ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Bayesian Analysis of Continuous Time Models of the Australian Short Rate

Andrew D. Sanford and Gael Martin

Working Paper 11/04

Bayesian Analysis of Continuous Time Models of the Australian Short Rate.*

Andrew D. Sanford†   Gael Martin‡

May 5, 2004

Abstract

This paper provides an empirical analysis of a range of alternative single-factor continuous time models for the Australian short-term interest rate. The models are indexed by the level effect parameter for the volatility in the short rate process. The inferential approach adopted is Bayesian, with estimation of the models proceeding via a Markov Chain Monte Carlo simulation scheme. Discrimination between the alternative models is based on Bayes factors, estimated from the simulation output using the Savage-Dickey density ratio. A data augmentation approach is used to improve the accuracy of the discrete time approximation of the continuous time models. An empirical investigation is conducted using weekly observations on the Australian 90 day interest rate from January 1990 to July 2000. The Bayes factors indicate that the square root diffusion model has the highest posterior probability of all the nested models.

Keywords: Interest Rate Models, Markov Chain Monte Carlo, Data Augmentation, Bayes Factors.

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. All numerical results in the paper have been produced using the MATLAB software. Programs are available from the authors on request.

†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. Phone (03) 9905-1147

‡Department of Econometrics and Business Statistics, Monash University.

1 Introduction

Correct modelling of the instantaneous short rate is of particular importance in finance, as it is this rate which is so fundamental to the pricing of fixed-income securities. Although there is now a large number of model specifications for the short rate process, which model is the most appropriate is still an open empirical question. One of the earliest papers to attempt a formal comparison of a number of single-factor models is Chan, Karolyi, Longstaff, and Sanders (1992). Using U.S. data, Chan *et al.* estimate a number of nested, single-factor short rate models using a Generalized Method of Moments (GMM) approach. Controversially, that study rejects the commonly adopted square root diffusion model of Cox, Ingersoll and Ross (1985), whereby the volatility is proportional to the square root of the level of the interest rate. Instead, their results favour a model in which volatility is more sensitive to the current level of the interest rate, specifying an exponent for the so-called level effect in the region of 1.5. More recent studies by Conley, Hansen, Luttmer, and Scheinkman (1997), and Jones (2003), based on U.S. Federal Fund interest rates and Eurodollar rates data respectively, have tended to confirm the findings of Chan *et al.*, whilst the analyses of Aït-Sahalia (1996) and Bliss and Smith (1998) provide more support for the square root diffusion model. In particular, Bliss and Smith (1998) find that catering for structural breaks in the U.S. interest rate series reduces the magnitude of the estimated level effect from the high value estimated by Chan *et al.* Treepongkaruna and Gray (2003a) estimate alternative single-factor models using data from several countries, including Australia. Although the majority of their empirical results tend to favour a level effect parameter that exceeds 0.5, the results are sensitive to the estimation techniques used, the frequency of observations and the sampling period.

Such inconclusive findings regarding the extent of the level effect in interest rate models shed some doubt on the validity of derivative pricing methods that assume a particular value for the level effect parameter. For instance, Cox *et al.* (1985), Chen and Scott (1992), Longstaff and Schwartz (1992) and Dai and Singleton (2000) adopt bond pricing and term structure models on the assumption of a square root process for the short rate, whilst Jamshidan (1987) and Cox *et al.* (1985) produce solutions for interest rate options assuming that the level effect parameter is 0 and 0.5 respectively. Treepongkaruna and Gray (2003b) demonstrate the impact on derivative pricing of different distributional assumptions for the short rate process, adopting numerical evaluation procedures when the level effect parameter differs from either 0 or 0.5.

The aim of this paper is to perform a comparative analysis of alternative short rate models for Australian interest rate data, with a view to determining, in particular, the extent of the level effect that prevails empirically. We adopt a Bayesian inferential approach, with the data augmentation method of Jones (1998, 2003), Elerian, Chib and Shephard (2001) and Eraker (2001), used to reduce the bias associated with estimating continuous time models with discretely observed data. The alternative models are nested in a general single-factor diffusion process for the short rate, with each alternative model indexed by the level effect parameter for the volatility. Estimation and model selection is performed using a hybrid Gibbs/Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm. The latent data used to augment the actual data observed at discrete intervals is integrated out via the simulation algorithm. Model selection is based on posterior model probabilities constructed from Bayes factors, calculated, in turn, using the Savage-Dickey density ratio; see Verdinelli and Wasserman (1995). The methodology is applied to weekly observations on the Australian 90 day interest rate from January 1990 to July 2000, with the results compared with other empirical results in the literature.

The remainder of the paper is organized as follows. In Section 2 we discuss the range of models under consideration. In Section 3 the Bayesian approach to estimation and model selection is outlined, along with the algorithms used to estimate the model parameters and the Bayes factors. In Section 4 we conduct an empirical analysis using Australian short-term interest rate data. Results from the investigation suggest that the square root model is given most support by the data, whilst the model that incorporates the high level effect reported by Chan *et al.* (1992) is assigned negligible posterior probability. Some conclusions are provided in Section 5.

2 The Models

This section outlines the models to be estimated, including details of their precise specification. We adopt as the general model in which all other models are nested, the following single-factor model for the short rate at time t, r_t, described by the stochastic differential equation (SDE),

$$dr_t = (\theta + kr_t)\, dt + \sigma r_t^{\delta} dW_t, \tag{1}$$

where k, $\mu = (\theta/k)$, σ and δ denote respectively the mean reversion, long term mean, volatility, and level effect parameter of the short rate process. The term dW_t in (1) represents the independent increments of a Wiener process, W_t. The alternative nested models are indexed by different values for the level effect parameter δ, and are designated as: M_1 $(\delta = 0)$, M_2 $(\delta = 0.5)$, M_3 $(\delta = 1.0)$ and M_4 $(\delta = 1.5)$. The first two models, M_1 and M_2, correspond to the Vasicek (1977) and Cox *et al.* (1985) (square root) models respectively. Model M_3 is a variation on the short rate model used by Courtadon (1982), whilst model M_4 corresponds to the empirical model estimated by Chan *et al.* (1992) for U.S. data. We denote the general, unrestricted model, in which δ is a free parameter, as M_0.

The numerical solution of the SDE in (1) requires that the model be represented in a discrete time form. We apply the simplest of the discretization schemes, known as the Euler scheme, with the resultant discrete time version of (1) given by

$$r_{t+\Delta t} - r_t = (\theta + kr_t)\,\Delta t + \sigma r_t^{\delta}\sqrt{\Delta t}\varepsilon_t, \tag{2}$$

where $\varepsilon_t \sim i.i.d.N\,(0,1)$ and Δt represents the time between each observation. When estimating the parameters of (2), the interval Δt should be made as small as possible to reduce the bias associated with using a discrete time approximation to the continuous time process in (1). This can be achieved by 'augmenting' the observed data set with higher frequency latent data, added in between each pair of successive discrete time observations. By increasing the number of augmented data points, the size of Δt can be made smaller, and (2) made to approximate (1) more accurately as a consequence; see Elerian, Chib and Shephard (2001) for further discussion of this point.

3 Bayesian Methodology

3.1 Estimation of Bayes factors

This section provides details of the Bayesian approach to estimation and model selection in the context of the five short rate models M_j, $j = 0, 1, \ldots, 4$, described above. Bayesian inference is characterized by the application of Bayes Theorem to produce the posterior distribution of the parameters and/or unobserved latent variables of a model, M_j, given the data. Since all inference is to be conducted in the context of the approximating model in (2), augmented by the latent data, model M_j is formally defined as the version of (2) associated with jth value of δ, $j = 1, 2, \ldots, 4$, with all posterior quantities also relating to (2). However, any posterior results produced regarding the jth version of (2) are viewed as evidence relating to the corresponding jth version of the exact (but intractable) model in (1).

Denoting by $r = (r_1, r_2, \ldots, r_T)'$ the $(T \times 1)$ vector of observations on the short rate, Bayes Theorem is expressed as

$$p\left(\phi_j \mid r, M_j\right) = \frac{L(\phi_j \mid M_j) \times p\left(\phi_j \mid M_j\right)}{p\left(r \mid M_j\right)}, \tag{3}$$

where $p\left(\phi_j \mid r, M_j\right)$ denotes the posterior probability density function (pdf) of the unknowns of model M_j, ϕ_j , conditioned on the observed data and the model M_j. As will be made clear in subsequent sections, ϕ_j comprises both the unknown fixed parameters that characterize M_j and the latent augmented data points that are introduced in order to render the discrete time approximation of (1) more accurate. The posterior pdf is equivalent to the product of the likelihood function, $L(\phi_j \mid M_j)$, and the prior pdf, $p\left(\phi_j \mid M_j\right)$, normalized by the marginal likelihood, $p\left(r \mid M_j\right)$, where the latter is defined as

$$p\left(r \mid M_j\right) = \int_{\phi_j} L(\phi_j \mid M_j) p\left(\phi_j \mid M_j\right) d\phi_j. \tag{4}$$

The marginal likelihood is a measure of the support for model M_j provided by the observed data, r.

Given the alternative models, M_j, with associated prior probabilities $P(M_j)$, $j = 0, 1, \ldots, 4$, where

$$P(M_0) + P(M_1) + \cdots + P(M_4) = 1, \tag{5}$$

incorporation of model uncertainty leads to the following expression for the posterior probability for each model,

$$P\left(M_j \mid r\right) = \frac{p\left(r \mid M_j\right) \times P\left(M_j\right)}{p\left(r\right)}, \tag{6}$$

where $j = 0, 1, \ldots, 4$, and

$$p\left(r\right) = \sum_{j=0}^{4} \left[p\left(r \mid M_j\right) \times P\left(M_j\right)\right]. \tag{7}$$

The ratio of the posterior probabilities for M_j and M_k,

$$\frac{P\left(M_j \mid r\right)}{P\left(M_k \mid r\right)} = \frac{P\left(M_j\right)}{P\left(M_k\right)} \times \frac{p\left(r \mid M_j\right)}{p\left(r \mid M_k\right)},\ j \neq k = 0, 1, \ldots, 4, \tag{8}$$

defines the posterior odds ratio for M_j versus M_k. Given the assumption of equal prior probabilities, $P(M_j) = P(M_k)$, $j \neq k = 0, 1, \ldots, 4$, the expression in (8) collapses to the ratio of marginal likelihoods, which is known as the Bayes factor,

$$BF_{jk} = \frac{p(r \mid M_j)}{p(r \mid M_k)}, \; j \neq k = 0, 1, \ldots, 4, \tag{9}$$

which measures the support in the data for M_j relative to M_k.

As is clear from the expression in (4), calculation of the marginal likelihood for any given model and, hence, calculation of the Bayes factor in (9) for each pair of models, may be difficult because of the need to evaluate a complex integral involving a large number of parameters and latent variables. In this paper we employ a simple method for estimating (9) based on the Savage-Dickey density ratio. Partition the vector of unknowns for the unrestricted model M_0, ϕ_0, as

$$\phi_0 = \left[\delta, \; \phi'_{0/\delta}\right]', \tag{10}$$

where δ is the scalar level parameter such that imposing the restriction $\delta = \delta_{(j)}$ in (2) defines model M_j, $j = 1, 2, \ldots, 4$, and $\phi_{0/\delta}$ represents the vector of parameters/latent factors in M_0 not including δ. The vector $\phi_{0/\delta}$ is common to all four nested models M_1 to M_4. On the condition that

$$p\left(\phi_{0/\delta} \mid \delta = \delta_{(j)}\right) = p_j\left(\phi_{0/\delta}\right), \tag{11}$$

where $p(.)$ denotes a prior under model M_0 and $p_j(.)$ a prior under model M_j, $j = 1, 2, \ldots, 4$, then the Bayes factor in (9), for M_j versus M_0, can be shown to collapse to the so-called Savage-Dickey density ratio,

$$BF_{j0} = \frac{p\left(\delta = \delta_{(j)} \mid r\right)}{p\left(\delta = \delta_{(j)}\right)}, \; j = 1, \ldots, 4, \tag{12}$$

where

$$p(\delta \mid r) = \int p(\phi_0 \mid r)\, d\phi_{0/\delta} \tag{13}$$

is the marginal posterior of δ under M_0, and

$$p(\delta) = \int p(\phi_0) d\phi_{0/\delta} \tag{14}$$

is the marginal prior of δ under M_0. Given the specification of a proper marginal prior density for δ, the denominator in (12) can be calculated analytically. Given output from the MCMC algorithm as applied to the unrestricted model, the ordinate in the numerator can be estimated in a manner to be explained below.[1]

Table 1 contains a useful aid, reproduced from Kass and Raftery (1995), and based on criteria first proposed by Harold Jeffreys, for the assessment of Bayes factors. Given equal prior odds, the posterior probability for each model can be readily produced from the set of four Bayes factors, with the models ranked according to the relative magnitudes of the probabilities.

<< Insert Table 1 here>>

3.2 Augmentation of the Short Rate Data

As previously mentioned, a Bayesian approach to estimating continuous time processes with discretely observed data, based on the introduction of latent augmented data, is presented in Jones (1998, 2003), Elerian, Chib, and Shephard (2001) and Eraker (2001). An application of this method within a term structure framework appears in Sanford and Martin (2004). The method derives its theoretical foundations from Pedersen (1995), who shows that the transition function of a discrete time approximation to a diffusion process provides an accurate approximate to the actual transition function of the diffusion, as long as the time increments of the approximation are sufficiently small. The approach adopted in the present paper involves simulating augmented data points between the observed short rate data. The inclusion of augmented data points reduces the time between observations, rendering the discrete time approximation to the continuous time model more accurate. The augmented short rate data is treated as a set of latent variables that are ultimately integrated out of

[1]For more detailed expositions of this approach to the calculation of Bayes factors see Verdinelli and Wasserman (1995) and Koop and Potter (1999). See also Han and Carlin (2001) for a comparative review of MCMC techniques for computing Bayes factors.

the joint posterior via the MCMC algorithm.

We denote the *actual* short rate data, observed at time points $t = 1$ to $t = \tilde{T}$, as

$$r^o = \left[r_1^o, r_2^o, \ldots, r_{\tilde{T}}^o\right]'. \tag{15}$$

We define a quantity h as the number of augmented observations added between each pair of actual observations. The augmented short rate data set is then denoted by the following $\left(\left[\left(\tilde{T}-1\right) \times h\right] \times 1\right)$ vector,

$$\begin{aligned} r^a \quad = \quad & [r^a_{1+\Delta t}, r^a_{1+2\Delta t}, \ldots, r^a_{1+h\Delta t}, r^a_{2+\Delta t}, \\ & \ldots, r^a_{2+h\Delta t}, r^a_{3+\Delta t}, \ldots, \ldots, r^a_{\tilde{T}-1+h\Delta t}]'. \end{aligned} \tag{16}$$

Combining the two vectors (15) and (16), the complete data set is given by

$$\begin{aligned} r \quad = \quad & [r^o_1, r^a_{1+\Delta t}, r^a_{1+2\Delta t}, \ldots, r^a_{1+h\Delta t}, r^o_2, r^a_{2+\Delta t}, \\ & \ldots, r^a_{2+h\Delta t}, r^o_3, r^a_{3+\Delta t}, \ldots, \ldots, r^a_{\tilde{T}-1+h\Delta t}, r^o_{\tilde{T}}]', \end{aligned} \tag{17}$$

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

$$r = [r_1, r_2, r_3, \ldots, r_{t-1}, r_t, r_{t+1}, \ldots, r_{T-1}, r_T]', \tag{18}$$

where the t subscript in (18) indicates the tth scalar element in r, with $t = 1, 2, \ldots, T$. For the purposes of estimation, it is not always necessary to distinguish between the observed and augmented data sets explicitly. Hence we drop the superscripts on the elements of the complete data set r and re-introduce them only if there is a need to identify the observed or augmented sets of data explicitly.

3.3 Gibbs-MH MCMC Algorithm

In this section we describe the MCMC sampling scheme used to estimate the parameters and the Bayes factors associated with the model in (2). As is clear from the expression in (12), all four Bayes factors are based on estimation of the unrestricted version of the model, M_0, with the marginal prior and posterior of δ then evaluated at the values associated with the four nested models. We now decompose the vector $\phi_{0/\delta}$ as

$$\phi_{0/\delta} = \left[r^{a\prime}, \omega_0'\right]',$$

with r^a as defined in (16) and $\omega_0 = [\theta, k, \sigma]'$. The joint posterior density for the full set of unknowns for M_0 can be expressed as

$$p\left(r^a, \omega_0, \delta \mid r^o\right) \propto \prod_{t=1}^{T-1} p\left(r_{t+1} \mid r_t, \omega_0, \delta\right) p(r^a) p\left(\omega_0\right) p\left(\delta\right), \tag{19}$$

where the elements r^a, ω_0 and δ are assumed to be *a priori* independent, with $p(r^a)$, $p\left(\omega_0\right)$ and $p\left(\delta\right)$ denoting respectively the associated marginal priors pdf's. The prior pdf for r^a is assumed to be uniform and the priors for ω_0 and δ are detailed below. The product of the component densities $p\left(r_{t+1} \mid r_t, \omega_0, \delta\right)$, $t = 1, 2, \ldots, T$, defines the joint distribution for the full vector r, where r comprises both observed and augmented data.

The joint posterior can be factored to reveal the full conditionals for each of the unknown components r^a, ω_0, and δ. To begin, we consider the conditional posterior for a single element of r^a, r^a_τ, $\tau = 1 + \Delta t, \ldots, 1 + h\Delta t, \ldots, \tilde{T} - 1 + \Delta t, \ldots, \tilde{T} - 1 + h\Delta t$,

$$\begin{aligned} p\left(r^a_\tau \mid r^a_{/\tau}, \omega_0, \delta, r^o\right) &= p\left(r^a_\tau \mid r_{\tau+\Delta t}, r_{\tau-\Delta t}, \omega_0, \delta\right) \\ &\propto p\left(r_{\tau+\Delta t} \mid r^a_\tau, \omega_0, \delta\right) p\left(r^a_\tau \mid r_{\tau-\Delta t}, \omega_0, \delta\right), \end{aligned} \tag{20}$$

where $r^a_{/\tau}$ denotes the vector of all augmented data other than r^a_τ. Given the Markovian nature of the model in (2), the conditional posterior in (20) is a function only of the two elements of the vector r that immediately precede and follow r^a_τ, $r_{\tau-\Delta t}$ and $r_{\tau+\Delta t}$ respectively. These conditioning elements may both constitute latent values, both constitute observed values, or may constitute one latent and one observed value, depending on the value of τ.

For the parameter vector ω_0, the conditional posterior is given by

$$p\left(\omega_0 \mid r^a, \delta, r^o\right) \propto \prod_{t=1}^{T-1} p\left(r_{t+1} \mid r_t, \omega_0, \delta\right) p\left(\omega_0\right), \tag{21}$$

whilst the conditional posterior for δ is defined by

$$p\left(\delta \mid r^a, \omega_0, r^o\right) \propto \prod_{t=1}^{T-1} p\left(r_{t+1} \mid r_t, \omega_0, \delta\right) p\left(\delta\right). \tag{22}$$

The Gibbs-based sampling scheme is implemented by sampling iteratively from each of the full conditionals (20), (21) and (22), until convergence. When the full conditional is a known, closed form distribution, then efficient, standard sampling algorithms are available.

When this is not the case, we sample from the full conditional using a Metropolis-Hastings (MH) algorithm. As described above, the data set is augmented with the higher frequency latent data in order to allow Δt to become smaller than the value associated with the observed data. The trade off associated with using greater augmentation is that as $\Delta t \longrightarrow 0$ convergence will occur more slowly. As Eraker (2001) points out, in the application of Gibbs sampling to discretized SDE's, in the limit, as $\Delta t \longrightarrow 0$, the sampler will not converge.

The following algorithm is applied to estimate the model parameters for the unrestricted model, and to calculate the Bayes factors for each of the nested models.

1. Specify initial values $\omega_0^{(0)}$, $\delta^{(0)}$ and $r_{/\tau}^{a(0)}$, $\tau = 1+\Delta t, \ldots, 1+h\Delta t, \ldots, \tilde{T}-1+\Delta t, \ldots, \tilde{T}-1+h\Delta t$;

2. Set $i = 1$;

3. Sample the latent augmented short rate variable $r_\tau^{a(i)}$ from the full conditional $p\left(r_\tau^{a(i)} \mid r_{/\tau}^{a(i-1)}, \omega_0^{(i-1)}, \delta^{(i-1)}, r^o\right)$; $\tau = 1 + \Delta t, \ldots, 1 + h\Delta t, \ldots, \tilde{T} - 1 + \Delta t, \ldots, \tilde{T} - 1 + h\Delta t$;

4. Sample $\omega_0^{(i)}$ from the full conditional $p\left(\omega_0^{(i)} \mid r^{a(i)}, \delta^{(i-1)}, r^o\right)$;

5. Sample the volatility exponent $\delta^{(i)}$ from the full conditional $p\left(\delta^{(i)} \mid r^{a(i)}, \omega_0^{(i)}, r^o\right)$;

6. Approximate $p(\delta \mid r)$ as a normal density function, with mean and variance calculated from the set of iterates of δ up to and including the current iterate; see Verdinelli and Wasserman (1995). For $j = 1, 2, \ldots, 4$, estimate the ordinate of $p(\delta \mid r)$ at $\delta_{(j)}$, $\widehat{p}^{(i)}\left(\delta = \delta_{(j)} \mid r\right)$ by calculating the ordinate of the approximating normal density;

7. Estimate the Bayes factor for M_j, $j = 1, \ldots, 4$, versus the unrestricted model, M_0, as
$$\widehat{BF}_{j0}^{(i)} = \frac{\widehat{p}^{(i)}\left(\delta = \delta_{(j)} \mid r\right)}{p\left(\delta = \delta_{(j)}\right)}; \tag{23}$$

8. Set $i = i + 1$;

9. Continue Steps 3. to 8. until convergence.

3.3.1 Priors

The choice of priors is guided by a desire to ensure that posterior computations are relatively straightforward and that, as far as possible, the observed data is allowed to 'speak' for itself without strong prior information being imposed. When performing Bayes factor analysis however, there is a requirement that the parameter(s) used to index the various nested models, in this case the level effect parameter δ, be assigned a proper prior.[2] Although not strictly necessary we have also opted to use proper priors for the vector of nuisance parameters, ω_0, specifically a conjugate normal-inverted gamma prior. A robustness analysis is carried out to identify the effect that changes to the nuisance parameter prior has on the Bayes factor results. The normal-inverted gamma prior (NIG) distribution for ω_0 is initially specified as

$$\begin{aligned} \omega_0 &\sim NIG\left(\bar{\nu}, \bar{\nu}\bar{s}^2, \left[\bar{\theta}, \bar{k}\right]', \sigma^2 A^{-1}\right) \\ &= NIG\left(1, 0.01, [1.0, -0.5]', \sigma^2 \times [0.0001 \times I_2]^{-1}\right). \end{aligned} \tag{24}$$

The NIG prior distribution can be factored into its two components, the conditional distribution for the drift parameters $[\theta, k]'$ given σ, which is normal, with mean

$$E\left([\theta, k]' \mid \sigma\right) = \left[\bar{\theta}, \bar{k}\right]'$$

and (2×2) covariance matrix

$$Cov\left[[\theta, k]' \mid \sigma\right] = \sigma^2 A^{-1},$$

and the marginal inverted gamma distribution for σ, with mean

$$E(\sigma) = \frac{\Gamma\left[(\bar{\nu} - 1)/2\right]}{\Gamma(\bar{\nu}/2)} \bar{s} \sqrt{\left(\frac{\bar{\nu}}{2}\right)} \qquad \bar{\nu} > 1 \tag{25}$$

and variance,

$$Var(\sigma) = \frac{\bar{\nu}\bar{s}^2}{\bar{\nu} - 2} - \left[E(\sigma)\right]^2 \qquad \bar{\nu} > 2, \tag{26}$$

with prior parameter $\bar{s}$. The component densities of the NIG prior are in turn expressed as

[2] See Kass and Raftery (1995) for an exposition of the impact of prior specification on Bayes factors.

$$p\left([\theta,k]' \mid \sigma\right) = (2\pi)^{-1}\sigma^{-2}|A|^{1/2} \times \exp\left[-\frac{1}{2\sigma^2}\left([\theta,k]' - [\bar{\theta},\bar{k}]'\right)' A\left([\theta,k]' - [\bar{\theta},\bar{k}]'\right)\right] \tag{27}$$

and

$$p(\sigma) = \frac{2}{\Gamma(\bar{v}/2)}\left(\frac{\bar{v}\bar{s}^2}{2}\right)^{\bar{v}/2}\frac{1}{\sigma^{\bar{v}+1}}\exp\left(-\frac{\bar{v}\bar{s}^2}{2\sigma^2}\right). \tag{28}$$

respectively. The marginal mode of σ is given by

$$\sigma_{mode} = \bar{s}\sqrt{(\bar{\nu}/(\bar{\nu}+1))}. \tag{29}$$

With reference to the conditional normal prior for $[\theta,k]'$, as the diagonal values of the matrix A^{-1} are decreased, for a given σ, the distribution of $[\theta,k]'$ becomes more concentrated around the prior mean of $[\bar{\theta},\bar{k}]'$ and the prior information about $[\theta,k]'$ is sharper as a consequence. With reference to the marginal inverted gamma prior for σ, as is demonstrated in Figures 1 and 2, as we decrease the value for $\bar{v}$ (from 1.0 through to 0.05), keeping $\bar{v}\bar{s}^2$ constant, the prior density for σ becomes more diffuse but the location remains constant. As we keep $\bar{v}$ constant and increase $\bar{v}\bar{s}^2$ both the location and dispersion of (28) change.

The alternative prior specifications for $[\theta,k]'$ and σ used in the robustness analysis in Section 4 are detailed in Table 2. As can be seen, movement from Prior 1 to Prior 2 constitutes a change in the prior location of the drift parameters $[\theta,k]$, whilst moving from Prior 1 to Prior 3 assumes that the prior beliefs regarding the dispersion of $[\theta,k]$ become less diffuse. With Priors 1 to 3, σ is located at a comparatively low value, with the prior mode in (29) equal to 0.07, a figure much lower than the value of σ estimated by Treepongkaruna and Gray (2003a) for Australian short rate data. By choosing a prior that specifies low values for σ, we are allowing δ to assume an increased role in capturing the volatility of the short rate, thereby giving more weight, a priori, to models that impose high values for δ. Alternatively, Prior 4, with a mode or 0.2 for σ, is imposing prior information on σ that reflects more closely the empirical estimates reported by Treepongkaruna and Gray. This, in turn, puts less emphasis on the role of δ, thereby giving more prior weight to models that impose lower values of δ.

<< Insert Figure 1 here >>

<< Insert Figure 2 here >>

<< Insert Table 2 here>>

For the level effect parameter δ, we select a uniform prior,

$$\delta \sim U\left(-0.5, 2.0\right), \tag{30}$$

where we have assumed an admissible domain of $(-0.5, 2.0)$. Motivated by the approach to Bayes factor construction adopted by Schotman and van Dijk (1991) for the parameter in a first order autoregressive model, we choose the boundaries of this domain in such a way that they encompass virtually all of the marginal posterior mass for δ.

3.3.2 Sampling the Latent Augmented Interest Rates

With MCMC algorithms, the blocking scheme used to sample the parameters and latent augmented data needs to be identified, as the unknowns can be simulated as individual scalars or grouped as vectors. We note that blocking highly correlated latent factors into higher dimensional components can be more efficient, as demonstrated by Carter and Kohn (1994) and Shephard and Pitt (1997), and as recommended by Elerian, Chib, and Shephard (2001). We choose, however, to keep the computational aspects of our algorithm as simple as possible, by following the approach of Jones (1998, 2003) and Eraker (2001), and sampling the latent augmented data one element at a time. Using the expression for the conditional posterior in (20) for r_{τ}^{a}, $\tau = 1 + \Delta t, \ldots, 1 + h\Delta t, \ldots, \tilde{T} - 1 + \Delta t, \ldots, \tilde{T} - 1 + h\Delta t$, it follows that the conditional density is given by

$$p\left(r_{\tau}^{a} \mid r_{\tau+\Delta t}, r_{\tau-\Delta t}, \omega_0, \delta\right) \propto \left(\sigma \left(r_{\tau}^{a}\right)^{\delta} \sqrt{\Delta t}\right)^{-1} \exp\left(-\frac{1}{2} Q_{\tau+\Delta t}^{2}\right) \times \left(\sigma r_{\tau-\Delta t}^{\delta} \sqrt{\Delta t}\right)^{-1} \exp\left(-\frac{1}{2} Q_{\tau}^{2}\right), \quad (31)$$

where

$$Q_{\tau+\Delta t} = \left(\frac{r_{\tau+\Delta t} - \left(r_{\tau}^{a} + \left(\theta + k r_{\tau}^{a}\right) \Delta t\right)}{\sigma \left(r_{\tau}^{a}\right)^{\delta} \sqrt{\Delta t}}\right) \quad (32)$$

and

$$Q_{\tau} = \left(\frac{r_{\tau}^{a} - \left(r_{\tau-\Delta t} + \left(\theta + k r_{\tau-\Delta t}\right) \Delta t\right)}{\sigma r_{\tau-\Delta t}^{\delta} \sqrt{\Delta t}}\right). \quad (33)$$

Note that we do not need to distinguish between observed and augmented conditioning values of in (31), (32) and (33), as the precise nature of these values has no relevance to the sampling of r_{τ}^{a}. The time between each element of r, whether observed or augmented, is Δt.

As (31) is nonstandard, we sample from it via an MH algorithm; see Chib and Greenberg (1995, 1996) for more details. The candidate density adopted is proportional to the second component of the conditional density in (31); see also Jones (1998, 2003). By using this candidate, the probability with which the candidate draw, $r_{\tau}^{a,cand}$, is accepted as a draw from the conditional posterior in (31) at iteration i in the MCMC algortihm is given by

$$\alpha\left(r_{\tau}^{a,(i-1)}, r_{\tau}^{a,cand}\right) = \min\left\{1, \frac{p\left(r_{\tau+\Delta t} \mid r_{\tau}^{a,cand}, \omega_0, \delta\right)}{p\left(r_{\tau+\Delta t} \mid r_{\tau}^{a,(i-1)}, \omega_0, \delta\right)}\right\}. \quad (34)$$

3.3.3 Sampling the Short Rate Process Parameters

The full conditional (21) for the parameter set $\omega_0 = [\theta, k, \sigma]'$ has a standard NIG form, given the use of the natural conjugate NIG prior in (24). To simplify the exposition, we define the drift parameter $\beta = [\theta, k]'$ and express the density associated with (24) as

$$p\left(\beta, \sigma\right) \propto \sigma^{-2-\bar{v}-1} \exp\left\{-\frac{1}{2\sigma^2}[\bar{v}\bar{s}^2 + \left(\beta - \bar{\beta}\right)' A \left(\beta - \bar{\beta}\right)\right\}, \quad (35)$$

where $\bar{\beta} = \left[\bar{\theta}, \bar{k}\right]'$. Defining the $(T \times 1)$ vector y as

$$y=\begin{bmatrix} (r_2-r_1)/r_1^{\delta}\sqrt{\Delta t} \\ \vdots \\ (r_{t+1}-r_t)/r_t^{\delta}\sqrt{\Delta t} \\ \vdots \\ (r_T-r_{T-1})/r_{T-1}^{\delta}\sqrt{\Delta t} \end{bmatrix}, \tag{36}$$

and the $(T \times 2)$ matrix X as

$$X=\begin{bmatrix} \sqrt{\Delta t}/r_1^{\delta} & r_1\sqrt{\Delta t}/r_1^{\delta} \\ \vdots & \vdots \\ \sqrt{\Delta t}/r_t^{\delta} & r_t\sqrt{\Delta t}/r_t^{\delta} \\ \vdots & \vdots \\ \sqrt{\Delta t}/r_{T-1}^{\delta} & r_{T-1}\sqrt{\Delta t}/r_{T-1}^{\delta} \end{bmatrix}, \tag{37}$$

standard algebra leads to a joint conditional posterior for β and σ with component densities

$$\begin{aligned} p(\beta \mid \sigma, y, X) &= (2\pi)^{-1}\sigma^{-2}\left|A+X'X\right|^{1/2} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2}\left[\left(\beta-\tilde{\beta}\right)(A+X'X)\left(\beta-\tilde{\beta}\right)\right]\right\} \end{aligned} \tag{38}$$

and

$$p(\sigma \mid y, X)=\frac{2}{\Gamma(\tilde{v}/2)}\left(\frac{\tilde{v}\tilde{s}^2}{2}\right)^{\tilde{v}/2}\sigma^{-(\tilde{v}+1)}\exp\left(-\frac{\tilde{v}\tilde{s}^2}{2\sigma^2}\right). \tag{39}$$

The posterior quantities in (38) and (39), $\tilde{\beta}$, $\tilde{v}\tilde{s}^2$ and $\tilde{v}$, are given respectively by

$$\tilde{\beta}=(A+X'X)^{-1}\left(A\bar{\beta}+X'y\right), \tag{40}$$

$$\tilde{v}\tilde{s}^2=\bar{v}\bar{s}^2+y'y-\tilde{\beta}'(A+X'X)\tilde{\beta}+\bar{\beta}'A\bar{\beta} \tag{41}$$

and

$$\tilde{v}=T+\bar{v}. \tag{42}$$

The product of the densities in (38) and (39) defines the conditional posterior for ω_0 in (21), with r^a, δ and r^o entering as conditioning values via y and X as defined in (36) and (37) respectively.

3.3.4 Sampling the Level Effect Parameter

The conditional in (22) for the level effect parameter δ, is of a nonstandard form. We adopt a random walk MH algorithm, based on a truncated normal candidate distribution, $N\left(\delta^{(i-1)}, \sigma^2_{tune,\delta}\right) I_{-0.5<\delta<2.0}$, located at the $(i-1)$th value of δ, $\delta^{(i-1)}$, and tuned with $\sigma^2_{tune,\delta}$. The indicator function $I_{-0.5<\delta<2.0}$ has the value 1 when δ is in the open interval $(-0.5, 2.0)$ and zero elsewhere. The candidate draw, δ^{cand}, is accepted with probability

$$\alpha\left(\delta^{(i-1)}, \delta^{cand}\right) = \min\left(1, \frac{p\left(\delta^{cand} \mid r^{a,(i)}, \omega_0^{(i)}, r^o\right)}{p\left(\delta^{(i-1)} \mid r^{a,(i)}, \omega_0^{(i)}, r^o\right)}\right), \tag{43}$$

with

$$p\left(\delta \mid r^a, \omega_0, r^o\right) \propto \left[\sigma^{-(T+1)} \prod_{t=1}^{T-1} \frac{1}{r_t^\delta} \times \exp\left(-\frac{1}{2}\left(\frac{r_{t+1} - \left(r_t + \left(\theta + kr_t\right)\Delta t\right)}{\sigma r_t^\delta}\right)^2\right)\right] \times I_{-0.5<\delta<2.0}. \tag{44}$$

3.3.5 Initialization, Convergence and Inefficiency Diagnostics

The latent augmented short rate data is initialized by linear interpolation between the observed rates. Parameters are initialized using perturbed values of previously published empirical results. Convergence of the MCMC chain is monitored graphically via a time series of cumulative means. Simulation inefficiency is determined by calculating the simulation inefficiency factors as described in Kim, Shephard and Chib (1998). The inefficiency factor represents the variance of the mean of the iterates from the MCMC sampling scheme, divided by the variance of the mean when it is assumed that draws are independent. This ratio, R_B, can be evaluated using the following expression,

$$\hat{R}_B = 1 + \frac{2B}{B-1} \sum_{n=1}^{B} K\left(\frac{n}{B}\right) \hat{\rho}\left(n\right), \tag{45}$$

where $\hat{\rho}\left(n\right)$ is the sample estimate of the autocorrelation at lag n, calculated as

$$\hat{\rho}\left(n\right) = \frac{\hat{\Gamma}\left(n\right)}{\hat{\Gamma}\left(0\right)}, \tag{46}$$

where

$$\hat{\Gamma}(n) = \frac{1}{M}\sum_{i=n+1}^{M}\left(z^{(i)} - \bar{z}\right)\left(z^{(i-n)} - \bar{z}\right), \qquad \bar{z} = \frac{1}{M}\sum_{i=1}^{M} z^{(i)}, \tag{47}$$

and $z^{(i)}$ represents the ith iterate of the relevant parameter. The parameter B in (45) is known as the bandwidth, and $K(.)$ is the parzen kernel defined as

$$\begin{aligned} K(x) &= 1 - 6x^2 + 6x^3, & x \in \left[0, \frac{1}{2}\right] \\ &= 2(1-x)^3, & x \in \left[\frac{1}{2}, 1\right] \\ &= 0, & otherwise. \end{aligned} \tag{48}$$

For this exercise we set the bandwidth at $B = 2000$. The maximum length of the lag n is also set at 2000. The numerical Monte Carlo error of the mean for each of the parameters is calculated as

$$MC\ Error = S.E. \times \sqrt{\hat{R}_B} \tag{49}$$

where $S.E.$denotes the standard error, calculated as the standard deviation of the iterates divided by the square root of the number of iterates.

4 Empirical Application: Australian Interest Rate Data.

4.1 Data Description.

The empirical investigation is based on 552 weekly observations on the Australian 90 day interest rate, sampled every Wednesday from 1 January 1990 to 26 July 2000. This period comprises a shift from historically high interest rates in the early 1990's to low interest rate levels in the latter part of the sample period, such as had not been experienced in Australia since the 1960's. As in the similar study by Treepongkaruna and Gray (2003a), we use the 90 day rate interest rather that the shorter 30 day rate as a proxy for the instantaneous short rate. Treepongkaruna and Gray comment that the use of 90 day rate is motivated by its high liquidity. The interest rate data is displayed in Figure 3 and the first differenced series in Figure 4. Summary statistics for both sets of data are provided in Table 3. The skewness and kurtosis statistics reported therein are sample estimates of the third and fourth moments respectively of the standardized random variable.

<<Insert Figure 3 here>>

<<Insert Figure 4 here>>

<<Insert Table 3 here>>

From Figures 3 and 4 it is clear that there is indeed a tendency for the volatility in the interest rate series to be positively correlated with the current level of the rate. This feature is particularly marked for the January 1990 to November 1991 period in which both the level and volatility of the rate are high. It is also relevant for the August 1997 to July 2000 period, in which, apart from a sharp jump in the level of interest rates on 10 June 1998, the level and volatility are both lower. The level effect is less marked over the 1991 to 1997 period, with the increased volatility observed in the late 1994 period appearing to be more closely aligned with the *shift* from a lower to a higher interest rate regime, rather than being associated with the latter specifically. These empirical features tend to tally with those reported in Brenner, Harjes and Kroner (1996) and Eraker (2001), with the former authors concluding that unexpected 'news' is important in understanding the volatility of interest rates.

The time-varying nature of the volatility that is evident in Figures 3 and 4 is associated, in turn, with an empirical distribution for the first differenced data that exhibits excess kurtosis, with the relevant kurtosis statistic reported in Table 3 being significantly greater than the value of 3 associated with the normal distribution. The negative skewness coefficient reported therein is also significantly less than the value of zero associated with the symmetric normal distribution, and is reflective of a 'leverage' effect of sorts, whereby interest rate falls are associated with higher volatility than increases of the same magnitude.

4.2 Empirical Results.

The estimation results for each of the four priors as described in Table 2 are reported in Tables 4, 5, 6 and 7 respectively. All results are based on a burn-in period of 100,000 iterations, followed by a further 500,000 iterations. Of the samples following burn-in, every tenth iterate is stored, resulting in a total of 50,000 iterates available for parameter estimation and convergence analysis. Augmentation is implemented by assigning values for h in (16) of 3, 1 and 0 respectively, corresponding, in turn, to values for Δt of 0.25, 0.5 and 1. This level of augmentation is considered adequate given that the observations are weekly. Jones (2003) comments that augmentation is most important when using monthly data for estimation, finding that daily data produces little discretization bias. This suggests that high levels of augmentation are unnecessary for our weekly observed data, thereby reducing the computational burden.

We first consider the results in Table 4, as associated with Prior 1. The first thing to note is the relative stability of both the location estimates (marginal posterior mean and 50th percentile (median)) and the posterior standard deviations, over different values for h. Only the inefficiency factors alter noticeably as the degree of augmentation increases, indicating that there is more correlation in the iterates of each of the four parameters as more augmented data points are inserted between the observed data; see also Sanford and Martin (2004) on this point. Inefficiency is also markedly higher for the diffusion parameters, σ and δ, than for the drift parameters, θ and k. For example, the value of 512.6 for δ ($h = 3$) indicates that approximately 51000 iterations of the chain are required in order to limit the variance of the mean of the iterates to be 1% of the variation due to the data (as measured by the posterior variance). The value of 1.4 for k ($h = 3$), on the other hand, indicates that only 1400 iterations are required in order to achieve the same degree of accuracy for k. Percentiles for the drift parameters, θ and k, show that the iterates are evenly dispersed above and below the mean, with the medium coinciding closely with the estimated mean. This symmetry can also be seen in the graphical outputs for the drift parameters in Figures 5 and 6 (as based on Prior 1). The estimate of the mean reversion parameter k implies a high persistence parameter of 0.99 for the weekly short rate data, which tallies with the near unit root behaviour evident in Figure 3. The long run mean of the short rate as implied by the estimates of k and θ is 5.51%.

The point estimates of δ reported in Table 4 differ little from the estimates reported by Andersen and Lund (1997), Eraker (2001) and Hurn, Lindsay and Martin (2003) of 0.676, 0.757 and 0.676 respectively, all as based on 90 day U.S. Treasury Bill data. Also, results reported by Dahlquist (1996), although varying across the different European economies investigated, favour δ values that are consistent with those reported here. In contrast, however, Chan *et al.* (1992) and Brenner, Harjes and Kroner (1996), estimate respective values for δ of 1.500 and 1.559, using US data. Similarly, Treepongkaruna and Gray (2003a) report high estimated values for δ for Australian data, ranging from 0.929 to 1.552 depending on both the data set and estimation procedure used. It is noteworthy that the data used by the latter authors covers a longer period of time than does our sample of Australian data, including a more extended period of high rates, with the mean value of our data being 7.13% compared with 10.62% for the Treepongkaruna and Gray data set. Treepongkaruna and Gray in fact conclude from their cross country evaluations that data sets with a high average value tend to produce higher level effect parameter estimates than those for which the average value is lower.

The results reported in Table 5 indicate that there is little impact on any of the estimates of adopting Prior 2 rather than Prior 1, apart from a slight increase (decrease) in the estimated values of δ (σ). The qualitative behaviour of the inefficiency factors across the different parameters is also the same in Table 4 as in Table 5. On the other hand, the results in Tables 6 and 7, as based on Priors 3 and 4 respectively, show that these particular prior specifications have had some impact on the posterior results. The impact is more pronouced for Prior 4 than for Prior 3. In particular, the results for σ and δ in Table 7, in which the prior information on k and θ is quite tight and the prior location for σ relatively high, differ from the results in Tables 4 and 5, in that the estimate for σ is substantially higher and the estimates of δ lower. Considering the results in all four tables, a negative correlation between the estimates of the volatility parameter σ and the level effect parameter δ is evident. This is understandable given that the overall model volatility is determined by the interaction of these two parameters. The tighter prior information on k and θ associated with the results in Tables 6 and 7 has produced smaller posterior standard deviations for these parameters. However, at the same time, the degree of correlation in the sampled iterates, as measured by the inefficency factors, has increased, as has the overall Monte Carlo error associated the mean estimates of each parameter.

<< Insert Table 4 here>>

<<Insert Table 5 here>>

<<Insert Table 6 here>>

<<Insert Table 7 here>>

<<Insert Figure 5 here>>

<<Insert Figure 6 here>>

<<Insert Figure 7 here>>

<<Insert Figure 8 here>>

The Bayes factors for each of the models are shown in Table 8, with the Bayes factor for model M_j ($j = 1, \ldots, 4$), relative to the unrestricted model M_0, denoted by BF_{j0}, calculated as the mean of the iterates of $\widehat{BF}_{j0}$ produced as described in Section 3.3. The MC Errors associated with the BF_{j0} values are calculated in a similar manner to those for the individual parameters. The Bayes factors support the unrestricted model, M_0, against each of the restricted models in virtually all cases. The exception to this is the support for the square root model, M_2, under Prior 1 for all levels of augmentation and under Prior 2 with no augmentation ($h = 0$). Based on the criteria in Table 1 however, it is only when using Prior 1, with $h = 0$, that M_2 has any substantial dominance over M_0.

Comparisons of the Bayes factors for the restricted models can be carried out in a straightforward way by noting that the following identity applies,

$$BF_{jk} = \frac{BF_{j0}}{BF_{k0}}, \quad j \neq k = 1, 2, \ldots, 4. \tag{50}$$

The results in Table 9 are based on results produced for all four priors, with augmentation set at $h = 2$. These results show that M_2 has the highest Bayes factor of all the nested models, for all priors. Even when the support for M_3 becomes more substantial on using Prior 2, the Bayes factor for M_2 versus M_3 is still approximately equal to six. Model M_1 obtains more support when Prior 4 is applied, but the Bayes factor in favour of M_2 is still approximately sixty. Model M_4 performs very poorly, with the Bayes factors indicating negligible support for this model over all other alternatives considered. Similar conclusions can be drawn from the results (not reported) based on $h = 3$ and $h = 0$.

An alternative representation of the information contained in Tables 8 and 9 is in terms of the posterior probability for each of the models. The posterior model probabilities are calculated from the Bayes factors as

$$p(M_j \mid r) = \frac{p(r \mid M_j)}{\sum_{k=0}^{k=4} p(r \mid M_k)}, \tag{51}$$

given the assumption of equal prior probabilities for all models $j = 0, \ldots, 4$. The results in Table 10 clearly highlight the lack of posterior support for models M_1 and M_4 and the small amount of support for M_3, and that only under Prior 2. M_2 is the dominant restricted model, with non-negligible posterior probability under all priors other than Prior 4. Under Prior 1 it clearly dominates even the unrestricted model, for all levels of augmentation. The results thus provide some support for the Cox *et al.* (1985) square root diffusion model (M_2), whilst providing no support for the model that corresponds to the pronounced level effect reported in Chan *et al.* (1992), namely M_4.

<< Insert Table (8) here >>

<< Insert Table (9) here >>

<< Insert Table (10) here >>

Finally, it is interesting to note the effect that an increase in augmentation has on the relative posterior probabilities of the two dominant models, namely M_0 and M_2. For Priors 3 and 4, the relativities remain fairly constant across different values of h. However, for Priors 1 and 2, as the degree of augmentation increases, the posterior probability of M_0 increases whilst that of M_2 decreases. Considering the case of Prior 2 in particular, when no augmentation is applied it is difficult to distinguish between M_0 and M_2. However, as augmentation is increased and the bias associated with approximating the continuous time model with a discrete time process is reduced, the support for the unrestricted model M_0 increases to the point where the relative support for it over M_2 is much clearer.

5 Conclusions

In this paper, we have compared a number of alternative models for the Australian short-term interest rate, all of which are restricted examples of a general continuous time model. The models are estimated using a Bayesian approach, with an MCMC algorithm used to draw iterates from the posterior densities of the parameters. Discretization bias associated with the Euler scheme used to approximate the continuous time model is reduced by incorporating latent augmented data. The iterates produced by the simulation algorithm are then used to estimate Bayes factors for each of the nested models using the Savage-Dickey density ratio. From the Bayes factors, we find that the Cox *et al.* (1985) square root diffusion model has the greatest support out of all of the nested models, whilst the Chan *et al.* (1992) model performs worst of all. Even when allowing for changes in the prior specifications for the parameters, the square root model still continues to perform substantially better than all other restricted model considered. For one particular prior specification the square root model has more posterior support than the model that allows the level parameter to be unrestricted. The results presented suggest therefore that the application of the analytical pricing equations made available under the Cox *et al.* model are not unreasonable in the Australian context. The differences in economic performance however, associated with using the analytical pricing versus the numerical pricing approach associated with an unrestricted level effect model for the short rate would still need to be assessed.

References

[1] Aït-Sahalia, Y., 1996, Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies*, Vol 9, No 2, 385-426.

[2] Andersen, T.G, and J. Lund, 1997, Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate, *Journal of Econometrics*, Vol 7, 343-377..

[3] Bliss, R.R., and D.C. Smith, 1998, The Elasticity of Interest Rate Volatility:Chan, Karolyi, Longstaff and Sanders Revisited, Working Paper (Working Paper 97-13a, Federal Reserve Bank of Atlanta).

[4] Brenner, R.J, R.H. Harjes, and K.F. Kroner, 1996, Another Look at Models of the Short-Term Interest Rate, *Journal of Financial and Quantitative Analysis*, Vol 31, No 1, 85-107.

[5] Carter, C.K. and R. Kohn, 1994, On Gibbs Sampling for State Space Models, *Biometrika*, Vol 81, No 3, 541-553.

[6] Chan, K. C., G. Karolyi, F. Longstaff, and A. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance*, 47, 1209-1227

[7] Chen, R., and L. Scott, 1992, Pricing Interest Rate Options in a Two-Factor Cox-Ingersoll-Ross Model of the Term Structure, *Review of Financial Studies* 5, 613-636.

[8] Chib, S. and E. Greenberg, 1995, Understanding the Metropolis-Hastings algorithm, *The American Statistician* 49, 327-335.

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

[10] Conley, T.G., L.P. Hansen, E.G.J. Luttmer, J.A. Scheinkman,1997, Short-Term Interest Rates as Subordinated Diffusions, *The Review of Financial Studies*, Vol 10, No 3, 525-577

[11] Courtadon, G., 1982, The Pricing of Options on Default-Free Bonds, *Journal of Financial and Quantitative Analysis*, Vol XVII, No1, 75-100

[12] Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985, A Theory of the Term Structure of Interest Rates, *Econometrica*, Vol53, No.2, 385-407.

[13] Dahlquist, M., 1996, On Alternative Interest Rate Processes, *Journal of Banking and Finance* 20, 1093-1119.

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

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

[16] Eraker, B. 2001, MCMC Analysis of Diffusion Models with Application to Finance, *Journal of Business & Economic Statistics*, April, Vol 19, No.2,177-191

[17] Han, C and B.P. Carlin, 2001, Markov Chain Monte Carlo Methods for Computing Bayes Factors : A Comparative Review, *Journal of the American Statistical Association*, Vol 96, No 455

[18] Hurn, A.S., K.A. Lindsay and V.L. Martin, 2003, On the Efficiency of Simulated Maximum Likelihood for Estimating the Parameters of Stochastic Differential Equations, *Journal of Time Series Analysis*, Vol 24, No 1, 45-63

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

[20] Jones, C., 2003, Nonlinear Mean Reversion in the Short-Term Interest Rate, *The Review of Financial Studies*, Vol 16, No 3. 793-843.

[21] Kass, R.E. and A.E. Raftery, 1995, Bayes Factors, *Journal of the American Statistical Association*, Vol 90, No 430, 773-795.

[22] Kim, S., N. Shephard, and S. Chib, 1998, Stochastic Volatility, Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies*, 65, 361-393.

[23] Koop G., and S.M. Potter, 1999, Bayes Factors and nonlinearity: Evidence from economic time series, *Journal of Econometrics* 88, 251-281

[24] Longstaff, F.A. and E.S. Schwartz, 1992, Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model, *Journal of Finance* 47, 1259-1282.

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

[26] Sanford, A., and G. Martin, 2004, Simulation-Based Bayesian Estimation of an Affine Term Structure Model. Forthcoming, *Computational Statistics and Data Analysis*, *Special Issue on Computational Econometrics.*

[27] Shephard, N and M.K. Pitt, 1997, Likelihood Analysis of Non-Gaussian Measurement Time Series, *Biometrika*, Vol 84, No 3., 653-667.

[28] Schotman P. and H.K. van Dijk, 1991, A Bayesian Analysis of the Unit Root in Real Exchange Rates, *Journal of Econometrics*, 49, 195-238.

[29] Treepongkaruna, S., and S. Gray, 2003a, On the robustness of short-term interest rate models, *Accounting and Finance* 43, 87-121.

[30] Treepongkaruna, S., and S. Gray, 2003b, Short-term interest rate models: valuing interest rate derivatives using Monte-Carlo approach, *Accounting and Finance* 43, 231-259.

[31] Verdinelli, I., and L. Wasserman, 1995, Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio, *Journal of the American Statistical Association*, Vol 90, No 430, 614618.

Table 1: Interpretation of Bayes factors

BF_{j0}	Evidence against M_0 and supporting M_j
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive

Table 2: Alternative priors for robustness analysis

	$\bar{\nu}$	$\bar{\nu}\bar{s}^2$	$[\overline{\theta}, \overline{k}]$	A
Prior 1	1	0.01	$[1.0, -0.5]$	$0.0001 \times I_2$
Prior 2	1	0.01	$[5, -2.5]$	$0.0001 \times I_2$
Prior 3	1	0.01	$[1.0, -0.5]$	$0.1 \times I_2$
Prior 4	1	0.1	$[1.0, -0.5]$	$0.1 \times I_2$

Table 3: Summary of short rate data

Variable	Mean	Median	Standard Deviation	Skewness	Kurtosis	Max. Value	Min. Value
r_t	7.132	5.970	2.885	1.682	5.106	17.400	4.680
Δr_t	−0.020	0.000	0.140	−0.413	11.826	0.860	−0.730

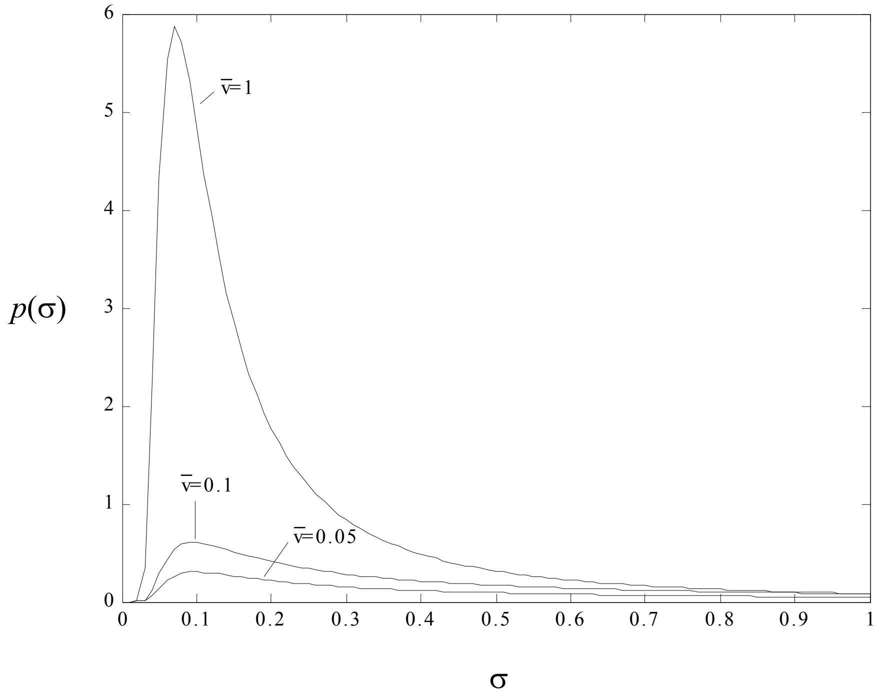

Figure 1: Inverted gamma prior distributions for σ for $\bar{v} = (1, 0.1, 0.05)$ and $\bar{v}\bar{s}^2 = 0.01$

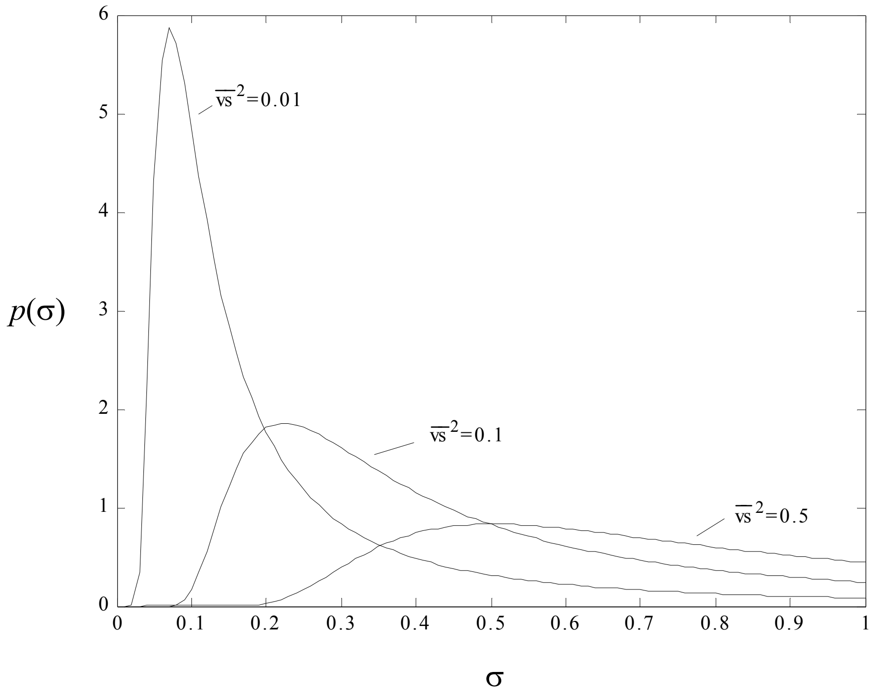

Figure 2: Inverted gamma prior densities for σ with $\bar{v} = 1$ and $\bar{v}\bar{s}^2 = (0.01, 0.1, 0.5)$.

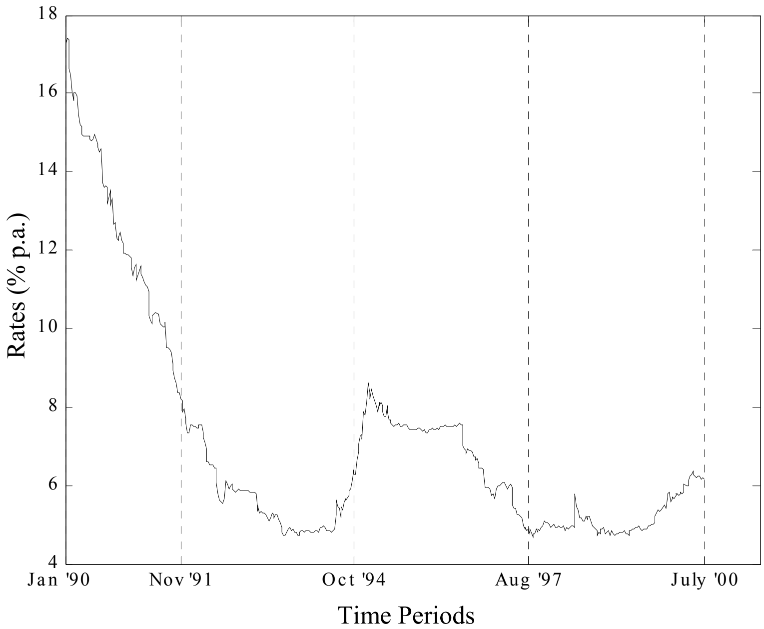

Figure 3: Australian interest rate data : Wednesday observations of 90 day rates from January 1990 to July 2000 (552 observations)

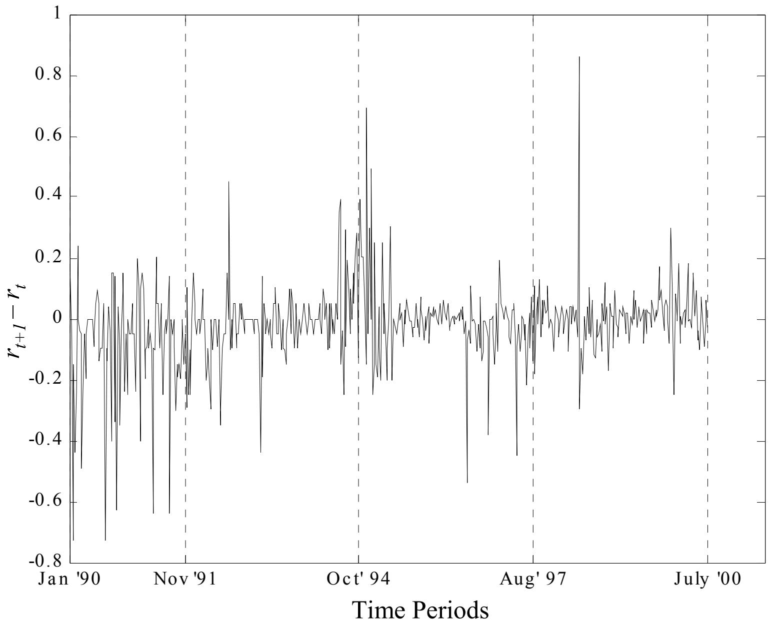

Figure 4: First differenced interest rate series

Table 4: Estimation results using Prior 1

	$h =$	Posterior Mean	Posterior Standard Deviation	Inefficiency Factor	MC Error	25th Perc.	50th Perc.	75th Perc.
k	3	−0.0124	0.0012	1.4	0.000006	−0.0132	−0.0124	−0.0116
	1	−0.0124	0.0012	0.7	0.000004	−0.0131	−0.0124	−0.0117
	0	−0.0123	0.0011	0.8	0.000004	−0.0130	−0.0123	−0.0116
θ	3	0.0685	0.0084	1.3	0.000044	0.0630	0.0685	0.0740
	1	0.0682	0.0078	0.7	0.000029	0.0631	0.0682	0.0733
	0	0.0677	0.0074	0.8	0.000029	0.0629	0.0677	0.0724
σ	3	0.0362	0.0055	480.5	0.000536	0.0322	0.0358	0.0398
	1	0.0381	0.0057	189.5	0.000352	0.0341	0.0378	0.0417
	0	0.0400	0.0061	83.2	0.000249	0.0357	0.0395	0.0437
δ	3	0.669220	0.077553	512.6	0.007852	0.614460	0.668460	0.723660
	1	0.642160	0.077064	209.6	0.004990	0.589930	0.640480	0.694350
	0	0.617080	0.077549	84.6	0.003190	0.566130	0.617570	0.670130

Table 5: Estimation results using Prior 2

	h =	Posterior Mean	Posterior Standard Deviation	Inefficincy Factor	MC Error	25th Perc.	50th Perc.	75th Perc.
k	3	−0.0124	0.0015	0.9	0.000006	−0.0134	−0.0124	−0.0115
	1	−0.0124	0.0015	1.2	0.000007	−0.0133	−0.0124	−0.0115
	0	−0.0123	0.0013	1.2	0.000006	−0.0132	−0.0123	−0.0115
θ	3	0.0685	0.0099	0.9	0.000042	0.0621	0.0685	0.0749
	1	0.0684	0.0099	1.2	0.000049	0.0621	0.0683	0.0746
	0	0.0679	0.0088	1.2	0.000043	0.0623	0.0678	0.0735
σ	3	0.0316	0.0048	464.8	0.000463	0.0283	0.0312	0.0346
	1	0.0321	0.0056	315.0	0.000442	0.0281	0.0318	0.0357
	0	0.0346	0.0057	77.5	0.000224	0.0306	0.0341	0.0381
δ	3	0.739660	0.078800	502.3	0.007902	0.685650	0.739590	0.790900
	1	0.732560	0.090700	310.5	0.007149	0.668250	0.728940	0.792140
	0	0.692630	0.085400	82.4	0.003467	0.635120	0.692590	0.749130

Table 6: Estimation results using Prior 3

	$h=$	Posterior Mean	Posterior Standard Deviation	Inefficiency Factor	MC Error	25th Perc.	50th Perc.	75th Perc.
k	3	−0.0133	0.0006	34.2	0.000015	−0.0136	−0.0132	−0.0129
	1	−0.0132	0.0006	14.3	0.000010	−0.0135	−0.0131	−0.0128
	0	−0.0130	0.0006	4.9	0.000006	−0.0134	−0.0130	−0.0126
θ	3	0.0746	0.0043	30.3	0.000107	0.0716	0.0744	0.0774
	1	0.0739	0.0043	12.7	0.000068	0.0710	0.0737	0.0766
	0	0.0729	0.0041	4.5	0.000039	0.0702	0.0727	0.0755
σ	3	0.0672	0.0067	256.9	0.000478	0.0625	0.0666	0.0715
	1	0.0683	0.0076	102.9	0.000344	0.0632	0.0677	0.0727
	0	0.0693	0.0075	30.1	0.000184	0.0640	0.0688	0.0740
δ	3	0.359660	0.049700	285.2	0.003756	0.324800	0.361530	0.3929
	1	0.350300	0.055000	108.3	0.002562	0.315840	0.351370	0.3864
	0	0.341920	0.054100	32.7	0.001385	0.305720	0.342020	0.3791

Table 7: Estimation results using Prior 4

	$h =$	Posterior Mean	Posterior Standard Deviation	Inefficiency Factor	MC Error	25th Perc.	50th Perc.	75th Perc.
k	3	−0.0130	0.0005	24.6	0.000010	−0.0133	−0.0130	−0.0127
	1	−0.0129	0.0005	9.7	0.000007	−0.0132	−0.0129	−0.0126
	0	−0.0128	0.0004	4.1	0.000004	−0.0131	−0.0128	−0.0125
θ	3	0.0726	0.0035	21.6	0.000073	0.0702	0.0725	0.0749
	1	0.0721	0.0035	8.6	0.000046	0.0698	0.0720	0.0744
	0	0.0713	0.0034	3.7	0.000029	0.0690	0.0712	0.0735
σ	3	0.0808	0.0081	254.1	0.000576	0.0752	0.0802	0.0860
	1	0.0812	0.0081	102.5	0.000368	0.0757	0.0806	0.0860
	0	0.0816	0.0083	38.1	0.000230	0.0758	0.0810	0.0868
δ	3	0.270200	0.049700	270.8	0.003656	0.236150	0.271070	0.304010
	1	0.267060	0.049900	107.5	0.002314	0.236040	0.267310	0.299320
	0	0.263420	0.050900	41.9	0.001474	0.229440	0.263880	0.297790

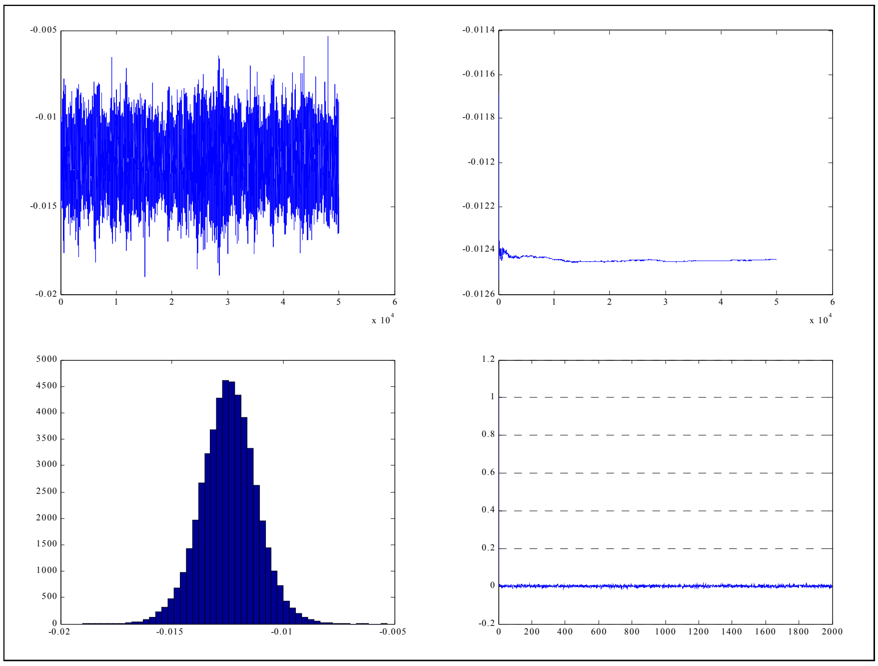

Figure 5: Graphical outputs for k estimated using Prior 1 settings: time series (top left); cumulative mean (top right); histogram (bottom left); and autocorrelation function (bottom right).

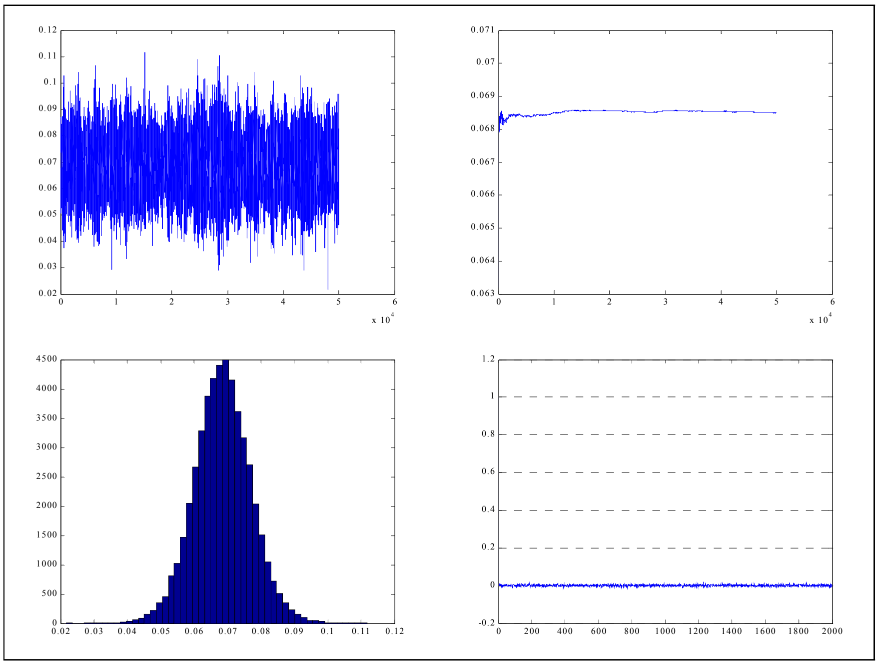

Figure 6: Graphical outputs for θ estimated using Prior 1 settings: time series (top left); cumulative mean (top right); histogram (bottom left); and autocorrelation function (bottom right).

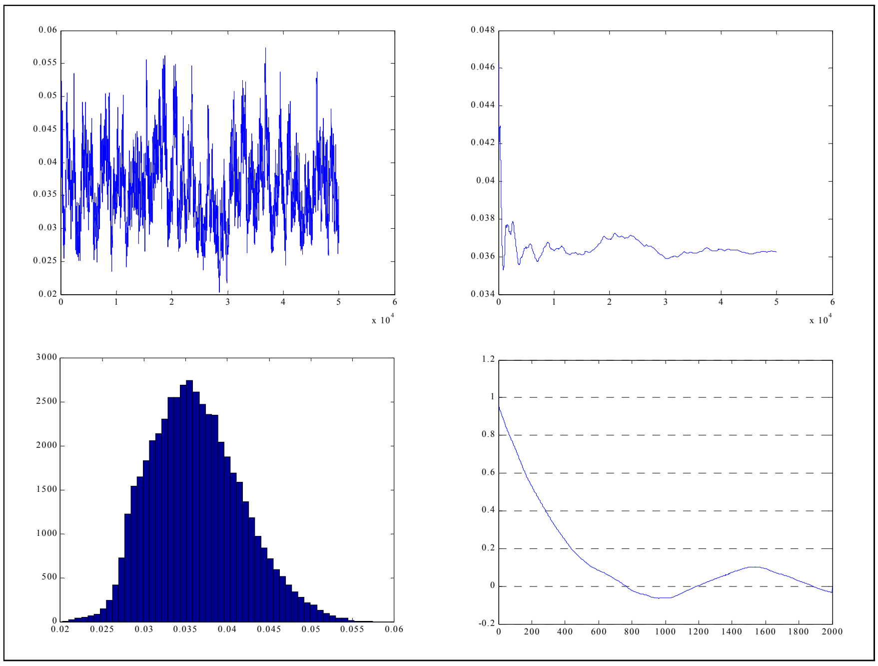

Figure 7: Graphical outputs for σ estimated using Prior 1 settings: time series (top left); cumulative mean (top right); histogram (bottom left); and autocorrelation function (bottom right).

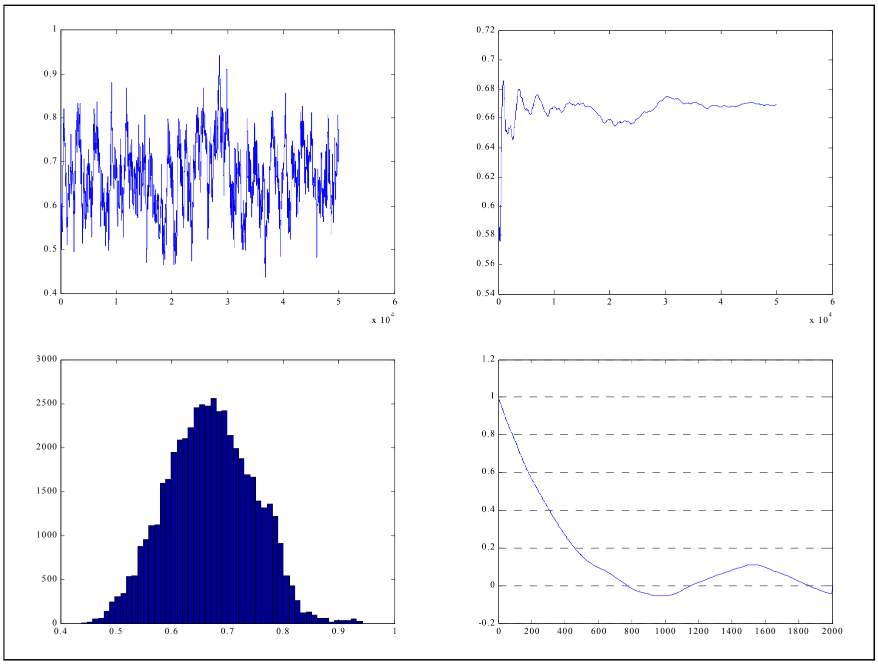

Figure 8: Graphical outputs for δ estimated using Prior 1 settings: time series (top left); cumulative mean (top right); histogram (bottom left); and autocorrelation function (bottom right).

Table 8: Bayes factors for all nested models against the unrestricted model

Model		Prior 1		Prior 2	
		BF_{j0}	MC Error	BF_{j0}	MC Error
	$h =$				
M_1	3	4.84×10^{-14}	1.05×10^{-14}	2.19×10^{-11}	1.70×10^{-11}
	1	7.73×10^{-13}	2.80×10^{-13}	2.02×10^{-13}	3.73×10^{-14}
	0	4.19×10^{-13}	4.70×10^{-14}	2.83×10^{-13}	2.90×10^{-14}
M_2	3	2.022900	0.080555	0.501870	0.054235
	1	2.747100	0.035273	0.483950	0.013306
	0	4.109200	0.026292	1.063300	0.013438
M_3	3	0.001062	0.000166	0.075787	0.003760
	1	0.000573	0.000060	0.125780	0.004820
	0	0.000084	0.000005	0.019627	0.000659
M_4	3	5.98×10^{-23}	2.27×10^{-23}	1.50×10^{-14}	1.16×10^{-14}
	1	5.79×10^{-23}	2.99×10^{-23}	3.68×10^{-15}	5.42×10^{-16}
	0	7.86×10^{-27}	3.21×10^{-27}	2.23×10^{-18}	3.78×10^{-19}

		Prior 3		Prior 4	
		BF_{j0}	MC Error	BF_{j0}	MC Error
	$h =$				
M_1	3	7.08×10^{-11}	1.79×10^{-11}	0.000006	4.19×10^{-7}
	1	6.65×10^{-8}	1.56×10^{-8}	0.000059	0.000008
	0	3.07×10^{-8}	1.86×10^{-9}	0.000031	6.86×10^{-7}
M_2	3	0.365730	0.009569	0.000354	0.000025
	1	0.457250	0.013030	0.000645	0.000021
	0	0.337430	0.009092	0.000381	0.000012
M_3	3	3.93×10^{-35}	1.44×10^{-35}	5.49×10^{-46}	2.82×10^{-46}
	1	1.53×10^{-27}	8.21×10^{-28}	5.41×10^{-41}	4.12×10^{-41}
	0	1.22×10^{-30}	2.48×10^{-31}	1.57×10^{-44}	6.41×10^{-45}
M_4	3	9.35×10^{-110}	8.88×10^{-110}	9.59×10^{-130}	7.03×10^{-130}
	1	1.32×10^{-84}	1.08×10^{-84}	2.07×10^{-112}	2.01×10^{-112}
	0	1.08×10^{-95}	6.09×10^{-96}	1.54×10^{-124}	1.39×10^{-124}

Table 9: Bayes factors for all nested nodels

Entry (k, j) in each panel indicates the Bayes factor in Favour of M_j Versus M_k, BF_{jk}

	Prior 1			
	M_1	M_2	M_3	M_4
M_k				
M_1	1.0	4.18×10^{13}	2.19×10^{10}	1.26×10^{-9}
M_2	2.39×10^{-14}	1.0	5.25×10^{-4}	2.96×10^{-23}
M_3	4.57×10^{-11}	1.91×10^{3}	1.0	5.63×10^{-20}
M_4	7.94×10^{8}	3.38×10^{22}	1.78×10^{19}	1.0

	Prior 2			
	M_1	M_2	M_3	M_4
M_k				
M_1	1.0	2.29×10^{10}	3.46×10^{9}	6.39×10^{-4}
M_2	4.36×10^{-11}	1.0	0.1510	2.99×10^{-14}
M_3	2.89×10^{-10}	6.6221	1.0	1.98×10^{-13}
M_4	1.56×10^{3}	3.35×10^{13}	5.05×10^{12}	1.0

	Prior 3			
	M_1	M_2	M_3	M_4
M_k				
M_1	1.0	5.17×10^{9}	5.55×10^{-25}	1.32×10^{-99}
M_2	1.94×10^{-10}	1.0	1.08×10^{-34}	2.56×10^{-109}
M_3	1.80×10^{24}	9.31×10^{33}	1.0	2.38×10^{-75}
M_4	7.57×10^{98}	3.91×10^{108}	4.20×10^{74}	1.0

	Prior 4			
	M_1	M_2	M_3	M_4
M_k				
M_1	1.0	59.0	9.15×10^{-41}	1.60×10^{-124}
M_2	0.0169	1.0	1.55×10^{-42}	2.71×10^{-126}
M_3	1.09×10^{40}	6.45×10^{41}	1.0	1.75×10^{-84}
M_4	6.26×10^{123}	3.69×10^{125}	5.72×10^{83}	1.0

Table 10: Posterior probabilities for all models.

		M_0	M_1	M_2	M_3	M_4
	$h =$					
Prior1	3	0.331	0.0	0.669	0.0	0.0
	1	0.277	0.0	0.733	0.0	0.0
	0	0.196	0.0	0.804	0.0	0.0
Prior 2	3	0.634	0.0	0.318	0.048	0.0
	1	0.621	0.0	0.300	0.078	0.0
	0	0.480	0.0	0.511	0.009	0.0
Prior 3	3	0.732	0.0	0.268	0.0	0.0
	1	0.686	0.0	0.314	0.0	0.0
	0	0.748	0.0	0.252	0.0	0.0
Prior 4	3	0.999	0.0	0.001	0.0	0.0
	1	0.999	0.0	0.001	0.0	0.0
	0	0.999	0.0	0.001	0.0	0.0