FACULTY OF
BUSINESS &
ECONOMICS

Melbourne Institute Working Paper Series

Working Paper No. 1/11

Predicting Short-Term Interest Rates: Does Bayesian Model Averaging Provide Forecast Improvement?

Chew Lian Chua, Sandy Suardi and Sarantis Tsiaplias

MELBOURNE INSTITUTE®
of Applied Economic and Social Research

Predicting Short-Term Interest Rates: Does Bayesian Model Averaging Provide Forecast Improvement?

Chew Lian Chua[†], Sandy Suardi[‡] and Sarantis Tsiaplias[§]

[†] Melbourne Institute of Applied Economic and Social Research,
The University of Melbourne

[‡] School of Economics and Finance, La Trobe University

[§] KPMG, Australia

Melbourne Institute Working Paper No. 1/11

ISSN 1328-4991 (Print)

ISSN 1447-5863 (Online)

ISBN 978-0-7340-4232-3

January 2011

Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
Victoria 3010 Australia
***Telephone* (03) 8344 2100**
***Fax* (03) 8344 2111**
***Email* melb-inst@unimelb.edu.au**
***WWW Address* http://www.melbourneinstitute.com**

Abstract

This paper examines the forecasting qualities of Bayesian Model Averaging (BMA) over a set of single factor models of short-term interest rates. Using weekly and high frequency data for the one-month Eurodollar rate, BMA produces predictive likelihoods that are considerably better than the majority of the short-rate models, but marginally worse off than the best model in each dataset. We observe preference for models incorporating volatility clustering for weekly data and simpler short rate models for high frequency data. This is contrary to the popular belief that a diffusion process with volatility clustering best characterizes the short rate.

JEL classification: G17, C11, C53

Keywords: Bayesian model averaging, out-of-sample forecasts

1 Introduction

The default-free short-term interest rate is one of the most commonly researched economic variables. It directly influences the short end of the term structure and, thus, has implications for valuing fixed income securities and derivatives. Furthermore, it is a general reference point for asset pricing on the basis that expected equilibrium returns on risky assets are expressed in terms of excess returns relative to the risk free rate. From a macroeconomic perspective, the short rate serves as an important input for business cycle analysis through the cost of credit, and its dynamics are to some degree governed by the stance of monetary policy and inflationary expectations. Given the vital role played by short-term interest rates in both the financial market and the economy, an enormous amount of work has been directed towards modelling and estimation of the short rate dynamics in the past three decades. Be that as it may, little consensus exists amongst financial practitioners about the appropriate choice of a short rate model both from a theoretical perspective and empirical application.

This paper examines the forecasting qualities of Bayesian Model Averaging (BMA) over a set of single factor models of short-term interest rates. An important contribution of the paper lies with the use of both low and high frequency short rate data. Using the same variable observed at apposite freqencies, we observe marked differences in the specifications of the models exhibiting the largest predictive likelihoods across the low and high frequencies. The differences in the preferred models across the two frequencies impact directly on the composition of the BMA measures of the short rate.

There are many contending short-term interest rate models which have been developed in the literature. The leading theoretical models specify continuous-time processes for the interest rate following the seminal work of Merton (1973) on the arithmetic Brownian motion representation. This specification, however, has been criticised for allowing negative interest rates and provides only, at best, a rough approximation to the actual process. The negative interest rate problem is overcome by Vasicek (1977) who imposes a mean-reversion (or stationarity) condition in the short rate model. Motivated by the observation that the discrete interest rate data display strong heteroskedasticity, Cox, Ingersoll and Ross (1985) (CIR) develop the square-root model of short rates which allows short rate volatility to peak with interest rate levels, the so-called 'level effect'. Both the Vasicek and CIR models provide closed form solutions and have been widely applied to discrete time data on short-term interest rates. Nevertheless, more flexible empirical specifications have been sought with the aim of obtaining an adequate characterisation of the actual short rate process. This has led Chan, Karolyi, Longstaff and Sanders (1992) (CKLS) to consider estimating the exponent parameter measuring the degree of level dependence in short rate volatility. They find a point estimate of $\frac{3}{2}$ which is in excess of unity, thus challenging the square-root model of CIR. More recently, evidence based on generalised autoregressive conditionally heteroskedastic (GARCH) models, developed by Engle (1982) and Bollerslev (1986), documents high degrees of volatility persistence in the interest rate process. Brenner, Harjes and Kroner (1994) (BHK) and Koedijk et al. (1994) nest the GARCH and approximate CKLS models under more general discrete-time specifications. These studies confirm the presence of rather extreme conditionally heteroskedastic volatility effects in the interest rate dynamics which tends to weaken the levels effect relative to the estimates from the CKLS model.

Although each of these models has been assessed using in-sample forecasts for their adequacy in characterising the behaviour of short-term interest rates, little is known about their relative out-of-sample forecast performance.[1] While it is possible that a particular short rate model may forecast interest rate movements more accurately in certain periods than other models, it is equally likely that its forecasting performance may diminish and be outperformed by another short rate model in other periods. In other words, the best forecasting model can change over time. An example is the high point estimate of the elasticity of variance in the CKLS model

[1] To our knowledge there has been no empirical study which has performed an evaluation of the out-of-sample forecast performance of short-rate models within a Bayesian framework. The work of Sanford and Martin (2006) explores a limited number of short-rate models that differ only in the level effect parameter for volatility. The forecast performance of the models is evaluated using Bayes factors that are, however, constructed using in-sample information.

which Bliss and Smith (1998) have argued is attributed to the volatile and high interest rate levels arising from a change in the Federal Reserve operating procedure from targeting interest rate levels to monetary aggregates during the 1979-1982 period. Using the CKLS model to forecast the short rate outside of the 1979-1982 period may not result in accurate forecasts given that the elasticity of variance estimate is likely to differ outside the aforementioned period. Another point of contention which often arises in the short rate literature is the linearity of the short rate drift specification. While a large proportion of the research reports a linear drift (Chan et al., 1992), others argue to the contrary (Ait Sahalia, 1996; Conley et al., 1997; Jones, 2003) finding nonlinear mean reversion. Bali and Wu (2006) provide evidence that the speed of mean-reversion for short-term interest rates at extremely high interest rates, such as in the Volcker (1979-1982), differs relative to that observed during periods of normal rates. It is these differences in the degree of mean-reversion at different interest rate levels which generate the nonlinearity in short-rate drift. Given the ambiguity in choosing a representative short-term interest rate model, and the problem of model and parameter uncertainty, examining a model's out-of-sample forecast performance and exploiting the potential gain in forecast accuracy by combining predictions from individual models are useful avenues of research from a financial practitioner's viewpoint.

To address these questions the BMA framework provides a suitable method for assessing the individual and combined forecast performance of the above mentioned models. The researcher does not know which of the short-term interest rate models is the true model. Using the researcher's chosen prior about which model is true, the posterior probability that a model is true can be computed. The combined forecast based on all available models can then be obtained by weighting the model forecasts using the model posterior probabilities. The flexibility offered by BMA through its judicious combination of information contained in different models has made it particularly attractive for forecasting. In economics and finance, we have observed pervasive applications of BMA in different areas including, amongst others, output growth forecasting (Koop and Potter, 2003), cross-country growth regressions (Doppelhofer et al., 2000;

Fernandez, 2001), exchange rate forecasts (Wright, 2008), portfolio management (Pesaran and Zaffaroni, 2004) and stock returns (Avramov, 2002; Cremers, 2002). By and large, all of these studies have shown that BMA provides improved out-of-sample predictive performance.

In this paper, we employ predictive likelihoods to compare model forecasts. As pointed out by Geweke and Whiteman (2006), the predictive likelihood which stores the out-of-sample prediction record of a model forms a basis for rigorous model evaluation. We also assess the BMA forecast based on five different methods of combining predictions from individual short rate models, one of which is the simple model average which assumes equal weighting across all models through time. The other four BMA methods adopt different posterior model probability assumptions; in principle, they are based on either the marginal likelihood over the estimation period (or in-sample fit) or the predictive likelihoods computed over the forecast period.

The remainder of the paper is structured as follows. Section 2 provides a survey of the short-term interest rate models. Section 3 describes the data and the generalised short rate model which nests twelve short rate models considered in this paper. Section 4 develops a Bayesian method for estimating parameters of discretely observed short rate processes. It also outlines the different methods of combining forecasts using BMA. Section 5 discusses the empirical results, and section 6 concludes.

2 Models of the Short-Term Interest Rate

A standard representation of the short-term interest rate dynamics is described by the following stochastic differential equation:

$$dr_t = \mu(r_t)dt + \sigma(r_t)dW_t, \tag{1}$$

which suggests that the change in the short rate can be decomposed into a drift over the time period $(t, t + dt)$ - namely, $\mu(r_t)dt$ - and a random shock represented by an increment of a

Brownian motion, dW_t, with an instantaneous diffusion of $\sigma(r_t)$. It is common in most empirical work to allow the drift and diffusion to depend on the short rate only.[2] In fact, various one-factor models have been constructed by specifying the drift, $\mu(r_t)$, and diffusion, $\sigma(r_t)$, of the short rate in different ways. This is best summarised through the generalised continuous time short rate model of Chan et al. (1992) which nests many single factor models. Specifically, the model is

$$dr_t = (\mu + \lambda r_t)dt + \sigma r_t^{\gamma} dW_t, \tag{2}$$

where all the parameters are generally supposed to be non-negative, apart from λ for which a negative value induces a mean reverting effect on r_t. The key differences between the models is summarised in Table 1.

– Table 1 about here –

The first two models namely, Merton (1973) and Vasicek (1977) imply Gaussian proceses with constant diffusion (or volatility). However, unlike the Merton (1973) model the Vasicek (1977) model is less likely to suffer from the drawback of negative interest rates for certain appropriate values of μ, λ and σ. Cox, Ingersoll and Ross (1985) relax the constant diffusion assumption by adopting a square-root term in the volatility of the short rate change. This process has a reflecting boundary at $r_t = 0$ if $2\mu \geq \sigma^2$, thus it precludes interest rates from being negative. Like the CIR model, the other short rate models define short rate diffusion as a function of interest rate levels with different elasticity of variance, γ, ranging from 0.5 (Cox et al., 1985) to 1.5 (Cox et al., 1980). In some of these models, such as the constant elasticity of variance (CEV) model of Cox (1975) and the CKLS model, the value of γ is estimated from the data.

All of the short-rate models reported in Table 1 assume a linear drift specification. A linear

[2] A two factor short rate model has featured in the literature such as the two factor Black, Derman and Toy (1990) model developed by Bali (2003). We do not consider this model as the focus of our paper is on a class of single factor short rate models.

drift implies that the strength of mean reversion is the same for all levels of the short rate. Even though there is no *a priori* economic intuition that would suggest the existence of a nonlinear drift, empirical research has shown that there is evidence of nonlinear drift in short-term interest rates, that is, mean reversion is stronger for extreme low or high levels of short rate. Ait-Sahalia (1996) advocates the use of a flexible functional forms to approximate the true unknown shapes of short rate process. He constructs a general specification test of a short rate model of the form

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \frac{\alpha_3}{r_t})dt + \sqrt{\beta_0 + \beta_1 r_t + \beta_2 r_t^{\beta_3}} dW_t, \tag{3}$$

and finds that the test rejects a linear drift in favour of models that imply no mean reversion for levels of the short rate between 4 and 22 percent, and strong mean reversion for extreme levels of the short rate. Conley et al. (1997) adopt the same drift parameterisation as Ait-Sahalia but keeps the CEV diffusion used by CKLS:

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \frac{\alpha_3}{r_t})dt + \sigma r_t^{\gamma} dW_t. \tag{4}$$

They find that the drift function displays mean-reversion only for rates below 3% or above 11%. Using a different estimation approach and data series but a comparable drift function, Stanton (1997) demonstrates results that are roughly consistent with those of Ait-Sahalia for high levels of interest rate. He finds that there is very little mean reversion for all rates below 15% but substantial negative drift for higher rates. Bali and Wu (2006), estimate a variant of the drift specification in (3) which includes a fifth order polynomial. They find that the statistical significance of nonlinearity in the drift function of the 3-month Treasury yield and the 7-day Eurodollar rate is reduced with the incorporation of GARCH volatility and the specification of non-normal innovation. In fact, these added features of the conditional variance and innovation specification can fully account for the nonlinearity observed in the drift dynamics of the 3-month Treasury yield and the 7-day Eurodollar rate.

Another criticism of short rate models in Table 1 is that, while it allows volatility to depend on interest rate levels, it does not incorporate the observation of volatility clustering in short rates. Nor does it allow past interest rate shocks to affect current and future volatility. To accommodate the strength of both the CEV and the GARCH models, Brenner, Harjes and Kroner (1996) (BHK) consider two ways of combining the two models into a more general form. One way is to adopt the functional form of the CEV model while allowing interest rate volatility to follow a GARCH process. Applying the Euler-Maruyama discrete time approximation to (2) gives

$$\Delta r_t = \mu + \lambda r_{t-1} + \varepsilon_t. \tag{5}$$

Let Ω_{t-1} represent the information set available at time $t-1$ and that $E\left(\varepsilon_t|\Omega_{t-1}\right) = 0$. Suppose h_t represent the conditional variance of the short-term interest rate changes then $E\left(\varepsilon_t^2|\Omega_{t-1}\right) \equiv h_t = \sigma^2 r_{t-1}^{\gamma}$. BHK relax the assumption of a constant σ^2 by allowing it to vary according to information arrival process following a GARCH(1,1) model:

$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2. \tag{6}$$

The innovation ε_t denotes a change in the information set from time $t-1$ to t and can be treated as a collective measure of unanticipated news. In (6) only the magnitude of the innovation is important in determining σ_t^2. We refer to this model as the GARCH-CEV model. Another way is by allowing information from unanticipated news and the one-period lagged interest rate levels to govern the dynamics of short rate volatility in the following manner:

$$\begin{aligned} \Delta r_t &= \mu + \lambda r_{t-1} + \varepsilon_t, \\ E\left(\varepsilon_t|\Omega_{t-1}\right) &= 0,\ E\left(\varepsilon_t^2|\Omega_{t-1}\right) = h_t, \\ h_t &= \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 h_{t-1} + b r_{t-1}^{\gamma}. \end{aligned} \tag{7}$$

We refer to this model as the GARCH-Affine model. Under the restriction $\beta_0 = \beta_1 = \beta_2 = 0$, the GARCH-Affine model collapses to the CKLS model where $b = \sigma^2$ and that volatility depends on interest rate levels alone. Furthermore when $b = 0$ then there is no levels effect.

Both the GARCH-CEV and GARCH-Affine models do not permit short rate volatility to respond asymmetrically to interest rate innovations of different sign. BHK relax the assumption of a symmetric GARCH process in both models. Specifically, the conditional variance specification in the GARCH-Affine model can be augmented as

$$h_t = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 h_{t-1} + b r_{t-1}^{\gamma} + \beta_3 \xi_{t-1}^2, \tag{8}$$

where $\xi_{t-1} = \max(0, \varepsilon_{t-1})$. Note that $\xi_t = \max(0, \eta_t)$ captures the asymmetric response of short-rate volatility to bad and good news, where bad news that is associated with an interest rate hike elicits greater volatility than good news of an interest rate cut of the same magnitude.

3 The Generalised Short Rate Model and Data

3.1 The Generalised Short Rate Model

For equation (1) to nest the nonlinear mean-reverting model of Ait-Sahalia (1996) and the combined GARCH and levels effects model of BHK (1996), we can express (1) in its discrete time approximation as follow:

$$\Delta r_t = X_{it} A_i \triangle t +_t \sqrt{f\left(Z_{it}, B_i\right) \triangle t}, \tag{9}$$

$$E(\eta_t^2 | \Omega_{t-1}) = g(W_{it} G_i),$$

where $f(\cdot)$ and $g(\cdot)$ denote some linear or nonlinear functions describing the relationships between Z_{it} and B_i, and W_{it} and G_i, respectively. The drift term $X_{it} A_i$ comprises X_{it} which is a vector of short-term interest rate variables with its corresponding vector of coefficients, A_i. The diffusion process is made up of two components namely, the elasticity of volatility and the news arrival process, each defined by $\sqrt{f\left(Z_{it}, B_i\right)}$ and $g(W_{it}G_i)$, respectively. The Z_{it} comprises lagged interest rates and W_{it} is made up of a vector of past conditional variance, squared innovations and lagged interest rates. B_i and G_i are coefficient vectors. We assume ε_t follows a Student's t distribution with degree of freedom v for $v > 2$ since the short rate distribution is known to depart from normality. For the purpose of our analysis we assume a specific function for both $f(\cdot)$ and $g(\cdot)$ such that (9) is

$$\Delta r_t = \left(\alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \frac{\alpha_3}{r_{t-1}}\right) \triangle t + \varepsilon_t \sqrt{\left(\beta_0 + \beta_1 r_t + \beta_2 r_t^{\beta_3}\right) \triangle t}, \tag{10}$$

$$E(\varepsilon_t^2|\Omega_{t-1}) = h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 h_{t-1} + \gamma_3 \xi_{t-1}^2 + \gamma_4 r_{t-1},$$

where the variables and parameters are defined in the same way as discussed in the previous section. It can be seen that by imposing certain restrictions in the parameters of equation (10), this generalised short-rate model nests twelve short-rate models which are outlined in section 2. Table 2 summarises the twelve short rate models which are nested in the generalised specification. The corresponding parameter restrictions to obtain a specific short-rate model are also provided.

– Table 2 about here –

3.2 Data Description

The empirical investigation is based on weekly and 15-minute tick observations on the annualised one-month Eurodollar deposit rates. The weekly data, which are obtained from Datastream, are sampled from February 1975 to December 2008. The 15-minute tick data are taken from the one-month Eurodollar futures prices for the period 19 May 2009 to 29 September 2009. This dataset, which is obtained from Thomson Reuters Tick History, is not available for the same period as the weekly data. Although the two datasets cover two non-overlapping periods, this does not impose any problem since our purpose is to determine whether the forecast performance of short rate models is sensitive to the data frequency employed. Previous studies on short rates have used different interest rates ranging from the Eurodollar deposit rate (AitSahalia, 1996; Bali, 2003; Christiansen, 2008) to the more commonly used Treasury bill rates (Andersen and Lund, 1997; Koedijk et al., 1997; Suardi, 2008). Furthermore, different maturity periods of the money market rates and Treasury bill rates have been employed in empirical studies; some use the 7-day rates while others use the 13-week (or 30-day) rates. Our choice of weekly frequency is motivated by the problem of discretization bias associated with lower frequency data like monthly data.[3] Stanton (1997) shows that when one-factor diffusion models are estimated with an Euler approximation, discretization error is negligible with daily and weekly data. Jones (2003) suggests that augmentation is most important when using monthly data for estimation, while daily and weekly data produce little discretization bias. Our weekly data period comprises a shift from historically high interest rates in the late 1970s to early 1980s during the Volcker monetary regime to low interest rate levels in the latter part of the sample period. The interest rate data and the first differenced series are presented in Figure 1. Summary statistics for the data set are provided in Table 3.

- Figure 1 about here -

In Figure 1(a) it can be seen that there is a tendency for the volatility in the interest rate series to be positively correlated with the current level of the rate. At the start of the sample period, the association between the level of interest rate and its volatility is visible. This feature becomes more apparent for the 1979-1983 period during which both the level and volatility of the rate are high. The level effect is not as obvious after the Volcker monetary regime. These empirical features tally with those reported in Brenner et al. (1996). The time-varying nature of the volatility in the sample is indicative that unexpected 'news' might be equally important

[3]Data augmentation involves manufacturing a higher-frequency dataset than that actually observed, usually by simulating the path of the stochastic process between observed data points (Gray, 2005).

in explaining the volatility of interest rates, in addition to the level effect. With the exception of the Volcker's monetary regime, the impact of Lehman's Brothers collapse which panicked global bankers and caused the Euro deposit rate to sky rocket on September 11, 2008 is most noticeable. The Euro deposit rate jumped from 2.6% to 6% on that day. Referring to Figures 1(a) and (b), it can be seen that the degree of volatility clustering is less apparent in high frequency data compared with weekly data, although there is evidence that the stylised feature of levels effect is still prevalent in high frequency data. Between observations 200 and 300, Euro deposit rate increases to about 1.7% and this is associated with acute volatility in short rate changes.

- Table 3 about here -

For both datasets, 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. The relevant kurtosis statistic reported in Table 3 is significantly greater than the value of 3 associated with the normal distribution. The positive skewness coefficient is also more than the value of zero associated with the symmetric normal distribution. This is reflective of a 'leverage' effect of sorts, whereby interest rate rises are associated with higher volatility than decreases of the same magnitude. The first differenced data exhibit strong correlation as shown by the Ljung-Box test statistic which overwhelmingly rejects the null hypothesis of no serial correlation at the fifth and tenth lag order. The interest rate series clearly possesses conditional heteroskedasticity as indicated by application of a formal fifth and tenth order LM test for ARCH to the residuals from an AR(10) regression of the interest rate data. The Jarque-Bera test strongly rejects the null of normality in the interest rate series. The ADF statistics indicate rejection of the null hypothesis of a unit root for the level of weekly Euro deposit rates at the 5% level. On the other hand, there is evidence to suggest that the 15-minute Euro deposit rates is non-stationary. However, for changes in the Euro deposit rates the unit root test statistic rejects the unit root null at the 1% level in both datasets.

4 Bayesian Inference

Consider a model based on (9) which we denote as model M_i. The likelihood function for M_i is given by $p(R_T|\theta_i, M_i)$, where R_T is the set of annualised one-month Eurodollar deposit rates $(r_1, r_2, ..., r_T)$ and θ_i is the parameter vector corresponding to M_i. Given the specification of a prior density $p(\theta_i|M_i)$, the posterior density of θ_i is available using

$$p(\theta_i|R_T, M_i) = \frac{p(R_T|\theta_i, M_i)p(\theta_i|M_i)}{\int p(R_T|\theta_i, M_i)p(\theta_i|M_i)d\theta_i}. \tag{11}$$

Since ε_t in (9) follows a Student-t distribution, the likelihood $p(R_T|\theta_i, M_i)$ is given by

$$p(R_T|\theta_i, M_i) = \prod_{t=2}^{T} p(r_t|R_{t-1,}\theta_i, M_i) = \prod_{t=2}^{T} f\left(X_{it}A_i \triangle t, \sqrt{g(W_{it}G_i)f\left(Z_{it}, B_i\right) \triangle t}, v_i\right), \qquad v > 2, \tag{12}$$

where R_{t-1} is the set of interest rates to time $t-1$, and $f\left(\mu, \sigma^2, v\right)$ is the student's t density function with mean μ, variance σ^2, and degrees of freedom v.

To provide a more tractable posterior density and facilitate estimation, we adopt the normal mixture representation of the Student-t distribution.[4] Pursuant to this representation, the independent Student-t density may be represented as a heteroscedastic normal model such that

$$p(r_t|R_{t-1,}\theta_i, M_i) = N\left(X_{it}A_i \triangle t, \frac{g(W_{it}G_i)f\left(Z_{it}, B_i\right) \triangle t}{\lambda_{it}}\right), \tag{13}$$

where the mixing variable, λ_{it}, is i.i.d. Gamma distributed

$$\lambda_{it}|v_i \sim Gamma(\frac{v_i}{2}, \frac{v_i}{2}).$$

The i subscripts for λ and v are omitted hereafter for notational convenience. This representation of the independent Student-t density facilitates estimation of the model parameters,

[4]Represent the Student-t distribution as a normal mixture model converts a non-log concave sampling problem into a log concave sampling problem (Polson, 1996).

subject to the introduction of a T-dimensional latent parameter vector $\lambda = (\lambda_1, \lambda_2, ...\lambda_T)$. The parameters to be estimated for M_i are, therefore, $\theta_i = \{A_i, B_i, G_i, \lambda, v\}$.

Prior elicitation for θ_i is reasonably straightforward. In our specification, the prior density is decomposed as

$$p(\theta_i|M_i) = p(A_i)\, p(B_i)\, p(G_i)\, p(v_i)\,, \tag{14}$$

where each density on the right hand side of (14) is proper but fairly uninformative.[5]

The prior for A_i, $p(A_i)$, is normally distributed with mean $\underline{A}_i$ and covariance $\underline{\Sigma}_{A_i}$. Each element of B_i, being $\beta_{0i}, \beta_{1i}, \beta_{2i}$ and β_{3i}, has the following inverse gamma prior

$$\beta_{ji}|v_{\beta_{ji}}, s_{\beta_{ji}} \sim Inverse\ Gamma(v_{\beta_{ji}}, s_{\beta_{ji}}), \tag{15}$$

for $j = 0, 1, 2, 3$. The prior for B_i, $p(B_i)$, is given by the product of these densities. We place non-informative but proper truncated uniform priors for G_i such that $\gamma_{0i} \sim U(0, 100)$, and $\gamma_{1i}, \gamma_{2i}, \gamma_{3i}, \gamma_{4i} \sim U(0, 1)$. Finally, we follow Geweke (1993) in specifying an exponential prior distribution for v

$$p(v_i) = \tau_i \exp(-v_i \tau_i) I_{v_i}, \tag{16}$$

where I_{v_i} is an indicator function such that $I_{v_i} = 1$ if $v_i \geq 2$ and $I_{v_i} = 0$ otherwise.

4.1 Sampling Scheme

As the posterior density cannot be sampled from using a known distribution, a combination of the Gibbs Sampler and the Metropolis-Hastings (MH) algorithm is used to obtain draws from the posterior $p(\theta_i|R_T, M_i)$. The sampling scheme involves iterating through the following five steps:

1. Draw A_i from $p(A_i|\theta_i \neq A_i, R_T, M_i)$ which is a multivariate normal distribution.

[5] Each of the densities on the right hand side of (14) is conditional on M_i. This conditioning is omitted here and in the remainder of this section for notational convenience.

2. Draw B_i from $p(B_i|\theta_i \neq B_i, R_T, M_i)$. This is an inverse gamma distribution for most of the models we consider.[6]

3. Draw λ_t from $p(\lambda_t|\theta_i \neq \lambda_t, r_t, M_i)$ for $t = 1, ..., T$. The conditional distribution of λ_t is a gamma distribution.

4. Draw G_i from $p(G_i|\theta_i \neq \gamma, R_T, M_i)$ using a MH algorithm.

5. Draw v_i from $p(v_i|\theta_i \neq v_i, R_T, M_i)$ using a MH algorithm.

Appendix A provides details about the drawing process at each step. Depending on the restrictions imposed on a particular model, some steps in the sampling scheme may not be performed. For instance, step 4 is not required for M_1 to M_{10} due to non-existence of GARCH processes. Table 4 summaries the actual steps taken for each model.

- Table 4 about here -

At the completion of each pass of the sampler a draw $\theta_i^{(j)}$ is obtained and, following convergence, $\theta_i^{(j)} \sim p(\theta_i|R_T, M_i)$.[7] We iterate through the sampling scheme until N draws of θ_i are available. To eliminate any dependence on the initial conditions of the Markovian chain, the first N_0 draws are discarded. Given $\left\{\theta_i^{(N_0+1)}, ...\theta_i^{(N)}\right\}$, estimates of moments and other quantities of functions of θ_i (assuming they exist) are straightforward to obtain; the expected value of a real-valued function $g(\theta_i)$ under the posterior distribution may be estimated using the sample mean of $g(\theta_i)$

$$\frac{1}{N - N_0} \sum_{j=N_0+1}^{N} g(\theta_i^{(j)}). \underset{a.s.}{\longrightarrow} E_{p(\theta_i|R_T, M_i)} g(\theta_i). \tag{17}$$

[6]The conditional pdf is not recognsible for models M_1, M_2 and M_6, thus we employ a MH alogarithm instead.

[7]See Geweke (1993) for a proof regarding convergence of the distribution of $\theta_i^{(j)}$ to the posterior distribution that is also applicable here.

4.2 Model Averaging

The marginal likelihood for M_i is the value of the likelihood function after integrating out the random parameter vector θ_i pursuant to

$$p(R_T|M_i) = \int p(R_T|\theta_i, M_i)p(\theta_i)d\theta_i. \tag{18}$$

Given $p(R_T|M_i), i = 1, 2, ...J$, pairwise comparisons between any two models M_i and M_j may be undertaken using Bayes factors and posterior odds ratios. Posterior probabilities for each of the J models under consideration are also straightforward to compute as

$$p(M_i|R_T) = \frac{p(R_T|M_i)p(M_i)}{\sum_{k=1}^{K} p(R_T|M_k)p(M_k)}, \tag{19}$$

where $p(M_i)$ is the probability that the data generating process is given by M_i before observation of the dataset R_T, and $p(M_i|R_T)$ is the probability that the data generating process is given by M_i after observation of the dataset R_T.

A related measure, the predictive likelihood, is used to determine the distribution of $\{r_{T+1}, ..., r_{T+S}\}$ prior to their observation (given R_T and M_i), in addition to providing the likelihood of $\{r_{T+1}, ..., r_{T+S}\}$ after these values are observed (again, given R_T and M_i). The predictive likelihood for $\{r_{T+1}, ..., r_{T+S}\}$ is

$$p(r_{T+1}, ..., r_{T+S}|R_T, M_i) = \int p(r_{T+1}, ..., r_{T+S}|\theta_i, R_T, M_i)p(\theta_i|R_T, M_i)d\theta_i. \tag{20}$$

It is clear from (20) that the predictive likelihood accounts for any parameter uncertainty in the posterior distribution of θ_i. In practice, an estimate of $p(r_{T+1}, ..., r_{T+S}|R_T, M_i)$ can be obtained from the posterior draws $\left\{\theta_i^{(N_0+1)}, ...\theta_i^{(N)}\right\}$ as

$$p(r_{T+1}, ..., r_{T+S}|R_T, M_i) = \frac{1}{N - N_0} \sum_{j=N_0+1}^{N} p(r_{T+1}, ..., r_{T+S}|\theta_i^{(j)}, R_T, M_i).$$

We can obtain a model-free estimate of the predictive likelihood by integrating across each model $\int p(r_{T+1}, ..., r_{T+S}, M_i|R_T)dM_i$. Since, there are a discrete number of models, this is equivalent to weighting each model's predictive likelihood according to its posterior probability

$$p(r_{T+1}, ..., r_{T+S}|R_T) = \sum_{k=1}^{K} p(r_{T+1}, ..., r_{T+S}|R_T, M_k)p(M_k|R_T). \tag{21}$$

Updating the posterior probability to generate model-free estimates of the predictive likelihood at each time step is straightforward. Given the current posterior probability $p(M_i|R_T)$, any new information associated with M_i is embedded in the one step-ahead predictive likelihood evaluated after the observation of r_{T+1}. Consequently, the posterior probability $p(M_i|R_{T+1})$ may be obtained as

$$p(M_i|R_{T+1}) = p(M_i|r_{T+1}, R_T) = \frac{p(r_{T+1}|R_T, M_i)p(M_i|R_T)}{\sum_{k=1}^{K} p(r_{T+1}|R_T, M_k)p(M_k|R_{T-1})}. \tag{22}$$

Liu and Maheu (2009) use this recursive BMA approach to forecast realised volatility.

A potential drawback associated with the use of (19) and (22) to construct model probabilities is that the probabilities may place too little emphasis on recent model performance. Since (19) depends on model performance for all t, a large posterior probability may be associated with a model that may have performed poorly in the most recent h time periods. This problem is more likely to appear when T is a large number (as is the case in this paper). An alternative is to limit the data available for computation of the posterior probability to a subset of recent predictive likelihoods. This ensures that posterior probabilities are heavily dependent on recent model performance. Pursuant to this logic, the posterior probability for M_i may be computed as

$$p(M_i|R_{T-h:T}) = \frac{p(R_{T-h:T}|M_i)p(M_i)}{\sum_{k=1}^{K} p(R_{T-h:T}|M_k)p(M_k)}, \tag{23}$$

where $p(R_{T-h:T}|M_i) = \prod_{t=T-h}^{T} p(r_t|R_{t-1}, M_i)$ depends only on the last $h+1$ predictive likelihoods.

5 Empirical Application and Results

5.1 Bayesian Model Averaging Application

We consider the performance of the short rates using in-sample and out-of-sample methods. To undertake the performance analysis, we divide the sample into three periods: a training period T_T, an estimation period T_E and a forecasting period T_F. The training period spans 9 January 1975 to 31 January 1985 and is used to construct a prior for the estimation period. The estimation period covers 7 February 1985 to 29 December 2005 and is used to obtain the posterior density of the parameter vector and an in-sample model comparison. The forecasting period is 5 January 2006 through to 18 December 2008 and is used to conduct a real time out-of-sample forecasting exercise.

The empirical application is carried out as follows. We produce $N = 20000$ posterior draws for θ_i using the data spanning the training period T_T. We apply a burn-in of $N_0 = 5000$ draws. The prior hyperparameters for the training period are largely uninformative, and are: 1) the individual elements of $A_i \sim N(0, 100)$, 2) $\beta_{ji} \sim Inverse\ Gamma(2, 1) \forall j$, and 3) $v_i \sim exponential(0.025) I_{v_i}$.

The density $p(\theta_i|T_T, M_i)$ based on the training sample is used to construct the prior for θ_i for the estimation period T_E. This approach is advocated in Geweke (1994) and generates a prior for the parameters that is almost always well defined irrespective of the prior specification for the training period (the protoprior). Given the length of the training period, we have observed that the effect of the choice of protoprior on the posterior density $p(\theta_i|T_E, M_i)$ is typically negligible. Specifically, the hyper parameters $\underline{A}_i$ and $\underline{\Sigma}_{A_i}$ for T_E are the posterior means and variances of A_i based on $p(\theta_i|T_T, M_i)$. To derive the hyperparameters $v_{\beta_{ji}}, s_{\beta_{ji}}$ used in the derivation of the prior for β_{ji} over the estimation period we hypothesise that there is a probability ϕ_{ji} of β_{ji} lying beyond $\overline{\beta^T_{ji}}$, where $\overline{\beta^T_{ji}}$ is the posterior mean of β_{ji} estimated using the training sample.

According to this hypothesis

$$p(\beta_{ji} \geq \overline{\beta_i^T}) = \phi_{ji}$$
$$\iff p\left(\frac{v_{\beta_{ji}} s_{\beta_{ji}}}{\beta_{ji}} \leq \frac{v_{\beta_{ji}} s_{\beta_{ji}}}{\overline{\beta_i^T}}\right) = p\left(\chi^2_{v_{\beta_{ji}}} \leq \chi^2_{v_{\beta_{ji}},1-\phi_{ji}}\right) = \phi_{ji}. \quad (24)$$

To satisfy (24) $s_{\beta_{ji}}$ is set to $v_{\beta_{ji}}^{-1}\overline{\beta_i^T}\chi^2_{v_{\beta_{ji}},1-\phi_{ji}}$, where $\chi^2_{v_{\beta_{ji}},1-\phi_{ji}}$ is obtained from a standard Chi-square table. The values ϕ_{ji} and $v_{\beta_{ji}}$ are determined by the researcher. In this paper, we assume that $v_{\beta_{ji}} = 2$ and $\phi_{ji} = 0.3$ for all j and i. By setting $v_{\beta_{ji}} > 1$, we impose the restriction that the first moment of the prior distribution of β_{ji} exists. The hyper parameter τ_i used to determine the prior for the degree of freedom parameter v is given by $\tau_i = \frac{1}{\overline{v_i^T}}$, where $\overline{v_i^T}$ is the posterior mean of v_i derived from $p(\theta_i|T_T, M_i)$.

The data spanning the estimation period T_E coupled with the training-based prior $p(\theta_i|T_T, M_i)$ is used to obtain the posterior density of the parameters $p(\theta_i|T_E, M_i)$. For each draw from $p(\theta_i|T_E, M_i)$ based on the sampler defined in Section 4.1, we generate up to $q = 8$ step ahead forecasts together with the associated predictive likelihoods. At each time step, we add an additional period of data to T_E before re-estimating the posterior $p(\theta_i|T_E, M_i)$ and generating another q-step ahead forecasts and the predictive likelihoods. This conventional recursive procedure continues until 23 October 2008.

As an alternative, we also repeat the exercise conditional on a diffuse prior for θ_i that does not depend on the observation of $p(\theta_i|T_T, M_i)$. This provides information concerning the sensitivity of the forecasts and predictive likelihoods to the specification of the prior. In this respect, we consider the following non-informative but proper prior specification: each element of A_i is normally distributed with zero mean and a variance of 100; for $\beta_{ji}\forall j$ there is a 99 per cent probability β_{ji} lying beyond $\overline{\beta_{ji}^T}$ given $v_{ji} = 2$; the τ_i hyper parameter is set to $1/100$.

To evaluate the predictive capacity of Bayesian model averaging, we consider five model averaging specifications denoted BMA_1 to BMA_4 and *Simple* MA. BMA_1 is derived from expressions (19) and (22) with the prior model probabilities computed from the first twenty 1-

step ahead predictive likelihoods for the 12 models under consideration. The marginal likelihood used to construct (19) is based on the data covering the period T_E and is estimated using the Modified Harmonic Mean (MHM) method of Gelfand and Dey (1994). The MHM method is detailed further in Appendix B. The posterior model probabilities are updated as new data arrives using (22). BMA_2 is derived in an analogous fashion to BMA_1 but subject to the adoption of equal prior model probabilities.

The third model averaging specification, BMA_3, is based on (23) with equal prior probabilities for each of the 12 models. The element $p(R_{T-h:T}|M_i)$ in (23) covers only the predictive likelihoods over the time period T_F. In other words, model choices are undertaken only by reference to the cumulative predictive capacity of the models over the forecast period. Fourth, BMA_4 is akin to BMA_3 except a rolling window of the last 20 predictive likelihoods is adopted (i.e. h fixed at 19). This imposes the restriction that only recent forecasting performance is considered in determining each model's weight. Lastly, *Simple MA* assumes that each model is always given an equal weight irrespective of its predictive likelihood values.

This exercise was also repeated for the high frequency data across the twelve specified models. The high frequency training, estimation and forecasting periods (i.e. the high frequency equivalents of T_T, T_E and T_F) are 19 May 2009 to mid day of 24 July 2009 (500 observations), mid day of 24 July 2009 to mid day of 22 September 2009, and mid day of 22 September 2009 to mid-day of 29 September 2009. This provides two sets of results, including two sets of marginal likelihoods and posterior model probabilities, which may be used to examine the impact (if any) of data frequency on both the choice of short rate model and the performance impact of BMA.

5.2 Results for Weekly Data

According to the model predictive likelihoods, models M_{11} and M_{12} perform best for the weekly Eurodollar rate over the forecast period. This forecast improvement is observed across each of the 8 step-ahead periods; consequently, there is little evidence to suggest that alternative models perform better at 1-step ahead forecasting relative to 8-step ahead forecasting notwithstanding the 7-week difference between the forecast horizons. Moreover, as mentioned above, the impact of the protopriors on the cumulative predictive likelihoods is small, with the order rankings being identical using either informative or diffuse priors. Consequently we present only the cumulated log predictive likelihoods for the informative protopriors (see Table 5).[8]

- Table 5 about here -

A few results are worth highlighting. Holding the diffusion process unchanged and comparing the cumulated predictive likelihood of models M_7 and M_3 (or M_6 and M_2) indicates that a short-rate model with linear mean-reverting drift is preferred to a non-stationary model. On the other hand, keeping the linear drift specification constant and allowing the elasticity of variance parameter to vary across models M_2 to M_5 shows that the model of Cox, Ingersoll and Ross (1985) performs the worst while the CKLS model yields the highest predictive likelihood value. M_1, which accommodates possible nonlinearities by adopting a more flexible function in both the drift and diffusion processes, gives rise to a lower predictive likelihood value than the CKLS model (i.e. M_2). The two models that give higher cumulated predictive likelihoods than M_2 are M_{11} and M_{12}. These are the models associated with BHK.

Perhaps unsurprisingly, M_{11} and M_{12} are the only two of the 12 nested models that account for news arrival and volatility clustering in the model innovations. The remaining models, which assume a constant variance for model innovations, exhibit consistently lower cumulated predictive likelihoods than M_{11} or M_{12}. Consequently, the relative underperformance of models M_1 to M_{10} is likely to be associated with their inability to account for the periodic presence of persistently high volatility in the short rate data. The news arrival process, therefore, appears to be pivotal in generating greater predictive likelihoods at the weekly frequency. This result corroborates the findings of Brenner et al. (1996) who examine this issue within a frequentist

[8] The cumulative predictive likelihoods based on diffuse priors are available from the authors upon request.

framework. The dramatic improvement in the predictive likelihood value of M_{11} over M_2 and M_6 also implies that the specification of a diffusion process which accounts for news arrival and volatility clustering bears greater importance on short rate model forecast performance than the stylised feature of levels dependent conditional variance.

These results indicate a marked preference for short rate models with GARCH components that approximate the news arrival process and the well known presence of volatility clustering in short rate data (a feature that is also present in equity returns, exchange rates and other financial data). The results also suggest that the modelling of the news arrival process is paramount in lower frequency data. It is uncertain, however, whether a similar set of conclusions would be reached given high frequency data; in other words, is the preference for models accounting for news arrival and volatility clustering a result of the selected data frequency.

5.3 Results for High Frequency Data

Table 6 presents the results of the exercise undertaken using high frequency data. As with the weekly data, only the cumulated log predictive likelihoods for the informative priors are shown (due to their similarity with the cumulated log predictive likelihoods when using diffuse priors). The results are markedly different to those obtained when using low frequency data. Model M_5, which was associated with the smallest predictive likelihood in the low frequency data estimation, is the best performing model at each forecast horizon when considering high frequency data. Model M_5, the Cox, Ingersoll and Ross model, is the most prominent and often cited paper in the theoretical short rate literature. In turn, the cumulated predictive likelihood for the best model using low frequency data, model M_{12}, is typically below that of the non-GARCH models and only slightly above the bottom four non-GARCH models (i.e. M_3, M_7, M_9 and M_{10}).

- Table 6 about here -

This striking change in results when adopting high frequency data indicates a clear contrast between the preferred short rate models across high and low frequencies data. The shift in performance towards the non-GARCH models indicates a decline in the importance of $g(W_{it}G_i)$ when modelling high frequency short rate data. Unlike the low frequency case, the appropriate characterisation of the news arrival process (and volatility clustering) appears less critical for high frequency data. Instead, the elasticity of volatility $\sqrt{f\left(Z_{it}, B_i\right)}$ appears to be the more important of the two diffusion factors in obtaining greater predictive likelihoods, with models M_8, M_4, M_2, M_6 and M_1(all of which incorporate level effects through $\sqrt{f\left(Z_{it}, B_i\right)}$ but have no GARCH element) being associated with greater cumulated predictive likelihoods relative to M_{12} (which incorporates both level effects through $\sqrt{f\left(Z_{it}, B_i\right)}$ and GARCH diffusion). Even in the case of the highly favoured M_{11} model which exhibits higher cumulated predictive likelihood value than some of the non-GARCH models, it is inferior in its forecast performance than the CIR model (i.e. M_5).

Taken together, these results provide little support for the prevailing consensus in the more recent short rate literature that earlier models, such as the CEV models, are inadequate in characterizing the short rate. Instead, the results suggest that this consensus may more appropriately be described as applying to the low frequency modelling of short rate data.

5.4 Results of BMA

There have been few applications of BMA to models of the short-rate, perhaps due to intensive computational requirements to obtain the posterior densities, predictive likelihoods, and posterior model probabilities for a set of short-rate models. Consequently, there is little evidence regarding the performance impact of the Bayesian weighting of short rate model predictions. Tables 5 and 6 present the results for BMA using the four approaches specified at the beginning of this section; BMA_1 and BMA_2 use posterior model probabilities based on the marginal likelihoods derived over the estimation period, whereas BMA_3 and BMA_4 are functions of posterior model probabilities based only on the predictive likelihoods computed over the forecast period.

The evidence in favour of the predictive benefits of model averaging using BMA_1 and BMA_2 is fairly weak, with cumulated predictive likelihoods that are clearly lower than the better performing models (especially for the high frequency data). The catalyst for this appears to be that the Bayesian model averages based on marginal likelihoods over the entire estimation period effectively collapse to model selection procedures; both BMA_1 and BMA_2 have probabilities close to unity for M_{12} in the low frequency case and M_1 in the high frequency case (and, therefore, close to zero for the remaining models). The posterior model probabilities therefore collapse to model selection based on estimation-period (or in-sample) fit. This problem is typically attributed to large sample size spanning the estimation period and is discussed further in Amisano and Geweke (2010) who experience a similar issue for BMA based on equity returns models. The cumulated predictive likelihoods of BMA_1 and BMA_2 are, however, greater than those of the simple model averaging approach (albeit only slightly in the high frequency case) suggesting that the construction of predictions assuming equal model weights is the least effective averaging procedure for short-rate data, especially at lower frequencies.

In contrast, the posterior model probabilities underlying BMA_3 and BMA_4, which are independent of the estimation-period marginal likelihood, do not collapse to unity for any single model. Figures 2 and 3 show plots of the posterior model probabilities for weekly and 15-minute data, respectively. For BMA_3 the probability of M_{11} dominates all models in the case of weekly data. However, for high frequency data we observe that the probability of M_5 exceeds that of M_{11} around about the start, the middle and thereafter of the forecasting period. Using a rolling window of the last 20 periods, we find that the results for BMA_4 are qualitatively unchanged. One noticeable difference in the posterior model probability plots of BMA_3 and BMA_4 is that other models like M_6 and M_8 begin to assume some importance in the BMA forecast, particularly with weekly data as their probabilities increase substantially in the last one-third of the forecasting period.

In Tables 5 and 6, the cumulated predictive likelihoods of BMA_3 and BMA_4 exceed those of BMA_1 and BMA_2 for both datasets. This is also true with the use of informative and uninformative priors. The performance of BMA_3 and BMA_4 exceeds that of all but the single best models across the two frequencies and, even then, are associated with cumulated predictive likelihoods only slightly below the best model across either frequency. Although the cumulated predictive likelihoods for BMA_3 are greater than those of BMA_4, the difference is not significant suggesting that the primary contribution to forecast improvement for short rate data does not stem from using a rolling window to determine model weights (at least for forecast periods of the length we have chosen).

This result indicates the predictive usefulness of restricting the data available to posterior model probabilities to the period covering the forecast horizon, thereby limiting the weight attached to the performance of the models over the estimation period. Given such a restriction, the evidence suggests that Bayesian model averaging produces predictive likelihoods that correspond closely to the better performing models while hedging prediction risk by attaching non-zero weight to the less likely outcomes generated by the remaining models. Indeed, BMA_3 and BMA_4 produce forecasts that perform better than all but the best performing models at every forecast horizon, implying that the performance gain is insensitive to forecast horizon.

6 Conclusion

This paper has investigated the usefulness of BMA for predicting US short-term interest rates observed at both weekly and 15-minute frequencies. We find that pooling forecasts from different short rate models using BMA yields clear forecast improvements at either frequency. In particular, the BMA forecasts based solely on recent predictive likelihoods rank above almost all individual short rate models. There is further evidence that BMA based on recent predictive likelihoods gives rise to better short rate forecasts than BMA that also uses in-sample data to determine the posterior probability associated with each model. These results are robust to the choice of prior for the parameters and the forecast horizon considered.

An interesting finding not previously documented in the literature is that, for the high frequency short rates, the results provide little support for the prevailing consensus in the short rate literature that earlier models not accounting for volatility clustering and news arrival, are inadequate in characterising short rate dynamics. On the contrary, simpler models such as the square root constant elasticity variance model of Cox et al. (1985) exhibit the largest predictive likelihoods for high frequency short rate data. In contrast, our results suggest that accounting for volatility clustering and news arrival is of primary importance in determining the predictive likelihood of models applied to lower frequency weekly short rates. Accordingly, the validity of the prevailing consensus appears to be limited to the case of lower frequency data.

7 Appendix A

Conditional Posterior pdf for A_i

It can be shown that the conditional posterior distribution of A_i, $p(A_i|\theta_i \neq A_i, r, M_i)$ is a normal distribution

$$A_i|\theta_i \neq A_i, r, M_i \sim N(\overline{A_i}, \overline{\Sigma}_{A_i})$$

with posterior mean and posterior covariance defined as

$$\overline{A_i} = \overline{\Sigma}_{A_i}(X_i' D_i^{-1} X \widehat{A} + \underline{\Sigma}_{A_i}^{-1} \underline{A}_i)$$

$$\overline{\Sigma}_{A_i} = (X_i' D_i^{-1} X_i + \underline{\Sigma}_{A_i}^{-1})^{-1}$$

, respectively where

$$D_i = \begin{bmatrix} \frac{f(Z_{i1}, B_i) g(W_{i1} G_i) \triangle t}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{f(Z_{iT}, B_i) g(W_{iT} G_i) \triangle t}{\lambda_T} \end{bmatrix}$$

and

$$\widehat{A} = \left(X_i' D_i^{-1} X_i\right)^{-1} X_i' D_i^{-1} r.$$

Conditional Posterior pdf for B_i

In general, the conditional posterior pdf for B_i is

$$p(B_i|\theta_i \neq B_i, r, M_i) \propto p(B_i) \prod_{t=1}^{T} \left(\frac{\lambda_t}{f(Z_{it}, B_i) W_{it} G_i \triangle t} \right)^{\frac{1}{2}} \exp\left(-\sum_{t=1}^{T} \frac{\lambda_t \left(\Delta r_t - X_i A_i \triangle t\right)^2}{2 f(Z_{it}, B_i) W_{it} G_i \triangle t} \right).$$

The distribution of $p(B_i|\theta_i \neq B_i, r, M_i)$ depends on the restriction placed on $f\left(Z_{it}, B_i\right)$. For $M_1, M_2,$ M_6 and M_{11}, we employed a random-walk Metropolis-Hastings algorithm based on a normal proposal distribution with variance k.[9] For other models the conditional posterior densities follow an inverted gamma distribution such that

$$\beta_i|\theta_i \neq \beta_i, r, M_i \sim Inverted\ Gamma(\overline{v}_{\beta_i}, \overline{s}_{\beta_i}) \qquad i = 0, 1, 2$$

where $\overline{v}_{\beta_i} = \frac{T}{2} + v_{\beta_i}$, $\overline{s}_{\beta_0} = s_{\beta_0} + \sum_{t=1}^{T} \frac{\lambda_t(\Delta r_t - X_i A_i \triangle t)^2}{2 W_{it} G_i \triangle t}$, $\overline{s}_{\beta_1} = s_{\beta_1} + \sum_{t=1}^{T} \frac{\lambda_t(\Delta r_t - X_i A_i \triangle t)^2}{2 r_{t-1} W_{it} G_i \triangle t}$ and $\overline{s}_{\beta_2} = s_{\beta_2} + \sum_{t=1}^{T} \frac{\lambda_t(\Delta r_t - X_i A_i \triangle t)^2}{2 r_{t-1}^{\beta_3} W_{it} G_i \triangle t}$. Note that β_3 in $\overline{s}_{\beta_2}$ is known. Similarly for models with β_2 and β_3 only, one could design a two-step sampling scheme. The first step is to draw β_2 conditional on β_3 by sampling from the inverse gamma distribution. The second step, which is to draw β_3 conditional on β_2, employs a normal random-walk Metropolis-Hastings with variance k.

Conditional Posterior pdf for λ_t

It can be shown that the posterior of λ_t, $t = 1, ..., T$, for any given model is a gamma distribution

$$\lambda_t|\theta_i \neq \lambda, r_t, M_i \sim Gamma(\frac{v+1}{2}, \frac{\left(\Delta r_t - X_{it} A_i \triangle t\right)^2}{2 f(Z_{it}, B_i) W_{it} G_i \triangle t} + \frac{v}{2}).$$

Conditional Posterior pdf for G_i

[9] The scaling parameter k is equal to 0.5 except for periods when slow mixing occurred. In such cases, we temporarily adjusted the scaling parameter.

$$p(G_i|\theta_i \neq \gamma, r, M_i) \propto I_{G_i} \prod_{t=1}^{T} \left(\frac{\lambda_t}{f(Z_{it}, B_i) W_{it} G_i \triangle t} \right)^{\frac{1}{2}} \exp \left(- \frac{\lambda_t \left(\Delta r_t - X_i A_i \triangle t \right)^2}{2 f(Z_{it}, B_i) W_{it} G_i \triangle t} \right), \quad (25)$$

where I_{G_i} is an indicator function such that $I_{G_i} = 1$ if $W_{it} G_i > 0$ for $t = 1, ..., G$. The restrictions $0 < \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 < 1$, $0 < \gamma_0 < 100$ and $0 < \gamma_1, \gamma_2, \gamma_3, \gamma_4 < 1$ are imposed.

The conditional posterior pdf for G_i is intractable so the independent Metropolis-Hastings algorithm is used to obtain draws of G_i. In this respect, we use a normal proposal distribution with the mean and variance parameters determined as the mode and the negative inverse Hessian of (25).

Conditional Posterior pdf for v

The conditional posterior density for the degree of freedom parameter is

$$p(v|\theta_i \neq v, r, M_i) \propto \Gamma\left(\frac{v}{2}\right)^{-T} \left(\frac{v}{2}\right)^{\frac{Tv}{2}} \exp\left(\frac{v}{2} \sum_{t=1}^{T} \left(\log(\lambda_t) - \lambda_t\right) - \tau\right).$$

To draw from the conditional density we use a random-walk Metropolis-Hastings algorithm based on the normal proposal distribution with variance k.[10] A rejection condition was imposed such that only values of v greater than two were accepted. A Griddy Gibbs sampler (Ritter and Tanner, 1992) may also be used to draw $p(v|\theta_i \neq v, r, M_i)$. From an implementation point of view, however, we found that the random-walk Metropolis Hastings algorithm performed better than Griddy Gibbs for the data used in this paper.[11]

8 Appendix B

In our paper, the marginal likelihoods cannot be computed analytically. To this end, we have adopted the Modified Harmonic Mean (MHM) method of Gelfand and Dey (1994) to obtain the marginal likelihoods. The advantage of this approach is that it uses the posterior para-

[10] Where mixing was slow, the variance parameter k was temporarily adjusted.

[11] The large sample size caused processor overflows when computing the densities required by the Griddy Gibbs procedure.

meter draws to obtain the marginal likelihood and can be employed alongside most sampling techniques. The inverse of the predicted marginal likelihood is computed as:

$$\widehat{p}(\mathcal{F}_T)^{-1} = \frac{1}{J}\sum_{j=1}^{J}\frac{f(\theta^{(j)},\rho^{(j)})}{p(\theta^{(j)},\rho^{(j)})\widehat{p}(z|\theta^{(j)},\rho^{(j)})}, \tag{26}$$

where $f(\theta,\rho)$ is a density function with supported constraint within the posterior support of (θ,ρ). $f(\theta,\rho)$ ideally approximates the posterior pdf. Geweke (1999) suggests a truncated multivariate normal distribution with different sets of truncation values such that $\delta \in (0,1)$ for $f(\theta,\rho)$:

$$\theta \sim N\left(\begin{array}{c}\widehat{\theta}\\ \widehat{\rho}\end{array}, \widehat{\Sigma}_{\theta,\rho}\right)_{I(\Gamma)} \tag{27}$$

where $\widehat{\theta} = J^{-1}\sum_{j=1}^{J}\theta^{(j)}$, $\widehat{\rho} = J^{-1}\sum_{j=1}^{J}\rho^{(j)}$,and $\widehat{\Sigma}_{\theta} = J^{-1}\sum_{j=1}^{J}\begin{bmatrix}\theta^{(j)}-\widehat{\theta}\\ \rho^{(j)}-\widehat{\rho}\end{bmatrix}\begin{bmatrix}\theta^{(j)}-\widehat{\theta}\\ \rho^{(j)}-\widehat{\rho}\end{bmatrix}'$. $I(\Gamma)$ is an indictor function such that $I(\Gamma)=1$ if $\left(\begin{bmatrix}\theta^{(j)}-\widehat{\theta}\\ \rho^{(j)}-\widehat{\rho}\end{bmatrix}'\widehat{\Sigma}_{\theta}^{-1}\begin{bmatrix}\theta^{(j)}-\widehat{\theta}\\ \rho^{(j)}-\widehat{\rho}\end{bmatrix}\right) \leq q$ where q is such that $P(\chi_a^2 < q) = \delta$ and a is the dimension of (θ,ρ). Note that in computing $f(\theta,\rho)$ an additional normalising constant δ is added to ensure $f(\theta,\rho)$ integrates to unity.

References

[1] Aït Sahalia, Y. 1996. Testing continuous-time models of the spot interest rates. *Review of Financial Studies*, 9, 385-426.

[2] Andersen, T.G. and J. Lund. 1997. Estimating continuous-time stochastic models of the short-term interest rate. *Journal of Econometrics*, 77, 343-377.

[3] Avramov, D. 2002. Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64, 423-458.

[4] Bali, T.G. 2003. Modeling the stochastic behaviour of short-term interest rates: Pricing implications for discount bonds. *Journal of Banking and Finance*, 27, 201-228.

[5] Bali, T. and L. Wu. 2006. A comprehensive analysis of the short-term interest rate dynamics. *Journal of Banking and Finance*, 30, 1269-1290.

[6] Bliss, R.R. and D.C. Smith. 1998. The elasticity of interest rate volatility: Chan, Karolyi, Longstaff and Sanders revisited. *Working paper 97-12a*, Federal Reserve Bank of Atlanta.

[7] Bollerslev, T. 1986. Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.

[8] Bollerslev, T., R. F. Engle and K. F. Kroner. 1992. ARCH modeling in finance. *Journal of Econometrics,* 52, 5-59.

[9] Brennan, M.J. and E.S. Schwartz. 1980. Analyzing convertible bonds. *Journal of Financial and Quantitative Analysis*, 15, 907-929.

[10] Brenner, R.J., R.H. Harjes and K.F. Kroner. 1996. Another look at models of the short term interest rate. *Journal of Financial and Quantitative Analysis*, 31, 85-107.

[11] Chan, K.C., G.A. Karolyi, F.A. Longstaff and A.B. Sanders. 1992. An empirical comparison of alternative models of the short-term interest rates. *Journal of Finance*, 389-403.

[12] Christofferson, C. 2008. Level-ARCH short rate models with regime switching: Bivariate modelling of U.S. and European short rates. *International Review of Financial Analysis*, 7, 925-948.

[13] Conley, T.G., L.P. Hansen , E.G.Z. Luttmer and J.A. Scheinkman. 1997. Short-term interest rates as subordinated diffusions. *Review of Financial Studies*, 10, 525-577.

[14] Cox, J. 1975. Notes on option pricing 1: constant elasticity of variance diffusions. Working Paper, Stanford University.

[15] Cox, J., J. Ingersoll and S. Ross. 1980. An analysis of variable rate loan contracts. *Journal of Finance*, 35, 389-403.

[16] Cox, J., J. Ingersoll and S. Ross. 1985. A theory of the term structure of interest rates. *Econometrica*, 53. 385-407.

[17] Cremers, K.J.M. 2002. Stock return predictability: A Bayesian model selection perspective. *Review of Financial Studies*, 15, 1223-1249.

[18] Doppelhofer, G., R.I. Miller and X. Sala-i-Martin. 2000. Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach, *NBER Working Paper* 7750.

[19] Dothan, L.U. 1978. On the term structure of interest rates. *Journal of Financial Economics*, 6, 59-69.

[20] Engle, R.F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50, 987-1008.

[21] Fernandez, C., E. Ley andM.F.J. Steel. 2001.Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16, 563-576.

[22] Geweke, J. 1993. Bayesian treatment of the independent Student's t linear model. Supplement: Special Issue on Econometric Inference Using Simulation Techniques. *Journal of Applied Econometrics*, 8: S19-S40.

[23] Geweke, J.,1994. Bayesian comparison of econometric models. *Federal Reserve Bank of Minneapolis Working Paper.*

[24] Geweke, J and G. Amisano, 2010. Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26, 216-230.

[25] Geweke, J. and C. Whiteman, 2006. Bayesian Forecasting, *Handbook of Economic Forecasting*, Elsevier.

[26] Gospodinov, N. 2005. Testing for threshold nonlinearity in short-term interest rates, *Journal of Financial Econometrics*, 3, 344-371.

[27] Gray, S.F. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42, 27.62.

[28] Jones, C.S. (2003), 'Nonlinear mean reversion in the short-term interest rate', *Review of Financial Studies*, 16, 793-843.

[29] Koedijk, K.C., F. Nissen, P.C. Schotman and C.C. Wolff, 1997. The dynamics of short-term interest rate volatility revisited. *European Finance Review*, 1, 105-130.

[30] Koop, G. and S. Potter. 2003. Forecasting in large macroeconomic panels using Bayesian model averaging. *Federal Reserve Bank of New York Staff Report* 163.

[31] Liu, C. and J.M. Maheu. (2009). Forecasting realized volatility: A Bayesian model averaging approach. *Journal of Applied Econometrics*, 24, 709-733.

[32] Merton, R.C. 1973. An intertemporal capital asset pricing model. *Econometrica*, 41, 867-887.

[33] Pesaran M.H. and P. Zaffaroni. 2005.Model averaging and value-at-risk based evaluation of large multi-asset volatility models for risk management. *CEPR Discussion Papers* 5279.

[34] Polson, N. 1996. Convergence of Markov Chain Monte Carlo Algorithms. *In Bayesian Statistics* 5, J.M. Bernardo et al (eds), Oxford University Press, Oxford, 297-323.

[35] Ritter, C. and M. A. Tanner, 1992, The Gibbs stopper and the griddy Gibbs sampler. *Journal of the American Statistical Association,* 87, 861-868.

[36] Sanford, A.D. and G.M. Martin. 2006. Bayesian comparison of several continuous time models of the Australian short rate. *Accounting and Finance*, 46, 309-326.

[37] Smith, D.R. 2002. Markov-switching and stochastic volatility diffusion models of short-term interest rates. *Journal of Business and Economics Statistics*, 20, 183-197.

[38] Stanton, R. 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance*, 52, 1973-2002.

[39] Suardi, S. 2008. Are levels effects important in out-of-sample performance of short rate models? *Economics Letters*, 99, 181-184.

[40] Vasicek, O. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5, 177-188.

[41] Wright, J.H. 2008. Bayesian averaging and exchange rate forecasts. *Journal of Econometrics*, 146, 329-341.

Table 1. Parameter Restrictions for Various Interest Rate Models

Model	Specification	μ	λ	γ
Merton (1973)	$dr_t = \mu dt + \sigma dW_t$		0	0
Vasicek (1977)	$dr_t = (\mu + \lambda r_t)dt + \sigma dW_t$			0
Cox, Ingersoll and Ross (1985)	$dr_t = (\mu + \lambda r_t)dt + \sigma\sqrt{r_t}dW_t$			1/2
Dothan (1978)	$dr_t = \sigma r_t dW_t$	0	0	1
Geometric Brownian Motion (GBM)	$dr_t = \lambda r_t dt + \sigma r_t dW_t$	0		1
Brennan and Schwartz (1980)	$dr_t = (\mu + \lambda r_t)dt + \sigma r_t dW_t$			1
Cox, Ingersoll and Ross (1980)	$dr_t = \sigma r_t^{3/2} dW_t$	0	0	3/2
Constant Elasticity of Variance, Cox (1975)	$dr_t = \lambda r_t dt + \sigma r_t^{\gamma} dW_t$	0		

Table 2. Summary of Short-Rate Models Nested in the Generalised Model

Model	$X_{it}A_i$	$f(Z_{it}, B_i)$	$g(W_{it}, G_i)$
M1: Ait-Sahalia (1996)	$\alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \frac{\alpha_3}{r_{t-1}}$	$\beta_0 + \beta_1 r_{t-1} + \beta_2 r_{t-1}^{\beta_3}$	1 $(\gamma_i = 0\ ,\ \forall i = 0,1,2,3,4,5)$
M2: CKLS (1992)	$\alpha_0 + \alpha_1 r_{t-1}$ $(\alpha_2{=}\alpha_3{=}0)$	$\beta_2 r_{t-1}^{\beta_3}$ $(\beta_0 = \beta_1 = 0)$	1
M3: Vasicek (1977)	$\alpha_0 + \alpha_1 r_{t-1}$ $(\alpha_2{=}\alpha_3{=}0)$	β_0 $(\beta_0 = \beta_2 = 0)$	1
M4: BS (1980)	$\alpha_0 + \alpha_1 r_{t-1}$ $(\alpha_2{=}\alpha_3{=}0)$	$\beta_2 r_{t-1}^2$ $(\beta_0 = \beta_1 = 0, \beta_3 = 2)$	1
M5: CIR (1985)	$\alpha_0 + \alpha_1 r_{t-1}$ $(\alpha_2{=}\alpha_3{=}0)$	$\beta_1 r_{t-1}$ $(\beta_0 = \beta_1 = 0, \beta_3 = 1)$	1
M6: CEV Cox (1975)	α_0 $(\alpha_1{=}\alpha_2{=}\alpha_3{=}0)$	$\beta_2 r_{t-1}^{\beta_3}$ $(\beta_0 = \beta_1 = 0)$	1
M7: Merton (1973)	α_0 $(\alpha_1{=}\alpha_2{=}\alpha_3{=}0)$	β_0 $(\beta_1 = \beta_2 = 0)$	1
M8: GBM	$\alpha_1 r_{t-1}$ $(\alpha_0{=}\alpha_2{=}\alpha_3{=}0)$	$\beta_2 r_{t-1}^2$ $(\beta_0 = \beta_1 = 0, \beta_3 = 2)$	1
M9: Dothan (1978)	0 $(\alpha_i{=}0\ \forall i = 0,1,2,3)$	$\beta_2 r_{t-1}^2$ $(\beta_0 = \beta_1 = 0, \beta_3 = 2)$	1
M10: CIR VR (1980)	0 $(\alpha_i{=}0\ \forall i = 0,1,2,3)$	$\beta_2 r_{t-1}^3$ $(\beta_0 = \beta_1 = 0)$	1
M11: BHK (1996)	$\alpha_0 + \alpha_1 r_{t-1}$ $(\alpha_2{=}\alpha_3{=}0)$	$r_{t-1}^{\beta_3}$ $(\beta_0 = \beta_1 = 0, \beta_2 = 1)$	$h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 h_{t-1} + \gamma_3 \xi_{t-1}^2$ $(\gamma_4 = 0)$
M12: BHK (1996)	$\alpha_0 + \alpha_1 r_{t-1}$ $(\alpha_2{=}\alpha_3{=}0)$	1 $(\beta_0 = 1, \beta_1 = \beta_2 = 0)$	$h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 h_{t-1}$ $+\gamma_3 \xi_{t-1}^2 + \gamma_4 r_{t-1}$

Note: The equalities in parenthesis are parameter restrictions of the specific functions.

Table 3. Summary Statistics of Δr_t

(a) **Weekly 1-Month Eurodollar Deposit Rates**			
Mean	-0.0048	Q(5)	46.548 (0.000)
Std. dev	0.3877	Q(10)	54.843 (0.000)
Skewness	0.8807	LM(5)	228.725 (0.000)
Kurtosis	34.2005	LM(10)	233.227 (0.000)
JB	86638.7 (0.000)	ADF (r_t)	-3.590 (0.031)
		ADF (Δr_t)	-20.477 (0.000)
(b) **15-Minute 1-Month Eurodollar Future Rates**			
Mean	-0.0002	Q(5)	89.964 (0.000)
Std. dev	0.0180	Q(10)	120.703 (0.000)
Skewness	5.6584	LM(5)	20.937 (0.001)
Kurtosis	134.1275	LM(10)	31.382 (0.000)
JB	981406.5 (0.000)	ADF (r_t)	-3.118 (0.103)
		ADF (Δr_t)	-12.999 (0.000)

Note: Q(5) and Q(10) are the Ljung-Box test statistics for serial correlation in short rate changes of order 5 and 10, respectively. LM(5) and LM(10) denotes the ARCH test of the residuals from an AR(10) regression of the interest rate data for lag order 5 and 10, respectively. JB is the Jarque-Bera test of normality of short rate distribution. ADF denotes the Augmented Dickey Fuller test statistic. Figures in parenthesis are p-value.

Table 4. Sampling Scheme for Interest Rate Models

Model		Step 1	Step 2	Step 3	Step 4	Step 5
M_1	Ait-Sahalia (1996)	✓	✓	✓		✓
M_2	CKLS (1992)	✓	✓	✓		✓
M_3	Vasicek (1977)	✓	✓	✓		✓
M_4	BS (1980)	✓	✓	✓		✓
M_5	CIR (1985)	✓	✓	✓		✓
M_6	CEV (1975)	✓	✓	✓		✓
M_7	Merton (1973)	✓	✓	✓		✓
M_8	GBM	✓	✓	✓		✓
M_9	Dothan (1978)		✓	✓		✓
M_{10}	CIR VR (1980)		✓	✓		✓
M_{11}	BHK1 (1996)	✓	✓	✓	✓	✓
M_{12}	BHK2 (1996)	✓		✓	✓	✓

Table 5. Cumulated log predictive likelihoods - weekly data

	Forecast Steps							
	1	2	3	4	5	6	7	8
M_1	-18.879	-19.495	-20.498	-20.408	-21.906	-25.672	-27.385	-29.341
M_2	-8.6556	-9.4788	-10.823	-10.545	-12.41	-16.709	-18.698	-21.192
M_3	-29.503	-29.718	-30.129	-30.183	-31.179	-33.695	-34.825	-35.743
M_4	-11.403	-12.028	-13.446	-13.229	-15.687	-21.243	-23.299	-26.19
M_5	-93.351	-93.385	-93.675	-93.517	-94.474	-97.873	-98.446	-100.05
M_6	-8.6281	-9.6076	-11.123	-10.971	-12.969	-17.197	-19.143	-21.728
M_7	-29.828	-30.108	-30.707	-30.833	-31.842	-34.123	-35.066	-35.923
M_8	-11.562	-12.25	-13.856	-13.683	-16.164	-21.741	-23.819	-26.819
M_9	-59.917	-60.56	-61.997	-61.841	-64.407	-69.911	-71.932	-74.85
M_{10}	-83.49	-84.015	-85.511	-85.202	-88.005	-94.128	-96.179	-99.398
M_{11}	2.957	1.9864	0.93958	0.99976	-1.2991	-6.361	-8.7797	-11.639
M_{12}	-2.165	-2.6759	-3.345	-3.2197	-5.0228	-9.0488	-10.992	-13.076
BMA_1	-2.169	-2.6799	-3.349	-3.2236	-5.0268	-9.0528	-10.996	-13.08
BMA_2	-2.1892	-2.7	-3.369	-3.2436	-5.0469	-9.0728	-11.016	-13.1
BMA_3	2.4243	1.4603	0.43705	0.51559	-1.7663	-6.8017	-9.2036	-12.047
BMA_4	0.69986	-0.15976	-1.5765	-1.4172	-3.6501	-8.8944	-11.171	-14.258
Simple MA	-24.811	-25.365	-26.397	-26.274	-28.09	-32.187	-33.935	-36.034

Note: Models M_1 to M_{12} are defined in Table 2. BMA 1 to 4 denote the four methods of pooling short rate forecasts, while Simple MA is a BMA method which assumes equal model weight.

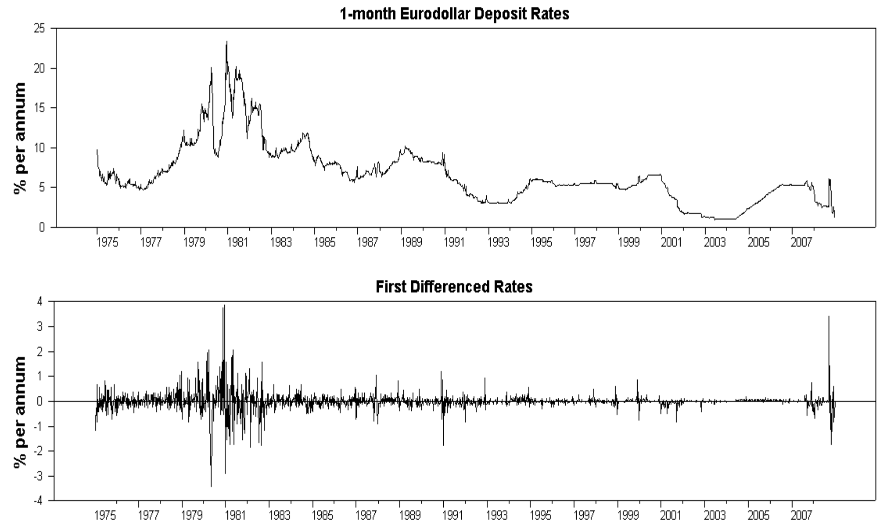
Table 6. Cumulated log predictive likelihoods - high frequency data

	Forecast Steps							
	1	2	3	4	5	6	7	8
M_1	198.76	198.85	198.83	198.94	198.92	198.95	200.78	200.72
M_2	204.66	204.68	204.7	204.75	204.8	204.83	207.13	207.13
M_3	185.92	185.91	185.92	185.92	185.9	185.91	187.27	187.26
M_4	205.79	205.82	205.86	205.91	205.96	205.99	208.3	208.31
M_5	210.59	210.6	210.62	210.66	210.69	210.73	213.14	213.14
M_6	204.47	204.46	204.43	204.44	204.44	204.45	206.59	206.59
M_7	186.04	186.04	186.04	186.04	186.03	186.03	187.36	187.36
M_8	206.13	206.11	206.08	206.09	206.07	206.07	208.24	208.25
M_9	180.43	180.4	180.35	180.34	180.3	180.29	182.44	182.44
M_{10}	186.76	186.71	186.61	186.59	186.55	186.52	189.05	189.04
M_{11}	207.75	208.24	208.48	208.61	208.75	208.89	208.97	209.06
M_{12}	188.76	188.93	189.01	189.04	189.05	189.08	190.59	190.59
BMA_1	198.76	198.85	198.83	198.94	198.92	198.95	200.78	200.72
BMA_2	198.76	198.85	198.83	198.94	198.92	198.95	200.78	200.72
BMA_3	209.08	209.25	209.34	209.45	209.55	209.89	210.94	211.01
BMA_4	208.8	209	209.04	209.04	208.97	209.23	210.04	209.38
Simple MA	198.42	198.49	198.51	198.55	198.57	198.59	200.12	200.12

Note: See note to Table 5.

Figure 1 Plots of Weekly and 15-minute Short Rates

(a) Weekly short rates (1 February, 1975 to 31 December, 2008)



(b) 15-minute short rates (19 May to 29 September, 2009)

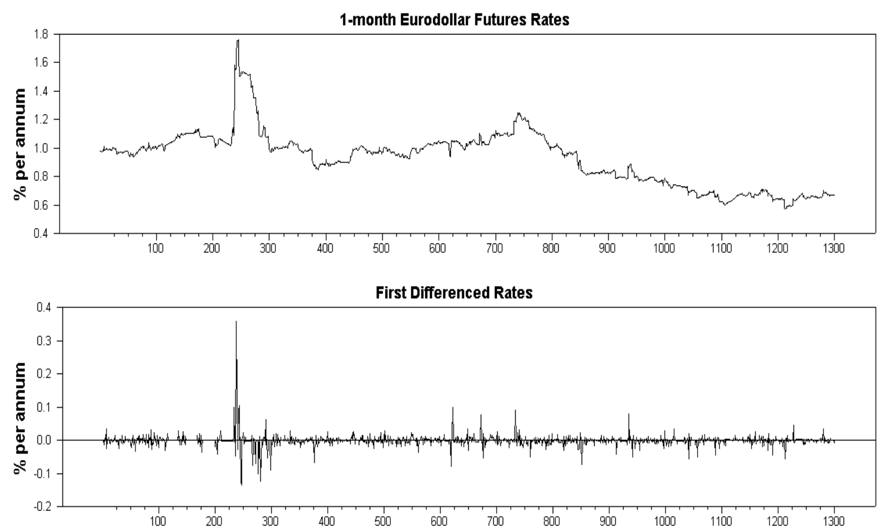
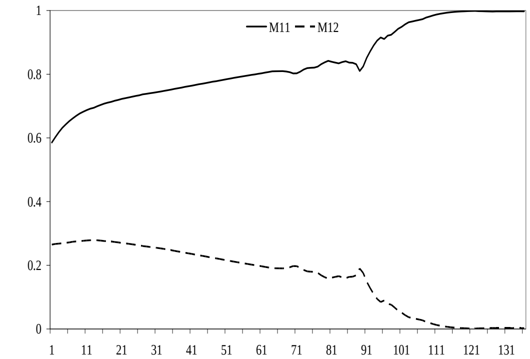
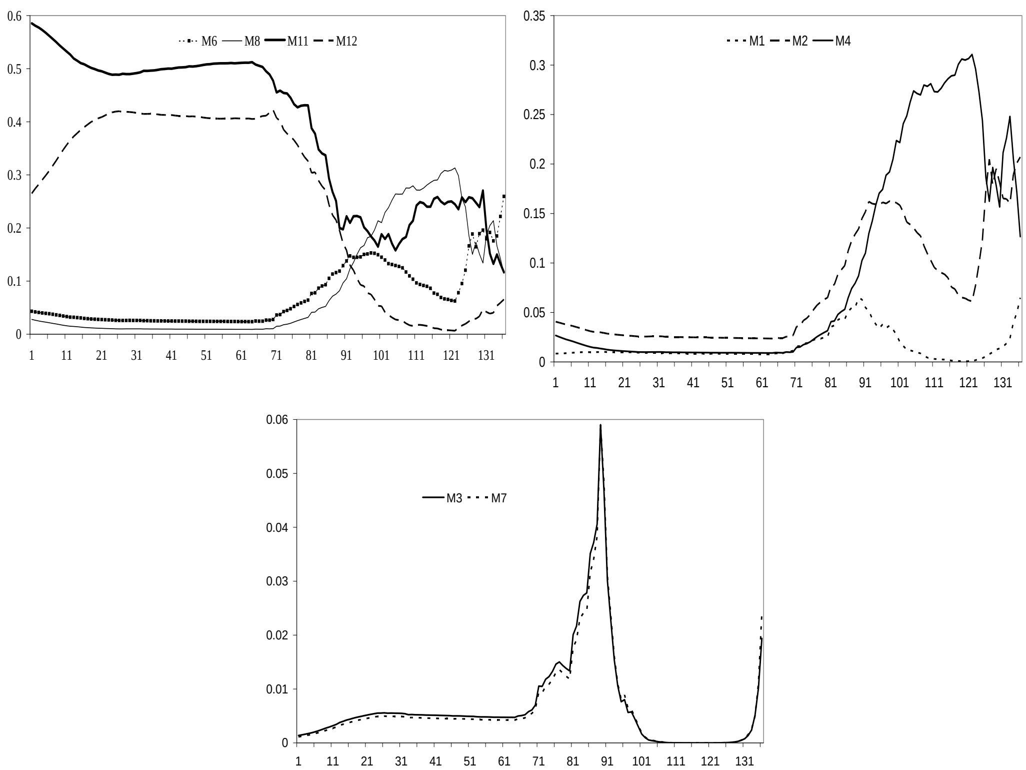


Figure 2. Plots of posterior model probabilities of BMA_3 and BMA_4 for weekly data

BMA_3



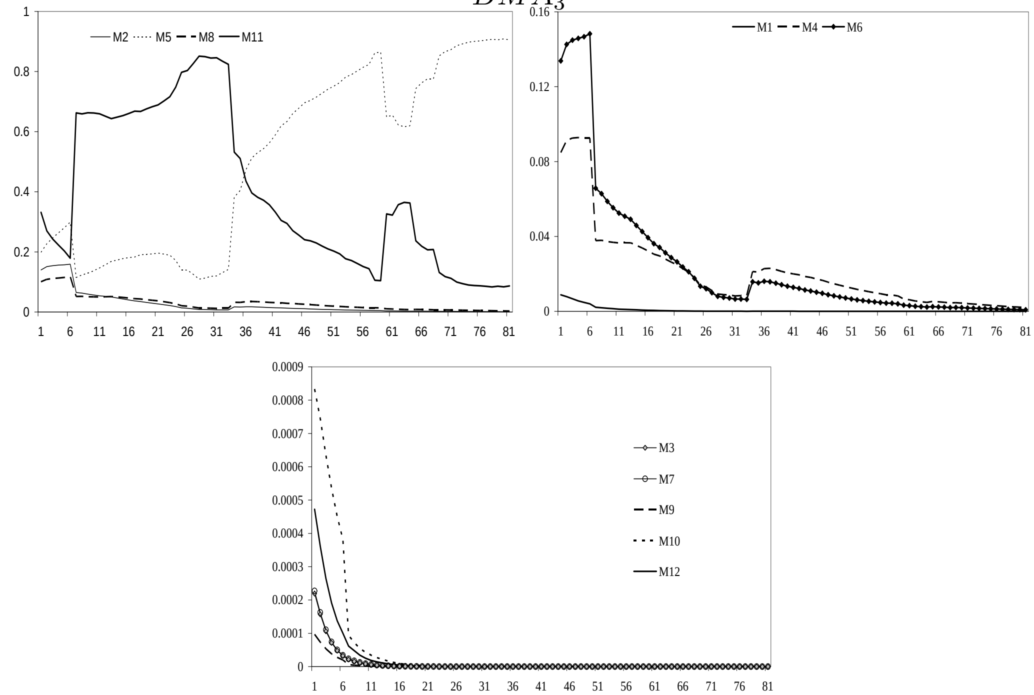
BMA_4



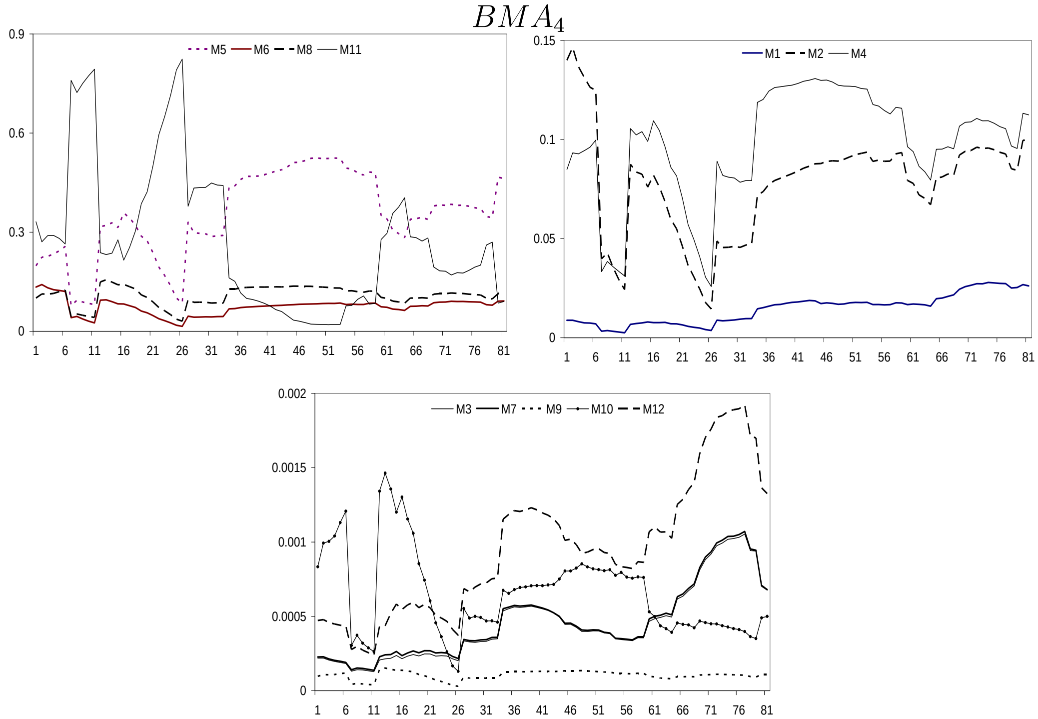
Note: We only plot the model probabilities that are non-zero.

Figure 3. Plots of posterior model probabilities of BMA_3 and BMA_4 for high frequency data

BMA_3



BMA_4



Note: See note to Figure 2.