Abstract

We examine the statistical accuracy and economic value of modeling and forecasting the term structure of interest rates using forecast combinations. We adopt five alternative methods to combine point forecasts from several univariate and multivariate autoregressive specifications, as well as from factor models for the yield curve such as the dynamic versions of the Nelson-Siegel and Svensson specifications. Moreover, we conduct a detailed performance evaluation based not only on statistical measures of forecast accuracy, but also an economic evaluation using Sharpe ratios of optimal mean-variance fixed income portfolios constructed based upon forecasts from individual models and their alternative combinations. Our empirical application based on a large panel of Brazilian interest rate future contracts with different maturities shows that combined forecasts consistently outperform individual models in several instances, specially when economic criteria is taken into account.

JEL: C53; E43; G17.
key-words: yield curve; dynamic factor models; forecast combinations; economic value of forecasts; Kalman filter.
1 Introduction

The ability to forecast the behavior of the term structure of interest rates is important for macroeconomists, financial economists and fixed income managers. More specifically, bond portfolio optimization, pricing of financial assets and their derivatives, as well as risk management, rely heavily on interest rate forecasts. Moreover, these forecasts are widely used by financial institution, regulators, and institutional investors to develop macroeconomic scenarios. However, until the seminal work of Diebold & Li (2006), little attention was given to yield curve forecasting, and previous theoretical developments were mainly focused on in-sample fit (see, for example, de Jong, 2000; Dai & Singleton, 2000). Diebold & Li (2006) have taken an out-of-sample perspective based on a dynamic version of the static approach proposed by Nelson & Siegel (1987) and have shown that this model produce accurate forecasts.

The seminal work of Diebold & Li (2006) on yield curve forecasting has been followed by a large number of studies that investigate the performance of alternative forecasting models; see, for instance, Diebold & Rudebusch (2013) for a text book review of these improvements. For instance, Diebold et al. (2006) jointly modeled the yield curve and macroeconomic variables, allowing for a bidirectional relationship between the yield curve and the macroeconomy to affect the forecasts; Christensen et al. (2011) developed an arbitrage-free version of this model and argued that this restriction may improve forecast; Haustsch & Yang (2012) show that allowing for stochastic volatility in the underlying yield factors enhances forecasting performance.

An interesting topic still unexplored in the literature is the implementation and performance evaluation of forecast combinations for the yield curve. Existing evidence has focused on the performance evaluation of individual forecast models. However, combined forecasts has been extensively and successfully applied in many areas; see Granger (1989), Clemen (1989), Granger & Jeon (2004), Timmermann (2006) and Wallis (2011) for reviews. Therefore, a natural question is: can forecast combinations deliver better forecasts for the yield curve? If so, are these forecasts economically relevant? The motivation to combine forecasts comes from an important result from the methodological literature on forecasting, which shows that a linear combination of two or more forecasts may yield more accurate predictions than using only a single forecast (Granger, 1989; Newbold & Harvey, 2002; Aiolfi & Timmermann, 2006). Moreover, adaptive strategies for combining forecasts might also mitigate structural breaks and model misspecification and thus lead to more accurate forecasts (Newbold & Harvey, 2002; Pesaran & Timmermann, 2007). In particular, there is recent evidence that combining forecast of nested models can significantly improve forecasting precision upon forecasts obtained from single model specifications (Clark & McCracken, 2009). However, to the best of our knowledge, there is no references in the
existing literature to the statistical and economic evaluation of forecast combination for the yield curve.

_Hendry & Clements (2004)_ point out a number of potential explanations for the good performance of combined forecasts vis-a-vis individual forecast models. First, if two models provide partial, but incompletely overlapping explanations, then some combination of the two might do better than either alone. Specifically, in particular, if two forecasts were differentially biased (one upwards, one downwards), then combining could be an improvement over either. Similarly, if all explanatory variables were orthogonal, and models contained subsets of these, an appropriately weighted combination could better reflect all the information. Second, averaging forecasts reduces variance to the extent that separate sources of information are used. Third, forecast combination can also be viewed as an application of the Stein-James shrinkage estimation. In this case, if the unknown future value is viewed as a meta-parameter of which all the individual forecasts are estimates, then averaging may provide a better estimate thereof. Finally, _Granger & Jeon (2004)_ point out that the benefits of pooling forecasts can be related to the portfolio selection problem, since a portfolio of assets is usually better than investing in a single asset.

Our paper also provides a comprehensive evaluation of the forecasting performance of forecast combinations vis-a-vis individual models in terms of both statistical accuracy and economic relevance, since the goal of interest rate forecasts is to improve decision making in portfolio allocation, pricing derivatives or managing financial risk. The existing literature, however, has focused mainly on statistical measures of forecast accuracy, taking forecasts out of the context of decision making in which they are ultimately used; see, for instance, _Pooter (2007), Diebold & Rudebusch (2013), Caldeira et al. (2010), and de Rezende & Ferreira (2013), amongst others._ As pointed out by _Granger & Pesaran (2000), Pesaran & Skouras (2002)_ and _Granger & Machina (2006),_ when forecasts are used in decision making, it is important to consider the decision process in the _ex post_ evaluation of these forecasts, maintaining the interaction between the forecasting model and the decision making task. Therefore, it is of main concern to both academics and market practitioners the extent to which forecasts of interest rates are useful to support economic decisions.

More specifically, we propose to evaluate the accuracy and economic relevance of yield curve forecasts based not only on statistical measures of forecasting performance, but also on mean-variance optimal portfolios as introduced by _Markowitz (1952)._ As a first step, we follow _Christoffersen & Diebold (1998), Hördahl et al. (2006),_ and _de Pooter et al. (2010)_ and carry out a traditional evaluation based on statistical measures of forecasting performance such as root mean squared forecast error (RMSFE) and trace root mean squared forecast error (TRMSFE). In a second step, different forecasts are used to construct mean-variance portfolios, which are then compared based on their Sharpe ratios. In order to solve the mean-variance optimization problem, we use
the different forecasts to derive estimates of expected bond returns and their covariance matrix, and use these estimators as inputs to obtain mean-variance portfolios for alternative levels of risk tolerance.

We obtain forecasts of the yield curve based on a broad set of alternative models usually considered in the literature on yield rates forecast, such as the random walk, the univariate autoregressive model, the vector autoregressive and the Bayesian vector autoregressive model. We also consider well-established yield curve factor models such as the dynamic versions of the Nelson-Siegel and Svensson specifications. More importantly, we implement five alternative forecast combination schemes. First, we consider the case of equally weighted forecasts. Various studies have demonstrated that simple averaging of a multitude of forecasts works well in relation to more sophisticated weighting schemes (Newbold & Harvey, 2002; Clark & McCracken, 2009). Second, we consider the thick modeling approach proposed by Granger & Jeon (2004) which consists of selecting the best forecasting models in the sub-sample period for model evaluation, according to the root mean square error (RMSE) criterion. In this case, we use the selection process of Granger and Jeon and subsequently compute weights by means of OLS regressions. Third, we use the rank-weighted combinations suggested by Aiolfi & Timmermann (2006). Finally, we implement the thick modeling approach with MSE-frequency and RMSE-frequency weights, which consists of selecting models by means of the thick-modeling approach and assigning to each individual forecast a weight equal to a model’s empirical frequency of minimizing the MSE and RMSE, respectively, over realized forecasts.

It is worth noting that our paper differs in several important aspects with respect to the few previous studies that address the question regarding the economic evaluation of yield curve forecasts. First, the data set used in this paper carries interesting characteristics, since it refers to a different marketplace, is sampled on a daily basis, and consists of high-liquidity fixed income future contracts that resembles zero-coupon bonds. Second, Carriero et al. (2012) consider trading strategies based only on one-step-ahead forecasts, whereas in our paper we consider the Sharpe ratios from optimal portfolios based on multi-step-ahead forecasts including 1-week-, 1-month-, 2-month-, and 3-month-ahead forecasts. Third, Carriero et al. (2012) obtain optimal mean-variance portfolio weights considering a single value for the risk aversion coefficient, whereas we provide results considering alternative levels of risk tolerance. Fourth, neither Carriero et al. (2012) nor Xiang & Zhu (2013) provide results regarding the statistical differences in the Sharpe ratios of the proposed approach with respect to the benchmark. We, on the other hand, employ a robust test for the Sharpe ratio based on the bootstrap procedure of Politis & Romano (1994). Fifth, none of the existing references focus on the performance of forecast combinations. In this paper, however, we provide a comprehensive evaluation by implementing five alternative forecast combination schemes, and check their performance in terms of statistical accuracy and
economic relevance.

Our empirical application is based on a large data set of constant-maturity future contracts of the Brazilian Inter Bank Deposit Future Contract (DI-futuro) which is equivalent to a zero-coupon bond and is highly liquid (293 million contracts worth US$ 15 billion traded in 2010). The market for DI-futuro contracts is one of the most liquid interest rate market in the world. Many banks, insurance companies, and investors use DI-futuro contracts as investment and hedging instruments. The data set considered in the paper contains daily observations of DI-futuro contracts traded on the Brazilian Mercantile and Futures Exchange (BM&F) with fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 42 and 48 months. To obtain 1-week-, 1-month-, 2-month-, and 3-month-ahead forecasts for each of the maturities available in the data set, we use individual forecasting models as well as alternative forecast combination schemes. The results show that combined forecasts consistently outperform individual models in terms of lower forecasting errors, specially for longer forecasting horizons. Moreover, we also observe that the Sharpe ratios of the mean-variance portfolios built upon combined forecasts are substantially and statistically higher than those obtained with the benchmark models. This result is also robust to the level of risk tolerance and to the portfolio re-balancing frequency. Finally, our results also suggest that, as long as yield curve forecasts are concerned, the differences in forecasting performance among candidate models based on statistical criteria are also economic meaningful as they generate optimal fixed income portfolios with improved risk-adjusted returns.

The paper is organized as follows. In Section 2 we describe the methods used to obtain forecasts for the yield rates, including both individual forecast models as well as forecast combinations. Next, in Section 3 we discuss the methodology used to evaluate forecast, both in terms of statistical criteria and in terms of economic relevance. Section 4 brings an empirical application. Section 5 concludes.

2 Methods used to forecast the yield curve

In this Section, we describe the methods used to forecast the yield curve. These methods are based on individual forecast models as well as alternative forecast combinations.

2.1 Random walk model

The main benchmark model adopted in the paper is the random walk (RW), whose $t + h$-step-ahead forecasts for an yield of maturity $\tau$ are given by:
\[ y_t(\tau_i) = y_{t-1}(\tau_i) + \varepsilon_t(\tau_i), \quad \varepsilon_t(\tau_i) \sim \mathcal{N}(0, \sigma^2(\tau_i)). \]  

(1)

In the RW, a \( h \)-step-ahead forecast for yield \( \hat{y}_{t+h}(\tau_i) \) is simply equal to the most recently observed value \( y_t(\tau_i) \). This model is a good benchmark for judging the relative prediction power of other models, since yields are usually nonstationary or nearly nonstationary. Thus, in practice, it is difficult to beat the RW in terms of out-of-sample forecasting accuracy. Many other studies that consider interest rate forecasting have shown that consistently outperforming the random walk is difficult (see, for example, Duffee, 2002; Ang & Piazzesi, 2003; Diebold & Li, 2006; Hördahl et al., 2006; Mönch, 2008).

2.2 Univariate autoregressive model

The second model that we consider is a first-order univariate autoregressive model, AR(1), which allows for mean-reversion. Yields at each maturity are predicted using an AR(1) that is estimated on the available data for that maturity:

\[ y_t(\tau) = \alpha + \beta y_{t-1}(\tau) + \varepsilon_t, \]  

(2)

for maturity \( \tau \). The 1-step ahead forecast is produced as \( \hat{y}_{t+1}(\tau) = \hat{\alpha} + \hat{\beta} y_{t-1}(\tau) \) the forecasts for \( h \)-step ahead horizon are obtained as:

\[ \hat{y}_{t+h|t}(\tau) = (1 + \hat{\beta} + \hat{\beta}^2 + \ldots + \hat{\beta}^{h-1})\hat{\alpha} + \hat{\beta}^h y_t(\tau). \]

2.3 Vector autoregressive model

The third and final competitor model is a first-order unrestricted vector autoregressive model, VAR(1), for yield levels. VAR forecasts are produced along the same lines of univariate AR forecasts. Specifically, VAR models allow for using the history of other maturities as additional information on top of any maturity’s own history. The regression model is:

\[ y_t = A + By_{t-1} + \varepsilon_t, \]  

(3)

where \( y_t = (y_t(\tau_1), y_t(\tau_2), \ldots, y_t(\tau_N))^\prime \). The 1-step ahead forecast is produced as \( \hat{y}_t = \hat{A} + \hat{B}y_{t-1} \), while the \( h \)-step ahead forecasts are obtained as:

\[ \hat{y}_{t+h|t} = (I + \hat{B} + \hat{B}^2 + \ldots + \hat{B}^{h-1})\hat{A} + \hat{B}^h y_t. \]  

(4)

As argued by de Pooter et al. (2010), a well-known drawback of using an unrestricted VAR model for yields
is the large number of parameters that need to be estimated. Since we want to construct forecasts for thirteen maturities, this results in a substantial number of parameters that need to be estimated.

2.4 Bayesian vector autoregressive model

The Bayesian vector autoregressive (BVAR) model employed in this paper is similar to the one adopted in Carriero et al. (2012). Consider the vector autoregressive model in (3):

\[ y_t = A + By_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim i.i.d. N(0, \Sigma), \tag{5} \]

where \( \varepsilon_t \) is a vector of Gaussian disturbances.

One well known stylized facts of interest rates is that yields, regardless of maturity, are very persistent processes; see Table 1. Existing evidence suggest that even a simple RW model or AR model can deliver good forecasts of the yields. Therefore, it is reasonable to think a priori that each of the yields in (7) obeys a univariate AR with high persistence, or equivalently, that the expected value of the matrix \( B \) is \( E[B] = \delta \times I \).

We also need to assess how strong is the belief we have in such a prior, i.e. we need to set a variance around the prior mean. It is assumed that \( B \) is conditionally normal, with first and second moments given by:

\[ E[B^{ij}] = \begin{cases} 
\delta_i & \text{if } i = j \\
0 & \text{if } i \neq j 
\end{cases}, \quad \text{Var}[B^{ij}] = \frac{\theta \sigma_i^2}{\sigma_j^2}, \tag{6} \]

where \((B^{ij})\) denotes the element in position \((i,j)\) in the matrix \( B \), and where the covariance among the coefficients in \( B \) are zero. The shrinkage parameter \( \theta \) measures the tightness of the prior: when \( \theta \to 0 \) the prior is imposed exactly and the data do not influence the estimates, while as \( \theta \to \infty \) the prior becomes loose and the prior information does not influence the estimates, which will approach the standard OLS estimates. The factor \( \sigma_i^2/\sigma_j^2 \) is a scaling parameter which accounts for the different scale and variability of the data. To set the scale parameters \( \sigma_i^2 \) we follow common practice (see Carriero et al., 2012) and set it equal to the variance of the residuals from a univariate autoregressive model for the variables.

The prior specification is completed by assuming a diffuse normal prior on \( A \) and an inverted Wishart prior for the matrix of disturbances \( \Sigma \sim iW(v_0, S_0) \), where \( v_0 \) and \( S_0 \) are the prior scale and shape parameters, and are set such that the prior expectation of \( \Sigma \) is equal to a fixed diagonal residual variance \( \Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2) \).

Banbura et al. (2010) suggest to set \( \delta_i = 1 \) for all \( i \), reflecting the belief that all the variables are characterized by high persistence. We believe it is reasonable to think a priori that each of the yields obeys a univariate AR
with high persistence, or equivalently, $\mathbb{E}[B] = 0.99 \times I$.

We can re-write more compactly the VAR as:

$$ Y = X F + E, \quad (7) $$

where $Y = [y_1, \ldots, y_T]$ is a $T \times N$ matrix containing all the data points in $y_t$, $X = [1, Y_{-1}]$ is a $T \times M$ matrix containing a vector of ones (1) in the first columns and one lag of $Y$ in the remaining columns, $F = [A, B]'$ is a $M \times N$ matrix, and $E = [\varepsilon_1, \ldots, \varepsilon_T]'$ is a $T \times N$ matrix of disturbances. As only one lag is considered we have $M = N + 1$. The Normal inverted Wishart prior has the form:

$$ \text{vec} (F) | \Sigma \sim N (\text{vec} (F_0), \Sigma \otimes \Omega_0), \quad \Sigma \sim iW(S_0, \alpha_0) \quad (8) $$

where the prior parameters $F_0, \Omega_0, S_0$ and $\alpha_0$ are chosen so that prior expectations and variances of $F$ coincide with those implied by equation (6), $\mathbb{E}[F] = F_0$ e $\text{var}[F] = \Sigma \otimes \Omega_0$, where $\Sigma$ is the variance matrix of the disturbances and the elements of $\Omega_0$ are given by $\text{var} [B^{ij}]$ in (6). As the used $iW$ prior is conjugate, the conditional posterior distribution of this model is also Normal-Inverted Wishart:

$$ \text{vec} (F) | \Sigma \sim N (\bar{F}, \Omega^{-1}, S, \bar{\alpha}), \quad \Sigma | Y \sim IW(\bar{S}, \bar{\alpha}), \quad (9) $$

where the bar denotes that the parameters are those of the posterior distribution. Defining $\hat{F}$ and $\hat{E}$ as the OLS estimates, we have that $F = (\Omega_0^{-1} + X'X)^{-1} (\Omega_0^{-1}F_0 + X'Y)$, $\bar{\Omega} = (\Omega_0^{-1}F_0 + X'X)^{-1}$, $\bar{\pi} = \alpha_0 + T$ and $\bar{S} = \hat{F}'X'X\hat{F} + F_0'\Omega_0^{-1}F_0 + F_0' + \hat{E}'\hat{E} - \hat{F}'\bar{\Omega}^{-1}\hat{F}$.

In order to perform inference and forecasting one needs the full joint posterior distribution and the marginal distributions of the parameters $F$ and $\Sigma$. One could use the conditional posteriors in (9) as a basis of a Gibbs sampling algorithm that drawing in turn from the conditionals $\Sigma | Y$ and $F | \Sigma, Y$ would eventually produce a sequence of draws from the joint posterior $F | \Sigma, Y$ and the marginal posteriors $\Sigma | Y$ e $F | \Sigma, Y$, as well as of the posterior distribution of any function of these coefficients. If one is interested only in the posterior distribution of $F$ there is an alternative to simulation: by integrating out (9), the marginal posterior distribution of $F$ is a multivariate $t$:

$$ \text{vec} (F) | Y \sim N (\text{vec} (F), \bar{\Omega}^{-1}, \bar{S}, \bar{\alpha})^1. \quad (10) $$

1 For details see Carriero et al. (2012).
The expected value for this distribution is given by:

$$F = (\Omega^{-1} + X'X)^{-1} \left( \Omega^{-1}F_0 + X'Y \right)$$  \hspace{1cm} (11)

Recalling that $\hat{F}$ is the estimator, and using the normal equations $(X'X)^{-1}\hat{F} = X'Y$ we can rewrite this as:

$$\bar{F} = (\Omega_{0}^{-1} + X'X)^{-1} \left( \Omega_{0}^{-1}F_{0} + X'X\hat{F} \right),$$  \hspace{1cm} (12)

which shows that the posterior mean of $F$ is a weighted average of the OLS estimator and of the prior mean $\hat{F}_0$, with weights proportional to the inverse of their respective variances. In the presence of a tight prior ($\theta \to 0$) the posterior estimate will collapse to $\hat{F} = \hat{F}_0$, while with a diffuse prior $\theta \to \infty$ the posterior estimate will collapse to the unrestricted OLS estimate. Given the posterior mean $\bar{F} = [\hat{A}, \hat{B}]'$, it is straightforward to produce forecasts up to $h$ steps ahead simply by setting:

$$\hat{Y}_{t+h} = \hat{A} + \hat{B} \cdot \hat{Y}_{t+h-1},$$  \hspace{1cm} (13)

where $\hat{Y}_{t+h-1}$ is computed as in (4).

### 2.5 Dynamic factor models

Nelson & Siegel (1987) have shown that the term structure can be surprisingly well fitted at a particular point in time by a linear combination of three smooth functions. The Nelson-Siegel model is give by:

$$y(\tau) = \beta_1 + \beta_2 \left( \frac{1 - e^{-\lambda_1 \tau}}{\lambda_1 \tau} \right) + \beta_3 \left( \frac{1 - e^{-\lambda_2 \tau}}{\lambda_2 \tau} - e^{-\lambda_2 \tau} \right) + \epsilon_{\tau}$$  \hspace{1cm} (14)

where $\beta_1$ can be interpreted as the level of the yield curve, $\beta_2$ as its slope, and $\beta_3$ as its curvature. The parameter $\lambda$ determines the exponential decay of $\beta_2$ and of $\beta_3$.

Svensson (1994) proposed an extension of the original Nelson-Siegel model by adding an extra smooth function to improve the flexibility and fit of the model. The proposed by Svensson (1994) is

$$y(\tau) = \beta_1 + \beta_2 \left( \frac{1 - e^{-\lambda_{11} \tau}}{\lambda_{11} \tau} \right) + \beta_3 \left( \frac{1 - e^{-\lambda_{12} \tau}}{\lambda_{12} \tau} - e^{-\lambda_{12} \tau} \right) + \beta_4 \left( \frac{1 - e^{-\lambda_{22} \tau}}{\lambda_{22} \tau} - e^{-\lambda_{22} \tau} \right) + \epsilon_{\tau}.$$  \hspace{1cm} (15)

In an important contribution, Diebold & Li (2006) have introduced dynamics into the original Nelson-Siegel model, and showed that its dynamic version has good forecasting ability. The dynamic Nelson-Siegel model
(henceforth DNS) is given by:

\[ y_t(\tau) = \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda_1 \tau}}{\lambda_1 \tau} \right) + \beta_{3t} \left( \frac{1 - e^{-\lambda_2 \tau}}{\lambda_2 \tau} - e^{-\lambda_2 \tau} \right) + \epsilon_t(\tau). \]  

Similarly to what Diebold & Li (2006) have done for the Nelson-Siegel model, a dynamic version of Svensson model (hereafter DSV) can be written as:

\[ y_t(\tau) = \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda_1 \tau}}{\lambda_1 \tau} \right) + \beta_{3t} \left( \frac{1 - e^{-\lambda_2 \tau}}{\lambda_2 \tau} - e^{-\lambda_2 \tau} \right) + \beta_{4t} \left( \frac{1 - e^{-\lambda_3 \tau}}{\lambda_3 \tau} - e^{-\lambda_3 \tau} \right) + \epsilon_t(\tau). \]  

The fourth factor in the dynamic version of the Svensson model can be interpreted as a second curvature. Svensson (1994) argues that the additional factor provides a better in-sample fit, especially for a richer structure of yields, and therefore provides better estimations of forward rates.

The dynamic versions of both the Nelson-Siegel and Svensson models can be interpreted as dynamic factor models (see, for example, Diebold et al., 2006). More specifically, consider an \( N \times T \) matrix of observable yields. The observation at time \( t \) is denoted by \( y_t = (y_{1t}, \ldots, y_{Nt})' \), for \( t = 1, \ldots, T \), and \( y_{it} \) is \( i \)-th variable in the vector \( y_t \) at time \( t \). The dynamic factor models considered are of the form

\[ y_t = \Lambda(\lambda)f_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \Sigma), \quad t = 1, \ldots, T, \]  

where \( \Lambda(\lambda) \) is the \( N \times K \) matrix of factor loadings that depends on the decaying parameter \( \lambda \), \( f_t \) is a \( K \)-dimensional vector containing the coefficients \( \beta_{1t}, \ldots, \beta_{Kt} \) for \( K = \{3, 4\} \), \( \varepsilon_t \) is the \( N \times 1 \) vector of disturbances and \( \Sigma \) is an \( N \times N \) diagonal covariance matrix of the disturbances. The dynamic factors \( f_t \) are modeled by the following stochastic process:

\[ f_t = \mu + A f_{t-1} + \eta_t, \quad \eta_t \sim \text{NID}(0, \Omega), \quad t = 1, \ldots, T, \]  

where \( \mu \) is a \( K \times 1 \) vector of constants, \( A \) is the \( K \times K \) transition matrix, and \( \Omega \) is the conditional covariance matrix of disturbance vector \( \eta_t \), which are independent of the residuals \( \varepsilon_t \) \( \forall t \). Note that equations (18) and (19) characterize a linear and Gaussian state space model.

**Estimation of the dynamic factor models**

Given the state space formulation of the dynamic factor model presented in (18) and (19), the Kalman filter can be used to obtain the likelihood function via the prediction error decomposition, as well as filtered
estimates of the states and of their covariance matrices, (see Durbin & Koopman, 2012, for more details regarding the Kalman filter). However, the computational burden associated with the Kalman filter recursions depends crucially on the dimension of both the state and observation vectors. Moreover, the dimension of the observation vector \((N \times 1)\) is often much larger than that of the state vector \((K \times 1)\) in yield curve models. In these circumstances, Jungbacker & Koopman (2008) have shown that significant computational gains can be achieved by a simple transformation. First, define the \(N \times N\) and the \(K \times N\) matrices:

\[
A = \begin{bmatrix} A_L \\ A_H \end{bmatrix}, \quad A^L = C \Lambda(\lambda)' \Sigma^{-1},
\]

respectively, where \(C\) can be any \(K \times K\) invertible matrix, and \(A_H\) is chosen to guarantee that \(A^H\) is full rank and \(A^L \Sigma A^H = 0\), thus \(A^H \Lambda(\lambda) = 0\). Selecting \(C = (\Lambda(\lambda)' \Sigma^{-1} \Lambda(\lambda))^{-1}\) implies:

\[
Ay_t = \begin{pmatrix} A^L y_t \\ A^H y_t \end{pmatrix} = \begin{pmatrix} f_t \\ 0 \end{pmatrix} + \begin{pmatrix} A^L \varepsilon_t \\ A^H \varepsilon_t \end{pmatrix}, \quad \begin{pmatrix} A^L \varepsilon_t \\ A^H \varepsilon_t \end{pmatrix} \sim N \left( 0, \begin{bmatrix} C & 0 \\ 0 & \Sigma^H \end{bmatrix} \right).
\]

The law of motion of the factors in (19) is not affected by the transformation. The interesting point to note is that the part \(A^H y_t\) is neither dependent on \(f_t\), nor correlated with \(A^L y_t\), and therefore does not need to be considered for the estimation of the factors. This implies that the Kalman filter only needs to be applied to the low-dimensional subvector \(A^L y_t\) for signal extraction, generating large computational gains when \(N >> K\) (see Table 1 of Jungbacker & Koopman, 2008).

Denote \(l(y)\) the log-likelihood function of the untransformed model in (18) and (19), where \(y = (y_1', \ldots, y_T')'\). Evaluation of \(l(y)\) can also take advantage of the transformations presented above. Jungbacker & Koopman (2008) show that the log-likelihood of the untransformed model can be represented as

\[
l(y) = c + l(y^L) - \frac{T}{2} \log \left| \Sigma \right| - \frac{1}{2} \sum_{t=1}^{T} e_t' \Sigma^{-1} e_t,
\]

where \(c\) is a constant independent of both \(y\) and the parameters, \(y^L = (A^L y_1', \ldots, A^L y_T')'\), and \(e_t = y_t - \Lambda(\lambda) A^L y_t\). Note that computation of matrix \(A^H\) is not required at any point, as proved in Lemma 2 of Jungbacker & Koopman (2008).
2.6 Combined forecasts

An important result from the methodological literature on forecasting is that a linear combination of two or more forecasts may yield more accurate predictions than using only a single forecast (Granger, 1989; Newbold & Harvey, 2002; Aiolfi & Timmermann, 2006). Adaptive strategies for combining forecasts might also mitigate structural breaks and model misspecification and thus lead to more accurate forecasts (Newbold & Harvey, 2002; Pesaran & Timmermann, 2007). In particular, there is recent evidence that combining forecast of nested models can significantly improve forecasting precision upon forecasts obtained from single model specifications (Clark & McCracken, 2009).

Assuming we are combining forecasts from $M$ different forecast models, a combined forecast for a $h$-month horizon for the yield with maturity $\tau_i$ is given by

$$\hat{y}_{t+h|t,\tau_i} = \sum_{m=1}^{M} w_{t+h|t,m}^{(\tau_i)} \hat{y}_{t+h|t,m}^{(\tau_i)}$$

where $w_{t+h|t,m}^{(\tau_i)}$ denotes the weight assigned to the time-$t$ forecast from the $m^\text{th}$ model, $\hat{y}_{t+h|t,m}^{(\tau_i)}$. Most of the forecast combination schemes considered are adaptive, meaning that the forecasts included in $M : \{\hat{y}_{t+h|t,m}^{(\tau_i)}\}$ and/or corresponding weights $m^\text{th}$ are based on alternative selection criteria within a sub-sample of realized observations.

Note that since a forecaster would only have information available up to the forecast origin $\omega$, the sub-sample for forecast selection and computation of weights must contain data on or before that period. Thus, we start by setting equal weights to all forecasts until the selection of forecasts and weighting schemes could be based on the evaluation of realized forecast errors. This procedure guarantees that we use only information available up to a particular period $\omega$ to set weights of forecasts for period $\omega+h$. The following 5 alternative combination strategies $\mathcal{M} = \{FC-EW, FC-OLS, FC-RANK, FC-MSE, FC-RMSE\} = \{1, 2, \ldots, 5\}$ are considered:

1. **Equally weighted forecasts (FC-EW)**: Various studies have demonstrated that simple averaging of a multitude of forecasts works well in relation to more sophisticated weighting schemes (Newbold & Harvey, 2002; Clark & McCracken, 2009). Therefore, the first forecast combination method we consider assigns equal weights to the forecasts from all individual models, i.e. $w_{t+h|t,m}^{(\tau_i)} = 1/M$ for $m = 1, \ldots, M$. We denote the resulting combined forecast as Forecast Combination - Equally Weighted (FC-EW). As explained in Timmermann (2006), this approach is likely to work well if forecast errors from different models have similar variances and are highly correlated.
2. **Thick modeling approach with OLS weights (FC-OLS):** A study by Granger & Jeon (2004) proposes the so-called thick modeling approach (TMA) which consists of selecting the $z$-percent of the best forecasting models in the sub-sample period for model evaluation, according to the root mean square error (RMSE) criterion. We use the selection process of Granger and Jeon and subsequently compute weights by means of OLS regressions along with the constraint that the weights are all positive and sum up to one. The $z$-percent of top forecasts selected is set to 2 (i.e. about $z=35\%$).

3. **Rank-weighted combinations (FC-RANK):** The FC-RANK scheme, suggested by Aiolfi & Timmermann (2006), consists of first computing the RMSE of all models in the sub-sample period for evaluation. Defining $RANK_{t+h|t,m}^{-1}$ as the rank of the $m^{th}$ model based on its historical RMSE performance up to time $t$ for horizon $h$, the weight for the $m^{th}$ forecast is then calculated as:

$$\hat{w}_{t+h|t,m}(\tau_i) = RANK_{t+h|t,m}^{-1} / \sum_{m=1}^{M} RANK_{t+h|t,m}^{-1}.$$  

4. **Thick modeling approach with MSE-Frequency weights (FC-MSE):** This scheme consists of selecting models by means of the thick modelling approach and assigning to each $m^{th}$ forecast a weight equal to a model’s empirical frequency of minimizing the squared forecast error over realized forecasts. The weight for model $m$ is computed as:

$$\hat{w}_{t+h|t,m}(\tau_i) = \frac{1/MSE_{t+h|t,m}(\tau_i)}{\sum_{m=1}^{M} 1/MSE_{t+h|t,m}(\tau_i)}.$$  

5. **Thick modeling approach with RMSE-weights (FC-RMSE):** This scheme consists of selecting models by means of the thick modelling approach, then computing the RMSE of all selected models $m$ and setting:

$$\hat{w}_{t+h|t,m}(\tau_i) = \frac{1/RMSE_{t+h|t,m}(\tau_i)}{\sum_{m=1}^{M} 1/RMSE_{t+h|t,m}(\tau_i)}.$$  

3 **Forecast evaluation**

We describe in this Section the methodology used to evaluate the yield curve forecasts obtained from the econometric specifications discussed in Section 2. We first describe the statistical-based evaluation and then the assessment of the economic value of the forecasts based on the Sharpe ratios of the fixed income portfolios.
3.1 Statistical performance measures

In order to evaluate out-of-sample forecasts, we compute popular error metrics. Given a sample of $M$ out-of-sample forecasts for a $h$-period-ahead forecast horizon, we compute the root mean squared forecast error (RMSFE) for maturity $\tau_i$ and for model $m$ as follows:

$$
\text{RMSFE}_m(\tau_i) = \sqrt{\frac{1}{M} \sum_{t=1}^{M} (\hat{y}_{t+h|t,m}(\tau_i) - y_{t+h}(\tau_i))^2}
$$

(22)

where $y_{t+h}(\tau_i)$ is the yield for the maturity $\tau_i$ observed at time $t + h$, and $\hat{y}_{t+h|t,m}(\tau_i)$ is the corresponding forecasting made at time $t$.

Following Christoffersen & Diebold (1998), Hördahl et al. (2006), and de Pooter et al. (2010), we also summarize the forecasting performance of each model by computing the trace root mean squared forecast error (TRMSFE). For each forecast horizon, we compute the trace of the covariance matrix of the forecast errors across all $N$ maturities. Hence, lower TRMSFE indicate more accurate forecasts. The TRMSFE can be computed as

$$
\text{TRMSFE}_m(\tau_i) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{M} (\hat{y}_{t+h|t}(\tau_i) - y_{t+h}(\tau_i))^2}
$$

(23)

The drawback of using RMSFE and TRMSFE is that these are single statistics summarizing individual forecasting errors over an entire sample. Although often used, they do not give any insight as to where in the sample a particular model makes its largest and smallest forecast errors. Therefore, we also graphically analyze the cumulative squared forecast errors (CSFE) used in Welch & Goyal (2008) and de Pooter et al. (2010). These cumulative prediction errors series clearly depicts when a model outperforms or underperforms a given benchmark. The CSFE is given by

$$
\text{CSFE}_{m,T}(\tau_i) = \sum_{t=1}^{T} \left[ (\hat{y}_{t+h|t,\text{benchmark}}(\tau_i) - y_{t+h}(\tau_i))^2 - (\hat{y}_{t+h|t,m}(\tau_i) - y_{t+h}(\tau_i))^2 \right]
$$

(24)

In the case a model outperforms the benchmark, the $\text{CSFE}_{m,T}$ will be an increasing series. If the benchmark produces more accurate forecasts, then $\text{CSFE}_{m,T}$ will tend to be decreasing.

Finally, we use the Giacomini & White (2006) test to assess whether the forecasts of two competing models are statistically different. The Giacomini-White (GW) test is a test of conditional forecasting ability and is constructed under the assumption that forecasts are generated using a moving data window. This is a test of equal forecasting 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. The test is based on the loss differential \( d_{m,t} = (e_{rw,t})^2 - (e_{m,t})^2 \), where \( e_{mt} \) is the forecast error of model \( m \) at time \( t \). We assume that the loss function is quadratic but it can be replaced by other loss functions depending on the goal of forecast. The null hypothesis of equal forecasting accuracy can be written as

\[ H_0 : E[d_{m,t+h} | \delta_{m,t}] = 0, \]  

(25)

where \( \delta_{m,t} \) is a \( p \times 1 \) vector of test functions or instruments and \( h \) is the forecast horizon. If a constant is used as instrument, the test can be interpreted as an unconditional test of equal forecasting accuracy. The \( GW \) test statistic \( GW_{m,t} \) can be computed as the Wald statistic:

\[
GW_{m,n} = n \left( \sum_{t=\omega+1}^{n-h} \delta_{m,t} d_{m,t+h} \right)' \hat{\Omega}^{-1} \left( \sum_{t=\omega+1}^{n-h} \delta_{m,t} d_{m,t+h} \right) \overset{d}{\to} \chi^2_{\text{dim}(\delta)} \]  

(26)

where \( \hat{\Omega}_n \) is a consistent \( HAC \) estimator for the asymptotic variance of \( \delta_{m,t} d_{m,t+h} \), and \( n = (T - \omega) \) the number of out-of-sample observations. Under the null hypothesis given in (25), the test statistic \( GW_{i,t} \) is asymptotically distributed as \( \chi^2_p \). Interestingly, by requiring that the forecasts be constructed using a small, finite, rolling window of observations, Giacomini & White (2006) are able to substantially weaken many of the most important assumptions needed for the results in West (1996), McCracken (2000) and McCracken (2007).

3.2 Assessing the economic value of forecasts

We follow Carriero et al. (2012) and Xiang & Zhu (2013) and analyze how useful the different forecasting models are when used as a basis for optimal fixed income portfolio allocation. There are alternative answers to the question of what an optimal portfolio means. The approach adopted in this paper is the mean-variance method proposed by Markowitz (1952), which is one of the milestones of modern finance theory. In this framework, individuals choose their allocations in risky assets based on the trade-off between expected return and risk. We consider the case of an investor that has a \( h \)-period investment horizon and re-balances her portfolio also on a \( h \)-period basis. The formulation of the mean-variance optimization problem is given by

\[
\begin{align*}
\text{minimize} & \quad w_t' \Sigma_r w_t - \frac{1}{2} w_t' \mu_{r,t+h} \\
\text{subject to} & \quad w_t' \iota = 1
\end{align*}
\]  

(27)
where \( \mu_{r|t-h} \) is the \( h \)-period-ahead vector of expected bond returns, \( \Sigma_r \) is the variance-covariance matrix of bond returns, \( w_t \) is the vector of portfolio weights at time \( t \) chosen at time \( t - h \), \( \iota \) is an \( N \times 1 \) vector of ones, and \( \delta \) is the coefficient of risk aversion. To improve the robustness of our empirical study, we solve the mean-variance optimization problem considering alternative values for the risk aversion coefficient \( \delta \). In particular, we solve the optimization problem in (27) for \( \delta = \{0.1, 0.5, 1.0, 2.0\} \). Finally, we focus in the case in which short-sales are restricted by adding to (27) a constraint to avoid negative weights, i.e. \( w_t \geq 0 \). Previous works show that adding such a restriction can substantially improve performance, especially reducing the turnover of the portfolio, see Jagannathan & Ma (2003), among others. In this case, the optimization problem in (27) is solved using numerical methods.

In order to perform fixed income portfolio optimization according to Markowitz’s mean-variance framework, one needs estimates of the vector of expected bond returns, \( \mu_{r|t-h} \), and of the covariance matrix of bond returns, \( \Sigma_r \); see (27). However, the forecasting models discussed in Section 2 are designed to model only bond yields. Nevertheless, Caldeira et al. (2012) show that it is possible to compute expected bond returns based on yield curve forecasts in a straightforward way. Taking into account that the price of a bond at time \( t \), \( P_t(\tau) \), is the present value at time \( t \) of \$1 receivable \( \tau \) periods ahead, and letting \( y_{t|t-h} \) denote the \( h \)-step-ahead forecast of its continuously compounded zero-coupon nominal yield-to-maturity, we obtain the vector of expected bond prices \( P_{t|t-h} \) for all maturities,

\[
P_{t|t-h} = \exp (-\tau \odot y_{t|t-h}),
\]

where \( \odot \) is the Hadamard (elementwise) multiplication and \( \tau \) is the vector of maturities. Using the log-return expression, we obtain

\[
r_{t|t-h} = \log \left( \frac{P_{t|t-h}}{P_{t-h}} \right) = \log P_{t|t-h} - \log P_{t-h} = -\tau \odot (y_{t|t-h} - y_{t-h}).
\]

Finally, taking expectation in (29), the vector of \( h \)-period-ahead expected bond returns, \( \mu_{r_{t|t-h}} \), is given by

\[
\mu_{r_{t|t-h}} = -\tau \odot y_{t|t-h} + \tau \odot y_{t-h}.
\]

Without loss of generality, we further assume that the model does not specify conditional variance dynamics, so that the conditional variance of \( r_{t|t-h} \) simply equals the unconditional variance-covariance matrix of the \( N \) bond returns. In this sense, \( \text{Var}_t[r_{t|t-h}] = \Sigma_r \).

The performance of optimal mean-variance portfolios is evaluated in terms of Sharpe ratio (SR), which
is defined as the ratio of the realized portfolio returns over its standard deviation. In order to assess the relative performance of the optimal mean-variance portfolios, we consider as benchmark policy the mean-variance portfolios obtained with estimates of expected bond returns based on the random walk specification discussed in Section 2. Even though Carriero et al. (2012) and Xiang & Zhu (2013) assessed the economic value of yield curve forecasts, neither of them measured the statistical significance of these differences. In order to determine if differences between Sharpe ratios are statistically different, we use the stationary bootstrap of Politis & Romano (1994) with \( B = 1000 \) bootstrap samples and block size \( b = 5 \). We use the methodology suggested in Ledoit & Wolf (2008, Note 3.2) to compute the resulting bootstrap \( p \)-values.

4 Empirical application

4.1 Data

The data set analyzed consists of yields of Brazilian Inter Bank Deposit Future Contract (DI-futuro)\(^2\), which is one of the largest fixed-income markets among emerging economies, collected on a daily basis. We use time series of daily closing yields of the DI-futuro contracts. In practice, contracts with all maturities are not observed on a daily basis. Therefore, based on the observed rates for the available maturities, the data were converted into fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 42 and 48 months, using a spline based method. More specifically, we fit the curve of zero-coupon yields with a cubic spline interpolation \( \hat{y}(\tau, \Psi) \), where \( \Psi \) is the vector of spline coefficients, \( \tau \) is the maturity, and \( \tau_{\text{min}} \leq \tau \leq \tau_{\text{max}} \), with \( \tau_{\text{min}} \) and \( \tau_{\text{max}} \) denoting the nearest and farthest maturities, respectively. We define \( P_i(\tau), i = 1, \ldots, N \), to be the time-\( t \) market price of a bond with maturity \( \tau_i \) and define \( \hat{P}_i(\tau) \) to be the price of the same bond computed by discounting its coupon and principal payments at the discount rate \( \hat{y} \). Next, we choose \( \Psi \) to solve the problem

\[
\min_{\Psi} \left( \sum_{i=1}^{N} \left( P_i(\tau) - \hat{P}_i(\tau) \right)^2 + \int_{\tau_{\text{min}}}^{\tau_{\text{max}}} \lambda(\tau)\hat{y}''(\tau, \Psi)^2 d\tau \right). \tag{31}
\]

This approach is also employed by Dai et al. (2007) and Andersen & Benzoni (2010), except that we fit the smoothed cubic spline directly on the zero-coupon yields curve, thus similar to McCulloch (1975) and Andersen & Benzoni (2010), while Dai et al. (2007) fit the smoothed spline on the forward rates curve.

The DI-futuro contract with maturity \( \tau \) is a zero-coupon future contract in which the underlying asset is the DI-futuro interest rate accrued on a daily basis, capitalized between trading period \( t \) and \( \tau \).\(^3\) The

\(^2\)The Brazilian Mercantile and Futures Exchange (BM&F) is the entity that offers the DI-futuro contract and determines the number of maturities with authorized contracts.

\(^3\)The DI-futuro rate is the average daily rate of Brazilian interbank deposits (borrowing/lending), calculated by the Clearinghouse
contract value is set by its value at maturity, $R$100,000.00, discounted according to the accrued interest rate negotiated between the seller and the buyer. In 2010 the DI-futuro market traded a total of 293 million contracts corresponding to US$ 15 billion. The DI-futuro contract is very similar to the zero-coupon bond, except for the daily payment of margin adjustments. The data set contains maturities with highest liquidity for January 2006 to December 2012, yielding a total of $T = 1488$ daily observations. The data source is the Brazilian Mercantile and Futures Exchange (BM&F)

Table 1 reports descriptive statistics for the Brazilian interest rate yield curve based on the DI-futuro market. For each time series we report the mean, standard deviation, minimum, maximum and the lag-1 sample autocorrelation. The summary statistics confirm some stylized facts common to yield curve data: the sample average curve is upward sloping and concave, volatility is decreasing with maturity, and autocorrelations are very high.

Figure 1 displays a three-dimensional plot of the data set and illustrates how yield levels and spreads vary substantially throughout the sample. The plot also suggests the presence of an underlying factor structure. Although the yield series vary heavily over time for each of the maturities, a strong common pattern in the 14 series is apparent. For most months, the yield curve is an upward sloping function of time to maturity. For example, last year of the sample is characterized by rising interest rates, especially for the shorter maturities, which respond faster to the contractionary monetary policy implemented by the Brazilian Central Bank in the first half of 2010. It is clear from Figure 2 that not only the level of the term structure fluctuates over time but also its slope and curvature. The curve takes on various forms ranging from nearly flat to (inverted) S-type shapes.

4.1.1 Implementation details

The forecasting exercise is performed in pseudo real time, i.e. we never use information which is not available at the time the forecast is made. For computing our results we use a rolling estimation window of 500 daily observations (2 years). We produce forecasts for 1-week, 1-month, 2-month, and 3-month ahead. The choice of a rolling scheme is suggested by two reasons. First, it is a natural way to avoid problems of instability, see

for Custody and Settlements (CETIP) for all business days. The DI-futuro rate, which is published on a daily basis, is expressed in annually compounded terms, based on 252 business days. When buying a future DI-futuro contract for the price at time $t$ and keeping it until maturity $\tau$, the gain or loss is given by:

$$
100,000 \left( \frac{\prod_{i=1}^{\zeta(t,\tau)} (1 + y_i \frac{252}{365})^{\frac{365}{252}}}{(1 + ID^\tau)^{\frac{252}{365}}} - 1 \right),
$$

where $y_i$ denotes the DI-futuro rate, $(i - 1)$ days after the trading day. The function $\zeta(t, \tau)$ represents the number of working days between $t$ and $\tau$. 

18
Second, having fixed the number of observations used to compute the forecasts and therefore the resulting time series of the forecast errors allows the use of the Giacomini & White (2006) test for comparing forecast accuracy. Such a test is valid provided that the size of the estimation window is fixed. It is worth noting that we use iterated forecasts instead of direct forecasts for the multi-period ahead predictions. Marcellino et al. (2005) argue that iterated forecasts are more efficient when the model is correctly specified.

4.2 Results

4.2.1 Statistical evaluation

Table 2 report statistical measures of the out-of-sample forecasting performance of six alternative individual models and five combination schemes for four forecast horizons. The first line in each panel of the Table reports the value of TRMSFE and RMSFE (expressed in basis points) for the random walk model (RW), while all other lines reports statistics relative to the RW. Bold values indicate relative differences below one, which means that a particular model outperforms the random walk. In order to assess the statistical significance of these differences in forecast, we use the test of conditional predictive ability proposed by Giacomini & White (2006). The following model abbreviations are used in Table 2: AR(1) for the first-order univariate autoregressive model, VAR(1) for the first-order vector autoregressive model, BVAR refers to Bayesian VAR, DNS for the dynamic Nelson-Siegel model with a VAR specification for the factors, DSV for the dynamic Svensson model with a VAR specification for the factors. FC-EW and FC-OLS stand for forecast combinations based on equal weights and OLS-based weights, respectively, and FC-RANK for forecast combinations using rank-weighted combinations. FC-RMSE and FC-MSE stand for the forecast combinations based on the thick modeling approach with RMSE-frequency and MSE-frequency weights, respectively.

Panels (a) and (b) of Table 2 show results of 1-day-ahead and 1-week-ahead forecasts, respectively. As documented in the literature, it is very difficult to outperform the RW for short horizons, since the near unit root behavior of the yields seems to dominate and model-based information add little (Diebold & Li, 2006; de Pooter et al., 2007; Nyholm & Vidova-Koleva, 2012; Xiang & Zhu, 2013). Nevertheless, we observe that some individual models and combination schemes are able to outperform the RW mainly for short-term maturities. For instance, in the case of 1-week-ahead forecasts we observe that the BVAR and DSV specifications deliver lower TRMSFE in comparison to the RW. Moreover, we observe that all forecast combination schemes outperform the RW in terms of TRMSFE, and that the FC-RANK achieved the lowest TRMSFE among all individual models and combination schemes. We also find that the forecast performance of both individual models and combination schemes deteriorate as maturity grows longer.
The results for the 1-month-ahead forecasts are shown in panel (c) of Table 2. We observe that for short-term maturities several individual models and all combination schemes outperform the RW in terms of lower RMSFE. The best overall performance is achieved by the DSV model, as it outperforms the RW for all maturities. Moreover, this model delivers the lowest TRMSFE among all individual specifications and combination schemes.

Panel (d) of Table 2 reports the results for the 2-month-ahead forecasts. We find that for maturities lower than 18 months, all forecast combination schemes outperform the RW in terms of lower RMSE. Similar as in the 1-month-ahead forecasts, the best overall performance is achieved by the DSV model, as it achieved a much lower TRMSFE in comparison to the RW and also in comparison to the remaining competing specifications. Similar results are observed in the panel (e) of Table 2 which reports the results for the 3-month-ahead forecasts. In this case, the best performing specification in terms of lower TRMSFE is also the DSV yield curve model.

To explore the accuracy of the forecasts in different time intervals, we follow Welch & Goyal (2008) and plot the difference in cumulative squared forecast errors between each of the prediction models and the RW along the out-of-sample evaluation period. Figures 2 to 5 display CSFE’s for each forecast horizon. Each line in the graph represents a different model and shows how a particular model performs relative to the random walk benchmark. In particular, an increasing CSFE indicates outperformance whereas a decreasing CSFE indicates underperformance with respect to the RW.

During the financial turmoil of the 2008-2009 period, interest rates initially went up and then declined sharply from roughly 14% to a level of 8.5% for the short rate accompanied by a substantial widening of spreads between long and short rates (see Figure 1). The CSFE graphs allow us to examine in detail how different models perform during this period on a day-to-day basis. The CSFE graphs reveal that most models perform poorly when the spread between long and short interest rates is high, however, for longer horizons some individual specifications, such as the DSV yield curve model, are able to maintain performance even in the crisis period. In most cases, the DSV model is able to beat the RW. The VAR(1) model does well for shorter maturities, while the AR(1) model shows a strikingly poor performance overall with decreasing CSFE.

The CSFE graphs also indicate that the forecast combination schemes consistently outperform the RW in the vast majority of the cases. We observe in the graphs that the outperformance of the combined forecasts is consistent throughout the entire out-of-sample period. This result suggests that the combination of forecasts results in an improvement in forecast accuracy with respect to the benchmark model, and also with respect to individual models.
4.2.2 Economic value of forecasts

In the previous subsection, we showed that alternative individual prediction models as well as forecasting combination schemes are able to deliver more accurate forecasts with respect to the benchmark when considering statistical criteria. We observe, however, that in some instances the improvement in forecasting performance (as indicated by lower forecasting errors) is small in magnitude. Therefore, a question that remains unanswered is whether or not this statistical gain is also economically meaningful.

Table 3 reports the Sharpe ratios of the mean-variance portfolios composed of Brazilian DI-futuro contracts. We observe that in vast majority of the instances, the mean-variance portfolios obtained with individual models and with forecast combinations achieve statistically higher Sharpe ratios in comparison to the mean-variance portfolios obtained with the random walk model. Moreover, this result is robust to the value of the risk aversion coefficient and to the portfolio re-balancing frequency. We also observe that the forecast combinations deliver higher Sharpe ratios in comparison to those obtained with individual models in most of the cases. For instance, the highest Sharpe ratio obtained under daily re-balancing is achieved by the FC-RMSE (1.366), whereas for the weekly, monthly, and bimonthly re-balancing the highest Sharpe ratio is achieved by the FC-RANK (1.336), FC-RANK (1.363), FC-OLS (0.951), respectively. These values are substantially higher than those obtained by the individual models, including the benchmark specification.

4.2.3 Discussion

The results discussed in the previous Sections reveal that the statistical and economic evaluation of forecast performance might be related, at least to some extent. We observe that those specifications that deliver lower forecasting errors (measured in terms of RMSFE and TRMSFE) also tend to deliver mean-variance portfolio with higher Sharpe ratios in comparison to the benchmark. However, how strong (or weak) is this relation? In order to address this question, we plot in Figure 6 a two-dimensional graph with the RMSFE in the $x$-axis and the Sharpe ratios in the $y$-axis for each forecast horizon. Moreover, we also report in the graph the value of the correlation coefficient between both variables.

Figure 6 indicates that the correlation coefficient between RMSFE and Sharpe ratios is negative for all forecasting horizons. This result corroborates our previous findings and suggests that lower forecasting errors is in fact associated to higher Sharpe ratios. In some instances, however, this relation is weaker since the correlation coefficient between RMSFE and Sharpe ratios is closer to zero. For example, the correlation coefficient for the 1-week-ahead is -0.116 whereas for the 2-month-ahead is -0.810. This finding suggest that in some cases statistical and economic evaluation can provide different answers about which candidate model is better. In
fact, the results discussed in Tables 2 and 3 seem to corroborate this finding. We observe in these tables that the ranking of specifications based on statistical performance can differ from the ranking based on economic criterion. As we discussed previously, the statistical evaluation in Table 2 indicate that the forecast combinations outperformed the remaining specification for shorter horizons, while for the longer horizons the DSV model deliver more accurate forecasts. However, the results for the economic evaluation in Table 2 reveal that forecast combinations tend to outperform the remaining specification in the vast majority of the cases. Therefore, the ranking of specifications may change when we look at statistical or economic criteria. In this sense, our evidence corroborate the previous results in the literature, such as those reported in Leitch & Tanner (1991), Carriero et al. (2012), and Cenesizoglu & Timmermann (2012).

5 Concluding remarks

This paper examines the statistical accuracy and economic value of modeling and forecasting the term structure of interest rates using forecast combinations. Combined forecasts has been extensively and successfully applied to forecast many economic time series (see, for example, Granger, 1989; Clemen, 1989; Granger & Jeon, 2004; Timmermann, 2006; Wallis, 2011). However, the literature on yield curve forecasting has focused on performance evaluation of individual forecast models. Moreover, the majority of yield curve studies considers only statistical evaluation of point forecasts, while the final goal of interest rate forecasts is not to minimize mean squared prediction error but to improve financial decision making.

Thus, we conduct a detailed forecast performance evaluation of several univariate and multivariate models based not only on statistical measures of accuracy, but also an economic evaluation using Sharpe ratios of optimal mean-variance fixed income portfolios constructed based upon forecasts from individual models and their alternative combinations. Differently from other yield curve studies, we use bootstrap methods to determine if differences between Sharpe ratios are statistically different.

The results reveal that the statistical and economic evaluation of forecast performance might be related. However, the correlations between RMSFE and Sharpe ratios are very low and often close to zero. Moreover, the ranking of specifications based on statistical performance can differ from the ranking based on economic criterion. More specifically, while the forecast combination schemes outperformed in terms of RMSFE the remaining specification only for shorter horizons, they outperform individual models in the vast majority of cases when Sharpe ratio was considered.
References


---

**Note:** The text is a list of references formatted according to a typical academic style. Each entry includes the author(s), year of publication, title of the work, and other relevant details such as the journal or book title, volume, issue, and page numbers. This list is likely to be part of a larger document, possibly a research paper or a dissertation, given the context of an academic reference list.
Table 1: Descriptive statistics for the term structure of interest rates

The Table reports summary statistics for DI-futuro yields over the period the sample period is January 2007 - December 2012 (1488 daily observations). We examine daily data, constructed using the cubic-spline method. Maturity is measured in months. We show for each maturity mean, standard deviation, minimum, maximum and a selection of autocorrelation (Acf, $\hat{\rho}(1)$, $\hat{\rho}(5)$, and $\hat{\rho}(21)$, respectively) and partial autocorrelation (Pacf, $\hat{\alpha}(2)$ and $\hat{\alpha}(5)$) coefficients.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
<th>Skew</th>
<th>Kurt</th>
<th>Acf $\rho(1)$</th>
<th>Acf $\rho(5)$</th>
<th>Acf $\rho(21)$</th>
<th>Pacf $\alpha(2)$</th>
<th>Pacf $\alpha(5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10.66</td>
<td>1.75</td>
<td>6.97</td>
<td>14.13</td>
<td>-0.337</td>
<td>2.219</td>
<td>0.998</td>
<td>0.987</td>
<td>0.936</td>
<td>-0.016</td>
<td>-0.012</td>
</tr>
<tr>
<td>3</td>
<td>10.66</td>
<td>1.79</td>
<td>7.02</td>
<td>14.52</td>
<td>-0.286</td>
<td>2.296</td>
<td>0.998</td>
<td>0.989</td>
<td>0.940</td>
<td>-0.017</td>
<td>-0.025</td>
</tr>
<tr>
<td>6</td>
<td>10.73</td>
<td>1.86</td>
<td>6.91</td>
<td>15.32</td>
<td>-0.201</td>
<td>2.472</td>
<td>0.998</td>
<td>0.989</td>
<td>0.939</td>
<td>-0.012</td>
<td>-0.023</td>
</tr>
<tr>
<td>9</td>
<td>10.85</td>
<td>1.90</td>
<td>6.86</td>
<td>16.04</td>
<td>-0.132</td>
<td>2.650</td>
<td>0.998</td>
<td>0.988</td>
<td>0.934</td>
<td>0.004</td>
<td>-0.010</td>
</tr>
<tr>
<td>12</td>
<td>11.00</td>
<td>1.92</td>
<td>6.87</td>
<td>16.40</td>
<td>-0.091</td>
<td>2.784</td>
<td>0.997</td>
<td>0.987</td>
<td>0.929</td>
<td>0.019</td>
<td>0.000</td>
</tr>
<tr>
<td>15</td>
<td>11.16</td>
<td>1.92</td>
<td>6.90</td>
<td>16.91</td>
<td>-0.053</td>
<td>2.907</td>
<td>0.997</td>
<td>0.985</td>
<td>0.923</td>
<td>0.024</td>
<td>0.003</td>
</tr>
<tr>
<td>18</td>
<td>11.31</td>
<td>1.88</td>
<td>7.00</td>
<td>17.12</td>
<td>-0.046</td>
<td>3.002</td>
<td>0.997</td>
<td>0.984</td>
<td>0.917</td>
<td>0.020</td>
<td>0.002</td>
</tr>
<tr>
<td>21</td>
<td>11.42</td>
<td>1.85</td>
<td>7.14</td>
<td>17.26</td>
<td>-0.042</td>
<td>3.118</td>
<td>0.997</td>
<td>0.983</td>
<td>0.912</td>
<td>0.023</td>
<td>-0.002</td>
</tr>
<tr>
<td>24</td>
<td>11.52</td>
<td>1.81</td>
<td>7.32</td>
<td>17.44</td>
<td>-0.032</td>
<td>3.243</td>
<td>0.996</td>
<td>0.982</td>
<td>0.907</td>
<td>0.020</td>
<td>-0.001</td>
</tr>
<tr>
<td>27</td>
<td>11.60</td>
<td>1.77</td>
<td>7.49</td>
<td>17.62</td>
<td>-0.004</td>
<td>3.375</td>
<td>0.996</td>
<td>0.981</td>
<td>0.901</td>
<td>0.015</td>
<td>-0.002</td>
</tr>
<tr>
<td>30</td>
<td>11.66</td>
<td>1.74</td>
<td>7.65</td>
<td>17.78</td>
<td>0.028</td>
<td>3.504</td>
<td>0.996</td>
<td>0.980</td>
<td>0.895</td>
<td>0.010</td>
<td>-0.006</td>
</tr>
<tr>
<td>36</td>
<td>11.75</td>
<td>1.66</td>
<td>7.91</td>
<td>17.83</td>
<td>0.080</td>
<td>3.731</td>
<td>0.995</td>
<td>0.978</td>
<td>0.887</td>
<td>0.015</td>
<td>-0.013</td>
</tr>
<tr>
<td>42</td>
<td>11.83</td>
<td>1.59</td>
<td>8.13</td>
<td>17.93</td>
<td>0.104</td>
<td>3.937</td>
<td>0.995</td>
<td>0.976</td>
<td>0.880</td>
<td>0.017</td>
<td>-0.006</td>
</tr>
<tr>
<td>48</td>
<td>11.89</td>
<td>1.56</td>
<td>8.29</td>
<td>18.00</td>
<td>0.181</td>
<td>4.156</td>
<td>0.995</td>
<td>0.975</td>
<td>0.873</td>
<td>0.018</td>
<td>-0.011</td>
</tr>
</tbody>
</table>

| Level    | 12.41 | 1.42    | 9.34 | 18.99| 0.673  | 5.185 | 0.992         | 0.957         | 0.800         | -0.064         | -0.014         |
| Slope    | -1.92 | 1.95    | -6.37| 2.84 | 0.267  | 2.814 | 0.995         | 0.974         | 0.866         | -0.039         | -0.036         |
| Curvature| -1.00 | 3.94    | -9.48| 8.60 | 0.120  | 2.192 | 0.997         | 0.978         | 0.903         | -0.160         | -0.011         |
Table 2: Relative (Trace)-Root Mean Squared Forecast Errors

The Table reports the relative root mean squared forecast errors (RMSFE) and trace RMSFE (TRMSFE) relative to the random walk model obtained by using individual yield models and different forecast combination methods, for the 1-day, 1-week, 1-month, 2-month, and 3-month forecast horizons. The evaluation sample is 2009:1 to 2012:12 (988 out-of-sample forecasts). The first line in each panel of the table reports the value of RMSFE and TRMSFE (expressed in basis points) for the random walk model (RW), while all other lines reports statistics relative to the RW. The following model abbreviations are used in the table: AR(1) for the first-order univariate autoregressive model, VAR(1) for the first-order vector autoregressive model, BVAR refers to Bayesian VAR, DNS for the dynamic Nelson-Siegel model with a VAR specification for the factors, DSV for the dynamic Svensson model with a VAR specification for the factors. FC-EW and FC-OLS stand for forecast combinations based on equal weights and OLS-based weights, respectively, and FC-RANK for forecast combinations using rank-weighted combinations. FC-RMSE and FC-MRMSE stand for the forecast combinations base on the thick modeling approach with RMSE-weights and MSE-Frequency weights, respectively. Numbers smaller than one (shown in ** bold) indicate that models outperform the RW, whereas numbers larger than one indicate underperformance. The stars on the right of the cell entries indicate the level at which the Giacomini and White (2006) test rejects the null of equal forecasting accuracy (*, and ** mean respectively rejection at 10%, and 5% level).

<table>
<thead>
<tr>
<th>Models</th>
<th>TRMSFE</th>
<th>1-Month</th>
<th>3-Month</th>
<th>6-Month</th>
<th>9-Month</th>
<th>12-Month</th>
<th>15-Month</th>
<th>18-Month</th>
<th>21-Month</th>
<th>24-Month</th>
<th>27-Month</th>
<th>30-Month</th>
<th>36-Month</th>
<th>42-Month</th>
<th>48-Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW</td>
<td>0.074</td>
<td>0.038</td>
<td>0.040</td>
<td>0.052</td>
<td>0.060</td>
<td>0.066</td>
<td>0.071</td>
<td>0.076</td>
<td>0.079</td>
<td>0.081</td>
<td>0.084</td>
<td>0.086</td>
<td>0.088</td>
<td>0.091</td>
<td>0.094</td>
</tr>
<tr>
<td>AR(1)</td>
<td>1.002</td>
<td>0.992</td>
<td>0.994</td>
<td>0.998</td>
<td>1.000</td>
<td>1.001</td>
<td>1.002*</td>
<td>1.002*</td>
<td>1.003*</td>
<td>1.003</td>
<td>1.003</td>
<td>1.003</td>
<td>1.003</td>
<td>1.003</td>
<td>1.003</td>
</tr>
<tr>
<td>VAR(1)</td>
<td>1.043</td>
<td>0.905**</td>
<td>0.931**</td>
<td>0.993</td>
<td>1.020*</td>
<td>1.030*</td>
<td>1.046**</td>
<td>1.048**</td>
<td>1.056**</td>
<td>1.057**</td>
<td>1.054**</td>
<td>1.052**</td>
<td>1.052**</td>
<td>1.054**</td>
<td>1.054**</td>
</tr>
<tr>
<td>BVAR</td>
<td>0.999</td>
<td>0.907**</td>
<td>0.927**</td>
<td>0.973**</td>
<td>0.990</td>
<td>0.998</td>
<td>1.002</td>
<td>1.004</td>
<td>1.005</td>
<td>1.004</td>
<td>1.004</td>
<td>1.005</td>
<td>1.005</td>
<td>1.005</td>
<td>1.005</td>
</tr>
<tr>
<td>DNS</td>
<td>1.638</td>
<td>2.633**</td>
<td>0.917**</td>
<td>1.975**</td>
<td>2.046**</td>
<td>1.841**</td>
<td>1.569**</td>
<td>1.112**</td>
<td>1.021**</td>
<td>1.065**</td>
<td>1.082**</td>
<td>1.056**</td>
<td>1.071</td>
<td>1.016</td>
<td>1.102**</td>
</tr>
<tr>
<td>DSV</td>
<td>1.196</td>
<td>2.266**</td>
<td>0.913**</td>
<td>1.234**</td>
<td>1.033**</td>
<td>1.076**</td>
<td>1.086**</td>
<td>1.056**</td>
<td>1.031**</td>
<td>1.056**</td>
<td>1.074**</td>
<td>1.043*</td>
<td>1.007</td>
<td>1.067**</td>
<td></td>
</tr>
<tr>
<td>FC-EW</td>
<td>1.023</td>
<td>1.428**</td>
<td>0.927**</td>
<td>1.075**</td>
<td>1.101**</td>
<td>1.066**</td>
<td>1.027**</td>
<td>1.006</td>
<td>1.002</td>
<td>0.999</td>
<td>0.998</td>
<td>0.995</td>
<td>0.997</td>
<td>1.000</td>
<td>1.011</td>
</tr>
<tr>
<td>FC-OLS</td>
<td>0.996</td>
<td>0.909**</td>
<td>0.914**</td>
<td>0.977**</td>
<td>0.989**</td>
<td>0.996**</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.001</td>
<td>1.001</td>
</tr>
<tr>
<td>FC-RANK</td>
<td>0.996</td>
<td>0.917**</td>
<td>0.915**</td>
<td>0.983**</td>
<td>0.991**</td>
<td>0.995*</td>
<td>0.997</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.001</td>
<td>1.001</td>
</tr>
<tr>
<td>FC-RMSE</td>
<td>0.995</td>
<td>0.910**</td>
<td>0.916**</td>
<td>0.981**</td>
<td>0.99**</td>
<td>0.994**</td>
<td>0.997</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.001</td>
<td>1.001</td>
</tr>
<tr>
<td>FC-MRMSE</td>
<td>0.995</td>
<td>0.907**</td>
<td>0.917**</td>
<td>0.976**</td>
<td>0.988**</td>
<td>0.994</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.001</td>
<td>1.001</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Models</th>
<th>TRMSFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC-RMSE</td>
<td>0.963</td>
</tr>
<tr>
<td>FC-MRMSE</td>
<td>0.961</td>
</tr>
</tbody>
</table>

**PANEL C: Horizon = 1-month ahead**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>1.001</th>
<th>1.006</th>
<th>1.012</th>
<th>1.017</th>
<th>1.023</th>
<th>1.032</th>
<th>1.035</th>
<th>1.028</th>
<th>1.041</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC-EW</td>
<td>0.966</td>
<td>0.810**</td>
<td>0.787**</td>
<td>0.897**</td>
<td>0.939**</td>
<td>0.958**</td>
<td>0.975**</td>
<td>0.982</td>
<td>0.986</td>
<td>0.982</td>
<td>0.981</td>
<td>0.984</td>
<td>0.988</td>
<td>0.995</td>
<td>1.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC-OLS</td>
<td>0.969</td>
<td>0.759**</td>
<td>0.746**</td>
<td>0.885**</td>
<td>0.921**</td>
<td>0.941**</td>
<td>0.999**</td>
<td>0.975</td>
<td>1.002</td>
<td>1.000</td>
<td>1.002</td>
<td>0.995</td>
<td>1.001</td>
<td>1.006</td>
<td>1.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC-RANK</td>
<td>0.959</td>
<td>0.785**</td>
<td>0.746**</td>
<td>0.886**</td>
<td>0.908**</td>
<td>0.934**</td>
<td>0.962**</td>
<td>0.974</td>
<td>0.976</td>
<td>1.001</td>
<td>1.002</td>
<td>0.971*</td>
<td>0.982</td>
<td>1.006</td>
<td>0.997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC-RMSE</td>
<td>0.963</td>
<td>0.757**</td>
<td>0.751**</td>
<td>0.882**</td>
<td>0.925**</td>
<td>0.948**</td>
<td>0.975**</td>
<td>0.976</td>
<td>0.984</td>
<td>0.999</td>
<td>1.000</td>
<td>0.980*</td>
<td>0.988</td>
<td>1.004</td>
<td>0.997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC-MRMSE</td>
<td>0.961</td>
<td>0.731**</td>
<td>0.752**</td>
<td>0.881**</td>
<td>0.924**</td>
<td>0.950**</td>
<td>0.966**</td>
<td>0.974**</td>
<td>0.98**</td>
<td>0.999**</td>
<td>1.000</td>
<td>0.975*</td>
<td>0.985</td>
<td>1.004</td>
<td>0.997</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PANEL D: Horizon = 2-month ahead**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>1.001</th>
<th>1.006</th>
<th>1.012</th>
<th>1.017</th>
<th>1.023</th>
<th>1.032</th>
<th>1.035</th>
<th>1.028</th>
<th>1.041</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC-EW</td>
<td>0.966</td>
<td>0.584**</td>
<td>0.655**</td>
<td>0.742**</td>
<td>0.796**</td>
<td>0.835**</td>
<td>0.862**</td>
<td>0.881**</td>
<td>0.891*</td>
<td>0.895</td>
<td>0.898</td>
<td>0.906</td>
<td>0.928</td>
<td>0.943</td>
<td>0.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC-OLS</td>
<td>0.969</td>
<td>0.447**</td>
<td>0.547**</td>
<td>0.655**</td>
<td>0.711**</td>
<td>0.756**</td>
<td>0.815**</td>
<td>0.843**</td>
<td>0.944</td>
<td>0.965</td>
<td>0.952</td>
<td>0.964</td>
<td>0.976</td>
<td>0.963</td>
<td>1.013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC-RANK</td>
<td>0.959</td>
<td>0.436**</td>
<td>0.541**</td>
<td>0.648**</td>
<td>0.706**</td>
<td>0.753**</td>
<td>0.799**</td>
<td>0.825</td>
<td>0.836</td>
<td>0.847</td>
<td>0.849</td>
<td>0.859</td>
<td>0.899</td>
<td>0.925</td>
<td>0.921</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC-RMSE</td>
<td>0.963</td>
<td>0.419**</td>
<td>0.533**</td>
<td>0.650**</td>
<td>0.715**</td>
<td>0.765**</td>
<td>0.819**</td>
<td>0.842**</td>
<td>0.881</td>
<td>0.888</td>
<td>0.892</td>
<td>0.901</td>
<td>0.923</td>
<td>0.941</td>
<td>0.952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC-MRMSE</td>
<td>0.961</td>
<td>0.417**</td>
<td>0.536**</td>
<td>0.654**</td>
<td>0.719**</td>
<td>0.770**</td>
<td>0.828**</td>
<td>0.841**</td>
<td>0.890*</td>
<td>0.902*</td>
<td>0.910</td>
<td>0.910</td>
<td>0.936</td>
<td>0.957</td>
<td>0.964</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PANEL E: Horizon = 3-month ahead**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>1.196</th>
<th>1.190</th>
<th>1.154</th>
<th>1.106</th>
<th>1.084</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC-EW</td>
<td>0.966</td>
<td>0.632**</td>
<td>0.710**</td>
<td>0.785**</td>
<td>0.831**</td>
<td>0.860**</td>
<td>0.879**</td>
<td>0.895*</td>
<td>0.906*</td>
<td>0.913*</td>
<td>0.922*</td>
<td>0.934</td>
<td>0.955</td>
</tr>
<tr>
<td>FC-OLS</td>
<td>0.969</td>
<td>0.463**</td>
<td>0.561**</td>
<td>0.648**</td>
<td>0.698**</td>
<td>0.735**</td>
<td>0.765**</td>
<td>0.884</td>
<td>0.903</td>
<td>0.907</td>
<td>0.913</td>
<td>0.930</td>
<td>0.933</td>
</tr>
<tr>
<td>FC-RANK</td>
<td>0.959</td>
<td>0.466**</td>
<td>0.560**</td>
<td>0.641**</td>
<td>0.729**</td>
<td>0.758**</td>
<td>0.801</td>
<td>0.819</td>
<td>0.830</td>
<td>0.842</td>
<td>0.879</td>
<td>0.904</td>
<td>0.911</td>
</tr>
<tr>
<td>FC-RMSE</td>
<td>0.963</td>
<td>0.455**</td>
<td>0.563**</td>
<td>0.658**</td>
<td>0.714**</td>
<td>0.754**</td>
<td>0.784**</td>
<td>0.844</td>
<td>0.858</td>
<td>0.865</td>
<td>0.874</td>
<td>0.900</td>
<td>0.913</td>
</tr>
<tr>
<td>FC-MRMSE</td>
<td>0.961</td>
<td>0.456**</td>
<td>0.565**</td>
<td>0.657**</td>
<td>0.713**</td>
<td>0.754**</td>
<td>0.829**</td>
<td>0.858*</td>
<td>0.877</td>
<td>0.878</td>
<td>0.883</td>
<td>0.920</td>
<td>0.928</td>
</tr>
</tbody>
</table>
The Table reports annualized Sharpe ratios of the optimal mean-variance portfolios using DI-futuro contracts with maturities equal to 1, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 36, 42 and 48 months. The following model abbreviations are used in the table: AR(1) for the first-order univariate autoregressive model, VAR(1) for the first-order vector autoregressive model, BVAR refers to Bayesian VAR, DNS for the dynamic Nelson-Siegel model with a VAR specification for the factors, DSV for the dynamic Svensson model with a VAR specification for the factors, FC-EW and FC-OLS stand for forecast combinations based on equal weights and OLS-based weights, respectively, and FC-RANK for forecast combinations using rank-weighted combinations. Asterisks indicate that the Sharpe ratio is statistically different with respect to that of the mean-variance portfolio obtained with the random walk specification at a significance level of 10%.

<table>
<thead>
<tr>
<th>Model</th>
<th>RW</th>
<th>AR(1)</th>
<th>VAR(1)</th>
<th>BVAR</th>
<th>DNS</th>
<th>DSV</th>
<th>FC-EW</th>
<th>FC-OLS</th>
<th>FC-RANK</th>
<th>FC-MSE</th>
<th>FC-RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Daily re-balancing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>δ = 0.01</td>
<td>0.493</td>
<td>0.542</td>
<td>0.693</td>
<td>1.021*</td>
<td>1.035*</td>
<td>0.691</td>
<td>0.879*</td>
<td>0.691</td>
<td>0.996*</td>
<td>1.126*</td>
</tr>
<tr>
<td>Risk Aversion</td>
<td>δ = 0.10</td>
<td>0.557</td>
<td>0.932*</td>
<td>0.595</td>
<td>0.912*</td>
<td>0.836</td>
<td>0.646</td>
<td>0.603</td>
<td>0.956*</td>
<td>1.048*</td>
<td>1.019*</td>
</tr>
<tr>
<td></td>
<td>δ = 0.50</td>
<td>0.657</td>
<td>1.079*</td>
<td>0.784</td>
<td>1.153*</td>
<td>0.555</td>
<td>0.547</td>
<td>0.830</td>
<td>1.177*</td>
<td>1.311*</td>
<td>1.248*</td>
</tr>
<tr>
<td></td>
<td>δ = 1.00</td>
<td>0.705</td>
<td>0.853</td>
<td>0.831</td>
<td>1.298*</td>
<td>0.757</td>
<td>0.671</td>
<td>0.883</td>
<td>1.240*</td>
<td>1.317*</td>
<td>1.297*</td>
</tr>
<tr>
<td></td>
<td>δ = 2.00</td>
<td>0.728</td>
<td>1.145*</td>
<td>1.013*</td>
<td>1.390*</td>
<td>0.867</td>
<td>0.801</td>
<td>0.930*</td>
<td>1.243*</td>
<td>1.524*</td>
<td>1.237*</td>
</tr>
<tr>
<td></td>
<td>Weekly re-balancing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>δ = 0.01</td>
<td>0.553</td>
<td>0.574</td>
<td>0.592</td>
<td>0.667</td>
<td>0.867*</td>
<td>0.898*</td>
<td>0.823*</td>
<td>0.703</td>
<td>0.794</td>
<td>0.826*</td>
</tr>
<tr>
<td>Risk Aversion</td>
<td>δ = 0.10</td>
<td>0.744</td>
<td>0.694</td>
<td>0.796</td>
<td>0.687</td>
<td>0.878</td>
<td>0.899</td>
<td>0.829</td>
<td>0.723</td>
<td>0.789</td>
<td>0.816</td>
</tr>
<tr>
<td></td>
<td>δ = 0.50</td>
<td>0.576</td>
<td>0.579</td>
<td>0.680</td>
<td>0.678</td>
<td>0.844*</td>
<td>0.802*</td>
<td>0.740</td>
<td>0.608</td>
<td>0.713</td>
<td>0.775</td>
</tr>
<tr>
<td></td>
<td>δ = 1.00</td>
<td>0.620</td>
<td>0.635</td>
<td>0.605</td>
<td>0.890*</td>
<td>0.943*</td>
<td>0.787</td>
<td>0.816</td>
<td>0.834*</td>
<td>0.835*</td>
<td>0.797</td>
</tr>
<tr>
<td></td>
<td>δ = 2.00</td>
<td>0.728</td>
<td>1.145*</td>
<td>1.013*</td>
<td>1.390*</td>
<td>0.867</td>
<td>0.801</td>
<td>0.930*</td>
<td>1.243*</td>
<td>1.524*</td>
<td>1.237*</td>
</tr>
<tr>
<td></td>
<td>Monthly re-balancing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>δ = 0.01</td>
<td>0.644</td>
<td>0.732</td>
<td>0.766</td>
<td>0.687</td>
<td>0.864*</td>
<td>0.872*</td>
<td>0.650</td>
<td>0.717</td>
<td>0.880*</td>
<td>0.699</td>
</tr>
<tr>
<td>Risk Aversion</td>
<td>δ = 0.10</td>
<td>0.658</td>
<td>0.750</td>
<td>0.718</td>
<td>0.667</td>
<td>0.881*</td>
<td>0.876*</td>
<td>0.675</td>
<td>0.753</td>
<td>0.890</td>
<td>0.730</td>
</tr>
<tr>
<td></td>
<td>δ = 0.50</td>
<td>0.634</td>
<td>0.741</td>
<td>0.690</td>
<td>0.678</td>
<td>0.965*</td>
<td>0.983*</td>
<td>0.760</td>
<td>0.826</td>
<td>0.935*</td>
<td>0.787</td>
</tr>
<tr>
<td></td>
<td>δ = 1.00</td>
<td>0.577</td>
<td>0.741</td>
<td>0.697</td>
<td>0.632</td>
<td>1.149*</td>
<td>1.199*</td>
<td>0.879*</td>
<td>0.871*</td>
<td>1.015*</td>
<td>0.914*</td>
</tr>
<tr>
<td></td>
<td>δ = 2.00</td>
<td>0.553</td>
<td>0.535</td>
<td>0.521</td>
<td>0.702</td>
<td>0.795*</td>
<td>0.775*</td>
<td>0.700</td>
<td>0.912*</td>
<td>0.711</td>
<td>0.901*</td>
</tr>
<tr>
<td></td>
<td>Bimonthly re-balancing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>δ = 0.01</td>
<td>0.533</td>
<td>0.535</td>
<td>0.521</td>
<td>0.702</td>
<td>0.795*</td>
<td>0.775*</td>
<td>0.700</td>
<td>0.912*</td>
<td>0.711</td>
<td>0.901*</td>
</tr>
<tr>
<td>Risk Aversion</td>
<td>δ = 0.10</td>
<td>0.543</td>
<td>0.543</td>
<td>0.521</td>
<td>0.702</td>
<td>0.805*</td>
<td>0.764*</td>
<td>0.667</td>
<td>0.928*</td>
<td>0.714</td>
<td>0.911*</td>
</tr>
<tr>
<td></td>
<td>δ = 0.50</td>
<td>0.562</td>
<td>0.554</td>
<td>0.521</td>
<td>0.718</td>
<td>0.865*</td>
<td>0.730</td>
<td>0.679</td>
<td>0.930*</td>
<td>0.850*</td>
<td>0.903*</td>
</tr>
<tr>
<td></td>
<td>δ = 1.00</td>
<td>0.782</td>
<td>0.756</td>
<td>0.736</td>
<td>0.850</td>
<td>0.817</td>
<td>0.717</td>
<td>1.106*</td>
<td>1.264*</td>
<td>1.363*</td>
<td>1.284*</td>
</tr>
<tr>
<td></td>
<td>δ = 2.00</td>
<td>0.551</td>
<td>0.554</td>
<td>0.618</td>
<td>0.625</td>
<td>0.794*</td>
<td>0.779*</td>
<td>0.798*</td>
<td>0.951*</td>
<td>0.816*</td>
<td>0.799*</td>
</tr>
<tr>
<td></td>
<td>Quarterly re-balancing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>δ = 0.01</td>
<td>0.647</td>
<td>0.662</td>
<td>0.934*</td>
<td>0.670</td>
<td>0.815</td>
<td>0.727</td>
<td>0.677</td>
<td>0.895</td>
<td>0.838*</td>
<td>0.867*</td>
</tr>
<tr>
<td>Risk Aversion</td>
<td>δ = 0.10</td>
<td>0.686</td>
<td>0.667</td>
<td>0.843</td>
<td>0.670</td>
<td>0.815</td>
<td>0.727</td>
<td>0.621</td>
<td>0.895*</td>
<td>0.837*</td>
<td>0.867*</td>
</tr>
<tr>
<td></td>
<td>δ = 0.50</td>
<td>0.827</td>
<td>0.670</td>
<td>0.707</td>
<td>0.850</td>
<td>0.825</td>
<td>0.727</td>
<td>0.605</td>
<td>0.908</td>
<td>0.838</td>
<td>0.866</td>
</tr>
<tr>
<td></td>
<td>δ = 1.00</td>
<td>0.915</td>
<td>1.019</td>
<td>0.403</td>
<td>0.694</td>
<td>0.870</td>
<td>0.707</td>
<td>0.651</td>
<td>0.911</td>
<td>0.829</td>
<td>0.810</td>
</tr>
<tr>
<td></td>
<td>δ = 2.00</td>
<td>0.694</td>
<td>0.598</td>
<td>0.688</td>
<td>0.672</td>
<td>0.883*</td>
<td>0.728</td>
<td>0.747</td>
<td>0.884*</td>
<td>0.833*</td>
<td>0.854*</td>
</tr>
</tbody>
</table>
Figures

Figure 1: Evolution of the yield curve

Note: The figure plots the evolution of term structure of interest rates (based on DI-futuro contracts) for the time horizon of 2006:01-2012:12. The sample consisted of the daily yields for the maturities of 1, 3, 4, 6, 9, 12, 15, 18, 24, 27, 30, 36, 42 and 48 months.
Figure 2: Cumulative squared forecast errors (1-week forecast horizon)

Note: Figures (a) and (b) show the cumulative squared forecast errors (CSFE), relative to the random walk, of individual yield-only models in Panel (a) and of forecast combinations schemes in Panel (b). Figures shows CSFEs for a 3-month forecast horizon. The evaluation sample is 2009:1 to 2012:12 (988 out-of-sample forecasts). Grey bars highlight recession periods.

(a) Individual models

(b) Forecast combinations
Figure 3: Cumulative squared forecast errors (1-month forecast horizon)

Note: Figures (a) and (b) show the cumulative squared forecast error (CSFE), relative to the random walk, of individual yield-only models in Panel (a) and of forecast combinations schemes in Panel (b). Figures shows CSFEs for a 3-month forecast horizon. The evaluation sample is 2009:1 to 2012:12 (988 out-of-sample forecasts). Grey bars highlight recession periods.

(a) Individual models

(b) Forecast combinations
Figure 4: Cumulative squared forecast errors (2-month forecast horizon)

Note: Figures (a) and (b) show the cumulative squared forecast error (CSFE), relative to the random walk, of individual yield-only models in Panel (a) and of forecast combinations schemes in Panel (b). Figures shows CSFEs for a 3-month forecast horizon. The evaluation sample is 2009:1 to 2012:12 (988 out-of-sample forecasts). Grey bars highlight recession periods.

(a) Individual models

(b) Forecast combinations
Figure 5: Cumulative squared forecast errors (3-month forecast horizon)

Note: Figures (a) and (b) show the cumulative squared forecast error (CSFE), relative to the random walk, of individual yield-only models in Panel (a) and of forecast combinations schemes in Panel (b). Figures shows CSFEs for a 3-month forecast horizon. The evaluation sample is 2009:1 to 2012:12 (988 out-of-sample forecasts). Grey bars highlight recession periods.

(a) Individual models

(b) Forecast combinations
Figure 6: Relation between RMSFE and Sharpe ratios

Note: This figure presents scatter plots of out-of-sample economic performance measures against RMSFE (root mean squared forecast error). The evaluation sample is 2009:1 to 2012:12 (988 out-of-sample forecasts).