

PROOF COVER SHEET

Author(s):	W. Keener Hughen
Article title:	The use of Bayes factors to compare interest rateterm structure models
Article no:	RQUF-593541
Enclosures:	1) Query sheet 2) Article proofs

Dear Author,

1. Please check these proofs carefully. It is the responsibility of the corresponding author to check these and approve or amend them. A second proof is not normally provided. Taylor & Francis cannot be held responsible for uncorrected errors, even if introduced during the production process. Once your corrections have been added to the article, it will be considered ready for publication.

For detailed guidance on how to check your proofs, please see <http://journalauthors.tandf.co.uk/production/checkingproofs.asp>.

2. Please review the table of contributors below and confirm that the first and last names are structured correctly and that the authors are listed in the correct order of contribution. This check is to ensure that your name will appear correctly online and when the article is indexed.

Sequence	Prefix	Given name(s)	Surname	Suffix
1		W. Keener	Hughen	
2		Carmelo	Giaccotto	
3		Po-Hsuan	Hsu	

Queries are marked in the margins of the proofs. Unless advised otherwise, submit all corrections and answers to the queries using the CATS online correction form, and then press the “Submit All Corrections” button.

AUTHOR QUERIES

General query: You have warranted that you have secured the necessary written permission from the appropriate copyright owner for the reproduction of any text, illustration, or other material in your article. (Please see <http://journalauthors.tandf.co.uk/preparation/permission.asp>.) Please check that any required acknowledgements have been included to reflect this.

	No queries

How to make corrections to your proofs using Adobe Acrobat

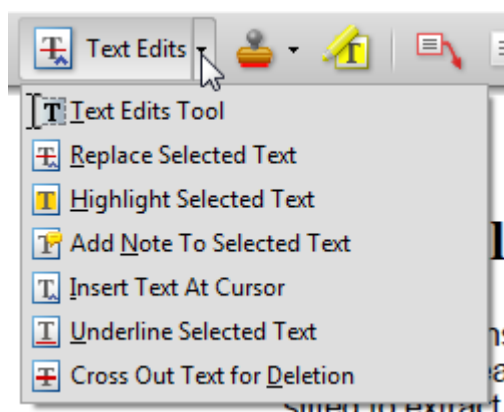
Taylor & Francis now offer you a choice of options to help you make corrections to your proofs. Your PDF proof file has been enabled so that you can edit the proof directly. This is the simplest and best way for you to ensure that your corrections will be incorporated. To do this, please follow these instructions:

1. Check which version of Adobe Acrobat you have on your computer. You can do this by clicking on the “Help” tab, and then “About”.

If Adobe Reader is not installed, you can get the latest version free from <http://get.adobe.com/reader/>.

- If you have Adobe Reader 8 (or a later version), go to “Tools”/ “Comments & Markup”/ “Show Comments & Markup”.
- If you have Acrobat Professional 7, go to “Tools”/ “Commenting”/ “Show Commenting Toolbar”.

2. Click “Text Edits”. You can then select any text and delete it, replace it, or insert new text as you need to. It is also possible to highlight text and add a note or comment.



3. Make sure that you save the file when you close the document before uploading it to CATS. A full list of the comments and edits you have made can be viewed by clicking on the “Comments” tab in the bottom left-hand corner of the PDF.

If you prefer, you can make your corrections using the CATS online correction form.

The use of Bayes factors to compare interest rate term structure models

5 W. KEENER HUGHEN*†, CARMELO GIACCOTTO‡ and PO-HSUAN HSU‡§

†University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, NC 28223, USA

‡University of Connecticut, 2100 Hillside Road, Storrs, CT 06269, USA

§University of Hong Kong, Pokfulam Road, Hong Kong

(Received 26 March 2009; in final form 28 April 2011)

10 Studies of the term structure of interest rates try to explain the relationship between the yield to maturity on zero-coupon bonds and their time to maturity. Over the years, many theoretical models have been developed to explain the stylized facts of U.S. Treasury yields; however, model comparison, parameter estimation and hypothesis testing remain thorny issues. The purpose of this paper is to show that Bayesian methods and Markov Chain Monte Carlo (MCMC) methods, in particular, may help resolve a number of these problems, especially those related to model comparison. We use MCMC to compare the seminal models of Vasicek and Cox, Ingersoll and Ross (CIR). The most surprising result of our analysis is that one of these two models is almost 50 000 times more likely than the other. In contrast, results in the previous literature have been much more ambiguous because they are based on a variety of goodness-of-fit measures. A Monte Carlo study shows that these results are not spurious: the MCMC method is able to select the correct data generation model, whereas goodness-of-fit measures are virtually indistinguishable regardless of whether the data were generated from Vasicek or CIR.

Keywords: Bayesian analysis; Bond yields; Monte Carlo methods; Term structure

25 1. Introduction

Studies of the term structure of interest rates try to explain the relationship between the yield to maturity on zero-coupon bonds and their time to maturity. For obvious reasons these studies are of intense interest to the economics/finance professions as well as Wall Street firms. To explain the term structure, a large number of bond pricing models have been developed following two seminal papers by Vasicek (1977) and Cox, Ingersoll and Ross (CIR 1985). More recent explanations can be found in texts such as Karatzas and Shreve (1988, 1991), Cochrane (2001), or the review paper by Maes (2004).

While theoretical models of the term structure have proliferated over the years, parameter estimation and hypothesis testing remain thorny issues (e.g., the recent exchange between Duffee and Stanton (2004) and Tauchen (2004)). Estimation problems were highlighted early on by Ball and Torous (1996); their study was one of many to point out difficulties in estimating the coefficient

of mean reversion in the instantaneous risk-free rate for both the CIR and other two-factor models. In contrast, Chib and Greenberg (1994) show that estimation in the presence of unit roots presents no particular challenges if one uses a Bayesian approach. Another problem is that most theoretical pricing models are specified in continuous time while actual bond prices or yields are observed in discrete time and this introduces a number of approximation errors. A third problem is how to compare the performance of competing term structure models. The purpose of this paper is to show that Bayesian methods and Markov Chain Monte Carlo (MCMC) methods, in particular, may help resolve a number of these problems, especially those related to model comparison.

Researchers have compared the Vasicek, CIR and other term structure models using various goodness-of-fit measures, with mixed results. For example, Chan *et al.* (1992) compute GMM minimized criterion (χ^2) values for various single-factor interest rate models and find that CIR has a slightly lower χ^2 value than the Vasicek model. This would suggest CIR fits observed yields slightly better than Vasicek. However, this result is based on one-month

*Corresponding author. Email: whughen@uncc.edu

Treasury bill yields as a proxy for the unobserved short rate. This approach has at least two points of contention: first, there is documented idiosyncratic variation in short maturity Treasury bill yields (Duffee 1994) which would suggest the one-month yield may not be good proxy for the short rate; and, second, this approach ignores information about the short rate imbedded in longer maturity Treasury bills. Chen and Yang (1995) also use relatively short maturity (three-month) Treasury bill yields as a proxy for the short rate, and run regressions to estimate the parameters of each model in the true, physical measure. Ten-year Treasury bond prices are then used to estimate a risk-premium parameter for each of the two models. They find little difference between the mean squared errors in the model-implied and the observed 10-year bond prices for the two models.

In the statistics literature, several papers, most notably Chib (1995), Chib and Jeliazkov (2001) and Chen (2005), have investigated methods for estimating the marginal likelihood using draws from the posterior distributions of the state variables and parameter vectors. We use these methods to compare the seminal models of Vasicek and CIR. In particular, we use the Gibbs Sampler and the Metropolis–Hastings algorithm to obtain the conditional distribution $p(\theta|r, Y)$ of the parameters θ given the instantaneous interest rate r and the bond yield data Y . We then employ the Kalman filter to obtain the conditional distribution $p(r|\theta, Y)$ for r given θ and Y for the Vasicek model; for the CIR model, since the interest rate r appears in the conditional variance of the interest rate evolution, the Kalman filter does not apply and we use a Metropolis–Hastings step to sample from the conditional distribution $p(r|\theta, Y)$. Sequentially sampling from the two conditional distributions $p(\theta|r, Y)$ and $p(r|\theta, Y)$ gives a Markov chain $\{\theta^{(g)}, r^{(g)}\}$ whose limiting distribution is the posterior distribution $\pi(\theta, r|Y)$ of interest. With the draws $\{\theta^{(g)}, r^{(g)}\}$ from the posterior distribution in hand, we apply the method described by Chib (1995) and Chen (2005) to compute the marginal likelihood for each model and obtain the posterior odds ratio. We assume the two models are equally probable *a priori*; the posterior odds ratio is then used to compare the likelihood of the two models given the Treasury bond yield data.

The most surprising result of our analysis is that the CIR model is almost 50 000 times more likely than the Vasicek model. Using the Kass and Raftery (1995) scale, the evidence is very strongly in favor of CIR. In contrast, results in the previous literature have been much more ambiguous because they are based on a variety of goodness-of-fit measures, usually mean square errors as in Chen and Yang (1995).

The main point of our paper is to show that measures such as mean square errors may not be powerful enough to detect the true underlying data generation process. A small Monte Carlo study shows that these results are not spurious: the MCMC method is able to select the correct data generation model, whereas goodness-of-fit measures are virtually indistinguishable regardless of whether the data were generated from Vasicek or CIR.

It is well known that neither model offers a perfect fit for U.S. Treasury yields. Nevertheless, due in large part to their analytic tractability, both models are still widely used. For example, researchers interested in pricing contingent claims on real assets, including commodities and real estate, choose either the single-factor Vasicek or CIR model for the interest rate so that their analyses remain manageable. Schwartz (1997) and Cassassus and Collin-Dufresne (2005) use the single-factor Vasicek model for the interest rate in conjunction with Gaussian models for commodity price processes in their analyses to price futures and options on various commodities. In the mortgage literature, Titman and Torous (1989) use the single-factor CIR model for the instantaneous rate to value commercial mortgages. Kau *et al.* (1992, 1995) also use the single-factor CIR model for the instantaneous rate to value fixed-rate mortgages, while Collin-Dufresne and Harding (1999) use the single-factor Vasicek model to value fixed-rate residential mortgages.

We wish to make clear that the dynamics of the term structure of U.S. Treasury yields are better explained by more advanced multi-factor models such as the classes of affine models that have been the focus of much of the recent empirical work investigated, for example, by Dai and Singleton (2000, 2002), Duffee (2002), Bester (2004), Duffee and Stanton (2004) and Collin-Dufresne *et al.* (2009). As the proposed models become more advanced and more numerous, an important enterprise will be to find the class of models that the data favor more. This paper is simply a first step in this direction. In subsequent research we hope to extend the analysis and compare the different classes of three- and four-factor affine models.

The remainder of this article is organized as follows. In section 2 we review the two models, and describe the empirical methodology and the MCMC algorithms in some detail. In section 3 we discuss the data and present the results: we find that U.S. Treasury bond yield data overwhelmingly favor the CIR model over the Vasicek model. We also explore various goodness-of-fit measures for the yield curve and the difficulty these measures have in distinguishing the two models. In section 4 we provide robustness checks of our results. Section 5 concludes.

2. Methodology

2.1. Term structure models

Models of the term structure of interest rates specify a stochastic process for the instantaneous interest rate r . Typically, the instantaneous rate is a function $r_t = f(X_t, t)$ of one or more state variables X that are assumed to follow a diffusion process

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW. \quad (1)$$

Assuming the existence of an equivalent martingale, or risk-neutral, measure, the price at time t of a zero-coupon bond that pays \$1 at time $t + \tau$ is

$$B(X_t, t, \tau) = E_t^*[e^{-\int_t^{t+\tau} r_s ds}], \quad (2)$$

70

75

80

85

90

95

100

105

110

115

120

125

130

135

140

145

150

155

160

165

170

175

180 and by the Feynman–Kac Theorem the bond price must satisfy the partial differential equation

$$\frac{\partial B}{\partial t} + \frac{\partial B}{\partial X} \mu^*(X, t) + \frac{1}{2} \text{trace} \left(\frac{\partial^2 B}{\partial X^2} \sigma(X, t) \sigma(X, t)' \right) = f(X, t) B, \quad (3)$$

185 where E_t^* denotes expectation with respect to the risk-neutral measure conditional on information available at time t , and $\mu^*(X, t)$ denotes the drift of the state variables in the risk-neutral measure. By Girsanov's Theorem, the risk-neutral drift μ^* is the physical drift μ less the risk premium whose functional form must be specified by the model.

190 The yield on a bond with maturity τ is minus the natural log of the bond price divided by τ : $Y_t(\tau) = -\ln B/\tau$. The term structure of interest rates, or yield curve, then displays the relationship between the yields and time to maturity. The instantaneous rate r is the limit of the yield as time to maturity goes to zero; because bonds of arbitrarily small maturities are not traded, the instantaneous rate is unobserved.

200 The term structure can be computed by solving (3) with boundary condition $B = 1$ for $\tau = 0$. Given a panel set of zero-coupon bond prices (or yields), this then allows estimation of the parameters conditional on the state variables. However, the state variables are typically unobserved and must themselves be estimated from the bond price data by filtering or other techniques. For many term structure models, the stochastic differential equation (1) is very difficult to solve; if the bond price data is observed at high frequency, then the solution to (1) may be approximated by the Euler discretization and in this case the distribution of the state variables at time $t + 1$ conditional on the state variables at time t is Gaussian.

2.2. The Vasicek and CIR models

215 Vasicek (1977) first adopted the mean-reverting Gaussian process to model the evolution of the instantaneous interest rate. It was one of the first models to allow mean reversion in the interest rate—at high interest rate levels economic activity would be hampered and equilibrium forces would tend to pull the rate lower, and conversely at low interest rate levels equilibrium forces would tend to pull the rate back higher—a property that has been empirically documented in observed rates.

220 In particular, the Vasicek model specifies that the instantaneous interest rate r follows an Ornstein–Uhlenbeck process:

$$dr_t = (\mu - \kappa r_t) dt + \sigma dW_t, \quad (4)$$

225 where μ , κ , and σ are constants. In his original paper, Vasicek assumes the risk premium is constant, so that the risk-neutral drift of the instantaneous rate is $\mu^* - \kappa r_t$. This is an unnecessary and stringent assumption. Duffee (2002) and Dai and Singleton (2002) show that the model is much improved under a variety of measures if the risk premium is allowed to vary over time. We assume the risk

premium is a linear function of r_t , so that the risk-neutral drift of the instantaneous rate is $\mu^* - \kappa^* r_t$.

The price B of a zero-coupon bond satisfies the differential equation

$$\frac{\partial B}{\partial t} + (\mu^* - \kappa^* r) \frac{\partial B}{\partial r} + \frac{1}{2} \frac{\partial^2 B}{\partial r^2} \sigma^2 = rB, \quad (5)$$

with boundary condition $B = 1$ when $\tau = 0$. The solution is of exponential affine form and is well known; the corresponding formula for the yield is

$$Y_t(\tau) = \alpha_V(\tau) + \beta_V(\tau) r_t, \quad (6)$$

where

$$\beta_V(\tau) = \frac{1 - e^{-\kappa^* \tau}}{\kappa^* \tau}, \quad (7)$$

$$\alpha_V(\tau) = \left(\frac{\sigma^2}{2\kappa^{*2}} - \frac{\mu^*}{\kappa^*} \right) (\beta_V(\tau) - 1) + \frac{\sigma^2 \tau}{4\kappa^*} \beta_V^2(\tau). \quad (8)$$

245 Following common practice, we include an additive pricing error in the yield formula, so that the observation equation is

$$Y_t = \alpha_V + \beta_V r_t + \epsilon_t, \quad t = 1, \dots, T, \quad (9)$$

250 where Y_t is the vector of yields at each observation time, and ϵ_t is normally distributed with zero mean and homoscedastic variance (uncorrelated across time and maturity). This pricing error captures some of the possible misspecification of the Vasicek model; alternatively, it may represent reporting or measurement error in the bond yields.

255 An undesirable feature of the Vasicek model is that it implies the instantaneous interest rate and yields are negative with positive probability. Cox *et al.* (1985) develop a model that is internally consistent in which the instantaneous interest rate follows a square-root process:

$$dr_t = (\mu - \kappa r_t) dt + \sigma \sqrt{r_t} dW_t. \quad (10)$$

260 Provided the parameters μ and κ are non-negative and satisfy $2\mu \geq \sigma^2$, there is a non-negative solution to this stochastic differential equation. To prevent arbitrage opportunities the risk-premium is assumed to be a constant multiple of r_t , so that the price of a zero-coupon bond satisfies

$$\frac{\partial B}{\partial t} + (\mu - \kappa^* r) \frac{\partial B}{\partial r} + \frac{1}{2} \frac{\partial^2 B}{\partial r^2} \sigma^2 r = rB, \quad (11)$$

270 with boundary condition $B = 1$ when $\tau = 0$. The solution to this differential equation is also of exponential affine form, and the corresponding formula for the yield is

$$Y_t(\tau) = \alpha_C(\tau) + \beta_C(\tau) r_t,$$

where

$$\beta_C(\tau) = \frac{2(e^{\gamma\tau} - 1)}{\tau(\kappa^* + \gamma)(e^{\gamma\tau} - 1) + 2\gamma\tau}, \quad (12)$$

$$\alpha_C(\tau) = -\frac{2\mu^*}{\tau\sigma^2} \log \left[\frac{2\gamma e^{(\kappa^* + \gamma)\tau/2}}{(\kappa^* + \gamma)(e^{\gamma\tau} - 1) + 2\gamma} \right], \quad (13)$$

$$\gamma = \sqrt{\kappa^{*2} + 2\sigma^2}. \quad (14)$$

235

240

245

250

255

260

265

270

275

We again include an additive pricing error, so that the observation equation is

$$Y_t = \alpha_C + \beta_C r_t + \epsilon_t, \quad t = 1, \dots, T. \quad (15)$$

In both models, the instantaneous interest rate is an unobserved state variable that must be estimated along with the model parameters from the observed bond yield data $Y = \{Y_1, \dots, Y_T\}$. The parameters of each model are collected in a vector θ and consist of the physical measure drift parameters μ and κ , the risk-neutral measure drift parameters μ^* (for the CIR model, $\mu = \mu^*$) and κ^* , the volatility parameter σ , the variance of the bond yield measurement error $\sigma_y^2 I$, and the initial instantaneous interest rate r_0 .

The bond yield data are observed at discrete times with frequency, say, Δt . Although the exact transition densities are known for both the Vasicek and CIR models, we find it more convenient to work with the Euler discretizations for which the transition densities are Gaussian and the posterior densities of the physical drift parameters are also Gaussian. Previous studies have investigated the error due to discretization and found no discernable error with observation frequencies up to one month (Bester 2004). In this case, both models may be cast in state space form

$$\begin{aligned} r_{t+1} &= \mu \Delta t + (1 - \kappa \Delta t) r_t + \xi_t, \\ Y_t &= \alpha + \beta r_t + \epsilon_t, \end{aligned}$$

where ξ_t is normal with mean zero and variance either σ^2 for the Vasicek model or $\sigma^2 r_t$ for the CIR model.

2.3. The Bayes factor

Many of the term structure models are not encompassed by a general likelihood function that can be used for classical model comparison. Although there do exist classical-based tests for non-nested models, the classical approach is not as intellectually appealing as the Bayesian approach. The classical approach attempts to reject one model in favor of the other, with the implicit assumption that one of the models is the ‘true’ model. As a practical matter, this requires the arbitrary designation of one of the models as the null hypothesis. The Bayesian approach, on the other hand, simply compares the two hypotheses given the data.

The Bayesian approach to model comparison is to compute the posterior odds ratio that one model is more favorable than the other, given the data. The posterior odds ratio is the ratio of the *a posteriori* probabilities of the two models, and measures the degree to which one model is more likely than the other to have generated the data.

By Bayes Theorem, the *a posteriori* probability $P(\mathfrak{M} | Y)$ for each model may be written

$$P(\mathfrak{M} | Y) = \frac{P(Y | \mathfrak{M})P(\mathfrak{M})}{P(Y | \mathfrak{M}_1)P(\mathfrak{M}_1) + P(Y | \mathfrak{M}_2)P(\mathfrak{M}_2)},$$

so that the posterior odds ratio is equal to the product of the Bayes factor and the prior odds ratio:

$$\frac{P(\mathfrak{M}_1 | Y)}{P(\mathfrak{M}_2 | Y)} = \frac{P(Y | \mathfrak{M}_1)P(\mathfrak{M}_1)}{P(Y | \mathfrak{M}_2)P(\mathfrak{M}_2)}. \quad (16)$$

The Bayes factor,

$$BF = \frac{P(Y | \mathfrak{M}_1)}{P(Y | \mathfrak{M}_2)},$$

is the ratio of the marginal likelihoods. When the two models are equally probable *a priori*, so that $P(\mathfrak{M}_1) = P(\mathfrak{M}_2) = 0.5$, then the Bayes factor is equal to the posterior odds ratio.

To compute the Bayes factor BF , the marginal likelihood $P(Y | \mathfrak{M})$ must be computed for each of the two competing models. If each model has unknown parameters θ and latent state variables X , the marginal likelihood for each model is

$$P(Y | \mathfrak{M}) = \int L(\theta, X | Y)P(X | \theta)P(\theta)dXd\theta, \quad (17)$$

where $L(\theta, X | Y)$ is the likelihood function, $P(X | \theta)$ is the probability density of the state variables given the model parameter vector θ , and $P(\theta)$ is the prior density of the parameter vector. Thus the Bayes factor is similar to the likelihood ratio statistic except the parameters θ are eliminated by integration rather than maximization.

The integral that defines the marginal likelihood in equation (17) is typically very complicated and extremely difficult, if not impossible, to compute analytically. In general, the difficulty of computing the marginal likelihood is well documented in the statistics literature (see, for example, Kass and Raftery (1995), Chib (1995), Chib and Jeliazkov (2001), Chen (2005)). One problem is that, for moderate to large sample sizes, the integrand in (17) is highly peaked around its maximum, and many numerical methods for estimating the integral have difficulty finding the region where the mass is accumulating. This is especially true for comparisons of interest rate models; because interest rates are highly persistent a large sample size is needed to get reliable estimates of the parameters. Another problem for many interest rate term structure models is that the state variables are unobserved and the parameter vectors are of high dimension, making ordinary Monte Carlo sampling approaches difficult to implement.

Recently, however, several papers have investigated indirect methods for estimating marginal likelihoods using output from Gibbs sampling and the Metropolis–Hastings algorithm. In particular, the approach of Chib (1995) exploits the fact that the log of the marginal likelihood can be written

$$\log P(Y | \mathfrak{M}) = \log L(\theta^* | Y) + \log P(\theta^*) - \log P(\theta^* | Y), \quad (18)$$

where θ^* is an arbitrary value of θ (generally chosen to be a point of high posterior density),

$$L(\theta^* | Y) = \int L(\theta^*, X | Y)P(X | \theta^*)dX \quad (19)$$

380 is the parameter likelihood function, $P(\theta^*)$ is the prior density, and $P(\theta^* | Y)$ is the posterior ordinate at θ^* .

The parameter likelihood $L(\theta^* | Y)$ may be computed using the particle filter method, and the prior density $P(\theta^*)$ is available directly. To estimate the posterior ordinate $P(\theta^* | Y)$ we follow the method of Chib and Jeliazkov (2001), who show how to estimate $P(\theta^* | Y)$ using draws $\{\theta^{(g)}, X^{(g)}\}_{g=1}^G$ from the posterior distribution $P(\theta, X | Y)$. Markov chain Monte Carlo (MCMC) methods may be used to obtain these draws.

2.4. Markov chain Monte Carlo (MCMC)

390 MCMC generates random samples from the posterior distribution $P(\theta, X | Y)$ by sequentially sampling from the conditional distributions $P(X | \theta, Y)$ and $P(\theta | X, Y)$, which are typically easier to characterize than the higher-dimensional distribution $P(\theta, X | Y)$. These conditional distributions uniquely determine the joint distribution $P(\theta, X | Y)$, due to the Clifford–Hammersley theorem (see Johannes and Polson (2006), for example). The MCMC algorithm is as follows. First, choose an initial value for θ , $\theta^{(0)}$; draw $X^{(1)} \sim P(X | \theta^{(0)}, Y)$; draw an updated θ , $\theta^{(1)} \sim P(\theta | X^{(1)}, Y)$; repeat for some number G of iterations. This generates a Markov chain $\{\theta^{(g)}, X^{(g)}\}_{g=1}^G$ whose distribution converges, under very mild regularity conditions, to the target distribution $\pi(\theta, X | Y)$.

405 For the Vasicek model the instantaneous rate follows a homoscedastic Gaussian process and draws from $P(r | \theta, Y)$ may be obtained using the Kalman filtering recursions. Specifically, the Kalman filter is run to obtain the moments of $P(r_t | \theta, Y_1, \dots, Y_t)$ for $t = 1, \dots, T$; the final state r_T is sampled from $P(r_T | \theta, Y_1, \dots, Y_T)$; then each earlier state r_t is sampled from $P(r_t | r_{t+1}, \theta, Y_1, \dots, Y_t)$ recursively. This procedure generates a draw r_1, \dots, r_T from

$$P(r | \theta, Y) = P(r_T | \theta, Y) \prod_{t=1}^{T-1} P(r_t | r_{t+1}, \theta, Y_1, \dots, Y_t)$$

415 in a single block. This procedure is described in detail by Carter and Kohn (1994).

For the CIR model, since the interest rate r appears in the conditional variance of the interest rate evolution, the Kalman filter does not apply and we use a Metropolis–Hastings step to sample from $P(r | \theta, Y)$. Since r is a first-order Markov process, by the law of total probability the conditional density for the time t short rate r_t , given the short rate at all the other times, the parameters, and the data, is

$$P(r_t | \theta, Y, \text{short rate at all other times}) \propto P(Y_t | r_t, \theta) P(r_t | r_{t-1}, \theta) P(r_{t+1} | r_t, \theta). \quad (20)$$

425 The Metropolis–Hastings algorithm consists of specifying a proposal density q from which a candidate \hat{r}_t for r_t is drawn; this candidate is accepted with probability given by

$$\min \left(\frac{P(\hat{r}_t | \theta, Y, \text{rate at all other times}) q(r_t^{(g)} | \hat{r}_t)}{P(r_t^{(g)} | \theta, Y, \text{rate at all other times}) q(\hat{r}_t | r_t^{(g)})}, 1 \right), \quad (21)$$

where $r_t^{(g)}$ is the previously accepted draw of r_t . We use random-walk Metropolis–Hastings, and draw the candidate from a symmetric fat-tailed proposal distribution q with mean $r_t^{(g)}$. The particular fat-tailed distribution we use is a generalization of Tukey’s lambda distribution as developed by Dudewicz *et al.* (1979), for which the variance is chosen so that the acceptance rate is around 40% as recommended by Johannes and Polson (2006) and Chib and Greenberg (1995). If a draw \hat{r}_t is negative, it is simply discarded and a new candidate is drawn. Because the proposal density q is symmetric in this case, q drops out of the quotient in (21) and the candidate is accepted with probability

$$\min \left(\frac{P(Y_t | \hat{r}_t, \theta) P(\hat{r}_t | r_{t-1}, \theta) P(r_{t+1} | \hat{r}_t, \theta)}{P(Y_t | r_t^{(g)}, \theta) P(r_t^{(g)} | r_{t-1}, \theta) P(r_{t+1} | r_t^{(g)}, \theta)}, 1 \right). \quad (22)$$

This generates a short rate series that is correlated across time since the probability that the candidate is accepted depends on the previously accepted short rate at times $t - 1$ and $t + 1$.

450 To obtain draws from $P(\theta | r, Y)$, we use the Gibbs sampler and the Metropolis–Hastings algorithm. It is not possible to directly sample from $P(\theta | r, Y)$, but the Clifford–Hammersley theorem may again be applied to show that this conditional distribution is uniquely characterized by the set of conditional distributions of each element, or block of elements, of θ . We decompose the parameter vector θ into four blocks: (μ, κ, μ^*) , (σ^2, κ^*) , σ_Y^2 , and r_0 . An outline of the procedure to draw from the conditional distribution for each block is given below.

460 The conditional density $P(\mu, \kappa, \mu^* | \sigma^2, \kappa^*, \sigma_Y^2, r_0, r, Y)$ of the drift parameters (μ, κ, μ^*) is proportional to the product

$$P(\mu, \kappa, \mu^* | \sigma^2, \kappa^*, \sigma_Y^2, r_0, r, Y) \propto P(\mu, \kappa, \mu^*) P(r | \mu, \kappa, \sigma^2, r_0) P(Y | \mu^*, \kappa^*, \sigma^2, \sigma_Y^2, r).$$

465 $P(r | \mu, \kappa, \sigma^2, r_0)$ is the product of the interest rate transition densities $P(r_t | \mu, \kappa, \sigma^2, r_{t-1})$; with the discretized version of each model, these transition densities are Gaussian and their means are linear functions of μ and κ and therefore $P(r | \mu, \kappa, \sigma^2, r_0)$ is a Gaussian density function of these two parameters. Similarly, because μ^* appears linearly in the yield formula, the likelihood function $P(Y | \mu^*, \kappa^*, \sigma^2, \sigma_Y^2, r)$ is a Gaussian density function of μ^* . With a Gaussian prior for (μ, κ, μ^*) the conditional density $P(\mu, \kappa, \mu^* | \sigma^2, \kappa^*, \sigma_Y^2, r_0, r, Y)$ is also Gaussian and these parameters are updated by drawing from a normal distribution.

475 The conditional density $P(\sigma^2, \kappa^* | \mu, \kappa, \mu^*, \sigma_Y^2, r_0, r, Y)$ of the volatility parameter σ^2 and the risk-neutral drift parameter κ^* is proportional to the product

$$P(\sigma^2, \kappa^* | \mu, \kappa, \mu^*, \sigma_Y^2, r_0, r, Y) \propto P(\sigma^2, \kappa^*) P(r | \mu, \kappa, \sigma^2, r_0) P(Y | \mu^*, \kappa^*, \sigma^2, \sigma_Y^2, r).$$

The instantaneous variance of the interest rate is proportional to σ^2 in either model and thus the interest rate

480 probability density $P(r | \mu, \kappa, \sigma^2, r_0)$ is proportional to an
Inverse Gamma density function of σ^2 . We assume the
prior $P(\sigma^2, \kappa^*) = P(\sigma^2)P(\kappa^*)$ is Inverse Gamma for σ^2 and
Gaussian for κ^* . The product $P(\sigma^2) P(r | \mu, \kappa, \sigma^2, r_0)$
485 is then Inverse Gamma in σ^2 ; however, the posterior density
is not recognizable. We therefore use a Metropolis–
Hastings step and first draw a candidate $\hat{\sigma}^2$ from the
Inverse Gamma density $P(\sigma^2) P(r | \mu, \kappa, \sigma^2, r_0)$ and then a
candidate $\hat{\kappa}^*$ from the student- t density with mean $m(\hat{\sigma}^2)$
and variance $V(\hat{\sigma}^2)$ given by

$$m(\hat{\sigma}^2) = \arg \max_{\kappa^*} \log P(Y | \mu^*, \kappa^*, \hat{\sigma}^2, \sigma_Y^2, r),$$

$$V(\hat{\sigma}^2) = - \left(\frac{\partial^2 \log P(Y | \mu^*, \kappa^*, \hat{\sigma}^2, \sigma_Y^2, r)}{\partial \kappa^{*2}} \right)^{-1} \Bigg|_{\kappa^*=m}.$$

490 The candidate $(\hat{\sigma}^2, \hat{\kappa}^*)$ is accepted with probability

$$\min \left(\frac{P(Y | \mu^*, \hat{\kappa}^*, \hat{\sigma}^2, \sigma_Y^2, r) P(\hat{\kappa}^*) St(\kappa^*; m(\sigma^2), V(\sigma^2))}{P(Y | \mu^*, \kappa^*, \sigma^2, \sigma_Y^2, r) P(\kappa^*) St(\hat{\kappa}^*; m(\hat{\sigma}^2), V(\hat{\sigma}^2))}, 1 \right).$$

This is a combination of independence Metropolis–
Hastings and the ‘tailored’ Metropolis–Hastings steps as
495 described by Chib and Greenberg (1995).

The variance of the pricing error σ_Y^2 appears in the yield
formula only. The conditional density for σ_Y^2 is propor-
tional to the product of the prior and the yield probability
density, which is an Inverse Gamma density function
500 of σ_Y^2 . We follow Chib and Ergashev (2009) and assume a
flat prior for σ_Y^2 , so that this parameter is updated by
drawing from an Inverse Gamma distribution.

The conditional density of the initial interest rate r_0 is
proportional to the product of the prior and the transition
505 density $P(r_1 | r_0, \mu, \kappa, \sigma^2)$. For the Vasicek model the
transition density is a Gaussian density function of r_0 , but
for the CIR model the transitional density is not a
recognizable density function of r_0 . We assume a
Gaussian prior for r_0 , so that this parameter may be
510 drawn from a normal distribution for the Vasicek model.
We use random-walk Metropolis for the CIR model.

The techniques we use to compute the Bayes factor
require that the priors for the parameters be proper
distributions. We have therefore chosen proper priors for
515 all the parameters. Furthermore, consistent with the work
of Chib and Ergashev (2009), we impose four additional
requirements. First, the yield curve implied by the priors
should be upward sloping on average. Second, the
instantaneous rate should be highly persistent—a well-
520 known characteristic. Third, the priors are chosen in such
a way that the means are near the values of previous
empirical results so that this information is included.
Fourth, we ensure the priors are relatively diffuse so that
the data have sufficient influence on the posterior
525 distribution.

In general, Markov chains generated by Metropolis–
Hastings algorithms (including the Gibbs sampler) have
special properties which allow convergence conditions to
be verified readily. A sufficient condition is that the
530 proposal distributions each has positive density on the
same support as the corresponding target distribution.
This condition is satisfied for our Metropolis–Hastings

algorithms described above, and therefore the resulting
Markov chain $\{\theta^{(g)}, r^{(g)}\}_{g=1}^G$ has equilibrium distribution
535 equal to the target distribution $P(\theta, r | Y)$.

An important practical problem is the determination of
the number of iterations to run. MCMC algorithms
generate dependent samples and therefore convergence
may be difficult to formally diagnose from the realized
540 output of the chain, even though the theory is clear that
the chains converge. We follow common practice and
analyse the correlation structure of draws by computing
the autocorrelation function, although it is possible for
non-convergent chains to have low autocorrelation.
545 Nevertheless, this should provide some evidence of
convergence. We also test for convergence by analysing
MCMC output from simulated data. A minimum
requirement is that the first moments (i.e. the sample
means) converge to their true values (used to generate the
550 simulated data). Furthermore, we verify that the Bayes
factor estimate does favor the correct model on our
simulated data sets.

3. Results

3.1. Data

The data are month-end yields on zero-coupon U.S.
555 Treasury bonds. They are obtained from the Center for
Research in Security Prices (CRSP), in the Fama–Bliss
files. Specifically, we use the three-month rate in the Fama
Risk-free Rates File, and the one-year and five-year yields
560 in the Fama–Bliss Discount Bond File. The sample period
is June 1964 to December 2005, which contains 499
monthly observations.

3.2. Prior specification

For the various parameter blocks, we used the following
565 specification.

- μ and κ are normally distributed with means
0.01 and 0.1, respectively. The variances are set
at 0.001 and 0.005, respectively, for both
models. This specification is applied also to
570 the risk-neutral parameters μ^* and κ^* .
- σ^2 is distributed Inverse Gamma, with mean
0.0004 and variance 0.001 for the Vasicek
model, and mean 0.004 and variance 0.001 for
the CIR model.
- As in Chib and Ergashev (2009), we assume flat
575 priors on the measurement errors.

We took 10000 draws of the parameters, and then
simulated three interest rate time series and three panel
sets of yields for each model, according to the
580 5-percentile, 50-percentile, and 95-percentile. For exam-
ple, the first panel set is at time series of yields (for $t = 1$ to
500 months, in line with our actual data) with eight
maturities: 1 month, 3 months, 6 months, 1 year, 2 years,
3 years, 4 years, and 5 years, corresponding to the fifth
585 percentile of the parameter draws. These are averaged

Table 1. Descriptive statistics based on the prior draws for the parameters for the two models. The slope of the term structure is measured as the difference between the 5-year and 3-month yields and is based on 10 000 draws of the parameters. For each draw, the yield curves are simulated then averaged over the entire period of 480 months.

Parameter	Vasicek				CIR			
	Mean	St. dev.	5%	95%	Mean	St. dev.	5%	95%
μ	0.0100	0.0316	-0.0420	0.0620	0.0100	0.0316	-0.0420	0.0620
κ	0.2000	0.0707	0.0837	0.3163	0.2000	0.0707	0.0837	0.3163
$\sigma^2 \times 10^3$	0.4846	0.8833	0.1038	1.3386	5.1686	10.2387	1.1034	14.0287
μ^*	0.0100	0.0316	-0.0420	0.0620	0.0100	0.0316	-0.0420	0.0620
κ^*	0.0500	0.0707	-0.0663	0.1663	0.0500	0.0707	-0.0663	0.1663
Slope	0.0112	0.0747	-0.1098	0.1345	0.0475	0.0524	-0.0017	0.1428

Table 2. Descriptive statistics based on the posterior draws for the parameters for the two models using U.S. Treasury bond yield data.

Parameter	Vasicek				CIR			
	Mean	St. dev.	5%	95%	Mean	St. dev.	5%	95%
μ	0.0108	0.0041	0.0042	0.0176	0.0095	0.0004	0.0089	0.0101
κ	0.1786	0.0550	0.0884	0.2703	0.1658	0.0330	0.1116	0.2205
$\sigma^2 \times 10^3$	0.2501	0.0222	0.2152	0.2884	3.4428	0.2977	2.9771	3.9588
μ^*	0.0090	0.0003	0.0085	0.0096	0.0095	0.0004	0.0089	0.0101
κ^*	0.0685	0.0047	0.0609	0.0763	0.0747	0.0048	0.0669	0.0826
$\sigma_Y^2 \times 10^4$	0.3144	0.0129	0.2938	0.3360	0.4046	0.0171	0.3771	0.4331
r_0	0.0316	0.0054	0.0226	0.0406	0.0311	0.0038	0.0252	0.0377

over the time dimension to generate three yield curves for each model. The results, listed in table 1, are consistent with our requirements: first, a positively sloped yield curve; second, a high degree of variability; and third, the instantaneous interest rate is highly persistent, and has ‘volatility’ around 2%.

A negative aspect of this prior specification is that some yields may be negative even with the CIR model. Recall that unless the Feller condition $\mu > \sigma^2/2$ holds, there is no guarantee that the interest rate process will stay positive. Since μ is normally distributed, the draws near the left tail will be below $\sigma^2/2$ (and in fact may be negative). This problem was noted also by Chib and Ergashev (2009) even with a much more sophisticated interest rate model.

3.3. Bayesian analysis

For each model a Markov chain was generated of length 125 000 and the first 25 000 draws were discarded to negate the effects of initial conditions. To address concerns that the posterior draws are too highly autocorrelated, only one out of every five iterations of the chain were saved, leaving 20 000 draws from the posterior distribution. The first-order autocorrelations of the physical drift parameters, σ^2 and σ_Y^2 , are all less than 0.25 and fall to less than 0.01 for the fifth-order autocorrelation. Although the first-order autocorrelations of the risk-neutral drift parameters are not trivial (about 0.7 for μ^* and 0.6 for κ^* for both models), the second-order autocorrelations are less than 0.1 and the

tenth-order autocorrelations are all less than 0.02. In fact, the inefficiency factor for each of the parameters (Chib and Ergashev 2009),

$$1 + 2 \sum_{k=1}^{500} \left(1 - \frac{k}{500}\right) \rho(k), \quad (23)$$

where $\rho(k)$ is the autocorrelation at lag k of the MCMC draws of that parameter, is less than 6.0 for every parameter.

Table 2 lists descriptive statistics on the posterior draws for the parameters: the means, standard deviations, and 5- and 95-percentiles. The point estimates are mostly consistent with previous findings. Table 3 provides a summary of parameter estimates from some earlier studies. These studies include Chan, Karolyi, Longstaff, and Sanders (CKLS 1992), Chen and Yang (1995), de Jong and Santa-Clara (1999), de Jong (2000), and Duffee and Stanton (2004). CKLS (1992) and Chen and Yang (1995) use proxies for the short rate, whereas the other three studies use panel data and treat the short rate as an unobserved state variable. The sample periods, observation frequencies, and estimation methods vary across these studies.

We see from table 2 that the risk-neutral drift parameters are estimated with much greater precision than the drift parameters in the physical measure. There are two reasons for this. First, since the short rate is not observed and must be filtered, the physical drift parameters are estimated using filtered estimates of the short rate. This is

590

595

600

605

610

615

620

625

630

635

640

Table 3. Summary of parameter estimates from earlier studies.

Study	Model	μ	κ	σ^2	μ^*	κ^*
CKLS	Vasicek	0.0154	0.1779	0.0004		
	CIR	0.0189	0.2339	0.0073		
Chen and Yang	Vasicek	0.0159	0.2456	0.00083	0.2877	0.2456
	CIR	0.0159	0.2456	0.02247	0.0159	0.1166
deJong and Santa-Clara	Vasicek	0.0082	0.1191	0.00014	0.1021	0.1191
	CIR	0.0122	0.1862	0.00231	0.0122	0.1120
Duffee and Stanton	Vasicek	0.0134	0.205	0.00031	0.0084	0.065
	CIR	0.0075	0.131	0.0064	0.0075	0.063
deJong	Vasicek	0.0016	0.0222	0.00020	0.0029	0.0222
	CIR	0.0025	0.0429	0.00217	0.0025	0.0116
Ours	Vasicek	0.0108	0.1786	0.00025	0.0090	0.0685
	CIR	0.0095	0.1658	0.00344	0.0095	0.0747
Ours, using proxy for r	Vasicek	0.0140	0.2323	0.00039	0.0108	0.0887
	CIR	0.0102	0.1681	0.00418	0.0102	0.0824

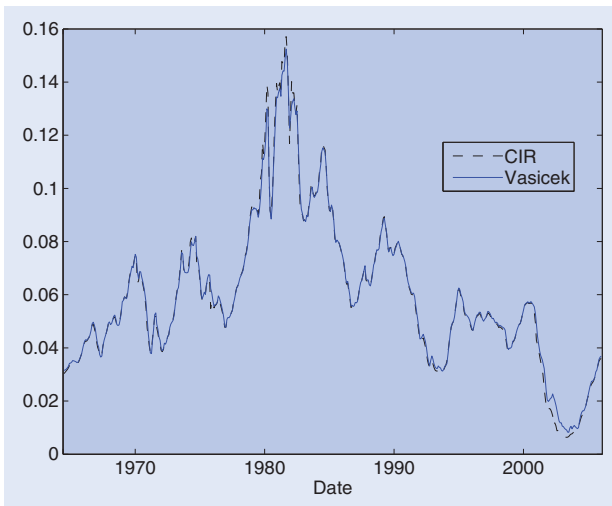


Figure 1. Model implied short rates. The 20000 draws of the short rate time series for each model are averaged to obtain the implied short rate series.

a problem with any asset pricing model that uses latent state variables. For example, Schwartz (1997) estimates several latent state variable, Gaussian models of commodity prices using futures data and reports standard errors of five to 10 times greater for the physical drift parameters than for the risk-neutral drift parameters. The second reason for the imprecise estimates of the physical drift parameters is that interest rates are highly persistent; even if the short rate were observable, larger data sets than are currently available would be required to estimate these parameters with much greater precision.

Plots of the implied short rate series are shown in figure 1. For each model, the 20000 draws of the short rate time series are averaged to obtain the point estimates of the plotted implied short rate series. The two implied short rate series are virtually indistinguishable; they differ by at most 50 or so basis points, while most of the time they only differ by less than 10 basis points. On average, at each observation time the two implied short rate series differ in magnitude by about five basis points.

We estimate the marginal likelihood $P(Y | \mathfrak{M})$ for each model following the method of Chib (1995) and Chib and Jeliazkov (2001) as outlined in section 2.3. We write the log marginal likelihood as the sum

$$\log P(Y | \mathfrak{M}) = \log L(\theta^* | Y) + \log P(\theta^*) - \log P(\theta^* | Y). \quad (24)$$

We use the particle filter method to estimate the parameter likelihood

$$L(\theta^* | Y) = P(Y_1 | \theta^*) \prod_{t=2}^T P(Y_t | Y_1, \dots, Y_{t-1}, \theta^*) \quad (25)$$

as follows. Given a random sample $\{r_{t-1}^{(m)}\}_{m=1}^M$ from $P(r_{t-1} | Y_1, \dots, Y_{t-1}, \theta^*)$, sample $\hat{r}_t^{(m)}$ from the transition density $P(r_t | r_{t-1}^{(m)}, \theta^*)$ and estimate $P(Y_t | Y_1, \dots, Y_{t-1}, \theta^*)$ as the Monte Carlo average,

$$P(Y_t | Y_1, \dots, Y_{t-1}, \theta^*) \approx \frac{1}{M} \sum_{m=1}^M P(Y_t | \hat{r}_t^{(m)}, \theta^*). \quad (26)$$

The initial sample $\{r_0^{(m)}\}$ is drawn from the prior $P(r_0 | \theta^*)$. The subsequent samples $\{r_t^{(m)}\} \sim P(r_t | Y_1, \dots, Y_t, \theta^*)$ for $t \geq 1$ are then obtained recursively. For the Vasicek model, the sample $\{r_t^{(m)}\}$ may be obtained using the standard Kalman filter recursions, which compute the mean and variance of the normally distributed $(r_t | Y_1, \dots, Y_t, \theta^*)$. For the CIR model, we employ the auxiliary particle filter as described by Pitt and Shephard (1999) and Chib *et al.* (2002): Given $\{r_{t-1}^{(m)}\} \sim P(r_{t-1} | Y_1, \dots, Y_{t-1}, \theta^*)$, first calculate $\hat{r}_t^{*(m)} = \mu \Delta t + (1 - \kappa \Delta t) r_{t-1}^{(m)}$ and $w_m = P(Y_t | \hat{r}_t^{*(m)}, \theta^*)$. Then sample R times the integers $1, 2, \dots, M$ with probability proportional to $\{w_m\}$, and let the sampled indexes be k_1, \dots, k_R and associate these with $\hat{r}_t^{*(k_1)}, \dots, \hat{r}_t^{*(k_R)}$. Next, for each value of k_j , simulate $r_t^{*(j)} \sim P(r_t | \hat{r}_t^{*(k_j)}, \theta^*)$ for $j = 1, \dots, R$. Finally, re-sample $\{r_t^{*(1)}, \dots, r_t^{*(R)}\}$ M times with probabilities proportional to the ratio $P(Y_t | r_t^{*(j)}, \theta^*) / P(Y_t | \hat{r}_t^{*(k_j)}, \theta^*)$ for $j = 1, \dots, R$ to produce the filtered sample $\{r_t^{(1)}, \dots, r_t^{(M)}\}$ from $(r_t | Y_1, \dots, Y_t, \theta^*)$.

The results, using $M = 20\,000$, are

$$\begin{aligned}\log L(\theta^* | Y)_{\text{VAS}} &= 5305.3, \\ \log L(\theta^* | Y)_{\text{CIR}} &= 5356.5.\end{aligned}$$

695 The prior density $P(\theta^*)$ is available directly using the priors specified in section 3.2. The results are

$$\begin{aligned}\log P(\theta^*)_{\text{VAS}} &= 16.1, \\ \log P(\theta^*)_{\text{CIR}} &= 10.8.\end{aligned}$$

700 To estimate the posterior ordinate $P(\theta^* | Y)$ we follow the method of Chib and Jeliazkov (2001) and decompose the posterior ordinate as

$$P(\theta^* | Y) = P(\theta_1^* | Y)P(\theta_2^* | Y, \theta_1^*)P(\theta_3^* | Y, \theta_1^*, \theta_2^*), \quad (27)$$

705 where $\theta_1 = (\sigma^2, \kappa^*)$, $\theta_2 = (\mu, \kappa, \mu^*)$, and $\theta_3 = \sigma_Y^2$. With our MCMC procedure above, the full conditional densities of θ_2 and θ_3 are known and θ_1 is drawn using Metropolis–Hastings. Let $\alpha(\theta_1, \theta_1^* | Y, r, \theta_2, \theta_3)$ denote the probability of a move for θ_1 in the Metropolis–Hastings step and let $q(\theta_1, \theta_1^* | Y, r, \theta_2, \theta_3)$ denote the proposal density. It may be shown (Chib and Jeliazkov 2001) that

$$P(\theta_1^* | Y) \approx \frac{\left\{ \sum_{g=1}^G \alpha(\theta_1^{(g)}, \theta_1^* | Y, r^{(g)}, \theta_2^{(g)}, \theta_3^{(g)}) \times q(\theta_1^*, \theta_1^* | Y, r^{(g)}, \theta_2^{(g)}, \theta_3^{(g)}) \right\}}{\sum_{k=1}^G \alpha(\theta_1^{(k)}, \theta_1^* | Y, r^{(k)}, \theta_2^{(k)}, \theta_3^{(k)})},$$

710 where the parameter draws used in the numerator are from the full MCMC run. For the denominator, the draws $(r^{(k)}, \theta_2^{(k)}, \theta_3^{(k)})$ are from a reduced MCMC run in which θ_1 is first fixed at θ_1^* and then for each k a variate $\theta_1^{(k)}$ is drawn from $q(\theta_1^*, \theta_1 | Y, r^{(k)}, \theta_2^{(k)}, \theta_3^{(k)})$. The reduced conditional ordinate $P(\theta_2^* | Y, \theta_1^*)$ may be expressed as

$$\begin{aligned}P(\theta_2^* | Y, \theta_1^*) &= \int P(\theta_2^* | Y, r, \theta_1^*, \theta_3)P(\theta_3, r | Y, \theta_1^*)d\theta_3 dr, \\ &\approx \frac{1}{G} \sum_{k=1}^G P(\theta_2^* | Y, r^{(k)}, \theta_1^*, \theta_3^{(k)}),\end{aligned}$$

720 the average of Gaussian densities, using the draws from the reduced run in the previous calculation. Similarly, the reduced conditional ordinate $P(\theta_3^* | Y, \theta_1^*, \theta_2^*)$ may be estimated as the average of Inverse Gamma densities

$$P(\theta_3^* | Y, \theta_1^*, \theta_2^*) \approx \frac{1}{G} \sum_{j=1}^G P(\theta_3^* | Y, r^{(j)}, \theta_1^*, \theta_2^*),$$

using draws from a final reduced run fixing θ_1 at θ_1^* and θ_2 at θ_2^* . The results are

$$\begin{aligned}\log P(\theta^* | Y)_{\text{VAS}} &= 41.3, \\ \log P(\theta^* | Y)_{\text{CIR}} &= 23.8.\end{aligned}$$

725 Our estimate of the log Bayes factor is therefore

$$\log \widehat{BF} = -18.4. \quad (28)$$

730 The numerical standard error is an estimate of the variance of the marginal likelihood estimate and gives the variation that can be expected if the simulation were to be repeated. For each of the two models, the variance of the log parameter likelihood may be estimated as follows. Using equations (25) and (26), the log parameter

likelihood is estimated as

$$\log \hat{L}(\theta^* | Y) = \sum_{t=1}^T \log \hat{P}(Y_t | Y_1, \dots, Y_{t-1}, \theta^*),$$

where

$$\hat{P}(Y_t | Y_1, \dots, Y_{t-1}, \theta^*) = \frac{1}{M} \sum_{m=1}^M P(Y_t | \hat{r}_t^{(m)}, \theta^*).$$

We use the Delta method to estimate the variance of the log parameter likelihood:

$$\text{var}(\log \hat{L}(\theta^* | Y)) \approx v' \text{var}(\hat{P})v, \quad (29)$$

740 where \hat{P} is the T -vector whose t th component \hat{P}_t is $\hat{P}(Y_t | Y_1, \dots, Y_{t-1}, \theta^*)$, v is the T -vector whose t th component is the reciprocal of \hat{P}_t , and $\text{var}(\hat{P})$ is given by the Newey–West (1987) estimate of the variance of \hat{P} using a lag length of 20. The results are

$$\begin{aligned}\text{var}(\log \hat{L}(\theta^* | Y))_{\text{VAS}} &= 1.46, \\ \text{var}(\log \hat{L}(\theta^* | Y))_{\text{CIR}} &= 0.51.\end{aligned}$$

745 Similarly, an estimate of the variance of the log posterior ordinate for each model may be obtained using the Delta method (see Chib and Jeliazkov (2001) for details). The results are

$$\begin{aligned}\text{var}(\log \hat{P}(\theta^* | Y))_{\text{VAS}} &= 3.26 \times 10^{-4}, \\ \text{var}(\log \hat{P}(\theta^* | Y))_{\text{CIR}} &= 17.76 \times 10^{-4};\end{aligned}$$

750 for each model, the variance of the log posterior ordinate is negligible compared with the variance of the log parameter likelihood. Therefore, the variance of the estimate of the log Bayes factor is approximately

$$\text{var}(\log \widehat{BF}) \approx 1.46 + 0.51 = 1.97,$$

755 and the t -statistic for the estimate of the log of the Bayes factor is

$$t = \frac{-18.4}{\sqrt{1.97}} = -13.1.$$

760 In other words, the CIR model is almost 50 000 times more likely than the Vasicek model to have generated U.S. Treasury bond yield data, and according to the scale of Kass and Raftery (1995), this is evidence very strongly in favor of the CIR model over the Vasicek model.

3.4. Classical goodness-of-fit analysis

765 Previous research that compares the performances of the Vasicek and CIR models has primarily relied on various measures of goodness of fit. Below we examine the various measures of each model's goodness of fit of the yield curve. Specifically, we compare each model's bias and root mean squared error (RMSE) of the model implied yield curve with the actual yield curve.

770 The bias b_t is defined as the difference between the actual yield Y_t and the model implied yield $\hat{Y}_t = \alpha + \beta r_t$, where α and β are given by (7) and (8) for the Vasicek

Table 4. Goodness-of-fit measures for the yield curve. The mean bias \bar{b} is the average bias in the fitted yields to the actual yields. Numbers in parentheses are Newey–West standard errors computed using a lag length of 20. We also report t -statistics for the difference in squared bias $b_{V,t}^2 - b_{C,t}^2$. For each bond yield n , the root mean squared error (RMSE) is the square root of the average squared bias: $RMSE(n) = \sqrt{(1/T) \sum_{t=1}^T (Y_t(n) - \hat{Y}_t(n))^2}$.

	\bar{b}_{VAS}	\bar{b}_{CIR}	t -Statistic	RMSE _{VAS}	RMSE _{CIR}
3-Month	-0.00109* (0.0008)	-0.00110* (0.0008)	-0.52	0.00500	0.00500
1-Year	0.00131* (0.0004)	0.00126* (0.0004)	0.05	0.00327	0.00327
5-Year	-0.00023 (0.0013)	-0.00040 (0.0014)	-0.47	0.00639	0.00645

model and (12)–(14) for the CIR model:

$$\text{bias: } b_t = Y_t - \hat{Y}_t.$$

The point estimates in table 2 are used to compute α and β and the short rate series r is the average of the 20 000 draws from the posterior distribution for each model. For each bond yield n ($n=3$ months, 1 year, 5 years), the sample mean of the bias $\bar{b}(n) = (1/T) \sum b_t(n)$ is reported in table 4. We also report Newey–West (1987) standard errors, and compute t -statistics to test the hypothesis that the mean bias is zero. The standard errors are calculated using a lag length of 20. If a reported mean bias is different from zero with a 5% statistical significance, that value is marked with an asterisk.

We also report in table 4 the root mean squared error (RMSE) for each yield for each of the two models. The RMSE is the square root of the sample mean of the squared bias:

$$RMSE(n) = \sqrt{\frac{1}{T} \sum_{t=1}^T b_t^2(n)}.$$

These results reveal that there is very little difference in yield curve fit for the two models with respect to bias and RMSE. On average, both models overestimate the 1-year bonds and underestimate the 3-month and 5-year bonds with nearly the same errors. While the mean bias is significant for the 3-month and the 1-year yields, from an economic point of view these findings do not suggest gross miss-pricing: the largest average bias is only 13.1 basis points. Interestingly, the RMSE for each model is lower for the 1-year bond than for the other two bonds, while the bias is largest for the 1-year bond for both models. The RMSEs for the two models are identical for the 3-month and 1-year bonds, and differ by less than one basis point for the 5-year bond.

To statistically compare the RMSEs for the two models, we follow Diebold and Mariano (1995). We compute t -statistics for the difference in squared bias $b_{V,t}^2 - b_{C,t}^2$; a significantly positive mean indicates the superiority of the CIR model over the Vasicek model with respect to the RMSE fit. Standard errors are again

computed using the method of Newey–West (1987) with lag length of 20. The resulting t -statistics for the three bonds imply no significant difference; the respective values are -0.52 , 0.05 , and -0.47 . These findings suggest that the RMSE does not have sufficient power to distinguish between the two models.

4. Robustness checks

The results in the previous section are consistent with earlier research that suggests that there is no clear consensus on which model is more likely to have generated Treasury bond data. In light of such long historical evidence, our finding (28) that the CIR model is much more likely than the Vasicek model may seem quite remarkable! In this section we examine the robustness of this result to potential problems arising from (1) the MCMC method and (2) the prior specification; we also (3) repeat the analysis using a proxy for the instantaneous rate.

4.1. Monte Carlo analysis

We use a Monte Carlo study to examine the robustness of our MCMC methods. For each model, we simulate 40 years of monthly interest rates and bond yield data using the point estimates of the parameters from table 2. Specifically, for the Vasicek model we use

$$dr_t = (0.0108 - 0.1768r_t)dt + 0.0158 dW_t, \quad (30)$$

and the point estimates $\mu^* = 0.0090$, $\kappa^* = 0.0685$, $\sigma_Y^2 = 0.4 \times 10^{-4}$, and $r_0 = 0.03$ to generate 480 monthly observations of the short rate series $\{r_t\}$ and the yield data $\{Y_t\}$ according to (7)–(6). For the CIR model, we use

$$dr_t = (0.0095 - 0.1658r_t)dt + 0.0587\sqrt{r_t} dW_t, \quad (31)$$

along with point estimates $\mu^* = 0.0095$, $\kappa^* = 0.0747$, $\sigma_Y^2 = 0.4 \times 10^{-4}$, and $r_0 = 0.03$ to generate the data according to (12)–(15). We repeat this procedure nine additional times, for a total of 20 simulated data sets.

For each of the simulated data sets, we treat the time series of yields Y_t as the observed data and repeat the analyses from the previous two sections. Clearly, one would expect the Bayes factor to be strongly in favor of the model that generated the data; in contrast, the root mean square should be of similar magnitude regardless of the data-generating model. The results are summarized in tables 5 and 6. In table 5 we report the averages over the 10 Vasicek data sets of the descriptive statistics of the MCMC results, and in table 6 we report the averages over the 10 CIR data sets. We observe that the average point estimates are all very close to their actual values. As expected, the risk-neutral parameters are estimated with greater precision than the physical measure parameters.

The average root mean squared errors for the two models using the Vasicek data sets are 0.0058, 0.0060, and 0.0059 for the three bonds, respectively, for the Vasicek model, and 0.0059, 0.0061, and 0.0060 for the CIR model.

Table 5. Monte Carlo statistics for 10 data sets generated with the Vasicek interest rate model.

	Vasicek				CIR			
	Mean	St. dev.	5%	95%	Mean	St. dev.	5%	95%
Panel A: True parameter values								
$\mu = 0.0108$	0.0129	0.0050	0.0047	0.0210	0.0087	0.0005	0.0079	0.0095
$\kappa = 0.1768$	0.2414	0.0825	0.1060	0.3773	0.1605	0.0484	0.0809	0.2401
$\sigma^2 \times 10^3 = 0.2561$	0.2650	0.0172	0.2380	0.2945	5.8999	0.3820	5.304	6.5529
$\mu^* = 0.0090$	0.0091	0.0004	0.0084	0.0097	0.0087	0.0005	0.0079	0.0095
$\kappa^* = 0.0685$	0.0690	0.0066	0.0582	0.0801	0.0608	0.0079	0.0481	0.0739
$\sigma_Y^2 \times 10^4 = 0.4$	0.3976	0.0126	0.3674	0.4290	0.5314	0.0199	0.4996	0.5649
$r_0 = 0.03$	0.0273	0.0056	0.0195	0.0354	0.0248	0.0041	0.0167	0.0301
Panel B: RMSE								
3 Months	0.0058				0.0059			
1 Year	0.0060				0.0061			
5 Years	0.0059				0.0060			
Panel C: Bayes factor 35.2								

Table 6. Monte Carlo statistics for 10 data sets generated with the CIR interest rate model.

	Vasicek				CIR			
	Mean	St. dev.	5%	95%	Mean	St. dev.	5%	95%
Panel A: True parameter values								
$\mu = 0.0095$	0.0126	0.0043	0.0055	0.0198	0.0095	0.0005	0.0086	0.0103
$\kappa = 0.1658$	0.2300	0.0704	0.1144	0.3454	0.1785	0.0368	0.1188	0.2386
$\sigma^2 \times 10^3 = 3.4428$	0.1972	0.0128	0.1772	0.2191	3.2991	0.2139	2.9645	3.6677
$\mu^* = 0.0095$	0.0097	0.0005	0.0089	0.0105	0.0095	0.0005	0.0086	0.0103
$\kappa^* = 0.0747$	0.0778	0.0077	0.0651	0.0905	0.0741	0.0085	0.0603	0.0881
$\sigma_Y^2 \times 10^4 = 0.4$	0.3792	0.0182	0.3595	0.4109	0.5093	0.0172	0.4817	0.5382
$r_0 = 0.03$	0.0281	0.0052	0.0148	0.0401	0.0278	0.0037	0.0219	0.0341
Panel B: RMSE								
3 Months	0.0059				0.0060			
1 Year	0.0061				0.0061			
5 Years	0.0059				0.0060			
Panel C: Bayes factor -17.4								

Similarly, using the CIR data sets, the root mean squared errors on average are 0.0059, 0.0061, and 0.0059 for the Vasicek model and 0.0060, 0.0061, and 0.0060 for the CIR model. Thus, for each data set the RMSEs of the two models are virtually indistinguishable, differing at most by only 1 or 2 basis points.

The Bayesian test result is unequivocal: on average, the log of the Bayes factor for the simulated Vasicek data is 35.2. The actual values range from 12.9 to 55.6, thus without exception the Bayesian methodology picks the right data-generating model. Similarly, for the CIR data the log of the Bayes factor is -17.4, on average. The actual values range from -39.2 to -8.2. These results confirm our empirical findings concerning the interest rate model being consistent with the bond yield data. Perhaps a more important conclusion is that Bayesian methods offer a more powerful tool for model selection.

4.2. Choice of priors

Our choice of priors was motivated by economic considerations related to the typical shape of the yield curve and

the dynamic behavior of the short rate. We were careful also to specify diffuse priors. Nevertheless, to rule out the possibility that our major result is a simple consequence of our prior specification we repeat the analysis with a less informative prior. Specifically, first we multiply the prior variance of each parameter by a factor of 10, and second we use flat priors on all parameters.

The results, displayed in table 7, reveal that our findings are very robust to the choice of priors. Without exception, the parameter estimates for the two models fall within the confidence intervals reported in table 2. The root mean square errors for all three yields are also virtually indistinguishable. Last, the Bayes factors of -17.4 and -18.8 are within one standard error of our original estimate.

4.3. Short rate proxy approach

Many of the previous studies that compare the two models, including some of those mentioned in the introduction, use a short maturity yield as a proxy for the unobserved instantaneous interest rate. To better

870

875

880

885

890

895

900

905

Table 7. Results using alternative priors. Prior 1 corresponds to 10 times the prior variances specified in section 3.2; prior 2 corresponds to flat priors on all parameters.

Parameter	Prior 1		Prior 2	
	Vasicek	CIR	Vasicek	CIR
μ	0.0095	0.0094	0.0092	0.0097
κ	0.1560	0.1563	0.1512	0.1597
$\sigma^2 \times 10^3$	0.2504	3.4797	0.2567	3.1856
μ^*	0.0090	0.0094	0.0090	0.0097
κ^*	0.0691	0.0728	0.0690	0.0765
$\sigma_Y^2 \times 10^4$	0.3148	0.4081	0.3148	0.4075
r_0	0.0311	0.0311	0.0311	0.0309
RMSE				
3 Months	0.00499	0.00481	0.00498	0.00498
1 Year	0.00327	0.00324	0.00326	0.00341
5 Years	0.00639	0.00653	0.00639	0.00647
Bayes factor	-17.4		-18.8	

that Treasury yield data very strongly favor the CIR model over the Vasicek model.

As argued by Johannes and Polson (2006), MCMC methods are particularly well-suited for term structure model calibration and offer several potential advantages over more traditional methods. One advantage is that MCMC directly computes the distribution of the latent variables and parameters given the observed data and does not rely on applying approximate filters or noisy latent variable proxies. Another advantage is that MCMC is based on conditional simulation, therefore avoiding any optimization or unconditional simulation. A third advantage is that it does not rely on asymptotic analyses and is therefore more immune to small sample biases. Yet another major advantage is that MCMC output may be used to compute the Bayes factor, an attractive alternative to classical- or frequentist-based tests for non-nested model comparison.

In subsequent research we hope to extend the analysis of this paper to multi-factor term structure models. In particular, we hope to compare the different classes of three- and four-factor affine models using these techniques.

Acknowledgements

The authors thank Ming-Hui Chen, Ren-Raw Chen, Joe Golec, John Harding, Scott Roark, Jimmy Sfiridis, John Vernon, Yazhen Wang, and Feng Zhao for many helpful comments.

References

Ball, C. and Tourus, W., Unit roots and the estimation of interest rate dynamics. *J. Empir. Finance*, 1996, **3**, 215–238.
 Bester, A., Random field and affine models for interest rates: An empirical comparison, Working Paper, Duke University, 2004.
 Carter, C.K. and Kohn, R., On Gibbs sampling for state space models. *Biometrika*, 1994, **81**, 541–553.
 Casassus, J. and Collin-Dufresne, P., Stochastic convenience yield implied from commodity futures and interest rates. *J. Finance*, 2005, **60**, 2283–2331.
 Chan, K., Karolyi, A., Longstaff, F. and Sanders, A., An empirical comparison of alternative models of the short-term interest rate. *J. Finance*, 1992, **47**, 1209–1227.
 Chen, M., Computing marginal likelihoods from a single MCMC output. *Statistica Neerlandica*, 2005, **59**, 16–29.
 Chen, R. and Yang, T., The relevance of interest rate processes in pricing mortgage-backed securities. *J. Hous. Res.*, 1995, **6**, 315–332.
 Chib, S., Marginal likelihood from the Gibbs output. *J. Am. Statist. Assoc.*, 1995, **90**, 1313–1321.
 Chib, S. and Ergashev, B., Analysis of multi-factor affine yield curve models. *J. Am. Statist. Assoc.*, 2009, **104**, 1324–1337.
 Chib, S. and Greenberg, E., Bayes inference in regression models with ARMA(p, q) errors. *J. Econometr.*, 1994, **64**, 183–206.
 Chib, S. and Greenberg, E., Understanding the Metropolis–Hastings algorithm. *Am. Statist.*, 1995, **49**, 327–335.
 Chib, S. and Greenberg, E., Markov chain Monte Carlo simulation methods in econometrics. *Econometr. Theory*, 1996, **12**, 409–431.

945

950

955

960

910

915

920

925

930

935

940

compare our findings with the findings of these earlier studies, we repeat our analyses using the three-month yield as a proxy for the instantaneous rate.

We set r equal to the three-month yield and let \tilde{Y} denote the one- and five-year yields, so that $Y = \{r, \tilde{Y}\}$. The parameter likelihood function,

$$L(\theta^* | Y) = P(r | \theta^*)P(\tilde{Y} | \theta^*, r),$$

is now a product of Gaussian densities and may be computed directly. The posterior ordinate $P(\theta^* | Y)$ is also simpler to compute. Because r is observed, $P(\theta_1^* | Y)$ and $P(\theta_2^* | \theta_1^*, Y)$ are estimated using the MCMC output with r fixed; furthermore, $P(\theta_3^* | \theta_2^*, \theta_1^*, Y)$ may be computed directly.

The point estimates from the MCMC output are given in table 3. For the parameters in the physical measure, the estimates are closer to those of CKLS (1992) and Chen and Yang (1995). The RMSEs for the 1- and 5-year yields are 0.00469 and 0.01000, respectively, for the Vasicek model, and 0.00470 and 0.01000, respectively, for the CIR model. For each of the two yields, the RMSEs again have very similar values across the two models and are virtually indistinguishable. The Bayes factor, however, is very much in favor of the CIR model:

$$\log \widehat{BF} = -126.2,$$

with an estimated standard error of 0.02. We again interpret this result as evidence that the RMSE is not a sufficiently refined measure to distinguish the two models.

5. Conclusion

This paper adds to the growing literature that uses Bayesian methods to estimate dynamic term structure models. In particular, we use MCMC methods to estimate the Vasicek and CIR models from U.S. Treasury bond yield data. We then demonstrate, and this is our main contribution, how to use the MCMC output to compare the two models by computing the Bayes factor. We find

- 1000 Chib, S. and Jeliazkov, I., Marginal likelihood from the Metropolis–Hastings output. *J. Am. Statist. Assoc.*, 2001, **96**, 270–281.
- Chib, S., Nardari, F. and Shephard, N., Markov chain Monte Carlo methods for stochastic volatility models. *J. Econometr.*, 2002, **108**, 281–316.
- 1005 Cochrane, J., *Asset Pricing*, 2001 (Princeton University Press: Princeton, NJ).
- Collin-Dufresne, P. and Harding, J., A closed form formula for valuing mortgages. *J. Real Estate Finance Econ.*, 1999, **19**, 133–146.
- 1010 Collin-Dufresne, P., Goldstein, R. and Jones, C., Can interest rate volatility be extracted from the cross section of bond yields? An investigation of unspanned stochastic volatility. *J. Financial Econ.*, 2009, **94**, 47–66.
- 1015 Cox, J., Ingersoll, J. and Ross, S., A theory of the term structure of interest rates. *Econometrica*, 1985, **53**, 385–407.
- Dai, Q. and Singleton, K., Specification analysis of affine term structure models. *J. Finance*, 2000, **55**, 1943–1978.
- Dai, Q. and Singleton, K., Expectation puzzles, time-varying risk premia, and affine models of the term structure. *J. Financial Econ.*, 2002, **63**, 415–441.
- 1020 de Jong, F., Time series and cross-section information in affine term-structure models. *J. Business Econ. Statist.*, 2000, **18**, 300–314.
- 1025 de Jong, F. and Santa-Clara, P., The dynamics of the forward interest rate curve: A formulation with state variables. *J. Financial Quant. Anal.*, 1999, **34**, 131–157.
- Diebold, F. and Mariano, R., Comparing predictive accuracy. *J. Business Econ. Statist.*, 1995, **82**, 339–350.
- 1030 Dudewicz, E., Mykytka, E., Ramberg, J. and Tadikamalla, P., A probability distribution and its uses in fitting data. *Technometrics*, 1979, **21**, 201–214.
- Duffee, G., Idiosyncratic variation of Treasury bill yields. *J. Finance*, 1994, **51**, 527–551.
- 1035 Duffee, G., Term premia and interest rate forecasts in affine models. *J. Finance*, 2002, **57**, 405–443.
- Duffee, G. and Stanton, R., Estimation of dynamic term structure models. Working Paper, Haas School of Business, University of California, 2004.
- Johannes, M. and Polson, N., MCMC methods for continuous-time financial econometrics. In *Handbook of Financial Econometrics*, edited by Y. Ait-Sahalia and L. Hansen, 2006 (Elsevier: Amsterdam).
- Karatzas, I. and Shreve, S., *Methods of Mathematical Finance, Applications of Mathematics*, 1988 (Springer: Berlin).
- 1045 Karatzas, I. and Shreve, S., *Brownian Motion and Stochastic Calculus*, 2nd ed., 1991 (Springer: Berlin).
- Kass, R. and Raftery, A., Bayes factors. *J. Am. Statist. Assoc.*, 1995, **90**, 773–795.
- Kau, J., Keenan, D., Muller, W. and Epperson, J., A generalized valuation model for fixed-rate residential mortgages. *J. Money, Credit Bank.*, 1992, **24**, 279–299.
- 1050 Kau, J., Keenan, D., Muller, W. and Epperson, J., The valuation at origination of fixed-rate mortgages with default and prepayment. *J. Real Estate Finance Econ.*, 1995, **11**, 5–36.
- 1055 Maes, K., Modeling the term structure of interest rates: Where do we stand? Working Paper, National Bank of Belgium, 2004.
- Newey, W. and West, K., A simple, positive-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 1987, **55**, 703–708.
- 1060 Pitt, M. and Shephard, N., Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.*, 1999, **94**, 590–599.
- Schwartz, E., The stochastic behavior of commodity prices: Implications for valuations and hedging. *J. Finance*, 1997, **52**, 922–973.
- 1065 Tauchen, G., The EMM reply to “Estimation of dynamic term structure models”. Working Paper, Duke University, 2004.
- Titman, S. and Torous, W., Valuing commercial mortgages: An empirical investigation of the contingent-claims approach to pricing risky debt. *J. Finance*, 1989, **44**, 345–373.
- 1070 Vasicek, O., An equilibrium characterization of the term structure. *J. Financial Econ.*, 1977, **5**, 177–188.