Affine State Dependent Variance Models

Nicholas G. Polson [1]
Graduate School of Business, University of Chicago

Jonathan R. Stroud
Department of Statistics, Wharton School, University of Pennsylvania

Peter Müller
Department of Biostatistics, University of Texas M.D. Anderson Cancer Center

In this paper we provide a methodology for inference in affine state dependent variance models (ASDVs). We consider an application to a subclass of models known as Affine Term Structure Models (ATSMs). The latter are commonly used for modeling financial time series and for modeling the term structure of interest rates. Standard statistical methodology for state space models are inappropriate due to nonlinearities inherent in the state dependent variances. We develop a methodology that addresses these issues which uncovers the state smoothing and parameter distributions of interest. We illustrate our methodology by modeling the term structure of daily U.S. interest rates from 1996–1999. We use a two factor ATSM with states given by the instantaneous short-term interest rate and its volatility. We provide sharper parameter estimates of the short interest rate and volatility dynamics versus standard interest rate models. This is due to the fact that our model incorporates all the pricing information about the parameters implicit in the panel of interest rates.

Keywords: Affine State Dependent Variance Models, Affine Term Structure Models, State-Space Models, State-Dependent Variance, Nonlinear Time Series, MCMC, Interest rates, Yield Curve Modeling, Stochastic Volatility.

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

1 Introduction

State dependent variance (SDV) models allow for a stochastic evolution of a vector of unobserved states $\underline{x}$ and data $\underline{y}$ of the form

$$\begin{aligned} y_t &= F(x_t, \Theta) + \sigma(x_t)\epsilon_t \\ x_{t+1} &= G(x_t, \Theta) + \Sigma(x_t)\omega_t \end{aligned}$$

where $\sigma(x_t)$ and $\Sigma(x_t)$ determine the state dependent variance functions. In this paper, we consider the class of affine state dependent variance (ASDV) models that make the additional assumption that the conditional mean functions are affine in the state variable x_t, namely $F(x_t, \Theta) = F(\Theta)x_t$ and $G(x_t, \Theta) = G(\Theta)x_t$. Our approach to inference and smoothing is simulation based and we discuss the extensions to filtering. The main advantage of our methodology is that we allow the states to be updated in block fashion.

Our methodology is related to the block sampling schemes of Carter and Kohn (1994), Frühwirth-Schnatter (1994), and deJong and Shephard (1995). We deal with parameter nonlinearities and non-normality of the errors using the methods in Carlin, Polson and Stoffer (1992). Despite the importance of ASDV models in financial time series analysis, there are no existing approaches for efficient posterior simulation in such models. While brute force methods for filtering in such models, for example, the particle filter (Pitt and Shephard, 1999; Liu and West, 1999; Doucet et al., 2000) are available, our approach has the benefit of block updating of states in an efficient MCMC algorithm.

An important area of application of ASDV models is in financial time series analysis. For example, affine term structure models (ATSMs) are particularly useful for modeling interest rates, and they can be complemented using the methodology described here. Current statistical methodologies for ATSMs (see, for example, Duffie and Kan, 1996; Dai and Singleton, 2000; Singleton, 2001; Andersen, Benzoni and Lund, 2001) provide only parameter estimates and are based on the inefficient simulated method of moments do not uncover the smoothing distribution of the state vector. A further caveat is that they are intractable for filtering.

To illustrate our methodology, we describe how to implement ATSMs whose primary application is term structure modeling and the evolution of interest rate time series. Interest rates and bond yields are assumed to be affine in the unobserved state variables, for example the instantaneous short-term interest rate and its volatility.

Parameter estimation has remained problematic due to the state and parameter nonlinearities inherent in the model. State variable smoothing requires computation of the posterior distribution of the state at each time point. As an ATSM assumes that yields are linear in these variables, it is not complicated by nonlinearities. However this information has to be combined with the state dependent variances that occur in the evolution equation.

Our methodology has two main advantages over existing methods. First, we uncover the full posterior distribution of the unobserved state vector. Secondly, we provide sharper parameter estimates of the short interest rate and volatility dynamics than standard models. For a discussion of standard models for estimating continuous time short-rate processes, see Aït-Sahalia (1996) and Barndorff-Nielsen and Shephard (2001). Our results concerning sharp parameter estimates are also in line with financial theory where bond yields are derived conditional on the parameters of the state dynamics.

The rest of the paper is outlined as follows. Section 2 describes the underlying methodology of ASDVs and ATSMs. Section 3 describes how to construct our MCMC algorithm for uncovering the unobserved states and the parameters of the model. 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 using a Metropolis step to re-weight these samples. Section 4 illustrates our approach with a two-factor ATSM for U.S. bond yields from 1996–1999. We use two unobserved states given by the unobserved instantaneous short-term interest rate and the short-rate volatility. The data consist of a panel of daily bond yields across different maturities. Finally, Section 5 concludes.

2 Affine State Dependent Variance Models

Affine State Dependent Variance (ASDV) models provide a convenient modeling approach for the dynamics of a data vector $\underline{y} = (y_1, \ldots, y_T)$ as a function of unobserved state variables $\underline{x} = (x_1, \ldots, x_T)$ and parameters Θ. More specifically, the observation and evolution equation are given by

$$\begin{aligned} y_t &= F(\Theta)x_t + \sigma(x_t)\epsilon_t \\ x_{t+1} &= G(\Theta)x_t + \Sigma(x_t)\omega_t \end{aligned} \quad (1)$$

where ϵ_t and ω_t are typically standard normals. ASDV models contain a number of commonly used models: stochastic volatility models (Jacquier, Polson and Rossi, 1994), where $\sigma(x_t) = \exp(x_t/2)$, $\Sigma(x_t) = \sigma_v$, $F(x_t, \Theta) = \mu$, $G(x_t, \Theta) = \kappa(\theta - x_t)$; the square-root stochastic volatility model (Heston, 1993; Barndorff-Nielsen and Shephard, 2001) with $\sigma(x_t) = \sqrt{x_t}$, $\Sigma(x_t) = \sqrt{x_t}$, $F(x_t, \Theta) = \kappa_y(\theta_y - y_t)$ and $G(x_t, \Theta) = \kappa_x(\theta_x - x_t)$. Finally, it contains the class of affine term structure models (ATSMs).

Statistical inference is given by the full joint posterior distribution of the states and parameters $(\underline{x}, \Theta)$ given $\underline{y}$, namely the joint posterior, $p(\underline{x}, \Theta|\underline{y})$. This posterior distribution can be used to perform smoothing $p(\underline{x}|\underline{y})$, By Bayes' rule this posterior is given by

$$p(\underline{x}, \Theta|\underline{y}) \propto p(\underline{y}|\underline{x}, \Theta)p(\underline{x}|\Theta)p(\Theta)$$

Posterior simulation from $p(\underline{x}, \Theta|\underline{y})$ requires implementation of a nonlinear state-space model with state-dependent variance. This can be achieved as follows. **[FIX]** Now, break the problem into the two conditionals $p(\underline{x}|\Theta, \underline{y})$ and $p(\Theta|\underline{x}, \underline{y})$ we see that nonlinearities are still apparent in both. Dealing with $p(\Theta|\underline{x}, \underline{y})$ can generally be handled with standard MCMC techniques as inference on the static parameters is typically straightforward given the states. Generating the conditional of the states $p(\underline{x}|\Theta, \underline{y})$ given the parameters in an efficient block sampling approach is much harder. Here we give a brief description of our approach which is based on Stroud, Müller and Polson (2001). Section 3 provides an explicit derivation for a two factor ATSM model.

The proposed block sampling approach exploits the following results. Our method is based on finding an auxiliary model from which we can then use a Metropolis proposal. We now describe the method for finding the auxiliary model. First, conditional on Θ, replacing the state dependent variance terms $\sigma(x_t)$ and $\Sigma(x_t)$ by fixed values σ^* and Σ^*, and assuming normal errors, the ASDV model reduces to a standard normal linear state space model. The corresponding normal linear state space model approximates (1). The quality of the approximation can be greatly improved by using a mixture of such models.

Consider then a finite grid of values $\{\sigma^*(k), \Sigma^*(k),\ k = 1, \ldots, K\}$ and replace the evolution by a locally weighted mixture model. To do this, introduce a state dependent weight $w_k(x_t)$ in the mixture, and use a mixture indicator $z_t \in \{1, \ldots, K\}$

for each time point. This leads to an auxiliary mixture model of the form

$$\begin{aligned} y_t &= F(\Theta)\, x_t + \sigma^*(z_t)\, \epsilon_t \\ x_{t+1} &= G(\Theta)\, x_t + \Sigma^*(z_t)\, \omega_t, \quad Pr(z_t = k | x_t) = w_k(x_t). \end{aligned}$$

Therefore, using the locally weighted mixture with state dependent weights $w_k(x_t)$ we can achieve close approximation to the model and hence a proposal distribution to MCMC (1). However, with the state dependent weights the model loses the normal linear state space form. However, a simple trick still allows us to use the efficient block sampling schemes for normal linear state space models. If the weights $w_k(x_t)$ are chosen proportional to Gaussian kernels in x_t, then the posterior distribution $p(\underline{x}|\underline{z}, \Theta, \underline{y})$ in the mixture model is almost a normal linear state space model except there is an additional factor due to the normalization of the weights $w_k(\cdot)$ across k. This can be handled by placing it into the acceptance probability mixture model with state-dependent weights. It is this mixture model which we use to define an efficient block sampling approach for inference. A number of other authors have proposed mixture models as proposal distribution, however they assume constant weights. The flexibility of our approach is the state-dependence of $w_t(x_t)$. As in any Metropolis-Hastings algorithm, the stationary distribution is always exactly the desired posterior distribution from the original model, namely $p(\underline{x}, \Theta|\underline{y})$.

3 Affine Term Structure Models

Affine Term Structure Models (ATSMs) are used to model bond yields of different maturities over time. ATSMs were introduced by Duffie and Kan (1996) and have been analysed by a number of authors Duffie and Singleton (1999). More specifically, the yield $y_{\tau,t}$ of a bond with maturity τ at time t is assumed to be driven by a state variable x_t and parameter vector Θ leading to a functional relationship: $y_{t,\tau} = F(\Theta)x_t$. The explicit functional form of $F(\Theta)$ is typically non-analytic and requires computation of the coefficients of the state variables that are solutions of ordinary differential equations (ODEs). Notice that the relationship to the state vector x_t is affine.

The ATSM assumes that the state vector x_t follow an affine diffusion equation, specifically

$$dx_t = \mathcal{K}\,(\theta - x_t)\, dt + \Sigma(x_t)\, dW_t \quad \text{with} \quad \Sigma(x_t) = V S_t. \tag{2}$$

$\mathcal{K}$ and Σ are $N \times N$ matrices, θ 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 assume to depend on the unobservable state vector x_t in a linear fashion, with the i-th diagonal element $S_{t,ii} = \alpha_i + \beta_i' x_t$. The dynamics (2) are known as the *physical process* for the state variables x_t given the parameters Θ. We partition the parameter vector into $\Theta = (\Lambda, \mathcal{K}, \theta, V)$, with Λ occurring only in the sampling distribution for y_τ, and $\Omega = (\mathcal{K}, \theta, V)$ parameterizing the dynamics of the state vector x_t.

$$dx_t = \tilde{\mathcal{K}}(\tilde{\theta} - x_t)dt + \Sigma(x_t)\, dW_t \quad \text{with} \quad \Sigma(x_t) = VS_t. \tag{3}$$

In the derivation of the pricing equation, we will need to refer to a second probability model for the evolution of the state process x_t, namely the risk neutral dynamics. The risk-neutral dynamics are described by the same affine diffusion as (2), but with parameters $(\tilde{\mathcal{K}}, \tilde{\theta}, V)$ transformed from the physical process parameters $(\mathcal{K}, \theta, V)$ through the equations

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

The i-th row of the vector Φ is given by $\lambda_i \beta_i'$. Recall that β_i are the coefficients in $V_{t,ii} = \alpha_i + \beta_i' x_t$.

The coefficients $\Lambda = (\lambda_1, \ldots, \lambda_N)$ are known as market price of risk parameters and are additional parameters that need to be estimated, which transform the physical process dynamics (2) to the risk-neutral dynamics. Put simply, they transform the drift while leaving the covariance structure unchanged. See Dai and Singleton (2000) for further discussion.

3.1 The Likelihood and Pricing Equation

We are now ready to discuss details of the likelihood function $p(y_{t,\tau}|x_t, \Theta)$ which is derived from the pricing equation.

Given an evolution of the unobserved instantaneous short interest rate process r_t and a risk neutral evolution of the state variable x_t the yield $y_{\tau,t}$ is given by

$$y_{\tau,t} = -\frac{1}{\tau} \log E^Q \left[e^{-\int_t^\tau r(s)ds} | x_t, \Theta \right] \tag{4}$$

where $E_t^Q [\cdot | x_t, \Theta]$ denotes the conditional expectation under the risk-neutral probability distribution Q given the information at time t.

The main advantage of the ATSM class is that yields $y_{\tau,t}$ can be solved in near-closed form and the resulting pricing equation is affine in the state vector x_t. Specifically, we find $y_{\tau,t} = -\frac{1}{\tau}\left[A(\tau,\Theta) - B(\tau,\Theta)'x_t\right]$. For the purpose of inference we add a small normal pricing error $\sigma\epsilon_t$ to avoid stochastic singularities in the state vector.

The coefficients $A(\tau,\Theta)$ and $B(\tau,\Theta)$ are derived from the pricing equation and the assumption that the instantaneous short rate r_t is an affine function of the N state variables $x_t = (x_{1t},\ldots,x_{Nt})$, that is $r_t = \delta_0 + \delta_x' x_t$. Typically, r_t is simply chosen as one of the state variables, giving $\delta_0 = 0$ and $\delta_x = (1,0,\ldots,0)$.

Using the parameters of the risk neutral dynamics the coefficients $A(\tau,\Theta)$ and $B(\tau,\Theta)$ are then determined by the following ordinary differential equations. Recall that (α_i,β_i) parameterize the elements of S in equation (2).

$$
\begin{aligned}
\frac{dA(\tau,\Theta)}{d\tau} &= -\tilde{\theta}'\tilde{\mathcal{K}}'B(\tau,\Theta) + \frac{1}{2}\sum_{i=1}^{N}\left[V'B(\tau)\right]_i^2 \alpha_i \;-\; \delta_0, \\
\frac{dB(\tau,\Theta)}{d\tau} &= -\tilde{\mathcal{K}}'B(\tau,\Theta) - \frac{1}{2}\sum_{i=1}^{N}\left[V'B(\tau,\Theta)\right]_i^2 \beta_i \;+\; \delta_x, \qquad (5)
\end{aligned}
$$

with initial conditions $A(0) = 0$ and $B_i(0) = 0$ for $i = 1,\ldots,N$. In general, equations (5) have to be solved by numerical methods.

3.2 A Two-Factor ATSM

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 $x_t = (r_t, v_t)$. The observed data is a multivariate panel of bond yields denoted by $y_t = (y_{1t},\ldots,y_{Nt})$, where $y_{it} \equiv y_{\tau_i,t}$ is the observed bond yield at maturity τ_i. Assuming independent normal pricing errors ϵ_{it} and substituting (4) we obtain the multivariate observation equation

$$
y_{it} \;=\; -\frac{1}{\tau_i}\left[A(\tau_i,\Theta) - B_1(\tau_i,\Theta)\,r_t - B_2(\tau_i,\Theta)\,v_t\right] \;+\; \sigma\,\epsilon_{it}, \;\text{ for } i = 1,\ldots,n. \qquad (6)
$$

Discretizing the diffusion equations (2) for $x_t = (r_t, v_t)$ we get evolution equations

$$
\begin{aligned}
r_{t+1} &= r_t + \kappa_r(\theta_r - r_t) + \sqrt{v_t}\,\omega_{1t} \qquad (7) \\
v_{t+1} &= v_t + \kappa_v(\theta_v - v_t) + \sigma_v\sqrt{v_t}\,\omega_{2t}. \qquad (8)
\end{aligned}
$$

The evolution equations (7) and (8) describe the dynamics of the (unobserved) instantaneous short rate and its volatility.

In the special case of the two-factor model (6) the coefficient $B_1(\tau, \Theta)$ is available in closed form, $B_1(\tau, \Theta) = (1 - e^{\kappa_r \tau})/\kappa_r$. For the coefficients $A(\tau)$ and $B_2(\tau)$ we solve (5) numerically using a simple Euler discretization scheme and substituting for $B_1(\cdot)$. The inference problem is complicated by the fact that at each step we need to solve for these ODEs.

We now describe our MCMC algorithm for implementing a two factor ATSM. The problem is to sample from the joint posterior distribution of the state variables $(\underline{r}, \underline{v})$ and the parameters $\Theta = (\Omega, \Lambda)$:

$$p(\underline{r}, \underline{v}, \Theta | \underline{y}) \quad \propto \quad p(\underline{y} | \underline{r}, \underline{v}, \Theta) \; p(\underline{r}, \underline{v} | \Theta) \; p(\Theta), \tag{9}$$

where $p(\Theta) = p(\Omega, \Lambda) = p(\Omega)p(\Lambda)$ and $p(\underline{r}, \underline{v}|\Theta) = p(\underline{r}, \underline{v}|\Omega)$. The likelihood function is the distribution of the increments in the bond yields

$$p(\underline{y}|\underline{r}, \underline{v}, \Theta) = \prod_{t=1}^{T} p(y_t | r_t, v_t, \Theta).$$

It is defined by the pricing equation (6), assuming independent normal errors ϵ_{it}. The prior on the state parameters $x_t = (r_t, v_t)$ is the discretized physical process model (7) and (8).

Many authors have considered the problem of direct estimation of x_t and Θ given just the time series of the short interest rates. For example, Andersen, Benzoni and Lund (2001) estimate a short-rate model with stochastic volatility and jumps; Barndorff-Nielsen and Shephard (2001) provide an excellent summary of implications of Ornstein-Uhlenbeck processes for short-rate dynamics and pricing implications. Our methodology will deal with the general case where a vector of bond yields is observed.

Finally, the approximation error ϵ_{it} is due to the Euler discretization of the the model. However, Pritsker (1998) shows that this is dwarfed by the estimation error for interest rate observations on a daily basis. The methods developed in Eraker (2001) and Elerian, Shephard and Chib (2001) are also applicable in our context.

4 MCMC Inference

4.1 Volatility Inference

For the volatility states, the complete conditional posterior, $p(\underline{v}|\underline{r}, \Omega, \Lambda, \sigma, \underline{y})$, is a SDV model with state-dependent variances in the evolution equation (8) and the observation equations. The relevant observation equation is the joint distribution of the short rate process r_t and bond prices y_t given v_t, defined in (7) and (6). The evolution and the short rate equation involve SDV, while the pricing equation is a standard linear regression. We now describe a sampling strategy for v_t in this SDV model. The strategy is explained in the specific context of the ATSM. But the method is valid for any SDV model, even including non-linear structure.

To update the vector of volatility states v we define a model augmentation with additional latent variables $\underline{z} = (z_1, \ldots, z_T)$. In this augmented model we use a Metropolis-Hastings step (Tierney, 1994) to update the volatility states v. Before we describe details of the model augmentation and the MCMC algorithm, we outline some important features of the proposed scheme.

Let $\pi(\underline{v})$ denote the posterior distribution $p(\underline{v}|\underline{z}, \underline{r}, \Omega, \Lambda, \sigma, \underline{y})$. Subject to some technical constraints only, the choice of the proposal distribution $q(\underline{\tilde{v}})$ in a Metropolis-Hasting step to update $\underline{v}$ is essentially arbitrary. An appropriate acceptance probability ensures the desired stationary distribution $\pi(\underline{\tilde{v}})$. One strategy to achieve a fast mixing Markov chain is to choose a proposal distribution which mimics the desired target distribution, $q(\underline{\tilde{v}}) \approx \pi(\underline{\tilde{v}})$. At the same time, $q(\underline{\tilde{v}})$ should allow efficient random variate generation. The art of constructing a good MCMC algorithm is to balance these two competing aims. We achieve this balance by considering a locally weighted mixture of normal linear state-space models. Introducing indicators $\underline{z}$ to break this mixture defines the already mentioned model augmentation. Conditional on $\underline{z}$ the candidate $\underline{\tilde{v}}$ is easily generated by using the forward filtering, backward sampling algorithm. Key to the efficient r.v. generation from $q(\underline{\tilde{v}})$, is the use of Gaussian kernels as weights in the locally weighted mixture. At the same time, the flexible form of the mixture model allows to achieve a good approximation $q(\underline{\tilde{v}}) \approx \pi(\underline{\tilde{v}})$.

We now describe details of the algorithm. For typographical ease we suppress the conditioning on $(\Omega, \Lambda, \sigma)$ in the notation. Let $(v^*(1), \ldots, v^*(K))$ denote a grid of v values. We start by considering an auxiliary mixture model $p^a(v_t|v_{t-1})\, p^a(r_t, y_t|v_t)$ defined by replacing (8) and (7) with mixture models

$$p^a(v_{t+1}|v_t) = \sum_{k=1}^{K} p^a(v_{t+1}|v_t, z_t = k)\, p^a(z_t = k|v_t) \tag{10}$$

and

$$p^a(r_t|v_t) = \sum_{k=1}^{K} p^a(r_t|z_t = k)\, p^a(z_t = k|v_t). \tag{11}$$

The sampling model for y_t remains unchanged, i.e., $p^a(y_t|v_t) = p(y_t|v_t)$ as in (6). We define $p^a(v_{t+1}|v_t, z_t = k)$ and $p^a(r_t|z_t = k)$ by substitute the grid value $\sqrt{v^*(k)}$ for the scale factor $\sqrt{v_t}$ in (7) and (8). Specifically, $p^a(v_{t+1}|v_t, z_t = k)$ is a normal distribution with expectation as in (8) and variance $\sigma_v^2\, v^*(k)$, and $p^a(r_t|z_t = k)$ is a normal distribution with mean as in (7) and variance $v^*(k)$. Let $\phi(x; m, s)$ denote a normal density with moments (m, s^2), evaluated at x. For the weight function $p^a(z_t|v_t)$ we use

$$p^a(z_t = k|v_t) = \phi(v_t; v^*(k), \sigma^*)/c(v_t) \tag{12}$$

with denominator $c(v_t) = \sum_{k=1}^K \phi(v_t; v^*(k), \sigma^*)$. The chosen weight function favors grid values $v^*(k)$ close to the currently imputed volatility v_t, using smooth interpolation between grid points. The mixture model $p^a(\cdot)$ motivates the following MCMC scheme. In words, we first augment the probability model $\pi(\underline{v})$ by drawing indicators $\underline{z} \sim p^a(\underline{z}|\underline{v})$. Then we propose new volatility states $\underline{\tilde{v}}$. The proposal distribution $q(\underline{\tilde{v}})$ is the conditional distribution $p^a(\underline{v}|\underline{z}, \underline{y}, \underline{r})$ in the auxiliary model, ignoring the denominator $c(v_t)$ in (12). Generating $\underline{\tilde{v}}$ is greatly simplified by the fact that $q(\underline{\tilde{v}})$ takes the form of a normal dynamic linear model. Finally, with the appropriate Metropolis-Hasting acceptance probability the proposal $\underline{\tilde{v}}$ is accepted as new volatility state. Formally the MCMC simulation is defined by the following three steps.

1. Generate mixture indicators $\underline{z} = (z_1, \ldots, z_T)$ from $p^a(\underline{z}|\underline{v})$. Given $\underline{v}$ the indicator variables $z_1, \ldots, z_T$ are conditionally independent, and can be sampled independently from multinomial distributions with probabilities $p^a(z_t = k|v_t, r_t, y_t) \propto p^a(r_t|z_t = k, v_t)\, p^a(z_t = k|v_t)$.

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

$$p^a(\underline{v}|\underline{z}, \underline{y}, \underline{r}) \;\propto\; p(v_0) \prod_{t=1}^{T} p^a(v_t|v_{t-1}, z_{t-1})\, p(y_t|v_t)\, p^a(z_t|v_t).$$

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. Substituting (12) for $p^a(z_t|v_t)$ we get

$$p^a(\underline{v}|\underline{z}, \underline{y}) \;\propto\; \underbrace{p(v_0) \prod_{t=1}^{T} p(v_t|v_{t-1}, z_{t-1})\; p(y_t|v_t)\; \phi(v_t;\; v^*(z_t), \sigma^{*2})}_{q(\underline{v})}\; \frac{1}{c(v_t)}. \tag{13}$$

The first factor, denoted by $q(\underline{v})$ in (13), corresponds to the smoothing distribution in another linear, Gaussian state space model. This function $q(\underline{v})$ will serve as proposal distribution in a Metropolis-Hastings step. The importance of augmenting our model by a mixture of linear regressions with Gaussian mixture 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 $\underline{\tilde{v}}$ of the state vector, $\underline{\tilde{v}} \sim q(\underline{\tilde{v}})$. The algorithm is known as forward filtering, backward sampling (FFBS). Let $1/\tilde{s}_t^2 = 1/\sigma^2 \sum_{i=1}^n 1/\tau_i^2 + 1/\sigma^{*2}$ and

$$\tilde{y}_t = \tilde{s}^2 \left\{ \frac{1}{\sigma^2} \sum_{i=1}^n \left[\tau_i y_{it} + A(\tau_i, \Theta) - B_1(\tau_i, \Theta) v_t - B_2(\tau_i, \Theta) v_t\right] / \tau_i + \frac{1}{\sigma^{*2}} v^*(z_t) \right\}$$

The normal, linear state-space model

$$\begin{aligned} \tilde{y}_t &= v_t + \tilde{s}_t \, \epsilon_t \\ v_{t+1} &= v_t + \kappa_v(\theta_v - v_t) + \sigma_v \sqrt{v^*(z_t)} \, \omega_t \end{aligned}$$

has a smoothing distribution equal to $q(\underline{v})$. The full smoothing distribution $q(\underline{v})$ of this model can be sampled directly using FFBS. We use it to generate a proposal for the state variables $\underline{\tilde{v}} \sim q(\underline{\tilde{v}})$.

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

$$a(\underline{v}, \underline{\tilde{v}}) = \min \left\{ 1, \prod_{t=1}^T \frac{p(r_t|\tilde{v}_t)}{c(\tilde{v}_t)\; p^a(r_t|\tilde{v}_t)} \frac{c(v_t)\; p^a(r_t|v_t)}{p(r_t|v_t)} \right\}. \tag{14}$$

With probability $a(\underline{v}, \underline{\tilde{v}})$ replace the currently imputed state parameters $\underline{v}$ by $\underline{\tilde{v}}$. Otherwise discard the proposal $\underline{\tilde{v}}$ and leave $\underline{v}$ unchanged.

Use of the acceptance probability $a(\underline{v}, \underline{\tilde{v}})$ ensures an ergodic distribution equal to $p(\underline{v}|\underline{r}, \underline{y})$, as desired. This is seen by considering an augmentation of $p(\underline{v}|\underline{r}, \underline{y})$ to $p(\underline{v}, \underline{z}|\underline{r}, \underline{y}) \equiv p(\underline{v}|\underline{r}, \underline{y}) \cdot\, p^a(\underline{z}|\underline{v}, \underline{r}, \underline{y})$, i.e., add z to the probability model $p(\underline{v}|\underline{r}, \underline{y})$ by defining the conditional distribution for $\underline{z}$ given $\underline{v}$ and $(\underline{r}, \underline{y})$ as in model $p^a(\cdot)$. Steps 1 through 3 define a Markov chain with ergodic distribution $p(\underline{v}, \underline{z}|\underline{r}, \underline{y})$. Step 1 replaces $\underline{z}$ by sampling from the complete conditional distribution $p^a(\underline{z}|\underline{v}, \underline{r}, \underline{y}) = p^a(\underline{z}|\underline{v}, \underline{r}, \underline{y})$. Step 2 generates a Metropolis-Hastings proposal $\underline{\tilde{v}} \sim q(\underline{\tilde{v}})$. Step 3 accepts the proposal with the correct Metropolis-Hastings acceptance probability $\min\left\{1, p(\underline{\tilde{v}}|\underline{z}, \underline{r}, \underline{y})\, q(\underline{v}) / [p(\underline{v}|\underline{z}, \underline{r}, \underline{y})\, q(\underline{\tilde{v}})]\right\}$. To verify expression (14) note that

$$p(\underline{v}|\underline{z}, \underline{r}, \underline{y}) \propto p(\underline{v}|\underline{r}, \underline{y})\, p^a(\underline{z}|\underline{v}, \underline{r}, \underline{y})$$

$$p(\underline{v}|\underline{z}, \underline{r}, \underline{y}) \propto p(v_0) \prod_{t=1}^{T} p(v_t|v_{t-1})\ p(r_t, y_t|v_t)\ \frac{p^a(r_t, y_t|z_t, v_t)\phi(v_t;\ \mu_{(t)}, \sigma^2_{(t)})}{\sum_{k=1}^{K} p^a(r_t, y_t|z_t = k, v_t)\phi(v_t;\ \mu_k, \sigma^2_k)}.$$

This completes the specification of our algorithm for updating the unobserved short-rate volatility states.

4.2 Instantaneous Short Rate and Parameter Inference

In this section, we describe the complete conditional posterior distributions for the unobserved short-term interest rate; the market price of risk parameters; the parameters of the physical process evolution; and the pricing error variance. These conditional distributions are used to complete the specification of the our MCMC algorithm.

First, the complete conditional posterior distribution of the unobserved short rate follows a heteroscedastic normal linear state-space model with evolution equation (7) and observation equation (6). We simulate from $p(r|\underline{v}, \Theta, \underline{y})$ using the FFBS algorithm of Carter and Kohn (1994) and Frühwirth-Schnatter (1994)

$$p(\underline{r}|\underline{v}, \Theta, \underline{y}) = p(r_T|\underline{v}, \Theta, \underline{y}) \prod_{t=1}^{T} p(r_t|r_{t+1}, \underline{v}, \Theta, \underline{y}).$$

Both the system equation (7) and the bond pricing equation (6) are linear in r_t. Thus we can run the Kalman filter to get the moments of $p(r_T|\underline{v}, \Theta, \underline{y})$. Also, linearity and normality of evolution and observation equation imply that $p(r_T|\underline{v}, \Theta, \underline{y})$ is normal. Then we proceed recursively backwards through time to sample $p(r_t|r^{(g)}_{t+1}, \underline{v}, \Theta, \underline{y})$.

Secondly, let $\pi(\Lambda) = p(\Lambda|\Theta, \sigma^2, \underline{x}, \underline{y})$ denote the conditional posterior distribution of market price of risk parameters. Since the market price of risk parameter Λ appears only in the observation equation we find

$$\pi(\Lambda)\ \propto\ p(\underline{y}|\underline{x}, \Theta, \sigma^2, \Lambda)\ p(\Lambda).$$

with

$$p(\underline{y}|\underline{x}, \Theta, \sigma^2, \Lambda)\ \propto\ \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - A(\tau, \Lambda) - B_2(\tau, \Lambda)x_t)^2\right\}$$

where $A(\tau)$ and $B_2(\tau)$ are solutions to the ODE defined by (5). We use a random-walk Metropolis step to update $\Lambda = (\lambda_r, \lambda_v)$, generating a proposal $\tilde{\Lambda} \sim N(\Lambda, \tau^2 I)$, with τ^2 calibrated to achieve an acceptance rate between 20% and 40%.

Thirdly, let $\pi(\Theta) = p(\Theta|\Lambda, \sigma^2, \underline{x}, \underline{y})$ denote the complete conditional of the physical process parameters. Since Θ appears in the evolution equation and the observation we find

$$\pi(\Theta) \ \propto \ p(\underline{y}|\underline{x}, \Theta, \sigma^2, \Lambda) \, p(\underline{x}|\Theta) \, p(\Theta).$$

Now, let $q(\theta) \propto p(\underline{x}|\Theta)\, p(\Theta)$, which is known in closed form from standard conjugate theory. Hence, we can use a Metropolis-Hastings algorithm with $q(\Theta)$ as the proposal.

Finally, there are a number of ways of dealing with the pricing error standard deviation σ. The observation equation assumes that $y_{it} = F(\tau_i, \Theta)x_t + \sigma\epsilon_{it}$. Here σ represents a pricing error and for simplicity we assume that the pricing errors are independent. Generalization to correlated errors is straightforward. One approach is to set σ to a predetermined level, which may be representative of the bid/ask spread for a particular price. An alternative is to perform inference on σ^2. To do this, let the conditional posterior distribution of the pricing error variance be denoted by $\pi(\sigma^2) = p(\sigma^2|\Theta, \Lambda, \underline{x}, \underline{y})$. Since σ^2 appears only in the pricing equation, we find

$$\pi(\sigma^2) \ \propto \ p(\underline{y}|\underline{x}, \Theta, \sigma^2, \Lambda) \, p(\sigma^2).$$

Assuming a conjugate inverse-gamma prior, $p(\sigma^2) = IG(a/2, b/2)$, the conditional posterior is also inverse-gamma $\pi(\sigma^2) = IG(a_1/2, b_1/2)$ with parameters $a_1 = a + T$ and $b_1 = b + \sum(y_t - A(\tau) + B_2(\tau)x_t)^2$. This completes the specification of the conditional posterior distributions needed for our algorithm.

5 Application: US Bond Yields

5.1 Yield Curve Modeling: US Interest Rates, 1996-1999

The main application of ATSM models is in the modeling of the term structure of interest rates. See Duffie and Singleton (1999) and Andersen et al. (2001) for recent applications. The data consists of a multivariate panel of daily bond yields. Let $y_{\tau,t}$ denote the bond yield for maturity τ at time t. Our panel includes maturities of $\tau_i \in \{1, 2, 5, 10, 30\}$ years. Figure 1 the time series plots of the five yields, along with the 1-month and 6-month yields from 1974 to 1999.

Figure 2 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 October 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%.

Figures 1 and 2 about here

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 (7) and (8) 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 $\mathcal{K}, \theta, V$, and the volatilities v_t. Using these initial values we fit the two-factor ATSM described in Section 5.1 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 $\mathcal{K}, \theta, V$ 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 turn to our estimation and smoothing results for (v_t, r_t). Figure 3 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 3 about here

Figure 4 plots the posterior distributions for the parameters of the volatility evolution (8) 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 σ_v does not match under the two models.

Figure 4 about here

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 σ_v towards larger values to accommodate these jumps. Secondly, the ATSM includes a pricing error σ, 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 σ_v would tend towards higher values.

Table 1 about here

Figure 5 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 October 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 October 1998.

Figure 5 about here

Figure 6 plots the posterior distribution for the market price of risk parameters $\Lambda = (\lambda_r, \lambda_v)$ and the pricing error parameter σ. The market risk parameters have the expected signs: λ_r is negative, and λ_v is positive. The posterior mean of the pricing error is 0.07%, or 7 basis points.

Figure 6 about here

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.

6 Conclusions

In this paper, we develop a general likelihood-based approach for inference in affine state dependent variance (ASDV) models. We apply the methodology to interest rate modeling and the class of affine term structure models (ATSMs). These models are becoming commonplace in financial time-series applications and in pricing of derivative securties. Our simulation-based methodology provides samples from the full smoothing distribution of the unobserved states x_t 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.

One area for future research is implementing this methodology to models that incorporate jumps in the state variables. Liechty and Roberts (2001) provide a methodology for estimating such models in continuous time. Johannes and Polson (2001) provide a discrete-time version and show how the pricing equation is affected by the addition of jumps and regime switches in the state variables. Finally, on-line filtering methods such as those developed in Müller, Polson and Stroud (2001) can also be applied in this setting.

References

Aït-Sahalia, Y. (1996) Nonparametric pricing of interest rate derivative securities. *Econometrica*, **64**, 527–560.

Andersen, T., Benzoni, L. and Lund, J. (2001) An empirical investigation of continuous-time equity return models. *Tech. rep.*, National Bureau of Economic Research.

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.

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.

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.

Duffie, D. and Singleton, K. (1999) Modeling term structure models of defaultable bonds. *Review of Financial Studies*, **12**, 687–720.

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.

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

Heston, S. (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, **6**, 327–343.

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

Johannes, M. and Polson, N. (2001) MCMC methods for financial econometrics. *Tech. rep.*, Graduate School of Business, University of Chicago.

Liechty, J. and Roberts, G. (2001) Markov chain Monte Carlo methods for switching diffusion models. *Biometrika*, **88**, 299–315.

Liu, J. and West, M. (1999) Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice* (eds. A. Doucet, J. de Freitas and M. Niranjan). Kluwer.

Müller, P., Polson, N. and Stroud, J. (2001) Practical filtering for nonlinear state-space models. *Tech. rep.*, Department of Statistics, Wharton School, University of Pennsylvania.

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.

Singleton, K. (2001) Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, **102**, 111–141.

Stroud, J., Müller, P. and Polson, N. (2001) Nonlinear state-space models with state-dependent variances. *Tech. rep.*, Institute of Statistics and Decision Sciences, Duke University.

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

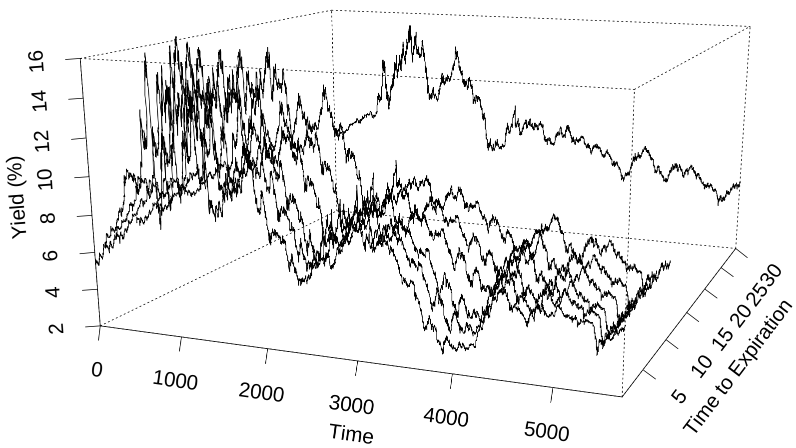

Figure 1: U.S. constant maturity bond yields (October 1995 – October 1999)

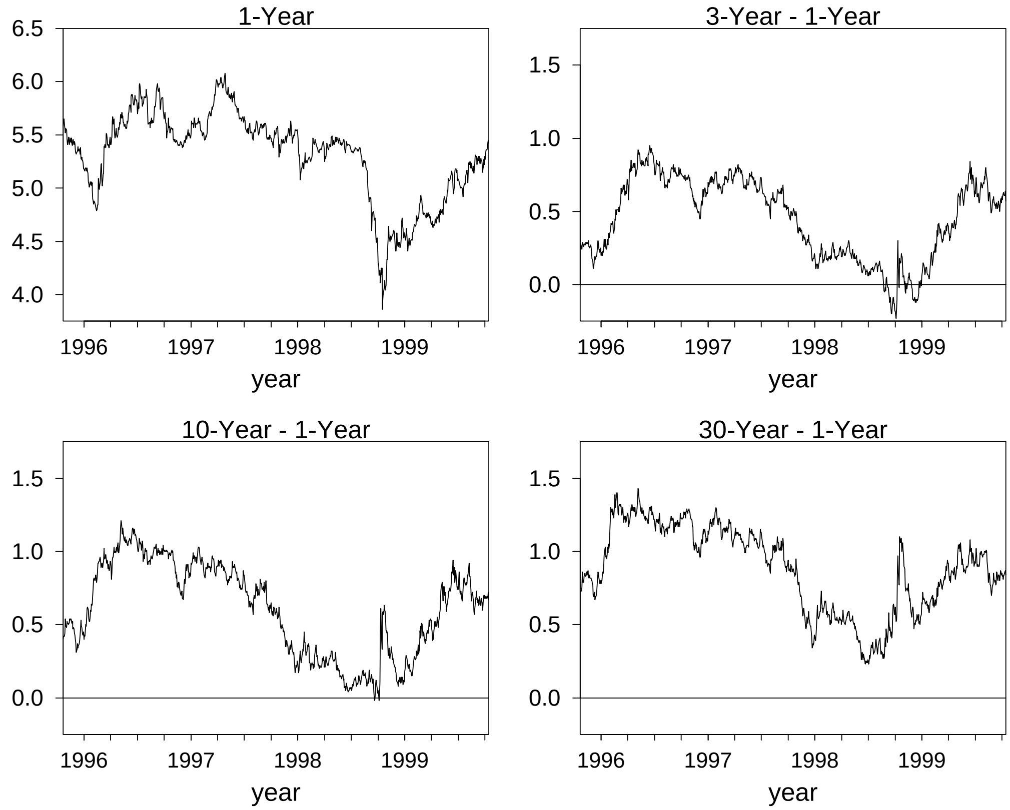

Figure 2: U.S. constant maturity bond yields (1995–1999). 1-year (top left); difference between 5-year and 1-year yields (top right); difference between 10-year and 1-year yields (bottom left); difference between 30-year and 1-year yields (bottom left). Notice the sharp drop in October 1998.

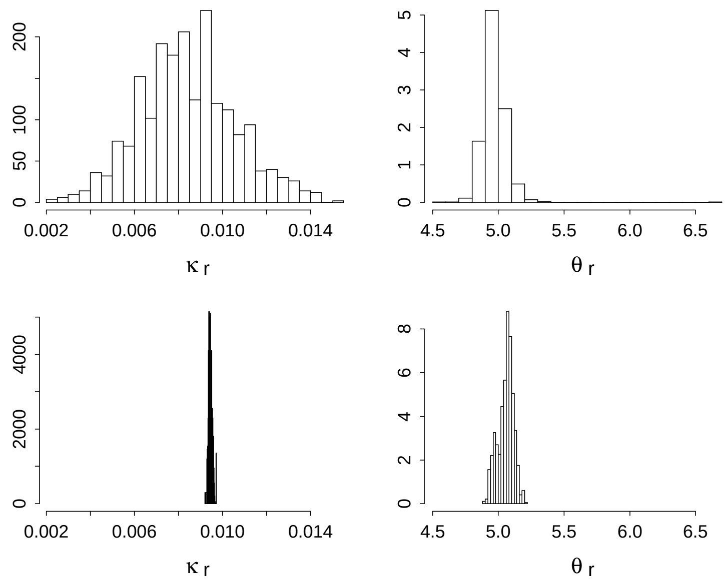

Figure 3: Posterior distributions of (κ_r, θ_r) under the short-rate SDV model (top), and the ATSM for daily yields (bottom). Put simply, compared to the uncertainties in the analysis of the observed short-rate time series only, the market has practically perfect knowledge of the short-rate dynamics. Here, the market is represented by the daily bond yields.

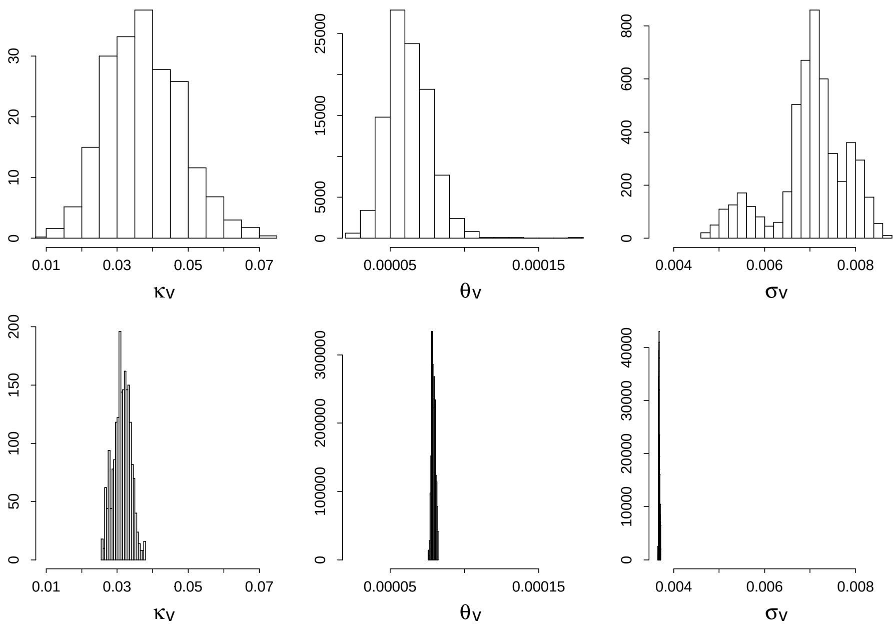

Figure 4: Posterior distributions of $(\kappa_v, \theta_v, \sigma_v)$ under the short-rate SDV model (top), and the ATSM (bottom).

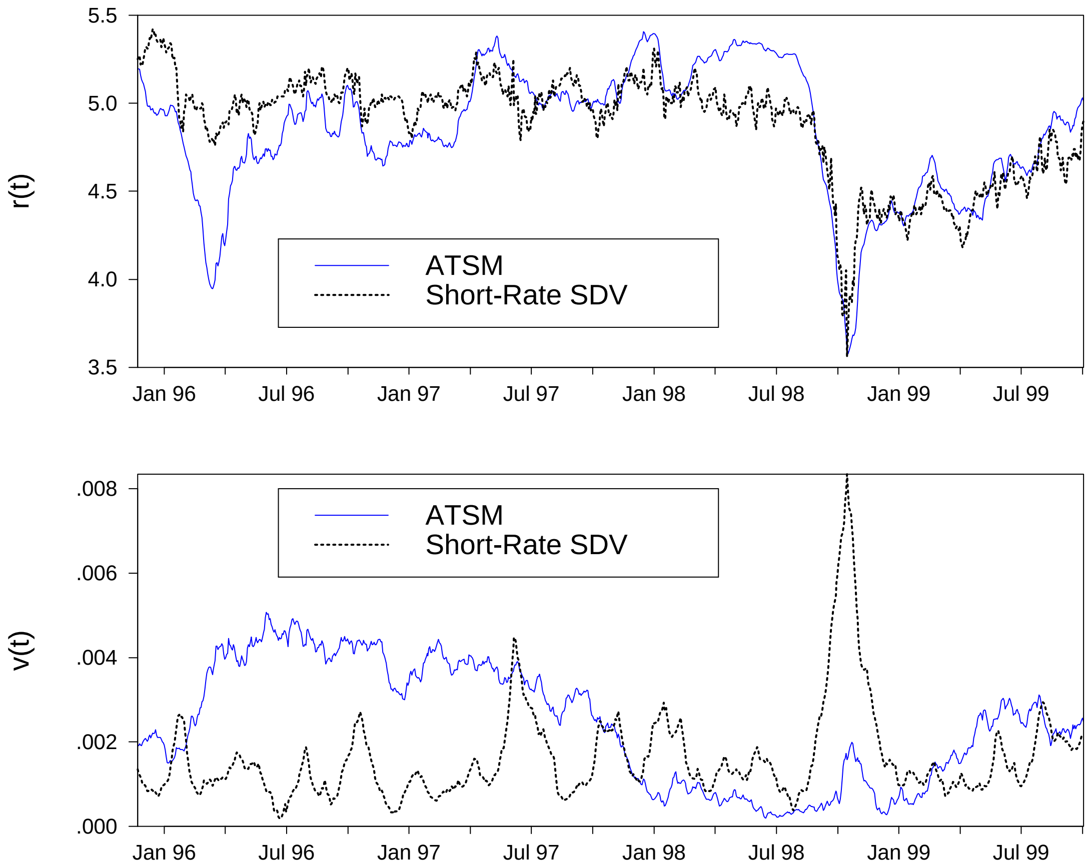

Figure 5: Top: Smoothed state variable r_t versus observed 3-month T-Bills $\hat{r}_t$. Bottom: Smoothed state volatilities $E(v_t|\underline{y})$ versus estimated volatility from the short-rate SDV model.

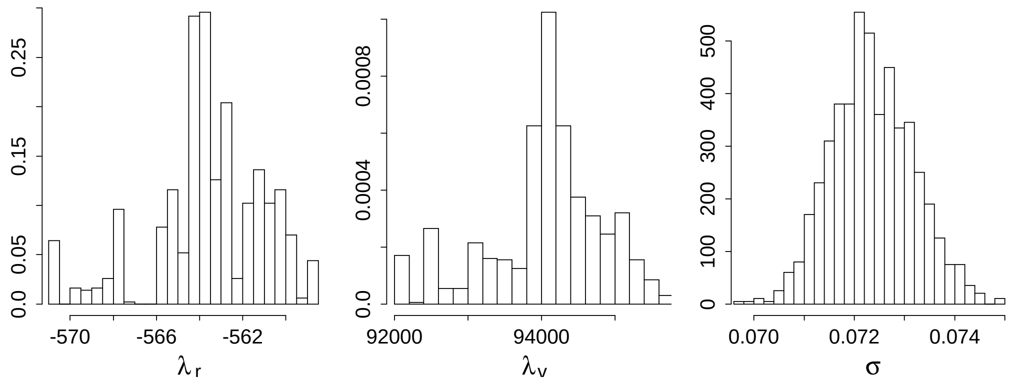

Figure 6: Posterior distribution of market price of risk and pricing error parameters.

Parameter	ATSM		Short-rate SDV	
λ_r	564	(2.0)	——	——
λ_v	94200	(1000)	——	——
σ	.072	(.001)	——	——
κ_r	.0095	(.0002)	.0082	(.002)
θ_r	5.05	(.110)	4.95	(.155)
κ_v	.033	(.005)	.040	(.015)
θ_v	.000075	(.000003)	.00006	(.00002)
σ_v	.0036	(.0001)	.007	(.001)

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