A Bayesian Comparison of Several Continuous Time Models of the Australian Short Rate∗†

Andrew D. Sanford‡ and Gael M. Martin§

∗The authors would like to thank the Editor and a referee for some very thoughtful and constructive comments on an earlier draft of the paper. This research has been supported by 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.

†JEL Codes: C11, C15, C52, E43. Keywords: Bayesian analysis; Interest rate models; Bayes factors; Markov chain Monte Carlo; Data augmentation.

‡School of Business Systems and Department of Accounting and Finance, Monash University. Corresponding author: E-mail address: Andrew.Sanford@buseco.monash.edu.au; Postal Address: Department of Accounting and Finance, Faculty of Business and Economics, Monash University, Clayton 3800, Australia; Phone: 9905-4938; Fax: 9905-5475.

§Department of Econometrics and Business Statistics, Monash University.
A Bayesian Comparison of Several Continuous Time Models of the Australian Short Rate

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 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. 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. 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 models considered.
1 Introduction

Correct modelling of the instantaneous short rate is of particular importance in finance, as it is this rate that is so fundamental to the pricing of fixed-income securities. Although many alternative model specifications for the short rate process have been proposed, 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 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.

The empirical results relating specifically to Australian rates have also been mixed. For example, Tse (1995), using monthly observations of 90 day money market rates, estimates values for the level parameter that are supportive of the square root diffusion model. Results reported by Li (2000) are also supportive of the square root model. In contrast, the Australian studies of Gray (1996), Brailsford and Maheswaran (1998) and Gray (2005) estimate higher values for the level parameter. 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
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. In a similar spirit to the independent work of Gray (2005), we adopt a Bayesian inferential approach. This is in contrast with the other Australian analyses cited above, in which classical inferential methods have been used. In adopting a Bayesian approach, uncertainty regarding all unknown quantities is quantified via posterior probabilities, which reflect both sample and prior information about the interest rate process. In particular, and in contrast with Gray, probabilities are produced not just for the unknown parameters, but also for alternative models, with the models ranked according to the magnitudes of the probabilities so assigned. 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. Crucially, we also exploit the ability of the Bayesian approach to augment the set of unknowns to include latent (or unobserved) random variables. Specifically, we reduce the bias associated with estimating continuous time interest rate models with discretely observed data via the introduction of higher frequency latent data in between each pair of successive discrete time observations.\footnote{Jones (1998, 2003), Elerian, Chib and Shephard (2001) and Eraker (2001) provide detailed expositions of this methodology, whilst Gray (2005) and Sanford and Martin (2005) employ simulation experiments to gauge the effectiveness of augmentation in a univariate interest rate model and term structure model respectively.}

Estimation and model selection is performed using a hybrid Gibbs/Metropolis-Hastings (MH) 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 Bank Acceptance Bill (BAB) 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 algorithm 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 essentially
assigned zero posterior probability. Some conclusions are provided in Section 5.

2 The Short Rate 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),

\[
dr_t = (\theta + kr_t) \, dt + \sigma r_t^\delta \, dW_t,
\]

where \(-k, \mu = -(\theta/k), \sigma \) and \( \delta \) 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 \( \delta \), 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 a model with the value of \( \delta \) as estimated by
Chan et al. (1992) for U.S. data imposed. The high value of \( \delta \) in \( M_4 \) is also broadly consistent with
empirical results reported in the Australian analyses of Gray (1996), Brailsford and Maheswaran
(1998) and Gray (2005). We denote the general, unrestricted model, in which \( \delta \) 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

\[
r_{t+\Delta t} - r_t = \left( \theta + kr_t \right) \Delta t + \sigma r_t^\delta \sqrt{\Delta t} \varepsilon_t,
\]

where \( \varepsilon_t \sim i.i.d.N(0,1) \) and \( \Delta t \) represents the time between each observation. When estimating
the parameters of (2), the interval \( \Delta t \) 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 $\Delta t$ 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, with particular emphasis given to the production of the Bayes factors and posterior model probabilities that underlie the model selection process; see Zellner (1996) and Koop (2003) for more comprehensive treatments. As is detailed below, Bayes Theorem is initially used to produce the posterior probability distribution of the parameters and/or unobserved latent variables of the $j$th model, $M_j$, conditional on the observed data and the model. Posterior probabilities are then assigned to each model in the model set. Given the assignment of equal prior probabilities to each model, the posterior probabilities measure the support in the data for the respective models, with the model with the highest probability ‘selected’ accordingly.

Denoting by $r^{obs}$ the vector of observations on the short rate, and by $\phi_j$ the set of unknowns associated with model $M_j$, the joint probability density function (pdf) of $r^{obs}$ and $\phi_j$, conditional on $M_j$, can be expressed as

$$p(r^{obs}, \phi_j | M_j) = p(r^{obs} | \phi_j, M_j) \times p(\phi_j | M_j)$$

A simple rearrangement of the expression in (3) leads to

$$p(\phi_j | r^{obs}, M_j) \propto L(\phi_j | M_j) \times p(\phi_j | M_j),$$

whereby the posterior pdf of $\phi_j$, conditioned on the observed data, $r^{obs}$, $p(\phi_j | r^{obs}, M_j)$, is equated with the product of the pdf of the data generating process under $M_j$, $p(r^{obs} | \phi_j, M_j)$, and the prior pdf for $\phi_j$, $p(\phi_j | M_j)$, normalized by the marginal likelihood of the data under $M_j$,

$$p(r^{obs} | M_j) = \int_{\phi_j} p(r^{obs} | \phi_j, M_j) p(\phi_j | M_j) d\phi_j.$$ 

The equality in (4) is referred to as Bayes Theorem, often equivalently written as

$$p(\phi_j | r^{obs}, M_j) \propto L(\phi_j | M_j) \times p(\phi_j | M_j),$$

6
where \( L(\phi_j \mid M_j) \propto p(r^{\text{obs}}|\phi_j, M_j) \) denotes the likelihood function for \( \phi_j \). As has already been alluded to, and as will be further clarified in subsequent sections, \( \phi_j \) 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 in (6) encompasses all information about this full set of unknowns, both as derives from the data, via \( L(\phi_j \mid M_j) \), and from prior knowledge, as expressed in \( p(\phi_j \mid M_j) \).

Incorporation into the analysis of model uncertainty, in addition to parameter (or latent variable) uncertainty, entails the manipulation of the joint distribution of \( r^{\text{obs}}, \phi_j \) and \( M_j \). Similar decompositions to those used in (3) above (see Zellner, 1996 Chp.10 for details), lead to the following expression for the posterior probability for each model,

\[
P\left( M_j \mid r^{\text{obs}} \right) = \frac{P\left( r^{\text{obs}} \mid M_j \right) \times P\left( M_j \right)}{p\left( r^{\text{obs}} \right)},
\]

where \( P\left( M_j \right) \) denotes the prior probability associated with \( M_j \), with \( P(M_0) + P(M_1) + \cdots + P(M_4) = 1 \), and

\[
p\left( r^{\text{obs}} \right) = \sum_{j=0}^{4} \left[ P\left( r^{\text{obs}} \mid M_j \right) \times P\left( M_j \right) \right],
\]

with \( p\left( r^{\text{obs}} \mid M_j \right) \) as defined in (5). Given (7), the ratio of posterior probabilities for \( M_j \) and \( M_k \), referred to as the posterior odds ratio for \( M_j \) versus \( M_k \), can be expressed as

\[
\frac{P\left( M_j \mid r^{\text{obs}} \right)}{P\left( M_k \mid r^{\text{obs}} \right)} = \frac{P\left( M_j \right)}{P\left( M_k \right)} \times \frac{p\left( r^{\text{obs}} \mid M_j \right)}{p\left( r^{\text{obs}} \mid M_k \right)}, \quad j \neq k = 0, 1, \ldots, 4.
\]

Given the assumption of equal prior probabilities, \( P\left( M_j \right) = P\left( M_k \right), \quad j \neq k = 0, 1, \ldots, 4 \), the expression in (8) collapses to the ratio of marginal likelihoods, which is known as the Bayes factor,

\[
BF_{jk} = \frac{p\left( r^{\text{obs}} \mid M_j \right)}{p\left( r^{\text{obs}} \mid M_k \right)}, \quad j \neq k = 0, 1, \ldots, 4.
\]

The Bayes factor, in measuring the ratio of the ordinates of the pdfs of the data generating processes under \( M_j \) and \( M_k \) respectively, at the observed data \( r^{\text{obs}} \), provides a measure of the support in the data for \( M_j \) relative to \( M_k \). In what follows, all Bayes factors are expressed relative to the general model, \( M_0 \), in which the level parameter \( \delta \) is left unrestricted. Hence, we compute

\[
BF_{j0} = \frac{p\left( r^{\text{obs}} \mid M_j \right)}{p\left( r^{\text{obs}} \mid M_0 \right)}, \quad j = 1, 2, \ldots, 4,
\]

with each Bayes factor measuring the support in the data for the particular restriction being imposed on \( \delta \).
Table 1 contains a useful aid, reproduced from Kass and Raftery (1995), and based on criteria first proposed by Harold Jeffreys, for the interpretation of Bayes factors. Given a finite model set, however, a more direct assessment of the relative support in the data for the alternative models occurs via the posterior model probabilities. On the assumption of equal prior probabilities for all models $j = 0, 1, \ldots, 4$, and given the restriction that all five posterior probabilities add to one, the posterior probability for each model can be readily produced from the Bayes factors as

$$p(M_j | x^{obs}) = \frac{BF_{j0}}{\sum_{k=0}^{4} BF_{k0}}, \; j = 0, 1, \ldots, 4,$$

where $BF_{j0} = 1$ for $j = 0$. The models are ranked according to the relative magnitudes of the probabilities in (10).

As is clear from the expression in (5), calculation of the marginal likelihood for any given model and, hence, calculation of the Bayes factor in (9) for each value of $j$, may be difficult because of the need to evaluate a complex integral involving a large number of parameters and latent variables. We employ a simple representation of (9) based on the Savage-Dickey density ratio; see Verdinelli and Wasserman (1995). Details are provided in Appendix A.

### 3.2 Augmentation of the Short Rate Data

As previously noted, 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), with applications of the methodology to interest rate models appearing in Gray (2005) and Sanford and Martin (2005). 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 approximation of 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 are treated as latent variables that are ultimately integrated out of the joint posterior via the MCMC algorithm.

The actual short rate data is assumed to be observed at time points \( t = 1 \) to \( t = \tilde{T} \). Hence, the vector of observations, \( r_{\text{obs}} \), defined in the previous section, can be expressed explicitly as

\[
 r_{\text{obs}} = [r_{1}^{\text{obs}}, r_{2}^{\text{obs}}, \ldots, r_{\tilde{T}}^{\text{obs}}]' . \tag{11}
\]

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 \(((\tilde{T} - 1) \times h) \times 1\) vector,

\[
 r_{\text{aug}} = [r_{1+\Delta t}^{\text{aug}}, r_{1+2\Delta t}^{\text{aug}}, \ldots, r_{1+h\Delta t}^{\text{aug}}, r_{2+\Delta t}^{\text{aug}}, \ldots, r_{2+h\Delta t}^{\text{aug}}, \ldots, r_{\tilde{T}-1+h\Delta t}^{\text{aug}}]' . \tag{12}
\]

Combining the two vectors (11) and (12), the complete data set is given by

\[
 r = [r_{1}^{\text{obs}}, r_{1+\Delta t}^{\text{aug}}, r_{1+2\Delta t}^{\text{aug}}, \ldots, r_{1+h\Delta t}^{\text{aug}}, r_{2}^{\text{obs}}, r_{2+\Delta t}^{\text{aug}}, \ldots, r_{2+h\Delta t}^{\text{aug}}, r_{3}^{\text{obs}}, r_{3+\Delta t}^{\text{aug}}, \ldots, r_{\tilde{T}-1+h\Delta t}^{\text{aug}}, r_{\tilde{T}}^{\text{obs}}]' , \tag{13}
\]

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

\[
 r = [r_{1}, r_{2}, r_{3}, \ldots, r_{t-1}, r_{t}, r_{t+1}, \ldots, r_{T-1}, r_{T}]' , \tag{14}
\]

where the \( t \) subscript in (14) indicates the \( t \)th 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 Savage-Dickey density ratio representation of the Bayes factor in (A1), all four Bayes factors are based on estimation of the unrestricted version of the model, \( M_{0} \), with the marginal prior and posterior of \( \delta \) then evaluated at the values associated with the four nested models. Hence, all estimation details relate explicitly
to $M_0$, with all distributions below implicitly conditioned on $M_0$ as a consequence. For notational
simplicity, however, we dispense with this conditioning.

Defining $\omega = [\theta, k, \sigma]^T$, the joint posterior density for the full set of unknowns for $M_0$ can be
expressed as

$$p(r_{\text{aug}}, \omega, \delta \mid r_{\text{obs}}) \propto \prod_{t=1}^{T-1} p(r_{t+1} \mid r_t, \omega, \delta) p(r_{\text{aug}}) p(\omega) p(\delta),$$  

(15)

where the elements $r_{\text{aug}}$, $\omega$ and $\delta$ are assumed to be a priori independent, with $p(r_{\text{aug}})$, $p(\omega)$
and $p(\delta)$ denoting respectively the associated prior pdfs. The product of the component densities
$p(r_{t+1} \mid r_t, \omega, \delta)$, $t = 1, 2, \ldots, T$, in (15) defines the joint distribution for the full vector $r$, where
$r$ comprises both observed and augmented data. In terms of the notation used in the previous
section, this joint distribution represents the likelihood function for the unknowns associated with
model $M_0$, $L(\phi_0 \mid M_0)$, whilst the product of priors represents $p(\phi_0 \mid M_0)$.

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 set 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 $\delta$, be assigned a proper prior. We choose to use a truncated uniform prior
for $\delta$, $\delta \sim U(-0.5, 2.0)$, 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 $\delta$. Although not strictly
necessary we have also opted to use proper priors for the vector of nuisance parameters, $\omega$, namely
a conjugate normal-inverted gamma prior; see Zellner (1996). The prior is parameterized in a way
that implies very diffuse prior information on $w_0$. The prior for $r_{\text{aug}}$ is assumed to be uniform.

The joint posterior in (15) is intractable. Marginal posterior densities for the parameters of
interest cannot be obtained either by analytical means or by sampling directly from the joint
distribution. Instead, we make use of a Gibbs-based MCMC algorithm, whereby simulation from
the joint posterior occurs indirectly via iterative simulation from the conditional posteriors for each
of the unknown components $r_{\text{aug}}$, $\omega$, and $\delta$; see Chib and Greenberg (1995, 1996) for an exposition
of MCMC schemes. To begin, we consider the conditional posterior for a single element of $r_{\text{aug}},$
$r_{r_{\text{aug}}}$, $t = 1 + \Delta t, \ldots, 1 + h\Delta t, \ldots, \tilde{T} - 1 + \Delta t, \ldots, \tilde{T} - 1 + h\Delta t,$

\[2\text{See Kass and Raftery (1995) for an exposition of the impact of prior specification on Bayes factors.}\]
\[
p\left(r_{\tau}^{\text{aug}} \mid r_{\tau}^{\text{aug}}, \omega, \delta, r^{\text{obs}}\right) = \frac{p\left(r_{\tau} \mid r_{\tau} + \Delta t, r_{\tau} - \Delta t, \omega, \delta\right)}{p\left(r_{\tau} + \Delta t \mid r_{\tau}^{\text{aug}}, \omega, \delta\right) p\left(r_{\tau}^{\text{aug}} \mid r_{\tau} - \Delta t, \omega, \delta\right)}, \\
\text{(16)}
\]

where \(r_{\tau}^{\text{aug}}\) denotes the vector of all augmented data other than \(r_{\tau}^{\text{aug}}\). Given the Markovian nature of the model in (2), the conditional posterior in (16) is a function only of the two elements of the vector \(r\) that immediately precede and follow \(r_{\tau}^{\text{aug}}, r_{\tau} - \Delta t\) 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 \(\tau\).

For the parameter vector \(\omega\), the conditional posterior is given by

\[
p\left(\omega \mid r_{\tau}^{\text{aug}}, \delta, r^{\text{obs}}\right) \propto \prod_{t=1}^{T-1} p\left(r_{t+1} \mid r_{t}, \omega, \delta\right) p\left(\omega\right), \\
\text{(17)}
\]

whilst the conditional posterior for \(\delta\) is defined by

\[
p\left(\delta \mid r_{\tau}^{\text{aug}}, \omega, r^{\text{obs}}\right) \propto \prod_{t=1}^{T-1} p\left(r_{t+1} \mid r_{t}, \omega, \delta\right) p\left(\delta\right). \\
\text{(18)}
\]

The Gibbs-based sampling scheme is implemented by sampling iteratively from each of the full conditionals (16), (17) and (18), until convergence. The latent augmented short rate data is initialized by linear interpolation between the observed rates. All parameters are initialized using perturbed values of previously published empirical results. When the full conditional is a known, closed-form distribution, then standard sampling algorithms are available. When this is not the case, we sample from the full conditional using an MH algorithm. As described above, the data set is augmented with the higher frequency latent data in order to allow \(\Delta t\) to become smaller than the value associated with the observed data. The trade off associated with using greater augmentation is that as \(\Delta t \to 0\) convergence will occur more slowly. As Eraker (2001) points out, in the application of Gibbs sampling to discretized SDE’s, in the limit, as \(\Delta t \to 0\), 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. All additional computational details associated with the implementation of Steps 3, 4 and 5 are available from the corresponding author on request. Further technical details relating to convergence of the algorithm are also available on request.
1. Specify initial values $\omega^{(0)}$, $\delta^{(0)}$ and $r_{\tau}^{\text{aug}(0)}$, $\tau = 1 + \Delta t, \ldots, 1 + h\Delta t, \ldots, \tilde{T} - 1 + \Delta t, \ldots, \tilde{T} - 1 + h\Delta t$;

2. Set $i = 1$;

3. Sample the latent augmented short rate variable $r_{\tau}^{\text{aug}(i)}$ from the full conditional $p\left(r_{\tau}^{\text{aug}(i)} | r_{\tau}^{\text{aug}(i-1)}, \omega^{(i-1)}, \delta^{(i-1)}, r_{\text{obs}}^{\tau}\right)$; $\tau = 1 + \Delta t, \ldots, 1 + h\Delta t, \ldots, \tilde{T} - 1 + \Delta t, \ldots, \tilde{T} - 1 + h\Delta t$;

4. Sample $\omega^{(i)}$ from the full conditional $p\left(\omega^{(i)} | r^{\text{aug}(i)}, \delta^{(i-1)}, r_{\text{obs}}\right)$;

5. Sample the volatility exponent $\delta^{(i)}$ from the full conditional $p\left(\delta^{(i)} | r^{\text{aug}(i)}, \omega^{(i)}, r_{\text{obs}}\right)$;

6. Estimate the Bayes factor for $M_j$, $j = 1, 2, \ldots, 4$, versus the unrestricted model, $M_0$, as detailed in Appendix B;

7. Set $i = i + 1$;


The Monte Carlo (MC) standard error is calculated for each parameter in the manner detailed by Kim, Shephard and Chib (1998), amongst others. It measures the accuracy of the simulation-based estimate of the posterior mean of the parameter, taking into account the correlation in the chain 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 BAB 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 BAB rate interest rather than the shorter 30 day rate as a proxy for the instantaneous short rate. Treepongkaruna and Gray comment that their use of the 90 day rate is motivated by its high liquidity. The interest rate data

---

3The data was provided to the corresponding author by the treasury unit of a major Australian bank.
and the first differenced series are presented in Figure 1. Summary statistics for both sets of data are provided in Table 2. The skewness and kurtosis statistics reported therein are sample estimates of the third and fourth moments respectively of the standardized random variable.

<<Insert Figure 1 here>>

<<Insert Table 2 here>>

From Figure 1 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. Indeed, the increased volatility associated with the high interest rate period of January 1990 to November 1991 also coincides with a marked downward shift in rates over this same time period. 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’ may be important in understanding the volatility of interest rates, in addition to the level effect being explicitly modelled here.

The time-varying nature of the volatility that is evident in Figure 1 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 2 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 empirical results are reported in Table 3. 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 (12) of 3, 1 and 0 respectively, corresponding, in turn, to values for \( \Delta t \) 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.

The first thing to note is the relative stability of both the location estimates (marginal posterior mean and 50th percentile, or median) and the posterior standard deviations, over the different values for \( h \). Percentiles for the drift parameters, \( \theta \) and \( k \), show that the iterates are evenly dispersed above and below the mean, with the median coinciding closely with the estimated mean, for each value of \( h \). The estimates of the mean reversion parameter \( k \) all imply a high persistence parameter of 0.99 for the weekly short rate data, which tallies with the near unit root behaviour evident in Figure 1. The long run mean of the short rate as implied by the estimates of \( k \) and \( \theta \) is approximately 5.5%.

The point estimates of \( \delta \) 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 \( \delta \) values that are consistent with those reported here. Our estimates of \( \delta \) are also broadly consistent with the estimates of 0.676 and 0.824 produced respectively by Tse (1995) and Li (2000) using Australian data.

In contrast, however, Chan et al. (1992) and Brenner, Harjes and Kroner (1996), estimate respective values for \( \delta \) of 1.500 and 1.559, using US data. Similarly, Gray (1996), Brailsford and Maheswaran (1998), Treponkaruna and Gray (2003a) and Gray (2005) all report high estimates of \( \delta \) for Australian data, ranging from 0.929 to 1.704 depending on both the data set and estimation procedure used. It is noteworthy, however, that the Australian data sets used by all of these latter authors cover longer periods of time than does our sample of Australian data, including a more extended period of high rates. For example, the mean value of our data is 7.13%, compared with 10.62% for the Treponkaruna and Gray data set. When the latter authors conduct sub-sample
analysis of their Australian data, the estimate of $\delta$ for the second half of the sample, which overlaps closely with our sample, is 0.774, compared with the higher estimate of 0.968 for the first half of the sample. In general, Treepongkaruna and Gray 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.

<< Insert Table 3 here >>

The Bayes factors for each of the models are shown in Table 4, with the Bayes factor for model $M_j$ ($j = 1, 2, \ldots, 4$), 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 Appendix B. 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 square root model, $M_2$, for all levels of augmentation. Based on the criteria in Table 1 however, it is only when $h = 0$ that $M_2$ has any substantial dominance over $M_0$. 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 relative to $M_2$.

The results reported in Table 5, based on (10), demonstrate that to four decimal places only $M_2$ and $M_0$ have non-zero posterior probability. Hence, the results provide clear support for the Cox et al. (1985) square root diffusion model ($M_2$), whilst providing no support for the other restricted models, including the model that corresponds to the pronounced level effect reported in Chan et al. (1992), Brailsford and Maheswaran (1998) and Gray (2005), amongst others, namely $M_4$.

<< Insert Table (4) here >>

<< Insert Table (5) here >>
5 Summary and 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 models considered, whilst the Chan et al. (1992) model is essentially given no support. 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, at least for the time period under consideration here.

A Appendix A: Savage-Dickey Density Ratio Representation of $BF_{j0}$

Partition the vector of unknowns for the unrestricted model $M_0$, $\phi_0$, as

$$\phi_0 = \left[ \delta, \phi_{0/\delta} \right]',$$

where $\delta$ 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 $\delta$. The vector $\phi_{0/\delta}$ is common to all four nested models $M_1$ to $M_4$. On the condition that

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

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

$$BF_{j0} = \frac{p \left( \delta = \delta(j) \mid \mathbf{z}^{obs} \right)}{p \left( \delta = \delta(j) \right)}, \quad j = 1, 2, \ldots, 4, \quad (A1)$$

where
\[ p \left( \delta \mid r^{\text{obs}} \right) = \int p \left( \phi_0 \mid r^{\text{obs}} \right) d\phi_0/\delta \]

is the marginal posterior of \( \delta \) under \( M_0 \), and

\[ p (\delta) = \int p(\phi_0)d\phi_0/\delta \]

is the marginal prior of \( \delta \) under \( M_0 \). For more detailed expositions of this approach to the calculation of Bayes factors see Verdinelli and Wasserman (1995) and Koop and Potter (1999).

**B Appendix B: Estimation of \( BF_{j0} \) Using the MCMC Output**

Given the specification of a proper marginal prior density for \( \delta \), the denominator in (A1) can be calculated analytically. We follow Verdinelli and Wasserman (1995) by approximating the marginal posterior, \( p \left( \delta \mid r^{\text{obs}} \right) \), as a normal density function, with mean and variance calculated from the set of iterates of \( \delta \) up to and including the current iterate. For \( j = 1, 2, \ldots, 4 \), at iteration \( i \) in the MCMC algorithm, we estimate the ordinate of \( p \left( \delta \mid r^{\text{obs}} \right) \) at \( \delta^{(j)} \), \( \tilde{p}^{(i)} (\delta = \delta^{(j)} \mid r^{\text{obs}}) \), by calculating the ordinate of the approximating normal density. The Bayes factor for \( M_j \), \( j = 1, 2, \ldots, 4 \), versus the unrestricted model, \( M_0 \), is then estimated at iteration \( i \) in the MCMC algorithm as

\[ BF_{j0}^{(i)} = \frac{\tilde{p}^{(i)} (\delta = \delta^{(j)} \mid r^{\text{obs}})}{p (\delta = \delta^{(j)})}. \]  

(B1)


**References**


Table 1: Interpretation of Bayes factors

<table>
<thead>
<tr>
<th>$BF_{j0}$</th>
<th>Evidence against $M_0$ and supporting $M_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 to 3.2</td>
<td>Not worth more than a bare mention</td>
</tr>
<tr>
<td>3.2 to 10</td>
<td>Substantial</td>
</tr>
<tr>
<td>10 to 100</td>
<td>Strong</td>
</tr>
<tr>
<td>&gt; 100</td>
<td>Decisive</td>
</tr>
</tbody>
</table>

Table 2: Summary statistics for the short rate data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Standard Deviation</th>
<th>Skewness $^{(a)}$</th>
<th>Kurtosis $^{(b)}$</th>
<th>Max. Value</th>
<th>Min. Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_t^{obs}$</td>
<td>7.132</td>
<td>5.970</td>
<td>2.885</td>
<td>1.682</td>
<td>5.106</td>
<td>17.400</td>
<td>4.680</td>
</tr>
<tr>
<td>$\Delta r_t^{obs}$</td>
<td>-0.020</td>
<td>0.000</td>
<td>0.140</td>
<td>-0.413$^{(c)}$</td>
<td>11.826$^{(d)}$</td>
<td>0.860</td>
<td>-0.730</td>
</tr>
</tbody>
</table>

(a) skewness coefficient = $(1/N)\sum_t [(x_t - \bar{x})/s.d.(x_t)]^3$ where $\bar{x}$ denotes the sample mean of the relevant variable, $s.d.(x_t)$ denotes the standard deviation and $N$ is the number of observations of $x_t$.

(b) kurtosis coefficient = $(1/N)\sum_t [(x_t - \bar{x})/s.d.(x_t)]^4$ where $\bar{x}$ denotes the sample mean of the relevant variable, $s.d.(x_t)$ denotes the standard deviation and $N$ is the number of observations of $x_t$.

(c) The skewness coefficient for $\Delta r_t^{obs}$ is significantly different from zero at the 5% level, using the standard error of $\sqrt{6/N} = \sqrt{6/551} = 0.104$.

(d) The kurtosis coefficient for $\Delta r_t^{obs}$ is significantly different from 3 at the 5% level, using the standard error of $\sqrt{24/N} = \sqrt{24/551} = 0.209$. 
Figure 1: Australian short rate data: levels and first differences
Table 3: Estimation results

<table>
<thead>
<tr>
<th>$h^{(a)}$</th>
<th>Posterior Mean</th>
<th>Posterior Standard Deviation</th>
<th>MC Error$^{(b)}$</th>
<th>25th Perc.$^{(c)}$</th>
<th>50th Perc.$^{(c)}$</th>
<th>75th Perc.$^{(c)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$-0.0124$</td>
<td>0.0012</td>
<td>0.000006</td>
<td>$-0.0132$</td>
<td>$-0.0124$</td>
<td>$-0.0116$</td>
</tr>
<tr>
<td>1</td>
<td>$-0.0124$</td>
<td>0.0012</td>
<td>0.000004</td>
<td>$-0.0131$</td>
<td>$-0.0124$</td>
<td>$-0.0117$</td>
</tr>
<tr>
<td>0</td>
<td>$-0.0123$</td>
<td>0.0011</td>
<td>0.000004</td>
<td>$-0.0130$</td>
<td>$-0.0123$</td>
<td>$-0.0116$</td>
</tr>
<tr>
<td>$\theta$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0685</td>
<td>0.0084</td>
<td>0.000044</td>
<td>0.0630</td>
<td>0.0685</td>
<td>0.0740</td>
</tr>
<tr>
<td>1</td>
<td>0.0682</td>
<td>0.0078</td>
<td>0.000029</td>
<td>0.0631</td>
<td>0.0682</td>
<td>0.0733</td>
</tr>
<tr>
<td>0</td>
<td>0.0677</td>
<td>0.0074</td>
<td>0.000029</td>
<td>0.0629</td>
<td>0.0677</td>
<td>0.0724</td>
</tr>
<tr>
<td>$\sigma$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0362</td>
<td>0.0055</td>
<td>0.000536</td>
<td>0.0322</td>
<td>0.0358</td>
<td>0.0398</td>
</tr>
<tr>
<td>1</td>
<td>0.0381</td>
<td>0.0057</td>
<td>0.000352</td>
<td>0.0341</td>
<td>0.0378</td>
<td>0.0417</td>
</tr>
<tr>
<td>0</td>
<td>0.0400</td>
<td>0.0061</td>
<td>0.000249</td>
<td>0.0357</td>
<td>0.0395</td>
<td>0.0437</td>
</tr>
<tr>
<td>$\delta$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.669220</td>
<td>0.077553</td>
<td>0.007852</td>
<td>0.614460</td>
<td>0.668460</td>
<td>0.723660</td>
</tr>
<tr>
<td>1</td>
<td>0.642160</td>
<td>0.077064</td>
<td>0.004990</td>
<td>0.589930</td>
<td>0.640480</td>
<td>0.694350</td>
</tr>
<tr>
<td>0</td>
<td>0.617080</td>
<td>0.077549</td>
<td>0.003190</td>
<td>0.566130</td>
<td>0.617570</td>
<td>0.670130</td>
</tr>
</tbody>
</table>

(a) $h$ denotes the number of augmented data points added between observations.

(b) The Monte Carlo (MC) standard error is calculated from the correlated iterates for each parameter as per Kim, Shephard and Chib (1998).

(c) The pth percentile is the value below which p% of the iterates fall.
Table 4: Bayes factors for all nested models against the unrestricted model

<table>
<thead>
<tr>
<th>Model</th>
<th>$BF_{j0}^{(a)}$</th>
<th>MC Error$^{(b)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h^{(c)}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M_1$</td>
<td>3</td>
<td>$4.84 \times 10^{-14}$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$7.73 \times 10^{-13}$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>$4.19 \times 10^{-13}$</td>
</tr>
<tr>
<td>$M_2$</td>
<td>3</td>
<td>$2.022901$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$2.747100$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>$4.109200$</td>
</tr>
<tr>
<td>$M_3$</td>
<td>3</td>
<td>$0.001062$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$0.000573$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>$0.000084$</td>
</tr>
<tr>
<td>$M_4$</td>
<td>3</td>
<td>$5.98 \times 10^{-23}$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$5.79 \times 10^{-23}$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>$7.86 \times 10^{-27}$</td>
</tr>
</tbody>
</table>

(a) $BF_{j0}$ is estimated as described in Appendices A and B.
(b) The Monte Carlo (MC) standard error is calculated from the iterates in (B1) as per Kim, Shephard and Chib (1998).
(c) $h$ denotes the number of augmented data points added between observations.

Table 5: Posterior probabilities for all models.

<table>
<thead>
<tr>
<th></th>
<th>$M_0$</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>$M_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h^{(a)}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.3310</td>
<td>0.0000</td>
<td>0.6690</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>1</td>
<td>0.2670</td>
<td>0.0000</td>
<td>0.7330</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0</td>
<td>0.1957</td>
<td>0.0000</td>
<td>0.8043</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

(a) $h$ denotes the number of augmented data points added between observations.