Research Division
Federal Reserve Bank of St. Louis
Working Paper Series

Forecasts of US Short-term Interest Rates:
A Flexible Forecast Combination Approach

Massimo Guidolin
and
Allan Timmermann

_____

Forecasts of US Short-term Interest Rates: A Flexible Forecast Combination Approach[*]

Massimo Guidolin[†]

Federal Reserve Bank of St. Louis, Research Division and
Manchester Business School, MAGF

Allan Timmermann[‡]

University of California San Diego

March 15, 2007

Abstract

This paper develops a flexible approach to combine forecasts of future spot rates with forecasts from time-series models or macroeconomic variables. We find empirical evidence that accounting for both regimes in interest rate dynamics and combining forecasts from different models helps improve the out-of-sample forecasting performance for US short-term rates. Imposing restrictions from the expectations hypothesis on the forecasting model are found to help at long forecasting horizons.

Keywords: forecast combinations; regime switches;short term interest rates; expectations hypothesis.

JEL.codes: C53, C32, E47.

1. Introduction

Accurate interest rate forecasts are crucial for investors' savings and investment decisions as well as in many monetary policy decisions. How best to model and predict the dynamics in interest rates is therefore an issue that has long occupied researchers in finance and econometrics. Three broad approaches can be identified. Since interest rates tend to be highly persistent, autoregressive models are commonly used to produce benchmark forecasts. Alternatively, under the expectations hypothesis, forward rates can be expected to provide optimal and−under additional restrictions−unbiased forecasts of future spot rates. A third class of forecasting models expands the information set and identifies macroeconomic variables as predictors.

Application of these forecasting models is made difficult by evidence of complicated non-linear dynamics in interest rates. This suggests that the relationship between spot and forward rates changes over time in a way that can sometimes involve discrete shifts, possibly due to changes in monetary policies.[1] Furthermore,

---

[1]Ait-Sahalia (1996) finds evidence of non-linear mean reversion in interest rates. Hall, Anderson and Granger (1992) find that the expectations hypothesis is rejected during periods that coincide with changes in the underlying monetary policy regime.

empirical findings suggest that the success of the expectations hypothesis and the ability of forward rates to forecast future spot rates may vary across forecast horizons: forward rates appear to be unbiased predictors of movements in future spot rates at longer horizons but perform worse at short horizons.

Building on the insight that the relative precision of different forecasting approaches varies both over time and across forecast horizons, we propose in this paper a flexible forecast combination approach that does not restrict the forecaster to always choose one econometric model over other alternatives. Instead, the approach allows forecasts from different models to be combined and lets the combination weights vary across a set of underlying states that get identified through the joint process governing spot, forward rates and other conditioning information. Since both the persistence of the states and the covariance between the forecasts can vary across states, the optimal combination weights will depend both on the current state probabilities and on the forecast horizon. In real time, as the underlying state probabilities are updated based on the arrival of new information, the optimal forecast combination weights will also change. In summary, our approach exploits information in short-term spot and forward rates and in a variety of macroeconomic variables to compute combined forecasts that are optimal given the assumed (joint) data generating process.

When applied to data on 1-month US interest rates over the period 1950-2003, our analysis reveals a number of interesting findings. Forecast combinations that incorporate information beyond spot rates are found to produce better predictions than traditional benchmarks both at short and long forecast horizons. Indeed, our results suggest that it is important both to combine information embedded in different forecasts and to allow for nonlinear (regime) dynamics in spot and forward rates. Gains in forecast precision can be sizeable both in statistical and economic terms.

The paper's main contributions are the following. First, our paper extends the forecast combination literature by proposing a flexible combination approach that extends earlier work in the combination literature by Bates and Granger (1969), Diebold and Pauly (1987), Deutsch, Granger and Terasvirta (1994), Stock and Watson (2001, 2005) and Aiolfi and Timmermann (2006). Our approach can also be viewed as an extension to earlier approaches to forecast combinations under regime switching (Elliott and Timmermann (2005)).

Second, we develop a bivariate model that captures nonlinear dynamics in the joint process for spot and forward rates. The nonlinearities appear to be well captured through a regime switching model characterized by four states with very different levels of volatility and long-run (unconditional) interest rates.

Third, we conduct an analysis of the out-of-sample predictability of spot rates that compares a wide range of approaches and shows how different sources of information—e.g., past sport rates, forward rates and macroeconomic variables—may be helpful at both short and long forecast horizons. We show that restrictions from the expectations hypothesis, when imposed on the forecasts, can be used to improve the out-of-sample forecasting performance compared to models that do not make use of these restrictions. We also find that model combination provides a useful approach to improve forecast accuracy for interest rates.

The plan of the paper is as follows. Section 2 introduces the multivariate hidden Markov model used in the paper and derives analytical results on optimal forecast combinations. Section 3 presents econometric estimates for the US interest rate data and characterizes the optimal combination weights using these estimates. Section 4 conducts an out-of-sample forecasting experiment and section 5 concludes.

Other papers documenting non-linearities in short-term interest rates include Hamilton (1988), Sola and Driffill (1994), Gray (1996), Ang and Bekaert (2002) and Bansal, Tauchen, and Zhou (2004).

2. A Flexible Forecast Combination Model

A large literature has found empirical evidence that combining forecasts from different models generally leads to improved out-of-sample forecasting performance relative to a strategy of selecting the single best forecasting model.[2] The classic Bates-Granger (1969) forecast combination regression takes the form:

$$y_{t+1} = \mu_y + \mathbf{a}'_y \hat{\mathbf{y}}_{t,t+1} + \varepsilon_{t+1}, \tag{1}$$

where y_{t+1} is the predicted variable and $\hat{\mathbf{y}}'_{t,t+1}$ is a vector of forecasts of y_{t+1} computed using information available at time t. This regression assumes that the combination weights are stable through time. However, it is easy to think of situations where models that, on average, generate superior forecasting performance may be slower to adapt in some states of the world than other models that generate higher average loss. Similarly, forecasting models that are superior at short horizons, may fail to be so at medium or long horizons. If the parameters of the underlying data generating process are unstable and the instability is sufficiently idiosyncratic (model-specific), it is plausible that there can be gains from combining forecasts from different econometric specifications.

Our paper is concerned with forecast combinations in situations such as these. We use a multivariate regime switching process to capture the existence of common, discrete factors driving both the stochastic process of the variable of interest (the 1-month spot rate) and a related market variable (the 1-month forward rate) that can be construed as a predictor of the target variable. Although the regime switching model only provides a reduced form for the underlying joint process, it can accommodate time-varying parameters and differences between the conditional (short-term) and unconditional (long-term) moments of the data generating process.[3]

2.1. *The Statistical Model*

To this end consider the following joint stochastic process for $\mathbf{z}_{t+1} \equiv (y_{t+1} \ \hat{\mathbf{y}}'_{t+1,t+2})'$, where $\hat{\mathbf{y}}'_{t+1,t+2}$ is a vector of forecasts of y_{t+2} produced at time $t+1$ so $\mathbf{z}_{t+1}$ is adapted to the filtration $\mathcal{F}_{t+1}$:

$$\begin{pmatrix} y_{t+1} \\ \hat{\mathbf{y}}_{t+1,t+2} \end{pmatrix} = \begin{pmatrix} \mu^*_{y,S_{t+1}} \\ \boldsymbol{\mu}^*_{S_{t+1}} \end{pmatrix} + \sum_{j=1}^{p} \begin{pmatrix} \mathbf{a}^*_{y,S_{t+1}} \\ \mathbf{A}^*_{j,S_{t+1}} \end{pmatrix} \begin{pmatrix} y_{t+1-j} \\ \hat{\mathbf{y}}_{t+1-j,t+2-j} \end{pmatrix} + \boldsymbol{\varepsilon}_{t+1}, \quad \text{or}$$

$$\mathbf{z}_{t+1} = \boldsymbol{\mu}_{S_{t+1}} + \sum_{j=1}^{p} \mathbf{A}_{j,S_{t+1}} \mathbf{z}_{t+1-j} + \boldsymbol{\varepsilon}_{t+1}. \tag{2}$$

The discrete state variable S_{t+1} (which is not known at time t, i.e. $S_{t+1} \notin \mathcal{F}_t$) takes integer values between 1 and k, $\boldsymbol{\mu}_{S_{t+1}}$ is the intercept in state S_{t+1}, $\mathbf{A}_{j,S_{t+1}}$ is the VAR(j) matrix in state S_{t+1}, and $\boldsymbol{\varepsilon}_{t+1} \sim N(\mathbf{0}, \boldsymbol{\Omega}_{S_{t+1}})$ is the vector of innovations to $\mathbf{z}_{t+1}$ which has zero mean and state-specific covariance matrix $\boldsymbol{\Omega}_{S_{t+1}}$. To complete the model we assume that S_{t+1} is driven by a first order, homogeneous Markov process with constant transition probability matrix $\mathbf{P}$,

$$\mathbf{P}_{[i,j]} = \Pr(S_{t+1} = j | S_t = i) = p_{ij}, \quad i,j = 1,..,k. \tag{3}$$

---

[2] Surveys have been provided by Diebold and Lopez (1996) and Timmermann (2006).

[3] Rudebusch and Wu (2004) argue that regimes may occur in term structure data as a consequence of monetary policy shifts.

Both the realized values of the variable of interest, y_{t+1}, and the vector of one-step-ahead forecasts, $\hat{\mathbf{y}}_{t+1,t+2}$, allow us to extract information about the unobserved (realized) states $\{s_1, ..., s_{t+1}\}$. The model assumes that the realization and predictions, aligned in time so they are adapted to a common information set, are driven by a common state variable, S_{t+1}.

To maintain sufficient generality and to ensure that our approach nests methods in common use and also is practical to implement, it is convenient to distinguish between those forecasts in $\hat{\mathbf{y}}_{t,t+1} \equiv (\hat{\mathbf{y}}'_{1,t,t+1} \ \hat{\mathbf{y}}'_{2,t,t+1})'$ that affect the identification of the regimes (labelled $\hat{\mathbf{y}}_{1,t,t+1}$) and those that do not (i.e. traditional Bates-Granger forecasting variables, $\hat{\mathbf{y}}_{2,t,t+1}$). We shall assume n_1 and n_2 of these forecasting variables, respectively, so $n = n_1 + n_2$ is the total number of forecasts. The forecasts in $\hat{\mathbf{y}}'_{2,t,t+1}$ serve as benchmark 'testers', since the last n_2 elements of $\hat{\mathbf{y}}_{t,t+1}$ should not matter under the assumption that the first n_1 predictions are optimal. Using only the first n_1 variables to identify the regimes can be motivated on economic grounds $-$ i.e. the identified regimes are easier to interpret if based on particular variables $-$ and further serves the purpose of reducing the number of parameters to be estimated.

Our extension of the Bates-Granger framework can be motivated in several ways. First, it is unlikely that the best forecasting model remains the same through time or even that a particular model has a constant bias. By allowing for regimes with different degrees of persistence, the ranking of the models and their usefulness (weight) in a combined forecast will also change over time. Indeed, it seems reasonable to assign different weights to forecasts that embody different sources of information in accordance with the underlying state of the economy or the monetary policy regime. Secondly, the extension of the conditional model of Bates and Granger to a multivariate VAR setting means that forecasts at arbitrary horizons, $h \geq 1$, can be generated through forward iteration even if the underlying forecasts use a shorter horizon than desired. Third, the framework readily allows for combining of pure time-series forecasts (reflected in lagged values of y) and forecasts that use information from other sources such as macroeconomic models (reflected in $\hat{y}$).

Our model nests many interesting special cases. It can be viewed as a generalization of (1) which emerges as the first equation of (2) in the special case where $p = k = 1$, $\mathbf{A}_1^* = \mathbf{O}$, and the first element of $\mathbf{a}_y^*$ is zero, i.e. when there is a single state, only the current forecast matters and past values of both the predicted and actual variable are excluded. Furthermore, (2) provides a flexible parametric representation from which the first and second conditional moments of $\mathbf{z}_{t+1}$ can readily be derived. The model allows for conditional heteroskedasticity, skew and kurtosis in the forecast errors and can also facilitate persistent sources of bias.

Another interesting special case is a regime-switching, heteroskedastic version of Vasicek's affine model. To see this, let y_{t+1} and $\hat{y}_{t+1,t+2}$ be the 1-period spot and forward rate, respectively. Setting $p = 1$, $a_{12S_{t+1}} = 0$, and assuming no contemporaneous correlation between shocks to the spot and forward rates, the first equation of (2) implies

$$y_{t+1} = \mu_{S_{t+1}} + a_{11S_{t+1}} y_t + \varepsilon_{y,t+1}, \quad \varepsilon_{y,t+1} \sim N(0, \sigma_{S_{t+1}}^2),$$

Hence (2) generalizes classic, linear term structure models to incorporate regime shifts and allows for a two-factor specification in which the short rate is also affected by shocks to the forward rate.

Finally, when $k = 1$, equation (2) reduces to a simple, homoskedastic VAR(p) model

$$\begin{pmatrix} y_{t+1} \\ \hat{\mathbf{y}}_{t+1,t+2} \end{pmatrix} = \begin{pmatrix} \mu_y^* \\ \boldsymbol{\mu}^* \end{pmatrix} + \sum_{j=1}^p \mathbf{A}_j \begin{pmatrix} y_{t+1-j} \\ \hat{\mathbf{y}}_{t+1-j,t+2-j} \end{pmatrix} + \boldsymbol{\varepsilon}_{t+1}, \quad \boldsymbol{\varepsilon}_{t+1} \sim N(\mathbf{0}, \boldsymbol{\Omega}). \tag{4}$$

4

2.2. *Forecast Combinations*

Forecast combinations seek to choose the $n \times 1$ vector of weights $\boldsymbol{\omega}_{t,t+h}$ that minimize the average expected loss given current state probabilities, $\hat{\boldsymbol{\pi}}_t$, $E[L(e_{t+h})|\hat{\boldsymbol{\pi}}_t]$ – denoted $E_t[L(e_{t+h})]$ – from h−period forecast errors defined as:

$$e_{t+h} \equiv y_{t+h} - \omega_{0,t,t+h} - \boldsymbol{\omega}'_{t,t+h}\hat{\mathbf{y}}_{t,t+h}.$$

The constant $\omega_{0,t,t+h}$ adjusts for possible biases in the combined forecast $\boldsymbol{\omega}'_{t,t+h}\hat{\mathbf{y}}_{t,t+h}$. Because of the conditional regime switching dynamics, the combination weights will generally be a highly non-linear function of past data. More general combination schemes could be considered, but we shall not do so here.

Formally the forecast combination problem takes the form:

$$\min_{\omega_{0,t,t+h}, \, \boldsymbol{\omega}_{t,t+h}} \quad E_t[L(y_{t+h} - \omega_{0,t,t+h} - \boldsymbol{\omega}'_{t,t+h}\hat{\mathbf{y}}_{t,t+h})]$$

$$s.t. \; \boldsymbol{\omega}_{t,t+h} \in \mathcal{C},$$

where $\mathcal{C}$ is an admissible region for the weights. When it can be assumed that the individual forecasts are unbiased, it is common to restrict the combination weights to be non-negative, sum to unity and impose that the intercept is zero:

$$\mathcal{C} \equiv \mathsf{X}_{i=1}^n [0,1] \quad \sum_{i=1}^n \boldsymbol{\omega}_{t,t+h}[i] = 1 \quad \omega_{0,t,t+h} = 0,$$

where $\boldsymbol{\omega}[i]$ is the i-th element of $\boldsymbol{\omega}$. Such restrictions may lead to efficiency gains in the estimation of the combination weights.

Throughout the paper we follow common practice and assume that loss is quadratic, i.e. $L(e_{t+h}) = e_{t+h}^2$. Thus the objective is to minimize mean squared forecast error (MSFE) loss:

$$
\begin{aligned}
E_t[e_{t+h}^2] \;=\;& Var_t[y_{t+h}] + \boldsymbol{\omega}'_{t,t+h}Var_t[\hat{\mathbf{y}}_{t,t+h}]\boldsymbol{\omega}_{t,t+h} - 2\boldsymbol{\omega}'_{t,t+h}Cov_t[y_{t+h},\hat{\mathbf{y}}_{t,t+h}] + \\
& + \{E_t[y_{t+h}] - \omega_{0,t,t+h}\}^2 + \{\boldsymbol{\omega}'_{t,t+h}E_t[\hat{\mathbf{y}}_{t,t+h}]\}^2 - 2(E_t[y_{t+h}] - \omega_{0,t,t+h})\boldsymbol{\omega}'_{t,t+h}E_t[\hat{\mathbf{y}}_{t,t+h}]. \quad (5)
\end{aligned}
$$

It is instructive to consider each of the elements in (5). For concreteness, assume that $n_1 = n_2 = 1$, $h = 1$, $\hat{\mathbf{y}}_{1,t,t+1} = f_{t,1}$, the one-month forward rate, and $\hat{\mathbf{y}}_{2,t,t+1} = ay_t$ (a simple random walk benchmark when $a = 1$). Let $\hat{\boldsymbol{\pi}}_t = \hat{\boldsymbol{\pi}}(\ddot{\mathbf{z}}_{t-1})$ be the conditional state probabilities given the entire history of $\mathbf{z}$, $\ddot{\mathbf{z}}_{t-1} \equiv \{\mathbf{z}_{t-j}\}_{j=1}^t$ and define the 3×3 matrix $\mathbf{V}_{t,t+1}$ with generic element $v_{i,j}$:

$$
v_{i,j} = \begin{cases}
\mathbf{e}'_i \left[\sum_{s=1}^k (\hat{\boldsymbol{\pi}}'_t \mathbf{P}\mathbf{e}_s)\boldsymbol{\Omega}_s + \sum_{s=1}^k (\hat{\boldsymbol{\pi}}'_t \mathbf{e}_s)\mathbf{C}_s^{1,1} \right] \mathbf{e}_j & i,j \leq 2 \\
a\mathbf{e}'_1 \left[\sum_{s=1}^k (\hat{\boldsymbol{\pi}}'_t \mathbf{e}_s)\mathbf{C}_s^{0,1} \right] \mathbf{e}_j & i = 3, \; j \leq 2 \\
a\mathbf{e}'_i \left[\sum_{s=1}^k (\hat{\boldsymbol{\pi}}'_t \mathbf{e}_s)\mathbf{C}_s^{0,1} \right] \mathbf{e}_1 & j = 3, \; i \leq 2 \\
a^2\mathbf{e}'_1 \left[\sum_{s=1}^k (\hat{\boldsymbol{\pi}}'_t \mathbf{P}\mathbf{e}_s)\boldsymbol{\Omega}_s + \sum_{s=1}^k (\hat{\boldsymbol{\pi}}'_t \mathbf{e}_s)\mathbf{C}_s^{1,1} \right] \mathbf{e}_1 & i = j = 3
\end{cases}, \quad (6)
$$

where $\mathbf{e}_s$ denotes a column vector of suitable dimension with a 1 in the s-th position and zeros elsewhere so $\hat{\boldsymbol{\pi}}'_t \mathbf{P}\mathbf{e}_s$ selects the conditional probability of the s-th state at time $t + 1$. The relevant conditional moments, conditioned on the state probabilities $\hat{\boldsymbol{\pi}}_t$ and denoted by $E_t[\cdot|\hat{\boldsymbol{\pi}}_t], Var_t[\cdot|\hat{\boldsymbol{\pi}}_t]$ and $Cov_t[\cdot|\hat{\boldsymbol{\pi}}_t]$, respectively (or

$E_t[\cdot]$, $Var_t[\cdot]$ and $Cov_t[\cdot]$ for short), are:

$$E_t[y_{t+1}] = \sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{P}\mathbf{e}_s)\left[\mu_{ys}^* + \mathbf{e}_1'\sum_{j=1}^{p}\mathbf{A}_{j,s}E_t\left(\begin{array}{c} y_{t+1-j} \\ \hat{\mathbf{y}}_{t+1-j,t+2-j} \end{array}\right)\right]$$

$$Var_t[y_{t+1}] = \mathbf{e}_1'\mathbf{V}_{t,t+1}\mathbf{e}_1$$

$$Var_t[\hat{\mathbf{y}}_{t,t+1}] = \mathbf{Q}_{n_1+1}'\mathbf{V}_{t,t+1}\mathbf{Q}_{n_1+1}$$

$$Cov_t[y_{t+1}, \hat{\mathbf{y}}_{t,t+1}] = \mathbf{e}_1'\mathbf{V}_{t,t+1}\mathbf{Q}_{n_1+1}$$

$$E_t[\hat{\mathbf{y}}_{t,t+1}] = \left(\begin{array}{c} \sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{e}_s)\mathbf{Q}_{n_1}'\left(\boldsymbol{\mu}_s + \sum_{j=1}^{p}\mathbf{A}_jE_t[\mathbf{z}_{t+h-1-j}]\right)\mathbf{Q}_{n_1} \\ a\sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{e}_s)\mathbf{e}_1'\left(\boldsymbol{\mu}_s + \sum_{j=1}^{p}\mathbf{A}_jE_t[\mathbf{z}_{t+h-1-j}]\right)\mathbf{e}_1 \end{array}\right), \tag{7}$$

where

$$\mathbf{C}_s^{q,h} \equiv \left((\mathbf{M}' - \overline{\boldsymbol{\mu}}_{t+q} \otimes \boldsymbol{\iota}_k') + \boldsymbol{\Delta}_{t+q}\right)\mathbf{D}_s^q\mathbf{P}^{h-q}\left((\mathbf{M} - \boldsymbol{\iota}_k \otimes \overline{\boldsymbol{\mu}}_{t+h}') + \boldsymbol{\Delta}_{t+h}'\right) \quad h \geq q \geq 0$$

$$\boldsymbol{\Delta}_{t+h} \equiv \left(\begin{array}{ccc} \sum_{j=1}^{p}\mathbf{e}_1'(\mathbf{A}_{j,1} - \bar{\boldsymbol{\alpha}}_j^h)\mathbf{z}_{t+1-j} & \cdots & \sum_{j=1}^{p}\mathbf{e}_1'(\mathbf{A}_{j,k} - \bar{\boldsymbol{\alpha}}_j^h)\mathbf{z}_{t+1-j} \\ \vdots & & \vdots \\ \sum_{j=1}^{p}\mathbf{e}_n'(\mathbf{A}_{j,1} - \bar{\boldsymbol{\alpha}}_j^h)\mathbf{z}_{t+1-j} & \cdots & \sum_{j=1}^{p}\mathbf{e}_n'(\mathbf{A}_{j,k} - \bar{\boldsymbol{\alpha}}_j^h)\mathbf{z}_{t+1-j} \end{array}\right)$$

$$\mathbf{D}_s^h \equiv diag\{\mathbf{e}_s'\mathbf{P}^h\mathbf{e}_1, \mathbf{e}_s'\mathbf{P}^h\mathbf{e}_2, ..., \mathbf{e}_s'\mathbf{P}^h\mathbf{e}_k\} \qquad \bar{\boldsymbol{\alpha}}_j^h \equiv (\hat{\boldsymbol{\pi}}_t' \otimes \mathbf{I}_n)(\mathbf{P}^h \otimes \mathbf{I}_n)\mathbf{A}_j$$

$$\mathbf{M} \equiv \left(\begin{array}{cccc} \boldsymbol{\mu}_1 & \boldsymbol{\mu}_2 & \cdots & \boldsymbol{\mu}_k \end{array}\right)', \qquad \overline{\boldsymbol{\mu}}_{t+h}' \equiv \hat{\boldsymbol{\pi}}_t'\mathbf{P}^h\mathbf{M}$$

$$\mathbf{A}_j \equiv \left(\begin{array}{cccc} \mathbf{A}_{j,1} & \mathbf{A}_{j,2} & \cdots & \mathbf{A}_{j,k} \end{array}\right)', \qquad \mathbf{Q}_n \equiv \left[\begin{array}{cccc} 0 & 0 & \cdots & 0 \\ & \mathbf{I}_n & & \end{array}\right] \tag{8}$$

and, for any matrix $\mathbf{B}$, $\mathbf{B}^q \equiv \prod_{i=1}^{q}\mathbf{B}$ and $\mathbf{B}^0 = \mathbf{I}$. The matrix $\mathbf{C}_s^{1,1}$ collects variance terms from discrete shifts in the conditional mean parameters, i.e. differences between the period $t+1$ regime-specific intercepts ($\boldsymbol{\mu}_s$ for $s = 1, ..., k$, the rows of $\mathbf{M}$), VAR(j) coefficients ($\mathbf{A}_{j,s}$ for $s = 1, ..., k$) and their predicted values at time $t+1$, $\overline{\boldsymbol{\mu}}_{t+1}$ and $\{\bar{\boldsymbol{\alpha}}_j^1\}_{j=1}^{p}$.[4] Similarly, the matrix $\mathbf{C}_s^{0,1}$ collects auto-covariance terms from discrete shifts in the conditional mean parameters.[5] The diagonal matrix $\mathbf{D}_s^h$ attaches probability weights to possible shifts in the mean parameters at time $t+h$.

---

[4] For instance, assuming that $k = 2$ and $p = 1$, we have

$$\overline{\boldsymbol{\mu}}_{t+i} \equiv \mathbf{M}'\mathbf{P}'\boldsymbol{\pi}_t = \left[\begin{array}{cc} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{array}\right]\left[\begin{array}{cc} p_{11} & p_{21} \\ p_{12} & p_{22} \end{array}\right]\left[\begin{array}{c} \pi_{1t} \\ \pi_{2t} \end{array}\right] = \left[\begin{array}{c} \mu_{11}(p_{11}\pi_{1t} + p_{21}\pi_{2t}) + \mu_{12}(p_{12}\pi_{1t} + p_{22}\pi_{2t}) \\ \mu_{21}(p_{11}\pi_{1t} + p_{21}\pi_{2t}) + \mu_{22}(p_{12}\pi_{1t} + p_{22}\pi_{2t}) \end{array}\right],$$

$$\bar{\boldsymbol{\alpha}}_j \equiv (\hat{\boldsymbol{\pi}}_t' \otimes \mathbf{I}_2)(\mathbf{P} \otimes \mathbf{I}_2)\mathbf{A}_j = \left[\begin{array}{cccc} \pi_{1t} & 0 & \pi_{2t} & 0 \\ 0 & \pi_{1t} & 0 & \pi_{2t} \end{array}\right]\left[\begin{array}{cccc} p_{11} & 0 & p_{12} & 0 \\ 0 & p_{11} & 0 & p_{12} \\ p_{21} & 0 & p_{22} & 0 \\ 0 & p_{21} & 0 & p_{22} \end{array}\right]\left[\begin{array}{c} \mathbf{A}_{1,1} \\ \mathbf{A}_{1,2} \end{array}\right]$$

$$= \left[\begin{array}{cc} (p_{11}\pi_{1t} + p_{21}\pi_{2t})a_{11,1} + (p_{12}\pi_{1t} + p_{22}\pi_{2t})a_{11,2} & (p_{11}\pi_{1t} + p_{21}\pi_{2t})a_{12,1} + (p_{12}\pi_{1t} + p_{22}\pi_{2t})a_{12,2} \\ (p_{11}\pi_{1t} + p_{21}\pi_{2t})a_{21,1} + (p_{12}\pi_{1t} + p_{22}\pi_{2t})a_{21,2} & (p_{11}\pi_{1t} + p_{21}\pi_{2t})a_{22,1} + (p_{12}\pi_{1t} + p_{22}\pi_{2t})a_{221,2} \end{array}\right],$$

where $a_{ij,1}$ is the $[i,j]$ element of $\mathbf{A}_{1,1}$.

[5] In general, $Cov(\mathbf{y}_t, \mathbf{y}_{t+h}|\hat{\boldsymbol{\pi}}_t) = \sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{P}\mathbf{e}_s)Cov(\mathbf{y}_t, \mathbf{y}_{t+h}|S_t = s)$, where

$$Cov(\mathbf{y}_t, \mathbf{y}_{t+h}|S_t = s) = \left((\mathbf{M}' - \overline{\boldsymbol{\mu}}_t \otimes \boldsymbol{\iota}_k') + \boldsymbol{\Delta}_t\right)\mathbf{D}_s^h\left((\mathbf{M} - \boldsymbol{\iota}_k \otimes \overline{\boldsymbol{\mu}}_{t+h}') + \boldsymbol{\Delta}_{t+h}'\right).$$

Using (7)-(8), the combination weights minimizing (5) emerge from the necessary and sufficient conditions for a minimum that take the form of a linear system in three equations and three unknowns:

$$\hat{\omega}_{0,t,t+1} - E_t[y_{t+1}] + \hat{\boldsymbol{\omega}}'_{t,t+1}E_t[\hat{\mathbf{y}}_{t,t+1}] = 0 \quad (9)$$

$$\hat{\boldsymbol{\omega}}'_{t,t+1}(\mathbf{Q}'_n\mathbf{V}_{t,t+1}\mathbf{Q}_n) - E_t[y_{t+1}](E_t[\hat{\mathbf{y}}_{t,t+1}])' + \hat{\omega}_{0,t,t+1}(E_t[\hat{\mathbf{y}}_{t,t+1}])' + \hat{\boldsymbol{\omega}}'_{t,t+1}E_t[\hat{\mathbf{y}}_{t,t+1}] - (\mathbf{e}'_1\mathbf{V}_{t,t+1}\mathbf{Q}_n) = \mathbf{0}.$$

Clearly $\hat{\omega}_{0,t,t+1}$ corrects for any biases in the combined forecast. The combination weights on $\hat{\mathbf{y}}_{1,t,t+1}$ and the tester $\hat{\mathbf{y}}_{2,t,t+1}$ depend not only on the usual covariance terms, but also reflect the means of the forecasts within and across each state (through $\mathbf{C}_s^{1,1}$ and $\mathbf{C}_s^{0,1}$) as well as the current state probabilities, $\hat{\boldsymbol{\pi}}_t$, the state transition probabilities, $\mathbf{P}$, and differences in the covariance terms across states. Compared to the standard forecast combination problem, this yields a richer optimization program that captures the idea of including a forecasting model in the combination because it acts as a hedge against model misspecification which takes the form of discrete shifts in the data generating process.

These results are naturally generalized in two ways. First, the tester forecast, $\hat{\mathbf{y}}_{2,t,t+1}$, may differ from ay_t. Allowing for this, the conditional MSFE is given by (dropping the second (horizon) subscripts from $\boldsymbol{\omega}_t$)

$$
\begin{aligned}
E_t[e_{t+1}^2] \;=\; & \mathbf{e}'_1\mathbf{V}_{t,t+1}\mathbf{e}_1 + \{E_t[y_{t+1}]\}^2 + (\omega_{0,t})^2 - 2\omega_{0,t}E_t[y_{t+1}] \\
& + (\boldsymbol{\omega}'_t E_t[\hat{\mathbf{y}}_{t,t+1}])^2 - 2E_t[y_{t+1}]\boldsymbol{\omega}'_t E_t[\hat{\mathbf{y}}_{t,t+1}] + 2\omega_{0,t}\boldsymbol{\omega}'_t E_t[\hat{\mathbf{y}}_{t,t+1}] \\
& + \boldsymbol{\omega}'_t(\mathbf{Q}'_n\mathbf{V}_{t,t+1}\mathbf{Q}_n)\boldsymbol{\omega}_t - 2\boldsymbol{\omega}'_t(\mathbf{e}'_1\mathbf{V}_{t,t+1}\mathbf{Q}_n),
\end{aligned}
\quad (10)
$$

where $\mathbf{V}_{t,t+1}$ generalizes (6).

When producing h-step forecasts for $h \geq 2$, equation (5) still describes the conditional MSFE, but the expression for the relevant moments are different. Forecasts are best understood by rewriting (2) as (assuming $p > h$)

$$\mathbf{z}_{t+h} = \boldsymbol{\mu}_{s_{t+h}} + \sum_{j=1}^{h-1}\mathbf{A}_{s_{t+h,j}}\mathbf{z}_{t+h-j} + \sum_{j=h}^{p}\mathbf{A}_{s_{t+h,j}}\mathbf{z}_{t+h-j} + \boldsymbol{\varepsilon}_{t+h},$$

and using that $\mathbf{z}_{t+h-j} \in \mathcal{F}_t$ for $j \geq h$. Only $\mathbf{z}_{t+1}, .., \mathbf{z}_{t+h}$ need to be predicted. Still, if the autoregressive matrices are state-dependent, the recursive expressions are generally path dependent. This can most easily be seen when $p = 1$, in which case we obtain the following expression:

$$\mathbf{z}_{t+h} = \sum_{j=0}^{h-1}\left(\prod_{i=1}^{j}\mathbf{A}_{s_{t+h+1-i}}\right)(\boldsymbol{\mu}_{s_{t+h-j}} + \boldsymbol{\varepsilon}_{t+h-j}) + \left(\prod_{i=0}^{h-1}\mathbf{A}_{s_{t+h-i}}\right)\mathbf{z}_t,$$

where $\prod_{i=1}^{0}\mathbf{A}_{s_{t+h+1-i}} \equiv \mathbf{I}_{n_1+1}$. This expression is most easily evaluated by numerical simulation (Granger and Terasvirta (1993)), which is not a problem since the Markovian form of the model makes it ideally suited for this type of method.

When the matrix of VAR coefficients, $\mathbf{A}$, is not state dependent, the multi-step forecasts simplify considerably. This is an important case to consider since the empirical analysis in the next section finds no evidence of such state dependence. The task is again easier when $n_2 = 1$, $\hat{\mathbf{y}}_{t,t+h}[n] = a^h y_t$, so that (again conditioning on $\hat{\boldsymbol{\pi}}_t$)

$$E_t[y_{t+h}] = \sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}'_t\mathbf{P}^h\mathbf{e}_k)\left[\mu_{y,s}^* + \mathbf{e}'_1\sum_{j=1}^{p}\mathbf{A}_j E_t[\mathbf{z}_{t+h-j}]\right]$$

7

$$
\begin{aligned}
Var_t[y_{t+h}] &= \mathbf{e}_1'\mathbf{V}_{t,t+h}\mathbf{e}_1 \\
Var_t[\hat{\mathbf{y}}_{t,t+h}] &= \mathbf{Q}_{n_1+1}'\mathbf{V}_{t,t+h}\mathbf{Q}_{n_1+1} \\
Cov_t[y_{t+h},\hat{\mathbf{y}}_{t,t+h}] &= \mathbf{Q}_{n_1+1}'\mathbf{V}_{t,t+h}\mathbf{e}_1 \\
E_t[\hat{\mathbf{y}}_{t,t+h}] &= \begin{pmatrix} \sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{P}^{h-1}\mathbf{e}_s)\mathbf{Q}_{n_1}'\left(\boldsymbol{\mu}_s + \sum_{j=1}^{p}\mathbf{A}_j E_t[\mathbf{z}_{t+h-1-j}]\right)\mathbf{Q}_{n_1} \\ a^h \sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{P}^{h-1}\mathbf{e}_s)\mathbf{e}_1'\left(\boldsymbol{\mu}_s + \sum_{j=1}^{p}\mathbf{A}_j E_t[\mathbf{z}_{t+h-1-j}]\right)\mathbf{e}_1 \end{pmatrix},
\end{aligned} \tag{11}
$$

where $\mathbf{V}_{t,t+h}$ now has generic element $v_{i,j}$ defined as

$$
v_{i,j} = \begin{cases}
\mathbf{e}_i'\left[\sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{P}^h\mathbf{e}_s)\boldsymbol{\Omega}_s + \sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{e}_s)\mathbf{C}_s^{h,h}\right]\mathbf{e}_j & i,j \leq n_1 \\
a^h\mathbf{e}_1'\left[\sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{e}_s)\mathbf{C}_s^{0,h}\right]\mathbf{e}_j & i = n_1+1,\ j \leq n_1 \\
a^h\mathbf{e}_i'\left[\sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{e}_s)\mathbf{C}_s^{0,h}\right]\mathbf{e}_1 & j = n_1+1,\ i \leq n_1 \\
a^{2h}\mathbf{e}_1'\left[\sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{P}^h\mathbf{e}_s)\boldsymbol{\Omega}_s + \sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{e}_s)\mathbf{C}_s^{h,h}\right]\mathbf{e}_1 & i = j = n_1+1 = n
\end{cases}, \tag{12}
$$

and $E_t[\mathbf{z}_{t+i-j}]$ follows a recursive structure:

$$
E_t[\mathbf{z}_{t+i-j}] = \begin{cases}
\mathbf{z}_{t+i-j} & \text{if } i \leq j \\
\sum_{s=1}^{k}(\hat{\boldsymbol{\pi}}_t'\mathbf{P}^{i-j}\mathbf{e}_k)\left[\boldsymbol{\mu}_s + \sum_{j=1}^{p}\mathbf{A}_{j,s}E_t[\mathbf{z}_{t+i-j-1}]\right] & \text{if } i > j
\end{cases}.
$$

The $\mathbf{C}_s^{h,h}$ matrix represents the contribution to the variance of $\mathbf{z}_{t+h}$ arising from the possibility of switches in the conditional mean parameters between periods t and $t+h$, while $\mathbf{C}_s^{0,h}$ collects $h-$step auto-covariances.

The objective function in the minimization program now becomes

$$
\begin{aligned}
E_t[e_{t+h}^2] &= \mathbf{e}_1'\mathbf{V}_{t,t+h}\mathbf{e}_1 + E_t[y_{t+h}]^2 + \omega_{0,t,t+h}^2 - 2\omega_{0,t,t+h}E_t[y_{t+h}] + (\boldsymbol{\omega}_{t,t+h}'E_t[\hat{\mathbf{y}}_{t,t+h}])^2 - 2E_t[y_{t+h}]\boldsymbol{\omega}_{t,t+h}'E_t[\hat{\mathbf{y}}_{t,t+h}] \\
&\quad + 2\omega_{0,t,t+h}\boldsymbol{\omega}_{t,t+h}'E_t[\hat{\mathbf{y}}_{t,t+h}] + \boldsymbol{\omega}_{t,t+h}'(\mathbf{Q}_{n_1+1}'\mathbf{V}_{t,t+h}\mathbf{Q}_{n_1+1})\boldsymbol{\omega}_{t,t+h} - 2\boldsymbol{\omega}_{t,t+h}'(\mathbf{Q}_{n_1+1}'\mathbf{V}_{t,t+h}\mathbf{e}_1),
\end{aligned}
$$

with first order conditions

$$
\begin{aligned}
\hat{\omega}_{0,t,t+h} &= E_t[y_{t+h}] - \hat{\boldsymbol{\omega}}_{t,t+h}'E_t[\hat{\mathbf{y}}_{t,t+h}], \\
\mathbf{0} &= (\mathbf{Q}_{n_1+1}'\mathbf{V}_{t,t+h}\mathbf{Q}_{n_1+1})\hat{\boldsymbol{\omega}}_{t,t+h} - E_t[y_{t+h}]E_t[\hat{\mathbf{y}}_{t,t+h}] + \hat{\omega}_{0,t,t+h}E_t[\hat{\mathbf{y}}_{t,t+h}] + \\
&\quad (\hat{\boldsymbol{\omega}}_{t,t+h}'E_t[\hat{\mathbf{y}}_{t,t+h}])E_t[\hat{\mathbf{y}}_{t,t+h}] - \mathbf{Q}_{n_1+1}'\mathbf{V}_{t,t+h}\mathbf{e}_1.
\end{aligned} \tag{13}
$$

3. Application to US Interest Rates

Having introduced the regime switching model and characterized the solution to the forecast combination problem, we next provide estimation results for this model applied to US spot and forward rates.

3.1. Data

We use the US Treasury Database from the University of Chicago Center for Research in Security Prices (CRSP). The sample period is January 1950 - December 2003, a total of 648 monthly observations. In addition to one-month T-bill rates, y_t, we collect data on the one-month forward yields, $f_{t,\varphi}$, implied by the term structure of spot rates at time t for the period starting at time $t+\varphi$, $\varphi = 0,...,5,11$. All rates are extracted from the CRSP 6- and 12-month files. The continuously compounded yields and forward rates are

standardized to a 30.4 day basis and calculated according to the formula $y_t \equiv \left[\ln \left(\frac{100}{P_{t,1}} \right) \right] \left(\frac{30.4}{\tau_{t,1}} \right)$, where $P_{t,1}$ is the average of the bid and ask prices for a one-month T-bill and $\tau_{t,1}$ is the time to expiration in calendar days.

Forward rates are calculated using the methodology in Fama (1984). At time t, the 1-month forward rate from period $t + \varphi$ to period $t + \varphi + 1$ is computed from the $\varphi + 1$ and φ-period spot yields, $y_{t:t+\varphi+1}$ and $y_{t:t+\varphi}$ as $f_{t,\varphi} = (\varphi + 1)y_{t:t+\varphi+1} - \varphi y_{t:t+\varphi}$.

3.2. *Econometric Estimates*

Following the analysis in Section 2, we set $n_1 = 1$ and model the bivariate system composed of the current 1-month spot and forward rates, i.e. $\mathbf{z}_t \equiv (y_t \ f_{t,1})'$. Even though theoretical term structure models can be used to constrain equation (2) (see e.g. Sola and Driffill (1994)), we adopt an unconstrained estimation strategy similar to Ang and Bekaert (2002) and instead test restrictions on the forecast combination weights.[6]

To select the number of states we considered values of the Schwartz and Hannan-Quinn information criteria for a range of models with different values for p and k. Both criteria suggested that a four-state VAR(1) specification is required to fit the data. While the covariance matrix varies strongly across states, the matrix of VAR(1) coefficients, $\mathbf{A}$, did not appear to be state-dependent. Hence we simplify the model by imposing that $\mathbf{A}_s = \mathbf{A}$ across regimes.

Panel A of Table 1 presents parameter estimates for the single-state VAR(1) model. Most of the estimates are statistically significant. Shocks to forward rates are slightly less volatile than those to spot rates and the simultaneous correlation between these shocks is quite high (0.88). The process is highly persistent with moduli of the eigenvalues of the estimated VAR matrix of 0.98 and 0.12.

Panel B presents maximum likelihood estimates for the four state model. The VAR(1) matrix in panel B is similar to the one reported in panel A with a largest modulus of 0.97. The correlation between shocks to the spot and forward rates does not seem to depend on the state and falls in the narrow interval [0.89, 0.92] under regime switching. Volatility levels and unconditional means differ greatly across the states, however, and therefore effectively identify the four regimes.[7]

Figure 1 plots the smoothed state probabilities. State 1 is associated with low and stable interest rates – the implied unconditional annualized means are 2.23 and 2.21% – as appeared during the early 1950s, part of the 1960s and, more recently, during the 2001-2003 recession. Although this stable state has an ergodic probability of only 0.12, it has an average duration of seven months.

State 2 identifies a regime with intermediate but stable interest rates: annualized means are 5.31 and

---

[6]Similarly, Diebold, Rudebusch, and Aruoba (2006) do not impose no-arbitrage restrictions. Ang and Piazzesi (2003) find that such restrictions at the estimation stage may improve forecasting performance.

[7]We tested for (regime-switching) ARCH effects using a model of the form

$$
\begin{pmatrix} y_t \\ f_{t,1} \end{pmatrix} = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_{j,S_t} \mathbf{z}_{t-j} + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{S_t})
$$

$$
\boldsymbol{\Sigma}_{S_t} = \mathbf{K}_{S_t} + \boldsymbol{\Delta}_{S_t} \boldsymbol{\varepsilon}_t' \boldsymbol{\varepsilon}_t \boldsymbol{\Delta}_{S_t}',
$$

where $\mathbf{K}_{S_t}$ is symmetric and positive definite and $\boldsymbol{\Delta}_{S_t}$ captures regime-dependent effects of past shocks on current volatility. Most coefficients failed to be significant and a likelihood ratio test of the restriction $\boldsymbol{\Delta}_s = \boldsymbol{\Delta}$, $s = 1, 2, 3, 4$ failed to reject the null of no ARCH effects.

5.14 percent, while the regime-specific (unconditional) monthly volatilities are only at three-quarters of their average levels. This state covered long spans of time such as 1953-1960, 1975-1978 and most of the 1990s, in total almost half (42%) of our 54-year long sample. However, the duration of this regime is only six months. This is explained by the tendency of the economy to frequently switch between regimes 2 and 3.

State 3 is associated with higher and relatively volatile interest rates. The mean interest rate in this state is above 10 percent and the volatility is close to its full-sample average. This regime occurs relatively frequently, with an average duration of five and a half months and a long-run probability of 0.36. Most of the 1970s and mid-1980s are best characterized by this regime which re-appears between 1999 and 2000, when the FED pulled the break on the US economy and interest rates gradually increased.

Finally, state four captures volatile market conditions. The unconditional mean implied by the estimated parameters for this state is very high as is the unconditional volatility. At a first glance, this state seems to be mostly dominated by the 1979-1982 'monetarist experiment' which is almost entirely captured by this state. Notice, however, that state four does much more than identify two structural breaks in short-term interest rates and is associated with short bursts of volatility in interest rates, as occurred during the Fall of 1984, during the FED contraction after October 1987, and during a few episodes in 1988 and 1989.

The last panel of Table 1 presents estimates of the transition probability matrix. All four states are mildly persistent with probabilities of staying in each state that vary between 0.81 and 0.84. Exits from the first regime are mostly to the second regime. From the second state it is possible to switch to both more volatile and higher interest rates (state 3), or back to state 1. However, the probability of a direct shift to state 4 is very small. From state 3, the economy can revert to lower and more stable interest rates (state 2) and there is also some chance of a switch to the turbulent market conditions associated with state 4. Finally, from state 4 the economy can only switch back to state 3. The transition matrix makes it possible for the economy to cycle for long periods between states 3 and 4−states with above-average and volatile interest rates−as occurred between 1978 and 1985.

Our analysis applies (2) to the level of US interest rates, rather than to their changes. This is an issue since a unit root is not rejected for the one-month spot rate: An augmented Dickey-Fuller (ADF) test produced a $p-$value of 0.14. However, the considerable persistence in interest rates may in part be due to shifts in the conditional mean parameters which can induce more persistent behavior. To investigate if this is indeed the case for our data, we simulated 1,000 time series of interest rates using the estimates of the four-state regime switching model from Table 1, assuming ergodic state probabilities for the initial observation and using the BIC for lag length selection. Figure 2 shows the histogram of the resulting $p-$values associated with the ADF statistic. 46% of the simulations produced a smaller ADF statistic (i.e. a higher $p-$value) than that observed in the data and hence weaker evidence against the null of a unit root. This supports our decision to model interest rate levels by means of a persistent Markov switching process.[8]

---

[8]When intercepts are non-zero, a regime switching specification for changes in interest rates implies persistent trends in the levels of interest rates. This is unlikely to be a plausible specification empirically. See also the discussion in Diebold and Kilian (2000) of the importance for forecasting performance of correctly identifying the order of integration.

3.3. *Combination Weights*

We next turn to three different but related issues. First, do combination weights depend on the underlying state probabilities? Second, do combination weights depend on the forecast horizon? Third, how big is the potential reduction in the expected loss from using forecast combinations rather than forecasts from the individual models?

To simplify the presentation, we start by considering as the testers in $\hat{\mathbf{y}}_{2,t,t+h}$ either (i) a random walk, $\hat{y}_{2,t,t+h} = y_t$ or (ii) $AR(p)$ forecasts $\hat{y}_{2,t,t+h} = \hat{a}_0 + \sum_{j=1}^{p} \hat{a}_j y_{t-j}$, where the coefficients $\{\hat{a}_j\}_{j=0}^{p}$ are estimated by OLS. The random walk is a classical benchmark in the interest rate literature. Duffee (2002) shows that this model is difficult to beat even for relatively flexible and widely used affine models.

Using the full-sample estimates from Table 1, Figure 3 plots optimal combination weights as a function of the forecast horizon for different values of the initial state probability, $\hat{\boldsymbol{\pi}}_t$. The plots assume that the initial states are known but that future states are not and so do not correspond to any one particular point in time in the sample. The figure assumes that $\hat{y}_{2,t,t+h}$ is obtained from an AR(4) model (selected by the BIC) but very similar results follow for the random walk benchmark and are omitted to save space. The combination weights are strongly dependent both on the state probabilities and on the forecast horizon. At short horizons the weights can be large in absolute value and can either have positive or negative signs depending on the initial state. States for which $\hat{\boldsymbol{\omega}}_{t,t+h}[1] >> 0$ are also characterized by $\hat{\boldsymbol{\omega}}_{t,t+h}[2] << 0$, and vice versa. Assuming a short horizon, states 1 and 3 assign large positive weights to the forward rate forecast and a large negative weight to the random walk forecasts. In state 2 the opposite happens as the weight on the time-series forecast is large and positive while the weight on the forward rate is negative.[9] As h grows the weight on the forward rate, $\hat{\boldsymbol{\omega}}_{t,t+h}[1]$, converges to unity, while the weight on the time-series forecast, $\hat{\boldsymbol{\omega}}_{t,t+h}[2]$, goes to zero. This means that the optimal weight on the forward rate, $\hat{\boldsymbol{\omega}}_{t,t+h}[1]$, is downward sloping when starting from state 1 and 3, while it is upward sloping when starting from state 2 or 4. We conclude from these findings that forward rates are particularly important for short-term forecasting of interest rates in regimes 1 and 3. Conversely, more backward-looking time-series forecasts seem to perform well in states 2 and 4.

Figure 3 also plots optimal weights when the initial state probabilities are set at their ergodic values – a scenario with high uncertainty about the current regime. At short horizons weights of 1.4 and -0.4 are obtained for the forward rate and AR(4) forecasts, respectively. As h grows, these weights approach unity and zero, respectively.

These findings have implications for tests of the expectations hypothesis (EH). Under the EH, long-term spot yields are given by an arithmetic average of one-month expected spot rates and future term premia for the different maturities. Forward rates, corrected for a term premium, should therefore be unbiased predictors of the future spot rate. In the regression

$$y_{t+h} = \alpha - T_h + \beta f_{t+h-1,1} + \boldsymbol{\gamma}' \hat{\mathbf{y}}_{2,t,t+1} + u_{t+h},$$

this implies that $\hat{\beta} = 1$ (unbiasedness), $\hat{\gamma} = 0$ (efficiency), while $\hat{\alpha}$ should provide an estimate of the risk premium. In short, only forward rates should be able to forecast future spot rates. Our forecast combination

---

[9]Assigning a negative weight to a forecast does not make it useless but – on the contrary – makes it potentially useful for minimizing expected loss through its covariance with other forecasts.

results reveal that for most configurations of the initial state probabilities, the EH is strongly rejected at short horizons since, for small h, $\hat{\boldsymbol{\omega}}_{t,t+h}[1] \neq 1$ and $\hat{\boldsymbol{\omega}}_{t,t+h}[2] \neq 0$, so there are advantages from using a combination of forward rates and time-series forecasts.[10] However, as h grows the restrictions implied by the EH become useful for forecasting purposes.

The fact that the EH is rejected at short forecast horizons is also clear from Figure 4 which compares the (in-sample) expected loss under the optimal forecast combination against the expected loss under the separate forecasts from the forward rate or AR(4) model. In-sample gains from using optimal forecasting methods can be viewed as a necessary but not sufficient condition for potential out-of-sample forecasting gains in the sense that if the in-sample gains are very small, we should not expect to find out-of-sample forecast improvements. It is therefore informative to study such in-sample loss differentials. Differences in expected loss are quite large, especially when comparing the optimal combination to the AR(4) model. The percentage decline in expected loss (relative to the benchmark) obtained by going from the AR(4) forecasts to a combination of time-series and forward rate forecasts exceeds 40% when $h \geq 4$, but is typically more modest at shorter horizons. Reductions in losses are smaller—between 5 and 15%—against the forward rate forecasts.

4. **Out-of-sample Forecasting Performance**

The analysis has so far demonstrated potentially large gains from forecast combinations that account for the sensitivity of the combination weights to the underlying state. However, there is no guarantee that such gains are empirically achievable since the results assumed that the regime switching model was correctly specified. To tackle this issue, we next conduct an out-of-sample forecasting exercise. Both individual forecasts and combinations of forward rates and testers are compared to a wide array of benchmarks from the literature.[11]

We proceed as follows. For each model, we obtain recursive parameter estimates over expanding samples starting with 1950:01 - 1980:01, 1950:01 - 1980:02, up to 1950:01 - 2003:12-h, where h is the forecast horizon. When $h = 1$, this gives a sequence of 287 sets of parameter estimates for each of the models. Only information available at the date when the forecast is formed is used. We refer to $\hat{y}_{t,t+h}^{(m)}$ as the $h-$step forecasts generated by model m and evaluate the accuracy of the forecasts through the forecast errors $e_{t,t+h}^{(m)} \equiv y_{t+h} - \hat{y}_{t,t+h}^{(m)}$ and the associated RMSFE and bias computed as

$$RMSFE_h^{(m)} \equiv 1200 \times \sqrt{\frac{1}{288 - h} \sum_{t=1980:01}^{2003:(12-h)} \left(y_{t+h} - \hat{y}_{t,t+h}^{(m)} \right)^2},$$

$$Bias_h^{(m)} \equiv 1200 \times \frac{1}{288 - h} \sum_{t=1980:01}^{2003:(12-h)} \left(y_{t+h} - \hat{y}_{t,t+h}^{(m)} \right).$$

We analyze the performance of several versions of the VAR regime switching model (2) fitted to the spot and forward rates measured at time t, $\mathbf{z}_t \equiv (y_t \; f_{t,1})'$. Therefore $n_1 = 1$ and $\hat{y}_{1,t,t+h}$ corresponds to some function of the (predicted) values of $\mathbf{z}_{t+h}$ specified below. Three alternative tester functions are used

---

[10]Empirical findings have generally been unfavorable to the EH, see e.g., Campbell and Shiller (1991).

[11]See also Egorov, Hong and Li (2005) for comprehensive evidence on the out-of-sample performance of affine term structure models.

$(n_2 = 1)$, namely the random walk (RW), $\hat{y}_{2,t,t+h} = y_t$, a recursively estimated AR(4) model (AR), $\hat{y}_{2,t,t+h} = \hat{a}_0 + \sum_{j=1}^{4} \hat{a}_j \hat{y}_{2,t,t+h-j}$, or forecasts from macroeconomic variables described further below (M):

1. The first method computes forecasts by using the optimal combination weights on the tester and the conditional forecast of the 1-period forward rate between time $t + h - 1$ and $t + h$, $E_t[f_{t+h-1,1}]$: $\hat{y}_{t,t+h}^{(1)} = \hat{\omega}_{0,t,t+h} + \hat{\boldsymbol{\omega}}_{t,t+h}[1]E_t[\mathbf{z}_{t+h-1}'\mathbf{e}_2] + \hat{\boldsymbol{\omega}}_{t,t+h}[2]\hat{y}_{2,t,t+h}$. Optimal combination weights are found using forecasts from either the random walk, AR or a macroeconomic model as testers, $\hat{y}_{2,t,t+h}$.

2. The second method is analogous to the first but restricts the combination weights so that $\hat{\omega}_{t,t+h}[1]$, $\hat{\omega}_{t,t+h}[2] \in [0,1]$ and $\hat{\boldsymbol{\omega}}_{t,t+h}'\boldsymbol{\iota}_2 = 1$: $\hat{y}_{t,t+h}^{(2)} = \hat{\omega}_{0,t,t+h} + \hat{\boldsymbol{\omega}}_{t,t+h}[1]E_t[\mathbf{z}_{t+h-1}'\mathbf{e}_2] + (1 - \hat{\boldsymbol{\omega}}_{t,t+h}[1])\hat{y}_{2,t,t+h}$.

3. The third method imposes the restriction $\hat{\omega}_{t,t+h}[1] = 1$ and $\hat{\omega}_{0,t,t+h} = \hat{\boldsymbol{\omega}}_{t,t+h}[2] = 0$, so $\hat{y}_{t,t+h}^{(3)} = E_t[\mathbf{z}_{t+h-1}'\mathbf{e}_2]$, the conditional forecast of the 1-step forward rate that will apply between time $t + h - 1$ and $t + h$. This is an iterated version of the expectations hypothesis.

4. The fourth method sets $\hat{y}_{t,t+h}^{(4)} = E_t[y_{t+h}] = E_t[\mathbf{z}_{t+h}'\mathbf{e}_1]$, i.e. the conditional forecast of the future spot rate at $t + h$. This forecast ignores the direct contribution of forward rates.

5. - 8. Methods five through eight are the single-state ($k = 1$) VAR versions of methods 1 - 4. For instance, method 5 combines the tester (i.e. forecasts from the random walk, AR(4) or macroeconomic model) and the conditional forecast of the 1-period forward rate between time $t + h - 1$ and $t + h$ when $\mathbf{z}_t$ follows the VAR(1) in (4).

9. Method 9 adopts the 'pure' expectations hypothesis, i.e. a model where $f_{t,h-1}$ is taken as an unbiased and efficient forecast of y_{t+h}, $\hat{y}_{t,t+h}^{(9)} = f_{t,h-1}$.

10. Method 10 is a modified version of the EH which corrects for possible biases in forward rates:

$$y_{t+h} = -T_h + \beta f_{t,h-1} + u_{t+h} \quad h = 1, 4, 12, \tag{14}$$

where T_h is an $h-$period term premium that is assumed to be time-invariant. The model is recursively estimated to generate $\hat{y}_{t,t+h}^{(10)} = -\hat{T}_h + \hat{\beta} f_{t,h-1}$. Since risk premia can change as a function of the horizon (T_h), we refer to this as the 'Liquidity Preference' model.

11. Method 11 is the random walk, $\hat{y}_{t,t+h}^{(11)} = y_t$.

12. Method 12 is a recursively estimated AR(4) specification, $\hat{y}_{t,t+h}^{(12)} = \hat{a}_0 + \sum_{j=1}^{4} \hat{a}_j E_t[y_{t+h-j}]$.

Table 2 shows results reported in annualized basis points. For example, a monthly RMSFE-value of 0.001 will be reported as an error of 120 basis point (b.p.) per year. A negative bias means that the forecasts on average exceed the realized spot rates.

Which method is best depends on the forecast horizon. When $h = 1$, it is optimal not only to model the presence of regimes in the joint distribution of spot and forward rates, but also to exploit information on the current state to combine forward rate and time-series forecasts. Forecasts from the regime switching model combined with an AR(4) tester generates the lowest RMSFE-value (34 annualized basis points) and a negligible bias of -2.5 b.p. Ignoring the presence of regimes and either directly forecasting off the simpler

VAR(1) model or computing optimal combinations assuming it is the data generating process increases the RMSFE to 62 b.p. or higher and hence reduces the forecast accuracy. Interestingly, this deterioration in the out-of-sample performance reflects both higher biases and more volatile forecast errors.[12]

All the proposed benchmarks fail at the short horizon. Even the best benchmark – the AR(4) model – produces a RMSFE-value more than double the value produced by method 1 and the random walk is not a particularly difficult benchmark to beat. The EH, either in the 'pure' form of model 9 ((14) with $\hat{T}_1 = 0$ and $\hat{\beta} = 1$), or in the modified form of method 10, delivers disappointing results, with RMSFE-values in excess of 85 b.p. and biases of almost 25 b.p. We conclude that, at short forecast horizons, imposing the EH directly leads to imprecise and biased forecasts. Utilizing the restrictions from the EH in the context of a regime switching model that allows forward rates to be combined with simpler time-series forecasts produces better forecasts.

As the forecast horizon grows, information from the filtered state probabilities becomes less useful. While there is some value in combining forecasts from (2) with either the random walk or AR(4) testers, at long horizons the minimum RMSFE is achieved by the VAR(1) regime switching model subject to the EH restriction (method 3). This method produces RMSFE-values between one-half and four-fifths of the value achievable through alternative methods. For instance, at $h = 12$, method 3 produces an RMSFE-value of 131 b.p. against the 183 b.p. of the best alternative benchmark, the AR(4) model; method 1 produces a RMSFE-value of 197 b.p. Even at long horizons, the performance of the pure EH and of the liquidity preference hypothesis (models 9 and 10) remains disappointing.[13] Conversely, when restrictions from the EH are applied on the forecasts from the regime switching model (method 3) or from the VAR model (method 7), forecast accuracy seems to improve. This effect is particularly important at the longer horizons. At these longer horizons, the EH restrictions on the combination weights lead to smaller biases and also reduce errors from parameter estimation, whereas at the short horizon ($h = 1$), the effect on the bias from ignoring the current state probabilities is more severe and so method 1 produces the best forecasts.

It is useful to consider the separate effects of combining versus allowing for regime dynamics and Table 2 allows us to do so. Comparing the combined forecast to the direct forecast of the spot rate from the single-state VAR models (method 5 versus method 8), benefits from combining mainly emerge at long horizons. Conversely, once regimes are introduced, combining seems to improve forecast accuracy at the short but not at the long horizons (method 1 versus method 4). Turning to the effect of allowing for regimes, comparisons of methods 1 and 5 or methods 4 and 8 suggest that the effect of regimes on forecast accuracy is of first order at most of the horizons.

Clearly, the performance of regime switching models depends not only on the trade-off between the flexibility of the model and the effect of parameter estimation error, but also on the ability of the regime switching framework to accurately track the latent states. A well-specified model will not only accurately identify regime shifts in the parameters governing the joint dynamics in spot and forward rates, but should also prove useful in forecasting future spot rates when such switches occur. We therefore proceed to calculate

---

[12]Imposing restrictions on the combination weights significantly worsens forecast accuracy at the short horizons, since the restrictions increase the resulting biases. Such restrictions affect the regime switching forecasts most adversely.

[13]Notice that imposing restrictions on the weights still leads to a deterioration in performance for $h = 4$, but seems not to matter for $h \geq 12$. This is consistent with the previous evidence since $\hat{\omega}^1_{t,t+h} \approx 1$ and $\hat{\omega}^2_{t,t+h} \approx 0$ at long horizons.

the following measure of precision in regime classification,

$$\rho(\hat{\boldsymbol{\pi}}_t) \equiv 1 - 4 \left(\max_{s=1,\ldots,k} \hat{\boldsymbol{\pi}}_t[s] \right) \left(\max_{s=1,\ldots,k} \{ \hat{\boldsymbol{\pi}}_t[s] \backslash \max_{s=1,\ldots,k} \hat{\boldsymbol{\pi}}_t[s] \} \right) \quad k \geq 2.$$

This measures is one minus four times the product of the two highest state probabilities estimated at time t.[14] Since $\max_{s=1,\ldots,k} \hat{\boldsymbol{\pi}}_t[s] \times \max_{s=1,\ldots,k} \{ \hat{\boldsymbol{\pi}}_t[s] \backslash \max_{s=1,\ldots,k} \hat{\boldsymbol{\pi}}_t[s] \} = 1/4$ when there is absolute uncertainty about regimes, the scalar 4 acts as a normalizing constant and in this case $\rho(\hat{\boldsymbol{\pi}}_t) = 0$. When any of the elements of $\hat{\boldsymbol{\pi}}_t$ equals one, investors perceive being in one of the regimes with certainty and $\rho(\hat{\boldsymbol{\pi}}_t) = 1$. Also in this respect, the four-state model seems to do a good job at describing the dynamics of short term interest rates, as the average value of the recursive estimates of $\rho(\hat{\boldsymbol{\pi}}_t)$ over the out-of-sample period 1980:01-2003:11 is 0.67. In 72% of the sample $\rho(\hat{\boldsymbol{\pi}}_t)$ exceeds 0.5 and in 39% of these months $\rho(\hat{\boldsymbol{\pi}}_t)$ exceeds 0.9, i.e. the regime is inferred with great precision.[15]

Our framework is sufficiently general to include information from additional testers. Following studies such as Diebold, Rudebusch, and Aruoba (2006), Ang and Piazzesi (2003) and Dai and Philippon (2004), we explore the role of macroeconomic variables in predicting future spot rates. Rather than including each individual macroeconomic variable, each month during the out-of-sample period, we estimate $h-$step predictive regression models

$$y_{t+h}^{macro} = b_{0,h} + \sum_{j=1}^{7} b_{j,h} x_{jt} + \gamma_h \hat{\sigma}_t + \theta_h y_t + u_{t+h}, \tag{15}$$

where $\hat{\sigma}_t$ is a measure of the 12-month rolling window volatility and x_{jt} is the jth macroeconomic variable. We consider seven macroeconomic variables, namely the rate of growth of capacity utilization (measured as the distance of real activity from a non-inflationary trend), the federal funds rate, the inflation rate, real personal consumption expenditure growth, the rate of growth of the Help Wanted Index, the growth rate in employment, and industrial production growth. All series are seasonally adjusted and available from FREDII at the Federal Reserve Bank of St. Louis.

The models differ in terms of which macroeconomic variables they include as we recursively estimate models by imposing a variety of zero coefficient restrictions. Each month the best $h-$period forecasting model is selected from a total of 2^9 models (all possible combinations of the regressors in (15)) by minimizing the SIC criterion. This criterion selects very parsimonious models and tends to include only capacity utilization, the federal funds rate, the inflation rate and volatility for most of the out-of-sample period. To keep the forecasting model as parsimonious as possible, we combine these h-step recursive forecasts from the macroeconomic model with the yield-based forecasts by taking the macro forecasts as given (i.e. not as part of model (2)) and then minimizing the RMSFE, using the variances and covariance between the forecast errors from the macro and yield models.

Macroeconomic spot rate forecasts generated from (15) were incorporated under methods 1 and 5 by setting $\hat{y}_{2,t,t+h} = \hat{y}_{t+h}^{macro}$ so that these additional testers do not influence the definition of the underlying regimes. Table 2 shows that combining iterated forward rate forecasts with macro-based predictions generates

---

[14]The operator $\max_{s=1,\ldots,k} \{ \hat{\boldsymbol{\pi}}_t[s] \backslash \max_{s=1,\ldots,k} \hat{\boldsymbol{\pi}}_t[s] \}$ picks the second highest probability over $s = 1, \ldots, k$.

[15]We also calculated correlations between the squares of the recursive, $h-$step forecast errors from a variety of regime switching models, starting from the one combining forward rate forecasts with AR(4) forecasts and the indicator variable $I_{\{\rho(\hat{\boldsymbol{\pi}}_t)>0.9\}}$. This generated correlations of -0.22 for $h = 1$ and -0.17 for $h = 4$ and 12 (all values are highly significant). Hence, the forecasting performance of the switching combination model is particularly good when the regime is well identified.

out-of-sample RMSFE-values that are slightly higher than when the simple AR(4) or random walk testers were used (e.g. 93 vs. 90 b.p. at $h = 4$).[16]

4.1. *Statistical Significance*

To address whether the out-of-sample performances are sufficiently different to allow us to draw any conclusions on the relative precision of the various forecasting methods, we implement the forecast accuracy comparison test proposed by Diebold and Mariano (1995). Define the differential loss of model m relative to the loss from model n

$$dif_{t,t+h}^{(m,n)} = \left(e_{t,t+h}^{(m)} \right)^2 - \left(e_{t,t+h}^{(n)} \right)^2 .$$

We test the significance of the differences between two sets of forecast errors based on the statistic

$$DM_h^{(m,n)} = \frac{\frac{1}{288-h} \sum_{t=1980:01}^{2003:(12-h)} dif_{t,t+h}^{(m,n)}}{\widehat{\sigma}(dif_{t,t+h}^{(m,n)})}, \tag{16}$$

where $\widehat{\sigma}(dif_{t,t+h}^{(m,n)})$ is an estimate of the standard error of the loss function differential. In practice, we use

$$\widehat{\sigma}(dif_{t,t+h}^{(m,n)}) = \sum_{j=-h}^{h} \widehat{Cov} \left[dif_{t,t+h}^{(m,n)}, dif_{t+j,t+j+h}^{(m,n)} \right].$$

Tables 3 and 4 provide pair-wise test statistics when $h = 1$ or $h = 12$. Each cell below the main diagonal reports the value of $DM_h^{(m,n)}$ for the models in row (m) and column (n). Negative (positive) numbers indicate that the row model out- (under-) perform the column model.

Unsurprisingly, the sharpest conclusions on the relative performance of our methods can be reached for $h = 1$. At this horizon, method 1 outperforms all other methods at a statistically significant margin, except for method 3. The forecast comparison offers less clear-cut results (higher p-values) for $h = 12$ months. Even so, both the pure EH and the Liquidity Hypothesis are systematically outperformed by all other models.

We address possible departures of the small-sample distribution of (16) from normality by block-bootstrapping the distribution of $DM_h^{(m,n)}$ for all pairs of models and $h = 1, 12$ months. The resulting bootstrapped $p-$values (using 50,000 trials) are reported above the main diagonal in Tables 3 and 4. At the 1-month horizon Table 3 indicates that forecast combinations based on regime switching models outperform single-state models by a statistically significant margin, with the majority of the p-values between 0.01 and 0.05. The only exception is the pairwise comparison with the AR(4) model. Although the regime-switching combination lowers the RMSFE of the AR(4) model by 37 b.p. (from 71 to 34 b.p.), the loss function differentials $(e_{t,t+1}^{(1)})^2 - (e_{t,t+1}^{(12)})^2$ are sufficiently volatile that the bootstrapped p-values exceed 0.25. Single-state forecasts (both combinations and spot and forward predictions) are dominated by the regime switching equivalents. Table 4 confirms the earlier impression that differences are less clear-cut at longer forecast horizons.

---

[16]We separately investigated the predictive accuracy of models based on the macroeconomic variables. In general, they perform quite poorly for $h \leq 12$ with RMSFE-values between 100 and 250 basis points. Furthermore, the bias associated with these models is large. At the long horizon ($h = 24$), the performance of the two macro-based models is similar to other benchmarks, although these models remain inferior to the best among the regime switching models. Macroeconomic forecasts are also dominated by the VAR forecast at the shorter horizons up to 12 months.

4.2. A Parametric Bootstrap

Caution should be exercised when interpreting these results. The small sample distribution of the statistic in (16) can show large departures from normality, thus complicating the task of conducting inferences on differential predictive accuracy. When $e_{t,t+h}^{(m)}$ and $e_{t,t+h}^{(n)}$ are generated from models whose parameters are estimated recursively, the test must be based on estimated values, using forecast errors $\hat{e}_{t,t+h}^{(m)}$ and $\hat{e}_{t,t+h}^{(n)}$, and this can affect the sampling distribution of the test. Clark and McCracken (2001) also show that the test depends on whether the forecasting models are nested or non-nested.

A partial solution to the fact that a number of the comparisons in Tables 3 and 4 involved models that are nested is to adopt the parametric bootstrap approach proposed by Kilian (1999). Consider the following VAR(4) representation of (spot and forward) interest rates and the macroeconomic variables that encompasses a number of the earlier forecasting models:

$$\mathbf{q}_{t+1} = \boldsymbol{\mu}^q + \sum_{j=1}^{4} \boldsymbol{\Phi}_j \mathbf{q}_{t+1-j} + \mathbf{u}_{t+1}, \tag{17}$$

where $\mathbf{q}_t \equiv (\mathbf{z}_t' \ \mathbf{x}_t')'$, $\mathbf{x}_t$ is a vector of macroeconomic factors and $\mathbf{u}_{t+1} \sim IIN(\mathbf{0}, \boldsymbol{\Sigma})$. The first equation in (17) is just an extension of (15) to the case where four lags of the macroeconomic variables forecast future spot rates, while the last equations pick up dynamics in $\mathbf{x}_t$. Such a process is flexible enough to provide a good description of the data at hand. The parametric bootstrap algorithm proceeds in four steps:

1. Estimate the parameters $\boldsymbol{\mu}^q$, $\{\boldsymbol{\Phi}_j\}_{j=1}^4$, and $\boldsymbol{\Sigma}$ in (17) over the full-sample 1954:01 - 2003:12. This yields the vector of residuals $\{\hat{\mathbf{u}}_t\}_{t=1954:05}^{2003:12}$.

2. Based on these estimates, generate a sequence of pseudo observations $\{\mathbf{q}_t^b\}_{t=1954:05}^{2003:12}$ from (17). We initialize the process at the unconditional means of $\mathbf{q}_t$ and discard the first 1,000 transients. The pseudo innovation terms $\{\hat{\mathbf{u}}_t^b\}_{t=1954:05}^{2003:12}$ are drawn randomly with replacement from the set of observed residuals $\{\hat{\mathbf{u}}_t\}_{t=1954:05}^{2003:12}$. We repeat this step B times.

3. For each of the B bootstrap replications $\{\mathbf{q}_t^b\}_{t=1954:05}^{2003:12}$ generated in the previous step, recursively estimate models 1, 3-5, 7-8, and 11-14 over the expanding periods 1950:01 - 1980:01 up to 1950:01 - 2003:12-h, where h is the forecast horizon.[17] This gives a sequence of h-step interest rate forecasts for each of the models. Compute the forecast errors $e_{t,t+h}^{(m)} \equiv y_{t+h} - \hat{y}_{t,t+h}^{(m)}$ and proceed to obtain B bootstrap replications of $\{\widehat{DM}_{h,b}^{(m,n)}\}_{b=1}^B$.

4. Using the small-sample simulated distribution, $\{\widehat{DM}_{h,b}^{(m,n)}\}_{b=1}^B$, the p-values for two-sided tests of the null of zero differential predictive accuracy is given by the percentage of bootstrapped simulations such that $|\widehat{DM}_{h,b}^{(m,n)}| > |DM_h^{(m,n)}|$, where $DM_h^{(m,n)}$ is the sample statistic reported in Tables 3 and 4.

Table 5 reports the results in the form of (signed) bootstrap p-values for $h = 1$ (below the main diagonal) and $h = 12$ (above the main diagonal). The values of $DM_h^{(m,n)}$ are identical to those appearing in Tables 3

---

[17]Models 2 and 6 were dropped because of our earlier finding that restricting the combination weights fails to improve the forecast performance. Models 9 and 10 cannot be simulated because they involve forward rates and hence would imply bootstrapping the overall dynamics of the term structure of interest rates, which is beyond the scope of our paper.

and 4. Consistent with Tables 3 and 4, p-values inherit the sign of the corresponding $DM_h^{(m,n)}$ statistic, so that a negative p-value indicates that a row model outperforms the column model; p-values that are less than or equal to 0.05 are boldfaced. At $h = 1$ results remain favorable to the regime switching forecasts, which are found to significantly outperform most of the other models. The dominance of regime switching models over single-state models is unchanged. VAR(1) forecasts prove superior to either the random walk or to the simple use of macroeconomic variables. At the 12-month horizon, pairwise comparisons between single- and multi-state models show that the latter outperform the former.

5. Conclusion

This paper proposed a four-state model to capture the dynamics in US spot and forward rates. We proposed a flexible approach that combines forecasts of future spot rates with other 'testers' that can be viewed as forecasts obtained from alternative model specifications. In an out-of-sample forecasting exercise we found evidence that, particularly at short horizons, combining regime switching forecasts with simpler, univariate time-series forecasts can help reduce the root mean squared forecast error. At longer horizons, we found that imposing theoretical restrictions from the expectations hypothesis linking future spot rates to forward rates helps improve forecasting accuracy. Although the expectations hypothesis is rejected using in-sample tests, it may still be helpful in improving out-of-sample prediction accuracy.

References

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

[2] Aiolfi, M., and Timmermann, A., 2006. Persistence of forecasting performance and combination strategies. Journal of Econometrics 135, 31-53.

[3] Ang, A., and Bekaert, G., 2002. Short rate nonlinearities and regime switches. Journal of Economic Dynamics and Control 26, 1243-1274.

[4] Ang, A., and Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 745-787.

[5] Bansal, R., Tauchen G., and Zhou, H., 2004. Regime-shifts, risk premiums in the term structure, and the business cycle. Journal of Business and Economic Statistics 22, 396-409.

[6] Bates, J., and Granger, C., 1969. The combination of forecasts. Operations Research Quarterly 20, 451-468.

[7] Campbell, J., and Shiller, R., 1991. Yield spreads and interest rate movements: A bird's eye view. Review of Economic Studies 58, 495-514.

[8] Clark, T., and McCracken, M., 2001. Tests of forecast accuracy and encompassing for nested models. Journal of Econometrics 105, 85-110.

[9] Dai, Q. and Philippon, T., 2004. Fiscal policy and the term structure of interest rates. NBER working paper No. 11574.

[10] Deutsch, M., Granger C., and Terasvirta, T., 1994. The combination of forecasts using changing weights. International Journal of Forecasting 10, 47-57.

[11] Diebold, F., and Kilian, L., 2000. Unit root tests are useful for selecting forecasting models. Journal of Business and Economics Statistics 18, 265-273.

[12] Diebold, F., and Lopez, J., 1996. Forecast evaluation and combination. In: Maddala and Rao (eds.), Handbook of Statistics, Elsevier, Amsterdam.

[13] Diebold, F., and Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253-263.

[14] Diebold, F., and Pauly, P., 1987. Structural change and the combination of forecasts. Journal of Forecasting 6, 21-40.

[15] Diebold, F., Rudebusch G., and Aruoba, B., 2006. The macroeconomy and the yield curve: a dynamic latent factor approach. Journal of Econometrics 131, 309-338.

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

[17] Egorov, A., Hong, Y., and Li, H., 2005. Forecasting the joint probability density of bond yields: can affine models beat random walk?, mimeo, Cornell University.

[18] Elliott, G., and Timmermann, A., 2005. Optimal forecast combination weights under regime switching. International Economic Review 46, 1081-1102.

[19] Fama, E., 1984. Forward and spot exchange rates. Journal of Monetary Economics 14, 319-338.

[20] Granger, C., and Terasvirta, T., 1993. Modeling Nonlinear Economic Relationships. Oxford University Press, Oxford.

[21] Gray, S., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27-62.

[22] Hall, A., Anderson, H., and Granger, C., 1992. A cointegration analysis of treasury bill yields. Review of Economics and Statistics 74, 116-126.

[23] Hamilton, J., 1988. Rational expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. Journal of Economic Dynamics and Control 12, 385-423.

[24] Kilian, L., 1999. Exchange rates and monetary fundamentals: what do we learn from long-horizon regressions? Journal of Applied Econometrics 14, 491-510.

[25] Rudebusch, G., and Wu, T., 2004. A macro-finance model of the term structure, monetary policy, and the economy. Mimeo, Federal Reserve Bank of St. Francisco.

[26] Sola, M., and Driffill, J., 1994. Testing the term structure of interest rates using a stationary vector autoregression with regime switching. Journal of Economic Dynamics and Control 18, 601-628.

[27] Stock, J., and Watson, M., 2001, A comparison of linear and nonlinear univariate models for forecasting macroeconomic time Series. In: Engle R., and White H., (eds.), Cointegration, Causality, and Forecasting: Festschrift in Honour of Clive Granger, Oxford University Press, Oxford, pp. 1-44.

[28] Stock, J., and Watson, M., 2005. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23, 405-430.

[29] Timmermann, A., 2006. Forecast combinations. In: Elliott, G., Granger C., and Timmermann A., (eds.), Handbook of Economic Forecasting, Elsevier Press, Amsterdam, pp. 135-196.

Table 1

Estimates of a Four-State VAR(1) Regime Switching Model

The table shows estimation results for the two-state vector autoregressive regime switching model:

$$\mathbf{z}_{t+1} = \mathbf{\mu}_{S_{t+1}} + \mathbf{A}\mathbf{z}_t + \mathbf{\varepsilon}_{t+1}$$

where $\mathbf{z}_{t+1} = [y_{t+1}\ f_{t,1}]'$, $f_{t,1}$ is the one-month forward rate. $\mathbf{\mu}_{S_{t+1}}$ is the intercept vector in state S_{t+1}, $\mathbf{A}$ is a matrix of autoregressive coefficients, $\mathbf{\varepsilon}_{t+1} = [\varepsilon_{1t+1}\ \varepsilon_{2t+1}]' \sim N(\mathbf{0}, \Omega_{S_{t+1}})$. The unobserved state variable S_{t+1} is governed by a first-order Markov chain that can assume four values. The first panel reports estimates for the single-state case $k = 1$. Asterisks on correlation coefficients refer to covariance estimates. For mean coefficients and transition probabilities, standard errors are reported in parentheses. Data is expressed as percentages (basis points). The sample period is 1950:01 – 2003:12.

	Panel A – Single State Model	
	Spot rate	Forward rate
1. Intercept	0.0097 (0.0041)	0.0059 (0.0033)
2. VAR(1) Matrix		
Spot rate	0.291 (0.081)	0.717 (0.085)
Forward rate	0.169 (0.067)	0.809 (0.070)
3. Correlations/Volatilities		
Spot rate	0.0501***	
Forward rate	0.8771***	0.0422***
	Panel B – Four State Model	
	Spot rate	Forward rate
1. Intercept		
State 1 (Low/Stable)	0.0008 (0.0017)	0.0003 (0.0015)
State 2	0.0098 (0.0025)	0.0007 (0.0031)
State 3	0.0141 (0.0066)	0.0027 (0.0090)
State 4 (High Volatility)	0.0416 (0.0204)	0.0077 (0.0116)
2. VAR(1) Matrix		
Spot rate	0.1653 (0.0715)	0.8002 (0.0729)
Forward rate	0.0980 (0.0622)	0.9008 (0.0623)
3. Correlations/Volatilities		
State 1 (Low/Stable):		
Spot rate	0.0078***	
Forward rate	0.9237***	0.0063**
State 2:		
Spot rate	0.0213***	
Forward rate	0.8993***	0.0170***
State 3:		
Spot rate	0.0483***	
Forward rate	0.8928***	0.0381**
State 4 (High Volatility):		
Spot rate	0.1257***	
Forward rate	0.9017***	0.1095***

4. Transition probabilities	State 1	State 2	State 3	State 4
State 1 (Low/Stable)	0.843	0.153	0.004	0.000
State 2	0.044	0.830	0.110	0.016
State 3	0.000	0.149	0.818	0.033
State 4 (High Volatility)	0.000	0.000	0.190	0.810

* denotes significance at the 10%, ** significance at the 5%, *** significance at the 1% level.

Table 2

Out-of-Sample Forecasting Performance

The table reports summary statistics for *h*-step-ahead recursive forecasts of spot rates under a variety of methods and testers such as random walk (RW), autoregressive (AR) and macro (M) forecasts. Bias and root mean squared forecast error (RMSFE) are expressed in basis points per annum. For each statistic and horizon, the best values are boldfaced.

	Tester	RMSFE (annual basis points)				Bias (annual basis points)			
		h=1	h=4	h=12	h=24	h=1	h=4	h=12	h=24
Regime Switching Models									
1. Switching VAR(1) Combination	RW	34.15	91.20	197.50	291.58	-2.46	-28.26	-76.35	-150.71
	AR	**34.14**	90.03	197.41	291.39	-2.47	-28.25	-76.41	-150.71
	M	68.18	93.48	197.29	291.30	**-2.24**	-23.84	-75.64	-150.06
2. Restricted Switching VAR(1) Combination	RW	68.29	103.44	197.50	291.58	-22.46	-38.26	-76.37	-150.71
	AR	68.28	103.44	197.48	291.53	-22.52	-38.25	-76.36	-150.69
	M	73.80	119.65	224.69	341.33	-2.77	-24.28	-79.40	-160.34
3. Switching VAR(1) Forward Forecast	—	34.95	**52.41**	**130.99**	**208.59**	4.55	**-11.56**	**-50.05**	-124.87
4. Switching VAR(1) Spot Forecast	—	66.81	104.50	198.37	292.61	-23.11	-39.32	-77.47	-152.21
Single-State VAR(1) Models									
5. VAR(1) Combination	RW	70.80	96.12	298.92	654.23	-21.72	-119.5	-172.5	-258.7
	AR	70.74	95.16	299.52	653.16	-21.76	-119.3	-172.2	-258.8
	M	78.82	99.94	376.09	809.15	-23.43	-124.4	-196.9	-271.7
6. Restricted VAR(1) Combination	RW	73.18	97.44	302.88	662.37	-21.59	-119.4	-172.1	-259.9
	AR	71.97	95.98	299.99	657.63	-21.78	-119.2	-172.2	-259.3
	M	76.56	118.75	358.99	685.92	-25.22	-152.9	-217.8	-280.5
7. VAR(1) Forward Forecast	—	62.43	86.52	270.24	595.32	4.64	-49.96	-164.6	-246.7
8. VAR(1) Spot Forecast	—	67.20	121.58	390.12	749.04	-21.83	-194.7	-272.5	-355.8
Benchmarks									
9. Pure EH	—	85.07	150.53	395.42	NA	29.37	47.78	128.20	NA
10. Liquidity preference Hypothesis	—	115.59	161.70	299.16	NA	-24.12	-47.41	-84.48	NA
11. Random Walk	—	80.43	105.69	184.98	267.37	-26.19	-39.49	-69.04	-128.53
12. AR(4)	—	70.55	107.89	182.94	239.72	-26.68	-38.69	-64.51	**-113.71**

Table 3

Comparison of Predictive Accuracy – 1 Month Horizon

The table reports Diebold-Mariano statistics for comparisons of the MSFE produced by different forecasting methods. The test is applied pairwise to forecast errors from recursive, 1-step forecasts of spot interest rates using a variety of univariate models and two possible tester forecasts, the random walk (RW) and an AR(4) univariate model for the spot rate. Statistics illustrate the comparative forecasting performance of the model in the row vs. the model in the column. Negative (positive) values indicate that the row model out- (under-) performs the column model. Diebold-Mariano statistics are shown below the main diagonal while p-values from a block bootstrap with 50,000 trials are reported above the diagonal.

			Switching VAR(1) Combination (1)		Restricted Switching VAR(1) Combination (2)		Switching VAR Forward (3)	Switching VAR(1) Spot Forecast (4)	VAR(1) Combination (5)		VAR(1) Forward (7)	VAR(1) Spot Forecast (8)	Pure EH (9)	Liquidity Preference Hypothesis (10)	Random Walk (11)	AR(4) (12)
			RW	AR	RW	AR			RW	AR						
Switching VAR(1) Combination	(1)	RW		0.845	0.546	0.142	0.991	0.048	0.029	0.028	0.053	0.030	0.008	0.001	0.049	0.284
		AR	-0.37		0.546	0.843	0.546	0.049	0.021	0.018	0.049	0.031	0.044	0.001	0.045	0.274
Restricted Switching VAR(1) Combination	(2)	RW	1.20	0.79		0.345	0.579	0.105	0.276	0.305	0.914	0.856	0.001	0.001	0.050	0.271
		AR	1.96	0.38	1.20		0.044	0.100	0.299	0.279	0.900	0.861	0.051	0.001	0.021	0.270
Switching VAR(1) Forward Forecast	(3)		0.07	1.06	-0.55	-2.07		0.154	0.015	0.014	0.018	0.012	0.020	0.007	0.010	0.304
Switching VAR(1) Spot Forecast	(4)		-2.27	-2.24	-2.07	-2.02	-1.72		0.040	0.038	0.037	0.035	0.039	0.001	0.041	0.542
VAR(1) Combination	(5)	RW	2.98	2.97	1.20	1.20	3.12	2.76		0.820	0.098	0.905	0.004	0.001	0.015	0.985
		AR	2.97	2.99	1.18	1.19	3.08	2.75	-0.16		0.100	0.785	0.004	0.000	0.013	0.984
VAR(1) Forward Forecast	(7)		2.11	2.14	-0.06	-0.05	2.86	2.55	-1.72	-1.72		0.279	0.006	0.000	0.020	0.712
VAR(1) Spot Forecast	(8)		2.86	2.89	-0.24	-0.22	3.30	2.71	0.34	0.41	-0.82		0.008	0.002	0.024	0.904
Pure EH	(9)		-8.43	-8.44	-8.39	-8.44	-4.90	-7.06	-5.38	-5.41	-4.88	-4.46		0.005	0.652	0.198
Liquidity Preference Hypothesis	(10)		-9.88	-9.86	-9.88	-9.85	-9.42	-9.88	-7.05	-7.12	-8.03	-5.65	9.85		0.002	0.003
Random Walk	(11)		-2.86	-2.88	-2.87	-2.82	-3.39	-3.06	-3.85	-3.88	-3.33	-3.05	-1.15	-9.86		0.027
AR(4)	(12)		-1.42	-1.40	-1.43	-1.42	-1.26	-0.70	0.05	0.04	-0.41	-0.24	-2.63	-9.84	-3.00	

Table 4

Comparison of Predictive Accuracy – 12 Month Horizon

The table reports Diebold-Mariano statistics for comparisons of the MSFE produced by different forecasting methods. The test is applied pairwise to forecast errors from recursive, 1-step forecasts of spot interest rates using a variety of univariate models and two possible tester forecasts, the random walk (RW) and an AR(4) univariate model for the spot rate. Statistics illustrate the comparative forecasting performance of the model in the row vs. the model in the column. Negative (positive) values indicate that the row model out- (under-) performs the column model. Diebold-Mariano statistics are shown below the main diagonal while p-values from a block bootstrap with 50,000 trials are reported above the diagonal.

			Switching VAR(1) Combination (1)		Restricted Switching VAR(1) Combination (2)		Switching VAR Forward (3)	Switching VAR(1) Spot Forecast (4)	VAR(1) Combination (5)		VAR(1) Forward (7)	VAR(1) Spot Forecast (8)	Pure EH (9)	Liquidity Preference Hypothesis (10)	Random Walk (11)	AR(4) (12)
			RW	AR	RW	AR			RW	AR						
Switching VAR(1) Combination	(1)	RW		0.402	0.570	0.683	0.095	0.925	0.105	0.101	0.184	0.045	0.011	0.042	0.161	0.121
		AR	-0.53		0.489	0.440	0.093	0.894	0.084	0.085	0.170	0.036	0.013	0.043	0.274	0.145
Restricted Switching VAR(1) Combination	(2)	RW	0.99	1.26		0.347	0.090	0.989	0.064	0.078	0.176	0.028	0.013	0.042	0.277	0.132
		AR	-0.46	1.51	-1.02		0.079	0.861	0.069	0.074	0.008	0.039	0.023	0.042	0.121	0.149
Switching VAR(1) Forward Forecast	(3)		-1.93	-1.91	-1.96	-1.93		0.096	0.041	0.039	0.106	0.008	0.027	0.044	0.332	0.105
Switching VAR(1) Spot Forecast	(4)		-0.17	-0.29	-0.03	-0.16	-1.90		0.298	0.308	0.300	0.096	0.010	0.039	0.154	0.124
VAR(1) Combination	(5)	RW	1.98	2.00	2.06	2.05	2.71	1.31		0.986	0.410	0.324	0.002	0.988	0.014	0.078
		AR	1.97	1.96	2.01	2.02	2.72	1.33	-0.03		0.354	0.319	0.001	0.965	0.013	0.069
VAR(1) Forward Forecast	(7)		1.66	1.69	1.79	1.80	2.98	1.28	-0.98	-1.05		0.117	0.006	0.210	0.009	0.093
VAR(1) Spot Forecast	(8)		2.45	2.51	2.58	2.57	3.59	1.98	-1.05	-1.10	-1.69		0.699	0.077	0.005	0.018
Pure EH	(9)		-3.58	-3.52	-3.55	-3.62	-3.66	-3.58	-5.38	-5.41	-4.88	-0.56		0.033	0.001	0.019
Liquidity Preference Hypothesis	(10)		-2.50	-2.49	-2.50	-2.49	-2.44	-2.47	-0.03	-0.09	-1.33	2.06	-3.54		0.023	0.041
Random Walk	(11)		1.06	1.02	1.13	1.08	-1.22	1.07	3.85	3.88	3.33	4.06	-3.50	-2.95		0.349
AR(4)	(12)		1.20	1.14	1.21	1.16	-1.47	1.21	1.89	1.91	1.86	2.76	-3.65	-2.74	-0.23	

Table 5

Predictive Accuracy Comparison Tests – Parametric Bootstrap P-Values

The table reports p-values for the Diebold-Mariano test based on pairwise comparisons of the MSFE produced by different forecasting methods. The test is applied to forecast errors from recursive, 1 and 12-step forecasts of spot rates using a variety of univariate models and three possible tester forecasts, the random walk (RW), an AR(4) univariate model for the spot rate, and macro forecasts (M). P-values are computed by computing the Diebold-Mariano statistic on recursive forecast errors when the data are simulated from a VAR(4) model for spot and forward interests rates extended to include macroeconomic predictor variables. Innovations are drawn from a bivariate regime switching model and the regression model with macroeconomic factors. The bootstrap uses 1,000 independent trials. Statistics illustrate the comparative forecasting performance of the model in the row vs. the model in the column: a negative (positive) value indicates that the row model out- (under-) performs the column model. Below the main diagonal we present results for 1-month forecasts; above the main diagonal, results are for 12-month forecasts. P-values inherit the sign of the corresponding DM statistic, so that a negative p-value indicates that a row model outperforms the column model. P-values that are equal or less than 0.05 are boldfaced.

			Switching VAR(1) Combination (1)			Switching VAR(1) Forward Forecast (3)	Switching VAR(1) Spot Forecast (4)	VAR(1) Combination (5)			VAR(1) Forward Forecast (7)	VAR(1) Spot Forecast (8)	Random Walk (11)	AR(4) (12)
			RW	AR	M			RW	AR	M				
							12-month horizon							
Switching VAR Combination	(1)	RW		0.787	0.680	0.079	0.785	-0.112	-0.056	-0.100	-0.192	**-0.032**	-0.205	-0.178
		AR	-0.954		0.579	0.085	0.750	**-0.049**	-0.052	-0.086	-0.199	**-0.003**	-0.066	-0.069
		M	0.351	0.279		0.073	0.698	**-0.039**	-0.072	**-0.048**	-0.140	**-0.011**	-0.077	-0.076
Switching VAR Forward	(3)		0.659	0.675	-0.079		0.085	**-0.006**	**-0.012**	**-0.020**	**-0.035**	**-0.001**	0.084	0.078
Switching VAR Spot	(4)		-0.265	-0.204	-0.340	-0.785		-0.252	-0.300	-0.450	-0.249	-0.100	-0.055	-0.115
VAR Combination	(5)	RW	**0.039**	**0.033**	**0.020**	**0.035**	0.052		0.919	-0.749	0.346	0.332	**-0.007**	-0.096
		AR	**0.033**	**0.024**	**0.022**	**0.014**	**0.048**	-0.816		-0.799	0.333	0.332	**-0.017**	-0.071
		M	**0.043**	**0.038**	**0.021**	**0.028**	**0.045**	0.759	0.800		0.258	0.210	**-0.041**	-0.082
VAR Forward	(7)		**0.038**	0.063	**0.050**	**0.016**	0.060	-0.115	-0.051	-0.089		0.079	**-0.013**	-0.070
VAR Spot Forecast	(8)		**0.037**	**0.044**	**0.011**	**0.045**	**0.043**	0.896	0.812	0.658	-0.275		**-0.019**	**-0.009**
Random Walk	(11)		**-0.035**	**-0.029**	**-0.037**	**-0.035**	**-0.003**	**-0.031**	**-0.049**	**-0.033**	**-0.030**	**-0.030**		0.569
AR(4)	(12)		**-0.015**	**-0.010**	-0.114	-0.079	**-0.047**	0.882	0.819	0.846	-0.701	-0.927	-0.699	

(Left vertical label: 1-month horizon)

Figure 1

Smoothed State Probabilities for the Four-State VAR(1) Regime Switching Model

The graphs plot the smoothed state probabilities for the multivariate Markov Switching model for spot and forward rates.
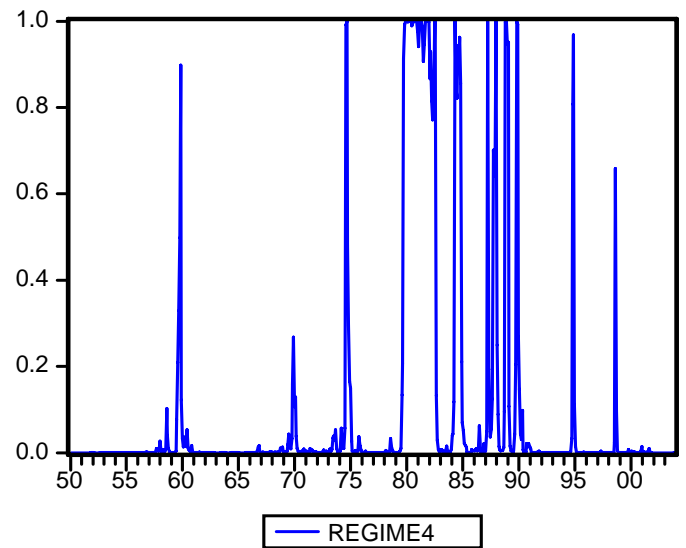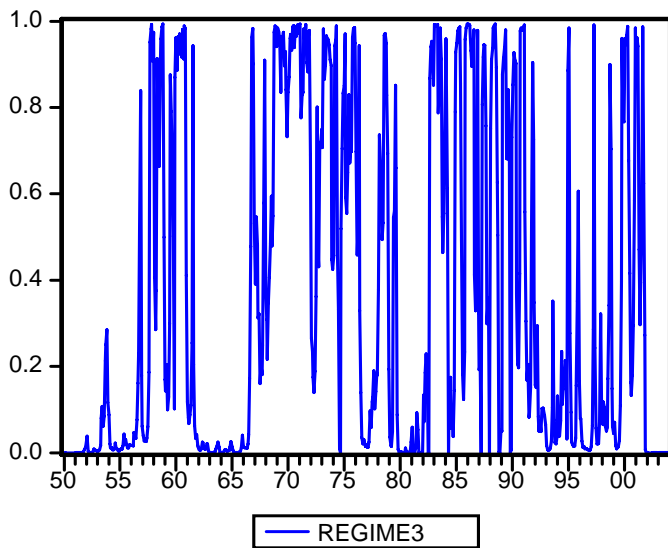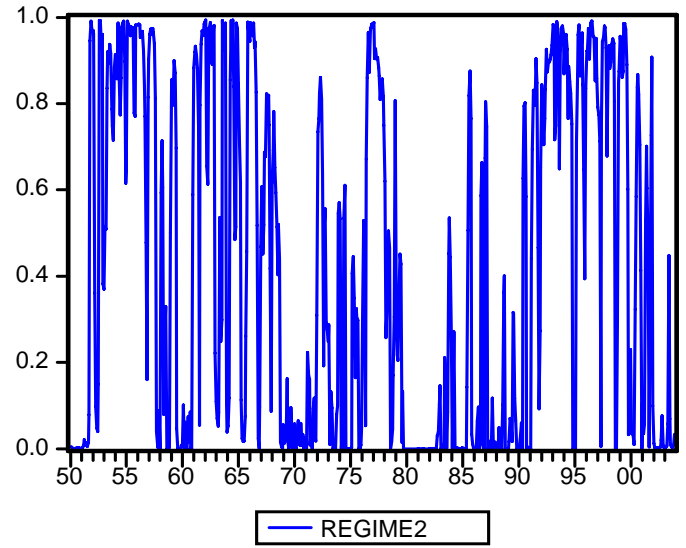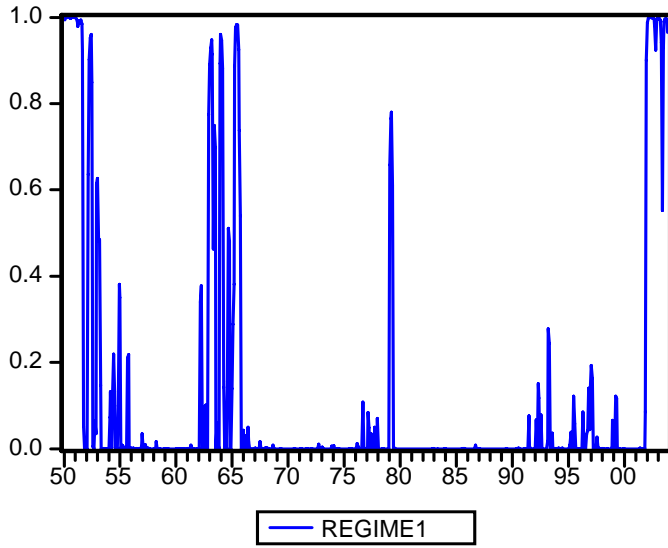
Figure 2

Simulated Distribution for Augmented Dickey-Fuller Test P-values

The graph displays the simulated distribution (over 1,000 trials) of the p-values obtained in ADF tests of the null of a unit root in the 1-month US T-bill rate. 648-month long time series are simulated from the bivariate four-state regime switching model after discarding 100 transients. The ADF test includes a constant while the number of lags is selected by minimizing the BIC for each simulation trial.
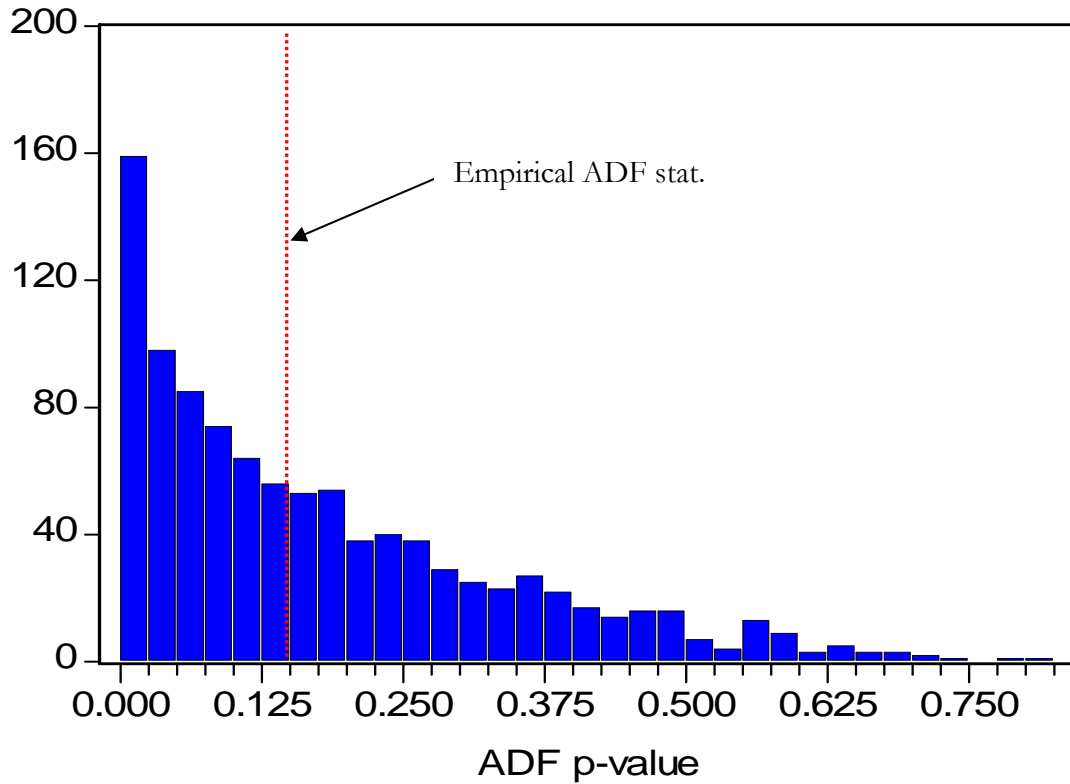
Figure 3

**Optimal Forecast Combination Weights as a Function of State Probabilities
– AR(4) Benchmark**

The graphs plot the values of ω_1 (the weight assigned to the forward rate forecasts) and ω_2 (the weight assigned to the AR(4) forecasts) that minimize the MSFE of the combined forecast as a function of the forecast horizon (h). Results are shown for different configurations of the initial state probabilities.
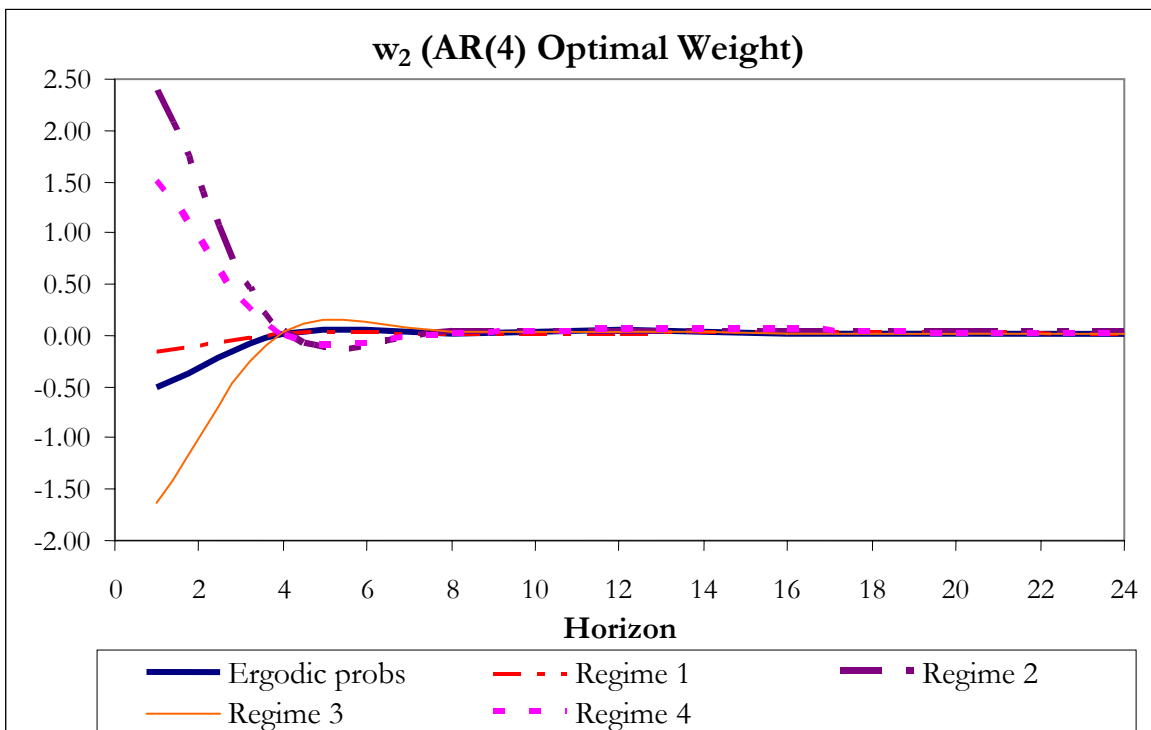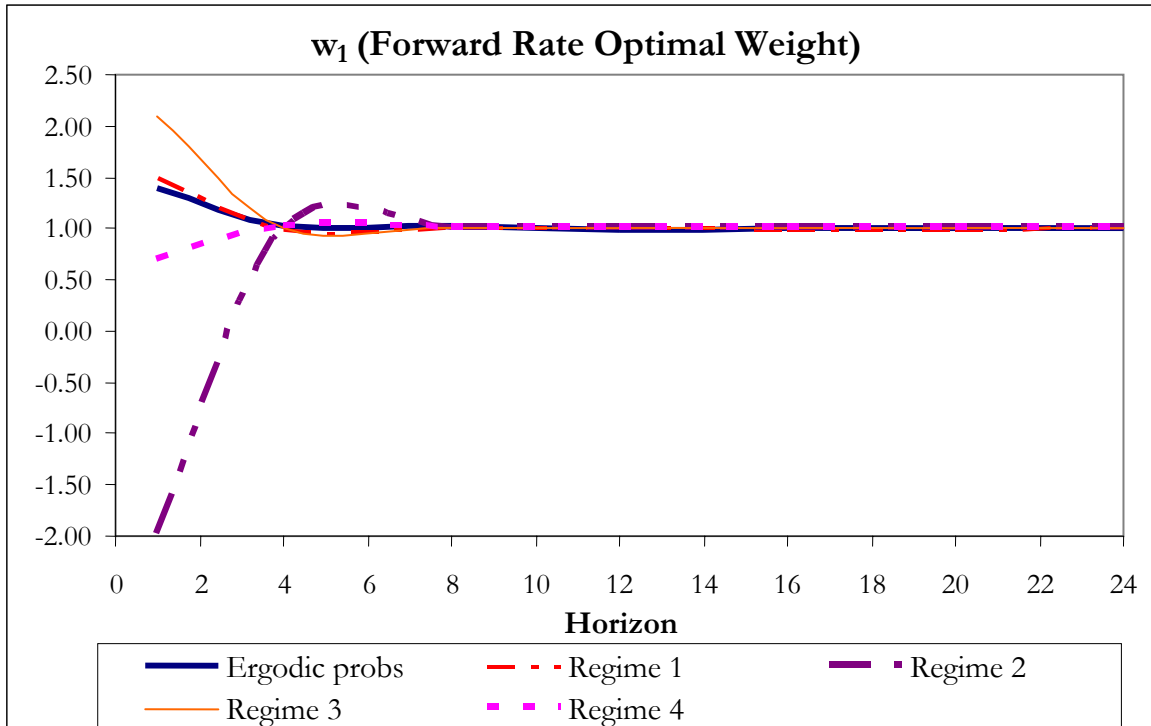
Figure 4
Improvement in forecast Precision Due to Combining Forecasts

The graphs show the percentage decline in root mean squared forecast error obtained by combining forward rates and AR(4) forecasts vs. pure forward forecasts ($\omega_0 = \omega_2 = 0$ and $\omega_1 = 1$) and AR(4) forecasts under a number of assumptions for the initial state.