ISSN 1440-771X

AUSTRALIA

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Simulation-Based Bayesian Estimation of Affine Term Structure Models

Andrew D. Sanford and Gael M. Martin

Working Paper 15/2003

Simulation-Based Bayesian Estimation of Affine Term Structure Models*

Andrew D. Sanford† and Gael M. Martin‡

September 19, 2003

Abstract

This paper demonstrates the application of Bayesian simulation-based estimation to a class of interest rate models known as Affine Term Structure (ATS) models. The technique used is based on a Markov Chain Monte Carlo algorithm, with the discrete observations on yields augmented by additional higher frequency latent data. The introduction of augmented yield data reduces the bias associated with estimating a continuous time model using discretely observed data. The technique is demonstrated using a one-factor ATS model, with the latent factor process that underlies the yields sampled via a single-move algorithm. Numerical application of the method is demonstrated using both simulated and empirical data. Extension of the method to a three-factor ATS model is also discussed, as well as the application of a multi-move sampler based on a Kalman Filtering and Smoothing algorithm.

Keywords: Interest Rate Models, Markov Chain Monte Carlo, Data Augmentation, Nonlinear State Space Models, Kalman Filtering.

JEL Classifications: C11, C15, E43.

*This research has been supported by an Australian Research Council Discovery Grant, No. DP0208333, and a Ph.D. Scholarship from the School of Business Systems, Monash University.

†School of Business Systems, Monash University. Corresponding author, E-Mail address: Andrew.Sanford@Infotech.monash.edu.au. Postal Address: School of Business Systems, Faculty of Information Technology, Monash University, Clayton Campus, Wellington Road, Clayton, Victoria 3800, Australia.

‡Department of Econometrics and Business Statistics, Monash University.

1 Introduction

In this paper we apply Bayesian simulation techniques to the estimation of Affine Term Structure (ATS) models. These models have become popular because of their tractability in asset and derivative pricing. A comprehensive specification framework for ATS models has been developed by Dai and Singleton (2000), under which all variants can be classified. Within this framework, parameter restrictions sufficient to ensure model stability and econometric identification are also described. We adopt the Dai and Singleton (DS) framework in this paper.

Two of the main difficulties associated with estimating ATS models are 1) the presence of unobservable latent factors; and 2) the bias introduced when estimating the parameters of the continuous-time specification with data observed only at discrete intervals. We overcome both problems by adopting a Bayesian approach to estimation, based on Markov Chain Monte Carlo (MCMC) simulation. The bias in estimation is reduced by augmenting the set of unknowns by higher frequency latent data placed at finite intervals between the observed yields. This aspect of the methodology extends into the term structure framework, the methods developed independently by Jones (1998), Elerian, Chib, and Shephard (2001) and Eraker (2001).

Bayesian approaches to estimating term structure models have been adopted by Frühwirth-Schnatter and Geyer (1998), Mikkelsen (2002) and Polson, Stroud and Müller (2002). None of these analyses however use data augmentation to approximate the underlying continuous time latent factor process. Frühwirth-Schnatter and Geyer and Mikkelsen use more restricted models for the factor process, such that exact solutions are produced, whilst Polson, Stroud and Müller adopt an Euler discretization of the continuous time process without adjustment for the induced bias.

The remainder of the paper proceeds as follows. Section 2 describes the general ATS specification that is adopted in the paper. Section 3 looks specifically at the one-factor ATS model and provides a discrete approximation to that model. Section 4 outlines the estimation methodology, including details of both the data augmentation and the algorithm used to sample from the joint posterior of the full set of unknowns. Section 5 reports the estimation results based both on yield data simulated from the assumed model and on empirical data on the Australian yield curve. Extension of the methodology to a three-factor ATS model is discussed in Section 6. In Section 7, a modification of the algorithm to incorporate multi-

move sampling of the latent state vector, based on the application of a Kalman Filtering and Smoothing algorithm to a state-space representation of the model, is discussed. Conclusions are provided in Section 8.

2 Affine Term Structure Models

In this section we describe ATS models using the general DS framework. ATS models assume that the instantaneous short rate at time t, r_t, is a linear function of an $(N \times 1)$ vector of latent factors, $X_t = (X_{1,t}, ..., X_{N,t})'$. This function is described by the following linear equation,

$$r_t = \delta_0 + \delta_X' X_t, \tag{1}$$

where δ_0 is a scalar and $\delta_X = (\delta_1, ..., \delta_N)'$ is an $(N \times 1)$ vector of coefficients relating r_t to the factors at time t. The value of r_t is effectively the return on an asset over an infinitesimally short time period, and is therefore unobservable. The dynamics of r_t are determined by the underlying X_t process, which in the ATS framework is assumed to be

$$dX_t = K^Q \left(\Theta^Q - X_t\right) dt + \Sigma \sqrt{S_t} dW_t^Q. \tag{2}$$

The process in (2) comprises a drift component, $K^Q \left(\Theta^Q - X_t\right) dt$, and a diffusion component, $\Sigma\sqrt{S_t}dW_t^Q$. The superscript Q indicates that the parameters are defined under the risk neutral measure. The risk neutral probability measure ensures that the future price of an asset, discounted by the risk free rate, will equate to its current price. The model parameters consist of K^Q, an $(N \times N)$ matrix of mean reversion parameters, Σ, an $(N \times N)$ scaling matrix for the diffusion component, and Θ^Q, an $(N \times 1)$ vector of the long-term means of the N latent factor processes. Both K^Q and Σ may be nondiagonal and asymmetric. The matrix S_t is an $(N \times N)$ diagonal matrix with the ith diagonal element given by

$$[S_t]_{ii} = \alpha_i + \beta_i' X_t, \tag{3}$$

where the parameter α_i is a scalar and β_i is an $(N \times 1)$ vector, $i = 1, 2, \ldots, N$. The components of β_i determine which of the latent factors in X_t influence the conditional variance of the N-factor process. Finally, dW_t^Q in (2) represents the instantaneous change in an N-dimensional independent standard Brownian motion. As can be seen from equations (2) and (3), the drift and volatility are affine in X_t.

The price of a zero coupon bond at time t that pays \$1 at time $t+\tau$ is defined as,

$$P(\tau, r_t) = E^Q\left(e^{-\int_t^{t+\tau} r_s ds}\right). \tag{4}$$

According to (4) the price of a zero coupon bond at time t is the expected value under a risk neutral probability measure of the value of \$1 discounted by the risk free instantaneous rate between t and $t+\tau$. For the ATS models, the price of a zero coupon bond at time t maturing at time $t+\tau$ and paying \$1 is related to the latent factors X_t by the following pricing equation,

$$P(\tau, X_t) = \exp\left(A(\tau) - B(\tau)' X_t\right), \tag{5}$$

where $A(\tau)$ is a scalar and $B(\tau) = (B_1(\tau), B_2(\tau), \ldots, B_N(\tau))'$ is an $(N \times 1)$ vector of coefficients. The functions $A(\tau)$ and $B(\tau)$ represent the solutions to two ordinary differential equations that are functions of the parameters of the risk neutral latent factor process (2) and the instantaneous short rate process (1). These equations are of the form,

$$\frac{dA(\tau)}{d\tau} = -\Theta^{'Q} K^{'Q} B(\tau) + \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma' B(\tau)\right]_i^2 \alpha_i - \delta_0. \tag{6}$$

$$\frac{dB(\tau)}{d\tau} = -K^{'Q} B(\tau) - \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma' B(\tau)\right]_i^2 \beta_i + \delta_X, \tag{7}$$

with initial conditions $A(0) = 0$ and $B_i(0) = 0$ for all $i = 1, 2, \ldots, N$. In some specific examples of ATS models, equations (6) and (7) have analytical solutions. However, numerical methods are needed in most cases to solve for $A(\tau)$ and $B(\tau)$; see Dai and Singleton (2000) and Bolder (2001) for further details. We employ a numerical solution in the case of the model that is used in this paper.

In this paper, we do not model the bond prices but rather their yields. Zero coupon bond prices and their continuously compounded yields are related by the identity,

$$P(\tau, X_t)\exp(\tau y(\tau, X_t)) = 1, \tag{8}$$

where $y(\tau, X_t)$ is the yield at time t of a zero coupon bond priced at $P(\tau, X_t)$ that pays \$1 at time $t+\tau$. Rearranging (8) as

$$y(\tau, X_t) = -\frac{1}{\tau}\ln P(\tau, X_t), \tag{9}$$

and substituting (5) into (9), results in the following linear relation between the yield and the latent factors X_t, for a given term to maturity τ,

$$y(\tau, X_t) = -\frac{1}{\tau}\left(A(\tau) - B'(\tau) X_t\right). \tag{10}$$

The link between the risk-neutral process in (2) and the physical process, which captures the actual time series dynamics of the latent factor process, occurs via incorporation of the market price of risk, Λ_t. The form of the market price of risk adopted by DS is given by the $(N \times 1)$ vector

$$\Lambda_t = \sqrt{S_t}\lambda, \tag{11}$$

where λ is an $(N \times 1)$ vector of parameters,

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix}. \tag{12}$$

The vector Λ_t is an adjustment made to the drift of the physical process to ensure that bonds of varying term to maturity, combined in a portfolio whose instantaneous payoff is known with certainty, has a return that is equivalent to the riskless return. This is therefore an environment where arbitrage opportunities are excluded. In equation (11), λ is a vector of static parameters that must also be estimated. The physical and risk neutral Wiener processes, dW_t and dW_t^Q respectively, are related as (see Maes, 2003),

$$dW_t^Q = dW_t + \Lambda_t dt. \tag{13}$$

Modifying the representation of the risk-neutral process as shown in equation (2) to make explicit the market price of risk, we use equation (13) to produce,

$$dX_t = K^Q\left(\Theta^Q - X_t\right)dt + \Sigma\sqrt{S_t}\Lambda_t dt + \Sigma\sqrt{S_t}dW_t, \tag{14}$$

which then yields the physical process as,

$$dX_t = K\left(\Theta - X_t\right)dt + \Sigma\sqrt{S_t}dW_t. \tag{15}$$

The physical mean reversion parameter, K, and physical long term mean, Θ, are related to the risk-neutral parameters by the relationships,

$$K = K^Q - \Sigma\Phi \tag{16}$$

and

$$K\Theta = K^Q\Theta^Q + \Sigma\Psi, \tag{17}$$

where Φ is an $(N \times N)$ matrix given by

$$\Phi = \begin{bmatrix} \lambda_1\beta_1' \\ \vdots \\ \lambda_i\beta_i' \\ \vdots \\ \lambda_N\beta_N' \end{bmatrix} \tag{18}$$

and Ψ is an $(N \times 1)$ vector given by

$$\Psi = \begin{bmatrix} \lambda_1\alpha_1 \\ \vdots \\ \lambda_i\alpha_i \\ \vdots \\ \lambda_N\alpha_N \end{bmatrix}. \tag{19}$$

Here λ_i is the ith element of λ in (12) and α_i and β_i are as defined in (3). The complete derivation of K and $K\Theta$ is provided in the Appendix.[1]

Adopting the DS terminology, $A_m(N)$ identifies an N-factor ATS model in which m of the latent factors determine the covariance of the latent process via the diffusion component S_t. The latent factors are partitioned as $X_t' = \left(X_t^{C\prime}, X_t^{D'}\right)$, where X_t^C is the $(m \times 1)$ vector of factors at time t that are included in the diffusion component of the model, whilst X_t^D is the $((N-m) \times 1)$ vector of factors at time t that do not enter the diffusion component. For $m > 0$, the mean reversion and long term mean of the $A_m(N)$ model are partitioned as follows,

$$\Theta = \begin{bmatrix} \Theta^C_{m\times 1} \\ 0_{(N-m)\times 1} \end{bmatrix} \tag{20}$$

and

$$K = \begin{bmatrix} K^{CC}_{m\times m} & 0_{m\times(N-m)} \\ K^{CD}_{(N-m)\times m} & K^{DD}_{(N-m)\times(N-m)} \end{bmatrix}. \tag{21}$$

The mean reversion matrix K, is partitioned into four components. The sub-matrix K^{CC} identifies those mean reversion parameters that relate to X_t^C, and that appear in the stochas-

[1] We note that more flexible forms for Λ_t have been suggested in the recent ATS model literature, including the 'Semi-Affine' models of Duarte (2000), and the 'Essentially Affine' models of Duffee (2000). However we retain the simpler Λ_t specification of DS to limit computational complexities, with the generalization of the representation in (11) left for future research.

tic differential equations that describe the dynamics of X_t^C. The sub-matrix K^{DD} identifies those mean reversion parameters related to X_t^D, that appear in the stochastic differential equations that describe the dynamics of X_t^D. Finally, the sub-matrix K^{CD}, identifies those mean reversion parameters related to X_t^C, when X_t^C appears in the stochastic differential equations that describe the dynamics of X_t^D. The sub-vector Θ^C contains the long term means for the components X_t^C in the stochastic differential equations that describe the dynamics of X_t^C. The zero restrictions in both (20) and (21) are imposed to ensure econometric identification of the unrestricted elements of Θ and K, as described in the DS paper.

Other parameters of the latent factor process are also restricted to ensure identification. Defining the $(N \times 1)$ vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)'$ and the $(N \times N)$ matrix $\Gamma = (\beta_1, \beta_2, \ldots, \beta_N)'$, we specify that

$$\alpha = \begin{bmatrix} 0_{m\times 1} \\ 1_{(N-m)\times 1} \end{bmatrix} \tag{22}$$

and

$$\Gamma = \begin{bmatrix} I_{m\times m} & \beta^{CD}_{m\times(N-m)} \\ 0_{(N-m)\times m} & 0_{(N-m)\times(N-m)} \end{bmatrix}, \tag{23}$$

where

$$\beta^{CD} = \begin{bmatrix} \beta^{CD}_{m+1} & \cdots & \beta^{CD}_{N} \end{bmatrix}. \tag{24}$$

Each $\beta_j^{CD} \geq 0$, $j = m+1, m+2, \ldots, N$, in (24) is an $(m \times 1)$ vector representing the parameters of the latent factors X_t^C that enter the diffusion components of the processes for X_t^D. We also impose the restriction that $\Sigma = I_N$. Other restrictions are expressed in terms of the following inequalities,

$$\delta_i \geq 0 \text{ where } m+1 \leq i \leq N, \tag{25}$$

$$K_i\Theta = \sum_{j=1}^{m} K_{i,j}\Theta_j > 0 \text{ where } 1 \leq i \leq m, \tag{26}$$

$$K_{i,j} \leq 0 \text{ where } 1 \leq j \leq m \text{ and } j \neq i, \tag{27}$$

$$\Theta_i \geq 0 \text{ where } 1 \leq i \leq m. \tag{28}$$

The restrictions (25), (26), (27) and (28) are defined in terms of parameters that relate to X_t^C and those that relate to X_t^D, by reference to the relevant indices, where the X_t^C factors

correspond to the indices 1 to m inclusive, and the X_t^D factors correspond to the indices $m+1$ to N inclusive. To illustrate, restriction (25) states that the factors X_t^D must have coefficients δ_i that are strictly non-negative in the instantaneous short rate process as shown in equation (1). The restriction (26) requires that the product of the mean reversion and long term mean parameters should be strictly positive, but only in the case of the parameters that relate to the X_t^C factors. Restriction (27) requires that all off-diagonal components of the mean reversion matrix should be less than or equal to zero for the X_t^C factors.[2]

3 Model Specification

3.1 Discussion

In this paper we demonstrate the application of Bayesian simulation techniques, including the techniques of data augmentation, to term structure models of the ATS form. We begin by demonstrating the methodology with reference to a one-factor model, denoted by $A_1(1)$, in which the single factor represents the instantaneous short rate. The $A_1(1)$ model is detailed in this section. In Section 4 we then outline the simulation algorithm that is applied to the model, in which the latent factor vector is sampled one element at a time. Since the extended algorithm presented in Section 7, based on the multi-move sampling of the latent vector, exploits the nonlinear state-space structure of the ATS model, we use that format to present the $A_1(1)$ model. Accordingly, the yield equation in (10) is viewed as the expected value of the observation equation of the state-space model, and the physical latent factor equation viewed as the system, or state equation; see also Mikkelsen (2002) and Polson, Stroud and Müller (2002).

3.2 State Equation

The physical latent factor process of the $A_1(1)$ model, where $m = 1$ and $N = 1$, has the following form,

$$dX_t = (\mu - kX_t)\,dt + \sqrt{X_t}dW_t. \tag{29}$$

The parameters of the physical process in (29) are re-defined as the mean reversion parameter k and parameter $\mu = k\theta$, where the scalar θ represents the long-run mean of X_t. With reference to (22), (29) implies that $\alpha_1 = 0$ and $\beta_1' = 1$. We approximate the continuous-time

[2] See Dai and Singleton (2000) for more motivation of the restrictions outlined here.

specification in (29) using an Euler scheme,

$$X_{t+\Delta} = \mu\Delta + (1 - k\Delta) X_t + \sqrt{X_t \Delta}\varepsilon_{t+\Delta}. \tag{30}$$

Under this discrete time approximation, the time between observations is denoted by Δ and the Wiener increments dW_t in (29) are approximated by multiplying a standard normal innovation term $\varepsilon_t \sim N(0,1)$ by $\sqrt{\Delta}$ to give a random variable $\varepsilon_t\sqrt{\Delta} \sim N(0,\Delta)$. The system equation in (30) approximates that in (29) for sufficiently small Δ. The $(T \times 1)$ vector $X = (X_1, X_2, \ldots . X_T)'$ denotes the vector of all values of the scalar factor X_t at times $t = 1, 2, \ldots, T$.

The parameters defining the process for the instantaneous short rate in (1), δ_X and δ_0, are both scalars, as is the market price of risk parameter, λ, in (11). The restrictions $\delta_0 = 0$ and $\delta_X = 1$ are imposed, with these restrictions implying that the factor X_t is equivalent to the instantaneous short rate, r_t, in (1). Due to the particular definition of the market price of risk that is adopted in the paper, λ influences the yield equation, which we define in the next subsection below, only via the risk-neutral parameters. Given the values for the diffusion component parameters, α_1 and β_1, as described above, Φ and Ψ defined in (16) and (17) are equal to λ and zero respectively. Further substituting $\Sigma = 1$ into equations (16) and (17) produces the parameters of the risk-neutral process as

$$k^Q = k + \lambda \tag{31}$$

and

$$\mu^Q = k^Q\theta^Q = k\theta = \mu. \tag{32}$$

Using (31) and (32), the long-run mean of the risk-neutral process, θ^Q, is given by

$$\theta^Q = \frac{k\theta}{k + \lambda}.$$

3.3 Observation Equation

The model specification is completed by the inclusion of the yield equation (10), which provides the link between the observed term structure and the underlying latent factors and parameters. The observed yields comprise a panel consisting of J yields on zero coupon bonds of different terms to maturity, τ_j, $j = 1, \ldots, J$. The number of terms used for estimation

are such that $J > 1$. The observed yield with maturity j at time period t is denoted by $y_{j,t}$. The cross section of yields at time t is a $(J \times 1)$ vector, $y_{.,t} = \left[\begin{array}{ccc} y_{1,t} & \ldots & y_{J,t} \end{array} \right]'$, with $y_{.,t}$ related to the latent factor at time t, X_t, as

$$y_{.,t} = \left[\begin{array}{c} -\frac{1}{\tau_1}[A(\tau_1) - B(\tau_1) X_t] \\ \vdots \\ -\frac{1}{\tau_J}[A(\tau_J) - B(\tau_J) X_t] \end{array} \right]. \tag{33}$$

The coefficients $A(.)$ and $B(.)$ are the solutions to the ordinary differential equations defined in equations (6) and (7) and are indexed by τ_j, $j = 1, 2, \ldots, J$. For the one-factor model, $A(.)$ and $B(.)$ are both scalars.

We define the observation equation as

$$y_{.,t} = d + ZX_t + \eta_{.,t}, \tag{34}$$

where d is the $(J \times 1)$ vector,

$$d = \left[\begin{array}{c} -\frac{1}{\tau_1} A(\tau_1) \\ -\frac{1}{\tau_2} A(\tau_2) \\ \vdots \\ -\frac{1}{\tau_J} A(\tau_J) \end{array} \right] \tag{35}$$

and Z is the $(J \times 1)$ vector,

$$Z = \left[\begin{array}{c} \frac{1}{\tau_1} B_1(\tau_1) \\ \frac{1}{\tau_2} B_1(\tau_2) \\ \vdots \\ \frac{1}{\tau_J} B_1(\tau_J) \end{array} \right]. \tag{36}$$

The innovation term in (34) is a $(J \times 1)$ vector, given by

$$\eta_{.,t} = \left[\begin{array}{c} \eta_{1,t} \\ \eta_{2,t} \\ \vdots \\ \eta_{J,t} \end{array} \right], \tag{37}$$

where

$$\eta_{.,t} \sim MN(0, H), \tag{38}$$

with $\eta_{.,t}$ independent of $\eta_{.,s}$ for $t \neq s$. The matrix H in (38) is specified as a $(J \times J)$ time-invariant diagonal variance matrix for the J yields,

$$H = \left[\begin{array}{cccc} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_J^2 \end{array} \right]. \tag{39}$$

The rational for adding an error term to the observation equation, which until this point has been a deterministic function of the latent factors, is, in part, to allow for market related pricing errors. It can also be viewed as reflecting the fact that the assumed model is only an approximation to the process that determines interests rates in practice; see also Mikkelsen (2002) on this point.[3] The allowance for maturity specific variances in (39) accommodates the fact that trading activity will vary across different terms to maturity, thereby impacting on the maturity specific bid-ask spreads; see Geyer and Pichler (1996).

4 Methodology

4.1 Data Augmentation

A Bayesian approach to estimating continuous-time processes with discretely observed data is presented in Jones (1998), Elerian, Chib, and Shephard (2001) and Eraker (2001). The method derives its theoretical foundations from Pedersen (1995), who proves that a stochastic differential equation for which the likelihood function is unknown or intractable, can be estimated using an approximate conditional Gaussian process, provided the time between observations is sufficiently small. Elerian (1999) provides a demonstration of the advantages of introducing auxiliary points between observations. The demonstration uses a Ornstein-Uhlenbeck (OU) process, which, given its analytical tractability, allows a comparison to be made between the closed-form solution and the augmented approximation. The expected path between fixed points approaches the true expected path of the OU process as the number of augmented points is increased. Hence, the discretization bias associated with using a discrete time approximation to estimate the parameters of the continuous time process, can, in principle, be reduced by increasing the degree of augmentation.

The approach adopted in the present paper involves simulating augmented data points between the observed data in the *yield curve* itself. The inclusion of augmented data points reduces the time between observations, rendering the discrete time approximation to the continuous time ATS model more accurate. In addition to the augmented yield data, additional values of the latent factor are also introduced. The augmented yield data is treated as a set of latent variables which is ultimately integrated out of the joint posterior via the MCMC algorithm. Due to the assumption of conditional independence for $y_{.,t}$ in (34), the

[3] Other papers that have introduced pricing errors into a term structure model include Chen and Scott (1995), Jegadeesh and Pennacchi (1996), Babbs and Nowman (1999) and Duffee and Stanton (2001).

augmented data points are independent of the remaining yields, given the augmented values of the latent factor X_t.

We describe the observed yield data, observed at time points $t = 1$ to $t = \tilde{T}$, as

$$y^o = \left[y^o_{.,1}, y^o_{.,2}, \ldots, y^o_{.,\tilde{T}} \right], \tag{40}$$

where $y^o_{.,t}$, $t = 1, ..., \tilde{T}$, represents a $(J \times 1)$ vector of continuously compounded yields that have been observed at time t, on a set of J zero coupon bonds of different terms to maturity $\tau_1, \ldots, \tau_J$. Hence y^o is a matrix of order $(J \times \tilde{T})$. We define a quantity h as the number of augmented observations added between each pair of actual observations. We denote this augmented yield data set using the following $(J \times [\left(\tilde{T} - 1\right) \times h])$ matrix,

$$\begin{aligned} y^a \;=\; & [y^a_{.,1+\Delta}, y^a_{.,1+2\Delta}, \ldots, y^a_{.,1+h\Delta}, y^a_{.,2+\Delta}, \\ & \ldots, y^a_{.,2+h\Delta}, y^a_{.,3+\Delta}, \ldots, \ldots, y^a_{.,\tilde{T}-1+h\Delta}]. \end{aligned} \tag{41}$$

Combining the two data sets (40) and (41) together, the complete data set is given by

$$\begin{aligned} y \;=\; & [y^o_{.,1}, y^a_{.,1+\Delta}, y^a_{.,1+2\Delta}, \ldots, y^a_{.,1+h\Delta}, y^o_{.,2}, y^a_{.,2+\Delta}, \\ & \ldots, y^a_{.,2+h\Delta}, y^a_{.,3+\Delta}, \ldots, \ldots, y^a_{.,\tilde{T}-1+h\Delta}, y^o_{.,\tilde{T}}], \end{aligned} \tag{42}$$

where y is of dimension $(J \times T)$, with $T = \tilde{T} + \left(\tilde{T} - 1\right) \times h$. For notational clarity we re-express y as

$$y = [y^o_{.,1}, y^a_{.,2}, y^a_{.,3}, \ldots, y^a_{.,n-1}, y^o_{.,n}, y^a_{.,n+1}, \ldots, y^a_{.,T-1}, y^o_{.,T}], \tag{43}$$

where the n subscript in (43) indicates the nth $(J \times 1)$ vector element in y, with $n = 1, 2, \ldots, T$. For the purposes of estimation, it is not always necessary to distinguish between the observed and augmented data sets explicitly. We therefore drop the superscripts to the panels for the complete data set y, and refer to each column vector as $y_{.,t}$ only, unless there is a need to identify the observed or augmented sets of data explicitly.

4.2 Bayesian Inference and MCMC

In Bayesian analysis, Bayes Theorem is applied to produce the posterior distribution of a set of unknowns in a model, conditional on the observed data. In order to characterize the properties of a particular parameter in the model, integration over the joint posterior is required to remove any 'nuisance' parameters or latent structure. In the case of the present model there are many nuisance parameters, in the form of the augmented yield data and

the unobserved latent factors. These nuisance parameters are integrated out of the joint posterior by using a hybrid Gibbs/Metropolis-Hastings MCMC algorithm.

To apply the Gibbs Sampler, it is necessary to derive the full conditional distributions for the model parameters and latent variables. The full conditional is the posterior of the unknown quantity of interest, conditioned on specified values for all other model parameters and latent factors. The joint posterior density function for $\varphi = (k, \mu)'$, H, λ, the vector of latent factors X, and the vector of augmented yields, y^a, conditional on the observed yields, y^o, factorizes as

$$\begin{aligned} p\left(X, \varphi, H, \lambda, y^a \mid y^o\right) \quad &\propto \quad p\left(y \mid X, \varphi, H, \lambda\right) \times p\left(X \mid \varphi\right) \\ &\quad \times p\left(\varphi\right) \times p\left(H\right) \times p\left(\lambda\right). \end{aligned} \tag{44}$$

Note that the component $p\left(X \mid \varphi\right)$ in equation (44) is independent of λ and H. The functions $p\left(\varphi\right)$, $p\left(H\right)$ and $p\left(\lambda\right)$ represent the prior distributions of φ, λ and H respectively, where all three (sets of) parameters are assumed to be *a priori* independent. Certain components on the right hand side of equation (44) are expanded below, the first being the density for the conditional distribution of y,

$$p\left(y \mid X, \varphi, H, \lambda\right) = \left\{ \prod_{t=1}^{T} p\left(y_{.,t} \mid X_t, \varphi, H, \lambda\right) \right\} \tag{45}$$

and the second being the transition density of the latent factors,

$$p\left(X \mid \varphi\right) = \left\{ \prod_{t=2}^{T} p\left(X_t \mid X_{t-1}, \varphi\right) \right\} \times p(X_1), \tag{46}$$

where $p(X_1)$ is specified as a diffuse Gaussian density. Given the distributional assumption for $y_{.,t}$ in equation (34), the component densities in (45), $p\left(y_{.,t} \mid X_t, \varphi, H, \lambda\right)$, are $J-$dimensional Gaussian densities. The component densities in (46), $p\left(X_t \mid X_{t-1}, \varphi\right)$, are univariate Gaussian, due to the assumption of normality for the error term in (30). The priors in (44) are standard noninformative priors, to be described fully in the following section.

From the joint probability distribution in equation (44), the full conditional posterior distribution can be derived up to the normalizing constant for each of the parameters and latent factors. The full conditional posterior for the latent factor vector X is given by

$$p\left(X \mid \varphi, H, \lambda, y\right) \propto \left\{\prod_{t=1}^{T} p\left(y_{.,t} \mid X_t, \varphi, H, \lambda\right)\right\} \times \left\{\prod_{t=2}^{T} p\left(X_t \mid X_{t-1}, \varphi\right)\right\} \times p(X_1). \tag{47}$$

The full conditional for the latent process parameters φ is given by

$$p\left(\varphi \mid X, H, \lambda, y\right) \propto \left\{\prod_{t=1}^{T} p\left(y_{.,t} \mid X_t, \varphi, H, \lambda\right)\right\} \times \left\{\prod_{t=2}^{T} p\left(X_t \mid X_{t-1}, \varphi\right)\right\} \times p\left(\varphi\right). \tag{48}$$

Similarly, the full conditional posteriors for λ and H respectively are defined as

$$p\left(\lambda \mid X, \varphi, H, y\right) \propto \left\{\prod_{t=1}^{T} p\left(y_{.,t} \mid X_t, \varphi, H, \lambda\right)\right\} \times p\left(\lambda\right) \tag{49}$$

and

$$p\left(H \mid X, \varphi, \lambda, y\right) \propto \left\{\prod_{t=1}^{T} p\left(y_{.,t} \mid X_t, \varphi, H, \lambda\right)\right\} \times p\left(H\right). \tag{50}$$

Finally, the full conditional posterior for the augmented yields is given by the multivariate Gaussian density

$$p\left(y^a_{.,t} \mid X^a_t, \varphi, H, \lambda\right) \propto \exp\left(-\frac{1}{2}\left(y^a_{.,t} - M^a_{X_t}\right)' H^{-1}\left(y^a_{.,t} - M^a_{X_t}\right)\right), \tag{51}$$

where

$$M^a_{X_t} = d + Z X^a_t, \tag{52}$$

with parameters d and Z as previously defined in (35) and (36). The notation X^a_t is used to denote the latent factor associated with an augmented data point, $y^a_{.,t}$.

The steps taken in implementing the Gibbs chain are as follows:

1. Specify initial values ($i = 0$) for the latent factors, $X^{(i)}$, the latent process parameters, $\varphi^{(i)}$, the pricing equation variance matrix, $H^{(i)}$, and the market price of risk parameter, $\lambda^{(i)}$;

2. Iteration $i = i + 1$, proceeding through steps $3 - 8$;

3. Generate the augmented data set based on current simulated values for $X^{(i-1)}$, $\varphi^{(i-1)}$, $H^{(i-1)}$ and $\lambda^{(i-1)}$, from

   $p\left(y_{.,t}^{a,(i)} \mid X_t^{a(i-1)}, \varphi^{(i-1)}, H^{(i-1)}, \lambda^{(i-1)}\right); t = 1, 2, \ldots, \left(\tilde{T} - 1\right) \times h.$

   As noted earlier, the independence assumption adopted for $\eta_{.,t}$ in (34) means that the generation of the augmented data set does not depend directly on the observed data set, y^o. Nor does it depend on any values of the latent factor other than those associated with the augmented data points.

4. Sample the latent factor, $X^{(i)} = \left[X_1^{(i)}, \ldots, X_T^{(i)}\right]$ from the full conditional

   $p\left(X^{(i)} \mid \varphi^{(i-1)}, H^{(i-1)}, \lambda^{(i-1)}, y^{a,(i)}, y^o\right);$

5. Sample the latent process parameters from the full conditional

   $p\left(\varphi^{(i)} \mid X^{(i)}, H^{(i-1)}, \lambda^{(i-1)}, y^{a,(i)}, y^o\right);$

6. Sample the pricing equation variance matrix from the full conditional

   $p\left(H^{(i)} \mid X^{(i)}, \varphi^{(i)}, \lambda^{(i-1)}, y^{a,(i)}, y^o\right).$

7. Sample the market price of risk parameter from the full conditional

   $p\left(\lambda^{(i)} \mid X^{(i)}, \varphi^{(i)}, H^{(i)}, y^{a,(i)}, y^o\right)$

8. Repeat from Step 2 until convergence.

In the cases where the conditionals are nonstandard and, hence, cannot be drawn from directly, a Metropolis-Hasting (MH) algorithm is used. The MH draws are inserted into the outer Gibbs algorithm, with convergence of the hybrid chain ensured as long as the usual convergence criteria are satisfied; see Chib and Greenberg (1996).

A summary of the marginal posterior distributions for each unknown can be produced using the simulated draws. For example, the marginal posterior mean and variance of the mean reversion parameter, k, in (29) can be estimated from the draws of k, $k^{(i)}$, as

$$E\left[k|y\right] \approx \frac{1}{L-q} \sum_{i=q+1}^{L} k^{(i)}, \tag{53}$$

and

$$Var\left[k|y\right] \approx \frac{1}{L-q-1}\sum_{i=q+1}^{L}\left(k^{(i)}-\bar{k}\right)^{2} \tag{54}$$

respectively. Here L represents the total number of iterations of the outer Gibbs chain, whilst q represents the number of iterations required for 'burn in' of the Markov chain. Marginal posteriors for the parameters can be estimated as averages of the conditional posteriors. For example, the marginal for k can be estimated as

$$p\left(k|y\right) \approx \frac{1}{L-q}\sum p\left(k \mid .\right), \tag{55}$$

where $p(k|.)$ is the normalized full conditional density for the parameter k.

Simulation efficiency can be assessed by calculating the inefficiency factor as described in Kim, Shephard and Chib (1998). The approach adopted therein is to calculate the ratio of the variance of the mean, taking into account the correlation within the sample, and the variance of the mean assuming no correlation. The inefficiency factor (R_B) is estimated as,

$$\hat{R}_B = 1 + \frac{2B}{B-1}\sum_{i=1}^{B} K\left(\frac{i}{B}\right)\hat{\rho}\left(i\right), \tag{56}$$

where $\hat{\rho}\left(i\right)$ is the sample estimate of the autocorrelations at lag i. The term B is used to denote the bandwidth, and $K(.)$ is the Parzen kernel function

$$\begin{aligned} K\left(x\right) &= 1-6x^{2}+6x^{3}, && x \in \left[0,\frac{1}{2}\right] \\ &= 2\left(1-x\right)^{3}, && x \in \left[\frac{1}{2},1\right] \\ &= 0, && otherwise. \end{aligned} \tag{57}$$

For the empirical investigation, we set the bandwidth to $B = 2000$ for all parameters. The numerical Monte Carlo (MC) error of the mean for each of the parameters is calculated as

$$MC\ Error = S.E. \times \sqrt{\hat{R}_B}, \tag{58}$$

where S.E. is defined as

$$S.E. = \sqrt{\frac{Var\left[k|y\right]}{L-q}}.$$

4.3 Priors and Constraints

Bayesian analysis requires specification of the prior distributions for all parameters. In order to allow the data to dominate the results, we have chosen standard noninformative priors for estimation of the $A_1(1)$ model. The priors for the latent process parameters φ, and the market price of risk λ, are all assumed to be independent and proportional to a constant. The priors for the pricing equation standard deviations, $\sigma_1, \sigma_2, \ldots, \sigma_J$, are also assumed to be independent and of the standard noninformative form,

$$p(\sigma_j) \propto \frac{1}{\sigma_j}. \tag{59}$$

The parameter constraints described for the general $A_m(N)$ model reduce to

$$k > 0,\ \theta > 0 \text{ and } \mu = k\theta > 0 \tag{60}$$

for the $A_1(1)$ case. In order to impose these constraints, we discard all draws that do not fall into the admissible regions. In the numerical applications of the method described in Section 5, we find that these constraints are not particularly binding, with very few discards occurring.

4.4 Generating Augmented Yields

As described in (51), the augmented yields $y^a_{\cdot,t}$ have a multivariate normal density, conditional on φ, H, λ and the latent factor X^a_t associated with the augmented data point. Hence, $y^a_{\cdot,t}$ can be generated using a standard simulation algorithm, for each of the $\left(\tilde{T}-1\right) \times h$ augmented data points.

4.5 Sampling the Latent Factor Vector X

Sampling of the latent factor vector X is performed using a random walk MH algorithm, with the vector sampled one element at a time. That is, a single-move sampler is used for the latent factors. For each time point t, the proposal distribution for X_t is chosen as $N(X_t^{(i-1)}, \sigma^2_{X_t})$, where the mean is the current value for the tth latent factor at iteration $(i-1)$ in the outer Gibbs chain, namely $X_t^{(i-1)}$, and the variance, $\sigma^2_{X_t}$, is chosen to tune sampler so that the acceptance rate of the chain is approximately 50-80%. The conditional posterior for X_t is given by

$$
\begin{aligned}
p\left(X_t \mid X_{t-1}, X_{t+1}, \varphi, H, \lambda, y_{.,t}\right) \quad \propto \quad & p\left(y_{.,t} \mid X_t, \varphi, H, \lambda\right) \\
& \times p\left(X_t \mid X_{t-1}, \varphi\right) \\
& \times p\left(X_{t+1} \mid X_t, \varphi\right). \qquad (61)
\end{aligned}
$$

For the initial time point $t = 1$, the posterior for X_1 is reduced to

$$
\begin{aligned}
p\left(X_1 \mid X_2, \varphi, H, \lambda, y_{.,t}\right) \quad \propto \quad & p\left(y_{.,1} \mid X_1, \varphi, H, \lambda\right) \\
& \times p\left(X_2 \mid X_1, \varphi\right) \times p(X_1), \qquad (62)
\end{aligned}
$$

whilst at time $t = T$, the posterior for X_T is

$$
\begin{aligned}
p\left(X_T \mid X_{T-1}, \varphi, H, \lambda, y_{.,t}\right) \quad \propto \quad & p\left(y_{.,T} \mid X_T, \varphi, H, \lambda\right) \\
& \times p\left(X_T \mid X_{T-1}, \varphi\right). \qquad (63)
\end{aligned}
$$

Given the relevant distributional assumptions, the density in (61) is given by

$$
\begin{aligned}
& p\left(X_t \mid X_{t-1}, X_{t+1}, \varphi, H, \lambda, y_{.,t}\right) \\
\propto \quad & |H|^{-1} \exp\left(-\frac{1}{2}\left(y_{.,t} - M_{X_t}\right)' H^{-1}\left(y_{.,t} - M_{X_t}\right)\right) \\
& \times \exp\left(-\frac{1}{2X_{t-1}\Delta}\left(X_t - \left(\mu\Delta + (1 - k\Delta) X_{t-1}\right)\right)^2\right) \\
& \times \exp\left(-\frac{1}{2X_t\Delta}\left(X_{t+1} - \left(\mu\Delta + (1 - k\Delta) X_t\right)\right)^2\right) \qquad (64)
\end{aligned}
$$

where

$$
M_{X_t} = d + ZX_t \qquad (65)
$$

and d and Z are as defined in (35) and (36). The densities in (62) and (63) are appropriately modified versions of (64).

4.6 Sampling the Parameter Vector φ

As with the sampling of the latent factor vector X, a random walk MH algorithm is used to draw samples from the full conditional posterior for φ. The conditional density is given by

$$
\begin{aligned}
& p\left(\varphi \mid X, H, \lambda, y\right) \\
\propto \quad & \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\left(\left(y_{.,t}-M_{X_t}\right)' H^{-1}\left(y_{.,t}-M_{X_t}\right)\right)\right) \\
& \times \exp\left(-\frac{1}{2}\sum_{t=2}^{T}\left(\frac{1}{X_{t-1}\Delta}\left(X_t-\left(\mu\Delta+\left(1-k\Delta\right)X_{t-1}\right)\right)^2\right)\right).
\end{aligned}
$$

The proposal density used is bivariate normal with mean set at $\varphi^{(i-1)}$. A tuning covariance matrix is selected to ensure that the acceptance rates are approximately 15-25%.

4.7 Sampling the Parameter λ

A random walk MH algorithm is again used to draw from the conditional posterior for λ. The conditional density is given by

$$
\begin{aligned}
& p\left(\lambda \mid X, \varphi, H, y\right) \\
\propto \quad & \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\left(\left(y_{.,t}-M_{X_t}\right)' H^{-1}\left(y_{.,t}-M_{X_t}\right)\right)\right).
\end{aligned}
$$

The proposal density used is normal with mean set at $\lambda^{(i-1)}$. A tuning variance is selected to ensure that the acceptance rates are approximately 15-25%.

4.8 Sampling the Yield Equation Variance Matrix H

As H is assumed to be a diagonal variance matrix, we can sample the components of H individually. Having assumed a normal distribution for the pricing equation innovations, a noninformative prior for each element of H and a-priori independence between H and the remaining parameters in the pricing equation, the posterior distribution of each standard deviation component is inverted gamma,

$$
p\left(\sigma_j \mid X, \varphi, \lambda, y_{j,.}\right) \propto \frac{1}{\sigma_{\tau_j}^{T+1}} \exp\left(-\frac{Ts_j^2}{2\sigma_j^2}\right), \tag{66}
$$

where

$$
Ts_j^2 = \sum_{t=1}^{T}\left(y_{j,t}-\left[-\frac{1}{\tau_j}A\left(\tau_j\right)+\frac{1}{\tau_j}B\left(\tau_j\right)X_t\right]\right)^2. \tag{67}
$$

The posterior distribution for σ_j in (66) is dependent only on $y_{j,.}$, which denotes the $(T \times 1)$ vector of yields for the term to maturity τ_j.

5 Numerical Applications

5.1 Simulated Data

5.1.1 Data Description

To test the performance of the estimation approach described in the previous section, we use data simulated from the model described in equations (30) and (34). Three sets of experiments are conducted, associated with different values for the parameters of the model. The impact of the degree of augmentation on the accuracy with which the true parameter values are estimated is gauged by varying the value of h in (41). The simulated yields are generated in the following manner. Values for the parameters of the factor process in (30) are chosen to reflect those found in the current empirical term structure literature. In the first two sets of experiments, for which the results are reported in Tables 1 and 2, the mean reversion parameter k in (30) is assigned a value of 2.5, for a set of simulated monthly observations. This value of k implies a daily persistence in r_t of approximately $(1 - 2.5/30) = 0.92$. The specified value for k, along with the value of 5% assigned to the long-run mean of r_t, θ, imply a value of $(5 \times 2.5) = 12.5$ for μ. The market price of risk parameter, λ, is set equal to -0.04. The negative value chosen for λ is in line with empirical estimates of this parameter in one-factor models. In the third set of experiments, reported in Table 3, parameter settings that are consistent with a model for weekly observations are used. The selected value for k implies a daily persistence for r_t of $(1 - 0.092/5) = 0.98$. This high degree of persistence is specified in order to reflect the near unit root behaviour that is often characteristic of observed interest rate data.

We begin by generating a time series for the latent factor X_t, based on the parameter settings described above, using $\Delta = 1/30$. This time interval is considered sufficiently small for $\varepsilon_t\sqrt{\Delta}$ in (30) to approximate a Wiener process. From the process in (30), 30 realizations of the latent factor are generated, based on successive increments of size Δ, with the $30th$ being stored as the value of the latent factor at time t. We repeat this 500 times until we have a time series of consisting of 500 values for X_t. Using the simulated values of the latent factor X_t, we then simulate 500 monthly (weekly) observed yield curves, based on five terms to maturity, expressed as a proportion of a year, $\tau_j = 0.25, 0.5, 1.0, 3.0, 5.0$. This simulation is performed using equation (34). Values for $A(\tau)$ and $B(\tau)$ in (6) and (7) respectively are calculated using the annualised parameter settings described above, in addition to $\delta_0 = 0$,

$\delta_X = 1, \Sigma = 1$, $\alpha_1 = 0$ and $\beta_1' = 1$. The *ode*45 $MATLAB^{®}$ function is used to solve the ordinary differential equations in (6) and (7).[4]

5.1.2 Initializing the Factors and Parameters

The parameters and latent factors need to be assigned initial values in order to implement the MCMC algorithm. To initialize the parameters, parameter values are used that differ from those used to generate the simulated yields. Even if these parameters are not close to those which underlie the artificial data, the MCMC algorithm should eventually reach the area of parameter space which corresponds to the true parameter settings. Furthermore, by using perturbed initial values, it gives some indication of how the algorithm performs when initial parameter values do not match those of the data generating process. The parameter restrictions detailed in (60) are imposed by discarding those simulated values that violate the restrictions. Initial augmentation of the yield data is carried out by linear interpolation between the observed data points. This approach is considered reasonable given that the simulated yields are highly persistent through time.

5.1.3 Numerical Results

The aim of the simulation exercise is to determine how well the MCMC algorithm performs in estimating the true parameters of the $A_1(1)$ model, as well as to gauge the impact on the performance of the scheme of the extent of data augmentation that is adopted. The simulation experiments are based respectively on $h = 0$ (no augmentation), $h = 1$ and $h = 3$ in (41). The results of the first set of experiments are reported in Table 1. The actual values for the parameters are given in the second column, with the starting values used for the algorithm reported in the third column. The fourth and fifth columns respectively report the marginal posterior mean and standard deviation for each parameter, associated with the three different values of h. The number of iterations used in the experiments varies from approximately 100,000 to over 1,000,000, depending on the convergence speed of the algorithm in each case. Since only every tenth iterate is saved in order to reduce the degree of correlation in the draws, the posterior estimates are based on a number of iterates that varies between approximately 10,000 and 100,000. Convergence is monitored via the computation of the cumulative means of the simulated values for each of the parameters; see Eraker, 2001.

[4] All computations are performed using $MATLAB^{®}$.

As is evident from the values reported in Table 1, an increase in the degree of augmentation of the yields produces an improvement, overall, in the accuracy of the posterior means as point estimates of the model parameters. The posterior interval estimates of certain of the parameters are marginally wider for the larger values of h, reflecting, in part, the added degree of uncertainty that is introduced as a consequence of using the artificially augmented data. However, the improved accuracy of the location of the marginal posterior densities, for the larger values of h, outweighs the impact of the slight increase in the posterior variation.

In Table 2 the results for the second set of experiments are reported. In this case, the values for the parameters as used in the first set of experiments are retained, except for the reduction of the values of all pricing error variances in (39) to 0.01. The effect of this change is seen to be negligible on the (relative) accuracy of the posterior mean estimates of all of the parameters. However, the associated reduction in the degree of uncertainly associated with the augmented data serves to decrease the posterior standard deviations associated with all parameters in virtually all cases.

In Table 3 the third set of results are reported, based on weekly observation parameter settings. As is evident, in the case of the parameters k, μ and λ, the proportionate difference between the true parameter value and the posterior mean is substantially lower than the corresponding difference associated with the results in Table 2, for any value of h. Also, even taking into account the different magnitudes of the parameters estimated in the two cases, the posterior standard deviations reported in Table 3 indicate much more precise estimation of the true parameters in the case of the weekly model, compared with the monthly model. These results reflect the fact that a given degree of augmentation of weekly observations produces more accurate estimates of the parameters of the underlying (approximate) continuous time process than does the same degree of augmentation of monthly observations.[5] There are negligible differences between the corresponding results in the two tables that relate to the pricing error variances.

<< Insert Table 1 here>>

[5] The method of simulating the approximate continuous time process, whereby an observation is recorded after 30 increments of $\Delta t = 1/30$, means that the process whose parameters are being estimated using (augmented) weekly data is based on a finer discretization than that estimated using the monthly data.

<< **Insert Table 2 here**>>

<< **Insert Table 3 here**>>

5.2 Empirical Data: the Australian Yield Curve

5.2.1 Data Description

In this section the parameters of an $A_1(1)$ model are estimated using weekly observations on the 90 day, 180 day, 1 year, 3 year and 5 year yields on Australian Treasury Bonds, for the period January, 1990 to July, 2000. The dataset constitutes 552 observations and is displayed in Figure 1. Figure 1 shows that during this period, interest rates in Australia declined from historically high levels in the order of 18% during the early 1990's, to rates as low as 5% in the more recent period. The high degree of persistence in each time series is also evident.

<< **Insert Figure 1 here**>>

5.2.2 Empirical Results

The empirical results are summarized in Table 4. The MCMC algorithm is based on a burn-in of 100,000, with latent and parameter values taken every tenth iteration. With a total number of 300,000 iterations produced, the results are thus based on 20,000 iterations. The acceptance rates of the Metropolis-Hastings algorithms are 72%, 18% and 18% for the latent factors, latent factor process parameters, and market price of risk parameter respectively.

Referring to Table 4, the estimated low mean reversion value is consistent with the high degree of persistence in the observed data. The estimated values for k and μ also suggest a value for the long term mean θ of approximately 5%, which corresponds closely with the sample average of approximately 6% calculated for the observed 3 month yields. The estimates of k and λ in particular, are consistent with those produced by related empirical studies, based on variants of the one-factor term structure model used here; see, for example, Martin and Pagan (1995), Fruhwirth-Schattner and Geyer (1998), Bolder (2001) and Nath and Nowman (2001).

Simulation output associated with the empirical application is displayed in Figures 2 to 5, as well as being reported in Table 4. The high autocorrelations at large lags and the high inefficiency factors for some parameters suggest that the single-move sampling of the latent factors may have had an adverse impact on the convergence performance of the chain. With this in mind, we demonstrate in Section 7 the modification of the algorithm required to sample the latent factor vector X in multi-move blocks.

<< **Insert Table 4 here** >>

<< **Insert Figure 2 here** >>

<< **Insert Figure 3 here** >>

<< **Insert Figure 4 here** >>

<< **Insert Figure 5 here** >>

A short-coming of the one-factor model, as evidenced by the high posterior standard deviations for the pricing errors shown in Table 4, is that it does not accurately price the longer term yields. This issue is also discussed by Mikkelsen (2001), who comments that the one-factor model is not sufficiently flexible to represent all the shapes that the yield curve can assume at both ends of the maturity spectrum. In related work, Litterman and Scheinkman (1991) show that 99% of the variation in a term structure can be accounted for by the first three principle components, namely the level, slope and curvature, thereby indicating the need for a three-factor model; see also Duffee (2000) and Duffee and Stanton (2001). Motivated by these findings, in the following section we demonstrate the modifications to the sampling algorithm outlined for the $A_1(1)$ model required to estimate the three-factor $A_1(3)$ model.

6 Extension 1: a Three-Factor Model

6.1 State Equation for the $A_1(3)$ Model

The physical latent factor process of the $A_1(3)$ model has the following form,

$$\begin{aligned} dX_{1,t} &= (\mu_1 - k_{11}X_{1,t})\,dt + \sqrt{X_{1,t}}dW_{1,t} \\ dX_{2,t} &= (\mu_2 - k_{21}X_{1,t} - k_{22}X_{2,t} - k_{23}X_{3,t})\,dt \\ &\quad + \sqrt{1+\beta_{12}X_{1,t}}dW_{2,t} \\ dX_{3,t} &= (\mu_3 - k_{31}X_{1,t} - k_{32}X_{2,t} - k_{33}X_{3,t})\,dt \\ &\quad + \sqrt{1+\beta_{13}X_{1,t}}dW_{3,t}. \end{aligned} \tag{68}$$

With reference to the notation used in Section 2, the physical process parameters in (68) are grouped as follows,

$$K = \begin{bmatrix} k_{11} & 0 & 0 \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix}, \tag{69}$$

$$\Theta = \begin{bmatrix} \theta_1 \\ 0 \\ 0 \end{bmatrix} \tag{70}$$

and

$$\Gamma = \begin{bmatrix} 1 & \beta_{12} & \beta_{13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{71}$$

where θ_1 in (70) denotes the long-run mean of $X_{1,t}$. Note that the values for θ_2 and θ_3 are both set to zero. This is required under the DS framework for econometric identification, to allow θ_1 and δ_0 to be free parameters. Also note that the partitioning of the mean reversion parameter matrix K in (69), and long term mean vector Θ in (70), are consistent with the notation defined in (21) and (20) respectively. We further define the (3×1) vector μ as

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} k_{11}\theta_1 \\ k_{21}\theta_1 \\ k_{31}\theta_1 \end{bmatrix}. \tag{72}$$

For the purposes of simulation, we block the parameters in K, μ and Γ together as $\varphi = (k_{11}, k_{21}, k_{22}, k_{23}, k_{31}, k_{32}, k_{33}, \mu', \beta_{12}, \beta_{13})'$.

The Euler approximation of the continuous-time specification in (68) is,

$$\begin{aligned} X_{1,t+\Delta} &= \mu_1\Delta + (1 - k_{11}\Delta) X_{1,t} + \sqrt{X_{1,t}\Delta}\varepsilon_{1,t+\Delta} \\ X_{2,t+\Delta} &= \mu_2\Delta - k_{21}\Delta X_{1,t} + (1 - k_{22}\Delta) X_{2,t} - k_{23}\Delta X_{3,t} \\ &\quad + \sqrt{(1 + \beta_{12}X_{1,t})\,\Delta}\varepsilon_{2,t+\Delta} \\ X_{3,t+\Delta} &= \mu_3\Delta - k_{31}\Delta X_{1,t} - k_{32}\Delta X_{2,t} + (1 - k_{33}\Delta) X_{3,t} \\ &\quad + \sqrt{(1 + \beta_{13}X_{1,t})\,\Delta}\varepsilon_{3,t+\Delta}. \end{aligned} \tag{73}$$

Once again, under this discrete time approximation the time between observations is denoted by Δ and the Wiener process increments, $dW_{i,t}$, in (68) are approximated by multiplying a standard normal innovation term, $\varepsilon_{i,t}$, by $\sqrt{\Delta}$ to give a random variable $\varepsilon_{i,t}\sqrt{\Delta} \sim N(0, \Delta)$, $i = 1, 2, 3$. The system equation in (73) will approximate that in (68) for sufficiently small Δ. Hereafter the cross section of latent factors in 68) at time t, is a (3×1) column vector, represented by $X_{.,t} = [X_{1,t}, X_{2,t}, X_{3,t}]'$ for each $t = 1, ..., T$. The entire time series of all factors is therefore a $(3 \times T)$ matrix defined as $X = [X_{.,1}, ..., X_{.,T}]$. Individual factor time series are a $(1 \times T)$ row vector, denoted as $X_{i,.} = [X_{i,1}, \ldots, X_{i,T}]$ for $i = 1, 2, 3$.

The parameters defining the process for the instantaneous short rate in (1) and market price of risk in (11) are defined respectively as

$$\delta_X = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} \text{ and } \lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix}, \tag{74}$$

with δ_0 in (1) a scalar. Whilst the parameters in (1) were held fixed in the outline of the algorithm for the $A_1(1)$ model, and the single factor X_t identified with r_t as a consequence, in the $A_1(3)$ model δ_0 and δ_X are treated as free parameters. For the purposes of simulation, we block δ_0, δ_X and λ together as $\psi = (\delta_0, \delta_X', \lambda')'$. With reference to (18) and (19), for the $A_1(3)$ model we have

$$\Phi = \begin{bmatrix} \lambda_1 & 0 & 0 \\ \lambda_2\beta_{12} & 0 & 0 \\ \lambda_3\beta_{13} & 0 & 0 \end{bmatrix} \text{ and } \Psi = \begin{bmatrix} 0 \\ \lambda_2 \\ \lambda_3 \end{bmatrix}, \tag{75}$$

with the diffusion matrix Σ set to the identity matrix I. The parameters of the risk-neutral process can be recovered from the physical process parameters using the identities (16) and (17).

6.2 Observation Equation for the $A_1(3)$ Model

The observation equation of the $A_1(3)$ model is given by

$$y_{.,t} = d + ZX_{.,t} + \eta_{.,t}, \tag{76}$$

where Z is now a $(J \times 3)$ matrix,

$$Z = \begin{bmatrix} \frac{1}{\tau_1} B'(\tau_1) \\ \frac{1}{\tau_2} B'(\tau_2) \\ \vdots \\ \frac{1}{\tau_J} B'(\tau_J) \end{bmatrix}, \tag{77}$$

with $B(\tau_j) = (B_1(\tau_j), B_2(\tau_j), B_3(\tau_j))'$, $j = 1, \ldots, J$, a (3×1) vector of parameter functions produced as the solution to the (vector) differential equation in (7). All other terms in (76) are as previously defined.

6.3 Priors and Constraints for $A_1(3)$ Model

As with the $A_1(1)$ model, the priors for the latent process parameters, φ, and the market price of risk and instantaneous short rate parameters, ψ, are all assumed, for computational convenience, to be independent, and proportional to a constant. The priors for the pricing

equation standard deviations remain as in (59). The parameter constraints described in equations (25), (26), (27) and (28) for the general $A_m(N)$ model are now to be interpreted with reference to the parameter matrices and vectors as defined above for the $A_1(3)$ model.

6.4 Sampling Algorithm for the $A_1(3)$ Model

An MCMC algorithm along the same lines as that outlined for the $A_1(1)$ model is applicable to the $A_1(3)$ model, apart from certain obvious modifications to cater for the increase in the number of factors. With regard to the details of the algorithm outlined in Section 4, we note that sampling the augmented yields in the three-factor model proceeds exactly as detailed for the one-factor case, except for the re-definition of $M^a_{X_t}$ in (51) as

$$M_{X_{.,t}} = d + ZX_{.,t}, \tag{78}$$

where Z in (78) is as defined in (77) above. The sampling of the pricing equation standard deviations for the three-factor model is also the same as described in Section 4.8, apart from the replacement of X_t with $X_{.,t}$ and the scalar $B(\tau_j)$ with the corresponding three-dimensional vector, in (66). Details of the changes required to the algorithms for the latent factors and the parameter vectors, φ and ψ, are provided in the following sub-sections.

6.4.1 Sampling the Latent Factors in the $A_1(3)$ Model

A single-move random walk MH algorithm along the lines of that described for the $A_1(1)$ model can be adopted to sample the latent factors of the $A_1(3)$ model. In this case, an appropriate proposal distribution for $X^{(i)}_{.,t}$, the vector of the three latent factors at time point t, is a three-dimensional normal distribution, with mean equal to the current value for the tth latent vector at iteration $i-1$ in the outer Gibbs chain, $X^{(i-1)}_{.,t}$, and the covariance matrix, Σ_X, chosen so as to produce an appropriate acceptance rate for the sampler. The conditional posterior of $X_{.,t}$, $t = 2, 3, \ldots, T-1$, is given by

$$\begin{aligned}
& p\left(X_{.,t} \mid X_{.,t-1}, X_{.,t+1}, \varphi, H, \psi, y_{.,t}\right) \\
\propto\ & |H|^{-1} \exp\left(-\frac{1}{2}\left(y_{.,t} - M_{X_{.,t}}\right)' H^{-1}\left(y_{.,t} - M_{X_{.,t}}\right)\right) \\
& \times |V_{.,t+1}|^{-1} \exp\left(-\frac{1}{2}\left(X_{.,t+1} - M_{.,t+1}\right)' V^{-1}_{.,t+1}\left(X_{.,t+1} - M_{.,t+1}\right)\right) \\
& \times |V_{.,t}|^{-1} \exp\left(-\frac{1}{2}\left(X_{.,t} - M_{.,t}\right)' V^{-1}_{.,t}\left(X_{.,t} - M_{.,t}\right)\right),
\end{aligned} \tag{79}$$

where $M_{X_{.,t}}$ is defined in (78),

$$M_{.,t} = \begin{bmatrix} \mu_1\Delta + (1 - k_{11}\Delta) X_{1,t-1} \\ \mu_2\Delta - k_{21}\Delta X_{1,t-1} + (1 - k_{22}\Delta) X_{2,t-1} - k_{23}\Delta X_{3,t-1} \\ \mu_3\Delta - k_{31}\Delta X_{1,t-1} - k_{32}\Delta X_{2,t-1} + (1 - k_{33}\Delta) X_{3,t-1} \end{bmatrix} \tag{80}$$

and

$$V_{.,t} = \begin{bmatrix} X_{1,t-1}\Delta & 0 & 0 \\ 0 & (1 + \beta_{12} X_{1,t-1})\Delta & 0 \\ 0 & 0 & (1 + \beta_{13} X_{1,t-1})\Delta \end{bmatrix}, \tag{81}$$

with $M_{.,t+1}$ and $V_{.,t+1}$ defined correspondingly. The conditional posteriors for $X_{.,1}$ and $X_{.,T}$ are appropriately modified versions of (79).

6.4.2 Sampling the parameters φ.

The conditional posterior for the vector φ is given by

$$\begin{aligned} & p(\varphi \mid X, \psi, H, y) \\ \propto \; & \det[H]^{-T} \exp\left(-\frac{1}{2} \sum_{t=1}^{T} \left(\left(y_{.,t} - M_{X_{.,t}} \right)' H^{-1} \left(y_{.,t} - M_{X_{.,t}} \right) \right) \right) \\ & \times \prod_{t=1}^{T} \det[V_{.,t}]^{-1} \exp\left(-\frac{1}{2} \left(X_{.,t} - M_{.,t} \right)' V_{.,t}^{-1} \left(X_{.,t} - M_{.,t} \right) \right), \end{aligned}$$

where $M_{X_{.,t}}$ is as defined in (78), $M_{.,t}$ is as defined in (80), and $V_{.,t}$ is as defined in (81). The proposal density used is multivariate normal with mean set at $\varphi^{(i-1)}$. A tuning covariance matrix is selected to ensure that reasonable acceptance rates are achieved.

6.4.3 Sampling the parameters ψ

The conditional posterior for the vector ψ is

$$\begin{aligned} & p(\psi \mid X, \varphi, H, y) \\ \propto \; & \det[H]^{-T} \exp\left(-\frac{1}{2} \sum_{t=1}^{T} \left(\left(y_{.,t} - M_{X_{.,t}} \right)' H^{-1} \left(y_{.,t} - M_{X_{.,t}} \right) \right) \right), \end{aligned}$$

with $M_{X_{.,t}}$ as defined in (78). The proposal density used is multivariate normal with mean set at $\psi^{(i-1)}$. Again, a tuning covariance matrix is selected to ensure that the chain produces reasonable acceptance rates.

7 Extension 2: a Multi-Move Sampler

Thus far, the sampling of the latent factors has been described in terms of a single-move blocking scheme, with that scheme underlying the numerical results in Section 5. As is well-known, the correlation between successive values of the latent factors can impinge on the performance of such algorithms; see for example, Kim, Shepherd and Chib (1998). This is substantiated by our own results for the empirical application, as reported in Section 5. A multi-move sampler can be applied to the ATS models using the Kalman Filtering and Smoothing algorithm to generate candidate latent values for all time periods. This approach takes advantage of the state space form described in equations (33) and (34) for the $A_1\,(1)$ model and equations (68) and (76) for the $A_1\,(3)$ model. These models are not standard linear state space models, due to the state dependent variance components. However, by approximating the variance of the system equation so as to remove the state dependence, the Kalman Filtering and Smoothing algorithm can be used to generate a block sample, with an MH step applied to ensure that the samples are drawn from the actual state dependent variance model. With reference to the system equations (68) for the $A_1\,(3)$ model, one possible approximation involves replacing the state dependent variance component, $X_{1,t}$, at iteration i in the Gibbs sampler, with the mean of the latent factor time series, $\bar{X}_1^{(i-1)}$, based on the time path for $X_{1,t}$ simulated in iteration $i-1$,

$$\bar{X}_1^{(i-1)} = \frac{1}{T}\sum_{t=1}^{T} X_{1,t}^{(i-1)}.$$

Alternatives to the mean include the first or last $X_{1,t}^{(i-1)}$, or any values of $X_{1,t}^{(i-1)}$ within the series that are considered representative. This approach can be described as a globally linear approximation to the target model, which can be contrasted with the locally linear approximation used by Polson, Stroud and Müller (2002), based on the introduction of an additional latent multinominal random variable. Along the lines suggested in Shephard and Pitt (1997), amongst others, the acceptance rate of the MH algorithm may be improved by sampling the full latent factor vector in sub-blocks.

The Kalman Filtering and Smoothing algorithm can also be applied to sample the entire time series of a latent factor when the latent factor process is not state dependent, that is, when the model is conditionally linear in the state vector. In the case of the $A_1(3)$ model, this applies to the sampling of the second and third factors, whereby the Kalman Filtering

and Smoothing algorithm can be used to simulate these state vectors directly from their full conditionals, without the need for the additional MH step.

8 Conclusions

This paper has extended the concept of data augmentation to the estimation of ATS models. An MCMC algorithm that incorporates the augmented yield data has been outlined and applied both to data simulated from a one-factor ATS model, and to empirical data for the Australian yield curve. Evidence has been produced to indicate that increasing the degree of augmentation in the yield equation does increase estimation accuracy. The results based on application of the methodology to the Australian yield curve data appear to be most reasonable, although do highlight the limitations of using a one-factor model in an empirical setting. The way in which the algorithm can be extended to cater for the more empirically relevant three-factor model has been detailed. In addition, a potentially more efficient multi-move sampler based on the application of the Kalman Filter and Smoother to an approximation to the state space representation of the ATS model has been proposed.

A Derivation of K^Q and u^Q

This Appendix provides details of the relationship between the physical and risk-neutral parameters for the $A_m(N)$ model. Given the relationship between the physical and risk-neutral Weiner processes,

$$dW_t^Q = dW_t + \Lambda_t dt,$$

and the formulation for the risk-neutral process,

$$dX_t = K^Q\left(\Theta^Q - X_t\right)dt + \Sigma\sqrt{S_t}dW_t^Q,$$

we have the physical process as follows,

$$\begin{aligned} dX_t &= K^Q\left(\Theta^Q - X_t\right)dt + \Sigma\sqrt{S_t}\Lambda_t dt + \Sigma\sqrt{S_t}dW_t \\ &= \left(K^Q\left(\Theta^Q - X_t\right) + \Sigma\sqrt{S_t}\Lambda_t\right)dt + \Sigma\sqrt{S_t}dW_t \\ &= \left(K^Q\left(\Theta^Q - X_t\right) + \Sigma S_t\lambda\right)dt + \Sigma\sqrt{S_t}dW_t, \end{aligned}$$

where,

$$\Sigma S_t \lambda = \Sigma \begin{bmatrix} \alpha_1 + \beta_1' X_t & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \alpha_N + \beta_N' X_t \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix}.$$

Therefore,

$$\begin{aligned} \Sigma S_t \lambda &= \Sigma \begin{bmatrix} \lambda_1 (\alpha_1 + \beta_1' X_t) \\ \vdots \\ \lambda_N (\alpha_N + \beta_N' X_t) \end{bmatrix} \\ &= \Sigma \begin{bmatrix} \lambda_1 \alpha_1 \\ \vdots \\ \lambda_N \alpha_N \end{bmatrix} + \Sigma \begin{bmatrix} \lambda_1 \beta_1' \\ \vdots \\ \lambda_N \beta_N' \end{bmatrix} X_t \\ &= \Sigma \Psi + \Sigma \Phi X_t. \end{aligned}$$

Hence, we can derive the relationships between the risk neutral parameters and physical process parameters as follows,

$$\begin{aligned} dX_t &= \left(K^Q \Theta^Q - K^Q X_t + \Sigma \Psi + \Sigma \Phi X_t \right) dt + \Sigma \sqrt{S_t} dW_t \\ &= \left(\left(K^Q \Theta^Q + \Sigma \Psi \right) - \left(K^Q - \Sigma \Phi \right) X_t \right) dt + \Sigma \sqrt{S_t} dW_t \\ &= \left(K\Theta - K X_t \right) dt + \Sigma \sqrt{S_t} dW_t, \end{aligned}$$

where

$$K = K^Q - \Sigma \Phi$$

and

$$K\Theta = K^Q \Theta^Q + \Sigma \Psi$$

as in (16) and (17).

References

[1] Babbs, S.H. and Nowman, K. B., 1999, *Kalman Filtering of Generalized Vasicek Term Structure Models*, Journal of Financial and Quantitative Analysis, 34, 115-130.

[2] Bolder, D.J., 2001, *Affine Term-Structure Models: Theory and Implementation,* Working Paper 2001-15, Bank of Canada.

[3] Chen, R. and Scott, L., 1995, *Multifactor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from a Kalman Filter Model*, Working Paper, University of Georgia.

[4] Chib, S. and Greenberg, E., 1996, *Markov Chain Monte Carlo Simulation Methods in Econometrics,* Econometric Theory, 12, 409-431.

[5] Dai, Q., and Singleton, K.J., 2000, *Specification Analysis of Affine Term Structure Models*, Journal of Finance, LV, 1943-1978.

[6] Duarte, J., 2000, *The Relevance of the Price of Risk in Affine Term Structure Models*, Working Paper, University of Chicago.

[7] Duffee, G.R., 2000, *Term Premia and Interest Rate Forecasts in Affine Models*, Working Paper, Haas School of Business, University of California.

[8] Duffee, G.R. and Stanton, R.H., 2001, *Estimation of Dynamic Term Structure Models*, Working Paper, Haas School of Business, University of California, Berkley.

[9] Elerian, O., 1999, *Simulation Estimation of Continuous-Time Models with Applications to Finance*, Unpublished PhD, Nuffield College, University of Oxford.

[10] Elerian, O., Chib, S., and Shephard N., 2001, *Likelihood Inference for Discretely Observed Nonlinear Diffusions*, Econometrica, 69, 959-993.

[11] Eraker, B., 2001, *MCMC Analysis of Diffusion Models with Application to Finance*, Journal of Business & Economic Statistics, 19, 177-191.

[12] Frühwirth-Schnatter, S. and Geyer, A.L.J., 1998, *Bayesian Estimation of Econometric Multifactor Cox Ingersoll Ross Models of the Term Structure of Interest Rates via*

MCMC Methods, Working Paper, Department of Statistics, Vienna University of Economics and Business Administration.

[13] Geyer, A. and Pichler, S., 1996, *A State Space Approach to Estimate the Term Structure of Interest Rates: Some Empirical Evidence*, Working Paper, University of Economics, Vienna.

[14] Jegadeesh, N. and Pennacchi, G., 1996, *The Behaviour of Interest Rates Implied by the Term Structure of Eurodollar Futures*, Journal of Money, Credit and Banking 28, 426-446.

[15] Jones, C., 1998, *Bayesian Estimation of Continuous-Time Finance Models,* Working Paper, Simon School of Business, University of Rochester, New York.

[16] Kim, S., Shepherd, N. and Chib, S., 1998, *Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, 65, 361-393.*

[17] Litterman, R., and Scheinkman, J.A., 1991, *Common Factors Affecting Bond Returns,* Journal of Fixed Income, 1, 54-61.

[18] Maes, K., 2003, *Modeling the Term Structure of Interest Rates: Where do we stand?*, Working Paper, University of Amsterdam.

[19] Martin, V.L. and Pagan, A., 1995, *Simulation-based estimation of some factor models in econometrics*, in Simulation-Based Inference in Econometrics: Methods and Applications, editors Mariano, R., Schuermann, T. and Weeks, M.J., Cambridge University Press.

[20] Mikkelsen, P., 2002, *MCMC Based Estimation of Term Structure Models*, Working Paper, Department of Finance, The Aarhus School of Business, Denmark.

[21] Nath, P. and Nowman, K.B., 2001, *Estimates of the continuous-time Cox-Ingersoll-Ross term structure model:Further results for the UK gilt-edged market,* Applied Economic Letters, 8, 85-88

[22] Pedersen, A.R., 1995, *A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations*, Scandinavian Journal of Statistics, 22, 55-71.

[23] Polson, N.G., Stroud, J.R., and Müller, P. 2002, *Affine State-Dependent Variance Models,* Technical. Report, Institute of Statistics and Decision Sciences, Duke University.

[24] Shepherd, N. and Pitt, M., 1997, *Likelihood Analysis of Non-Gaussian Measurement Time Series,* Biometrica, 84, 653-667.

Table 1: Simulation Experiment: Varying Degrees of Augmentation of the Yield Curve; Monthly Observations (T=500)

	Actual Value	Initial Value	Posterior Mean			Posterior Standard Deviation		
			$h=0$	$h=1$	$h=3$	$h=0$	$h=1$	$h=3$
k	2.5	2.6	1.238	2.160	2.821	0.0202	0.0397	0.0485
μ	12.5	15.6	6.205	10.649	14.034	0.0960	0.1830	0.2353
λ	-0.04	-0.06	-0.020	-0.067	-0.062	0.0064	0.0156	0.0253
σ_1^2	0.05	1.0	0.010	0.035	0.054	0.0088	0.0049	0.0045
σ_2^2	0.04	1.0	0.054	0.044	0.042	0.0046	0.0032	0.0029
σ_3^2	0.03	1.0	0.033	0.031	0.032	0.0022	0.0021	0.0021
σ_4^2	0.02	1.0	0.021	0.021	0.020	0.0013	0.0013	0.0013
σ_5^2	0.01	1.0	0.010	0.010	0.010	0.0006	0.0005	0.0006

Table 2: Simulation Experiment: Varying Degrees of Augmentation of the Yield Curve with Reduced Pricing Variances; Monthly Observations (T=500)

	Actual Value	Initial Value	Posterior Mean			Posterior Standard Deviation		
			$h=0$	$h=1$	$h=3$	$h=0$	$h=1$	$h=3$
k	2.5	2.6	1.402	2.133	2.877	0.0187	0.0289	0.0279
μ	12.5	15.6	7.056	10.399	14.338	0.1062	0.1506	0.1274
λ	-0.04	-0.06	-0.017	-0.089	-0.058	0.0040	0.0099	0.0146
σ_1^2	0.01	1.0	0.0003	0.001	0.011	0.0011	0.0013	0.0015
σ_2^2	0.01	1.0	0.014	0.013	0.011	0.0009	0.0009	0.0008
σ_3^2	0.01	1.0	0.011	0.010	0.010	0.0007	0.0007	0.0007
σ_4^2	0.01	1.0	0.010	0.010	0.010	0.0007	0.0006	0.0007
σ_5^2	0.01	1.0	0.010	0.009	0.010	0.0006	0.0006	0.0006

Table 3: Simulation Experiment: Varying Degrees of Augmentation of the Yield Curve; Weekly Observations (T=500)

	Actual Value	Initial Value	Posterior Mean			Posterior Standard Deviation		
			$h=0$	$h=1$	$h=3$	$h=0$	$h=1$	$h=3$
k	0.092	0.100	0.0979	0.0940	0.0940	0.0008	0.0016	0.0006
μ	0.54	0.600	0.5767	0.5789	0.5806	0.0050	0.0039	0.0037
λ	-0.046	-0.050	-0.0495	-0.0441	-0.0452	0.0007	0.0015	0.0004
σ_1^2	0.01	0.01	0.0103	0.0106	0.0109	0.0020	0.0020	0.0021
σ_2^2	0.01	0.01	0.0119	0.0118	0.0117	0.0013	0.0013	0.0013
σ_3^2	0.01	0.01	0.0108	0.0109	0.0109	0.0008	0.0008	0.0008
σ_4^2	0.01	0.01	0.0104	0.0104	0.0104	0.0007	0.0007	0.0007
σ_5^2	0.01	0.01	0.0096	0.0096	0.0096	0.0006	0.0006	0.0006

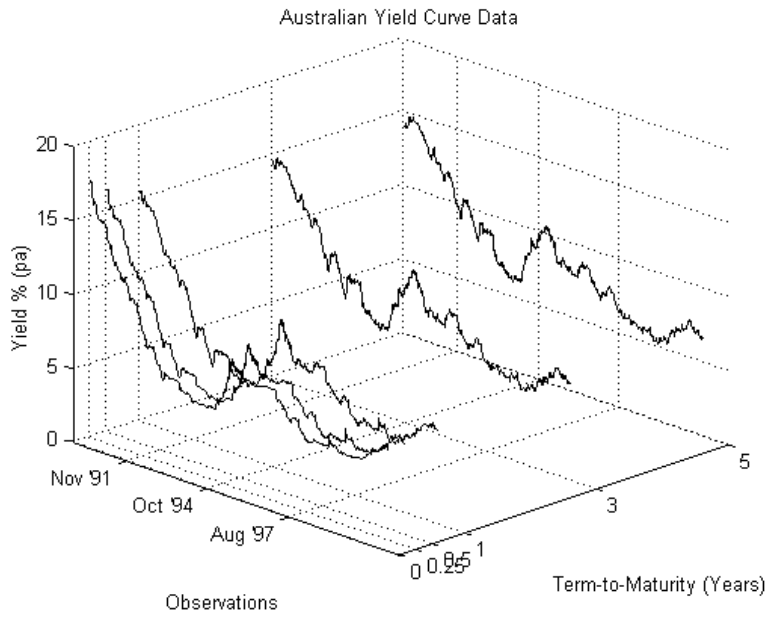

Figure 1: Weekly Australian Yield Curve Data from January 1990 to July 2000.

Table 4: Empirical Estimates Based on Weekly Australian Yeld Curve Data.

Parameter	Posterior Mean	Posterior Std. Dev.	Ineff. Factor	MC Error	25th Perc.	50th Perc.	75th Perc.
k	0.1017	0.0030	828.2	0.0003	0.0997	0.1018	0.1038
μ	0.5872	0.0085	1440.1	0.0010	0.5806	0.5870	0.5935
λ	−0.0331	0.0027	588.4	0.0002	−0.0350	−0.0331	−0.0315
σ_1^2	1.3634	0.1063	9.3	0.0010	1.2902	1.3598	1.4328
σ_2^2	0.3029	0.0250	52.0	0.0005	0.2854	0.3018	0.3191
σ_3^2	1.6527	0.1094	79.9	0.0030	1.5774	1.6485	1.7242
σ_4^2	4.2992	0.2603	9.1	0.0024	4.1207	4.2911	4.4687
σ_5^2	4.8177	0.2921	7.7	0.0024	4.6174	4.8079	5.0090

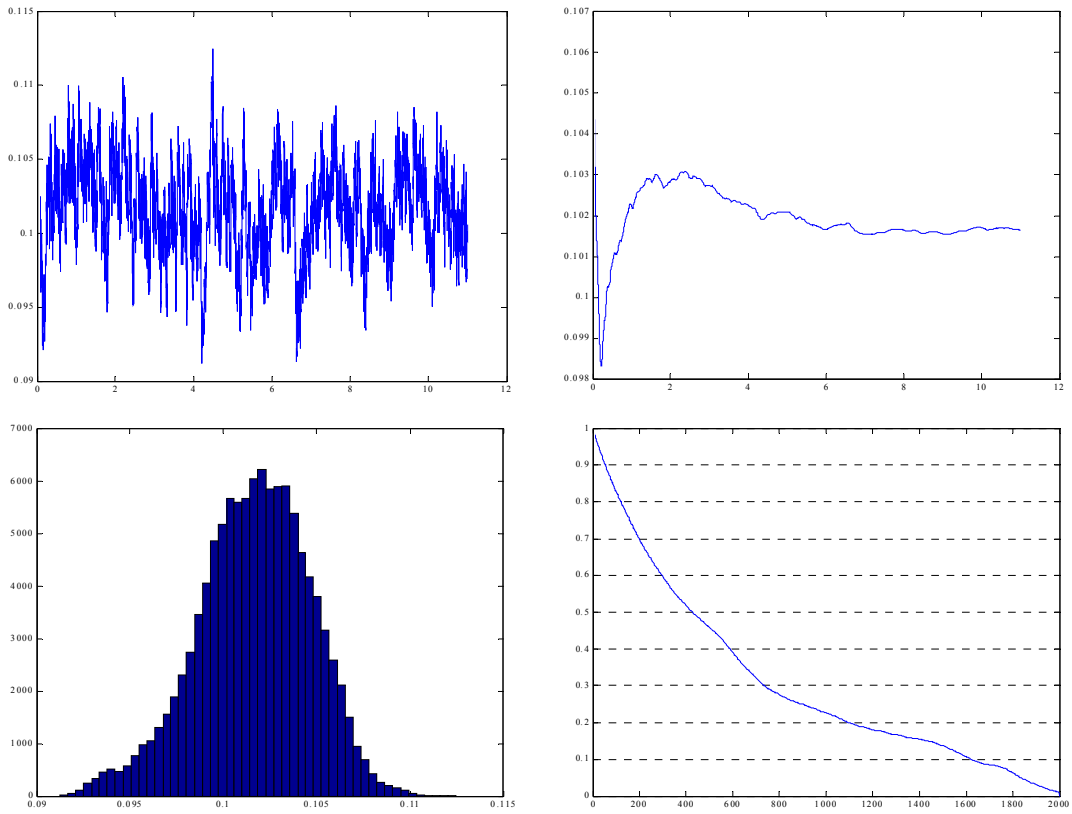

Figure 2: Empirical results for mean reversion parameter k. Time series (Top Left), Cumulative Mean (Top Right), Histogram (Bottom Left), Autocorrelations (Bottom Right)

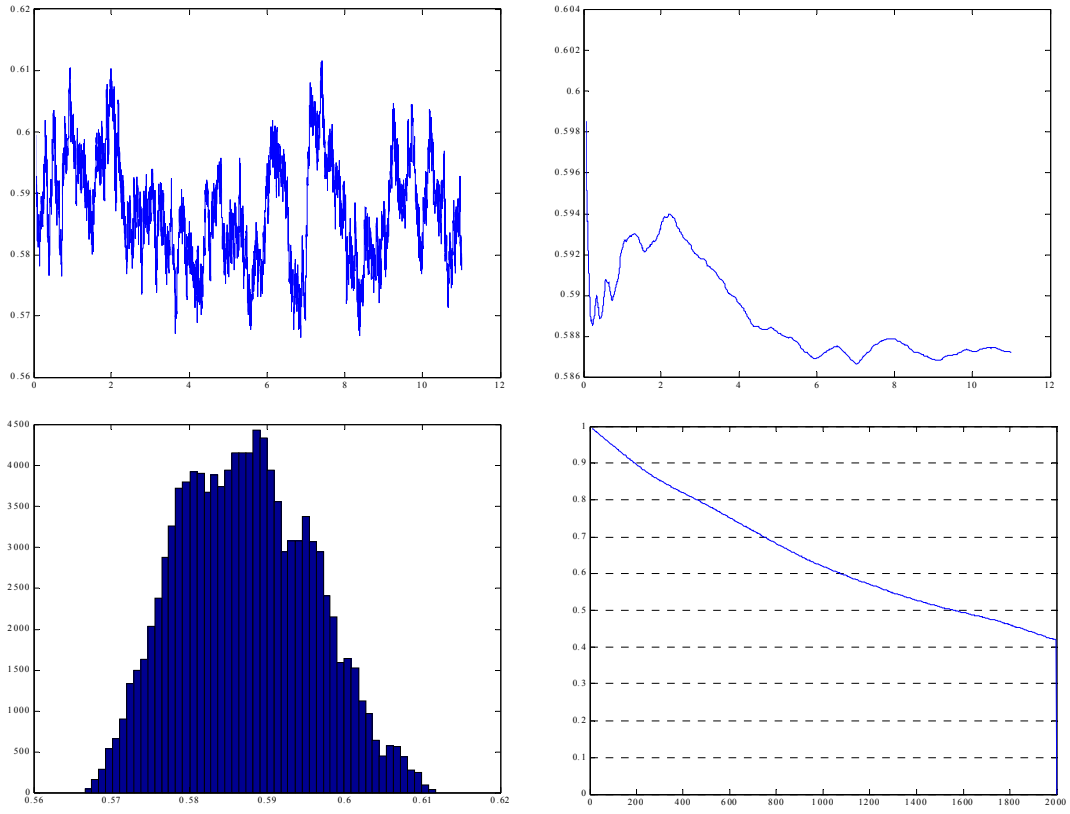

Figure 3: Empirical results for parameter μ. Time series (Top Left), Cumulative Mean (Top Right), Histogram (Bottom Left), Autocorrelations (Bottom Right)

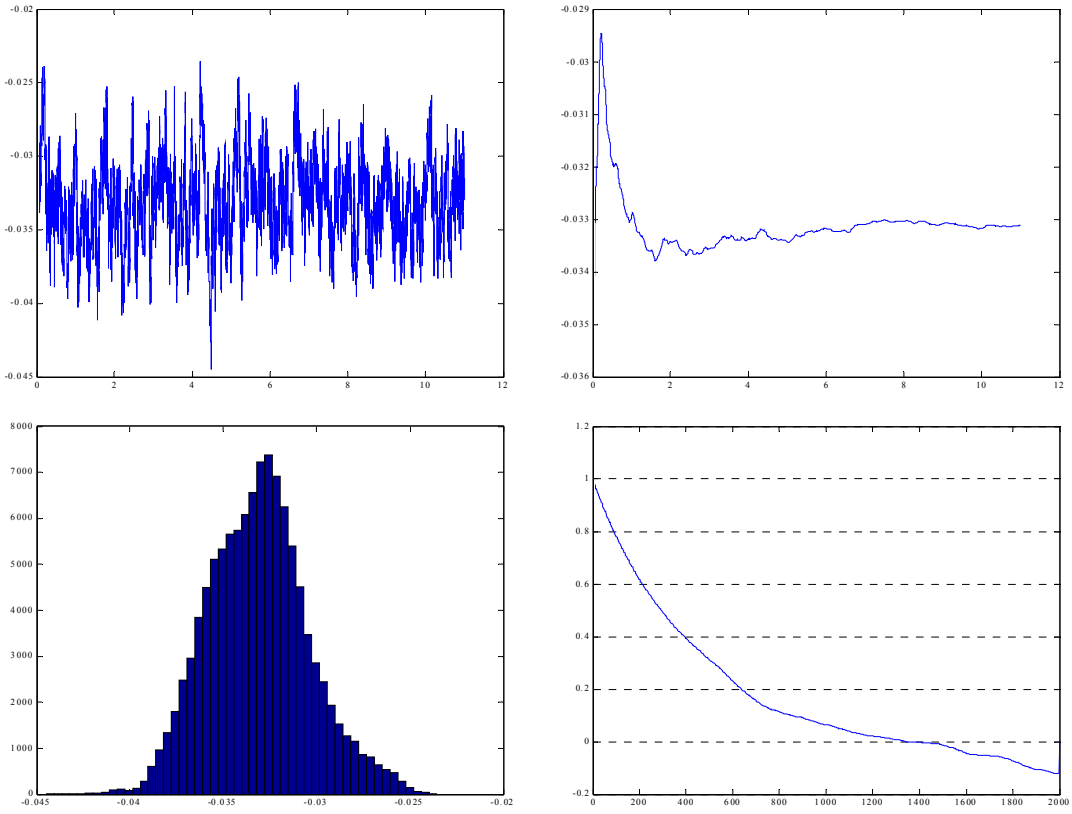

Figure 4: Empirical results for market price of risk parameter λ. Time series (Top Left), Cumulative Mean (Top Right), Histogram (Bottom Left), Autocorrelations (Bottom Right)

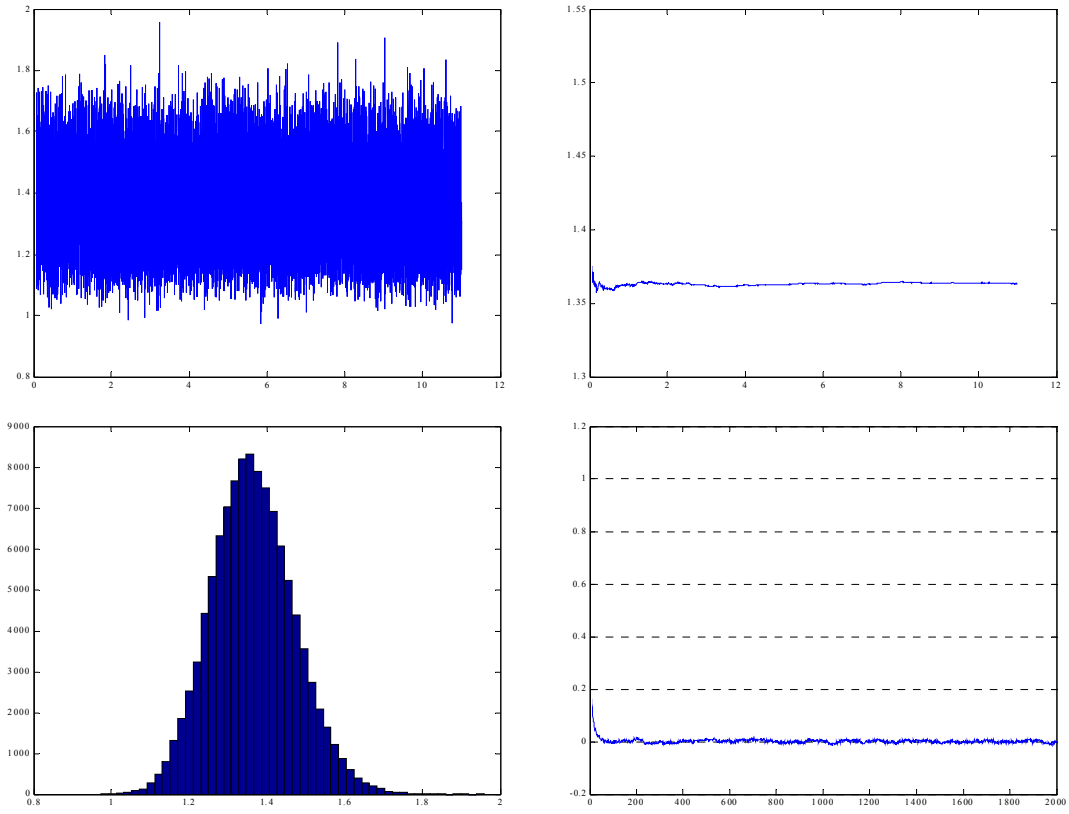

Figure 5: Empirical results for 90 day yield equation variance σ_1^2. Time series (Top Left), Cumulative Mean (Top Right), Histogram (Bottom Left), Autocorrelations (Bottom Right)