{K}, \theta, \Sigma)$  through the equations

$$\tilde{\mathcal{K}} = \mathcal{K} + \Sigma\Phi \quad \text{and} \quad \tilde{\theta} = \mathcal{K}^{-1}\tilde{\mathcal{K}}\theta.$$

The  $i$ -th row of  $\Phi$  is given by  $\lambda_i\beta'_i$ . The transformation involves the additional parameters  $\Lambda = (\lambda_1, \dots, \lambda_N)$  known as market price of risk parameters. See Dai and Singleton (2000) for further discussion. Put simply, the risk neutral dynamics are obtained from the physical process dynamics by adjusting the drift (or mean function) using market price of risk parameters. The covariance structure is left unchanged under this change of measure.

Using the parameters of the risk neutral dynamics the coefficients  $A(\tau)$  and  $B(\tau)$  in the pricing formula are then determined by the following ordinary differential equations. Recall that  $(\delta_0, \delta_x)$  are the coefficients in (11), and  $(\alpha_i, \beta_i)$  parameterize the elements of  $S$  in equation (13).

$$\begin{aligned} \frac{dA(\tau)}{d\tau} &= -\tilde{\theta}'\tilde{\mathcal{K}}'B(\tau) + \frac{1}{2} \sum_{i=1}^N [\Sigma' B(\tau)]_i^2 \alpha_i - \delta_0, \\ \frac{dB(\tau)}{d\tau} &= -\tilde{\mathcal{K}}'B(\tau) - \frac{1}{2} \sum_{i=1}^N [\Sigma' B(\tau)]_i^2 \beta_i + \delta_x, \end{aligned} \quad (17)$$

with initial conditions  $A(0) = B_1(0) = B_2(0) = 0$ . In general, equations (17) have to be solved by numerical methods. In the special case of the two-factor model (16) the coefficient  $B_1(\tau)$  is available in closed form,  $B_1(\tau) = (1 - e^{\kappa_r\tau})/\kappa_r$ . For the coefficients  $A(\tau)$  and  $B_2(\tau)$  we solve (17) numerically using a simple Euler discretization scheme. See Duffie and Kan (1996) for other methods for solving (17).

## 4 Applications

State Dependent Variance (SDV) models have many applications in time series. In this section we implement models that were introduced in Section 3. First we model the movements in the daily short-term interest rate in Hong Kong from 1986-2000 using a SVOL 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 to model periods of 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 standard stochastic volatility models. Next, we apply our methodology to affine term structure models (ATSM) for U.S. interest rates. We analyze the multivariate panel of daily US interest rate yields from Oct. 19, 1996 to Oct. 15,

1999, using a two factor ATSM. The unobservable factors, i.e., state variables, are now the instantaneous short rate  $r(t)$  and its volatility  $v(t)$ .

#### 4.1 Hong Kong Interest Rates (HIBOR)

Hong Kong short interest rates provide a good illustration of the need to incorporate jumps. Figure 2 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 reason for the existence of jumps in this series is that the Hong Kong dollar operates under a currency board and 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.

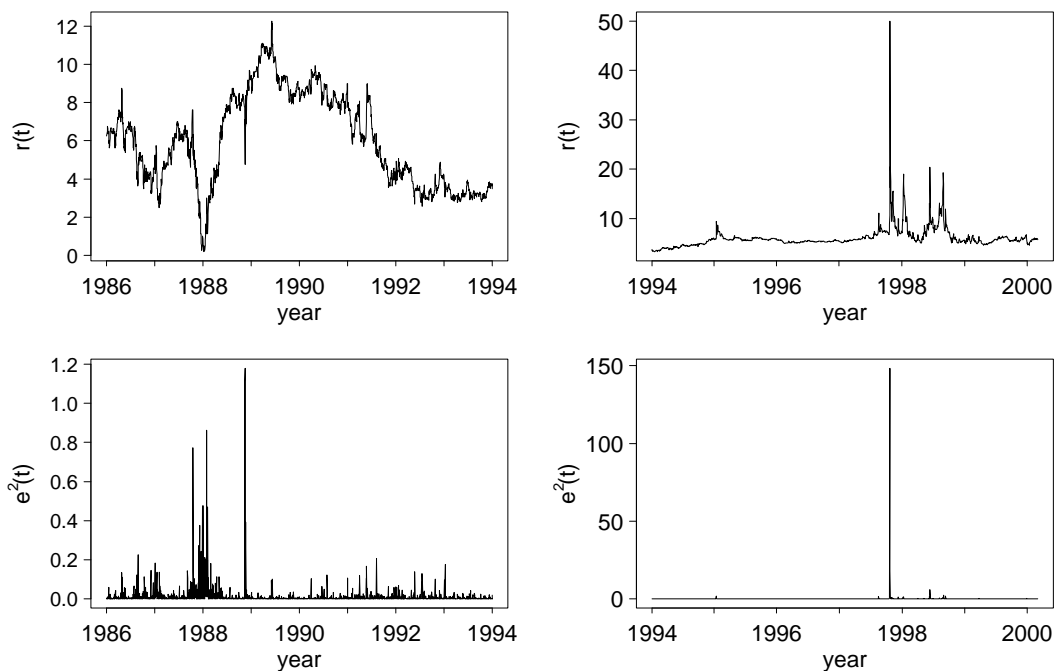


Figure 2: Hong Kong interest rate series, 1986–2000. The top panels show the short term interest rates, using a different scale before and after 1994. The bottom panels shows the squared residuals (as proxies for volatility) using a model without jumps. The plot illustrates the need to allow for jumps in the model.

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 stochastic volatility model (10) with jumps in the level of interest rates. For comparison we also analyze a two-factor stochastic volatility

Table 1: Posterior means and standard deviations of model parameters.

Model	$\kappa_r \theta_r$	$1 - \kappa_r$	$\kappa_v \theta_v$	$1 - \kappa_v$	$\sigma_v$
No Jumps	.02 (.09)	.99 (.03)	.001 (.008)	.900 (.103)	.025 (.005)
Jumps	.0218 (.0074)	.987 (.0043)	.0001 (.000011)	.935 (.011)	.0002 (.000032)

CIR model without jumps. Figure 2 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.

The first row of Table 1 shows the posterior means and standard deviations of the model parameters for the no-jump model. The model is forced to estimate a high level of  $\sigma_v$ . The residual plots (not shown) are highly non-normal indicating a very poor description to the underlying series. The second row in Table 1 shows the posterior moments under the model with jumps. The major difference is that, after allowing for jumps, the estimate of  $\sigma_v$  decreases, whereas our estimate of the mean reversion parameter  $\kappa_v$  increases, implying more persistence in the volatility sequence. This leads to very different pricing implications for the two models.

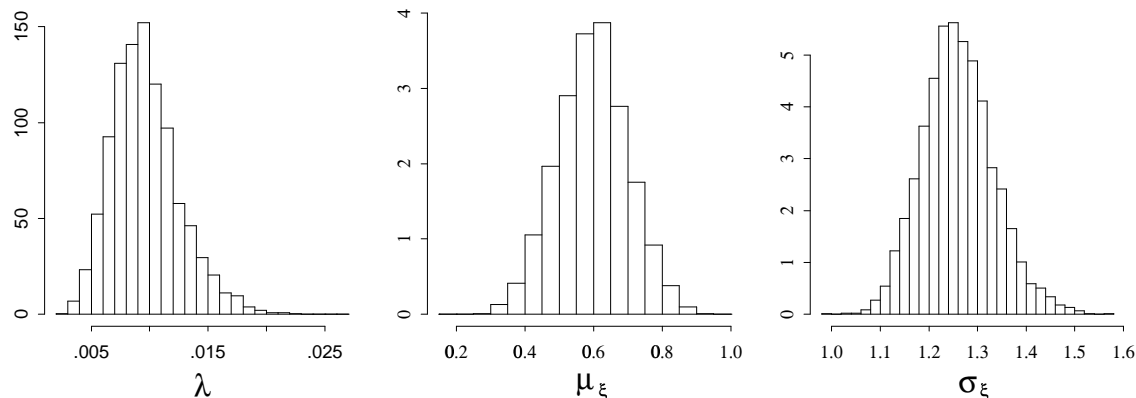


Figure 3: Posterior distribution of jump process parameters: jump probability  $\lambda$ , and mean  $\mu_\xi$  and standard deviation  $\sigma_\xi$  of the jump sizes.

Our methodology also provides inference for the jump component of the model. Figure 3 shows the posterior distributions  $p(\lambda|y)$  on the jump probability  $\lambda$  and mean  $\mu_\xi$  and standard deviation  $\sigma_\xi$  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 analysis, Figure 4 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 is also provided for comparison.

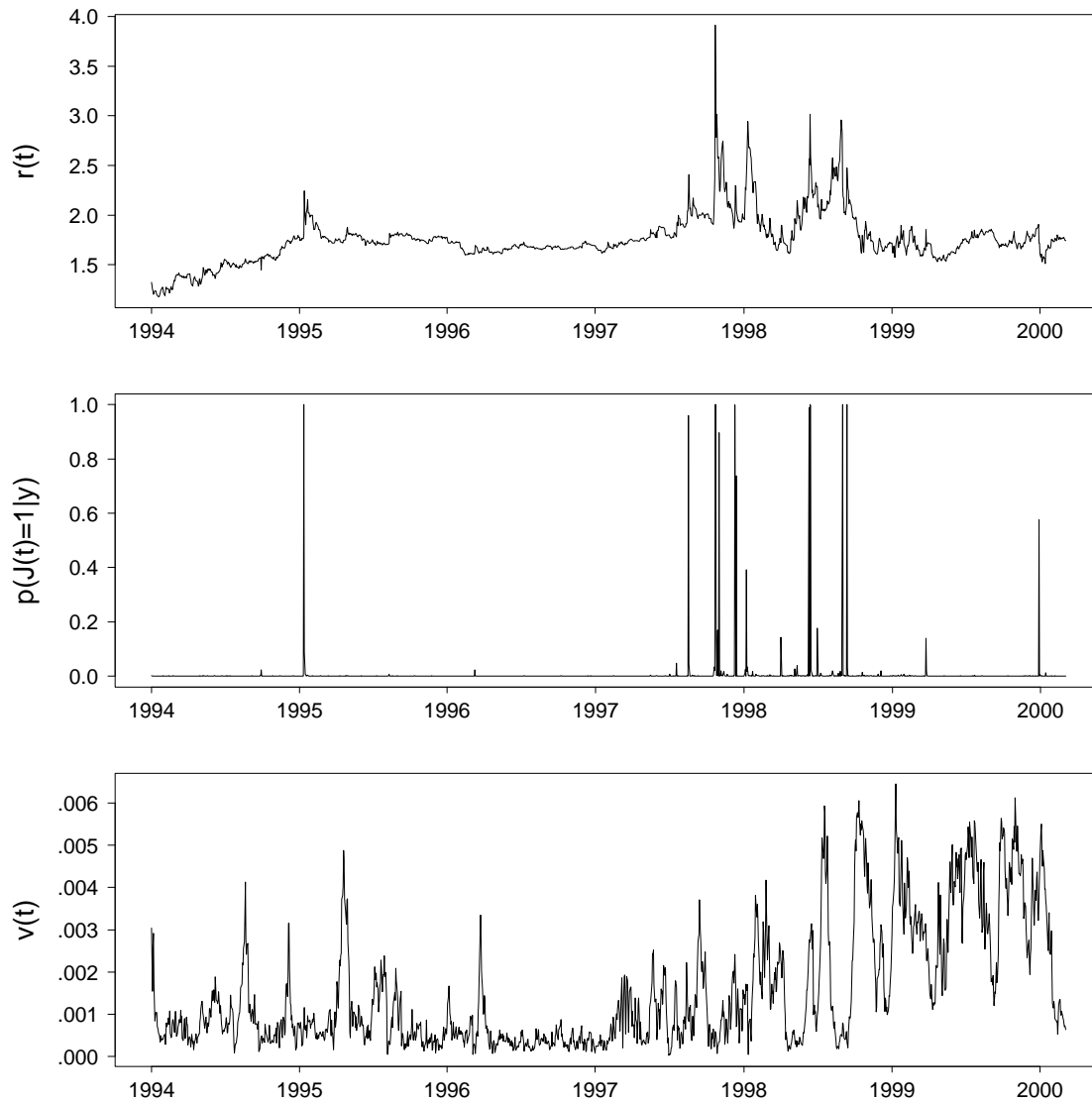


Figure 4: (a) Log HIBOR rates (1994-2000). (b) Posterior probability of jump occurrences  $p(J(t) = 1|y)$ . (c) Smoothed volatility estimates  $E(v(t)|y)$ .

## 4.2 Yield Curve Modeling: US Interest Rates, 1996-1999

To provide an illustration of the ATSM we implement a two-factor ATSM for a multivariate panel of daily bond yields. The panel includes maturities of  $\tau_1 = 1$ ,  $\tau_2 = 2$ ,  $\tau_3 = 5$ ,  $\tau_4 = 10$ , and  $\tau_5 = 30$  years. Figure 5 plots the one-year rate, and the differences between the 1-year and the 5-year, 10-year and 30-year bond yields, respectively. One period of special interest is late Oct. 1998, where the short interest rate fell from about 5% to 3% and the 30-year bond yields fell to a low of 4.7%.

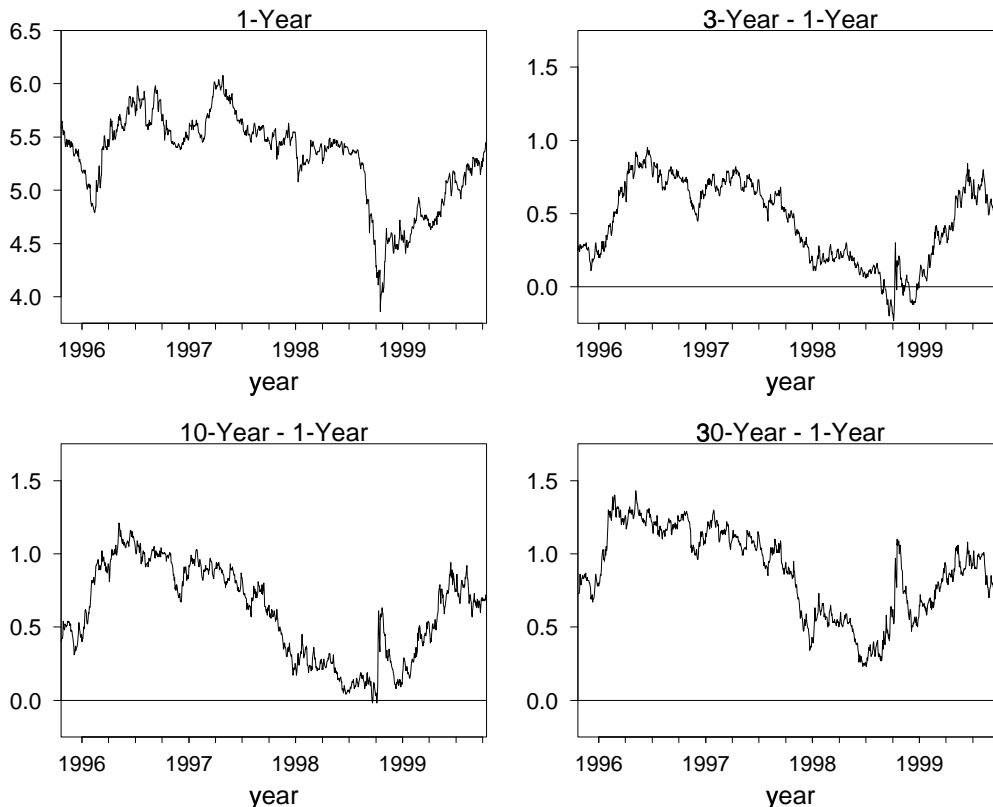


Figure 5: U.S. constant maturity bond yields (Oct. 1995 – Oct. 1999): 1-year (top left); difference between 5-year and 1-year yields (top right); difference between 10-year and 1-year yields (bottom left); difference between 30-year and 1-year yields (bottom left). Notice the sharp drop in Oct. 1998.

Our analysis of the two-factor ATSM proceeds in two steps. First, we use observed three-month T-bill rates  $\hat{r}(t)$  as proxies for the instantaneous short-rate  $r(t)$  and fit a univariate SDV model (14) and (15) for the observed  $\hat{r}(t)$ . We will refer to this model as the “short-rate SDV model”. This analysis provides us with initial values for  $\mathcal{K}, \theta, \Sigma$ , and the volatilities  $v(t)$ . Using these initial values we fit the two-factor ATSM described in Section 3.3 to the panel of bond yields.

If the two-factor model of the yield curve were exact, there would be no pricing error and the latent state variables  $r(t), v(t)$  and static parameters  $\mathcal{K}, \theta, \Sigma$  would be identical to those of the instantaneous short-rate process. Thus, to the extent to which the three-month T-bill rate is a good proxy for instantaneous short-rates, posterior inference under the two models should match. We will show some comparison in the following figures. In any case, inference under the short-rate SDV model provides reasonable starting values for MCMC simulation in the two-factor ATSM.

We now describe our MCMC algorithm. Following the notation of the previous sections, let  $y = (y(1), \dots, y(T))$  denote the time series of bond yields,  $r = (r(1), \dots, r(T))$  the vector of short rates, and  $v = (v(1), \dots, v(T))$  the vector of volatilities. We define  $\Omega = (\mathcal{K}, \theta, \Sigma)$  as the vector of parameters under the physical process (13), and  $\Lambda = (\lambda_r, \lambda_v)$  as the market price of risk parameters. The problem is to sample from the joint posterior distribution of the state variables  $(r, v)$  and the parameters.

The joint posterior distribution of the state variables and the parameters is given by

$$p(r, v, \Omega, \Lambda, \sigma | y) \propto p(y | r, v, \Omega, \Lambda, \sigma) p(r | v, \Omega) p(v | \Omega) p(\Omega, \Lambda, \sigma), \quad (18)$$

To implement MCMC, we need the complete conditionals, that is, the posterior conditional distribution of each parameter given the others (see, for example, Tierney, 1994). At each iteration of the MCMC, all static parameters and state variables are updated. We now describe each update in turn.

The complete conditional of the unobserved short rate,  $p(r | v, \Omega, \Lambda, \sigma, y)$ , follows a heteroscedastic normal linear state-space model with evolution equation given by (14) and observation equation given by the bond pricing equation (16). We simulate from this distribution using the forward filtering, backward sampling algorithm of Carter and Kohn (1994) and Frühwirth-Schnatter (1994). For the volatility states, the complete conditional,  $p(v | r, \Omega, \Lambda, \sigma, y)$ , corresponds to an SDV model with state-dependent variances in the evolution and observation equations. The evolution equation is given by (15), while the observation equation contains the short rate process (14) and the bond pricing equation (16). The short rate equation has an SDV, while the pricing equation is a standard linear regression. The vector of volatility states  $v$  is then updated using the algorithm described in Section 2.2. Both the observation and evolution equations are SDVs, so we use two mixtures with 10 equally-spaced knots placed according to quantiles of the stationary distribution of volatility.

For the complete conditionals for the parameters  $\Omega$  and the market price of risk parameters  $\Lambda$ , we use a component-wise random walk Metropolis step. The proposal distribution is a normal centered at the current value with a scale factor tuned to yield an acceptance probability of about 40%. The complete conditional for the pricing error variance  $\sigma^2$ ,

$p(\sigma^2|r, v, \Omega, \Lambda, y)$ , is a standard inverse-gamma distribution. Throughout, we use proper but vague prior distributions for all parameters.

We now turn to our estimation and smoothing results for  $(v(t), r(t))$ . Figure 6 plots the posterior distributions of the parameters  $(\theta_r, \kappa_r, \sigma_r)$  of the underlying short rate process  $r(t)$  for both the ATSM and the short-rate SDV model. Table 2 reports the corresponding posterior means and standard deviations. Estimation of the mean reversion parameter and the central tendency parameter under the ATSM leads to a markedly sharper posterior distribution. The multivariate panel of daily bond yields provides significantly more information about the short rate dynamics than analysis of the short-rate time series alone.

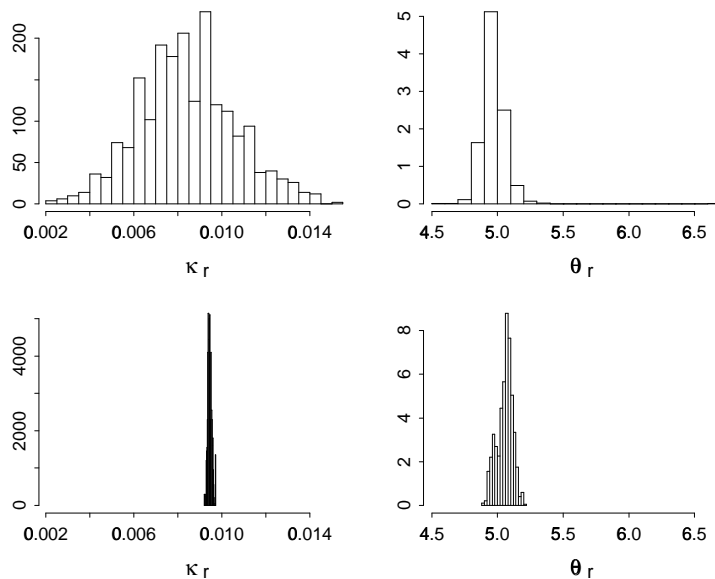


Figure 6: Posterior distributions of  $(\kappa_r, \theta_r)$  under the short-rate SDV model (top), and the ATSM for daily yields (bottom). Put simply, compared to the uncertainties in the analysis of the observed short-rate time series only, the market has practically perfect knowledge of the short-rate dynamics. Here, the market is represented by the daily bond yields.

Figure 7 plots the posterior distributions for the parameters of the volatility evolution (15) under the ATSM and the short-rate SDV model. Table 2 gives the posterior means and standard deviations. Again, the parameters  $(\kappa_v, \theta_v)$  are estimated very precisely under the ATSM. Only inference for the volatility of volatility parameter  $\sigma_v$  does not match under the two models.

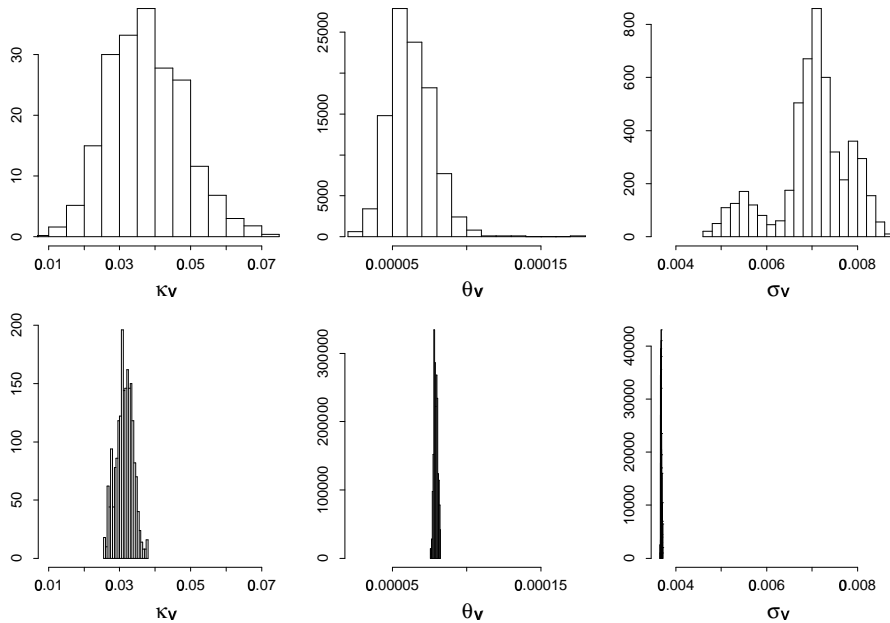


Figure 7: Posterior distributions of  $(\kappa_v, \theta_v, \sigma_v)$  under the short-rate SDV model (top), and the ATSM (bottom).

There are at least two important reasons for this conflict. First, the three-month rate is used as a proxy for the instantaneous short rate in the short-rate SDV model. The three-month rate contains a number of jumps (not modeled in this framework), which necessarily shifts the posterior distribution on  $\sigma_v$  towards larger values to accommodate these jumps. Secondly, the ATSM includes a pricing error  $\sigma$ , which accounts for some of the variability in the  $v(t)$  series. This suggests that if the pricing errors were smaller, the ATSM estimate of  $\sigma_v$  would tend towards higher values.

Table 2: Posterior means and standard deviations of model parameters.

Parameter	ATSM		Short-rate SDV	
$\lambda_r$	564	(2.0)	---	---
$\lambda_v$	94200	(1000)	---	---
$\sigma$	.072	(.001)	---	---
$\kappa_r$	.0095	(.0002)	.0082	(.002)
$\theta_r$	5.05	(.110)	4.95	(.155)
$\kappa_v$	.033	(.005)	.040	(.015)
$\theta_v$	.000075	(.000003)	.00006	(.00002)
$\sigma_v$	.0036	(.0001)	.007	(.001)

Figure 8 plots the smoothed series for  $r(t)$ ,  $v(t)$  and compares them with the observed three-month T-Bill rate and the volatility estimate under the short-term SDV model. Notice that for the short rate level  $r(t)$ , the latent factors in the ATSM fit the period in Oct. 1998 very precisely, but at the beginning of the series in 1996, the ATSM favors a higher level of volatility,  $v(t)$  rather than the lower level of interest rates  $r(t)$  to fit the observed yield curve. The volatility state variable,  $v(t)$ , has the general shape of that obtained using the three-month rate, but misses the large jump in Oct. 1998.

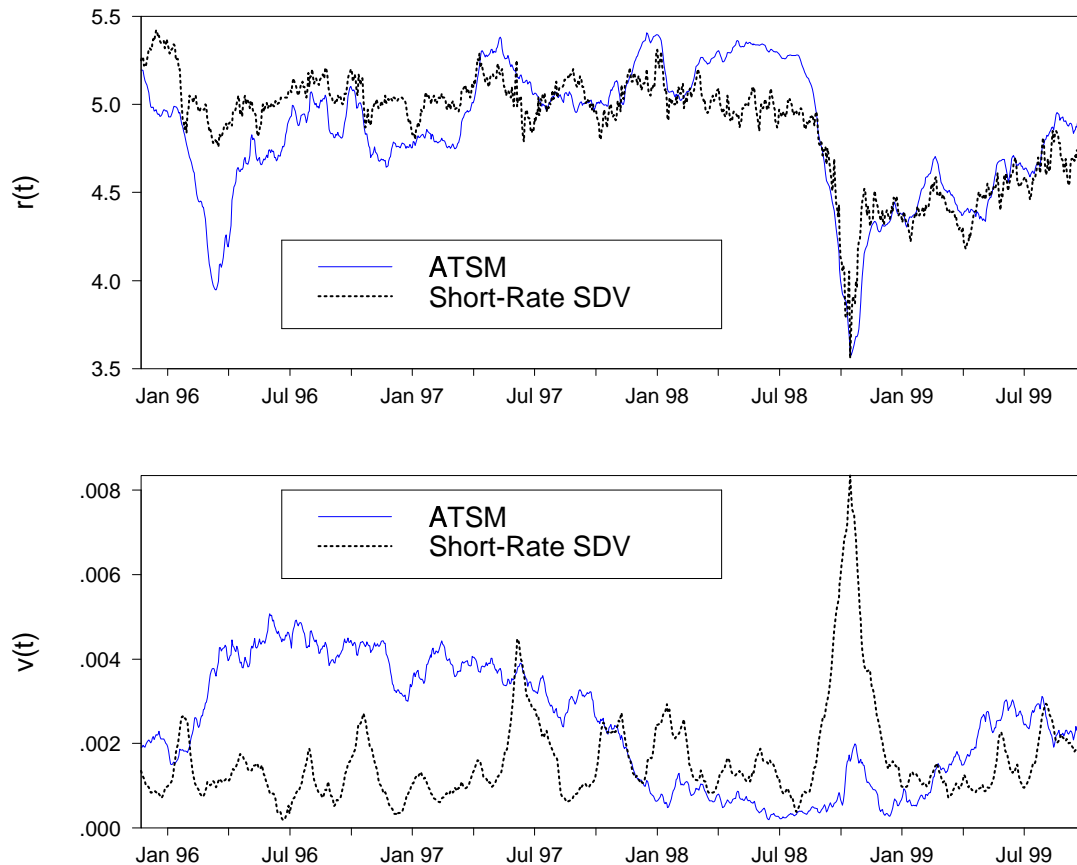


Figure 8: Top: Smoothed state variable  $r(t)$  versus observed 3-month T-Bills  $\hat{r}(t)$ . Bottom: Smoothed state volatilities  $E(v(t)|y)$  versus estimated volatility from the short-rate SDV model.

Figure 9 plots the posterior distribution for the market price of risk parameters  $\Lambda = (\lambda_r, \lambda_v)$  and the pricing error parameter  $\sigma$ . The market risk parameters have the expected signs:  $\lambda_r$  is negative, and  $\lambda_v$  is positive. The posterior mean of the pricing error is 0.07%, or 7 basis points (bps).

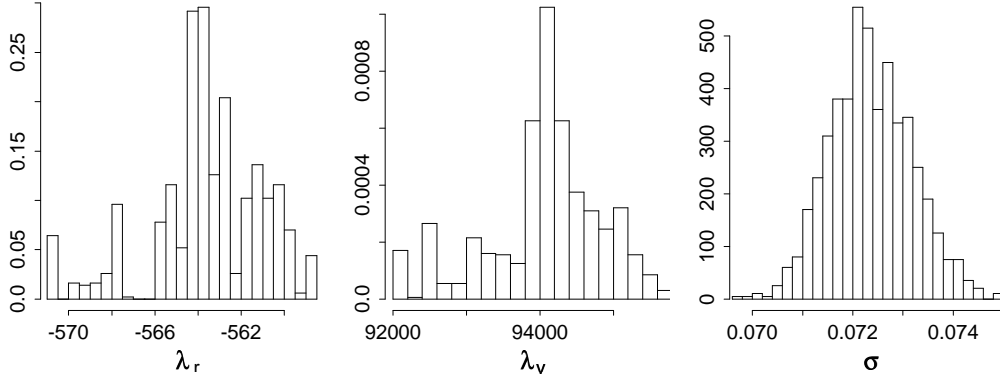


Figure 9: Posterior distribution of market price of risk and pricing error parameters.

There are a number of directions for extending the pricing error specification. One would be to include correlated errors, another would be a sensitivity analysis with respect to the distribution  $p(\sigma)$ . For example, imposing a smaller pricing error than estimated here would lead to smoothed states  $r(t)$  and  $v(t)$  that would more closely match those of the short-rate SDV model assuming that the two-factor ATSM provides an adequate description of the data.

## 5 Conclusions

In this paper, we develop a general likelihood-based approach for inference in state-space models with nonlinear mean functions and state-dependent variances (SDV). SDV models allow the variance functions of either the observation or evolution equation,  $V(x_t)$ ,  $W(x_t)$ , respectively, to depend on the unobserved state  $x_t$ . This class of models contains both stochastic volatility (SVOL) and affine term structure models (ATSMs). These models are becoming commonplace in time-series applications and in pricing of derivative securities. We propose a simulation-based methodology that requires a specification of an auxiliary mixture model. Samples from the full smoothing distribution of the unobserved states  $x_t$  are provided by our algorithm together with inference on the parameters and predictions for future observations. Our methodology provides an alternative to the simulated method of moments approach with the added benefit of smoothing for the unobserved states  $x_t$ .

We illustrate our methodology by modeling the evolution of interest rates in Hong Kong and the U.S. We use two SDV models: an SVOL model with jumps, and a two-factor ATSM. Our method is computationally efficient and allows the researcher to tackle problems such as on-line filtering and prediction for derivative pricing. Due to the block nature of our auxiliary mixture model and the use of simulation algorithms it is possible to extend our methodology to provide the sequence of filtered distributions of the state. Current

methodologies for on-line filtering are based on particle-filtering methods, or sequential Monte Carlo sampling. See, for example, Gordon, Salmond and Smith (1993), Carpenter, Clifford and Fearnhead (1997), Pitt and Shephard (1999) and Doucet, Godsill and Andrieu (2000).

## References

- Andersen, T., Benzoni, L. and Lund, J. (1998) Estimating jump diffusions for equity returns. *Tech. rep.*, Kellogg School of Management, Northwestern University.
- 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.
- Carpenter, J., Clifford, P. and Fearnhead, P. (1997) An improved particle filter for nonlinear problems. *IEE Proceedings – Radar, Sonar and Navigation*, **146**, 2–7.
- Carter, C. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika*, **81**, 541–553.
- Cox, J., Ingersoll, J. and Ross, S. (1985) An intertemporal general equilibrium model of asset prices. *Econometrica*, **53**, 363–384.
- Dai, Q. and Singleton, K. (2000) Specification analysis of affine term structure models. *Journal of Finance*, **55**.
- deJong, P. and Shephard, N. (1995) The simulation smoother for time series models. *Biometrika*, **82**, 339–350.
- Doucet, A., Godsill, S. and Andrieu, C. (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, **10**, 197–208.
- Duffie, D. and Kan, R. (1996) A yield-factor model of interest rates. *Mathematical Finance*, **6**, 379–406.
- Eraker, B., Johannes, M. and Polson, N. (2000) 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.

- Gordon, N., Salmond, D. and Smith, A. (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings*, vol. F-140, 107–113. IEE.
- Hürzeler, M. and Künsch, H. (1998) Monte Carlo approximations for general state-space models. *Journal of Computational and Graphical Statistics*, **7**, 175–193.
- Jacquier, E., Polson, N. and Rossi, P. (1994) Bayesian analysis of stochastic volatility (with discussion). *Journal of Business and Economic Statistics*, **12**, 371–417.
- Kitagawa, G. (1996) Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational 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.
- Shephard, N. and Pitt, M. (1997) Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, **84**, 653–667.
- 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.
- Zehnirith, B. (1988) A generalization of the Kalman filter for models with state-dependent observation variance. *Journal of the American Statistical Association*, **83**, 164–167.