# Department of Economics

## Forecasting the Yield Curve Using Priors from No Arbitrage Affine Term Structure Models

Andrea Carriero

Working Paper No. 612 October 2007 ISSN 1473-0278

# Forecasting the Yield Curve using Priors from No Arbitrage Affine Term Structure Models

Andrea Carriero
Queen Mary, University of London

September 2007

**Abstract**

In this paper we propose a strategy for forecasting the term structure of interest rates which may produce significant gains in predictive accuracy. The key idea is to use the restrictions implied by Affine Term Structure Models (ATSM) on a vector autoregression (VAR) as prior information rather than imposing them dogmatically. This allows to account for possible model misspecification. We apply the method to a system of five US yields, and we find that the gains in predictive accuracy can be substantial. In particular, for horizons longer than 1-step ahead, our proposed method produces systematically better forecasts than those obtained by using a pure ATSM or an unrestricted VAR, and it also outperforms very competitive benchmarks as the Minnesota prior, the Diebold-Li (2006) model, and the random walk.

Keywords: Bayesian methods, Forecasting, Term Structure.
JEL: C11, C53, E43, E47.

## 1 Introduction

Finance theory suggests that the current term structure should contain information about future term structures. Is it actually possible to use finance theory to extract this information and improve the forecasts of the Treasury yields? Giving an answer to this question is very important, as out of sample yield curve forecasting is key for the conduct of monetary policy as well as for bond portfolio management.

In the last years the finance literature has produced major advances in term structure modeling, building on the assumption of absence of arbitrage opportunities in the markets, but surprisingly only a few papers have studied the issue of forecasting the yield curve. Among these, De Jong (2000) and Dai and Singleton (2000) have focused

on *in sample* forecasting, while only Duffee (2002) and Ang and Piazzesi (2003) have focused on *out of sample* forecasting. Duffee (2002) has shown that beating a random walk with a traditional no arbitrage affine term structure model is difficult. Ang and Piazzesi (2003) show that imposing no-arbitrage restrictions and an essentially affine specification of market prices of risk improves out-of-sample forecasts from a VAR(12), but the gain respect to a random walk forecast is small.

Better results have been obtained outside the no-arbitrage paradigm by Diebold and Li (2006), who used the Nelson and Siegel (1987) exponential components framework to forecast the yield curve and found that it produces one-year-ahead forecasts that are noticeably more accurate than standard benchmarks. However, as stressed out by Duffee (2002), even if forecasting the yields is important in its own rights, having a model which is both consistent with finance theory and produces accurate forecasts is of key relevance, because explaining the time variation in expected returns means explaining the failure of the expectations hypothesis of the term structure of interest rates.

The absence of arbitrage ensures that the yield movements are consistent, controlling for risk, with the shape of the term structure at each point in time. As discussed by Diebold et al. (2005) this consistency will likely hold in deep and well-organized bond markets, but if the underlying term structure model is misspecifed the no arbitrage restrictions may actually degrade empirical performance. Most of the empirical implementations of no-arbitrage term structure models focused on the special case in which bond yields are affine functions of some state vector. As discussed in Piazzesi (2003) this assumption provides tractability, at the cost of some restrictive assumptions on the risk-adjusted dynamics of the state vector.

We believe that the reason behind the poor forecasting performance of no arbitrage affine term structure models is not the failure of no-arbitrage assumptions, but rather the failure of the additional specification restrictions implied by the model. As stressed by Sims (2003), if in the true data generating process a given set of restrictions holds up to some noise, then imposing the restrictions exactly would be suboptimal.

In this paper we propose to use a no arbitrage affine term structure model (ATSM) taking into account its possible misspecification. The key idea is to use the restrictions implied by the ATSM on a vector autoregression (VAR) as prior information rather than imposing them dogmatically. To implement the ATSM prior, we follow Del Negro and Schorfheide (2004) and use a set of dummy artificial observations which are first generated under the assumption of the validity of the ATSM and then added to the data.

When applied to a system of five US yields our proposed method produces substantial gains in predictive accuracy. In particular, while at the 1-step ahead horizon

the evidence is mixed, for longer forecast horizons the proposed method systematically outperforms the unrestricted VAR and a pure ATSM, and also outperforms very competitive benchmarks as the model by Diebold and Li (2006), the random walk, and the Minnesota prior.

The paper ties together two streams of research. The first is the literature on no arbitrage term structure models started from the seminal work of Vasicek (1977) and Cox et al. (1985), and continued to its modern multi-factor extensions by Duffie and Kan (1996), Dai and Singleton (2000), De Jong (2000), Duffee (2002), Ang and Piazzesi (2003). The second stream belongs to the macroeconomic literature and focuses on deriving priors from economic models for VARs. It has been initiated by Ingram and Whiteman (1987) who used a RBC model to derive priors for a VAR, and has been recently revived in a series of papers by Del Negro and Schorfheide (2004, 2005, 2007) who provided a full-blown approach to estimate VARs imposing priors from DSGE models.

Finally, we have to mention the fact that, starting from the seminal work of Ang and Piazzesi (2003), a series of papers has considered the joint modeling of the term structure of interest rates and macroeconomic factors. Prominent examples are Ang et al. (2004), Diebold et al (2006), Rudebush and Wu (2004), Mönch (2005), Favero et al (2007). Considering also macroeconomic information is beyond the scope of the present paper, but is on our agenda for further research.

The paper is organized as follows: in Section 2 we briefly describe the ATSM, while in Section 3 we discuss the econometric methodology. Section 4 presents estimation results, Section 5 focuses on forecasting. Section 6 concludes.

## 2 A Baseline Affine Term Structure Model

In this subsection we shall describe the ATSM used in the paper. We use the model of Ang and Piazzesi (2003). Formally, that model is a discrete-time version of the affine class introduced by Duffie and Kan (1996), where bond prices are exponential affine functions of underlying state variables. In particular we use the so called "yield-only" version of the model, in which only unobservable latent factors (and not also observable macroeconomic factors) are used as states. In what follows we just provide a brief summary of the model, the interested reader may refer to Ang and Piazzesi (2003) for a complete discussion and additional details.

The assumption of no arbitrage (Harrison and Kreps, 1979) guarantees the existence of a risk neutral measure $Q$ such that the price of an asset $V_t$ that does not pay any dividends at time $t+1$ satisfies $V_t = E_t^Q(\exp(-r_t)V_{t+1})$, where the expectation is taken

under the measure $Q$ and $r_t$ is the short term rate. The assumption of no arbitrage is equivalent to the assumption of the existence of the Radon-Nikodym derivative $\xi_{t+1}$, which allows to convert the risk neutral measure to the data generating measure as $E_t^Q(Z_{t+1}) = E_t^Q(\xi_{t+1}Z_{t+1})/\xi_t$, where $Z_t$ is a random variable. Assume $\xi_{t+1}$ follows a log-normal process:

$$\xi_{t+1} = \xi_t \exp(-0.5\Lambda_t'\Lambda_t - \Lambda_t'\varepsilon_{t+1}), \tag{1}$$

where $\Lambda_t$ are called market prices of risk and are an affine function of a vector of $k$ factors $F_t$:

$$\Lambda_t = \Lambda_0 + \Lambda_1 F_t, \tag{2}$$

where $\Lambda_0$ is a $k$-dimensional vector and $\Lambda_1$ a $k \times k$ matrix. Also the short term rate is assumed to be an affine function of $F_t$:

$$r_t = \delta_0 + \delta_1' F_t, \tag{3}$$

where $\delta_0$ is a scalar and $\delta_1$ a $k$-dimensional vector. We assume the factors follow a zero-mean stationary vector process:

$$F_t = \Psi F_{t-1} + \Omega\varepsilon_t, \tag{4}$$

where $\varepsilon_t \sim iidN(0, \Sigma_\varepsilon)$ with $\Sigma_\varepsilon = I$ with no loss of generality. The nominal pricing kernel is defined as:

$$m_{t+1} = \exp(-r_t)\xi_{t+1}/\xi_t = \exp(-\delta_0 - \delta_1' F_t - 0.5\Lambda_t'\Lambda_t - \Lambda_t'\varepsilon_{t+1}), \tag{5}$$

where the second equality comes using (3) and (1). The nominal pricing kernel prices all assets in the economy, so by letting $p_t^n$ denote the time $t$ price of a $n$-period zero coupon we have:

$$p_t^{n+1} = E_t(m_{t+1}p_{t+1}^n). \tag{6}$$

By using the above equations is possible to show that bond prices are an affine function of the state variables:

$$p_t^n = \exp(\bar{A}_n + \bar{B}_n' F_t), \tag{7}$$

where $\bar{A}_n$ and $\bar{B}_n$ are a scalar and a $k$-dimensional vector obeying to:

$$\begin{array}{c} \bar{A}_{n+1} = \bar{A}_n + \bar{B}_n'(-\Omega\Lambda_0) + 0.5\bar{B}_n'\Omega\Omega'\bar{B}_n - \delta_0 \\ \bar{B}_{n+1}' = \bar{B}_n'(\Psi - \Omega\Lambda_1) - \delta_1' \end{array}, \tag{8}$$

with $\bar{A}_1 = -\delta_0$ and $\bar{B}_1 = -\delta_1$. See Ang and Piazzesi (2003) for a derivation. The continuously compounded yield on a $n$-period zero coupon bond is:

$$y_t^{(n)} = -\ln p_t^n / n = A_n + B_n' F_t, \tag{9}$$

with $A_n = -\bar{A}_1/n$ and $B_n = -\bar{B}_1/n$, so yields are also an affine function of the factors. Equations (4) and (9) can be interpreted as a state and a space equation, so the state space for a vector of yields of $q$ different maturities can be written as:

$$\begin{aligned} F_t &= \Psi F_{t-1} + \Omega \varepsilon_t, \\ Y_t &= A + B F_t + v_t, \end{aligned} \tag{10}$$

where $Y_t = (y_t^{(1)}, y_t^{(2)}, ..., y_t^{(q)})'$ is a $q-$dimensional vector process collecting all the yields at hand, $A = (A_1, A_2, ...A_q)'$ and $B = (B_1, B_2, ..., B_q)'$ are functions of the structural coefficients of the model according to equation (8), and where $v_t$ is a vector of iid Gaussian measurement errors with variance $\Sigma_v$. For future reference we collect the coefficients in $\theta = \{\Lambda_0, \Lambda_1, \delta_0, \delta_1, \Psi, \Omega, \Sigma_v\}$ and we record the moments of the state space system in (10):

$$\begin{aligned} E[Y_t, Y_t'] &= AA' + B\Sigma_F B' + B\Omega\Sigma_{\varepsilon v} + \Sigma_{\varepsilon v}\Omega B' + \Sigma_v, \\ E[Y_t, Y_{t-h}'] &= AA' + B\Psi^h(\Sigma_F B' + \Omega\Sigma_{\varepsilon v}), \end{aligned} \tag{11}$$

where $\Sigma_F = E[F_t F_t']$ obeys to $\Sigma_F = \Psi\Sigma_F\Psi' + \Omega\Sigma_\varepsilon\Omega'$.

# 3 Econometric Methodology

In this section we briefly describe the econometric methodology used in the paper. We use the approach proposed by Del Negro and Schorfheide (2004), from which this section draws heavily and to which the interested reader may want to refer for a complete discussion. Basically we adapt to our case their framework which was originally designed for deriving priors from DSGE models.

Consider a VAR(p) representation of the $q-$dimensional vector process collecting all the yields at hand:

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + ... + \Phi_p Y_{t-p} + u_t \tag{12}$$

where $Y_t = (y_t^{(1)}, y_t^{(2)}, ..., y_t^{(q)})$ and $u_t$ is a vector of one-step-ahead forecast errors having a multivariate normal distribution with variance $\Sigma_u$. The VAR in (12) can be interpreted as an approximation of the Moving Average (MA) representation of $Y_t$. The approximation gets better as more dynamics is added to the system. Importantly, as is clear

from equation (10), the ATSM features a MA representation. As the ATSM depends on a vector of coefficients $\theta$ having much less elements than the elements of the coefficient matrices of the VAR, the validity of the ATSM imposes a set of restrictions on the VAR in (12).

The key idea is to impose such restrictions in a non-dogmatic way, and in particular to let them hold with some degree of uncertainty which can be controlled by the researcher. This amounts to treating the ATSM restrictions as a prior on the coefficients on the VAR. The tightness of the prior would naturally provide a measure of the uncertainty around the restrictions, i.e. of the misspecification of the model. To implement the ATSM prior, following Del Negro and Schorfheide (2004), we use a set of dummy artificial observations which are first generated under the maintained assumption of the validity of the ATSM and then added to the data. The ratio of dummy over actual observations will measure the weight of the prior information relative to the sample information.

Rewrite the VAR in the data-matrix notation:

$$Y = X\Phi + U, \tag{13}$$

where $Y$ is a $T \times q$ data-matrix with rows $Y_t'$, $X$ is a $T \times k$ (where $k = 1+qp$) data-matrix with rows $x_t = (1, Y_{t-1}', Y_{t-2}', ...Y_{t-p}')$, $\Phi = (\Phi_0, \Phi_1, ..., \Phi_p)'$, and $U$ is a $T \times q$ data-matrix with rows $u_t'$. Conditional on $Y_{1-p}, ..., Y_0$, the VAR has likelihood:

$$p(Y \mid \Phi, \Sigma_u) \propto |\Sigma_u|^{-T/2} \exp\{-0.5tr[\Sigma_u^{-1}(Y'Y - \Phi'X'Y - Y'X\Phi + \Phi'X'X\Phi)]\}. \tag{14}$$

Now assume there exist $T^* = \lambda T$ artificial observations $Y^*, X^*$, generated from the ATSM with a given $\theta$. The likelihood of the artificial data is:

$$p(Y^*(\theta) \mid \Phi, \Sigma_u) \propto |\Sigma_u|^{-\lambda T/2} \exp\{-0.5tr[\Sigma_u^{-1}(Y^{*\prime}Y^* - \Phi'X^{*\prime}Y^* - Y^{*\prime}X^*\Phi + \Phi'X^{*\prime}X^*\Phi)]\}. \tag{15}$$

The likelihood function of the combined sample of actual and artificial data is:

$$p(Y^*(\theta), Y \mid \Phi, \Sigma_u) = p(Y^*(\theta) \mid \Phi, \Sigma_u)p(Y \mid \Phi, \Sigma_u). \tag{16}$$

The term $p(Y^*(\theta) \mid \Phi, \Sigma_u)$ can be interpreted as a prior density for $\Phi$ and $\Sigma_u$ (conditional on $\theta$) as it contains information about the VAR parameters contained in the sample of artificial observations.

The next step is to remove the stochastic variation, which is undesirable, from $p(Y^*(\theta) \mid \Phi, \Sigma_u)$. To do so it is sufficient to replace the sample moments $Y^{*\prime}Y^*, Y^{*\prime}X^*$, and

$X^{*\prime}X^*$ with their expected values. Define $\Gamma^*_{yy}(\theta), \Gamma^*_{yx}(\theta)$, and $\Gamma^*_{xx}(\theta)$, which can be computed using equation (11), as the population moments under the validity of the ATSM. The expected values of $Y^{*\prime}Y^*, Y^{*\prime}X^*$, and $X^{*\prime}X^*$ are simply given by the corresponding population moments multiplied by the number of artificial observations, respectively $\lambda T\Gamma^*_{yy}(\theta), \lambda T\Gamma^*_{yx}(\theta)$ and $\lambda T\Gamma^*_{xx}(\theta)$. Provided that $\lambda T \geq k+q$ and $\Gamma^*_{xx}(\theta)$ is invertible, this gives a proper, nondegenerate, Normal-Inverted Wishart prior density $(\Phi, \Sigma_u \mid \theta)$:

$$\Phi \mid \Sigma_u,\ \theta \sim N(\Phi^*(\theta), \Sigma_u \otimes (\lambda T\Gamma^*_{xx}(\theta))^{-1}), \tag{17}$$

$$\Sigma_u \mid \theta \sim IW(\lambda T\Sigma^*_u(\theta), \lambda T - k, q), \tag{18}$$

where $\Phi^*(\theta)$ and $\Sigma^*_u(\theta)$ are the VAR coefficients under the maintained assumption of the validity of the ATSM:

$$\Phi^*(\theta) = \Gamma^{*-1}_{xx}(\theta)\Gamma^*_{xy}(\theta), \tag{19}$$

$$\Sigma^*_u(\theta) = \Gamma^*_{yy}(\theta) - \Gamma^*_{yx}(\theta)\Gamma^{*-1}_{xx}(\theta)\Gamma^*_{xy}(\theta). \tag{20}$$

Conditional on $\theta$ the ATSM prior and the likelihood function are conjugate, so the posterior distribution of $\Phi$ and $\Sigma_u$ is also Normal-Inverted Wishart (see Zellner 1971):

$$\Phi \mid Y,\ \Sigma_u,\ \theta \sim N(\tilde{\Phi}(\theta), \Sigma_u \otimes (\lambda T\Gamma^*_{xx}(\theta) + X'X)^{-1}), \tag{21}$$

$$\Sigma_u \mid Y,\ \theta \sim IW((\lambda+1)T\tilde{\Sigma}_u(\theta), (\lambda+1)T - k, q), \tag{22}$$

where $\tilde{\Phi}(\theta)$ and $\tilde{\Sigma}_u(\theta)$ are the maximum-likelihood estimates of $\Phi$ and $\Sigma_u$ based on both the artificial and the actual data:

$$\tilde{\Phi}(\theta) = (\lambda T\Gamma^*_{xx}(\theta) + X'X)^{-1}(\lambda T\Gamma^*_{xy}(\theta) + X'Y), \tag{23}$$

$$\tilde{\Sigma}_u(\theta) = \frac{1}{\lambda T + T}[(\lambda T\Gamma^*_{xy}(\theta) + Y'Y) - (\lambda T\Gamma^*_{yx}(\theta) + Y'X)(\lambda T\Gamma^*_{xx}(\theta) + X'X)^{-1}(\lambda T\Gamma^*_{xy}(\theta) + X'Y)]. \tag{24}$$

The parameter $\lambda = T^*/T$ is the ratio between artificial and actual data and it can be interpreted as the tightness of the prior. When $\lambda \to 0$, no artificial data are used and as is clear from (23) the posterior mean of $\Phi$ approaches the OLS estimate. On the other side, when $\lambda \to \infty$ the number of artificial data is very high compared to actual data, so the posterior mean of $\Phi$ approaches the prior mean $\Phi^*(\theta)$ (see equations (23) and (19)), and the procedure is equivalent to estimating the VAR subject to the ATSM

restrictions. Indeed it is possible to show that for fixed $T$ and $\lambda \to \infty$ the marginal likelihood function:

$$\begin{aligned} p(Y \quad | \quad \theta) &= p(Y \mid \Phi, \Sigma_u) p(\Phi, \Sigma_u \mid \theta) / p(\Phi, \Sigma_u \mid Y) \\ &= \frac{|\lambda T \Gamma_{xx}^*(\theta) + X'X|^{-\frac{q}{2}} \left|(\lambda+1) T \tilde{\Sigma}_u(\theta)\right|^{-\frac{(\lambda+1)T-k}{2}}}{|\lambda T \Gamma_{xx}^*(\theta)|^{-\frac{q}{2}} |\lambda T \Sigma_u^*(\theta)|^{-\frac{\lambda T-k}{2}}} \\ &\quad \times \frac{2\pi^{\frac{-qT}{2}} 2^{\frac{q((\lambda+1)T-k)}{2}} \Pi_{i=1}^q \Gamma[((\lambda+1)T-k+1-i)/2]}{2^{\frac{q(\lambda T-k)}{2}} \Pi_{i=1}^q \Gamma[(\lambda T-k+1-i)/2]}, \end{aligned} \tag{25}$$

where $\Gamma[.]$ denotes the gamma function, approaches the (quasi)-likelihood function of the ATSM model. See Del Negro and Schorfheide (2004) for the proof. Importantly, the approach allows to perform posterior inference also with respect to $\theta$. The marginal posterior $p(\theta \mid Y)$ can be obtained simply by combining the marginal likelihood $p(Y \mid \theta)$ with the prior $p(\theta)$, as explained in more detail in the next Section.

# 4 Estimation

In this section we estimate our VAR with ATSM prior. We start with describing the data, the parameterization, and the estimation procedure used. Then we provide the estimation results obtained using the whole sample. Results of the forecasting experiment are described in Section 5.

## 4.1 Data description

We estimate the model using monthly data (end of month rates) on US yields of five different maturities, 1-month, 3-month, 1-year, 3-year and 5-year. We use Treasury yield curve estimates of the Federal Reserve Board provided by Gurkaynak et al (2006) and publicly available on the website http://www.federalreserve.gov/pubs/feds/2006. A graph of the data is displayed in Figure 1.

To avoid problems of instability of the underlying data generating process during the so called Volker-era, we use data starting from January 1983 until December 2006, which provides us with a sample size of 24 years. As for the lag length of the VAR, the BIC selects 1 lag and the AIC 2 lags, but in both these cases the LM test statistic reported in Johansen (1995) rejects the null of no autocorrelation in the disturbances, therefore we choose a richer specification with 4 lags which eliminates this problem. To sum up, our baseline unrestricted VAR specification for the vector of yields $Y_t =$

$(y_t^{(1)}, y_t^{(3)}, y_t^{(12)}, y_t^{(24)}, y_t^{(60)})'$ is:

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \Phi_3 Y_{t-3} + \Phi_4 Y_{t-4} + u_t. \tag{26}$$

As $Y_t$ is $5 \times 1$, the matrices $\Phi_0, \Phi_1, \Phi_2, \Phi_3, \Phi_4$ contain a total of 105 coefficients to be estimated.

### 4.2 Choice of parameterization

Given that scaling, shifting, or rotation of the factors provides observational equivalence, a normalization is required. Following Dai and Singleton (2000) we impose the canonical normalization of the model, i.e. we identify the factors by assuming lower triangular structure for the matrix $\Psi$, and we set $\delta_1$ equal to a vector of ones. Given that we also assumed the factors have zero mean, $\delta_0$ equals the unconditional mean of the short term rate $r_t$, so it can be estimated consistently in a preliminary step by using the sample average of the 1-month yield $y_t^{(1)}$. As for second order coefficients, we assume $\Omega$ to be diagonal, $\Sigma_{\varepsilon v} = 0$, and $\Sigma_v = \sigma_v^2 I$.

In view of the simulation approach needed to estimate the model, we try to keep the number of parameter to be estimated as small as possible by assuming two factors. As discussed in Diebold et al. (2005) two factors suffice to capture the time-series variation and forecast yields, since the first two principal components account for almost all (99 percent) of the variation in yields, while more than two factors (typically three) will invariably be needed in order to obtain a close fit to the entire yield curve at any point in time for pricing derivatives. Note that the chosen normalization implies that the sum of the factors equals the (demeaned) short term rate, which is consistent with interpreting the first factor as a level and the second as a slope factor.

Given the above assumptions equation (2) becomes:

$$\Lambda_t = \underbrace{\begin{bmatrix} \Lambda_0^{(1)} \\ \Lambda_0^{(2)} \end{bmatrix}}_{\Lambda_0} + \underbrace{\begin{bmatrix} \Lambda_1^{(11)} & \Lambda_1^{(12)} \\ \Lambda_1^{(21)} & \Lambda_1^{(22)} \end{bmatrix}}_{\Lambda_1} \begin{bmatrix} F_{t-1}^{(1)} \\ F_{t-1}^{(2)} \end{bmatrix}, \tag{27}$$

and the state space in (10) simplifies to:

$$\begin{bmatrix} F_t^{(1)} \\ F_t^{(2)} \end{bmatrix} = \underbrace{\begin{bmatrix} \psi_{11} & 0 \\ \psi_{21} & \psi_{22} \end{bmatrix}}_{\Psi} \begin{bmatrix} F_{t-1}^{(1)} \\ F_{t-1}^{(2)} \end{bmatrix} + \underbrace{\begin{bmatrix} \omega_{11} & 0 \\ 0 & \omega_{22} \end{bmatrix}}_{\Omega} \begin{bmatrix} \varepsilon_{t-1}^{(1)} \\ \varepsilon_{t-1}^{(2)} \end{bmatrix} \tag{28}$$

$$Y_t = A(\Lambda_0, \Lambda_1, \Psi, \Omega) + B(\Lambda_0, \Lambda_1, \Psi, \Omega) F_t + v_t$$

where $A(\Lambda_0, \Lambda_1, \Psi, \Omega)$ and $B(\Lambda_0, \Lambda_1, \Psi, \Omega)$ obey to equations (8) and $var(v_t) = \sigma_v^2 I$ . The vector of structural parameters to be estimated is:

$$\theta = \{\psi_{11}, \psi_{21}, \psi_{22}, \omega_{11}, \omega_{22}, \sigma_v^2, \Lambda_0^{(1)}, \Lambda_0^{(2)}, \Lambda_1^{(11)}, \Lambda_1^{(12)}, \Lambda_1^{(21)}, \Lambda_1^{(22)}\}. \tag{29}$$

### 4.3 Posterior simulation

The posterior distribution of the ATSM coefficients $\theta$ and of the VAR coefficients $\Phi$ (for a given $\lambda$) is obtained by using MCMC methods, drawing in turn from $p(\theta \mid Y)$, $p(\Sigma_u \mid Y, \theta)$, and $p(\Phi \mid Y, \Sigma_u, \theta)$.

To draw from $p(\theta \mid Y) \propto p(Y \mid \theta)p(\theta)$ we use a Random Walk Metropolis (RMW) algorithm, which generates Markov chains with stationary distributions that correspond to the posterior distributions of interest. The algorithm requires the evaluation of $p(Y \mid \theta)p(\theta)$. The marginal likelihood $p(Y \mid \theta)$ is provided by equation (25), while the prior distribution $p(\theta)$ used throughout the paper is described in Table 1.

First, the algorithm is initialized by specifying an initial value $\theta^0$. Then for each iteration $s$ a proposal value $\vartheta$ for $\theta^s$ is drawn from the proposal distribution $N(\theta^{s-1}, c^2\Xi)$, where $c$ is a scaling parameter and $\Xi$ is a variance matrix. The jump from $\theta^{s-1}$ to $\vartheta$ is accepted with probability $\min\{1, r(\theta^{s-1}, \vartheta \mid Y)\}$, where

$$r(\theta^{s-1}, \vartheta \mid Y) = \frac{p(Y \mid \vartheta)p(\vartheta)}{p(Y \mid \theta^{s-1})p(\theta^{s-1})}. \tag{30}$$

If the jump is accepted, then the draw $\vartheta$ is kept and $\theta^s = \vartheta$, otherwise it is thrown away and $\theta^s = \theta^{s-1}$. The algorithm terminates when the final draw $\bar{s}$ is obtained.

We initialize the algorithm with $\theta^0$ equal to the maximum likelihood estimate of the state space system in (28), and $\Xi$ equal to the inverse of the Hessian in this point. The parameter $c$ is calibrated to 0.05, which provides a rejection percentage of about 35%. As the RWM algorithm generates draws that are influenced by the starting point and correlated, we discard the initial draws and we perform skip-sampling. In the next subsection we provide an informal assessment of the convergence of the algorithm in our application, for a review of the regularity conditions needed to guarantee the convergence of the Markov chain to the posterior distribution of interest see Geweke (2005).

Finally, given the sequence of draws $\{\theta^s\}_{s=1}^{\bar{s}}$ obtained from the RWM algorithm, we obtain posterior estimates of $\Phi$ by drawing from $p(\Sigma_u \mid Y, \theta)$ and $p(\Phi \mid Y, \Sigma_u, \theta)$ using (22) and (21).

### 4.4 Estimation results

The estimation results are conditional on the choice of the tightness parameter $\lambda$. We carry on our analysis for a grid of values of the tightness hyperparameter, i.e. $\lambda \in \{\infty, 10, 5, 4, 3, 2, 1, 0.5, 0.25, 0\}$. When $\lambda \to 0$ the prior is very loose and the posterior mean of $\Phi$ approaches the OLS estimate, while when $\lambda \to \infty$ the prior is very tight, which amounts to estimating the VAR subject to the ATSM restrictions. For a theoretical derivation of this result see Del Negro and Schorfheide (2004). To provide an intuitive illustration of this point Figure 2 plots posterior estimates of some elements of $\Phi$ (the coefficients on the first lag of the dependent variable for each equation of the VAR). As it is clear from the graph, when $\lambda \to \infty$ the data do not influence the posterior estimates, which stay close to the value implied by ATSM, while as $\lambda$ increases, the posterior mean moves towards the unrestricted OLS estimate of the coefficient.

In this section we briefly present the estimation results for the case $\lambda \to \infty$, which is interesting as it corresponds to the Bayesian estimation of the exact ATSM model. Estimation results for the other values of $\lambda$ show the same pattern (in terms of convergence) and are available upon request. To assess the convergence of our MCMC algorithm we run four independent Markov chains with 2000000 draws. As the RWM algorithm generates draws that are influenced by the starting point and correlated we discard the first 20%, and then keep every $10th$ draw among the remaining. This provides us with a total of 160000 draws.

Based on these draws, we plot in Figure 3 and Figure 4 the posterior distributions of the VAR coefficients $\Phi$ and of the ATSM coefficients $\theta$. As it is clear in Figure 3, the four independent Markov chains do converge to the same distributions for all the VAR coefficients. This is important because it means that the forecasts produced by the VAR would converge to the same distribution as well. As shown in Figure 4, also the posterior estimates of the coefficients governing the factors dynamics $(\psi_{11}, \psi_{21}, \psi_{22}, \omega_{11}, \omega_{22})$ are all converging to the same distribution. The implied posterior means of the factors are indistinguishable among the four chains, and are plotted in Figure 5. To sum up, all four chains converge to the same distribution of the VAR coefficients and factors dynamics.

Some problems arise for the coefficients in the price of risk equation: Figure 4 shows that the posterior distributions of the coefficients in the $\Lambda_1$ matrix are different among the four chains. This latter result is driven by the likelihood being flat along this dimension. Indeed, by looking at the rolling maximum likelihood estimates, the coefficients in $\Lambda_1$ are not identified, they jump and move together in couples with different combinations of their values providing roughly the same value for the likelihood. This is probably

due to the high nonlinearity of the model at hand and raises issues about maximum likelihood estimation of affine term structure models. Of course we may add additional identification restrictions, or exploit the Bayesian framework to tackle the problem by tightening the prior on $\Lambda_1$. However, we do not pursue this strategy for two reasons. First, we do not have enough information on the price of risk to specify a reasonable informative prior on it. Second, estimating the structural parameters governing the price of risk is not necessary for our aim, which is forecasting. As stressed above, to forecast we use the reduced form system given by our VAR, whose posterior estimates do not have problems of convergence.

# 5 Forecasting using the ATSM prior

We are now ready to perform a horse-race between the VAR with ATSM prior and several competitors. This section is divided into three subsection. The first describes the competing models, the second the forecasting exercise, the third provides results and discussion.

## 5.1 Competing models

We compare the forecasting performance of our proposed model, which we label ATSM-VAR, against six models: a random walk, two unrestricted VARs, the "exact" ATSM model, the model by Diebold and Li (2006), and a VAR with a Minnesota prior.

### Random Walk (RW)

Duffee (2002) has shown that beating a random walk with a traditional no arbitrage affine term structure model is difficult. Therefore we include in the comparison a simple random walk (RW) and use it as the benchmark model.

### "Exact" Affine Term Structure Model (ATSM)

It is natural to include in the comparison a VAR in which the ATSM restrictions hold exactly. This case corresponds to the ATSM-VAR with prior tightness $\lambda = \infty$ and is basically a two-factor version of the "yields only" model of Ang and Piazzesi (2003).

### Unrestricted VARs (UVAR4 and UVAR12)

We include in the comparison two unrestricted VARs with different lag length. The first is a VAR with 4 lags (UVAR4) which corresponds to the ATSM-VAR with prior

tightness set to zero ($\lambda = 0$). Then, we use a richer 12 lags specification (UVAR12), which corresponds to the benchmark UVAR used by Ang and Piazzesi (2003).

**Diebold-Li model (DL)**

A model which has proven to be good in forecasting, especially at long horizons, is that of Diebold and Li (2006), which is based on the Nelson and Siegel (1987) exponential components framework. In particular, we use the specification of Diebold et al. (2006) in which the model is written in its state space form before being estimated in a single step via maximum likelihood. The starting point is the Nelson and Siegel (1987) yield curve:

$$y_t(\tau) = \beta_1 + \beta_2 \frac{1 - e^{-\gamma\tau}}{\gamma\tau} + \beta_3 \left( \frac{1 - e^{-\gamma\tau}}{\gamma\tau} - e^{-\gamma\tau} \right), \tag{31}$$

where $\tau$ is the maturity and $\beta_1, \beta_2$,$\beta_3$ and $\gamma$ are parameters. Diebold and Li (2006) interpreted equation (31) in a dynamic fashion as a latent factor model in which $\beta_1, \beta_2$, and $\beta_3$ are time-varying level, slope, and curvature factors (label them $L_t, S_t, C_t$) and the terms that multiply these factors are factor loadings:

$$y_t(\tau) = L_t + S_t \frac{1 - e^{-\gamma\tau}}{\gamma\tau} + C_t \left( \frac{1 - e^{-\gamma\tau}}{\gamma\tau} - e^{-\gamma\tau} \right). \tag{32}$$

Once the dynamic movements of the factors $L_t, S_t, C_t$ are specified, then the model immediately forms a state-space system which can be estimated via maximum likelihood. We chose an univariate AR(1) representation for the three factors, as it has proven to be better in forecasting[1].

**VAR with Minnesota Prior (MP)**

Finally, to check whether the good forecasting performance is merely due to the use of a shrinkage estimator, we include in the comparison the Minnesota prior (Doan et al 1984). The Minnesota prior (MP) shrinks parameter estimates towards a random walk representation and has proven to be robustly good in forecasting (Litterman, 1986). In our case this amounts to shrinking the diagonal elements of $\Phi_1$ in (12) toward one and all the remaining coefficients toward zero.

We implement the Minnesota prior in the version proposed by Kadiyala and Karlsson (1997). This prior has a Normal-Inverted Wishart form such that the expectation of $\Sigma_u$

[1] For completeness we have also tried with a VAR(1) specification as in Diebold et al (2006) and it produces systematically worse results than the univariate AR(1) specification.

is equal to the fixed residual variance matrix of the Minnesota prior, and the prior expectation and variances of the coefficients $\Phi_k$ are:

$$E[\Phi_k^{(ij)}] = \begin{cases} 1, & for\ j = i,\ k = 1 \\ 0 & otherwise \end{cases} \quad ; \quad V[\Phi_k^{(ij)}] = \phi k^{-2} \frac{\sigma_i^2}{\sigma_j^2}. \tag{33}$$

The hyperparameter $\phi$ measures the tightness of the prior: when $\phi = 0$ the prior is imposed exactly and the data do not influence the estimates, while as $\phi \to \infty$ the prior becomes loose and the posterior estimates approach the OLS estimates. The factor $1/k^2$ is the rate at which prior variance decreases with increasing lag length while the ratio $\sigma_i^2/\sigma_j^2$ accounts for the different scale and variability of the data. Finally, the prior on the intercept is diffuse. Although the parameters should in principle be set using only prior knowledge we follow common practice (see e.g. Litterman, 1986; Sims and Zha, 1998) and set the scale parameters $\sigma_i^2$ equal the variance of a residual from a univariate autoregressive model for the variables. Posterior estimates can be easily obtained (via OLS) by implementing the prior in the form of dummy variable observations. For details see Kadiyala and Karlsson (1997).

### 5.2 Forecasting exercise

At each point in time we estimate the competing models and use them to produce forecasts up to 12-step ahead. We use a rolling estimation window of 10 years, so that the first estimation sample is 1983:1 to 1992:12 and the last is 1995:12 2005:12, while the first forecasting window is 1993:1 1993:12 and the last is 2006:1 2006:12.

Two of the models at hand, namely the ATSM-VAR and the MP, depend on tightness hyperparameters (respectively $\lambda$ and $\phi$) that need to be calibrated before estimation. We select ex-ante the optimal prior tightness by using a simple data-driven procedure based on past forecasting performance. In particular we do a grid search over the tightness hyperparameters selecting at each point in time, for each yield, the model producing the smallest forecast error in the previous period. The grid used for the ATSM-VAR is $\lambda \in \{\infty, 10, 5, 4, 3, 2, 1, 0.5, 0.25, 0\}$, that used for the MP is $\phi \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$.

For a given $\lambda$, the ATSM-VAR is estimated as described in Section 3, The UVAR(4) and the plain ATSM model are obtained as a by product imposing $\lambda = 0$ and $\lambda = \infty$. The UVAR(12) is estimated with OLS, the DL model with maximum likelihood, and the MP with OLS after rewriting the prior in the form of dummy observations.

We evaluate forecast accuracy using mean squared forecast error ($MSFE$) and mean

absolute forecast error ($MAFE$) as loss functions. To assess whether the difference in the forecasts is significant we use the test for predictive accuracy recently developed by Giacomini and White (2006). This is a test of equal forecasting *method* accuracy and as such can handle forecasts based on both nested and non-nested models, regardless from the estimation procedures used in the derivation of the forecasts, including Bayesian methods.

### 5.3 Empirical results

We report results for $MSFE$ and $MAFE$ of a given model $M$ relative to the random walk benchmark, at horizon $h$:

$$RMSFE_{M,h} = \frac{MSFE_M}{MSFE_{RW,h}}, \quad RMAFE_{M,h} = \frac{MAFE_{M,h}}{MAFE_{RW,h}}. \tag{34}$$

Relative Mean Squared Forecast Errors (RMSFE) are displayed in Table 2, while Relative Mean Absolute Forecast Errors (RMAFE) are displayed in Table 3. A number below 1 signals that a given model improves over the random walk. The tables report results for 1, 3, 6, 9 and 12 step-ahead forecasts, complete tables including all the forecast horizons are in the Appendix. The best forecasting models for each horizon and variable are highlighted in bold, while the stars on the right of the cell entries signal the level at which the Giacomini and White (2006) test rejects the null of equal forecasting method accuracy.

Several conclusions can be drawn from the tables.

First, the evidence for the 1-step ahead horizon is mixed. For the 1-month yield all the models but the UVAR12 are able to beat the RW, the best model being the MP, followed by the exact ATSM and the UVAR4. The MP is the best model also for the 3-month yield, followed by the UVAR4 and the ATSM-VAR, while the performance of the exact ATSM and of DL becomes worse than the RW. For yields with longer maturities (1-, 3-, 5- year) none of the models is able to beat the RW, with MP being the best for the 1-year yield, DL for the 3-year yield and DL or exact ATSM (depending on the loss function) for the 5-year yield.

Second, for longer horizons, the ATSM-VAR systematically outperforms all the other models, with substantial and significant gains in terms of both RMSFE and RMAFE. Importantly, it is the sole model able to systematically beat the RW. The MP shows a good performance for the 1-month yield but it deteriorates a lot for yields with longer maturities. The DL and the exact ATSM perform overall well, especially at very long

forecast horizons, but still they provide smaller gains than the ATSM-VAR and are occasionally outperformed by the RW.

Third, let us consider the results form the point of view of overparameterization. The UVAR12 forecasts very poorly. This is due to its overparameterization, as the more parsimonious UVAR4 is a clear improvement on it. Still, also the UVAR4 provides poor forecasts respect to the other models (except for 1-step ahead forecasts for short maturities). A further improvement is obtained by using shrinkage, i.e. using the MP.

Finally, related to the latter point, the inclusion of the Minnesota prior in the comparison also allows to check whether the good performance of the ATSM prior is merely due to the fact that it reduces the dimensionality of the VAR by shrinking coefficients. For forecast horizons above one-step ahead, the ATSM prior produces substantial gains against the Minnesota prior. Therefore, shrinkage in itself does help to improve forecast accuracy, but also the *direction* of the shrinkage (i.e. the assumed prior mean) plays a fundamental role. In other words, our results suggest that shrinking coefficients in the direction implied by ATSM models produces better forecasts than shrinking them to a random walk representation.

# 6 Conclusions

In this paper we proposed to forecast the yield curve by using no arbitrage affine term structure models, taking into account their possible misspecification. The key idea was to use the restrictions implied by an ATSM on a VAR as prior information rather than imposing them dogmatically. To do so we have used the approach proposed by Del Negro and Schorfheide (2004) for deriving priors from DSGE models for VARs.

Applying the proposed method to a system of five US yields provided substantial gains in forecast accuracy. In particular, while at the 1-step ahead horizon the evidence is mixed, for longer forecast horizons the proposed method systematically outperformed the unrestricted VAR and the exact ATSM, and it also outperformed very competitive benchmarks as the model proposed by Diebold and Li (2006), the random walk, and the Minnesota prior.

We believe that the inclusion of additional information in the model, and in particular information related to relevant macroeconomic variables as in Ang and Piazzesi (2003), could further improve the forecasting performance. Such extension would pose serious computational burdens and is in our agenda for future research.

## References


[1] Ang, A. and M. Piazzesi, 2003. A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables, Journal of Monetary Economics 50, 4, 745-787.

[2] Ang, A., M. Piazzesi, and M. Wei ,2004, "What does the Yield Curve Tell us about GDP Growth?" Journal of Econometrics 131, 359-403.

[3] Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407.

[4] Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 1943–1978.

[5] De Jong, F., 2000. Time series and cross section information in affine term structure models. Journal of Business and Economic Statistics 18, 300–314.

[6] Del Negro, M., Schorfheide, F, 2004, "Priors from General Equilibrium Models for VARs," *International Economic Review*, Vol. 45, 643-673.

[7] Del Negro, M., Schorfheide, F, 2005. "Policy Predictions if the Model Does Not Fit," Journal of the European Economic Association, MIT Press, 3(2-3), 434-443

[8] Del Negro, M., Schorfheide, F, 2007. Monetary Policy with Potentially Misspecified Models. NBER Working Paper 13099

[9] Diebold, F.X., Piazzesi, M. and Rudebusch, G.D., 2005. "Modeling Bond Yields in Finance and Macroeconomics," American Economic Review, 95, 415-420.

[10] Diebold, F. X., and Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337–364

[11] Diebold, F.X., Rudebusch, G.D. and Aruoba, B. 2006, "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach," Journal of Econometrics, 131, 309-338

[12] Doan T., Litterman R., Sims C., 1984, "Forecasting and Conditional Projection Using Realistic Prior Distributions", *Econometric Reviews*, Vol. 3, pp. 1-100.

[13] Duffee, G., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance 57, 405–443.

[14] Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical Finance 6, 379–406.

[15] Favero, C.A. Niu, L. Sala, L, 2007. "Term Structure Forecasting: No-Arbitrage Restrictions vs Large Information Set," CEPR Discussion Papers 6206, C.E.P.R. Discussion Papers.

[16] Geweke, J. (2005) *Contemporary Bayesian Econometrics and Statistics.* Wiley.

[17] Giacomini R., White H., 2006, "Tests of Conditional Forecast Accuracy", *Econometrica*, Vol. 74, pp. 1545-1578.

[18] Gurkaynak, Brian Sack, and Jonathan H. Wright, 2006, The U.S. Treasury Yield Curve: 1961 to the Present, FRB working paper, 2006-28

[19] Harrison, J.M., Kreps, D.M., 1979. Martingales and arbitrage in multiperiod securities markets. Journal of Economic Theory 2, 381–408.

[20] Ingram B., Whitemann C., 1994, "Supplanting the Minnesota prior. Forecasting Macroeconomic Time Series Using Real Business Cycle Priors", Journal of Monetary Economics, Vol. 34, pp. 497-510.

[21] Johansen S., 1995, *Likelihood-based Inference in Cointegrated Vector Auto-regressive Models*, Oxford University Press, Oxford.

[22] Kadiyala, K. R., and S. Karlsson 1997. "Numerical Methods for Estimation and Inference in Bayesian VAR-Models," Journal of Applied Econometrics, 12(2), 99–132.

[23] Litterman R.B., 1986, "Forecasting with Bayesian Vector Autoregressions-Five Years of Experience", *Journal of Business & Economic Statistics*, Vol. 4, pp. 25-38.

[24] Mönch, E. 2005, "Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augumented VAR Approach." ECB Working Paper No. 544.

[25] Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curve. Journal of Business 60, 473–489.

[26] Piazzesi, M. (2003). Affine term structure models. Manuscript prepared for the *Handbook of Financial Econometrics.*

[27] Rudebusch G, and Wu, T., 2004. "A macro-finance model of the term structure, monetary policy, and the economy," Proceedings, Federal Reserve Bank of San Francisco, issue Mar.

[28] Sims C.A., 2003, "Probability Models for Monetary Policy Decisions", manuscript, Princeton University.

[29] Sims, C. A., and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," International Economic Review, 39(4), 949–68.

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

# Tables and Figures

**Table1: Prior Distributions**

| Parameter | | Prior | |
|---|---|---|---|
| | Density | Para(1) | Para(2) |
| $\psi_{11}$ | Normal | 0.9 | 0.1 |
| $\psi_{21}$ | Normal | 0.1 | 0.1 |
| $\psi_{22}$ | Normal | 0.9 | 0.1 |
| $\omega_{11}$ | InvGamma | 1.0e-9 | 3 |
| $\omega_{22}$ | InvGamma | 0.5e-7 | 3 |
| $\Lambda_0^{(1)}$ | Normal | 0 | 1e+20 |
| $\Lambda_0^{(2)}$ | Normal | 0 | 1e+20 |
| $\Lambda_1^{(11)}$ | Normal | 0 | 1e+20 |
| $\Lambda_1^{(21)}$ | Normal | 0 | 1e+20 |
| $\Lambda_1^{(12)}$ | Normal | 0 | 1e+20 |
| $\Lambda_1^{(22)}$ | Normal | 0 | 1e+20 |
| $\sigma_v^2$ | InvGamma | 1.2e-9 | 3 |

The Table reports the prior densities used throughout the paper. Para(1) and Para(2) are mean and standard deviation for the Normal distributions and $s$ and $v$ for the Inverse Gamma distributions (where $p(\sigma \mid v, s) \propto \sigma^{-v-1} e^{-vs^2/2\sigma^2}$).

**Table 2: Forecasts Comparisons RMSFE**

| | 1-month yield | 3-month yield | 12-month yield | 36-month yield | 60-month yield |
|---|---|---|---|---|---|
| exact ATSM | | | | | |
| 1 | 0.673 | 1.011 | 1.960*** | 1.103** | 1.136* |
| 3 | 0.683*** | 0.951 | 1.269** | 1.030 | 1.019 |
| 6 | 0.783** | 0.931 | 1.082 | 0.967 | 0.970 |
| 9 | 0.829** | 0.907 | 0.979 | 0.913 | 0.917 |
| 12 | 0.849* | 0.890 | 0.918 | 0.881** | 0.886*** |
| UVAR(12) | | | | | |
| 1 | 1.557** | 2.167*** | 3.549*** | 3.665*** | 3.606*** |
| 3 | 1.483 | 2.337* | 3.792*** | 4.216*** | 4.217*** |
| 6 | 2.507 | 3.115* | 4.419** | 4.717*** | 4.522*** |
| 9 | 3.085 | 3.815* | 5.274** | 5.721** | 5.457*** |
| 12 | 4.079* | 4.811* | 6.067* | 6.306** | 5.954** |
| UVAR(4) | | | | | |
| 1 | 0.732 | 0.810 | 1.263 | 1.420*** | 1.408*** |
| 3 | 0.594** | 0.831 | 1.397* | 1.673*** | 1.710*** |
| 6 | 0.709 | 0.973 | 1.611* | 1.956** | 1.964** |
| 9 | 0.912 | 1.157 | 1.695 | 2.035* | 2.078* |
| 12 | 1.079 | 1.278 | 1.690 | 1.977 | 2.051* |
| ATSM-VAR | | | | | |
| 1 | 0.744 | 0.846 | 1.078 | 1.226*** | 1.191** |
| 3 | **0.455*** ** | **0.513** ** | **0.776** ** | **0.965** | **0.971** |
| 6 | **0.413** ** | **0.482** ** | **0.705*** ** | **0.915** | **0.920** |
| 9 | **0.423** ** | **0.472** ** | **0.649*** ** | **0.819*** | **0.833** |
| 12 | **0.434** ** | **0.506** ** | **0.641*** ** | **0.804** ** | **0.829*** |
| Minn. Prior | | | | | |
| 1 | **0.655** | **0.692*** ** | **1.032** | 1.160** | 1.165** |
| 3 | 0.459*** | 0.718 | 1.192 | 1.376* | 1.390** |
| 6 | 0.645 | 0.908 | 1.377 | 1.624* | 1.638* |
| 9 | 0.810 | 1.030 | 1.433 | 1.729 | 1.778 |
| 12 | 0.955 | 1.120 | 1.430 | 1.702 | 1.778 |
| Diebold-Li | | | | | |
| 1 | 0.815* | 1.361 | 1.668*** | **1.076** | **1.071** |
| 3 | 0.863** | 1.157* | 1.325** | 1.084 | 1.066 |
| 6 | 0.921** | 1.071 | 1.170 | 0.998 | 0.966 |
| 9 | 0.947* | 1.025 | 1.062 | 0.944 | 0.923 |
| 12 | 0.954 | 0.987 | 0.968 | 0.862 | 0.848 |

The Table reports Relative Mean Squared Forecast Error of a given model against the Random Walk benchmark. The stars * ** *** denote rejection (respectively at 10 ,5, and 1 percent) of the null of equal forecasting method accuracy according to the Giacomini-White (2006) test. The best model for each horizon and maturity is highlighted in bold.

**Table 3: Forecasts Comparisons RMAFE**

| | 1-month yield | 3-month yield | 12-month yield | 36-month yield | 60-month yield |
|---|---|---|---|---|---|
| exact ATSM | | | | | |
| 1 | 0.873** | 1.010 | 1.455*** | 1.057** | **1.040** |
| 3 | 0.822*** | 0.983 | 1.133** | 1.010 | 1.012 |
| 6 | 0.877*** | 0.967 | 1.025 | 0.989 | 0.995 |
| 9 | 0.898*** | 0.949 | 0.981 | 0.949 | 0.971 |
| 12 | 0.916*** | 0.943 | 0.967 | 0.931* | 0.943* |
| UVAR(12) | | | | | |
| 1 | 1.497*** | 1.570*** | 1.770*** | 1.781*** | 1.778*** |
| 3 | 1.212 | 1.520* | 1.877*** | 1.903*** | 1.882*** |
| 6 | 1.490 | 1.691* | 1.953** | 1.981*** | 1.982*** |
| 9 | 1.597 | 1.834* | 2.169** | 2.209*** | 2.260*** |
| 12 | 1.834 | 2.031* | 2.314** | 2.321*** | 2.317*** |
| UVAR(4) | | | | | |
| 1 | 0.908 | 0.936 | 1.117* | 1.156*** | 1.159*** |
| 3 | 0.759** | 0.901 | 1.168 | 1.250** | 1.250** |
| 6 | 0.845 | 0.999 | 1.217 | 1.295* | 1.312** |
| 9 | 0.946 | 1.076 | 1.285 | 1.364 | 1.436* |
| 12 | 1.025 | 1.135 | 1.310 | 1.371 | 1.426* |
| ATSM-VAR | | | | | |
| 1 | 0.913 | 0.925 | 1.025 | 1.079** | 1.077** |
| 3 | **0.672*** ** | **0.686*** ** | **0.842*** ** | **0.960** | **0.965** |
| 6 | **0.646*** ** | **0.679*** ** | **0.795*** ** | **0.917** | **0.906** |
| 9 | **0.642*** ** | **0.665*** ** | **0.748*** ** | **0.836** ** | **0.864*** |
| 12 | **0.626*** ** | **0.672** ** | **0.752*** ** | **0.839** ** | **0.878** ** |
| Minn. Prior | | | | | |
| 1 | **0.829*** ** | **0.872** ** | **1.019** | 1.070* | 1.077** |
| 3 | 0.673*** | 0.802 | 1.027 | 1.103 | 1.114 |
| 6 | 0.790 | 0.922 | 1.089 | 1.165 | 1.191 |
| 9 | 0.881 | 0.994 | 1.151 | 1.231 | 1.298 |
| 12 | 0.955 | 1.040 | 1.188 | 1.263 | 1.314 |
| Diebold-Li | | | | | |
| 1 | 0.908** | 1.101** | 1.312*** | **1.048** | 1.047 |
| 3 | 0.936** | 1.086** | 1.174*** | 1.044 | 1.026 |
| 6 | 0.951*** | 1.053 | 1.084 | 0.997 | 0.996 |
| 9 | 0.956** | 1.020 | 1.033 | 0.969 | 0.986 |
| 12 | 0.971* | 1.003 | 0.992 | 0.919 | 0.931 |

The Table reports Relative Mean Absolute Forecast Error of a given model against the Random Walk benchmark. The stars * ** *** denote rejection (respectively at 10 ,5, and 1 percent) of the null of equal forecasting method accuracy according to the Giacomini-White (2006) test. The best model for each horizon and maturity is highlighted in bold.

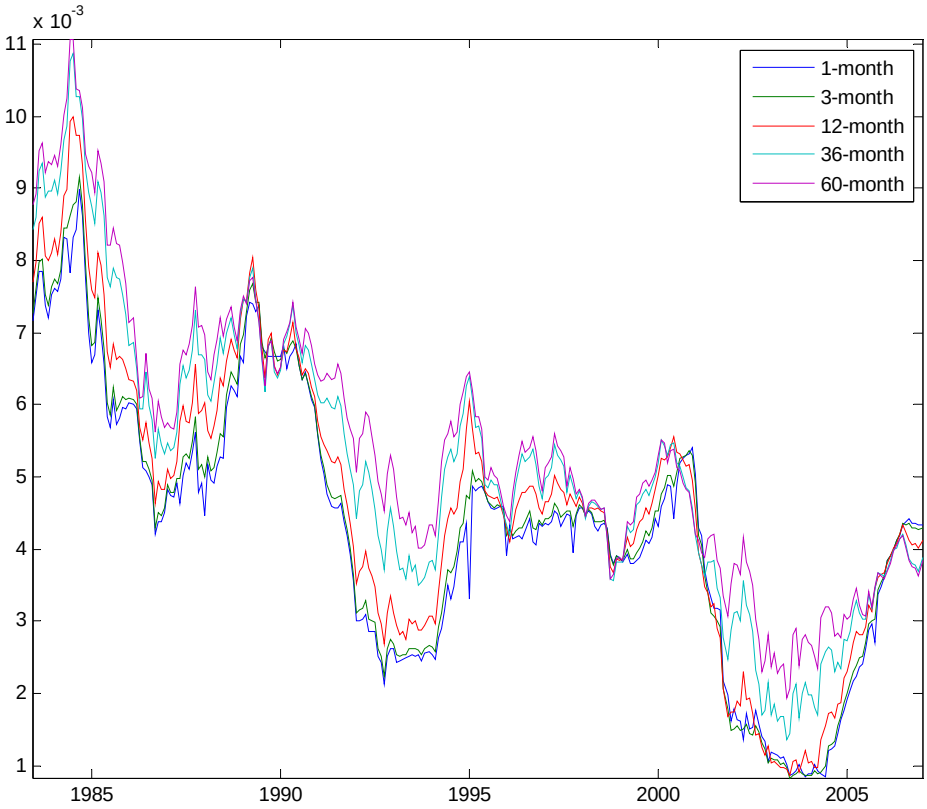


Figure 1: Data. Yields of 1-, 3-, 12-, 36-, 60- month maturities.

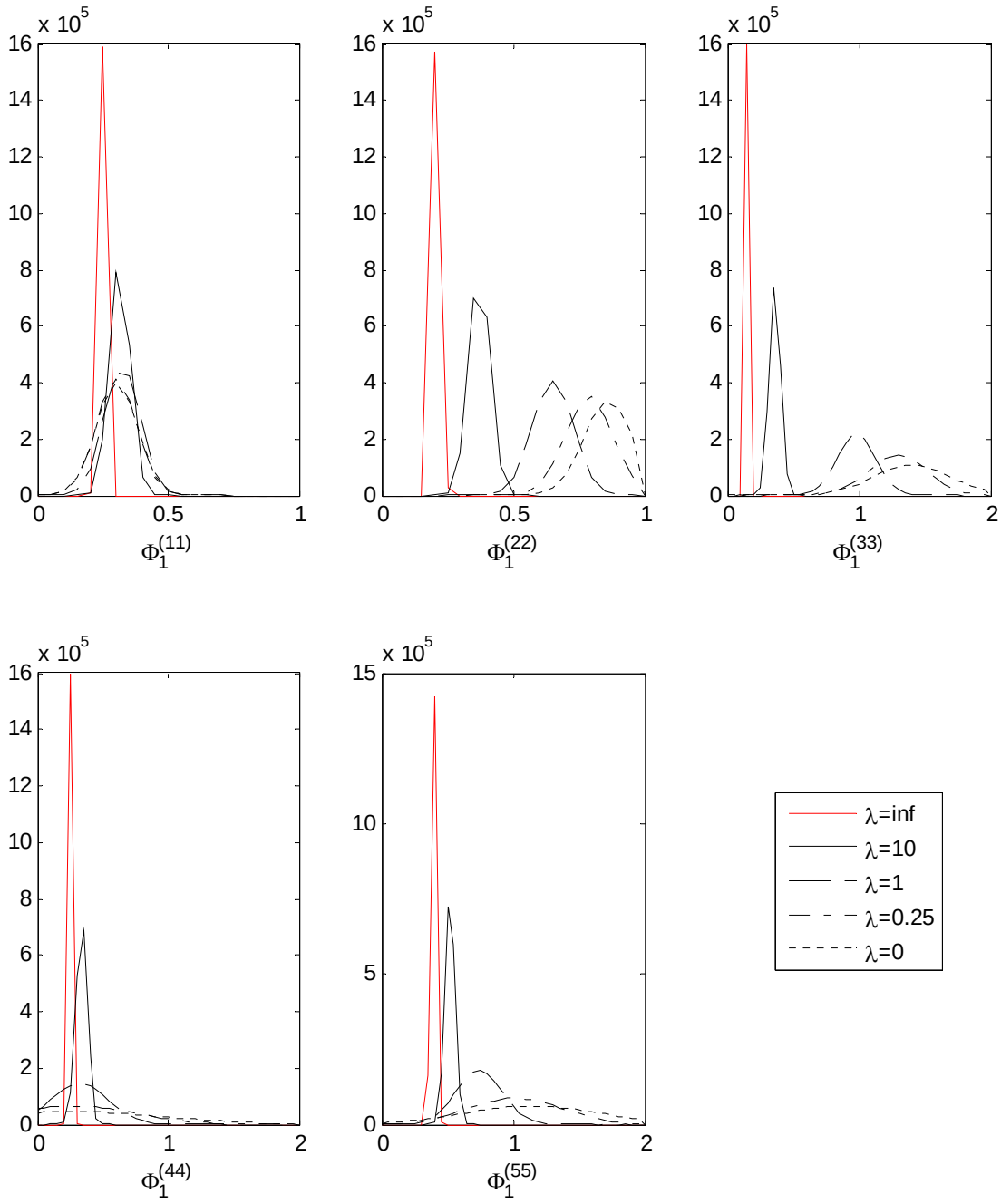


Figure 2: Posterior distribution of the VAR coefficients for different values of the prior tightness $\lambda$. The displayed coefficents are the diagonal elements of the matrix $\Phi_1$, i.e. the coefficients on the first lag of the dependent variable for each equation of the VAR.

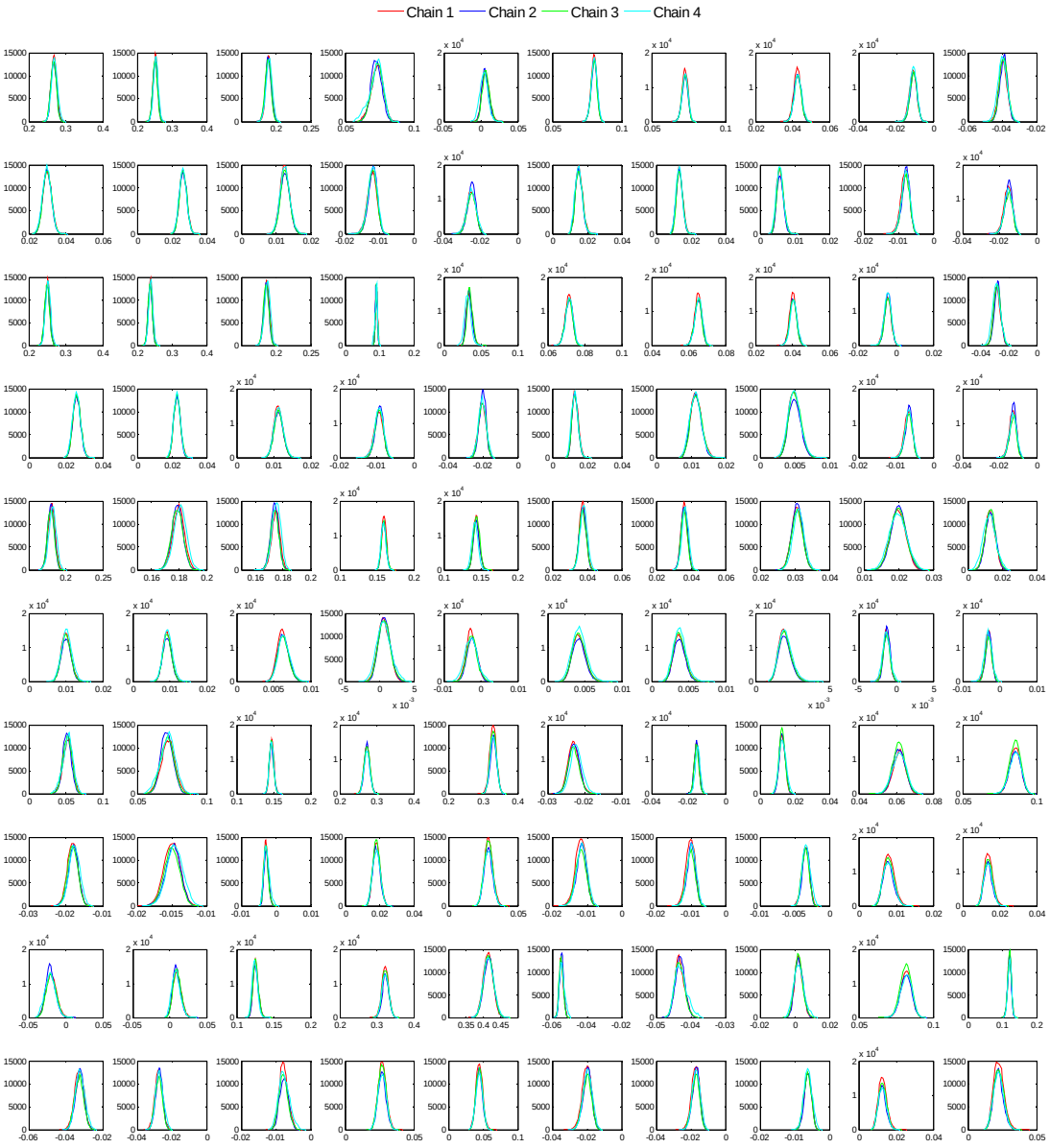


Figure 3: Posterior distribution of ATSM-VAR coefficients, obtained with 4 independent Markov chains. The matrices $\Phi_1, \Phi_2, \Phi_3, \Phi_4$ are $5 \times 5$ matrices containing a total of 100 coefficients. The first 25 graphs going from the top left to the right are the coefficients of $\Phi_1$, the following 25 are coefficients of $\Phi_2$, and so on. Within each group of 25 the coefficients of the generic matrix $\Phi_i$ , are ordered by row.

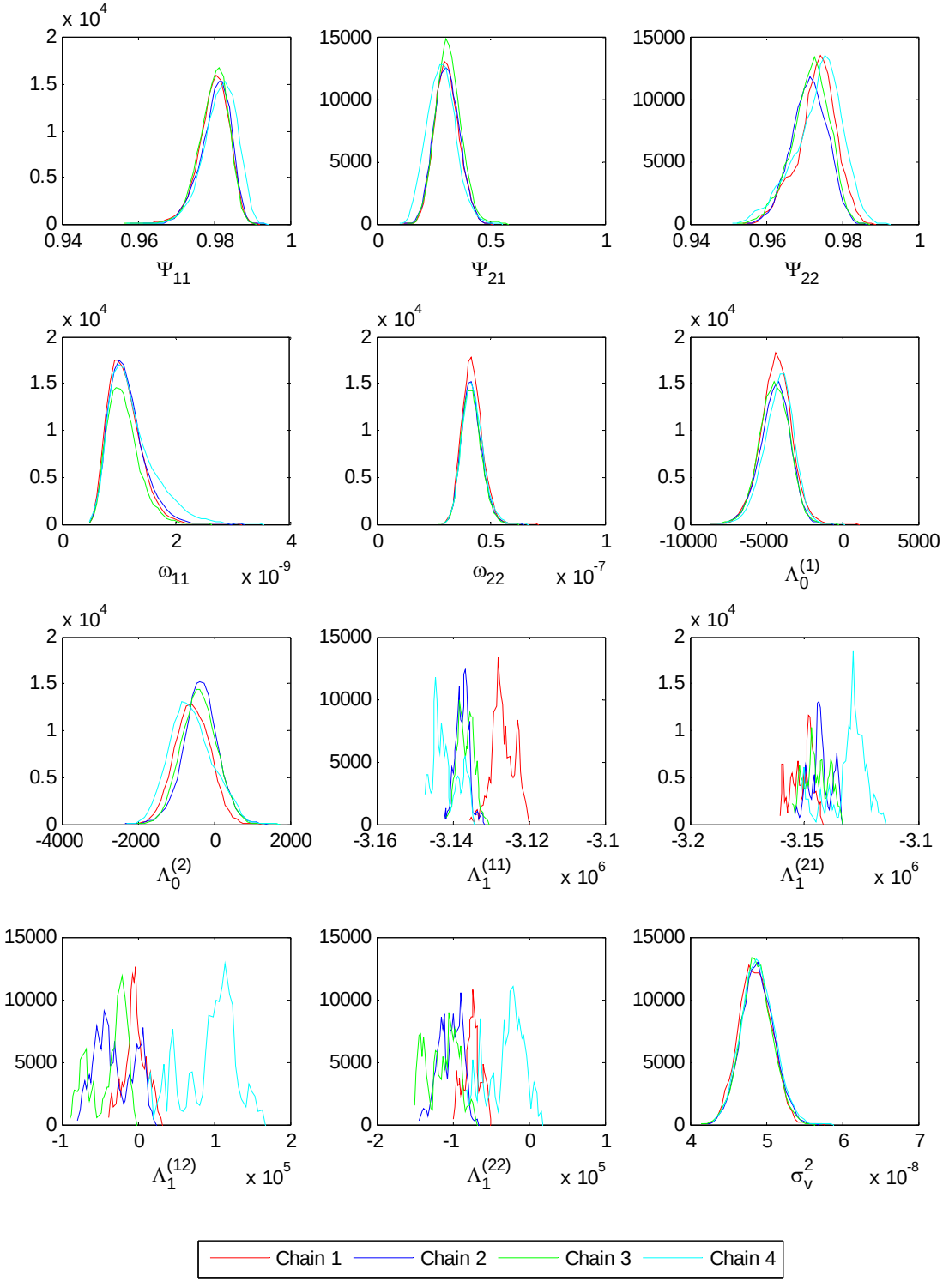


Figure 4: Posterior distribution of the structural ATSM coefficients, for 4 independent Markov chains (case $\lambda = \infty$).

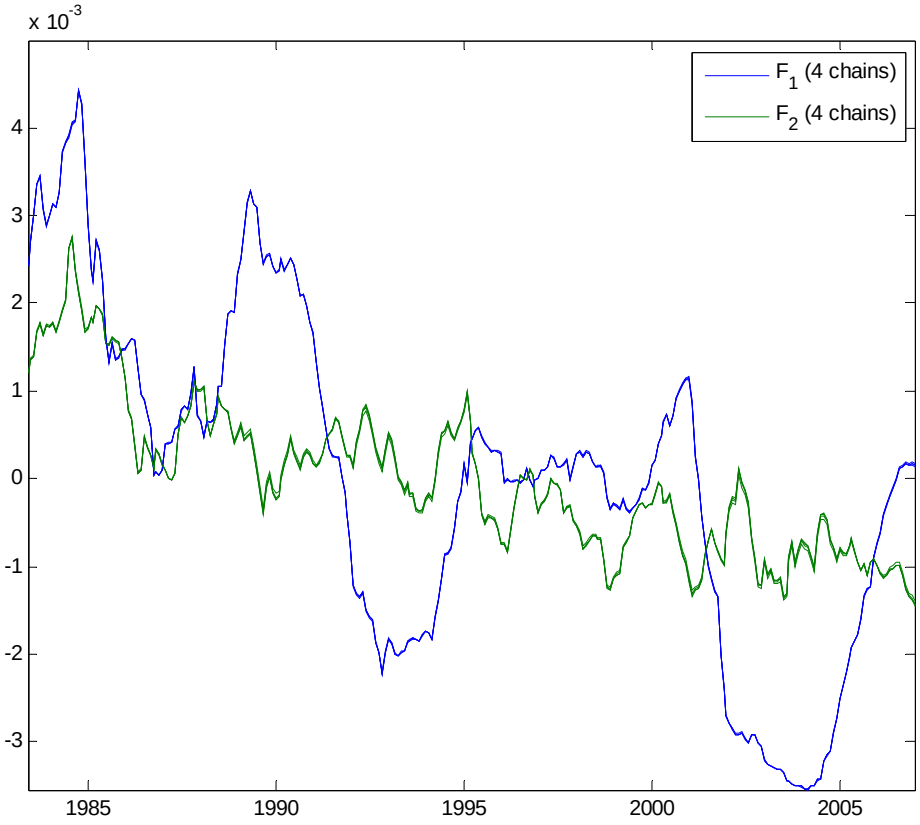


Figure 5: Estimated factors (case $\lambda = \infty$). The factors are given by $\hat{F}_t = (\hat{B}'\hat{B})^{-1}\hat{B}'(\hat{Y} - \hat{A})$, where the hats denote the posterior means. Note that the Figure displays four plots for each factor (one for each chain), but the plots are almost indistinguishable.

# Table Appendix.

The appendix reports Tables 2 and 3 with all the forecast horizons(1- to 12- step ahead).

**Table 2: Forecasts Comparisons MSFE**

| | 1-month yield | 3-month yield | 12-month yield | 36-month yield | 60-month yield |
|---|---|---|---|---|---|
| exact ATSM | | | | | |
| 1 | 0.673 | 1.011 | 1.960*** | 1.103** | 1.136* |
| 2 | 0.624*** | 0.944 | 1.447*** | 1.062 | 1.044 |
| 3 | 0.683*** | 0.951 | 1.269** | 1.030 | 1.019 |
| 4 | 0.714*** | 0.947 | 1.183* | 1.010 | 1.013 |
| 5 | 0.753*** | 0.940 | 1.129 | 0.989 | 0.987 |
| 6 | 0.783** | 0.931 | 1.082 | 0.967 | 0.970 |
| 7 | 0.801** | 0.920 | 1.036 | 0.941 | 0.939 |
| 8 | 0.819** | 0.912 | 1.004 | 0.923 | 0.920 |
| 9 | 0.829** | 0.907 | 0.979 | 0.913 | 0.917 |
| 10 | 0.828** | 0.897 | 0.952 | 0.898 | 0.904 |
| 11 | 0.832** | 0.891 | 0.933 | 0.887* | 0.887** |
| 12 | 0.849* | 0.890 | 0.918 | 0.881** | 0.886*** |
| UVAR(12) | | | | | |
| 1 | 1.557** | 2.167*** | 3.549*** | 3.665*** | 3.606*** |
| 2 | 1.385 | 2.187* | 3.675*** | 3.911*** | 3.867*** |
| 3 | 1.483 | 2.337* | 3.792*** | 4.216*** | 4.217*** |
| 4 | 1.699 | 2.621* | 4.121*** | 4.456*** | 4.344*** |
| 5 | 2.079 | 2.866* | 4.284** | 4.581*** | 4.411*** |
| 6 | 2.507 | 3.115* | 4.419** | 4.717*** | 4.522*** |
| 7 | 2.544 | 3.230* | 4.608** | 4.918*** | 4.706*** |
| 8 | 2.825 | 3.496* | 4.904** | 5.297** | 5.055*** |
| 9 | 3.085 | 3.815* | 5.274** | 5.721** | 5.457*** |
| 10 | 3.350 | 4.149* | 5.619** | 6.038** | 5.753** |
| 11 | 3.614 | 4.431* | 5.825** | 6.143** | 5.811** |
| 12 | 4.079* | 4.811* | 6.067* | 6.306** | 5.954** |
| UVAR(4) | | | | | |
| 1 | 0.732 | 0.810 | 1.263 | 1.420*** | 1.408*** |
| 2 | 0.582*** | 0.768 | 1.292** | 1.492*** | 1.502*** |
| 3 | 0.594** | 0.831 | 1.397* | 1.673*** | 1.710*** |
| 4 | 0.602** | 0.861 | 1.491* | 1.801*** | 1.820*** |
| 5 | 0.649* | 0.914 | 1.559* | 1.881*** | 1.888*** |
| 6 | 0.709 | 0.973 | 1.611* | 1.956** | 1.964** |
| 7 | 0.765 | 1.036 | 1.662* | 2.007** | 2.017** |
| 8 | 0.852 | 1.107 | 1.684 | 2.029* | 2.050** |
| 9 | 0.912 | 1.157 | 1.695 | 2.035* | 2.078* |
| 10 | 0.966 | 1.201 | 1.689 | 2.008 | 2.062* |
| 11 | 1.013 | 1.239 | 1.692 | 1.992 | 2.055* |
| 12 | 1.079 | 1.278 | 1.690 | 1.977 | 2.051* |

**Table 2 continued: Forecasts Comparisons MSFE**

| | 1-month yield | 3-month yield | 12-month yield | 36-month yield | 60-month yield |
|---|---|---|---|---|---|
| ATSM-VAR | | | | | |
| 1 | 0.744 | 0.846 | 1.078 | 1.226*** | 1.191** |
| 2 | 0.492*** | **0.558*** ** | **0.877** | **1.050** | **1.024** |
| 3 | **0.455*** ** | **0.513** ** | **0.776** ** | **0.965** | **0.971** |
| 4 | **0.427*** ** | **0.472*** ** | **0.727*** ** | **0.919** | **0.949** |
| 5 | **0.402*** ** | **0.507** ** | **0.730*** ** | **0.924** | **0.931** |
| 6 | **0.413** ** | **0.482** ** | **0.705*** ** | **0.915** | **0.920** |
| 7 | **0.408** ** | **0.480** ** | **0.662*** ** | **0.890** | **0.871** |
| 8 | **0.413** ** | **0.477** ** | **0.665** ** | **0.873** | **0.873** |
| 9 | **0.423** ** | **0.472** ** | **0.649*** ** | **0.819*** | **0.833** |
| 10 | **0.416** ** | **0.494** ** | **0.652** ** | **0.818*** | **0.830** |
| 11 | **0.400** ** | **0.492** ** | **0.645** ** | **0.809*** | **0.825** |
| 12 | **0.434** ** | **0.506** ** | **0.641*** ** | **0.804** ** | **0.829*** |
| Minn. Prior | | | | | |
| 1 | **0.655** | **0.692*** ** | **1.032** | 1.160** | 1.165** |
| 2 | **0.478*** ** | 0.660* | 1.138 | 1.301** | 1.303*** |
| 3 | 0.459*** | 0.718 | 1.192 | 1.376* | 1.390** |
| 4 | 0.480*** | 0.766 | 1.265 | 1.476* | 1.490* |
| 5 | 0.565** | 0.840 | 1.325 | 1.547* | 1.558** |
| 6 | 0.645 | 0.908 | 1.377 | 1.624* | 1.638* |
| 7 | 0.698 | 0.957 | 1.415 | 1.682* | 1.704* |
| 8 | 0.759 | 1.000 | 1.427 | 1.713 | 1.747* |
| 9 | 0.810 | 1.030 | 1.433 | 1.729 | 1.778 |
| 10 | 0.838 | 1.045 | 1.419 | 1.716 | 1.778 |
| 11 | 0.882 | 1.077 | 1.426 | 1.714 | 1.783 |
| 12 | 0.955 | 1.120 | 1.430 | 1.702 | 1.778 |
| Diebold-Li | | | | | |
| 1 | 0.815* | 1.361 | 1.668*** | **1.076** | **1.071** |
| 2 | 0.833** | 1.206* | 1.429*** | 1.088 | 1.061 |
| 3 | 0.863** | 1.157* | 1.325** | 1.084 | 1.066 |
| 4 | 0.885** | 1.124* | 1.259** | 1.059 | 1.043 |
| 5 | 0.906** | 1.094* | 1.209 | 1.019 | 0.991 |
| 6 | 0.921** | 1.071 | 1.170 | 0.998 | 0.966 |
| 7 | 0.935** | 1.055 | 1.129 | 0.982 | 0.955 |
| 8 | 0.945** | 1.038 | 1.095 | 0.967 | 0.943 |
| 9 | 0.947* | 1.025 | 1.062 | 0.944 | 0.923 |
| 10 | 0.949* | 1.012 | 1.024 | 0.911 | 0.893 |
| 11 | 0.951* | 1.000 | 0.994 | 0.883 | 0.865 |
| 12 | 0.954 | 0.987 | 0.968 | 0.862 | 0.848 |

**Table 3: Forecasts Comparisons MAFE**

| | 1-month yield | 3-month yield | 12-month yield | 36-month yield | 60-month yield |
|---|---|---|---|---|---|
| exact ASTM | | | | | |
| 1 | 0.873** | 1.010 | 1.455*** | 1.057** | **1.040** |
| 2 | 0.825*** | 0.979 | 1.242*** | 1.026 | **1.005** |
| 3 | 0.822*** | 0.983 | 1.133** | 1.010 | 1.012 |
| 4 | 0.859*** | 0.983 | 1.072 | 1.008 | 1.011 |
| 5 | 0.874*** | 0.977 | 1.048 | 1.000 | 0.993 |
| 6 | 0.877*** | 0.967 | 1.025 | 0.989 | 0.995 |
| 7 | 0.890*** | 0.964 | 1.007 | 0.970 | 0.984 |
| 8 | 0.894*** | 0.957 | 0.997 | 0.956 | 0.971 |
| 9 | 0.898*** | 0.949 | 0.981 | 0.949 | 0.971 |
| 10 | 0.893*** | 0.944 | 0.962 | 0.936* | 0.957 |
| 11 | 0.900*** | 0.942* | 0.965 | 0.928* | 0.944 |
| 12 | 0.916*** | 0.943 | 0.967 | 0.931* | 0.943* |
| UVAR(12) | | | | | |
| 1 | 1.497*** | 1.570*** | 1.770*** | 1.781*** | 1.778*** |
| 2 | 1.276* | 1.495** | 1.843*** | 1.809*** | 1.765*** |
| 3 | 1.212 | 1.520* | 1.877*** | 1.903*** | 1.882*** |
| 4 | 1.265 | 1.592* | 1.901*** | 1.952*** | 1.943*** |
| 5 | 1.380 | 1.627* | 1.912*** | 1.957*** | 1.946*** |
| 6 | 1.490 | 1.691* | 1.953** | 1.981*** | 1.982*** |
| 7 | 1.454 | 1.704 | 1.988** | 1.986*** | 2.011*** |
| 8 | 1.531 | 1.770* | 2.070** | 2.073*** | 2.120*** |
| 9 | 1.597 | 1.834* | 2.169** | 2.209*** | 2.260*** |
| 10 | 1.665 | 1.909* | 2.230** | 2.249*** | 2.281*** |
| 11 | 1.724 | 1.959* | 2.270** | 2.276*** | 2.299*** |
| 12 | 1.834 | 2.031* | 2.314** | 2.321*** | 2.317*** |
| UVAR(4) | | | | | |
| 1 | 0.908 | 0.936 | 1.117* | 1.156*** | 1.159*** |
| 2 | 0.784*** | 0.869 | 1.138* | 1.203*** | 1.196*** |
| 3 | 0.759** | 0.901 | 1.168 | 1.250** | 1.250** |
| 4 | 0.762** | 0.941 | 1.195 | 1.294** | 1.292** |
| 5 | 0.802* | 0.969 | 1.214 | 1.321** | 1.331** |
| 6 | 0.845 | 0.999 | 1.217 | 1.295* | 1.312** |
| 7 | 0.872 | 1.026 | 1.236 | 1.296* | 1.327* |
| 8 | 0.912 | 1.056 | 1.261 | 1.325 | 1.386** |
| 9 | 0.946 | 1.076 | 1.285 | 1.364 | 1.436* |
| 10 | 0.958 | 1.091 | 1.283 | 1.353 | 1.410* |
| 11 | 0.982 | 1.113 | 1.300 | 1.365 | 1.423* |
| 12 | 1.025 | 1.135 | 1.310 | 1.371 | 1.426* |

**Table 3 continued: Forecasts Comparisons MAFE**

| | 1-month yield | 3-month yield | 12-month yield | 36-month yield | 60-month yield |
|---|---|---|---|---|---|
| ASTM-VAR | | | | | |
| 1 | 0.913 | 0.925 | 1.025 | 1.079** | 1.077** |
| 2 | **0.708*** ** | **0.768*** ** | **0.923** | **1.018** | 1.009 |
| 3 | **0.672*** ** | **0.686*** ** | **0.842*** ** | **0.960** | **0.965** |
| 4 | **0.655*** ** | **0.679*** ** | **0.811*** ** | **0.926*** | **0.947** |
| 5 | **0.636*** ** | **0.713*** ** | **0.813*** ** | **0.934** | **0.931** |
| 6 | **0.646*** ** | **0.679*** ** | **0.795*** ** | **0.917** | **0.906** |
| 7 | **0.635*** ** | **0.681*** ** | **0.766*** ** | **0.878*** | **0.880*** |
| 8 | **0.637*** ** | **0.669*** ** | **0.768*** ** | **0.871*** | **0.891** |
| 9 | **0.642*** ** | **0.665*** ** | **0.748*** ** | **0.836** ** | **0.864*** |
| 10 | **0.621*** ** | **0.679** ** | **0.742*** ** | **0.834** ** | **0.846** ** |
| 11 | **0.598*** ** | **0.664*** ** | **0.748*** ** | **0.834** ** | **0.864** ** |
| 12 | **0.626*** ** | **0.672** ** | **0.752*** ** | **0.839** ** | **0.878** ** |
| Minn. Prior | | | | | |
| 1 | **0.829*** ** | **0.872** ** | **1.019** | 1.070* | 1.077** |
| 2 | 0.728*** | 0.813** | 1.052 | 1.141** | 1.127** |
| 3 | 0.673*** | 0.802 | 1.027 | 1.103 | 1.114 |
| 4 | 0.679*** | 0.841 | 1.056 | 1.135 | 1.140 |
| 5 | 0.737** | 0.886 | 1.081 | 1.174 | 1.200* |
| 6 | 0.790 | 0.922 | 1.089 | 1.165 | 1.191 |
| 7 | 0.811 | 0.947 | 1.112 | 1.171 | 1.195 |
| 8 | 0.847 | 0.978 | 1.138 | 1.195 | 1.245 |
| 9 | 0.881 | 0.994 | 1.151 | 1.231 | 1.298 |
| 10 | 0.887 | 1.002 | 1.162 | 1.238 | 1.295 |
| 11 | 0.916 | 1.023 | 1.176 | 1.258 | 1.317 |
| 12 | 0.955 | 1.040 | 1.188 | 1.263 | 1.314 |
| Diebold-Li | | | | | |
| 1 | 0.908** | 1.101** | 1.312*** | **1.048** | 1.047 |
| 2 | 0.923*** | 1.093** | 1.243*** | 1.054 | 1.042 |
| 3 | 0.936** | 1.086** | 1.174*** | 1.044 | 1.026 |
| 4 | 0.937*** | 1.077** | 1.141** | 1.039 | 1.020 |
| 5 | 0.949** | 1.066** | 1.110 | 1.017 | 1.009 |
| 6 | 0.951*** | 1.053 | 1.084 | 0.997 | 0.996 |
| 7 | 0.958*** | 1.047 | 1.066 | 0.969 | 0.972 |
| 8 | 0.955** | 1.030 | 1.048 | 0.957 | 0.970 |
| 9 | 0.956** | 1.020 | 1.033 | 0.969 | 0.986 |
| 10 | 0.958** | 1.013 | 1.006 | 0.931 | 0.949 |
| 11 | 0.963** | 1.010 | 0.996 | 0.919 | 0.942 |
| 12 | 0.971* | 1.003 | 0.992 | 0.919 | 0.931 |


This working paper has been produced by
the Department of Economics at
Queen Mary, University of London



Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm